

## **Der eHumanities Desktop als Werkzeug in der historischen Semantik: Funktionsspektrum und Einsatzszenarien**

---

Die *Digital Humanities* bzw. die *Computational Humanities* entwickeln sich zu eigenständigen Disziplinen an der Nahtstelle von Geisteswissenschaft und Informatik. Diese Entwicklung betrifft zunehmend auch die Lehre im Bereich der geisteswissenschaftlichen Fachinformatik. In diesem Beitrag thematisieren wir den *eHumanities Desktop* als ein Werkzeug für diesen Bereich der Lehre. Dabei geht es genauer um einen Brückenschlag zwischen Geschichtswissenschaft und Informatik: Am Beispiel der historischen Semantik stellen wir drei Lehrszenarien vor, in denen der *eHumanities Desktop* in der geschichtswissenschaftlichen Lehre zum Einsatz kommt. Der Beitrag schließt mit einer Anforderungsanalyse an zukünftige Entwicklungen in diesem Bereich.

### **1 Einleitung**

In einer Zeit aufkommender Bestrebungen um die Entwicklung einer *Digital Humanities* als wissenschaftliche Disziplin mit eigenem Erkenntnisinteresse und eigenen Untersuchungsmethoden nimmt es nicht Wunder, dass die Ergebnisse dieser Arbeit zunehmend auch im Bereich der Lehre erprobt werden. In diesen Kontext ordnen wir ähnlichlautende Bestrebungen um die Entwicklung des *Computing in the Humanities* bzw. der *Computational Humanities* ein. Der Unterschied dieser Bestrebungen resultiert aus einer stärkeren Betonung der computerbasierten Repräsentation geisteswissenschaftlicher Interpretationsstrukturen als zentrale Aufgabe der *Digital Humanities* gegenüber einer stärkeren Betonung der Algorithmisierung von Explorationen solcher Strukturen als zentrale Aufgabe der *Computational Humanities* (vgl. hierzu den Workshop von Heyer and Büchler, 2010). Während also geisteswissenschaftlich ausgerichtete Informatiker im Bereich der *Digital Humanities* Interpretationsresultate computerbasiert modellieren und verarbeiten, sind es Informatiker im Bereich der *Computational Humanities*, die auf die automatische Exploration solcher Interpretationsresultate zielen. Wegen des Fokus auf lehrorientierte Nutzungsszenarien des vorliegenden Themenhefts vernachlässigen wir diesen Unterschied im Rahmen dieses Beitrags und sprechen unterschiedslos von geisteswissenschaftlicher Fachinformatik, um diese beiden Bereiche zu denotieren.

Gemäß dem interdisziplinären Grundcharakter all dieser Ansätze steht ebenso außer Frage, dass die geisteswissenschaftliche Fachinformatik nicht allein die Lehre in der Informatik selbst betrifft. Vielmehr geht es darum, lehrorientierte Anwendungsfelder

insbesondere in den geisteswissenschaftlichen Einzeldisziplinen aufzutun. Hierzu zählen die sprachwissenschaftlichen Disziplinen, die infolge ihres frühen Bezugs zur Computerlinguistik längst im Austausch mit der Informatik stehen, ebenso wie jene geisteswissenschaftlichen Disziplinen, für welche die Anwendung computerbasierter Methoden nach wie vor eine Seltenheit darstellt. Zum gegenwärtigen Zeitpunkt ist die Geschichtswissenschaft weitgehend letzterem dieser beiden Lager zuzurechnen.

In diesem Kontext thematisiert das vorliegende Papier einen Brückenschlag ausgehend von der Informatik hin zur textorientierten Geschichtswissenschaft und umgekehrt. Damit bleibt die Anwendbarkeit des zu umreißenden Instrumentariums auch in anderen Disziplinen unbestritten.<sup>1</sup> Der Erfahrungshintergrund der Lehre in der geisteswissenschaftlichen Fachinformatik, der hier skizziert werden soll, betrifft jedoch vorrangig die Geschichtswissenschaft.

Allgemein gesprochen stehen wir bei diesem Brückenschlag vor einer Aufgabe, die in Abbildung 1 ihren schematisierten Niederschlag findet und wie folgt umschrieben werden kann:

1. Das eine Extrem bildet jene Auffassung von Informatik, für welche die Orientierung an formalisierbaren Repräsentationen und berechenbaren Algorithmen zur Erfassung von Strukturen und Prozessen des jeweiligen Gegenstandsbereichs kennzeichnend ist. Was nicht algorithmisierbar oder gar nicht formalisierbar ist, steht außerhalb dieser Auffassung des Gegenstandsbereichs der Informatik – dieser Lesart entspricht in etwa die klassische Auffassung der Kognitionswissenschaft bei Fodor and Pylyshyn (1988). Im Idealfall, sozusagen fernab von der Komplexität semiotischer Prozesse (Petitot et al., 2000; Rieger, 2003), ist die Informatik allein mit vollständig automatisierbaren Prozessen befasst, deren Qualitätsmessung strikten Regeln folgt und also nach dem Muster überwachter Lernprozesse abläuft.
2. Das andere Extrem bildet jene Auffassung von Geisteswissenschaft diesseits der Anwendung computerbasierter Methoden, die sich am Ideal einer hermeneutischen, rein intellektuellen, verstehensgeleiteten Wissenschaft orientiert, deren Erklärungsmodelle wegen ihres konstruktivistischen Gehalts im Extremfall nicht erschöpfend reproduzierbar sind. Die Aktualisierung solcher Modelle setzt vielmehr jene Art von Interpreten als unabdingbares Aktualisierungsmedium voraus, das keines seiner Verstehensschritte je an eine Maschine delegiert noch delegieren könnte.
3. In diesem Spannungsfeld ist die geisteswissenschaftliche Fachinformatik nun – vergleichbar mit der Computersemiotik (Rieger, 1989; Andersen, 1990) – als der Versuch aufzufassen, eine Position zwischen diesen Extremen einzunehmen, und zwar entlang zweier Orientierungen: dies betrifft zum einen die Annahme von der Formalisierbarkeit und Algorithmisierbarkeit nicht trivialer, erkenntnisfördernder Teile semiotischer Prozesse sowie die Forderung nach der Interpretierbarkeit

<sup>1</sup>Siehe beispielsweise Mehler et al. (2011) für eine Anwendung im Bereich der Romanistik oder Gleim et al. (2010) für eine Anwendung im Bereich der Bildwissenschaft.

jeglicher Formalisierungs- und Algorithmisierungsergebnisse, und zwar so, dass diese an die hermeneutische Tradition anknüpfbar sind. Einer solcherart verstandenen geisteswissenschaftlichen Fachinformatik geht es folglich darum, den Bereich jener Methoden zu vergrößern, die beides zugleich sind: formalisier- und berechenbar wie auch geisteswissenschaftlich interpretierbar.

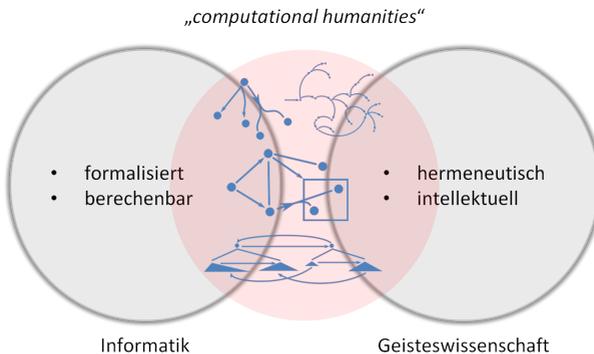
Offenbar adressiert die Umsetzung letzteren Programms eine Art von Algorithmus, der – im Sinne des *Human Computation* (von Ahn, 2008) – ein enges Wechselspiel von Automatisierung und geisteswissenschaftlicher Interpretation voraussetzt – fernab vom Ideal überwachten Lernens, aber auch fernab vom Paradigma des unüberwachten Lernens insofern dieses lediglich auf dem Mechanismus des *Relevance Feedback* beruht.<sup>2</sup>

Diesem Ansatz gemäß ist die geisteswissenschaftliche Fachinformatik genauso wenig eine Hilfswissenschaft irgendeiner Geisteswissenschaft wie auch umgekehrt geisteswissenschaftliche Objekte nicht einfach den Gegenstandsbereich der ansonsten unverändert bleibenden Informatik erweitern. Vielmehr ist die Vorstellung leitend, dass die Anwendung informationswissenschaftlicher Methoden nur dann *Humanities* als *Computational* kennzeichnet, wenn sie den Erkenntnishorizont der betroffenen Geisteswissenschaften erweitert, ohne bereits mit ihrem herkömmlichen, nicht-informationswissenschaftlichen Apparat erbringbar zu sein. Es geht also um eine Art von geisteswissenschaftlicher Disziplin, deren Gegenstandsbereich und Erkenntnisinteresse durch computerbasierte Methoden überhaupt erst ihre Prägung erhalten, und zwar jenseits etablierter Geisteswissenschaften. Als ein Beispiel für eine solche Blickrichtung können intertextuelle Strukturen gelten, die allein an solchen Textmengen beobachtbar sind, deren Größe anders als computerbasiert nicht zu bewältigen ist. In diesem Beispiel wird Intertextualität als Untersuchungsobjekt erst fassbar durch die computerbasierte Modellierung – und zwar in qualitativer wie auch quantitativer Hinsicht (Wagner et al., 2009).

Blickrichtungen dieser Art ist der eHumanities Desktop und seine Anwendung in Forschung und Lehre der Geschichtswissenschaft geschuldet. Im Folgenden wird dies an der historischen Semantik und ihrer Vermittlung in der Lehre exemplifiziert. Das Papier ist wie folgt organisiert: Sektion 2 beschreibt die historische Semantik als ein Teilgebiet der Geschichtswissenschaft. Dies betrifft ihr Erkenntnisinteresse ebenso wie ihre vorherrschende korpusbasierte Methodik. Sektion 3 erläutert eine Auswahl des Funktionsspektrums des eHumanities Desktops, jenes Systems also, das die Geschichtswissenschaft an der Goethe-Universität Frankfurt im Rahmen von drei Lehrszenarien eingesetzt hat bzw. einsetzt. Diese Szenarien werden in Sektion 4 beschrieben. Sektion 5 folgt mit einer Diskussion von Anforderungen, die sich aus diesen Erfahrungen an die Weiterentwicklung des eHumanities Desktop ergeben, ehe Sektion 6 mit einer Zusammenfassung und Schlussfolgerung schließt.

---

<sup>2</sup>Siehe Wagner et al. (2009) für einen solchen Ansatz im Bereich von *literary memory information systems*.



**Abbildung 1:** Die geisteswissenschaftliche Fachinformatik im Spannungsfeld von Informatik und Hermeneutik.

## 2 Historische Semantik: methodisch-theoretischer Hintergrund

In dieser Sektion skizzieren wir den erkenntnistheoretischen Rahmen, auf dem die Historische Semantik und damit letztlich die von Historikern angestrebte Nutzung des eHumanities Desktops ruht. Unter Verwendung der in den historischen Wissenschaften gängigen Fachterminologie wird dabei aufgezeigt, welchen Erkenntnisfortschritt der Einsatz des eHumanities Desktops – ob nun in Forschung oder Lehre – maßgeblich unterstützen soll. Dieser Fortschritt setzt an einer methodischen Verbindung von datengetriebener Korpuslinguistik, statistischer Textlinguistik und interpretatorischer Hermeneutik an, und macht somit die Zusammenarbeit von Historikern und Texttechnologern unabdingbar.

Die Historische Semantik widmet sich innerhalb der Geschichtswissenschaft der Erforschung von Bedeutung und insbesondere der Erforschung ihrer historischen Variabilität. Dabei nimmt sie verschiedene Medien und Objekte in den Blick, die sich zur Interpretation von Sinnzuschreibungsprozessen eignen. Hier geht es zunächst um eine textuell geprägte Historische Semantik.

Eine der Grundannahmen ist, dass Bedeutung jeweils im Gebrauch entsteht und immer wieder aktualisiert werden muss. So ist Bedeutung vor allem auch situativ variabel. Der Blick der Forschung richtet sich daher nicht allein auf bestehende Verknüpfungen von Zeichen und Sinn, sondern auf die Mechanismen der Sinnzuschreibung – Rieger (2003) spricht an dieser Stelle genauer von Prozessen der Bedeutungskonstitution. Diese gelten als von der Gesellschaft kontrollierbar, da sie es ist, die mit Hilfe von Sprache Sinn erzeugen muss. Es schließt sich die Annahme an, dass sprachliche Phänomene und Sinnzuschreibungsprozesse auch Aufschluss geben über ihre gesellschaftlichen Bedingungen. Eine der zentralen Fragen der Historischen Semantik ist die nach dem Verhältnis von Sprachwandel und gesellschaftlichen Wandlungsprozessen.

Die Geschichtswissenschaft kennt seit dem *linguistic turn* der 1970er Jahren verschiedene Traditionen, die sich mit der Erforschung von Sinn beschäftigt haben. Für

Deutschland ist in erster Linie die *Begriffsgeschichte* nach Reinhart Koselleck prägend gewesen. Koselleck versammelte eine Reihe von Kollegen um sich, mit denen er die *Geschichtlichen Grundbegriffe* schrieb, heute ein Standardwerk in der Geschichtswissenschaft. In Form von Handbucheinträgen widmeten sich die Autoren der politisch-sozialen Sprache, wie Koselleck (1972, XIII) sagt, den "Leitbegriffe[n] der geschichtlichen Bewegung, die, in der Folge der Zeiten, den Gegenstand der historischen Forschung ausmacht." Im Zentrum stehen dabei solche Begriffe, die aus Sicht des modernen Wissenschaftlers/der modernen Wissenschaftlerin gesellschaftliche Relevanz beanspruchen, und deren Geschichte. Dabei wird die Untersuchung des Begriffs immer rückgebunden an die sozialen Gegebenheiten einer Zeit. Daraus ergibt sich ein Geflecht aus Sprache und ihren gesellschaftlichen Bedingungen sowie, anders herum, aus gesellschaftlichen Strukturen und ihren sprachlichen Konstruktionen und Modalitäten.

Um den Blick der Moderne auf die Geschichte ihrer Begriffe zu lenken, hat die Historische Semantik ihre Methoden immer weiter verfeinert und auch quantitative Vorgehensweisen in ihr Konzept integriert. Ziel war die Abkehr von der Untersuchung verfassungsgeschichtlicher Termini und deren Geschichte, und zwar deutlicher als dies in den Anfängen der *Begriffsgeschichte* geschehen ist. Außerdem versprach der quantitative Zugriff eine Verbesserung der Materialbasis für sprachliche Analysen wie er von den Kritikern der Koselleckschen Schule und auch in der *Conceptual History* (vgl. Skinner, 2003) gefordert wird. Ausgehend von den Forschungsinteressen der Pioniere der Historischen Semantik – oder der Methoden zur sprachlichen Analyse gegenwärtiger und vergangener Wirklichkeitskonstruktionen – haben sich entsprechende Untersuchungen lange Zeit auf Korpora der Moderne bzw. der Frühen Neuzeit konzentriert. Steinmetz (2008) vermittelt einen umfassenden Überblick zu dieser Forschung und deren Entwicklung in den letzten 40 Jahren. Auch wenn die *Geschichtlichen Grundbegriffe* in ihren Artikeln immer einen Blick auf die Vormoderne (Antike und Mittelalter) geworfen haben, blieb es doch lange bei eben diesem kurzen Blick.<sup>3</sup> Es sind aber gerade historische Langzeitkorpora, die im Rahmen historisch-semantischer Versuchsaufbauten besonders interessante Ergebnisse versprechen. Für Sprachwandeluntersuchungen bietet es sich beispielsweise an, diachron angelegte Textkorpora heranzuziehen, die für die Vormoderne wesentlich längere Zeiträume abdecken als für die Moderne. Vor allem die quantitative Analyse kann von der Untersuchung solcher Langzeitkorpora profitieren (Mehler et al., 2010, 2011). Sie gewinnt in diesem Rahmen eine breite Materialbasis.

Ausgehend von der *Begriffsgeschichte* können so semasiologische, diachron angelegte Untersuchungen der Verwendungsmuster bestimmter Wörter konzipiert werden. Der engere methodische Zugriff ist dabei eine quantifizierende Kookkurrenz-Untersuchung im Umfeld der Gebrauchssituationen der ausgewählten Vokabel. Innerhalb spezifischer Verwendungsmuster markieren die neben der Zielvokabel gebrauchten Wörter den situativ aktualisierten Sinn. Im Zeitalter der *Digital Humanities*, also der zunehmenden Verfügbarkeit großer Textvolumina in digitaler Form, steigen auch die Möglichkeit en computergestützter, korpusorientierter Methoden. Für die serielle Untersuchung

---

<sup>3</sup>Einen Überblick über semantische Studien für die Vormoderne bietet Rexroth (2009).

von Gebrauchskontexten bestimmter Vokabeln oder anders gestalteter sprachlicher Muster können so Quantifizierungen im großen Stil erstellt werden.<sup>4</sup> Korpora, die für solche Studien herangezogen werden können, liegen quer zu den vorhandenen digitalen Editionen mittelalterlicher Texte wie etwa der *Patrologia Latina* (PL) (Migne, 1855), der *digitalen Monumenta Germaniae Historica* (dMGH) bzw. der *elektronischen Monumenta Germaniae Historica* (eMGH) (Radl, 2007). Diese Sammlungen stellen die Ressourcen zur Verfügung, aus denen Korpora erstellt werden müssen, die sich an den spezifischen Anforderungen je verschiedener Forschungskonzepte und -fragen orientieren (vgl. Jussen 2006). Ziel ist es, aus den Gebrauchssituationen bestimmter Sprachmuster eine “kollektive Haltung” jenseits einzelner Autoren und ihrer Texte zu erarbeiten. Um diese dann in diachroner Perspektive betrachten zu können und sie gleichzeitig mit den Hypothesen der politischen Ideengeschichte abzugleichen, ist der Untersuchung ein Textkorpus zu Grunde zu legen, das es ermöglicht, homogene Beobachtungstrecken über längere Zeiträume zu erstellen (Jussen, 2006).<sup>5</sup> Ein solches Korpus besteht vorzugsweise aus einzelnen Werken einer bestimmten Textsorte.<sup>6</sup> Rieger (1989) spricht in diesem Zusammenhang von dem Kriterium der pragmatischen Homogenität der den Gebrauchsanalysen zugrundezulegenden Korpora.

Neben der Entwicklung neuer Methoden und Werkzeuge bleibt es eine Herausforderung für die Historische Semantik, ihre Ergebnisse zu kontextualisieren. Für diese Aufgabe werden Theorien herangezogen, die sich dem Zusammenspiel von Sprache und Sozialstruktur widmen (vgl. Luhmann, 1993). Semantik kann so verstanden werden als “Vorrat an bereitgehaltenen Sinnverarbeitungsregeln” und die Gesellschaftsstruktur als Rahmen, der die Beliebigkeit dieses Vorrats für die Verarbeitung von Sinn einschränkt (Luhmann, 1993, 19). Semantik liefert also ein breites Spektrum an möglichen Regularitäten zur Verarbeitung für Umstände, Situationen und Handlungszusammenhänge, denen die Gesellschaft Sinn zuschreibt. Eine Untersuchung dieser sprachlich verfassten Phänomene verspricht neue Einsichten in die Strukturen historischer Gesellschaften. Für die Geschichtswissenschaft ergeben sich daraus Beschreibungsmuster und Deutungsschemata für die Vergangenheit, die in Dialog treten können mit gegenwärtigen Mechanismen der Weltdeutung.

### 3 Zum Funktionsspektrum des eHumanities Desktops

Der erkenntnistheoretisch-methodologische Grundriss, den die vorangehende Sektion von der historischen Semantik gegeben hat, steht aus computerlinguistischer Sicht in enger

<sup>4</sup>Vgl. zur Darstellung der technologischen Anforderungen Jussen et al. (2007) und Mehler et al. (2010).

<sup>5</sup>In einem zweiten Schritt würden diese dann untereinander verglichen, um textsortenspezifische Semantiken herausarbeiten zu können, die wiederum dem Konzept der “languages” in der *Conceptual History* ähneln.

<sup>6</sup>Im Rahmen des LOEWE-Schwerpunkt *Digital Humanities* (<http://digital-humanities-hessen.de>) wird in Kooperation mit dem BMBF-Projekts *Linguistic Networks* (<http://project.linguistic-networks.net>) derzeit der Versuch unternommen, alle Texte der *Patrologia Latina* nach Textsorten zu kategorisieren.

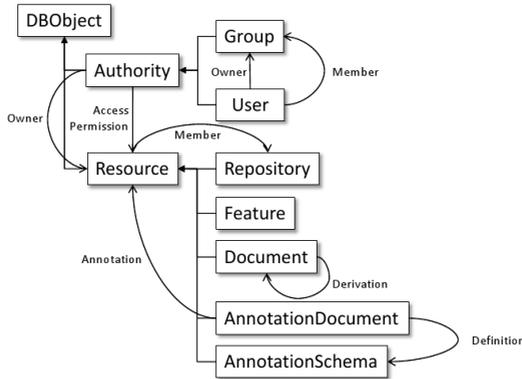
Nachbarschaft zur explorativen Korpuslinguistik (Rieger, 1998; Heyer et al., 2006; Evert, 2008). Folgerichtig ist das Methodenspektrum zur Bedienung dieses Erkenntnisinteresses durch kollokationsstatistische Methoden vorgezeichnet. In diesem Kontext beschreiben wir nun eine Auswahl von Funktionalitäten, welche der eHumanities Desktop für Forschung und Lehre bereitstellt (Gleim et al., 2009a,b, 2010): Ausgehend von dem zentralen Administrationsmodul des eHumanities Desktops (Sektion 3.1) betrifft dies den *Corpus Manager* (Sektion 3.2) für die Ressourcenverwaltung, das frei konfigurierbare Modul für die Annotation Ressourcen-orientierter Metadaten (Sektion 3.3), das zentrale Vorverarbeitungstool des Desktops (Sektion 3.4), das HSCM-Modul (Sektion 3.5), den *Lexicon Browser* (Sektion 3.6) als zentrale Schnittstelle für das Lexikon-orientierte Retrieval und schließlich das *Classifier Builder* genannte Modul, mit Hilfe dessen vektorbasierte SVM-Modelle gebaut werden können. Nachfolgend werden diese Module erläutert.

### 3.1 Das Administrator-Modul

Der produktive Einsatz eines Systems zur kollaborativen Korpusverwaltung und -analyse erfordert eine entsprechende Rechteverwaltung. Diese soll den Anforderungen einzelner Arbeitsgruppen bis hin zu verteilten Forschungsprojekten gerecht werden und die Kollaborationsstrukturen solcher Projekte widerspiegeln. Abbildung 2 zeigt schematisch die Stammdatenverwaltung des eHumanities Desktop. Grundsätzlich wird unterschieden zwischen Autoritäten und den Ressourcen, hinsichtlich derer diese Berechtigungen besitzen. Eine Autorität ist zunächst ein abstrakter Begriff, der Benutzer und Benutzergruppen umfasst. Ein Benutzer kann beliebig vielen Gruppen angehören und erhält dadurch deren Berechtigungen. Für jede Benutzergruppe ist ein Eigentümer definiert (z.B. der Projektleiter einer Gruppe), welcher die Mitgliedschaften unabhängig vom zentralen Administrator verwalten kann.

Eine Ressource ist wiederum ein Sammelbegriff, der vor allem Repositorien und Dokumente umfasst. Die Berechtigungen auf Ressourcen können abgestuft gesetzt werden, ausgehend von reinen Leserechten, über Schreib- und Löschberechtigungen bis hin zu Freigabeberechtigungen. In der Praxis bedeutet dies, dass zum Beispiel Mitglieder einer Arbeitsgruppe Textsammlungen aufbauen und bearbeiten können und im Rahmen von Kooperationen die Möglichkeit erhalten, Projektpartnern Leseberechtigung auf den Daten zu vergeben. Diese Zugriffskontrolle ist vor allem für die Arbeit mit urheberrechtlich geschütztem Material relevant. Darüber hinaus besteht die Möglichkeit, nicht nur Dokumente und Repositorien, sondern auch Programmfunktionen freizugeben bzw. zu sperren. Auf diese Weise kann einer Anwenderin bzw. einem Anwender gezielt der Funktionsumfang bereitgestellt werden, der für ihre bzw. seine Arbeit relevant ist. Mit dem Anwendungsmodul *Administrator* steht somit ein Werkzeug zur Verfügung, Einzelbenutzer- und Benutzergruppenberechtigungen zu verwalten.

Genau diese Art der Berechtigungsverwaltung dient nicht nur zur Unterstützung von Forschungsprojekten, sondern erweist sich auch im Zusammenhang von Lehrveranstaltungen von großem Nutzen. Dies betrifft die Binnendifferenzierung der Zugänglichkeit einzel-



**Abbildung 2:** Das Modell der Stammdatenverwaltung des eHumanities Desktops.

ner Ressourcen ebenso wie die Freistellung oder Einschränkung von Einzelstudierenden- oder Studierendengruppen-bezogenen Funktionalitäten.

### 3.2 Das Corpus Manager-Modul

Der *Corpus Manager* ist ein zentrales Modul, das den webbasierten Aufbau von Korpora sowie die Verwaltung von Ressourcen ermöglicht. Dokumente können entweder direkt vom Arbeitsplatzrechner aus hochgeladen oder durch die Angabe einer URL aus dem Internet heruntergeladen werden. Umgekehrt ist es möglich, von jedem Rechner mit Internetanbindung über den Webbrowser auf sämtliche Daten und Anwendungen des eHumanities Desktop zuzugreifen. Der Desktop eignet sich damit insbesondere auch für den gruppenarbeitsorientierten Betrieb in ganzen Rechnerräumen.

Für viele Dokumenttypen und -arten stellt der Desktop Programme zur Betrachtung bereit, etwa für TEI P5, HTML und auch für Bilder. Der graphische Aufbau sowie die Benutzerführung orientiert sich dabei am Windows Explorer zur Dokumentverwaltung. Anders als etablierte Dateisysteme setzt der eHumanities Desktop jedoch nicht auf eine feste Verbindung von Dokumenten und Verzeichnissen. Vielmehr kann ein Dokument in verschiedenen Repositorien verwendet werden, ohne dass das Dokument dafür jedesmal kopiert werden muss. Das bietet den Vorteil, dass unterschiedliche Korpora auf der gleichen Dokumentbasis speichereffizient aufgebaut werden können. Wird ein Dokument bearbeitet oder mit Zusatzinformationen annotiert, so sind diese automatisch in allen Repositorien verfügbar, denen das betreffende Dokument angehört. Die Annotation von Dokumenten, welche im *Corpus Manager* durchgeführt werden kann, umfasst mehrere Arbeitsschritte, die im Folgenden skizziert werden.

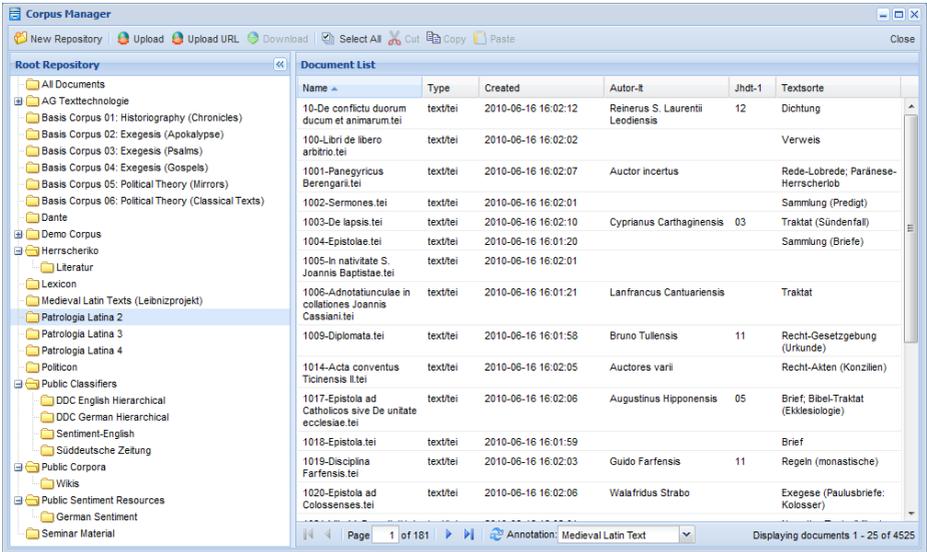


Abbildung 3: Ein Screenshot des *Corpus Manager* des *eHumanities Desktop*.

### 3.3 Das Annotator-Modul

Die Auszeichnung von Dokumenten mit Metadaten ist ein zentraler Baustein des Aufbaus und der anschließenden Analyse von Korpora. Der eHumanities Desktop trägt dieser Aufgabe durch Bereitstellung umfangreicher Funktionen Rechnung. So haben Anwender die Möglichkeit, selbstständig Datenfelder im Rahmen von Annotationsschemata zu definieren, um auf dieser Grundlage ihre Ressourcen auszuzeichnen. Es werden also keine Standardschemata vorgegeben. Nichtsdestotrotz reicht die Ausdrucksmächtigkeit des Systems dazu aus, gängige Standards, wie etwa *Dublin Core*<sup>7</sup>, abzubilden.

Abbildung 4 zeigt einen Screenshot des *Schema Editor* des Desktops, in dem ein Datenfeld des Schemas "Medieval Latin Text" bearbeitet wird. Neben einem Namen kann unter anderem der Datentyp (Text, Zahl, Datum etc.) und dessen Wertebereich festgelegt werden. Abbildung 5 zeigt die Auszeichnung eines Dokuments auf der Grundlage dieses Schemas. Die Datenfelder von Annotationen können mittels des *Corpus Manager* auch in Tabellenform angezeigt, durchsucht und sortiert werden. Dies wird in Abbildung 3 exemplifiziert. Aus der Sicht der Datenmodellierung werden Annotationen von Ressourcen ebenfalls als Ressource betrachtet und fallen somit unter die Rechteverwaltung. Es steht damit jeder Anwenderin bzw. jedem Anwender frei, eine Ressource zu annotieren und diese Annotation mit anderen Anwendern zu teilen. Auf diese Weise

<sup>7</sup><http://dublincore.org>

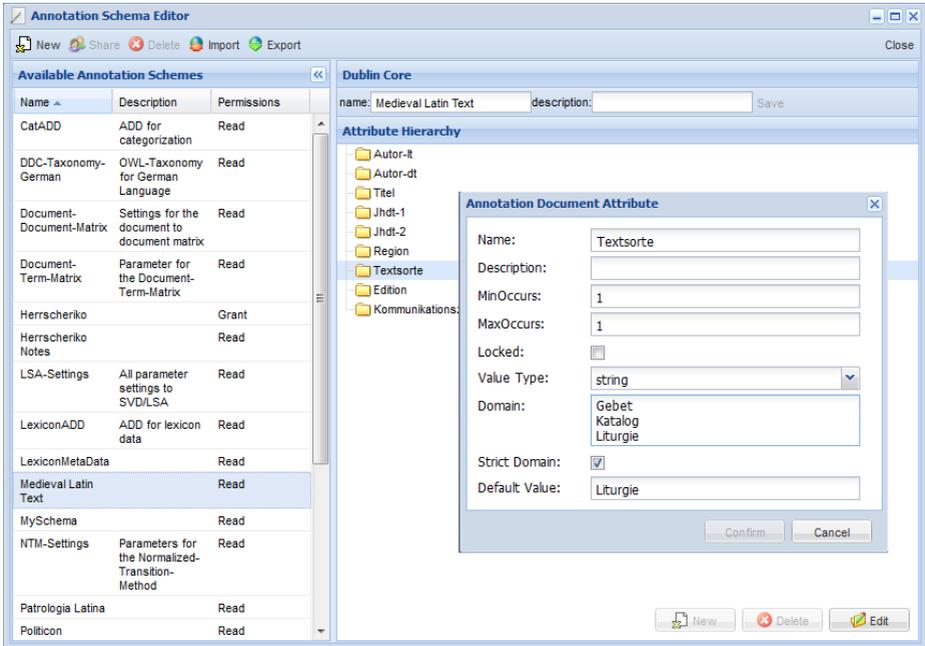


Abbildung 4: Ein Screenshot des *Annotation Schema Editor*.

wird eine kollaborative Dokumentauszeichnung unterstützt, was wiederum in der Lehre im Rahmen der kollaborativen Bearbeitung derselben Korpora eingesetzt werden kann.

### 3.4 Das Preprocessor-Modul

Für den sinnvollen Einsatz von Texten in korpuslinguistischen Studien ist eine linguistische Vorverarbeitung unabdingbar. Diese umfasst in der Regel die Tokenisierung, Lemmatisierung und Wortartenauszeichnung sowie die Erkennung logischer Dokumentstrukturen. Zur Repräsentation solcherart ausgezeichnete Dokumente verwendet der eHumanities Desktop den Dokumentrepräsentationsstandard der *Text Encoding Initiative*, das heißt TEI P5 (TEI Consortium, 2010). Darüber hinaus bietet das System mit dem *Preprocessor* ein Funktionsmodul, mit Hilfe dessen Anwender selbstständig Rohdaten (z.B. im PDF-, RTF- oder HTML-Format) vorverarbeiten und nach TEI P5 überführen können.

Aktuell wird die Verarbeitung von deutschen, englischen sowie von lateinischen Texten unterstützt (Waltinger, 2010; Mehler et al., 2008). Über die Benutzerschnittstelle können beliebig viele Eingangsdokumente im Stapelbetrieb automatisch verarbeitet werden.

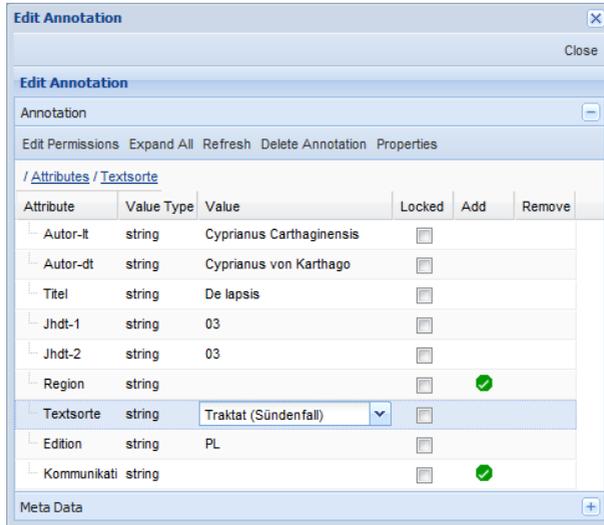
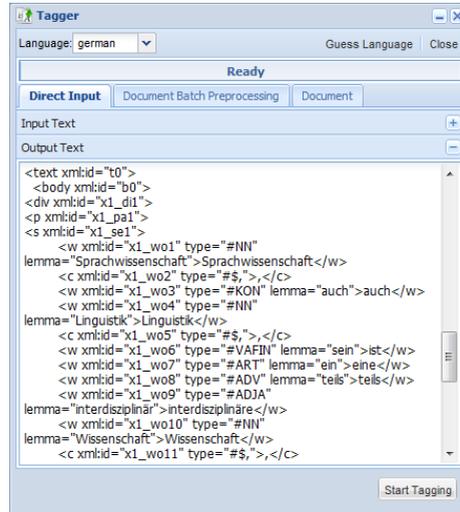


Abbildung 5: Ein Screenshot des ressourcenorientierten Annotationsdialogs.

Darüber hinaus können auch Texte direkt in ein Textfeld kopiert und vorverarbeitet werden. Abbildung 6 zeigt exemplarisch die Ausgabe einer solchen Verarbeitung in TEI P5.

### 3.5 Das Historical Semantics Corpus Management-System

Das *Historical Semantics Corpus Managements-System* (HSCM) zielt auf die text-technologische Repräsentation und quantitative Analyse zeitlich geschichteter Corpora (Mehler et al., 2010). Es ermöglicht die Durchsuchung von Textsammlungen nach Belegstellen für einzelne Wörter oder Phrasen. Neben einer Auflistung von Texten, welche der Suchanfrage entsprechen, können gezielt Treffersätze und deren Kontext angezeigt werden. Die erste Version des HSCM war als eigenständige Client/Server Anwendung konzipiert und für die Verarbeitung der *Patrologia Latina* hin optimiert worden (Jussen et al., 2007). Der eHumanities Desktop inkludiert nunmehr die PL in einer Fassung, die aufgrund ihres Annotationsumfangs weit über die ursprüngliche SGML-basierte Ausgabe hinausgeht (Mehler et al., 2009). Die Dokumente dieses Korpus wurden mit einer Reihe von Metadaten ausgezeichnet, die insbesondere deren Entstehungszeitraum thematisieren. Mit der Integration in den eHumanities Desktop ist nun auch die Arbeit mit beliebigen, in TEI P5 repräsentierten Textsammlungen möglich. Die Syntax für die Suchanfragen ist intuitiv und leicht erlernbar. Intern werden die Anfragen in *Apache*



**Abbildung 6:** Ein Screenshot des *Preprocessor* in Zusammenhang mit einer Beispielausgabe basierend auf TEI P5.

*Lucene*<sup>8</sup> übersetzt. Die vormalige Verwendung einer XML-Datenbank hat sich zwischenzeitlich als deutlich weniger effizient erwiesen. Abbildung 7 zeigt einen Screenshot des HSCM mit den Ergebnissen einer Suchanfrage nach “Helmut Schmidt” in einem Korpus von Artikeln der “Süddeutschen Zeitung”.

### 3.6 Der Lexicon Browser

Neben Texten und Bildern können auch Lexika als Dokumente im eHumanities Desktop einheitlich verwaltet werden. Der *Lexicon Browser* bietet eine Schnittstelle, die Anwender bei der Durchsuchung und Bearbeitung von Lexika unterstützt. Als Grundlage hierfür dient die einheitliche Repräsentation lexikalischer Ressourcen mit Hilfe des objektorientierten Datenmodells des eLexicon (Mehler et al., 2008). Für jedes Lexikon kann zudem ein Referenzkorpus angegeben werden, auf dessen Basis eine Reihe von Maßen zur Berechnung der Trennschärfe einzelner Lexikoneinträge berechnet wird. Dazu zählen die Frequenz, Textfrequenz und die inverse Dokumentfrequenz.

Da die Einträge der Lexika mit einer umfangreichen Ontologie von Wortarten ausgezeichnet sind, ermöglicht dies zusammen mit den Maßen eine differenzierte Suche nach Einträgen. Abbildung 8 zeigt einen Screenshot des Lexicon Browsers mit Ergebnissen einer Suchanfrage auf einem lateinischen Lexikon.

<sup>8</sup>Siehe <http://lucene.apache.org>

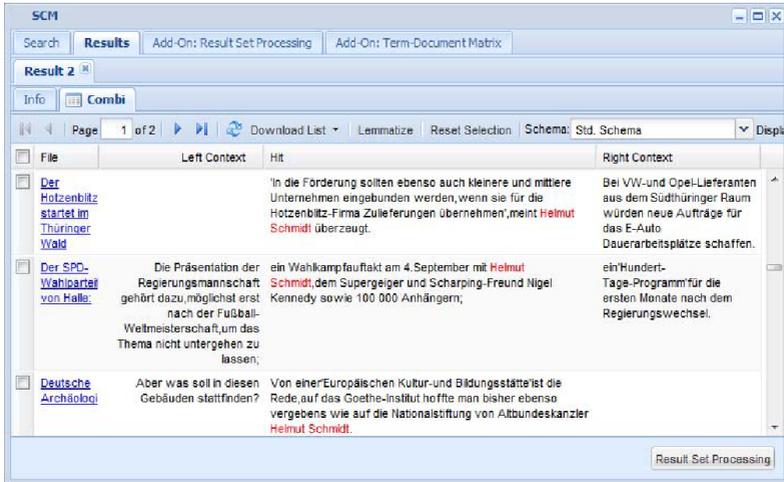


Abbildung 7: Ein Screenshot des HSCM mit den Ergebnissen einer Suchanfrage nach "Helmut Schmidt" im Korpus von Artikeln der Süddeutschen Zeitung.

### 3.7 Der Classifier Builder

Der *Categorizer* des eHumanities Desktop dient dazu, Textdokumente oder eingefügte Textausschnitte zu kategorisieren. Die Menge der hierfür verfügbaren Klassifikatoren ist nicht fest vorgegeben. Vielmehr können mit dem *Classifier Builder* eigene Klassifikatoren erstellt werden, um diese zur überwachten Textkategorisierung im Rahmen des Categorizer-Moduls heranzuziehen. Abbildung 9 exemplifiziert das Ergebnis der Kategorisierung eines Textausschnitts auf Basis der englischen *Dewey Decimal Classification* (DDC) (Waltinger et al., 2011). Die zutreffenden Kategorien sind nach ihrer Signifikanz absteigend sortiert und farblich abgesetzt.

## 4 Einsatzszenarien und Ziele in der Lehre

Nachdem nun der methodologische und erkenntnistheoretische Rahmen für Nutzung des eHumanities Desktops in der Geschichtswissenschaft skizziert wurde, geht es in dieser Sektion darum, Nutzungsszenarien des Desktops in der Lehre in der Geschichtswissenschaft zu beschreiben.

Die Historische Semantik ist an der Goethe-Universität bislang in drei Lehrformaten als Methoden- und Theorierahmen erprobt worden. Dabei standen ganz unterschiedliche Aspekte im Mittelpunkt. Dies betrifft eine Lehrveranstaltung zur "Politischen Sprache im Mittelalter" (durchgeführt von Bernhard Jussen und Silke Schwandt) wie auch, zweitens, eine Studiengruppe "Historische Semantik". Drittens sind mehrere "Anwender-

The screenshot shows the Lexicon Browser interface with the following search criteria:

- frequency: 5
- idf: 0
- length: 3
- text-freq: 5
- ridf: 0

The search results table is as follows:

Lemma	Lemma ID	PoS	Author	Procedure	Edition	Length	Frequency	Text Freq.	IDF	RIDF
apparere	138995	N (Nomen)	sysop	initialization	2010-04-01 21:05:09	9	6	5	7.438854007	0.181968802
apportare	140379	N (Nomen)	sysop	initialization	2010-04-01 21:05:10	9	6	5	7.438854007	0.181968802
deminutor	206459	N (Nomen)	sysop	initialization	2010-04-01 21:05:27	9	7	5	7.438854007	0.336060693
distortor	223223	N (Nomen)	sysop	initialization	2010-04-01 21:05:32	9	6	5	7.438854007	0.181968802
estas	233968	N (Nomen)	sysop	initialization	2010-04-01 21:05:35	5	7	5	7.438854007	0.336060693
fumigator	252529	N (Nomen)	sysop	initialization	2010-04-01 21:05:40	9	7	5	7.438854007	0.336060693
includere	266251	N (Nomen)	sysop	initialization	2010-04-01 21:05:44	9	6	5	7.438854007	0.181968802
mirare	293967	N (Nomen)	sysop	initialization	2010-04-01 21:05:52	6	6	5	7.438854007	0.181968802
saepes	411935	N (Nomen)	sysop	initialization	2010-04-01 21:06:24	6	8	5	7.438854007	0.469533299
iumentum	415017	N (Nomen)	sysop	initialization	2010-04-01 21:06:25	8	15	5	7.438854007	1.097730480
faenum	415720	N (Nomen)	sysop	initialization	2010-04-01 21:06:25	6	17	5	7.438854007	1.222776068

**Abbildung 8:** Ein Screenshot des *Lexicon Browser* mit Ergebnissen einer Suchanfrage basierend auf einem lateinischen Vollformenlexikon.

Workshops“ zu computergestützten Methoden durchgeführt worden, und zwar in Zusammenarbeit mit der Arbeitsgruppe Texttechnologie von Alexander Mehler. In Veranstaltungen dieser drei Lehrformate wurde der eHumanities Desktop dazu genutzt, praktische korpuslinguistische Arbeit in der Geschichtswissenschaft zu leisten.

#### 4.1 Das Format Übung/Seminar

Bei den Übungen und Seminaren mit Studierenden aus dem Grund- und Hauptstudium ging es in erster Linie darum, neue Ansätze zur Erforschung der “Politischen Sprache des Mittelalters” zu vermitteln. Nach einer Einführungsphase zum Konzept des Politischen und seiner Anwendbarkeit auf die Vormoderne oblag es den Studierenden, eigene semantische Untersuchungen zu einzelnen Wörtern zu unternehmen. Dazu wurde ihnen das HSCM-System (siehe Sektion 3.5) zur Verfügung gestellt. Diese Herangehensweise verfolgte zwei Ziele. Zum einen sollte den Studierenden deutlich gemacht werden, dass es

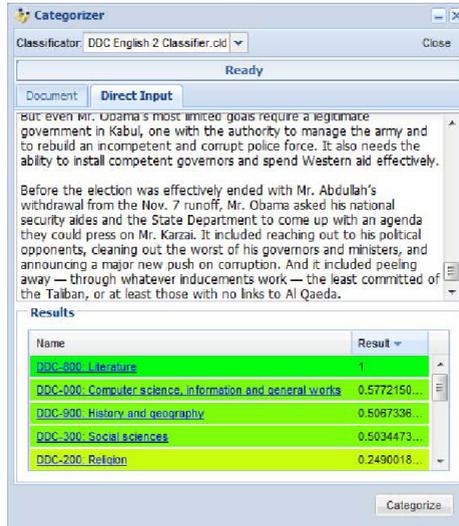


Abbildung 9: Ein Screenshot des *Categorizer* mit dem Ergebnis einer Kategorisierung.

lohnenswert ist, sich von altbekannten Methoden zu verabschieden und in der Forschung unbekanntes Terrain zu betreten. Nicht zuletzt auch, um mit den gewonnen Erkenntnissen bestehende Narrative der Forschung in Frage zu stellen. Zum anderen sollten Kompetenzen im Umgang mit korpuslinguistischen Fragestellungen, Methoden und Werkzeugen eingeübt werden. Dies geschah vor dem Hintergrund eines mediävistischen Erkenntnisinteresses und sollte so den Blick für neue Methoden und Theorien (etwa im Bereich der Historischen Semantik) öffnen.

Die Arbeit mit HSCM hat dabei das eigenständige Forschungsinteresse der Studierenden geweckt. Durch den erleichterten Zugriff auf empirische Daten für die geisteswissenschaftliche Forschung, die HSCM ermöglicht, waren die Teilnehmerinnen und Teilnehmer in der Lage, schnell eigene Ideen zu erproben und zu Ergebnissen zu gelangen. Auf diese Weise lassen sich Studierende besonders gut an Forschungskonzepte und an die Forschungspraxis heranführen.

## 4.2 Das Format der Studiengruppen

In der Studiengruppe “Historische Semantik”<sup>9</sup> werden gegenüber dem in Sektion 4.1 skizzierten Format der Übung andere Ziele verfolgt. Das Modell für dieses Lehrformat unterscheidet sich wesentlich von dem einer Übung oder auch eines Seminars. Ohne

<sup>9</sup>Dieses Lehrformat wird derzeit an der Goethe-Universität Frankfurt erprobt ([http://www.fzhg.org/front\\_content.php?idcat=39](http://www.fzhg.org/front_content.php?idcat=39) [letzter Zugriff 19.04.2011]).

hierarchische Strukturierung dient die Studiengruppe als ein Forum zur Diskussion von forschungsorientierten Problemen und Fragestellungen. Die Gruppe wählt sich ihre Themen selbst aus, trifft sich über mehrere Semester hinweg und verfolgt neben der wissenschaftlichen Erarbeitung eines Themas auch das Ziel der Vernetzung von Wissenschaftlerinnen und Wissenschaftlern an der Goethe-Universität. Sie richtet sich an Professorinnen und Professoren, Mitarbeiterinnen und Mitarbeiter sowie an Studierende. In diesem Zusammenhang wird die Historische Semantik als ein Forschungsprogramm diskutiert. Als Grundlage dieser Diskussionen dienen methodisch-theoretische Basistexte, eigene Arbeiten der Mitglieder, die zum Teil auf dem eHumanities Desktop basieren, sowie Vorträge von eingeladenen Gästen. Außerdem werden Workshops zum Thema der Studiengruppe organisiert.

Auch wenn die Diskussion breit angelegt ist, spielen neue, korpusorientierte und computergestützte Ansätze eine wesentliche Rolle. Durch die verschiedenen Forschungspositionen und Fragestellungen der Mitglieder ergeben sich immer wieder neue interessante Anwendungsbereiche für entsprechende Werkzeuge. In der Studiengruppe wird zudem die basale Erstellung von Korpora diskutiert, und zwar anhand forschungspraktischer Beispiele, um den Teilnehmerinnen und Teilnehmern einen unmittelbaren Einblick in die geisteswissenschaftliche Fachinformatik basierend auf dem eHumanities Desktop zu gewähren.

### **4.3 Das Workshop-Format**

Das dritte Format, in dem die computergestützte Historische Semantik als Lehrinhalt erprobt worden ist, betrifft reine Anwendungsworkshops. Das Ziel dieser Veranstaltungen besteht darin, die Teilnehmerinnen und Teilnehmer mit den Möglichkeiten und Funktionsweisen des eHumanities Desktops und insbesondere des HSCM-Systems vertraut zu machen. Diese Veranstaltungsform adressiert keine spezifische Zielgruppe, sondern richtet sich an Forscherinnen und Forscher, die mit den Werkzeugen arbeiten oder diese diskutieren wollen.

Dieses Veranstaltungsformat erweitert zum einen den Kreis derer, die den Desktop für ihre Forschungen verwenden. Zum anderen liefert es Anregungen zur weiteren Verbesserung und Erweiterung des eHumanities Desktops. Alles in allem dient das Format der Anwendungsworkshops insbesondere dazu, die Vorteile der Programme und Methoden für die geisteswissenschaftliche Forschung und ihre Anwendungsbereiche zu demonstrieren.

## **5 Diskussion**

Nachdem wir drei Einsatzszenarien des eHumanities Desktops in der Lehre umrissen haben, geht es in dieser Sektion um Anforderungen an die Nutzung dieses Systems. Betrachtet man den Kreis der Personen, welche den eHumanities Desktop nutzen, so lassen sich diese erfahrungsgemäß in zwei Gruppen einordnen:

- Auf der einen Seite steht die Zielgruppe der Geisteswissenschaftler, deren Arbeit im Wesentlichen verwendungsorientiert ist. Für diese Nutzergruppe stehen Funktionen aus den Bereichen Suchen, Browsen und Annotieren im Vordergrund. Es geht also um die computerbasierte Repräsentation von Daten, um deren Annotation weit diesseits einer automatischen Weiterverarbeitung. Als Nutzungsprämisse gilt die möglichst niedrige Einstiegsschwelle aller verwendeten Programme: nicht das maschinelle Lernen zur Exploration unbekannter Strukturen, sondern die Datenbankfunktionalität zur Verwaltung von Annotationen und deren Verteilung unter Arbeitsgruppenmitgliedern steht im Fokus.
- Die zweite Zielgruppe bilden die Informatiker, deren Vorgehen modellierungsorientiert ist. Für diese Nutzergruppe stehen Funktionen aus den Bereichen Entwerfen, Modellieren, Entwickeln, Testen und Optimieren im Vordergrund. Vergleichbar mit Systemen wie WEKA (Hall et al., 2009) oder GATE (Bontcheva et al., 2004) erwarten diese Nutzer eine Unterstützung hinsichtlich der Erzeugung automatischer Klassifikatoren oder zumindest die API-basierte Nutzbarkeit von Klassifikatoren von Arbeitsgruppen mit verwandter thematischer Ausrichtung (etwa im Hinblick auf den DDC-Klassifikator des Desktops – siehe Sektion 3.7). Als Nutzungsprämisse gilt die Erweiterbarkeit und prozedurale Vielfalt des Systems, denegenüber die Datenbankfunktionalität in den Hintergrund rückt.

Dieser Grobeinteilung zweier Nutzergruppen gemäß ergeben sich unterschiedliche Anforderungen an den Ausbau des eHumanities Desktops: auf der einen Seite betrifft dies die Verfügbarkeit von Daten in zunehmender Breite und Tiefe, ob nun auf der Ebene von Korpora oder Lexika. Bei der Analyse von althochdeutschen Texten (Gippert, 2001) tritt beispielsweise die Aufgabe zu Tage, lateinische Lehnwörter zu identifizieren oder Sätze in Althochdeutsch und Latein zu alignieren. Aufgaben dieser Art setzen die Verfügbarkeit von Lexika und Korpora für historische Sprachen voraus, für die vielfach noch immer digitalisierte Ressourcen fehlen. Die hinlängliche Dokumentation aller verwendbaren Werkzeuge wie auch die Kontrollierbarkeit und Rekonstruierbarkeit von Messresultaten stellen weitere typische geisteswissenschaftliche Anforderungen dar. Diese Anforderungen zu erfüllen bedeutet letztlich, vollständige Transparenz und unmittelbare Nachvollziehbarkeit jeder systemseitigen Berechnung zu gewährleisten, ohne auf irgendeine externe Ressource (und sei es eine Literaturreferenz) zurückgreifen zu müssen. In letzter Konsequenz lässt diese Anforderung nur direkte Messungen zu, unter Reduktion des Desktops auf dessen Datenbank- und Retrieval-Funktionalität – unter Ausklammerung jeglicher Form von explorativer Datenanalyse. Für viele geisteswissenschaftliche Anwendungen, auch für weite Teile der historischen Semantik, gilt dies als eine Nutzungsprämisse, so dass es nicht verwundert, dass Module wie das HSCM im Kern Kollokationsanalysen und mengenorientierte Textvergleiche anbieten. Der Grund hierfür besteht darin, dass weitergehende explorative Verfahren hinsichtlich ihrer statistischen Resultate in der geisteswissenschaftlichen Diskussion schwer vermittelbar sind.

An dieser Stelle ist eine Anforderung hervorzuheben, welche die Texttechnologie vor erhebliche Herausforderungen stellt. Dies betrifft jene Art von “pervasive computing”, welche sich Geisteswissenschaftler von der Nutzung von Werkzeugen wie dem eHumanities Desktop erwarten. Es geht dabei genauer um die prinzipielle Vernetzung sämtlicher lexikalischer, syntaktischer und textueller Ressourcen, so dass jedes Textsegment durch Aufruf sämtlicher seiner Belegstellen, ob nun in Lexika, Korpora oder Baumbanken, überall im System aktualisiert werden kann: wer beispielsweise in einem lateinischen Korpus wie der PL einen französischen Kommentar findet, will womöglich zu einem französischen Lexikon wechseln können, das die gefundenen Wortformen ausweist, um von dort aus zu einem französischen Korpus wechseln zu können, das weitere Belegstellen derselben Wortformen bereithält u.s.w. Dieses Überall-verfügbar-sein und Weiterverarbeiten-können von Ressourcen, wo immer man sie benötigt, kann als Argument für die Unabdingbarkeit von Anwendungen wie dem eHumanities Desktop gelten, da hier die webbasierte Verfügbarmachung einer objektorientierten Datenbank erst die Voraussetzung für diese Art der Ressourcenvernetzung schafft.

Anforderungen in einem typischerweise geisteswissenschaftlichen Kontext bilden die eine Seite des Anforderungsspektrums. Demgegenüber stehen stärker informatorische Anforderungen wie die prozedurale Erweiterbarkeit des Gesamtsystems und dessen Interoperabilität (Wittenburg et al., 2010). Zusammenfassend gesprochen ergeben sich somit zwei zentrale Forderungen an die Weiterentwicklung des eHumanities Desktops: hinsichtlich des Ausbaus seines prozeduralen Angebots und der Vernetzung von Ressourcen wie auch hinsichtlich der Geschwindigkeit, mit welcher der Desktop weiterentwickelt wird, um immer neuen Anforderungen an eine automatische Verarbeitung geisteswissenschaftlicher Informationsobjekte gerecht zu werden. Denn eines lehrt unsere nunmehr mehrjährige Kooperation von Geisteswissenschaft und Informatik: die Umsetzung hinkt stets der Anforderungsentwicklung hinterher.

## 6 Zusammenfassung und Schlussfolgerung

Der eHumanities Desktop ist eine webbasierte Schnittstelle für die texttechnologische Arbeit in Forschung und Lehre. Sein Funktionsumfang eignet ihn dazu, eine Vielfalt von Arbeitsfeldern in unterschiedlichen Bereichen der *Digital Humanities* zu bedienen: ausgehend von dem Browsing einzelner Korpora über das Annotieren von Dokumenten und das Bearbeiten einzelner Lexikoneinträge bis hin zur explorativen Analyse von Textvergleichen und dem maschinellen Lernen von Klassifikatoren. Damit adressiert der eHumanities Desktop einen Funktionsumfang, der ihn neben der Forschung auch für den Einsatz in der Lehre eignet. Infolge der webbasierten Architektur des Desktops und seines graphischen Interfaces geht es dabei nicht allein um die Lehre für Computerlinguisten, Texttechnologien oder Informatiker. Vielmehr zeigen die hier beschriebenen Nutzungsszenarien die Einsetzbarkeit des Desktops insbesondere auch für Geisteswissenschaftler. In diesem Zusammenhang haben wir drei Lehrformate diskutiert. Die Erfahrungen, die wir dabei mit dem Lehrinhalt *Historische Semantik* unter Einsatz des *eHumanities Desktops* sammeln konnten, unterscheiden sich je nach Lehrformat. In je-

dem Fall sollte die Anwendungsorientierung im Mittelpunkt jedes dieser Formate stehen. Ein Grund hierfür liegt darin, dass sich theoretische Ansätze und deren methodische Eigenheiten überhaupt erst in der Anwendung als ertragreich erweisen. Dabei bildet die Vermittlung von Fähigkeiten zur Interpretation quantitativer Ergebnisse eine zentrale Lehraufgabe. Erst im Kontext von historischem Grundwissen und wissenschaftlichen Fragestellungen werden die mit Hilfe des eHumanities Desktops ermittelten Zahlen und die damit verknüpften sprachlichen Befunde interpretierbar. Die hier erprobte Interdisziplinarität von Geschichtswissenschaft und Texttechnologie ist offenbar also unabdingbar, um diesen Brückenschlag zwischen qualitativen und quantitativen Methoden zu leisten.

### Literatur

- Andersen, P. B. (1990). *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*. Cambridge University Press, Cambridge.
- Bontcheva, K., Tablan, V., Maynard, D., and Cunningham, H. (2004). Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*, 10(3/4):349–373.
- Evert, S. (2008). Corpora and collocations. In Lüdeling, A. and Kytö, M., editors, *Corpus Linguistics. An International Handbook of the Science of Language and Society*. Mouton de Gruyter, Berlin/New York.
- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28((1-2)):3–71.
- Gippert, J. (2001). TITUS – Alte und neue Perspektiven eines indogermanistischen Thesaurus. *Studia Iranica, Mesopotamica et Anatolica*, 2:46–76.
- Gleim, R., Mehler, A., Waltinger, U., and Menke, P. (2009a). eHumanities Desktop – an extensible online system for corpus management and analysis. In *5th Corpus Linguistics Conference, University of Liverpool*.
- Gleim, R., Waltinger, U., Ernst, A., Mehler, A., Esch, D., and Feith, T. (2009b). The eHumanities Desktop – an online system for corpus management and analysis in support of computing in the humanities. In *Proceedings of the Demonstrations Session of the 12th Conference of the European Chapter of the Association for Computational Linguistics EACL 2009, 30 March – 3 April, Athens*.
- Gleim, R., Warner, P., and Mehler, A. (2010). eHumanities Desktop – an architecture for flexible annotation in iconographic research. In *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10), April 7-10, 2010, Valencia*.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Heyer, G. and Büchler, M. (2010). GI annual meeting 2010. Workshop eHumanities – how does computer science benefit? In Fähnrich, K.-P. and Franczyk, B., editors, *Proceedings of INFORMATIK 2010: Service Science, September 27 – October 01, 2010, Leipzig*, volume 2 of *Lecture Notes in Informatics*. GI.
- Heyer, G., Quasthoff, U., and Wittig, T. (2006). *Text Mining: Wissensrohstoff Text*. W3L, Herdecke.

- Jussen, B. (2006). “Ordo” zwischen Ideengeschichte und Lexikometrie. Vorarbeiten an einem Hilfsmittel mediävistischer Begriffsgeschichte. In Schneidmüller, B. and Weinfurter, S., editors, *Ordnungskonfigurationen im Hohen Mittelalter*, volume 64 of *Vorträge und Forschungen*, pages 227–256. Thorbecke, Sigmaringen.
- Jussen, B., Mehler, A., and Ernst, A. (2007). A corpus management system for historical semantics. *Sprache und Datenverarbeitung. International Journal for Language Data Processing*, 31(1-2):81–89.
- Koselleck, R. (1972). Einleitung. In Brunner, O., Conze, W., and Koselleck, R., editors, *Geschichtliche Grundbegriffe. Lexikon zur politisch-sozialen Sprache in Deutschland*, pages V–XXVII. Klett-Cotta, Stuttgart.
- Luhmann, N. (1993). *Gesellschaftsstruktur und Semantik. Studien zur Wissenssoziologie der modernen Gesellschaft*, volume 1. Suhrkamp, Frankfurt am Main.
- Mehler, A., Diewald, N., Waltinger, U., Gleim, R., Esch, D., Job, B., Küchelmann, T., Pustynikov, O., and Blanchard, P. (2011). Evolution of Romance language in written communication: Network analysis of late Latin and early Romance corpora. *Leonardo*, 44(3).
- Mehler, A., Gleim, R., Ernst, A., and Waltinger, U. (2008). WikiDB: Building interoperable wiki-based knowledge resources for semantic databases. *Sprache und Datenverarbeitung. International Journal for Language Data Processing*, 32(1):47–70.
- Mehler, A., Gleim, R., Waltinger, U., and Diewald, N. (2010). Time series of linguistic networks by example of the Patrologia Latina. In Fähnrich, K.-P. and Franczyk, B., editors, *Proceedings of INFORMATIK 2010: Service Science, September 27 – October 01, 2010, Leipzig*, volume 2 of *Lecture Notes in Informatics*, pages 609–616. GI.
- Mehler, A., Gleim, R., Waltinger, U., Ernst, A., Esch, D., and Feith, T. (2009). eHumanities Desktop – eine webbasierte Arbeitsumgebung für die geisteswissenschaftliche Fachinformatik. In *Proceedings of the Symposium “Sprachtechnologie und eHumanities”, 26.–27. Februar, Duisburg-Essen University*.
- Migne, J.-P., editor (1844–1855). *Patrologiae cursus completus: Series latina*, volume 1–221. Chadwyck-Healey, Cambridge.
- Petitot, J., Varela, F. J., Pachoud, B., and Roy, J.-M. (2000). *Beyond the Gap: An Introduction to Naturalizing Phenomenology*. Stanford University Press, Stanford.
- Radl, C. (2007). Die digitalen Monumenta Germaniae Historica. In historischer Forschungseinrichtungen in der Bundesrepublik Deutschland e.V., A., editor, *Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland: Berichtsjahr 2006*, pages 69–74. München.
- Rexroth, F. (2009). Politische Rituale und die Sprache des Politischen in der historischen Mittelalterforschung. In DeBenedictis, A., Corni, G., Mazohl, B., and Schorn-Schütte, L., editors, *Die Sprache des Politischen in actu. Zum Verhältnis von politischem Handeln und politischer Sprache von der Antike bis ins 20. Jahrhundert*, pages 71–90. Vandenhoeck & Ruprecht, Göttingen.
- Rieger, B. (1998). Warum fuzzy Linguistik? Überlegungen und Ansätze zu einer computerlinguistischen Neuorientierung. In Krallmann, D. and Schmitz, H. W., editors, *Perspektiven einer Kommunikationswissenschaft. Internationales Gerold Ungeheuer Symposium, Essen 1995*, pages 153–183. Nodus, Münster.

- Rieger, B. B. (1989). *Unschärfe Semantik: Die empirische Analyse, quantitative Beschreibung, formale Repräsentation und prozedurale Modellierung vager Wortbedeutungen in Texten*. Peter Lang, Frankfurt a. M.
- Rieger, B. B. (2003). Semiotic cognitive information processing: Learning to understand discourse. A systemic model of meaning constitution. In Kühn, R., Menzel, R., Menzel, W., Ratsch, U., Richter, M. M., and Stamatescu, I. O., editors, *Adaptivity and Learning. An Interdisciplinary Debate*, pages 347–403. Springer, Berlin.
- Skinner, Q. (2003). *Visions of Politics (reprint)*, volume 1: Regarding Method. Cambridge University Press, Cambridge.
- Steinmetz, W. (2008). Vierzig Jahre Begriffsgeschichte. The State of the Art. In Kämper, H. and Eichinger, L. M., editors, *Sprache – Kognition – Kultur. Sprache zwischen mentaler Struktur und kultureller Prägung [Vorträge der Jahrestagung 2007 des Instituts für Deutsche Sprache]*, pages 174–197. De Gruyter, Berlin.
- TEI Consortium, editor (2010). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium, <http://www.tei-c.org/Guidelines/P5/>.
- von Ahn, L. (2008). Human computation. In *IEEE 24th International Conference on Data Engineering (ICDE 2008)*, pages 1–2. IEEE.
- Wagner, B., Mehler, A., Wolff, C., and Dotzler, B. (2009). Bausteine eines Literary Memory Information System (LiMeS) am Beispiel der Kafka-Forschung. In *Proceedings of the Symposium “Sprachtechnologie und eHumanities”, 26.–27. Februar, Duisburg-Essen University*.
- Waltinger, U. (2010). *On Social Semantics in Information Retrieval*. Phd thesis, Bielfeld University, Germany.
- Waltinger, U., Mehler, A., Lösch, M., and Horstmann, W. (2011). Hierarchical classification of OAI metadata using the ddc taxonomy. In Bernardi, R., Chambers, S., Gottfried, B., Segond, F., and Zaihrayeu, I., editors, *Advanced Language Technologies for Digital Libraries (ALT4DL)*, LNCS. Springer-Verlag, Berlin.
- Wittenburg, P., Hinrichs, E. W., Broeder, D., and Zastrow, T. (2010). eHumanities – können wir die Komplexität beherrschen? In Fähnrich, K.-P. and Franczyk, B., editors, *Proceedings of INFORMATIK 2010: Service Science, September 27 – October 01, 2010, Leipzig*, volume 2 of *Lecture Notes in Informatics*, pages 530–541. GI.