Universität Stuttgart
Universitätsbibliothek

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

Karlsruher Institut für Technologie

ZAR
Zentrum für Angewandte Rechtswissenschaft

Von nützlich zu nutzbar – Zugang zu geschützten Forschungsdaten ermöglichen

XSample
access, reuse, advance

Sibylle Hermann, Felicitas Kleinkopf, Markus Gärtner

Motivation

XSample
access, reuse, advance

- Die Forschung an Texten ist in den digitalen Geisteswissenschaften durch Defizite des deutschen Urheberrechts konfrontiert
- Beschränkung der Weitergabe, Archivierung und Zugriffsmöglichkeiten auf Forschungskorpora des Text und Data Mining (TDM)
- Konsequenz: kaum Forschung auf urheberrechtlich geschützten Texten

§ 60c UrhG

XSample
access, reuse, advance

- § 60c UrhG erlaubt, zu Zwecken der nicht-kommerziellen wissenschaftlichen Forschung, Auszüge von bis zu 15% bzw. vollständige Werke geringen Umfangs zu nutzen

- Idee: Anwendung des § 60c auf § 60d UrhG

§ 60d UrhG

- Ab 01.03.2018: TDM - Erlaubnis zugunsten der nicht-kommerziellen wissenschaftlichen Forschung
- Auf Basis dieses Rechtsrahmens war es möglich, zu Forschungszwecken TDM-Analysen vorzunehmen
- Forschungsergebnisse mussten allerdings **gelöscht** oder einer **archivierenden Institution übergeben** werden
- Seit 07.06.2021: **Löschung der Daten** auf Datenträgern der Forschenden selbst ist nun **nicht mehr erforderlich**, allerdings ist auch die **ausdrückliche Erlaubnis der Datenweitergabe an archivierende Institutionen weggefallen**
- Defizit: Rechtslage zum Umgang mit den Korpora nach Abschluss der Forschungsarbeiten

Technischer Ansatz

XSample
access, reuse, advance

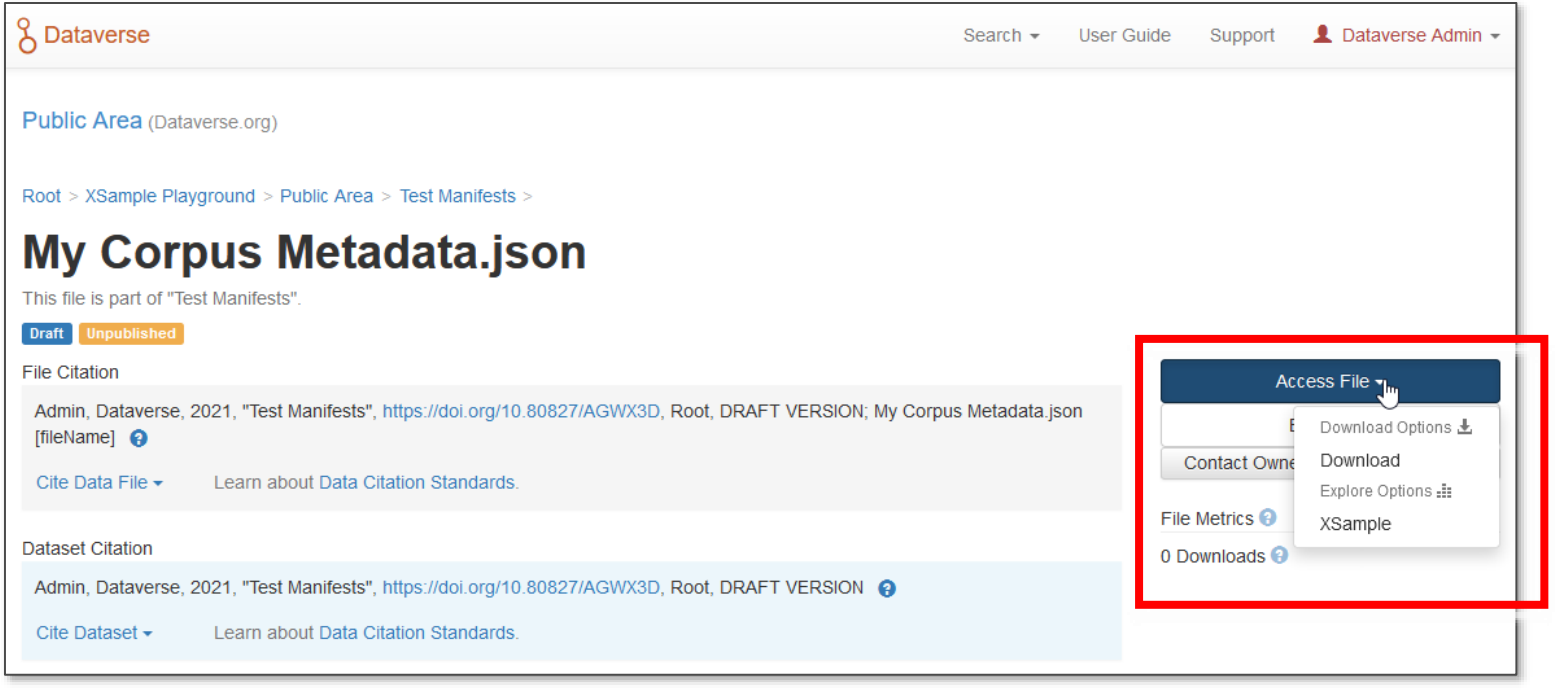
- Annahme: geschützte Ursprungsdaten und daraus erzeugte Korpora liegen zusammen mit Metadaten in einem Repositorium
- Einstiegspunkt: öffentlich auffindbare Metadaten
- Auszugserstellung: Über eine Plugin-Schnittstelle, direkt über die Weboberfläche des Repositoriums

Auszugerstellung

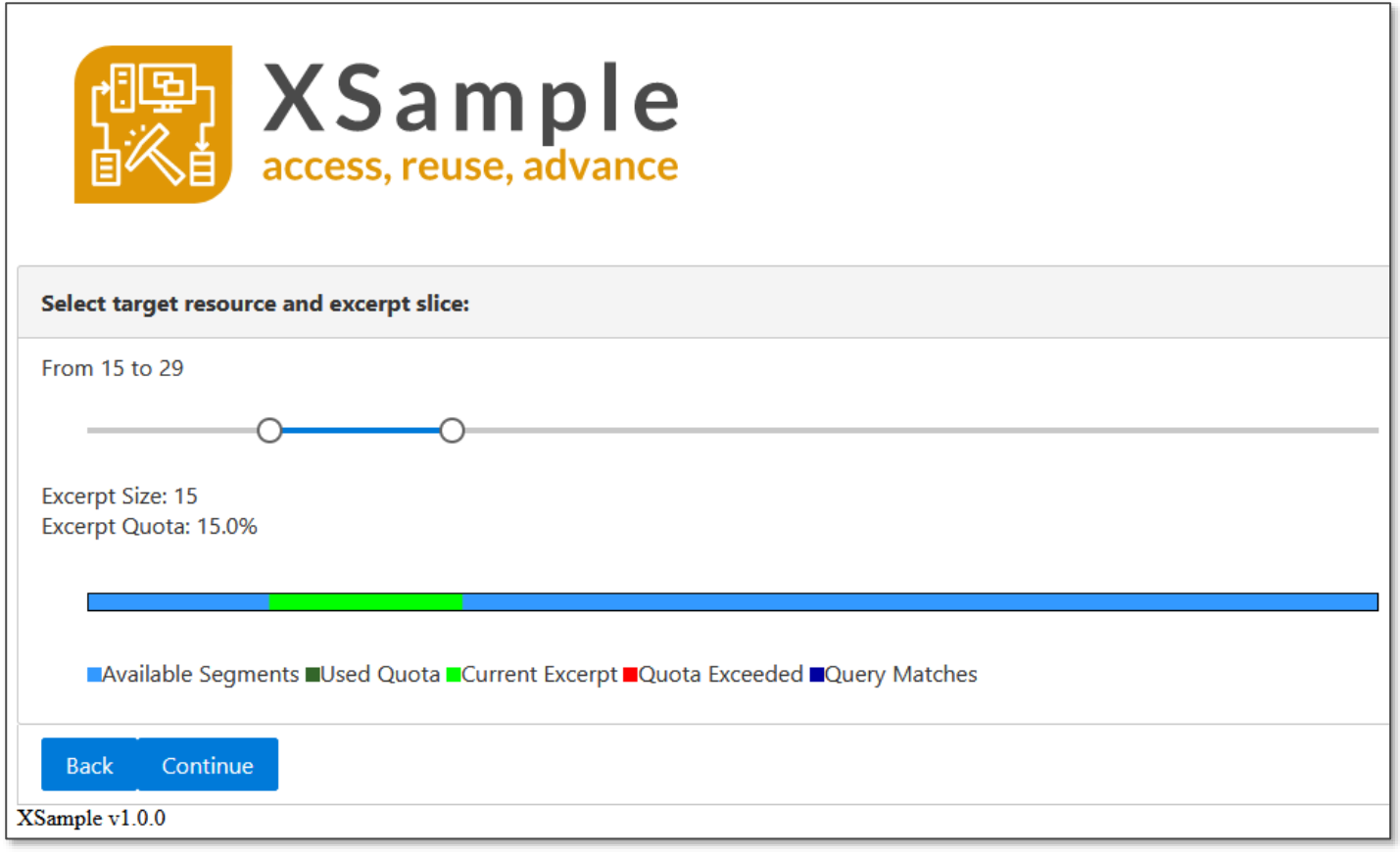
XSample
access, reuse, advance

- Statisch: z.B. die ersten 15% oder zusammenhängende Abschnitte (z.B. der Bereich 50-60%) eines Korpus als Auszug
- Dynamisch: feinere Abstufungen, die Suchanfragen auf Basis der im Korpus enthaltenen Annotationen ermöglichen
- Basierend auf diesen Suchergebnissen werden die auszugebenden Segmente der Primärdaten (zumeist Seiten) bestimmt

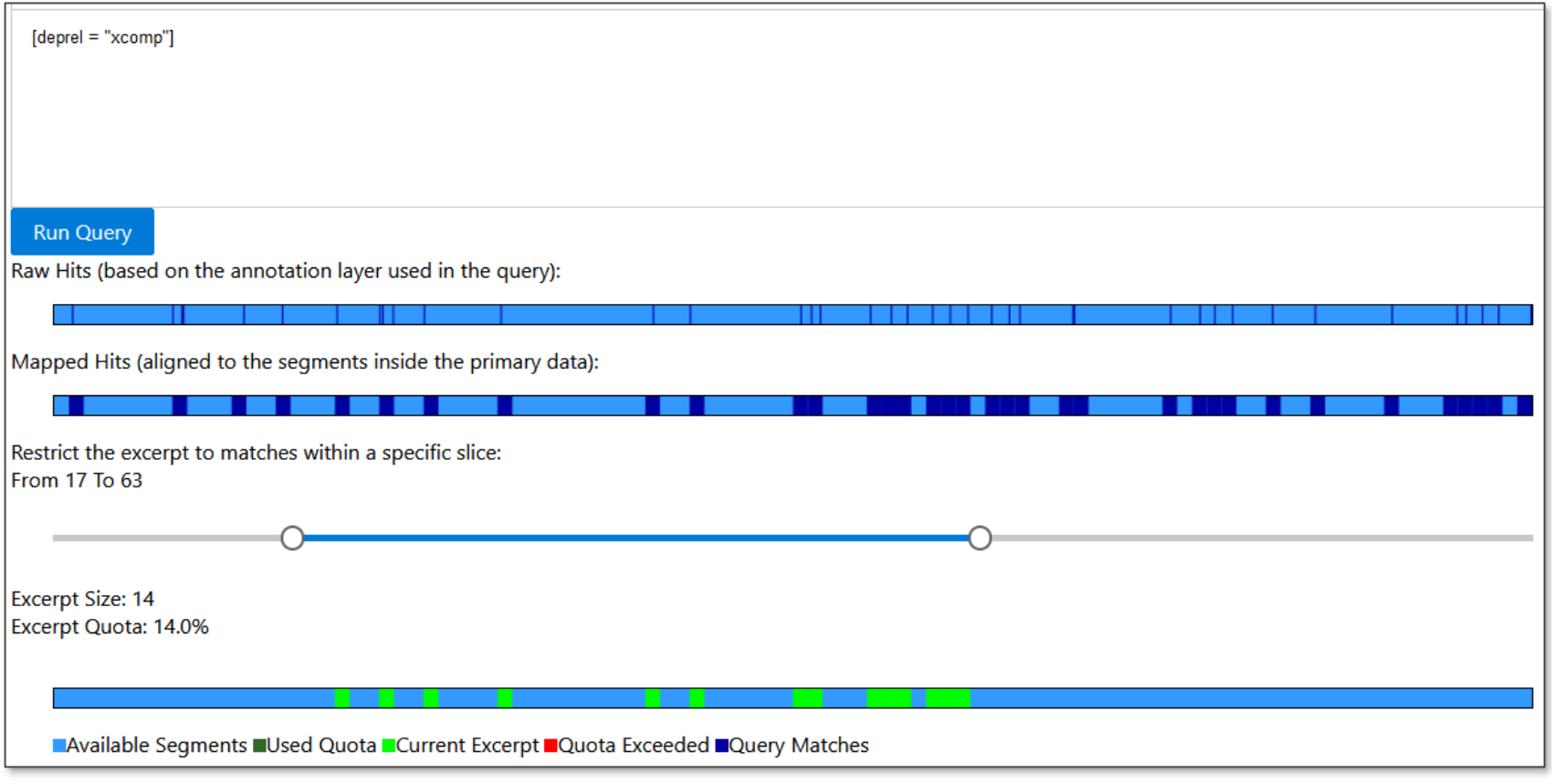
Auszugerstellung - Einstiegspunkt



Auszugerstellung - Abschnitt



Auszugerstellung - Query



Use Case 1: Unzuverlässiges Erzählen

Zeichnung von E. T. A. Hoffmann zu seinem Buch „Der Sandmann“

- Unzuverlässiges Erzählen ist ein in einigen literarischen Erzählungen auftretendes Phänomen, bei dem den Aussagen des Erzählers nicht zu trauen ist
- Faktenbezogene Unzuverlässigkeit, die mithilfe von automatischen und manuellen Annotationen von Primär- und Sekundärtexten untersucht wird
- Breites Spektrum unterschiedlicher literarischer Werke, sowohl urheberrechtlich geschützt als auch gemeinfrei und in diversen typischen Publikationsformaten

Use Case 2: Wissenschaftssprache

XSample
access, reuse, advance

- Wissenschaftssprache der geisteswissenschaftlichen Disziplinen Literaturwissenschaft, Linguistik und Philosophie werden in einem datengeleiteten Verfahren miteinander verglichen
- Datengrundlage: Korpus aus 45 Artikel Zeitschriftenartikeln pro Fach
- Die Daten werden mit linguistischen Annotationen zu Lemmata, Wortarten und syntaktischen Dependenzen versehen
- Auf Grundlage dieser Annotationen (und der einfachen Ebene der Wortformen) ist eine gezielte Auswahl von Textabschnitten möglich
- Zugriff auf die Kontexte, wie er durch das Auszugsprinzip geleistet werden kann, ist hierzu unbedingt notwendig

Weitere Informationen

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

https://mwk.baden-wuerttemberg.de/de/service/presse-und-oeffentlichkeitsarbeit/pressemitteilung/pid/bigdiwa-bibliotheken-gestalten-digitalen-wandel/

XSample
access, reuse, advance

https://bw-bigdiwa.bib.uni-mannheim.de/projekte/xsample-stuttgart/