

Suche nach bestimmten Texten im VLO

10/13/2024 10:12:50

FAQ Article Print

Category:	FAQ Clarin::Ressourcen	Votes:	0
State:	public (all)	Result:	0.00 %
Language:	de	Last update:	11:31:20 - 07/25/2016 (Europe/Berlin)

Keywords

Suche

Symptom (public)

Ich bin auf der Suche nach einer Sammlung von Texten zu einem bestimmten Thema, einer bestimmten Textsorte, eines bestimmten Autors, Zeitperiode oder Region. Wie kann ich feststellen, ob ein solches Textkorpus in der CLARIN-Umgebung zur Verfügung steht?

Problem (public)

Solution (public)

Die Standardsuchmaschine für Ressourcen in CLARIN ist das Virtual Language Observatory (VLO) (Erreichbar unter [1]<http://catalog.clarin.eu/ds/vlo>). Diese Suchmaschine nutzt die von den RessourcenanbieterInnen zur Verfügung gestellten Metadaten über Ressourcen (wie einzelne Texte, Korpora, Webservices usw.), extrahiert relevante Angaben und stellt sie in übersichtlicher Form auf einer Webseite zur Verfügung. Derzeit sind Metadaten zu ungefähr 500.000 Ressourcen über das VLO abfragbar (Stand Januar 2013).

Für die Suche nach einer konkreten Ressource unterstützt das VLO zwei Strategien: das Einschränken der Ergebnismenge über Facetten, und die Suche auf der getroffenen Vorauswahl über die Eingabe von Suchtermen (ähnlich einer Suchanfrage in üblichen Web-Suchmaschinen). Für jede Anfrage wird die Ergebnismenge angezeigt und kann durchsucht werden. Für jede einzelne Ressource steht eine Übersichtsseite zur Verfügung, die alle vorhandenen Metadaten zusammenfasst.

Diese Übersicht enthält unter anderem auch einen Link auf das ursprüngliche Metadatenfile sowie Links auf die eigentliche Ressource (falls verfügbar).

Derzeit unterstützt das VLO zehn verschiedene Facetten: Kollektion, Sprache, Kontinent, Genre, Land, Thema, Herkunftsorganisation, Ressourcentyp, Metadatenformat und Herkunftsprojekt. Durch die Nutzung dieser Facetten ist für einzelne Fragestellungen bereits die direkte Selektion relevanter Ergebnisse möglich; unter anderem wird die Einschränkung der Treffermenge nach Kriterien wie der Sprache der Ressource, Ressourcentyp oder des Anbieters der Ressource unterstützt.

Für Suchanfragen die nicht ausschließlich durch diese Selektion geklärt werden können bzw. für die keine Facetten zur Verfügung stehen, kann zusätzlich die Suche über das Suchfeld genutzt werden. Dabei werden die Beschreibungstexte nach den jeweiligen Suchtermen durchsucht. Falls die benötigten Informationen in den Metadaten enthalten sind, können somit erweiterte Suchanfragen wie die Suche nach bestimmten Autorennamen oder auch nach konkreten Jahresangaben durchgeführt werden.

Da das VLO ausschließlich Informationen enthält die aus den Metadaten gewonnen werden können, hängt die Qualität (und der Umfang) der Ergebnisse stark davon ab was und in welchem Umfang in den Metadaten zur Verfügung gestellt wird. Suchanfragen die durch die vorhandenen Metadaten nicht bearbeitet werden können, müssen somit durch manuelle Inspektion der Treffermenge bei den jeweiligen RessourcenanbieterInnen geklärt werden.

[1] <http://catalog.clarin.eu/ds/vlo>