

Digitale Perspektiven mediävistischer Quellenrecherche

VON ANDREAS KUCZERA VERÖFFENTLICHT 2014-04-18

Zusammenfassung: Forschern in den Geisteswissenschaften stehen immer mehr digitalisierte Volltexte zur Verfügung. Daraus ergeben sich Chancen, neue Fragestellungen werden möglich, deren Beantwortung wiederum neue Perspektiven eröffnet, aber auch Herausforderungen schafft. In diesem Aufsatz wird versucht, diese Probleme einzugrenzen und neue Möglichkeiten für die Arbeit mit Volltexten in den Geisteswissenschaften mit einem Fokus auf die Mediävistik zu skizzieren.

Nutzerverhalten — Nutzerperspektiven

Der Zugang, den ein Forscher im Hinblick auf eine Fragestellung in der Mediävistik wählt, formt sich zumeist erst in der intensiven und direkten Auseinandersetzung mit den Quellen. Das Spektrum der Auseinandersetzung reicht dabei vom Studium und der Transskription des Originals hin zur Benutzung von Editionen oder Regestenwerken. Letztendlich aber stellt der Forscher meist engen Kontakt zu den Quellen her, indem er sie deutet, einordnet und mit den aus ihnen geschöpften Aussagen arbeitet. Zusammenfassende und erschließende Hilfsmittel bei Regesten- und Editionswerken sind, soweit vorhanden, Personen-, Orts- und Sachregister. Seit den 1990er Jahren gab es erste Anstrengungen maßgebliche Quellenbestände, Editionen und Regestenwerke im Volltext im Internet für die Recherche zur Verfügung zu stellen. Die zunehmende Verfügbarkeit der Volltexte änderte auch langsam die Herangehensweise an die Quellen: Die Volltextsuche ergänzte zunächst das gesamthafte Selbststudium aller für eine wissenschaftliche Fragestellung relevanten Quellen. Im Laufe der Zeit verschoben sich die Arbeitsschwerpunkte vom Studium der Einzelquellen hin zu Suchanfragen über die Volltextsuche. Damit einher geht der zunehmende Verlust des direkten Kontakts zur Quelle oder ihrer bearbeiteten Fassung in einer Edition. Das bedeutet bei den heutigen Studenten (und damit auch den zukünftigen Geschichtswissenschaftlern), dass sie eine Quelle in ihrer ursprünglichen Form sowie ihren Repräsentationen in einer Editionen immer weniger wahrnehmen da sie das Original oder die Editionsbande kaum noch zur Hand nehmen. Sie beschäftigen sich vielmehr mit Suchergebnissen, die ihren ursprünglichen Kontext der Quelle kaum noch deutlich werden lassen. Andererseits können die neuen Möglichkeiten sie aber auch in die Lage versetzen, wesentlich größere Datenbestände gemeinsam zu durchsuchen und damit Fragestellungen aufzuwerfen, deren Behandlung einem Wissenschaftler unter den traditionellen „analogen“ Rahmenbedingungen allein auf Grund seiner beschränkten Lebenszeit und seines begrenzten Auffassungsvermögens gar nicht möglich war. Dies wird ohne Zweifel mittelfristig zu einem Perspektivenwechsel innerhalb der Geschichtswissenschaften führen, da neue Fragestellungen zu viel größer bemessene Quellenbestände möglich werden. So interessant die neuen Möglichkeiten für die Geschichtswissenschaften auch sein mögen, sind vor ihrer allgemeinen Akzeptanz doch noch methodische und qualitative Fragen zu klären.

Suchinterfaces — Ergebnisdarstellung: Nutzerperspektiven

CIN (concrete information need)	POIN (problem-oriented information need)
Thematische Grenzen sind klar abgesteckt.	Thematische Grenzen sind nicht exakt bestimmbar.
Die Suchfrageformulierung ist durch exakte Terme ausdrückbar.	Die Suchfrageformulierung lässt mehrere terminologische Varianten zu.
Eine Fakteninformation reicht aus, um den Bedarf zu decken.	In der Regel sind mehrere relevante Treffer möglich. Ob der Informationsbedarf damit abschließend gedeckt ist, bleibt offen.
Mit der Übermittlung der Fakteninformation ist das Informationsproblem erledigt.	Mit der Übermittlung der Ergebnisse wird ggf. das Informationsproblem modifiziert oder ein neuer Bedarf geweckt.

Geht man zunächst vom Nutzer einer Suchmaschine aus, wird in den Fachwissenschaften zwischen CIN (concrete information need) und POIN (problem-oriented information need) unterschieden.¹ In der Tabelle werden die beiden Ansätze auf einen fiktiven Nutzer der Regesta Imperii projiziert.² Als kurze Erläuterung für die Tabelle hier noch zwei Beispiele: Eine Suchanfrage im Sinne von CIN ist z. B. die Suche nach einem Regest, wobei dem Nutzer Band und Regestennummer bekannt sind. Eine POIN-Suche würde z.B. die Verteilung von Nennungen der Kurfürsten in den Regesten Kaiser Friedrichs III. umfassen. Diese Kategorisierungen beschreiben nicht vollständig alle Suchstrategien der Nutzer der Regesta Imperii Online, bilden aber zwei größere Nutzergruppen ab, deren Perspektiven in den folgenden Betrachtungen berücksichtigt werden sollen.

Beispiele zu Suchinterfaces von mediävistischen Quellenportalen

Den Zugang zu digitalen Quellen im Bereich Mediävistik gewähren unter anderem die Suchinterfaces von Editions- und Regestenprojekten. In einem ersten Schritt werden daher exemplarisch die Suchmasken von vier verschiedenen Projekten untersucht und verglichen.³ In einem zweiten Schritt werden die Funktionen für die Anzeige der Ergebnisse näher betrachtet.

Regesta Imperii

Die Online-Regestensuche der **Regesta Imperii Online** zeigt zunächst nur einen einfachen Suchschlitz an, um dem allgemeinen Nutzer einen möglichst flachen Einstieg zu ermöglichen. Die Expertensuche bietet dagegen wesentlich mehr Möglichkeiten, die Suche weiter einzuschränken und ermöglicht es, die Treffermenge auf ein zu bewältigendes Maß zu reduzieren. Insbesondere können flexibel mehrere Stichworte mit und/oder/nicht als Phrase oder auch als Ausstellungsort gesucht werden. Als Einstieg steht an prominenter Stelle eine umfangreiche Hilfeseite zur Verfügung.

Papstregesten

Expertensuche

Suche:

Testauswahl: Regestentext Überlieferung Kommentar Nachträge

Schreibweisen: toleranter Modus exakter Modus

Von Datum: Bis Datum:

Abteilung:

Band:

Nummern:

Relevanzgrenze:

Ähnlich strukturiert zeigt sich die Expertensuche des **Göttinger Akademienprojekts zu den mittelalterlichen Papsturkunden**. Es vereint eine einfache Suche und die Expertensuche auf einer Seite. Bis zu vier und/oder/und nicht-Verknüpfungen von Stichworten verschiedener Kategorien sind möglich.

Das Suchinterface des Göttinger Papsturkundenprojekts (www.papsturkunden.de).

ein Suchfeld für Stichwörter aber dieses bringt interessante Funktionalitäten mit sich: Es liefert eine Vorschau auf mögliche Stichwörter und bietet dem Nutzer Einblicke in die vorhandenen Stichwortlisten. Diese Funktion zeigt dem Nutzer vorkommende Schreibformen an und sensibilisiert ihn für die Stärken und Schwächen einer Volltextsuche. Auf den Hilfeseiten werden die nicht direkt aus dem Suchformular heraus erkennbaren Suchfunktionen erklärt.

INSCHRIFTENSUCHE

Das Suchinterface der Deutschen Inschriften (www.inschriften.net).

Struktur der zugrundeliegenden Daten wider. Alle Suchinterfaces bringen einen Nutzer, der sich mit dem Material auskennt, sehr schnell dem gewünschten Ziel näher. Zusammengefasst kommen hier Experten mit CIN-Anfragen sehr schnell zum Ziel.

Ergebnisdarstellung

Ergebnisanzeige der Regesta Imperii

meinem bisherige Kenntnisstand lässt sich die Reihung nicht beeinflussen. Große Treffermengen lassen sich hier nur schwer

Die Freitextsuche muss erst explizit in der Suchkategorisierung angewählt werden, voreingestellt sind Person, Ort etc. Zudem kann man die Suche trunkieren oder nach bestimmten Datumsangaben suchen, wobei aber zwingend ein Suchwort anzugeben ist. Eine kurze Erläuterung am Fuß der Seite gibt Hinweise zu den Suchformaten für Trunkierung oder Datumseinschränkungen, wobei sich letztere allerdings nicht sehr leicht erschließen. Zweifellos von Vorteil sind die zahlreichen durchsuchbaren Kategorien, die auf eine gut strukturierte Datengrundlage schließen lassen⁴ Interessant wäre hier in jedem Fall die Darstellung der kategorisierenden Begriffe, um dem Nutzer einen Überblick zu den vorkommenden Orten, Personenkreisen etc. zu bieten.

dMGH

Das Suchinterface der dMGH bietet nur

Das Suchinterface der MGH (www.dmggh.de).

Deutsche Inschriften Online

Unter www.inschriften.net findet man den Internetauftritt des Akademienprojekts "Die Deutsche Inschriften". Das Projekt bietet einen Google-Suchschlitz und auch eine Expertensuche, die nach Bänden, Zeiträumen etc. einschränken kann. Besonders gelungen sind auf dieser Seite die Tipps zur Suche, die informativ und kurz über die Suchmöglichkeiten informieren. So kann man z. B. mit Hilfe der Strg-Taste mehrere Bände auswählen und damit verbunden einen Zeitraum nach Treffern absuchen.

In der nebenstehenden Tabelle sind nochmal die verschiedenen Suchmöglichkeiten der Seiten zusammengestellt. Einige Suchmöglichkeiten sind bei allen Projekten gleich, andere sind jeweils auf die Eigenschaften des präsentierten Materials zugeschnitten und spiegeln die

	RI	dMGH	Page-Links	DBO
Einfacher Suchschlitz	x	x	x	x
Trunkierte Suche	x	x	x	x
Trunkierung Einzelwörter				
und-oder-nicht	x		x	x
Eingrenzung durchsuchter Datenbündel	x	x	x	
Einschränkung Band/ Abteilung etc.	x	x		x
Suche über mehrere Bände / Abteilungen				x
Datum von-bis	x	x	x	x
Exaktes Datum	x	x	ja, mit schonem Sensord B.	x
Hilfe zur Suchtechnik	vorhanden	fehlt	gut	sehr gut
Sortierung nach Datum	x	x	unklar	x
Relevanzsortierung	x	x		
Erfahrungen zu Relevanzkriterien	keine	keine		
Weitere Sortiermöglichkeiten				x
Orte	teilweise		x	
Personenregister	teilweise			
Material-Kriterium			x	

Zusammenfassung der Ergebnisse.

Anders sieht dies für Benutzer mit POIN-Anfragen aus. Im Gegensatz zu CIN-Nutzern können oder wollen sie das durchsuchte Material nicht zu stark eingrenzen, da sie einen Überblick zu den Treffern ihrer Suchanfrage erhalten möchten. Für POIN-Anfragen spielt also die Anzeige der Trefferliste eine wichtige Rolle.

Regesta Imperii: Die Regesta Imperii Online bieten bei der Ergebnisanzeige die Möglichkeit nach Datum, Herrschernamen und Regestennummern aufsteigend und absteigend zu sortieren. Die Zahl der pro Seite angezeigten Treffer kann auf 10, 20, 50 und 100 Treffer eingeschränkt werden. Zusätzlich wird die Ergebnisliste per Voreinstellung auf die 10.000 relevantesten Treffer beschränkt.⁵ Große Treffermengen lassen sich über die o.a. Möglichkeiten hinaus nicht weiter strukturieren.

Papsturkunden: Die Ergebnisse im Göttinger Papsturkunden-Projekt werden in einer zweiseitigen Liste angezeigt, deren Sortierung sich nicht sofort erschließt. Nach

verarbeiten.

dMGH: Die dMGH bieten bei der Trefferliste die Möglichkeit nach Relevanz, Jahr, Titel und Band/Sortierschlüssel auf- und absteigend zu listen und zeigt schlussendlich das jeweilige Druckbild der Trefferseite an. Darüber hinaus kann die Ergebnisliste per Facettierung eingeschränkt werden. Außerdem bieten die Angaben zu den einzelnen Facetten erste Hinweise auf die zeitliche, räumliche und inhaltliche Verteilung.

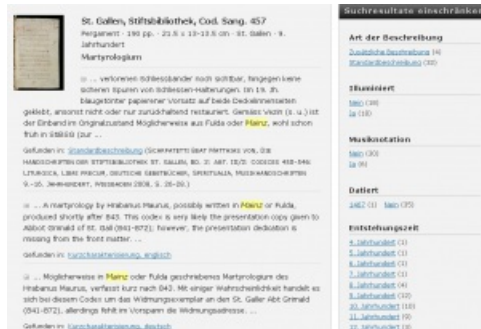


Ergebnisanzeige der dMGH mit Facettierungen rechts.

Datum, Landkreis/Stadt und Standort, auch in verschiedenen Kombinationen, zu sortieren. Die Trefferanzeige selbst lässt sich nicht weiter beeinflussen und bietet bei großen Treffermengen keine Möglichkeiten der Ergebnisstrukturierung.

Ergebnis: Werden die Suchkriterien so gewählt, dass im Anschluss die Trefferliste nicht zu groß wird, sind alle Seiten gut nutzbar. Nutzer mit CIN-Anfragen bekommen in kürzester Zeit Ergebnisse, mit denen sie weiter arbeiten können. Nutzer mit POIN-Anfragen haben es dagegen deutlich schwerer aus größeren Treffermengen Strukturen und Ansätze für die Fokussierung auf ihre Fragestellung herauszulesen. Lediglich bei den dMGH werden interessante Facettierungsansätze und auch die automatische Identifizierung von Personen und Orten gezeigt, deren Ausbau im Hinblick auf geographische und grafische Darstellung wünschenswert wäre.

Mehrfach-Facettierendes Suchinterface – der Spaziergang durch den Informationsdschungel



Beispiel für eine facettierte Trefferanzeige (<http://www.e-codices.unifr.ch/de>)

B. die Datumsangaben der Regesten, die Ausstellung inkl. Geo-Koordinaten und die Verteilung auf die verschiedenen Bände und Abteilungen denkbar. Eine Suchanfrage mit dem einzigen Suchbegriff Heinrich und den damit verbundenen über 17.000 Treffern könnte damit nicht nur in zeitlicher und räumlicher Perspektive sondern auch im Hinblick auf die den Regesta Imperii zugrunde liegende Projektstruktur visualisiert werden, indem zeitliche und räumliche Häufungen sichtbar werden.

Nutzer-beeinflusste Facettierungen: Die eben betrachteten Facettierungen beruhen auf den bereits in der Regestendatenbank der Regesta Imperii vorhandenen Datenstrukturen und könnten durch weitere, vom Nutzer selbst formulierte Facettierungskriterien ergänzt werden. Zu denken ist hier z.B. an Kriterien aus der Computerlinguistik, Begriffshäufungen oder Distanz von Suchbegriffen, ggf. in Verbindung mit regulären Ausdrücken und Trunkierungen.

Fazit

Zusammenfassend bleibt festzustellen, dass die Digitalisierung von mittelalterlichen Quellen für den Forscher große Vorteile mit sich

Im einzelnen werden als Facetten angeboten:

Abteilung (Treffervorkommen in den Abteilungen der MGH)

Reihe (Treffervorkommen in den Reihen der MGH)

Autor/Herausgeber

Jahr (Treffer pro Jahr) Erscheinungsjahr des Bandes

Automatische Personenerkennung (Liste der vorkommenden Personen nach Häufigkeit)

Automatisch Ortserkennung (Deutsch) (Liste der Orte nach Häufigkeit)

Die Durchsicht der automatisch erkannten Personen- und Ortsnamen brachte zwar einige Fehler, ist an sich aber ein interessantes Hilfsmittel.

Deutsche Inschriften: Bei den Deutschen Inschriften gibt es in der Expertensuche die

Möglichkeit die Anzeige der Treffer nach

Datum, Landkreis/Stadt und Standort, auch in verschiedenen Kombinationen, zu

sortieren. Die Trefferanzeige selbst lässt sich nicht weiter beeinflussen und bietet bei

großen Treffermengen keine Möglichkeiten der Ergebnisstrukturierung.

Ergebnis: Werden die Suchkriterien so gewählt, dass im Anschluss die Trefferliste nicht

zu groß wird, sind alle Seiten gut nutzbar. Nutzer mit CIN-Anfragen bekommen in

kürzester Zeit Ergebnisse, mit denen sie weiter arbeiten können. Nutzer mit POIN-

Anfragen haben es dagegen deutlich schwerer aus größeren Treffermengen Strukturen

und Ansätze für die Fokussierung auf ihre Fragestellung herauszulesen. Lediglich bei

den dMGH werden interessante Facettierungsansätze und auch die automatische

Identifizierung von Personen und Orten gezeigt, deren Ausbau im Hinblick auf

geographische und grafische Darstellung wünschenswert wäre.

Die Trefferanzeige der

Papsturkunden.

Die Durchsicht der automatisch erkannten Personen- und Ortsnamen brachte zwar einige

Fehler, ist an sich aber ein interessantes Hilfsmittel.

Deutsche Inschriften: Bei den Deutschen

Inschriften gibt es in der Expertensuche die

Möglichkeit die Anzeige der Treffer nach

Datum, Landkreis/Stadt und Standort, auch in verschiedenen Kombinationen, zu

sortieren. Die Trefferanzeige selbst lässt sich nicht weiter beeinflussen und bietet bei

großen Treffermengen keine Möglichkeiten der Ergebnisstrukturierung.

Ergebnis: Werden die Suchkriterien so gewählt, dass im Anschluss die Trefferliste nicht

zu groß wird, sind alle Seiten gut nutzbar. Nutzer mit CIN-Anfragen bekommen in

kürzester Zeit Ergebnisse, mit denen sie weiter arbeiten können. Nutzer mit POIN-

Anfragen haben es dagegen deutlich schwerer aus größeren Treffermengen Strukturen

und Ansätze für die Fokussierung auf ihre Fragestellung herauszulesen. Lediglich bei

den dMGH werden interessante Facettierungsansätze und auch die automatische

Identifizierung von Personen und Orten gezeigt, deren Ausbau im Hinblick auf

geographische und grafische Darstellung wünschenswert wäre.

SUCHERGEBNIS

IHRE SUCHE HAT 291 TREFFER ERGEBEN. SIE SEHEN DIE TREFFER: 1 BIS 100.

Trefferdarstellung der Deutschen Inschriften (www.inschriften.net).

Datenbankbasierte Facettierungen: Facettierte Suchinterfaces finden heute in den

Onlineshops Anwendung und bieten dem Nutzer die Möglichkeit, große Treffermengen

nach vorgegebenen Kriterien einzuzugrenzen. Dabei werden Facettierungen angeboten,

die in der Regel in der zu Grunde liegenden Datenbank bereits vorhanden sind, wie z.B.

der Preis des Produkts, Hersteller etc. Als Beispiel für die Implementierung einer

facettierten Ergebnisanzeige ist in der Abbildung links die Seite [http://www.e-](http://www.e-codices.unifr.ch/de)

[codices.unifr.ch/de](http://www.e-codices.unifr.ch/de) dargestellt, auf der am rechten Rand die Suchergebnisse einzelnen

Facetten zeitlicher und inhaltlicher Art zugeordnet werden. Dieser Ansatz für die

Trefferanzeige wäre für Nutzer mit POIN-Anfragen interessant, da hier eine große

Treffermenge nicht abschreckt sondern auch überblicksartig den Blick aus anderen

Perspektiven eröffnet. Für die Regesta Imperii wären als Facettierungsmöglichkeiten z.

bringt. So ist der Zugriff auf die Quelle sehr viel schneller als zu "analogen" Zeiten möglich. Mit Hilfe ausgereifter Expertensuchmasken bieten die hier betrachteten Portale schnellen und umfassenden Zugriff auf das Material.

Für einen sinnvollen Umfang der Treffermengen ist bei allen Portalen in der Regel die Nutzung der Expertensuche mit hinreichend einschränkenden Suchkriterien notwendig (CIN-Suche). Für explorative Suchanfragen, die von sich aus größere Treffermengen ergeben sind die Ergebnisanzeigen dagegen weniger geeignet. Auch wenn solche explorativen Suchen sich auf verschiedene der hier untersuchten Datenbanken erstrecken gibt es keine Möglichkeit, die Treffermengen als Ganzes zu untersuchen.

Festzustellen ist, dass projektübergreifende Suchanfragen unter vertretbarem Aufwand nicht über Suchinterfaces der einzelnen Projekte zu realisieren sind. Vielmehr wäre ein zweigleisiges Vorgehen sinnvoll. Zum einen arbeiten die Projekte weiter an der Bereitstellung mediävistischer Quellen im Internet und stellen interessierten Nutzern Schnittstellen für ihre Datenbanken zur Verfügung. Zum anderen arbeiten national oder EU-weit gelagerte Projekte (wie z.B. **DARIAH**) an Tools, mit denen umfangreiche Abfrageergebnisse sinnvoll und transparent visualisiert werden können.

Wie sehr für viele die Nutzung von Google, sei es die Suchmaschine, Google-Maps oder andere Dienste im Alltag schon unentbehrlich geworden ist, können die meisten in einer kritischen Selbstreflexion ergründen. Google liefert oft bei geringem Aufwand einen sehr guten Überblick zu den Suchergebnissen, wobei der Weg zu den Ergebnissen nicht sehr transparent ist. Meine Vorstellung von "digitalen Perspektiven" wäre daher die Formulierung neuer Fragestellungen an die digitalisierten Quellen mit Hilfe moderner, transparenter Suchtechniken, ergänzt um leicht und intuitiv verständliche Visualisierungsmethoden.

[Download PDF](#)

Empfohlene Zitierweise:

Andreas Kuczera: Digitale Perspektiven mediävistischer Quellenrecherche, in: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 18. April 2014, <http://mittelalter.hypotheses.org/3492> (ISSN 2197-6120).

Die CIN/POIN-Systematik wurde entwickelt von Frants, Valery I.; Shapiro, Jacob; Voiskunskii, Vladimir G.: Automated information retrieval: Theory and methods. Library and information science. San Diego: Academic Press 1997. [↔]

Vgl. hierzu Wolfgang Stock, Information Retrieval: Informationen suchen und finden. München 2007, dessen Tabelle auf S. 52 als Vorlage diente. [↔]

Eine intensive qualitative Analyse mit vergleichenden Suchanfragen an die verschiedenen Seiten würde den Rahmen dieser Untersuchung sprengen. Vielmehr standen beim Vergleich die dem Nutzer direkt dargebotenen Funktionen im Vordergrund. [↔]

Im Einzelnen kann man nach z. B. Jaffé-Nr., Pontificia-Nr., Regesta-Imperii-Nr., Pontifiakt, Papstunterschrift, Notarsunterschriften, Zeugen, Siegel, Überlieferung, Diplomatischer Kommentar und Überlieferung filtern. [↔]

Relevanz meint in diesem Fall die Relevanzkriterien der implementierten Suchmaschine **SPHINX**. [↔]

Zu den Relevanzkriterien heißt es in der Hilfe: Die Relevanz berechnet sich durch die mathematische Ähnlichkeit von Suchanfrage und Dokument (hier Buch). Die Bedeutung einzelner Wörter hängt von ihrer Häufigkeit und der Größe des Dokumentes ab. [↔]
