Aggregierende Dienste

Suchmaschinen

BASE - Bielefelder Academic Search Engine

CORE

Google Scholar

Dokumente werden in Google Scholar (meistens) gefunden - auch Google hat vieles indexiert.
Ein Problem ist, dass nicht die Links von einem Repositorium in Google Scholar gefunden werden, sondern der vorher woanders veröffentlichte Volltext.
Google merkt bei der Indexierung, wenn ein Dokument bereits woanders veröffentlicht wurde und indexiert wurde und schließt dann Seiten vom Index aus.

Google - Was kann ich tun, damit ich besser von Google indexiert werde?

BING

(Anwendertreffen 11/2020) **Vortragsfolien**

Frage:

In welchen Abständen aktualisiert BASE die bereits erfassten Dokumente?

Manchmal können sich ja nachträglich Metadatenfelder (Autoren etc.) ändern oder die Zugangsinfos (Upload von PDFs).

Antwort:
siehe FAQs in BASE: https://www.base-search.net/about/de/faq.php#chap0

Fragen zur Aktualisierung / Löschung von Inhalten: ...
Ich habe Änderungen an Dokumenten aus meiner Quelle vorgenommen. Warum werden die Änderungen in BASE nicht aktualisiert?
Jede nachträgliche Änderung eines Datensatzes, muss in Ihrer OAI-Schnittstelle gekennzeichnet werden, indem der "datestamp" des Dokuments aktualisiert wird.
Wird der "datestamp" nicht aktualisiert, ist eine automatische Aktualisierung im BASE-Index nicht möglich und das Dokument verbleibt unverändert und damit fehlerhaft im Index.
Eine Aktualisierung ist dann nur möglich, wenn die Inhalte des Datenlieferanten vollständig neu indexiert werden, was einige Zeit in Anspruch nimmt.

Antwort von BASE auf Anfrage des BSZ: https://base-search.net/about/de/contact.php

...OAI-PMH liefert Aktualisierungen bei Korrekturen an Metadatensätzen via inkrementeller Bereitstellung. Normalerweise funktioniert das auch.

Wir harvesten standardmässig via cronjob einmal pro Woche und ab Montag der Folgewoche sollten die Informationen im Index bereitstehen und suchbar sein.

Beim inkrementellen Harvesten gibt es aber immer mal Asynchronitäten, was nach meiner Beobachtung abhängig von der eingesetzten Software ist. Bei deutschen Quellen ist die Situation aber eher entspannt.

In diesem Kontext refreshen (also Neu-Holen, was solche Phaenomene bereinigen sollte) wir circa im Jahresabstand die Quellen....