

Judaica recherchieren – Unterstützung bei der Realisierung forschungsspezifischer Suchlösungen durch die generische Suche von DARIAH-DE

,
tobias.gradl@uni-bamberg.de
Universität Bamberg

,
lor@steinheim-institut.org
Salomon Ludwig Steinheim-Institut für deutsch-jüdische
Geschichte

,
andreas.henrich@uni-bamberg.de
Universität Bamberg

Einleitung

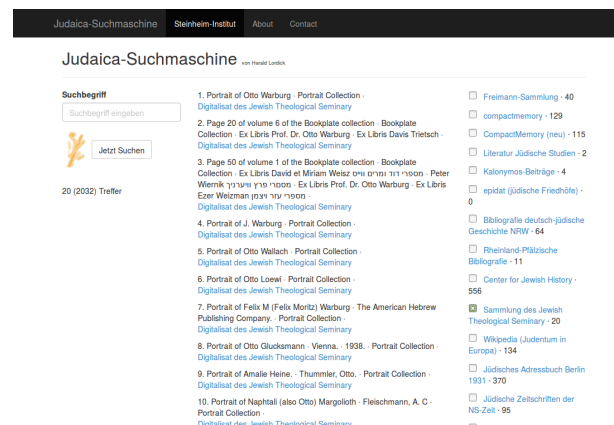
Jenseits der standardisierten und institutionalisierten Bereitstellung von Forschungsquellen und -literatur bedarf es weiterführender Recherchekonzepte und Anwendungen, die Geisteswissenschaftler_innen mit ihren spezifischen Forschungen und den *für sie verfügbaren und relevanten Daten* so weit wie möglich entgegenkommen. Integrative Suchlösungen wie OAIster und Europeana bieten Zugriff auf eine Vielzahl von Kollektionen und unterstützen breite, fachunabhängige Suchanfragen im strukturellen Rahmen integrativer Schemata (Hagedorn 2013; Peroni et al. 2013). Der semantisch tiefe und forschungsspezifische Zugang zu Ressourcen steht bei solchen Suchportalen dabei zumeist nicht im Fokus. Aus diesem Grund stehen den Suchportalen spezifische Lösungen gegenüber, die speziell an die Bedürfnisse des einzelnen Forschers und seine aktuellen Fragestellungen angepasst sind. Ein Beispiel bildet hierbei die am Steinheim-Institut entwickelte Judaica-Suchmaschine (Lordick 2013).

In diesem Beitrag stellen wir mit der generischen Suche von DARIAH-DE eine weitere breite Suchlösung vor, welche jedoch auch Funktionen spezifischer Ansätze realisiert, wie etwa die individuelle Aggregation und Filterung von Sammlungen, sowie differenzierte Möglichkeiten der Zugriffskontrolle. Auf Basis des Anwendungsfalls der Judaica-Suchmaschine zeigen wir exemplarisch die

Unterstützung individueller Suchbedürfnisse im Rahmen der generischen Suche und verdeutlichen dabei, wie neue, fachspezifische Suchen verfügbar gemacht werden können - ohne dass die hierfür notwendigen, technischen Aspekte durch den einzelnen Forscher umzusetzen sind.

Die Judaica-Suchmaschine als Anwendungsfall

Als Use-Case dient die langjährige Auseinandersetzung mit den Möglichkeiten und Grenzen einer übergreifenden *Judaica-Suchmaschine* im Steinheim-Institut. Sie war ursprünglich als Allegro-C-Katalog gestartet, der verschiedene institutsinterne Datenbanken in eine gemeinsame Datenbank mit dem Ziel der effizienteren fachspezifischen Recherche zusammenführte. Bald darauf wurde begonnen, auch passende externe Datenangebote einzubinden. Die Suchmaschine enthält zur Zeit ca. 500.000 Datensätze aus 20 filterbaren Katalogen, ist auch für mobile Geräte geeignet (Lordick 2014) und XML-basiert.



Über Standardschnittstellen lassen sich beispielsweise integrieren: die Freimann-Sammlung mit 8.772 Titeln (Teil von Judaica Frankfurt), CompactMemory 46.637 (seit 2014 ebenfalls Judaica Frankfurt), Center for Jewish History (New York) 65.667 oder Jewish Theological Seminary 9.257. Die ca. 15.000 digitalisierten Seiten der Sammlung Jüdische Zeitschriften der NS-Zeit der Deutschen Nationalbibliothek sind zwar 2012 abgeschaltet worden. Die Suchmaschine enthält aber noch die 34.346 erschließenden Metadaten dazu. Sie haben nun 'nur noch' bibliografischen Charakter, der um so wertvoller bleibt. Gleiches gilt für die Exilpresse digital: Deutsche Exilzeitschriften 1933–1945 mit nicht weniger als 231.548 Metadaten.

In diesen Kontext gehört auch der Ansatz, eigene Daten ebenfalls über etablierte Standardformate und Protokolle anzubieten, um sie zur Nachnutzung zur Verfügung zu stellen. So wurde ein OAI-PMH Testserver eingerichtet,

der digitale Sammlungen des Steinheim-Instituts zusammenführt: Universal-Kirchenzeitung, Kalonymos, Deutsch-jüdische Publizistik. Ebenfalls wurde das Verfahren der Bereitstellung mittels eines statischen OAI-PMH-Repositorys geprüft (Beer et al. 2013). Dieser Weg ist empfehlenswert, hinsichtlich des erforderlichen technischen Knowhows jedoch durchaus anspruchsvoll, erfordert (Zugriff auf) Infrastrukturkomponenten und Serverressourcen und eignet sich insbesondere für 'fertige', abgeschlossene Datensätze.

Viele digitale Quellen lassen sich jedoch unter den gegebenen Umständen gar nicht harvesten, etliche verfügbare fachlich interessante oder auch unentbehrliche Datensätze immerhin mittels individueller Programmierung in die Suchmaschine einbinden: das jüngst erschienene Jüdische Adressbuch Berlin 1931, die NS-Liste der verbannten Bücher, das Kapitel Judentum der Rheinland-Pfälzischen Bibliografie oder die 10.600 Artikel der Kategorie Judentum in Europa der Wikipedia etwa.

Umsetzung im Rahmen der generischen Suche

Der Umgang mit technischen Problemen zählt nicht zu den typischen Aufgaben von Forschern kultur- und geisteswissenschaftlicher Disziplinen. Eine auf individuelle Bedürfnisse zugeschnittene Integration von Daten oder die Realisierung von Softwarekomponenten ist jedoch für die Umsetzung spezifischer Anwendungen wie der Judaica-Suchmaschine erforderlich. Um Lösungen für technische Probleme in nachhaltiger und für anschließende Forschungsprojekte wiederverwendbarer Form bereitzustellen, beinhalten derzeit laufende Initiativen wie DARIAH oder CLARIN häufig Arbeitspakete zur Umsetzung von Infrastrukturkomponenten.

Obwohl die generische Suche von DARIAH-DE (Gradl 2011-2016) nicht als Infrastrukturkomponente, sondern als eigenständiger Dienst entwickelt wurde, basiert auch das Konzept der generischen Suche auf dem Ziel der Umsetzung wiederverwendbarer technischer Funktionalität zur Integration, Verarbeitung und Analyse von Daten. Im Folgenden stellen wir eine im Rahmen der generischen Suche implementierte Funktionalität vor, welche es Forschern ermöglicht, spezifisch angepasste Suchlösungen zu generieren und bereitzustellen - ohne die hierzu erforderlichen technischen Aspekte selbst lösen zu müssen.

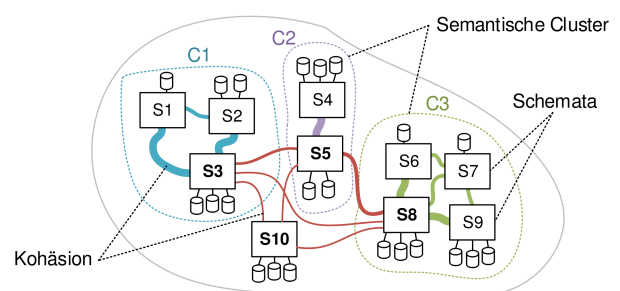
Hintergrund des generischen Konzepts

Konzept und Implementierung der generischen Suche resultieren aus der Zielsetzung der Realisierung einer kombinierten und flexibel anpassbaren Breiten- und Tiefensuche (Gradl / Henrich 2014). Dabei wurde zum einen eine universelle und einfache Möglichkeit (vgl.

Abbildung) zur übergreifenden Suche in einer Vielzahl digitaler Kollektionen eingerichtet, um Forscher bei der Suche und Analyse relevanter und potenziell unbekannter Ressourcen und Kollektionen zu unterstützen. Neben dieser disziplinunabhängigen Breiten Suche erlaubt die generische Suche auch eine kontextspezifische Anpassung #sowohl im Hinblick auf die Auswahl der zu durchsuchenden Datenquellen, als auch auf das verwendete Schema für die Integration der in diesen enthaltenen, zumeist heterogenen Ressourcen. Die so entstehende Tiefensuche erlaubt eine differenziertere Analyse und Suche in einer forschungsspezifisch begrenzten Datenbasis.



Die flexible Anpassbarkeit der generischen Suche an übergreifende oder spezifische Fragestellungen wird dabei durch das in untenstehender Abbildung exemplarisch angedeutete Konzept der forschungsorientierten Föderation digitaler Kollektionen erreicht (Gradl / Henrich 2014; Gradl et al. 2014).



Anstelle der für übergreifende Integrationssysteme typischen Harmonisierung heterogener Daten (vgl. Batini et al. 1986; Lenzerini 2002) basiert die generische Suche wesentlich auf folgenden Annahmen:

- Durch die semantische Assoziation von Kollektionen und darin verwendeten Schemata durch fachwissenschaftliche Experten können integrative Sichten generiert werden, die den Anforderungen konkreter Forschungsfragen entsprechen. Dies wird

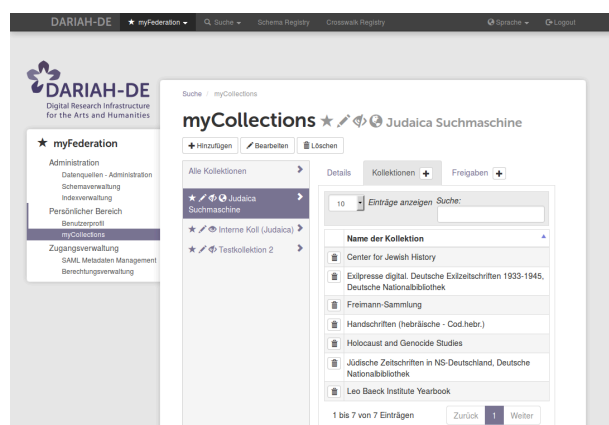
insbesondere dadurch erreicht, dass nicht technisch motivierte Integrationsziele wie die Vollständigkeit und Korrektheit eines Integrationsschemas (Batini et al. 1986) im Vordergrund stehen, sondern disziplinspezifische und auch konfliktäre Zusammenhänge modelliert werden können.

- Kohärente Bereiche mit eng assoziierten Schemata *S* und Kollektionen werden in obiger Abbildung durch *semantische Cluster* widergespiegelt und bilden die Voraussetzung für spezifische Tiefensuchen. Für die übergreifende Breitensuche werden wichtige Schemata der einzelnen Cluster (repräsentiert als *S3*, *S5* und *S8*) mit generischen Schemata, wie z. B. Dublin Core - in der Abbildung symbolisiert durch *S10* - assoziiert.

Die besondere Eigenschaft des Föderationskonzepts besteht darin, dass Daten in ihrer ursprünglichen Form analysiert und indexiert werden. Erst zum Anfragezeitpunkt und in Abhängigkeit von der einer Anfrage zu Grunde liegenden Zusammenstellung von Kollektionen werden die Daten zusammengeführt und integriert.

Abbildung der Cluster in der generischen Suche

Die generische Suche beschreibt einen im Rahmen des DARIAH-DE Projektes entwickelten Dienst, welcher im Hinblick auf seine Datenbasis auf die Einträge der DARIAH-DE Collection Registry zurückgreift (Plutte et al. 2014) und registrierten Benutzern die Aufnahme weiterer Datenquellen erlaubt. Neben der Assoziation von Schemata als wesentlicher Teilaspekt (Gradl / Henrich 2014) bildet die Möglichkeit der Auswahl und Gruppierung relevanter Kollektionen im Rahmen der generischen Suche (als *myCollections*) die Basis für die Erstellung angepasster Suchlösungen.

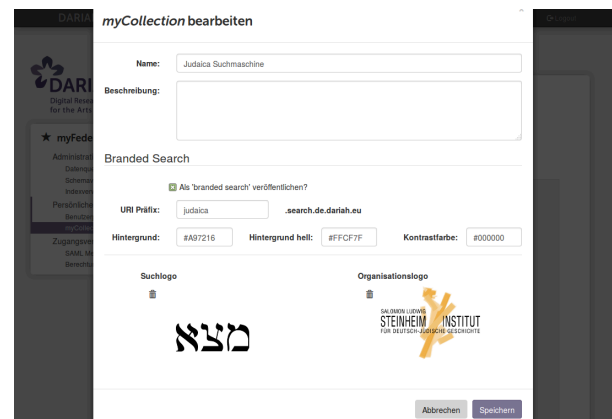


Der Bildschirmausschnitt zeigt exemplarisch drei solcher *myCollections* und verdeutlicht die derzeit implementierte Funktionalität:

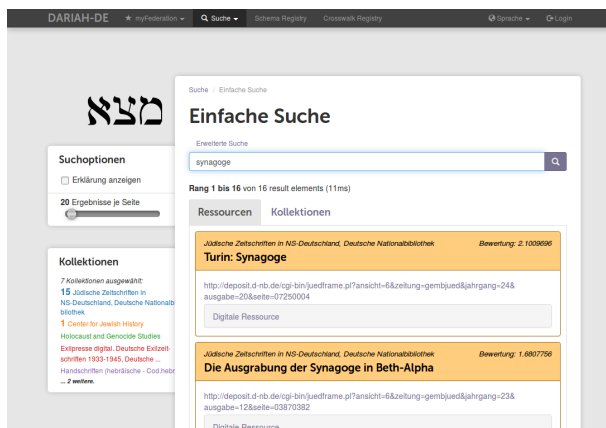
- Einträge von *myCollections* können durch authentifizierte Anwender angelegt werden und definieren eine Zusammenstellung einzelner Kollektionen.
- Diese *myCollections* erlauben eine Schnellauswahl von Kollektionen bei der Ausführung von Suchanfragen - sowohl im Rahmen des Benutzerinterfaces, als auch in Form der REST-basierten Schnittstelle.
- Möchte der Benutzer eine *myCollection* (z. B. im Rahmen eines gemeinsamen Forschungsinteresses) teilen, so können Freigaben für weitere Benutzer oder DARIAH Gruppen erteilt werden.

Spezifischer Zugang in Form einer *Branded Search*

Anknüpfend an den eingeführten Anwendungsfall der Judaica-Suchmaschine wird der Zusammenhang zwischen der generischen Suche von DARIAH-DE und forschungsspezifischen Suchlösungen abgebildet durch die Idee der benutzerdefinierten Zusammenstellung von Kollektionen: Wird eine *myCollection* - wie in der Abbildung unten dargestellt - als *Branded Search* ausgezeichnet, so wird dadurch eine eigene Suchoberfläche veröffentlicht, welche sowohl optisch als auch inhaltlich an spezifische Bedürfnisse angepasst und von der eigentlichen generischen Suche abgegrenzt ist.



Die folgenden Bildschirmausschnitte verdeutlichen insbesondere die optische Abgrenzung durch konfigurierbare Farbgebung und die Verwendung von Such- und Organisationslogos. Insbesondere der Vergleich der Wordclouds der Startseiten von generischer Suche und (Gradl / Lordick 2015-2016) deuten jedoch die jeweils unterschiedliche, zu Grunde liegende Datenbasis an: Die in einer *Branded Search* angebotenen Kollektionen spiegeln bei sämtlichen Such-, Analyse- und Visualisierungsaufgaben die von den Erstellern der Suche getroffene Kollektionsauswahl wider.



Im Fall der Veröffentlichung einer Branded Search bleibt diese auch als zugreifbare myCollection für berechnete Benutzer erhalten und kann im Hinblick auf die Kollektionsauswahl und die Assoziation der Schemata verändert werden #mit unmittelbaren Auswirkungen auf die entsprechende Branded Search. Von technischen Implementierungen an der Basis der generischen Suche, z. B. der Anbindung weiterer Quellenarten wie Wikipedia können schließlich sämtliche eingerichteten Branded Searches profitieren, sofern dies durch die jeweiligen Forscher gewünscht wird.

Ausblick

Ein ausgeprägt generischer Ansatz muss kein Gegensatz zu den in den Geisteswissenschaften vorherrschenden individuellen Fragestellungen und Forschungsansätzen sein. Indem sie entsprechende Freiheitsgrade, kreatives und kollaboratives Datenmanagement anbietet, erlaubt die generische Suche ihren Nutzern die Erstellung eigener, jeweils individuell ausgelegter Suchmaschinen.

Es ist das Knowhow der Forschenden, das die Relevanz der Daten, die sie zusammenstellen, filtern, teilen, auch ad-hoc zum Zwecke der Recherche bereitstellen, ausmacht. Ein solches Framework, verbunden mit der

fachspezifischen Kenntnis der Daten ist eine gute Basis für überraschende Funde und das Aufspüren unerwarteter Zusammenhänge.

Fußnoten

1. am Jülich Supercomputing Center, DARIAH-DE.

Bibliographie

Batini, Carlo / Lenzerini, Maurizio / Navathe, Shamkant Bhalchandra (1986): "A comparative analysis of methodologies for database schema integration", in: *ACM Computing Surveys* 18, 4: 323–364.

Beer, Nikolaos / Herold, Kristin / Kolbmann, Wibke / Kollatz, Thomas / Romanello, Matteo / Rose, Sebastian / Walkowski, Niels-Oliver (2013): *Recommendations for Interdisciplinary Interoperability (R 3.3.1)*. DARIAH-DE report <https://dev2.dariah.eu/wiki/download/attachments/14651583/R3.3.1.pdf?version=1&modificationDate=1366904278298&api=v2> [letzter Zugriff 07. Februar 2016].

Europeana Foundation (2008-2015): *Europeana Collections*. Den Haag <http://www.europeana.eu/portal/> [letzter Zugriff 07. Februar 2016].

Gräd, Tobias (2011-2016): *DARIAH-DE Generic Search* <http://search.de.dariah.eu> [letzter Zugriff 07. Februar 2016].

Gräd, Tobias / Henrich, Andreas (2014): "A novel approach for a reusable federation of research data within the arts and humanities", in: *Digital Humanities 2014*. Book of Abstracts 382–384 <http://dh2014.org/program/abstracts/> [letzter Zugriff 09. Oktober 2015].

Gräd, Tobias / Henrich, Andreas / Plutte, Christoph (2015): "Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Förderung von Kollektionen", in: Constanze Baum / Thomas Stäcker (eds.): *Grenzen und Möglichkeiten der Digital Humanities*. Sonderband der Zeitschrift für digitale Geisteswissenschaften 1 http://dx.doi.org/10.17175/sb001_020 [letzter Zugriff 09. Oktober 2015].

Gräd, Tobias / Lordick, Harald (2015-2016): *Judaica Search*. Branded Search in the DARIAH-DE Generic Search <http://judaica.search.de.dariah.eu> [letzter Zugriff 07. Februar 2016].

Hagedorn, Kat (2003): "OAIster: a 'no dead ends' OAI service provider", in: *Library Hi Tech* 21, 2: 170–181.

Henrich, Andreas / Gräd, Tobias (2013): "DARIAH(-DE): Digital research infrastructure for the arts and humanities - concepts and perspectives", in: *International Journal of Humanities and Arts Computing* 7: 47–58.

Lenzerini, Maurizio (2002): "Data integration: a theoretical perspective", in: *PODS'02 - Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of*

database systems : 233-246 <http://dl.acm.org/citation.cfm?doid=543613.543644> [letzter Zugriff 09. Oktober 2015].

Lordick, Harald (2010-2016): *Vieles finden*. Die Judaica-Suchmaschine im Steinheim-Institut. Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte an der Universität Duisburg-Essen <http://steinheim-institut.de/vf/> [letzter Zugriff 07. Februar 2016].

Lordick, Harald (2013): "Die judaica-bibliothek im web. ganz real oder noch immer virtuell?", in: *Kalonymos1*: 12–14 http://www.steinheim-institut.de/edocs/kalonymos/kalonymos_2013_1.pdf [letzter Zugriff 09. Oktober 2015].

Lordick, Harald (2014): "Jüdische Geschichte (mobil) recherchieren", in: *Kalonymos3*: 13 http://www.steinheim-institut.de/edocs/kalonymos/kalonymos_2014_3.pdf [letzter Zugriff 9. Oktober 2015].

OAister (2001-2016): *OAister*. OCLC WorldCat.org Services. Dublin / Ohio: OCLC Online Computer Library Center <http://oaister.worldcat.org/> [letzter Zugriff 07. Februar 2016].

Peroni, Silvio / Tomasi, Francesca / Vitali, Fabio (2013): "Reflecting on the europeana data model", in: Agosti, Maristella / Esposito, Floriana / Ferilli, Stefano / Ferro, Nicola (eds.): *Digital Libraries and Archives*. Berlin / Heidelberg: Springer 228–240 http://link.springer.com/chapter/10.1007%2F978-3-642-35834-0_23 [letzter Zugriff 09. Oktober 2015].

Plutte, Christoph (2011-2014): *DARIAH-DE Collection Registry*. Initial Prototype. <http://colreg.de.dariah.eu> [letzter Zugriff 07. Februar 2016].