

Search & Analytics SHI GmbH

Welches Fallbeispiel, welchen Use Case zeigen Sie auf dem CrossMedia-Forum digital?

Lessons Learned: Aufbau eines Dienstes zur automatischen Keyword-Extraktion Bibliographischer Metadaten

Welche Zielsetzungen wurden mit dem Projekt verfolgt?

Die richtigen Keywords sorgen für Umsatz: Sie sichern die Sichtbarkeit, optimieren die Auffindbarkeit und bewirken ein höheres Ranking von Publikation aller Art.

Die Metadatenpflege ist mühsam und Outsourcing teuer. Das war unsere Motivation, die im vergangenen Jahr den Startschuss zur Entwicklung von CAKE bildete – unserem Webservice zur Extraktion von Keywords aus Metadaten.

Was waren die Herausforderung bei der Umsetzung des Projektes?

Das richtige Werkzeug zu finden, um wiederum die richtigen Keywords zu extrahieren war eine große Herausforderung und die konstante Weiterentwicklung ist es bis heute.

Worin bestanden die Herausforderung im Hinblick auf interne Abläufe und Workflows?

- Integration von CAKE in das bestehende Metadaten-Management
- Abgesehen von den individuellen Prozessen innerhalb der Verlage wird auch die Qualität der Keywords individuell bewertet – teils sogar innerhalb eines Verlags.



Daniel Wrigley arbeitet an Such- und Advanced Analytics-Anwendungen mit einem starken Fokus auf Open-Source-Projekten wie NiFi, Kafka, Solr, Spark oder Zepelin, um für Unternehmen den größten Nutzen aus ihren Daten zu ziehen. Seine Erfahrungen haben ihm den Weg zu einem O'Reilly-Autor geebnet und ihn bis auf internationale Konferenzen geführt.

Wodurch zeichnet sich die Umsetzung vor allem aus?

Unser gewählter Ansatz basiert auf dem NEWBOOKS Stichwortindex. Dadurch wird im Automatisierung die Qualität der ermittelten Keywords sichergestellt. Zudem greifen moderne Mittel und Methoden aus ETL und Natural Language Processing. Diese Kombination ermöglicht eine wesentliche Arbeitserleichterung bei der Keyword-Vergabe.

Wie lässt sich die Lösung bzw. Ergebnis in drei Sätzen beschreiben?

- Passende Keywords verkaufen mehr Bücher.
- Verbesserte Sichtbarkeit: Stärkere Wahrnehmung der Titel von mehr Endkunden und Buchhändlern auf Onlineportalen

- und in Katalogen
- Optimale Auffindbarkeit: Erhöhte Auffindbarkeit bei Online- und Katalogrecherchen
- Relevante Platzierungen: Deutlich höhere Rankings der Titel in Suchergebnissen und bei Kaufempfehlungen
- Höherer Umsatz: Steigerung des Verkaufs durch Auffindbarkeit, Sichtbarkeit und vor allem Relevanz

Welchen Rat geben Sie Verlagen, wenn Sie ein solches Projekt angehen wollen vor allem in Bezug auf die Vorbereitung und das Projektmanagement?

Passende Keywords zu finden und zu vergeben ist eine anspruchsvolle Aufgabe. Diese Keywords über mehrere Personen hinweg konsistent zu vergeben ist rein manuell gesehen beinahe unlösbar. Einerseits helfen hier detaillierte Richtlinien innerhalb des Verlags, um einen gewissen Rahmen zu definieren, innerhalb dessen die Keywords vergeben werden. Andererseits hilft es, wenn in diesem Prozess Möglichkeiten geschaffen werden, um den manuellen Prozess zu unterstützen. Diese Möglichkeiten werden in den wenigsten Projekten vorgesehen und erschweren später die Anbindung intelligenter und automatischer Dienste wie CAKE, die Mitarbeiter unterstützen und entlasten können.

