

Geoharvester.ch

**Ein Suchmaschinen-Prototyp für
Schweizer Geodienste.**

**Un prototype de moteur de
recherche pour les géoservices
suisses.**

Prof. Pia Bereuter, Friedrich Striewski
Institut Geomatik
Fachhochschule Nordwestschweiz





Problemdefinition Définition du problème

- Es fehlt ein zentraler Zugangspunkt für Schweizer Geoservices.
Il manque un point d'accès central pour les géoservices suisses.
- Die Qualität von Metadaten und der Umfang von Dokumentation variiert sehr stark.
La qualité des métadonnées et le volume de la documentation est très variable.
- Die Indexierung von Services und ihre Aktualisierung ist zeitaufwändig
Indexer des services et leur mise à jour prend du temps



Projektziele

Objectifs du projet

- Entwicklung eines zentralen **Zugangspunktes**, um Geoservices zu finden (Portal und API)
Développement d'un point **d'accès central** pour trouver des géoservices (portail et API).
- Kontinuierliche, **automatisierte Updates** (Scraper und Daten-Präprozessierung)
Mises à jour automatiques et régulières (scraper et prétraitement des données)
- **NLP-Methodik** und «**fuzzy search**» um Datensätze zu finden
Méthodologie NLP et «**fuzzy search**» pour trouver des jeux de données.
- **Filterung** und **Sortierung**, um Ergebnisse auf relevante Datensätze einzugrenzen.
Filtrage et **classement** pour limiter aux résultats pertinents
- Bewertung von Metadaten mithilfe eines **Scores**
Évaluation des métadonnées par un **score**.

🔍 Webservices suchen... ✕ **SUCHEN**

FastAPI 0.2.0 OAS 3.1

</api/openapi.json>

default

GET /api Root

GET /api/getData Get Data

Schemas

EnumLangType > Expand all string

EnumProviderType > Expand all string

EnumServiceType > Expand all string

GeoserviceModel > Expand all object

HTTPValidationError > Expand all object

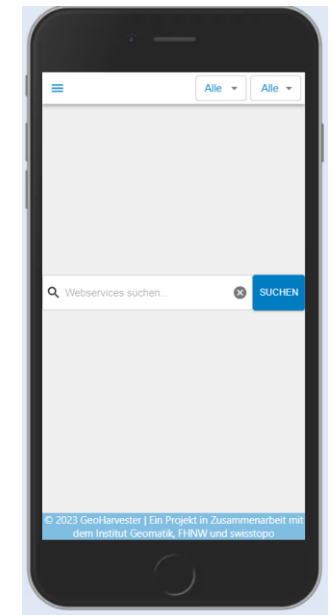
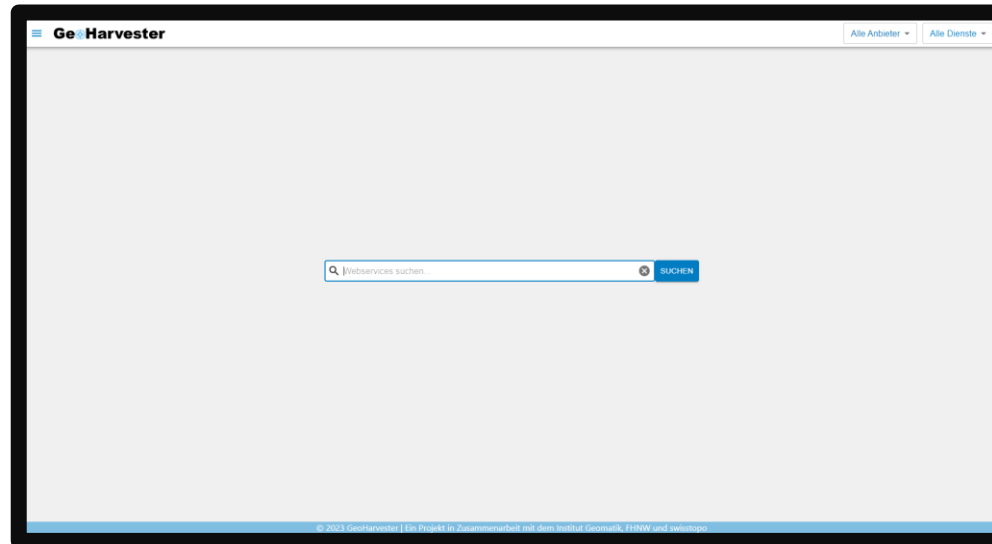
PageCustomized[GeoserviceModel] > Expand all object

ValidationError > Expand all object

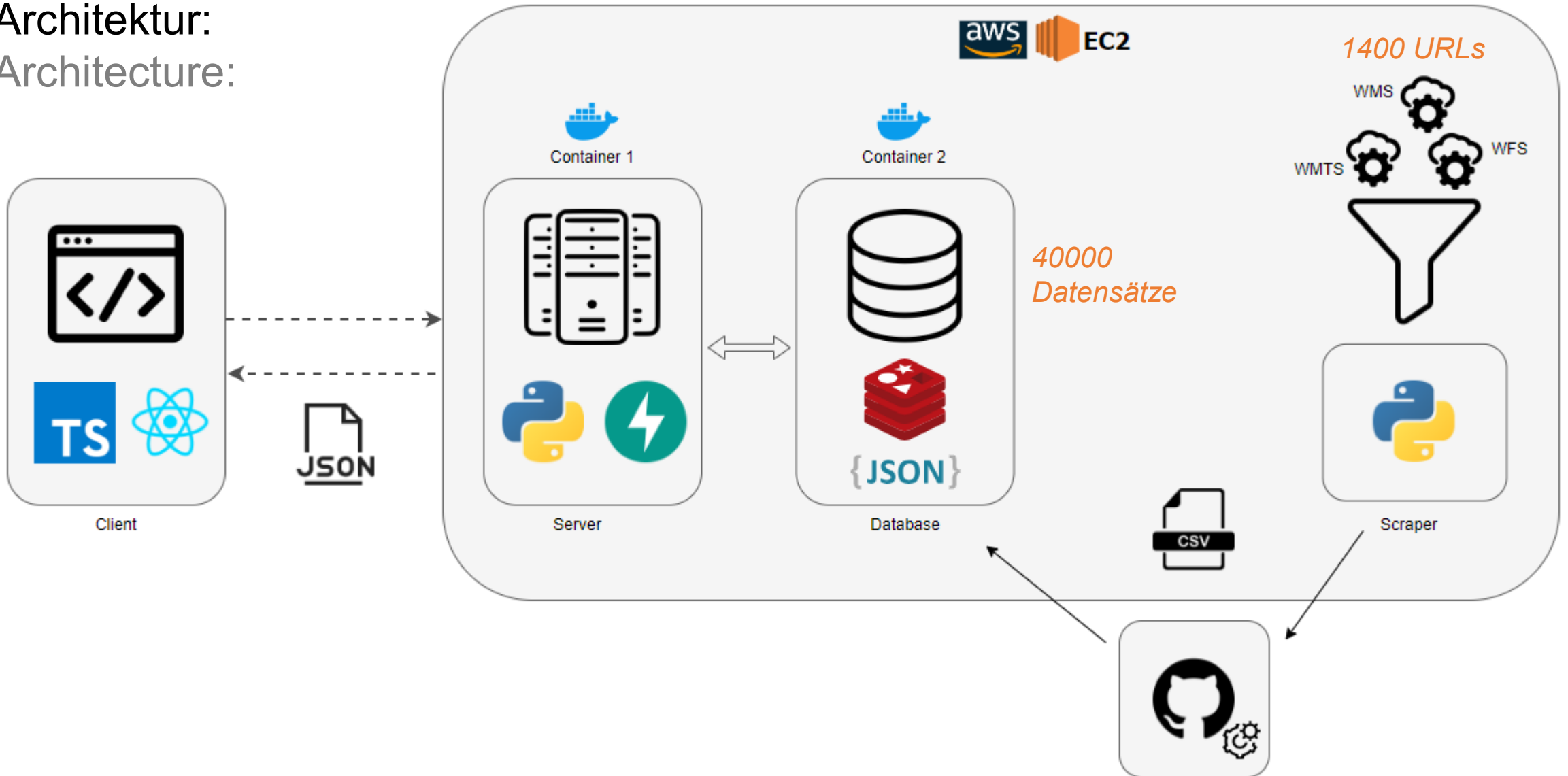
Geoharvester ist erreichbar unter:
Geoharvester est disponible à:



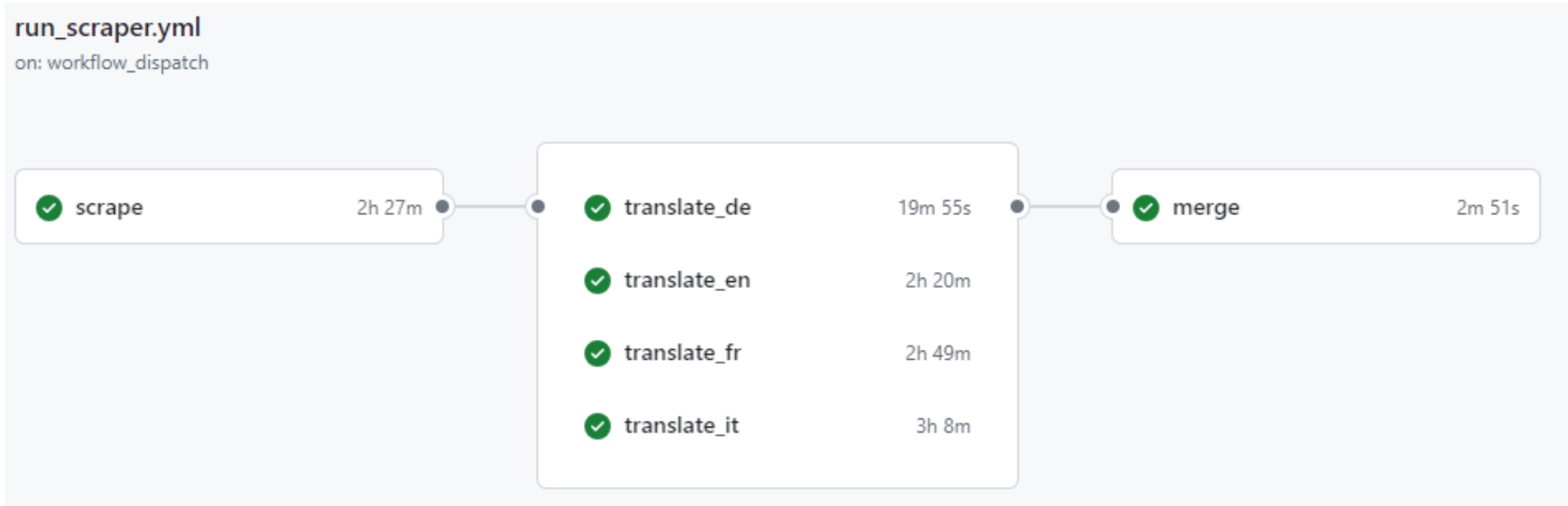
www.geoharvester.ch



Architektur: Architecture:



Präprozessierungspipeline / Pipeline de pré-processus



Scraping

Keyword-Extraktion

Übersetzung

Zusammenführen neuer / alter Daten

Säubern

Metadaten Score

Suche / Recherche

- Mehrere Suchbegriffe:
- Suchbegriff/e, Service-Typ, Kanton:
- Stopwörter werden entfernt
- Suchbegriffe werden gestemmt
- Suchbegriffe werden mit Datenbank abgeglichen

- Plusieurs mots de recherche :
- Mot(s) recherché(s), type de service, canton :
- Les mots d'arrêt sont supprimés
- Les mots de recherche sont balisés
- Les mots de recherche sont comparés
avec la base de données

ODER-Logik

UND-Logik

(z.B. Artikel, Konjunktionen, Präpositionen)

(z.B. «*Wörter*» → «*Wort*», «*amtlich*» → «*amt*»)

(Titel, Zusammenfassung, Schlüsselwörter)

logique **OU**

logique **ET**

(p.ex. articles, conjonctions, prépositions)

(p.ex. «mots» → «mot», «officiel» → «office»)

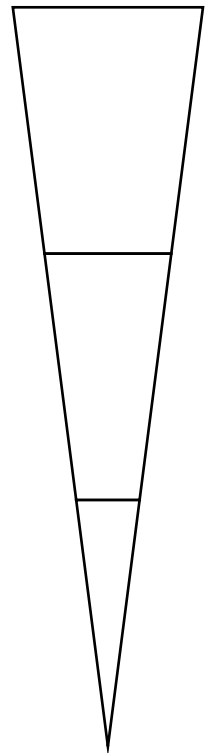
(titre, résumé, mots clés).

Ranking der Ergebnisse / Ranking des résultats

Suchbegriff/e hat: / le(s) mot(s) recherché(s) :

- exaktes Äquivalent in **Titel** oder **Schlüsselwörtern**
équivalent exact dans le **titre** ou les **mots-clés**
- ungefähres Äquivalent in **Titel** oder **Schlüsselwörtern**
équivalent approximatif dans le **titre** ou les **mots-clés**
- exaktes Äquivalent in **extrahierten** Schlüsselwörtern (aus Zusammenfassung)
équivalent exact dans les mots-clés **extraits** (du résumé)
- ungefähres Äquivalent in **extrahierten** Schlüsselwörtern (aus Zusammenfassung)
équivalent approximatif dans les mots-clés **extraits** (du résumé)
- Exaktes Äquivalent in Titel und *extrahierten* Schlüsselwörtern – **ohne Übersetzung**
Équivalent exact dans le titre et les mots-clés extraits - **sans traduction**

Gewichtung
pondération



Metadatenqualität / Qualité des métadonnées

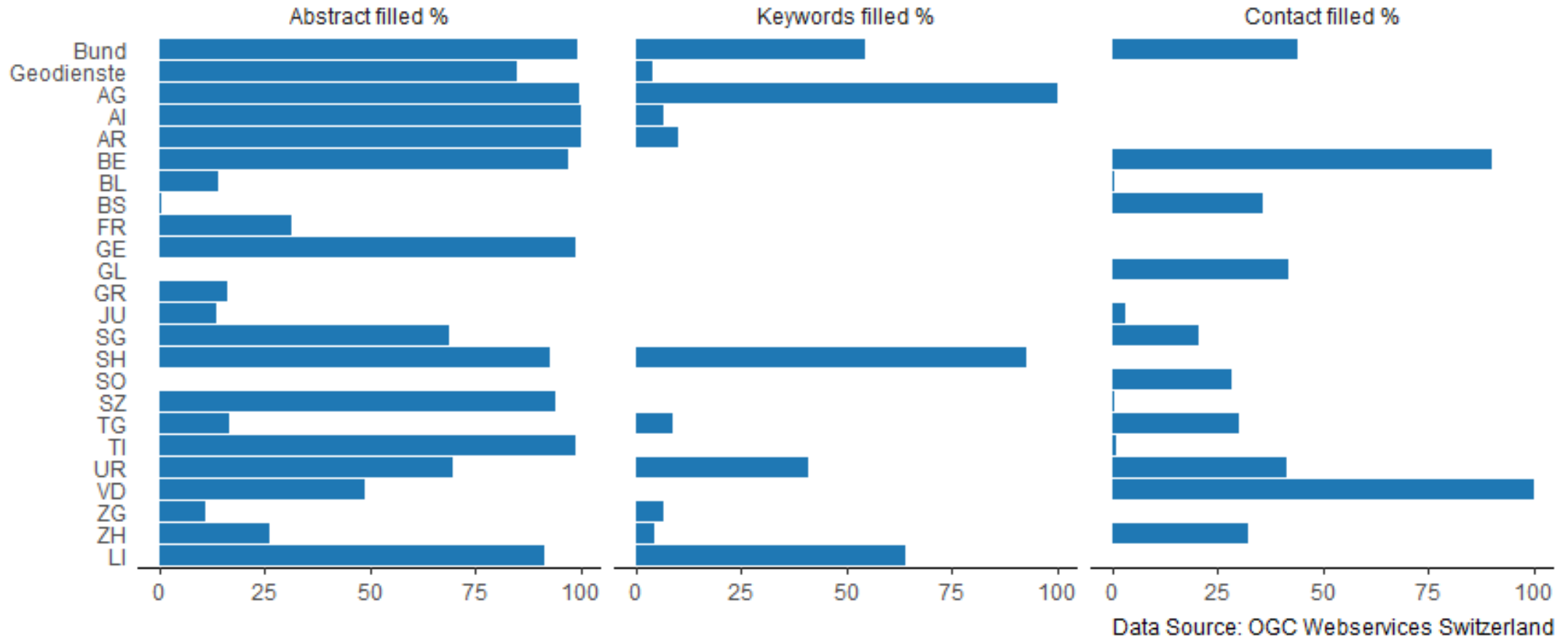
GeHarvester

Titel	Auszug	Anbieter	Dienst	Metaqualität
unklarer Titels 😬 titre peu clair	ungen (auch Ver schnitt in einem b h betrieben werd Abstract fast exakte Kopie des Titels 😬 L'abstract est la copie presque exacte du titre		WMS	100
	Kein Abstract 😬 Pas d'abstract			
		Fehlende Kontaktangaben 😬 Absence de coordonnées		
		Kontaktangaben als generischer URL 😬 Coordonnées sous forme d'URL générique		
		Fehlende Keywords 😬 Mots-clés manquants		
		Keine Angaben zu Metadatenportal 😬 Aucune information sur le portail de métadonnées		

Name: sh.verkehr.individual.
Contact:
Abstract: Verkehrsmessungen (auch Verkehrszählungen) d
können dauerhaft oder periodisch betrieben werd
unterschieden werden. Die ermittelten Daten sind Grundlage für
Tree:
Group:
Keywords: Mots clés en tant que lien
Metadata: https://www.geocat.ch/geonetwor
Endpoint: https://wms.geo.sh.ch/wms
Mapgeo: SERVICE IN MAPGEO ÖFFNEN
Legend: LEGENDE ÖFFNEN

Metadaten

Verwendung von Metadatenfeldern im OWS

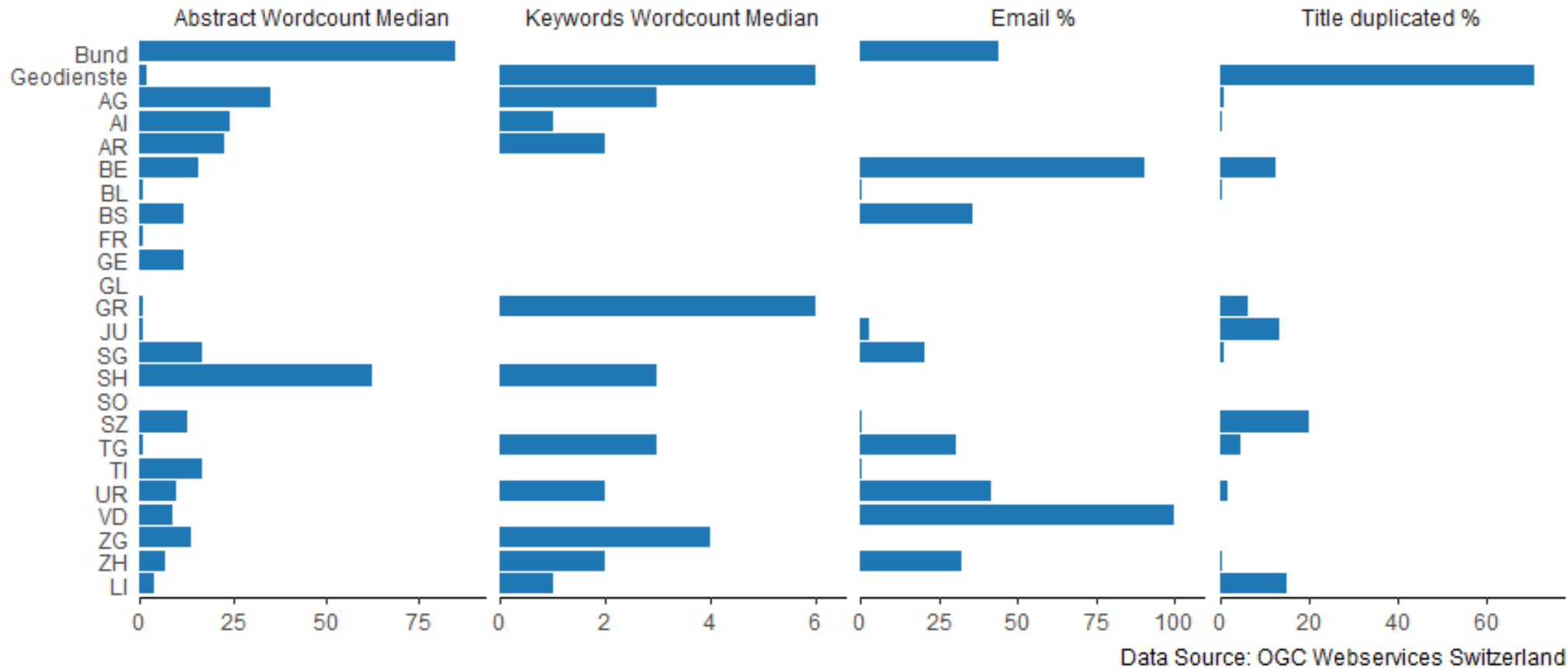


Metadatenvollständigkeit für Abstract, Contact und Keywords

Métadonnées complètes pour l'abstract, le contact et les mots-clés

Metadaten

Verwendung der Metadatenfeldern in OWS



Qualität der Metadatenbeschreibung mit dem Median der Anzahl Wörter für Abstrakt und Keywords, Anteil Keywords mit URL oder Ressourcen Angabe, sowie ob eine Emailadresse im Feld Kontakte aufgeführt wird.

Qualité de la description des métadonnées avec la médiane du nombre de mots pour l'abstrait et les mots-clés, la proportion de mots-clés avec indication d'URL ou de ressources, ainsi que si une adresse e-mail est mentionnée dans le champ Contacts.

Team

FHNW

- Prof. Pia Bereuter (*Projektleitung*)
- Fiona Tiefenbacher (*UI/UX*)
- Elia Ferrari (*NLP*)
- Friedrich Striewski (*Software-Entwicklung*)

Swisstopo

- David Oesch (*Idee, Projektkoordination*)
- Pasquale Di Donato (*Projektkoordination*)

Umfrage zur Nutzerfreundlichkeit GeoHarvesters.

Wir freuen uns auf Eure Inputs und Ideen

Enquête sur la fonctionnalité de GeoHarvester.

Nous nous remercions de vos contributions et
de vos idées.

