Open WebSearch

UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

OpenWebSearch.EU – Finding, Analysing and Utilizing Web Data

Prof. Dr. Michael Granitzer
Chair of Data Science University of Passau and
Coordinator OpenWebSearch.eu

Open Web Search

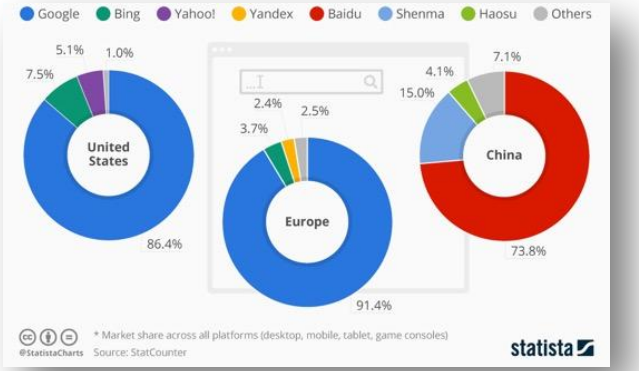
Funded by the European Union

SUPPORTED BY NGI

Todays Situation in Web Search: A Critical Infrastructure managed as Oligopoly

Two properties of Web Search that don't fit

- A critical infrastructure for society, comparable to satellite navigation
- A market oligopoly: i.e. "a market structure in which a market or industry is dominated by a small number of large sellers or producers." (Wikipedia)



Effects

- Digital Sovereignty
- Reduced User Experience (limited choice, locked-in ..)
- Limited Innovation Potential
- Web-data as driver for AI Innovation beyond Web Search

Challenges

- High infrastructure costs
- Broad range of technical skills required
- (Growing) Legal uncertainties

□ Especially small innovators and researchers are left behind

Our approach: Building an Open Index of the Web and a corresponding open infrastructure

What?

Restore an open search ecosystem as a basis for a new Internet Search

→ empower Europe's researchers, innovators and businesses to systematically tap into the Web as business and innovation resource

→ lay a foundation for a new Internet search

→ contribute to Europe's digital sovereignty

→ Support Web-data analytics and AI usage

How?

Four Objectives

1. Open Technology Stack
2. Resource provision by a network of infrastructure providers
3. Added value services
4. Bootstrapping the ecosystem plus third party calls

14 Partners plus Third Party Calls

Open WebSearch .eu

UNIVERSITÄT PASSAU

Webis.de

UNIVERSITÄT LEIPZIG

Bauhaus-Universität Weimar

Radboud Universiteit

DLR

TU Graz

Research

lrz

CERN

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

CSC ICT Solutions for Brilliant Minds

Infrastructure

open search foundation

nlnet FOUNDATION

suma-ev

NGOs

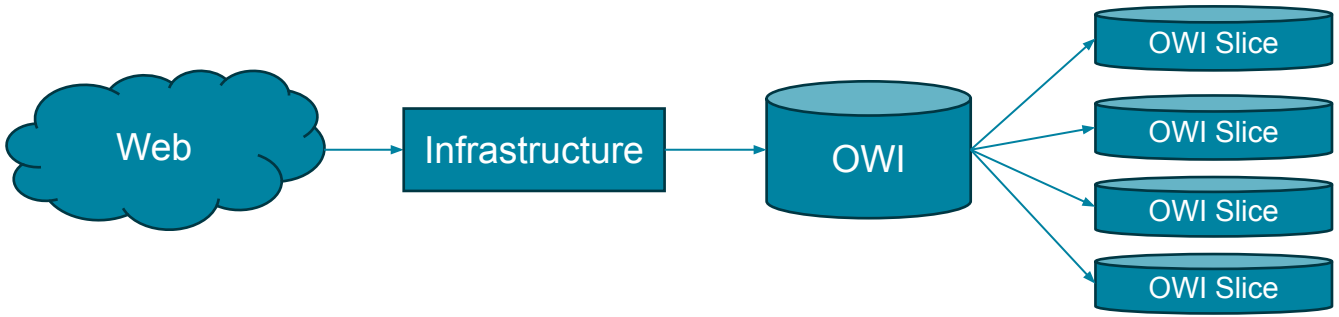
A1

Businesses

The Open Web Index (OWI)

A Web Index a data structure for fast query-based access and ranking of web documents – the core of every web search engine

Our Proposition: A collaboratively created, open and transparent Web Index for empowering smaller innovators and creating an ecosystem for Web search, Web-data Analytics and AIs



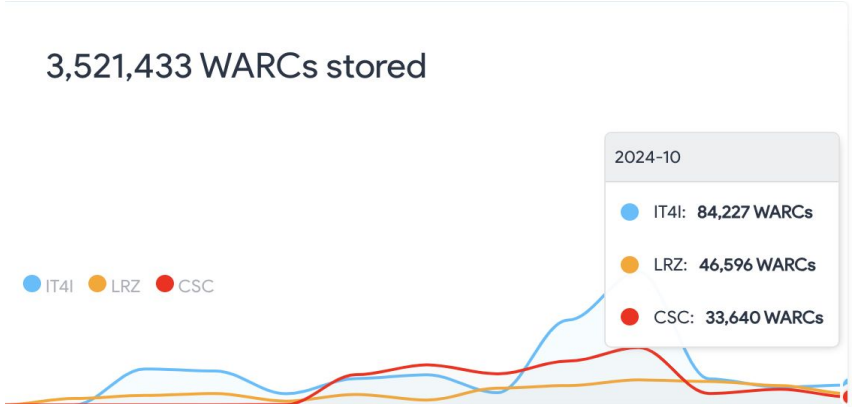
Temporal, topical slices of the index for web search, analytics and AI

Granitzer, Michael, et al. "Impact and development of an Open Web Index for open web search." *Journal of the Association for Information Science and Technology* (2023).

Status

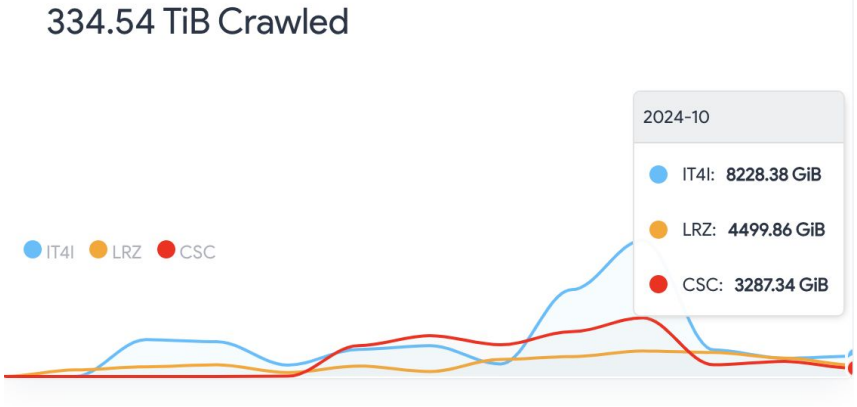
Statistics (since 08/24)

- 10 Billion Websites from ~ 10 Million Domains
- 185 Languages
- 334 TiB with 1.5 TiB / day
- Filter for malicious and adult content



Goals

- Scaling up to 10 TiB / day and getting to ~50% of the Web
- Scaling crawled and indexed data to the Petabyte scale
- Legal compliance and Coordinating Crawls



Example Open Geo-Science Search @DLR Prototype

Open WebSearch .eu

Up-to-date information

News articles

Scientific information

Publications & literature

Geo-spatial information

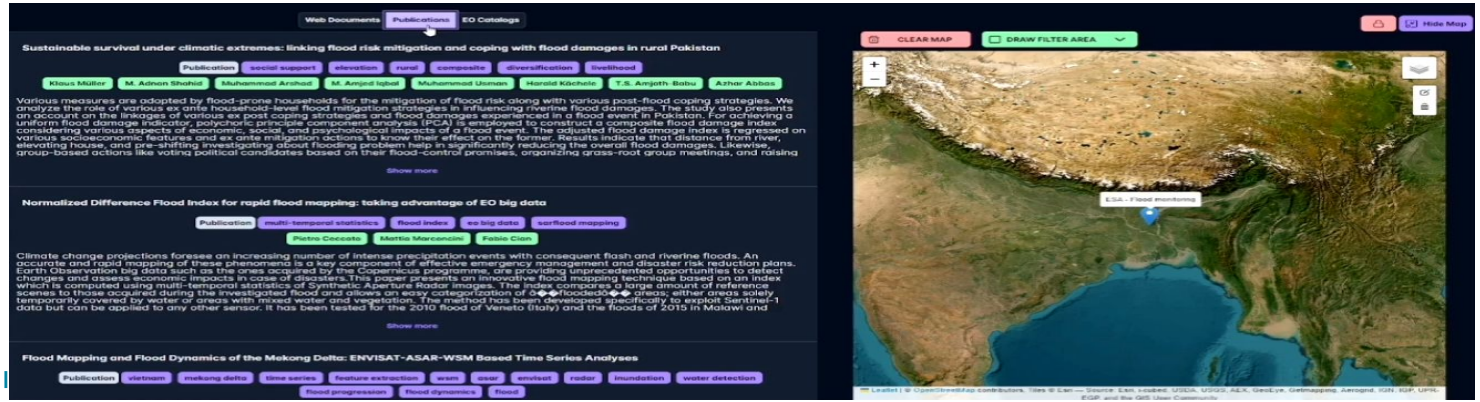
Earth observation catalogs

Three types of information sources

OpenWebSearch.eu Search service

DLR digital library of publication data

Planetary Computerm, EOC Geoservice, Terrabyte STAC



Prototype search application

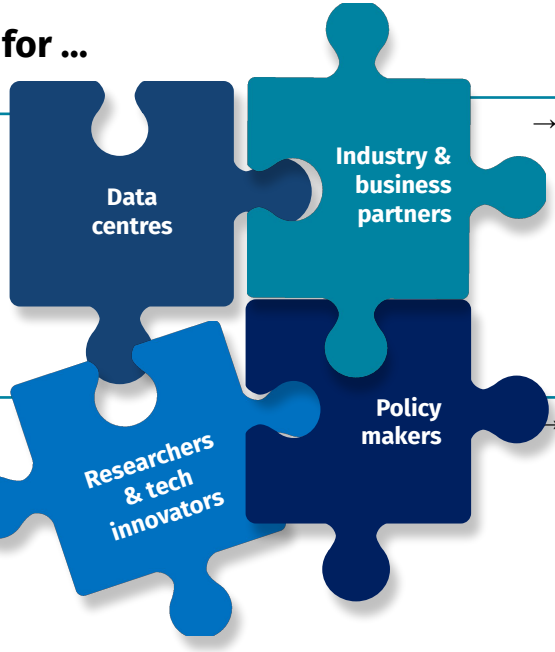
Thanks. Questions?

Open WebSearch .eu

Contact us:

To keep in touch with these possibilities or to join us send an email to **join@openwebsearch.org**

We are looking for ...



→ help hosting a distributed Open Web Index

→ discover the business models of an Open Web Index

→ develop new search & retrieval paradigms and content analysis algorithms

→ help shaping the governance of an open search ecosystem

→ We will also offer small grants for potential contributors.

Meet us at our Open Search Symposium at LRZ October 2024