

# Clas-Digital

In den letzten Monaten haben wir *clas-digital* mit viel Energie (und auch Freude) überarbeitet, verbessert und hier und da den letzten (?) Feinschliff gegeben.

## **Zu den neuen Features gehören:**

### *clas-digital-server:*

- Anzeigen und Hochladen von auf mehreren Festplatten verteilten Büchern.
- Sichereres User-Interface
- Hochladen von PDFs
- Beheben von Bugs im Katalog
- Strikte Trennung von Darstellung, Suche und Daten.

### *clas-digital-suche:*

- Verbesserung der Suche-Geschwindigkeit
- Verbesserung der Such-Ergebnisse
- Preview zeigt nun Originaltext an (keine lower-case Konvertierung, keine Konvertierung von Umlauten)
- Verlässliche anzeige von Previews ("Preview not found" kommt nicht mehr vor)
- Previews in Metadaten zeigt an, in welchem Tag das Wort gefunden wurde.
- Suche in beliebigen Metadaten-Tags
- Verbessertes Ranking-System ([https://en.wikipedia.org/wiki/Okapi\\_BM25](https://en.wikipedia.org/wiki/Okapi_BM25))
- Verbesserte Vorschläge
- Such-Daten in separater Datenbank (Sqlite)

### *Allgemeine Verbesserungen:*

- Code refactoring
- Code documentation (unabgeschlossen)
- Code formatting ([https://en.wikipedia.org/wiki/Okapi\\_BM25](https://en.wikipedia.org/wiki/Okapi_BM25)) (unabgeschlossen)

### *Development Server:*

- Der development server läuft mittlerweile auf dem original Datensatz.
- Der Development server läuft als read-only, um die Daten auf jeden Fall zu schützen.
- (Dadurch ist allerdings das Hochladen von Büchern auf dem Developement server nicht testbar. Dies soll ink Zukunft noch verbessert werden)

### *Schritte um clas-digital erweiterbar und für Drittnutzer verfügbar zu machen:*

- Plugin-System
- Flexibles Metadaten-System für *clas-digital-server* und *-suche*.

## **Bewertungs- und Fragebogen**

Um einen Eindruck davon zu bekommen, inwieweit die Verbesserungen für Sie sinnvoll sind, und wo hier und dort an den Stellschrauben gedreht werden sollte, haben wir einen kleinen Bewertungs-/ Fragebogen zusammengestellt.

Um *clas-digital* zu verbessern, würden wir uns freuen, wenn Sie sich die Zeit nehmen könnten diesen auszufüllen.

Die neueste Version kann hier getestet werden: [clas-digital-development-server \(clas-digital.uni-frankfurt.de:9991/\)](http://clas-digital-development-server (clas-digital.uni-frankfurt.de:9991/))

## Fragen

### Ranking

Fürs Ranking verwenden wir nun eine leichte Abwandlung des Okapi\_BM25-Algorithmus, welcher das Ranking eines Dokuments in Abhängigkeit anderer Dokumente berechnet. Kurz: ein Dokument wird höher gerankt, wenn der Suchbegriff in Abhängigkeit zu anderen Dokumenten und der Länge des Dokuments besonders häufig vorkommt. Selten vorkommende Begriffe sind dadurch besonders stark gewichtet.

*Feedback:*

Das führt unter anderem dazu, dass Bücher ohne Corpus stärker gewichtet werden und oft zunächst nur Bücher ohne Corpus angezeigt werden.

Dies bewerten wir allerdings als positiv, da die ersten Treffer tatsächlich sehr relevante Treffer sind. Das nun zuerst Bücher ohne Corpus angezeigt werden betrachten wir nicht als ein Problem der Gewichtung, sondern als ein Filter-Problem, daher haben wir die Option "Show only Metadaten-/ Corpus-Matches" wieder hinzugenommen.

*Feedback:*

### Welche Metadaten sollen durchsucht werden

Es gibt nun die Möglichkeit in einer Konfigurations-Datei flexibel die zu durchsuchenden Metadaten anzugeben, in der Preview wird außerdem angezeigt, wo, der Treffer gefunden wurde, z.B.: Title: ...auf, auf, auf. Die Wilden Hunde ,penthesileas'...

Als erste Vorentscheidung haben wir uns für die folgenden Tags entschieden:

- Title
- Autoren (Vor- und Nachnamen)
- Editoren (Vor- und Nachnamen)
- Datum
- Book-Key

Gibt es Tags, die Sie gerne außerdem hinzufügen würden?

*Feedback:*

**Wie gut sind die Treffer?**

Wir haben unter anderem die Fuzzy-Search angepasst, und außerdem Treffer stärker gewichtet, wenn das gesuchte Wort am Anfang oder Ende des gefundenen Wortes auftaucht ("Jagd**hund**" > "Jahr**h**undert")

*Feedback:*

**Wie gut sind die Previews?**

Wir haben uns bemüht die Erstellung der Previews und die Auswahl der besten Previews für ein Wort in einem Buch zu verbessern.

*Feedback:*

**Wie gut ist die Suchgeschwindigkeit?**

Wir haben uns bemüht auch die Suchgeschwindigkeit und das Erstellen der Previews durch bessere Algorithmen und das Benutzen einer Datenbank zu verbessern.

*Feedback:*

**Platz für sonstige Anmerkungen**

Haben Sie sonst Anmerkungen, Wünsche, Verbesserungsvorschläge?

*Feedback:*