

Artikelindex

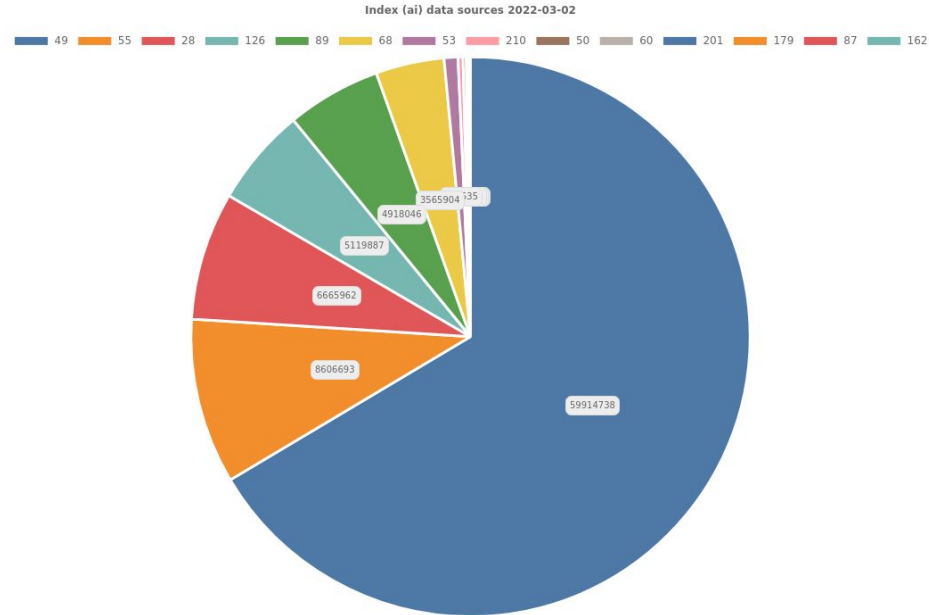
MMM UB Leipzig 2022-03-16 Martin Czygan

Artikelindex: Service

- Teil des Index-Angebots der UB Leipzig
- Technische Beiträge von: David Aumüller, Evelyn Weiser, Frank Morgner, Martin Czygan, Robert Schenk, Ronny Gey, Stefanie Nooke, Thomas Gängler, Thomas Gersch, Tracy Hoffmann, sowie: Andreas Czerniak, David Glueck, Gunnar Þór Magnússon, Justin Kelly, Natanael Arndt, und weitere.
- Open Source Entwicklung seit 2014-06-23, s.a. <https://github.com/ubleipzig>
- als Service im Betrieb seit Q2/2015, (Primo Central “Substitut”)

Artikelindex: Zahlen

- 90,157,793 Artikelmetadaten (151,228,082), 14 Datenquellen (offene, liz., ...)



Artikelindex: Projekte / Ansigelung

- aktuell 32 Bibliotheken und Projekte
- alle Projekte teilen sich einen Index, individualisierbar durch “Ansigelung”
- Einrichtungen konfigurieren Datenquellen (Listen, KBART-Dateien, ...)
 - EZB → AMSL oder FOLIO → Index
- Deduplizierung über DOI per Einrichtung
- Lizenzierung dynamisch, Deduplizierung nötig, daher komplette “Berechnung” der Sichten / Ansigelung per Index-Update

Artikelindex: Tools

- Viele, kleinere Prozessierungsschritte: Datenakquise, Normalisierung, Ansigelung, Deduplizierung, Indexierung
- Datenakquise: API-Abfragen, FTP-Kopien, OAI-PMH Abfragen, ...
 - Tools: metha, srufetch, solrdump, esdump, ...
- Normalisierung:
 - intermediate schema
- Indexierung:
 - solrbulk, esbulk
- Nachnutzung einiger dieser Werkzeuge außerhalb der UB (z.B. BASE Bielefeld, Internet Archive, ...)

Artikelindex: Fin

- Alle Entwicklungen als Open Source (Sichtbarkeit, Kollaboration, Nachnutzung, ...)
- Regelmäßige Updates und Anpassungen
- Umzug auf neue, geteilte Hardware, später: geteilter (solrcloud) Index mit SLUB
- Sonstiges:
 - Offene Kataloge, Daten und Suchmaschinen / <https://fatcat.wiki>, <https://scholar.archive.org> (32M+ Volltexte)
 - Zitationsdaten, Open Citations, MAG/OpenAlex, Refcat / <https://github.com/slub/labe>, <https://opencitations.net>, <https://arxiv.org/abs/2110.06595>