

Faherwerk zur ersten Einführung in Berlin 1886.

*Leopold Dornschütz, Maschinenbau-Ingenieur
Frankfurt*

Ranking in VuFind

Einführung

Julia Beck | FID Darstellende Kunst
j.beck@ub.uni-frankfurt.de

2023-09-20

Ranking in VuFind

Disclaimer



- ▶ ich bin keine Expertin in Relevance-Ranking...
- ▶ ...aber lange genug in der Community, um zu wissen, wohin man sich wenden kann
- ▶ dient als Materialsammlung
- ▶ Hilfe zur Selbsthilfe



Typische Fragen:

- ▶ Warum kommt bei meiner Suche Dokument X in der Trefferliste an Stelle Y? Warum ist Dokument X nicht höher gerankt?
- ▶ Was passiert eigentlich beim Relevance Ranking?
- ▶ Wie kann ich das Ranking ändern/optimieren? (z.B. ein bestimmtes Format oder aktuelle Records höher ranken)

VuFind nutzt die Fähigkeit von Solr, durch **Relevance Ranking** bei einer Suche bestimmte Dokumente regelbasiert zu ranken.

Welche Records in der Trefferliste am höchsten gerankt werden, sollte anhand der eigenen **Zielgruppe** bestimmt werden.



Wo finde ich Hilfe?:

- ▶ [VuFind Doku](#) | [VuFind Github](#) | [VuFind Mailing Liste](#) | [VuFind Slack](#)
- ▶ [VuFind Book - Chapter Understanding Solr](#)
- ▶ [VuFind Doku - Search Customization](#)
- ▶ [Video Tutorial zur Suche in VuFind](#)
- ▶ [Solr Dokumentation / Tutorials](#)
- ▶ Vorträge auf Anwendertreffen (Achtung: evtl. unvollständige Liste, manche evtl. veraltet):
 - ▶ [Relevance Ranking Revisited](#) (BSZ, 2014)
 - ▶ [Grouping/Deduplizierung in Solr](#) (BSZ, 2018)
 - ▶ [Relevanzranking aus der Nähe](#) (TU Hamburg, 2017)
 - ▶ [Relevance Ranking und Query Type Detection](#) (SUB Hamburg, 2013)
- ▶ [Artikel von Mukhopadhyay](#), 2020



Wichtige **Konfigurationsdateien** in diesem Zusammenhang in VuFind:

- ▶ `$VUFIND_HOME/config/vufind/searches.ini` (Welche Suchoptionen bekommt der User?)
- ▶ `$VUFIND_HOME/config/vufind/searchspecs.yaml` (Wie verhält sich die Suche genau? Welche Felder werden durchsucht und wie werden sie gerankt?)

`searchspecs.yaml` beginnt mit Kommentarzeile, die die verschiedenen Konfigurationsmöglichkeiten erklärt.

Alle folgenden Sections sind unter ihren jeweiligen **Search Handlern** (Arten von Suchen, die durchgeführt werden können) subsummiert.

Sidenote: Eigene Konfigs immer in `$VUFIND_LOCAL_DIR` ablegen.
Achtung bei YAML-Dateien! Whitespace und Einrückungen beachten.



DisMax = **Disjunction Max**, für Menge von Feldern

Title:

DismaxFields:

- title_short^500
- title_full_unstemmed^450
- title^300
- title_new^100
- title_old
- series^100

DismaxHandler: edismax

- ▶ **DismaxFields** (Welche Felder sollen mit welcher relativen Gewichtung durchsucht werden? Solr kümmert sich um die Berechnung.)
- ▶ **DismaxHandler** (Query Mode: z.B. edismax = Extended Dismax)

Felder werden in ihrer Relevanz gegenübergestellt. In diesem Fall ist z.B. title_new 100x wichtiger als title_old.



Title:

DismaxFields:

- title_short^500
- title_full_unstemmed^450
- title^300
- title_new^100
- title_old
- series^100

DismaxHandler: edismax

DismaxParams:

- [bq, format:Book]

- ▶ **DismaxParams** (weitere Parameter: u.a. Boost Queries (bq) oder Phrase Fields (pf), um bestimmte Formate oder Kombinationen höher zu ranken; Minimum Matching (mm), Beeinflussung der Matching Toleranz)
- ▶ **CustomMunge** (Sonderfall: Datenmanipulation der Eingabe, bevor sie bei Solr ankommt, z.B. Bereinigung von führenden 0)

Die "richtige" Konfiguration **für die eigene Zielgruppe** zu finden, ist meist Trial & Error und ggf. ein Kompromiss (gut dokumentieren!)



Troubleshooting des Relevance Rankings:

Bisher v.a. in VuFinds Debug Mode (`debug = true` in `config.ini`).

Und in Solr: `debug-explainOther`

Neues Tool **Explain Addon** von HEBIS zur Visualisierung der Ranking-Berechnungen zur besseren Nachvollziehbarkeit des Rankings. ([Folien](#); Fragen an: p2_fragen@hebis.de; Code aktuell noch nicht veröffentlicht, [Pull Request](#), [Demo](#)):

HEBIS-Vortrag bei der WOLFCOn 2023:

[Please explain the rank of the title. Why is it found at all?](#)

(**Sidenote:** Die aktuelle Version basiert auf dem Scoring **TF-IDF** und (noch) nicht auf **BM25** (in neuer Solr Version genutzt) → [s.a.](#))

Erweitert



Allgemeine **Empfehlung** die [WOLFCon VuFind Playlist](#) anzusehen!

Interessante Vorträge u.a. zu den Themen

- ▶ Accessibility
- ▶ CMS-Funktionalitäten
- ▶ hierarchische Facetten
- ▶ Security in VuFind
- ▶ ... sowie ...
- ▶ "State of the Project"
- ▶ Roadmapping

