



Projektbericht GeoHarvester Weiterentwicklung

1. Projektziele

Zusammenfassung: Geodienste zu finden ist nicht so effizient wie eine Websuche und benötigt Kenntnisse, wie die Nationale Geodaten Infrastruktur (NGDI) der Schweiz aufgebaut ist. Dies, da Geodienste nicht von Suchmaschinen indiziert werden und die Server der Geodienste bei den Geodatenanbietenden (Bund, Kantone, Gemeinden und Dritte) verteilt betrieben werden. Daher fehlt ein zentraler und einfach zu nutzende Einstiegspunkt für die Suche nach Geodiensten.

Der GeoHarvester PoC ist eine einfach zu nutzende Suchmaschine mit einer offenen API für die Suche von Geodiensten. Der GeoHarvester durchsucht die Metadaten der bekannten Geodienste und indiziert die einzelnen Geodienste der Geodatenprovider der Schweiz. Er strukturiert diese anhand der im Dienst aufgeführten Metadaten und reichert diese mit einem Metadatenqualitätsindex und weiteren Attributen zur Unterstützung des Rankings bei Suchanfragen an. Die Suchresultate werden auf Basis eines Rankings entsprechend den Metadateninhalten und der Metadatenqualität der Dienste ausgegeben. Die Suche ist in einer schlichten Oberfläche gehalten und die Ergebnisse werden den Nutzenden strukturiert in einer Tabelle mit zusätzlich weiterführenden Links zu Darstellung in map.geo.admin.ch, WMS Server URL sowie Layerdefinitionsdateien für die direkte Einbindung in QGIS und ArcGIS zur Verfügung gestellt.

Im Rahmen der Weiterentwicklung der GeoHarvesters wurde das Webdesign für mobile Endgeräte optimiert, Mehrsprachigkeit in den Suchresultaten integriert, die Suche über sprachgestützte Vorprozessierung verbessert, sowie Wartung und Dissemination des Projekts realisiert.

Summary: Finding geospatial web services is not as efficient as an online web search and requires knowledge on how the National Geodata Infrastructure (NGDI) is structured. This is because geospatial services are not indexed by search engines, and the servers of the geospatial services are operated in a distributed manner by the geodata providers (federal government, cantons, municipalities and third parties). Therefore, there is no central and easy-to-use entry point of access for searching OGC Web Services.

The GeoHarvester PoC is an easy-to-use search engine with an open API for searching OGC geospatial services. The GeoHarvester searches or harvests the metadata of the known geospatial services (WMS, WMTS and WFS) and creates an index of these services by the geodata providers in Switzerland. It structures this based on the metadata listed in the service and enriches it with a metadata quality index and other attributes to support the ranking in search queries. The search results are output based on a ranking according to the metadata content and the metadata quality of the services. The search has a user friendly interface, and the results are provided structured as a table including further links to map.geo.admin.ch, WMS server URL and layer definition files for direct integration into QGIS and ArcGIS. Future steps for the GeoHarvester include adapting the harvester for mobile devices, supporting multilingualism, and optimizing search results.

As part of the further development of Geoharvesters, the web design was optimised for mobile devices, multilingualism was integrated into the search results, the search was

improved via language-supported pre-processing, and maintenance and dissemination of the project were implemented.

2. Arbeitspakete

1.1. AP 1 – Optimierung für responsives Webdesign

Für den mobilen “UseCase” des GeoHarvester-Portals wurde eine responsive UI implementiert, welche Bildschirmgrößen unter 600 Pixel Breite unterstützt. Dazu werden einige der zentralen Bedienelemente neu angeordnet sowie weniger wichtige Funktionalitäten, z.B. der Datenexport in Desktop GIS, ausgeblendet, um Platz zu schaffen. Als vorbereitende Arbeit für AP2 wurde zudem das Portal in drei weitere Sprachen (Englisch, Französisch, Italienisch) übersetzt. Die Anzeigesprache wird anhand der Browsereinstellungen ermittelt, alternativ kann die Sprache der GUI auch im Menü umgestellt werden. Die ausgewählte Sprache wird als Parameter für API Requests verwendet und an den Server übermittelt.

Implementation siehe <https://github.com/FHNW-IVGI/Geoharvester/pull/85> und <https://github.com/FHNW-IVGI/Geoharvester/pull/90>

1.2. AP 2 – Mehrsprachigkeit in den Suchresultaten

Die Benutzerfreundlichkeit wurde durch die Integration der Mehrsprachenunterstützung sowohl im Front-End als auch im Back-End verbessert. Dazu wurden Funktionalitäten für die Vorprozessierung der Metadaten entwickelt, welche relevante Metadaten (Titel, Beschreibung, Keywords und NLP-extrahierte Keywords) in vier Sprachen (Deutsch, Französisch, Italienisch, Englisch) übersetzen und in die Suchdatenbank integrieren. Ebenfalls wurden die vier Sprachen in das Frontend integriert. Dies ermöglicht eine sprachunabhängige Suche mit dem GeoHarvester, so dass die BenutzerInnen die Sprache verwenden können, mit der sie am besten vertraut sind.

In diesen Zusammenhang wurde das Thema für ein Vertiefungsprojekt im Masterstudiengang entwickelt, um die Fähigkeiten von Large Language Model für die Mehrsprachenunterstützung zu untersuchen. Dieses Vertiefungsprojekt ist noch in Bearbeitung und wird Anfangs 2025 fertiggestellt:

A Large Language Model LLM goes GeoHarvester, Ken Kaufmann, Vertiefungsprojekt MSE Master of Science in Engineering, Profile Geomatics, Herbstsemester 2024, in Bearbeitung

Implementation siehe <https://github.com/FHNW-IVGI/Geoharvester/pull/94> und <https://github.com/FHNW-IVGI/Geoharvester/pull/96>.

1.3. AP 3 – Verbesserung der Suchalgorithmen, User Testing und Entwicklung weiterer Funktionalitäten

Die Mehrsprachunterstützung in AP2 stellte zusätzliche Herausforderungen bei der Suche nach Begriffen, die aus mehreren Wörtern bestehen. Da die Suchergebnisse in verschiedenen Sprachen identisch und in der gleichen Reihenfolge dargestellt werden müssen, wurden in

diesem Arbeitspacket die Such- und Rankingalgorithmen durch den Einsatz von Knowledge Graphs und Sprachmodellen erweitert und optimiert. Die Schlüsselbegriffe wurden mithilfe von Large Language Models (LLM), insbesondere Llama 3.1, extrahiert, übersetzt und abschliessend manuell überprüft.

Diese Daten bilden die Grundlage für die Erstellung eines Knowledge Graphs (KG), der mithilfe von Knoten und Kanten die Beziehungen zwischen Übersetzungen und Synonymen abbildet. Der KG wurde in der Suchfunktion integriert und dient als Katalysator für bekannte Begriffe. Außerdem erleichtert er das Ranking der Ergebnisse, indem er bekannten Begriffen mehr Gewicht verleiht. Die Struktur wurde so gestaltet, dass die Begriffe im KG jederzeit erweitert oder angepasst werden können.

Implementation siehe <https://github.com/FHNW-IVGI/GeoHarvester/pull/121>.

1.4. AP 4 – Wartung und Dissemination des GeoHarvester

Das Entwicklungsteam sicherte den Betrieb und die Wartung des Systems und führte mehrere Optimierungen der Prozessierungspipeline durch, welche den neu integrierten Präprozessierungsschritte für die Integration der mehrsprachigen Suche Rechnung tragen. Der GeoHarvester Poc wurde an mehreren Schweizer GIS-Events wie GeoSummit, KGK oder der GeoPython Konferenz vorgestellt und die Erkenntnisse der Entwicklung in einem internationalen Journal publiziert.

Publikationen und Präsentationen

Paper

Ferrari, E., Striewski, F., Tiefenbacher, F., Bereuter, P., Oesch, D., & Di Donato, P. (2024). *Search Engine for Open Geospatial Consortium Web Services Improving Discoverability through Natural Language Processing-Based Processing and Ranking*. ISPRS International Journal of Geo-Information, 13(4), Article 4. <https://doi.org/10.3390/ijgi13040128>

Präsentationen

GeoHarvester - Revolutionizing Geospatial Web Service Discovery, Elia Ferrari, GeoPython 2024, 28.05.2024, Muttenz

GeoHarvester.ch, Ein Suchmaschinen-Prototyp für Schweizer Geodienste. Pia Bereuter, Friedrich Striewski, 02.09.2024, KGK Olten

GeoHarvester.ch, Ein Suchmaschinen-Prototyp für Schweizer Geodienste. Pia Bereuter, Friedrich Striewski, Elia Ferrari, 10.10.2024, Webinar Geosummit, Online

KI in der Geoinformation, Pia Bereuter, 21.11.2024, GIS-Verbund Thurgau, Frauenfeld

3. Ausblick

Das Auffinden von Geodiensten der NDGI ist ein wichtiges Fundament der Nationalen Geodaten Infrastruktur der Schweiz. Umso mehr, wenn das benötigte Fachwissen zur Struktur und Aufbau der Infrastruktur fehlt, bietet ein single point of entry, respektive ein Katalog der öffentlich verfügbaren Geodienste mit der Sichtbarkeit einen zusätzlichen Mehrwert. Mit einem

öffentliche verfügbaren Repository ist der Geodienst Harvester langfristig verfügbar und für Weiterentwicklungen offen.

Weitere Schritte für die Weiterentwicklung des GeoHarvester PoC sind die (semi)automatisierte Wartung und Prüfung des Index der Endpoints der Geodienste, die Optimierung der Suche in mehreren Sprachen, Optimierung und Erweiterung des Knowledge Graphen, Einbindung weiterer Datendienste oder eine interaktive, explorative Abbildung der Service/Datenlandschaft der Schweiz über den Knowledge Graphen.