Bausteine eines *Literary Memory Information System* (LiMeS) am Beispiel der Kafka-Forschung

Benno Wagner[1], Alexander Mehler[2], Christian Wolff[3], Bernhard Dotzler[3]

[1]Fachbereich 3 – Sprach-, Literatur- und Medienwissenschaften
Universität Siegen
wagner@lit-wiss.uni-siegen.de

[2]Geisteswissenschaftliche Fachinformatik
Universität Frankfurt / Main
mehler@em.uni-frankfurt.de

[3]Institut für Information und Medien, Sprache und Kultur
Universität Regensburg
{bernhard.dotzler, christian.wolff}
@sprachlit.uni-regensburg.de

> Ich rede von denen, die je nach der verschiedenen Stufe ihrer Kenntnisse ganz verschiedene Bücher lesen, ohne bestimmten Plan, unaufhörlich wechselnd, selten in einem Buch lang ausruhend, getrieben von einer unausgesetzten, nie recht gestillten Sehnsucht. [...] sie suchen ja von Buch zu Buch, was der Inhalt keines ihrer tausend Bücher ihnen geben kann: sie suchen etwas, was zwischen den Inhalten aller einzelnen Bücher schwebt, was diese Inhalte in eins zu verknüpfen vermöchte. Sie schlingen die realsten, die entseeltesten aller Literaturen hinunter und suchen etwas höchst Seelenhaftes. [...]. Aber die Sehnsucht [...] geht durchaus nicht auf den Dichter. Es ist der Mann der Wissenschaft, der diese Sehnsucht zu stillen vermag.
> *Hugo von Hofmannsthal (1907)*

> Letztlich reicht es nicht aus, auf Seiten der Objektbasis unablässig neu digitales Material zu akkumulieren. Parallel dazu müsste auf Seiten der Forschung die Bereitschaft zum aktiven Einsatz technologisch und methodisch innovativer Verfahren gefördert werden. Die Digitalisierung allein ohne eine begleitende Theoriedebatte und ohne ein verfeinertes methodisches Rüstzeug betreiben zu wollen dürfte zu verkürzten Ergebnissen führen. Beide Seiten, das Material und die Forschung, die Technik und die Methodologie, sind aufs engste miteinander verbunden.
> *Michael Embach/Andrea Rapp (2008)*

1 Einleitung

In dem Paper beschreiben wir Bausteine eines *Literary Memory Information System* (LiMeS), das die literaturwissenschaftliche Erforschung von so genannten *Matrixtexten* – das sind Primärtexte eines bestimmten literarischen Gesamtwerks – unter dem Blickwinkel großer Mengen so genannter *Echotexte* (Topia 1984; Wagner/Reinhard 2007) – das sind Subtexte im Sinne eines literaturwissenschaftlichen Intertextualitätsbegriffs – ermöglicht. Den Ausgangspunkt dieses computerphilologischen Informationssystems bildet ein Text-Mining-Modell basierend auf dem Intertextualitätsbegriff in Verbindung mit dem Begriff des *Semantic Web* (Mehler, 2004b, 2005a, b, Wolff 2005). Wir zeigen, inwiefern dieses Modell über bestehende Informationssystemarchitekturen hinausgeht und schließen einen Brückenschlag zur derzeitigen Entwicklung von Arbeitsumgebungen in der geisteswissenschaftlichen Fachinformatik in Form eines *eHumanities Desktop*.

2 Literaturwissenschaftliche Perspektive: Transbiblionome Räume

Moderne Literatur funktioniert essentiell intertextuell und intermedial. Mit dem Heraufkommen der neuen ‚Konkurrenz'-Medien (Photographie und Film, Telegraph und Telephon) und mit der Ablösung der Referenzinstitution Bibliothek durch globale Informationssysteme (Rayward

2008) sowie das alle Lebensbereiche durchdringende Wissen moderner Verwaltungen (Wagner 2006a) existiert Schrift nurmehr im Spannungsverhältnis zwischen Buchgebundenheit und einem zunehmend transbiblionom organisierten kulturellen Kontext. Der literarische Text gerät auf diese Weise einerseits zum „geometrischen Ort eines hors-texte", zu einem „Kreuzungspunkt von Schichten, die Myriaden von Horizonten entspringen" (Topia 1984, 103). Anderseits wird Literatur unter diesen Bedingungen zu einer besonderen Instanz des kulturellen Gedächtnisses. Sie lässt sich als komplexer „Spurenkörper" (Pêcheux 1983, 55) beschreiben, oder – jedenfalls in ihren raffiniertesten und reflektiertesten Schreibweisen – auch als „hypermnemische Maschine" (Derrida 1984, 147), als dynamischer Erinnerungsapparat, dessen virtuelles Verweispotential auf andere Texte die Ordnungsraster realer Wissensspeicher (Enzyklopädien, Bibliotheken, Archive) durchkreuzt.

So schrieb Derrida in der *Grammatologie* zunächst: „Es geht [...] nicht darum, der Buchhülle noch nie dagewesene Schriften einzuverleiben, sondern endlich das zu lesen, was in den vorhandenen Bänden schon immer zwischen den Zeilen geschrieben stand. Mit dem Beginn einer zeilenlosen Schrift wird man auch die vergangene Schrift unter einem veränderten Organisationsprinzip lesen. [...] Was es heute zu denken gilt, kann in Form der Zeile oder des Buches nicht niedergeschrieben werden; ein derartiges Unterfangen käme dem Versuch gleich, die moderne Mathematik mit Hilfe einer Rechenschiebermaschine zu bewältigen." Stattdessen avisiert er, diesmal mit Leroi-Gourhan, „eine andere[n], bereits vorstellbare[n] Art der Speicherung [...], deren rasche Verfügbarkeit der des Buches überlegen sein wird: die große ‚Magnetothek' mit elektronischer Auswahl wird in naher Zukunft vorselektierte und sofort verfügbare Informationen liefern" (Derrida 1967, 154f.).

Knapp zwei Jahrzehnte später hingegen, in einem Aufsatz aus dem Jahre 1984, konfrontiert uns Derrida mit einem ganz anderen und scheinbar diametral entgegengesetzten Szenario. Im Bezug auf den *Ulysses* von Joyce heißt es nun: „for there be no simple confusion between him [Joyce] and a sadistic demiurge, setting up a hypermnesiac machine, there in advance, decades in advance, to compute you, control you, forbid you the slightest inaugural syllable because you can say nothing that is not programmed on this 1000th generation computer [...] beside which the current technology of our computers and micro-computerfied archives and translating machines remains a bricolage of a prehistoric child's toys" (Derrida 1984, 147). Hier hat sich offenbar das Komplexitätsgefälle zwischen Druckschrift und elektronischem Speicher verkehrt. Die lineare Schrift ist nicht länger Komplexitäts-Engpass, sondern sie fungiert selbst als Quelle einer überbordenden Komplexität. *Sie* ist nun der "Computer der 1000. Generation", im Vergleich zu dem die elektronischen Speichermedien als Problem erscheinen, als eine dem Gegenstand der Druckschrift unangemessene, prähistorische Spielerei.

Betrachtet man nun den Einsatz von Computern zu Zwecken der Literaturforschung seit den 1990er Jahren, so drängt sich der Eindruck auf, als habe jedes der beiden Zitate eine Arbeitsperspektive eröffnet, die von der jeweils anderen nichts zu wissen scheint. So haben Autoren wie George Landow und Jay Bolter, in einer eigentümlichen Einebnung des Unterschieds zwischen der syntagmatischen und der paradigmatischen Text-Dimension, den elektronischen Hypertext kurzerhand zu jenem Medium deklariert, mit dessen Hilfe sich das intertextuelle Verweispotential eines literarischen Textes restlos implementieren, der vieldimensionale literarische Text sich aus dem

Zwang der Zeile befreien ließe.1 Dass freilich Hypertext im besten Falle als funktional limitiertes Instrument zur Darstellung von Intertextualitätsbeziehungen dienen kann, belegen ex negativo die an diese kurzschlüssigen Verheißungen anschließenden Forschungsprojekte. Als paradigmatischer Gegenstand dient hier in erster Linie das Werk von James Joyce, wobei sich die *Hypermedia Joyce Studies* (http://www.geocities.com/hypermedia_joyce/) mit einer Vielzahl von text*theoretisch* anregenden Joyce-Lektüren als dynamisches Zentrum etabliert haben. Ein anderes Bild ergibt sich, wenn man sich die *HJS*-Liste der „hypermedia projects" ansieht (http://www.geocities.com/hypermedia_joyce/biblio_ht.html). Die dort aufgelisteten *praktischen* Versuche, den intertextuellen Raum etwa des *Ulysses* mittels Hypertext darzustellen und nutzbar zu machen, erinnerten deutlich an Derridas „prehistoric child's toy"-Verdikt, bevor sie mehrheitlich durch eine file-not-found-Meldung ersetzt wurden.

Sehr viel erfolgreicher gestaltet sich computergestützte Literaturforschung immer dann, wenn sie die transbiblionome Dimension der Intertextualität von vorherein aus ihrem Gegenstandsbereich ausschließt. Dies geschieht zumeist implizit, ohne weitere Erörterung und Reflexion, bisweilen aber auch mit programmatischem Nachdruck, wenn sich etwa die in Deutschland etablierte Computerphilologie explizit auf die Befassung mit „traditionellen philologischen Gegenständen" und damit das Potential des Computer auf die Optimierung biblionomer Funktionen sowie auf die Herstellung dezentraler Forschungskollaborationen beschränkt.

Eine *intertextuell-transbiblionome Computerphilologie* bleibt unter diesen Bedingungen *Desiderat*. Ihre methodische und technische Entwicklung hätte sich vor dem skizzierten Erfahrungshintergrund an drei Leitlinien zu orientieren:

(1) *Zielsetzung*: Transbiblionome Computerphilologie zielt auf die computergestützte Erschließung und Darstellung der intertextuellen und intermedialen Dimension literarischer Texte jenseits einer Beschränkung auf einen biblionomen Forschungshorizont und der Fixierung auf das Potential von Hypertext. Als technische Grundlage hierfür hätte eine gegenstandsspezifische digitale Arbeitsumgebung zu dienen, die zugleich die dezentrale Kollaboration von Experten(gruppen) und eine nutzerspezifisch differenzierte Aufbereitung der Forschungsergebnisse ermöglicht.

(2) *Theorie*: Unter den genannten Bedingungen kann und muss sich das zugrundeliegende Intertextualitäts-Modell der methodischen Alternative entziehen, mittels derer die printorientierte Methodendiskussion (insbesondere der 70er und 80er Jahre) sich um eine „Zähmung" (Lachmann 1984, 137) des transbiblionomen Potentials ‚moderner' Intertextualität bemüht hatte: hier eine ‚geschlossene', durch unterstellte Verknüpfungsabsichten des Autors oder faktische Verknüpfungsoperationen des Lesers begrenzte, dort eine unmittelbar auf das System der langue bezogene ‚offene' und daher, zumal unter Bedingungen einer printfixierten Forschung, forschungspraktisch niemals einholbare Intertextualität. Baßlers Entwurf eines ‚archivimmanenten Strukturalismus' (Baßler

1 Hierzu konstatiert Baßler: „Landows Parallelisierung von Hypertext mit jenem poststrukturalistischen Textbegriff, den Barthes in S/Z entwickelt, setzt sich über den elementaren Unterschied von syntagmatischer und paradigmatischer Textdimension großzügig hinweg. [...] Dabei handelt es sich jedoch um zwei vollkommen verschiedene Dinge, denn ein Hypertext mag so nonlinear sein wie er will – das betrifft doch immer nur die Sequenz, in seiner paradigmatischen Dimension dagegen unterscheidet er sich nicht vom normalen Text" (Baßler 2005, 307f.). Bei Bolter scheint dieses ebenso banale wie fatale Missverständnis zu der Auffassung zu führen, dass der von Joyce ‚stark' bewirtschaftete intertextuelle Raum des *Ulysses*, dass also das vielschichtige Paradigma des Romans (seine „several layers of allusions") eine wundersame Vermehrung seines Textsyntagmas (seiner „simple storyline") bewirkt. Das geht so weit, dass am Ende der schrille Rückkopplungseffekt vieler amerikanischer Hypertext-Panels hier einmal in Druckschrift gebannt wird: „[Joseph] Frank's characterization of James Joyce remains appropriate for the hyperfictions of Michael Joyce" (Bolter 2001, 174).

2005), der den intertextuellen Raum (das ‚Paragrammaire' nach J. Kristeva) eines Bezugstextes auf eine historische Positivität von Kontext-Dokumenten bezieht, die er ‚Archiv' nennt, kann hier als fundierte und konstruktive Alternative dienen, deren Begrifflichkeit sich unmittelbar auf die Konzepte und Leistungen einer digitalen Texttechnologie beziehen lässt. Präzisierungen und Erweiterungen der Theorie werden dort anzustreben sein, wo das 'Archiv' nicht nur den biblionomen Raum, sondern zugleich den der Textualität überschreitet, indem es sich multimedial konstituiert.

(3*) Methode*: Intertextuelle Computerphilologie dieser transbiblionomen Art zielt nicht auf die *Implementierung* literarischer Intertextualität: ihre ‚Befreiung' aus dem ‚Gefängnis' der Druckzeile und ‚vollständige Entfaltung' im barrierefreien digitalen Schreibraum, sondern auf ihre *Supplementierung*: auf die forschungstechnische Unterstützung der selbstverständlich2 stets selektiven und (projekt- und methodenspezifisch) perspektivischen Erschließung des intertextuellen Potentials eines je gegebenen literarischen Texts. Bei der Entwicklung einer zweckmäßigen Arbeitsumgebung hätte die Kooperation zwischen Philologie, Medienwissenschaft, Texttechnologie und Informatik einer Logik *pragmatischer Schnittstellenbildung* zu folgen. Statt entweder digitale Lösungsmöglichkeiten mit philologischen Problemstellungen zu überfordern, oder umgekehrt von vornherein die philologischen Problemstellungen an das Leistungsvermögen digitaler tools anzupassen, wären für jede Teilaufgabe die Schnittstellen zwischen humaner Intelligenz und künstlicher Intelligenz präzise zu definieren, um die Leistungsvermögen von Menschen und Rechnern möglichst effizient miteinander zu verschalten.

Ausgehend von diesen Überlegungen projektieren wir die Entwicklung eines *Literary Memory Information System* (LiMeS) als einer literaturwissenschaftlichen Forschungsumgebung, die *literarische Texte* nicht einfach als Gegenstände, sondern *als Medien des kulturellen Gedächtnisses* behandelt, indem sie ihr intertextuelles Verweispotential erschließbar, darstellbar und für unterschiedliche Verwertungszusammenhänge nutzbar macht. Die texttechnologischen Entwicklungen der letzten Jahre bieten u. E. eine tragfähige Basis für die Konzeption und Implementierung einer digitalen Arbeitsumgebung für eine solche intertextuell und transbiblionom orientierte Literaturforschung.

3 Kafka als Paradigma

Das Werk Franz Kafkas bietet einen idealen Testgegenstand für ein solches Vorhaben. Benjamins Bemerkung, nach der „Kafkas ganzes Werk einen Kodex von Gesten darstellt" (Benjamin 1981, 18), lässt sich nämlich ohne weiteres von Kafkas Protagonisten auf seinen Text selbst übertragen. Dies ist im Weiteren näher auszuführen.

3.1 Kafkas literarische Gesten

Wenn sich die sekundäre Sprache der Literatur (J. Lotman) aus einem Ensemble *„literarischer Gesten"* („gestes littéraires") zusammen, die wiederum auf einen Horizont von Archetypen verweisen, den sie beständig reproduzieren, transformieren und überschreiten (Jenny 1976, 257), dann wäre diese Definition für Kafkas literarische Gesten zunächst zu modifizieren. Ihr Zuschnitt und ihre Verweis-Funktion auf den kulturellen Horizont beruhen bereits auf der Erkenntnis, dass dieser Horizont im Zeitalter der Hyperliterarisierung und der aufkommenden elektronischen Medien längst zerfallen ist.

2 Oder nicht ganz so selbstverständlichen, wie Stephan Porombka (2001, 104; 127ff.) in seiner Rekonstruktion der mit dem Hypertext verknüpften Vollständigkeits- und Totalitäts-Phantasien zeigt...

Zunächst sind es die vielgestaltigen, bereits von Massenmedien getragenen Nationalismen der Donaumonarchie, die die Zerfallsmasse dieses Horizonts als Steinbruch fragmentierter „Archaismen" ausbeuten, denen sie, je nach Standpunkt und Anlass, „einen ‚aktuellen Sinn' zu geben versuchen" (Deleuze/Guattari 1976, 35). Das intertextuelle Resonanzfeld der literarischen Gesten Kafkas – seine mauerbauenden Chinesen, nachahmenden Affen, forschenden Hunde, etc. – erstreckt sich durch alle diese Steinbrüche hindurch und setzt so okzidentale und orientalische Religionen und Mythologie, historische Narrative, die Philosophie und die modernen Wissenschaften, und nicht zuletzt das schier unüberschaubare Verbreitungsfeld massenmedialer Stereotypen. Dabei beschränken sich Kafkas literarische Gesten keineswegs darauf, die Versatzstücke ihres kulturellen Kontexts zu speichern und zu rearrangieren (zu letzterem v.a. Neumann 1996; Wagner 2006a). Sie implizieren vielmehr eine dauernd mitlaufende Reflektion der medialen und institutionellen Anschlusspunkte ihrer Operationen. Wo immer Kafka über Telefonzentralen (*Der Verschollene*), Schreibapparate (*In der Strafkolonie*), oder über bürokratische Aktenzirkulation schreibt (*Das Schloß*), transkribiert er die Medientechniken und -dispositive seiner Zeit in selbstreferentielle Protokolle seiner eigenen, nicht auf Sinnbildung oder -zerstörung, sondern auf intermediale Verknüpfung heterogener Kontexte abstellenden Schreibweise. So statten diese Gesten die Schrift mit einer monomedialen Intermedialität (also: Hypermedialität) aus, indem sie zum einen Inhalte aller denkbaren medialen Träger kopieren wie zum anderen diesen Vorgang literaler Inkorporation als mediale Operation schlechthin – nämlich die der Konnektivierung – explorieren (Dotzler 2008). Aufgrund des immensen Reichtums seines intertextuellen Resonanzfelds und der hohen Reflexionsschärfe für den Zusammenhang von Diskursivität, Medialität und Poetizität stellt Kafkas Werk eine ideale Herausforderung für die Entwicklung einer digitalen Arbeitsumgebung für transbiblionome Literaturforschung dar.

3.2 *Literarische Gesten* als Schnittstelle von Intertextualitätstheorie und Texttechnologie

Unter den skizzierten Voraussetzungen liegt es auf der Hand, dass substantialistische Konzeptionen von Intertextualität (wie etwa die der ‚Montage' oder des ‚Palimpsests') durch eine zugleich dynamische und relationale Konzeption zu ersetzen sind. So ließen sich die eingangs beschriebenen literarischen Gesten auch als Ensemble „kybernetischer Schlüssel" beschreiben, die jeweils „ganze Serien von Referenzen, Reminiszenzen, Konnotationen, Echos, Zitate, Pseudo-Zitate, Parallelen und Reaktivierungen" auslösen und rearrangieren (Topia 1984, 103). Während solche kybernetischen Effekte für alle Ebenen des literarischen Texts beschrieben werden können (Topik, Stil, Narrativik, Genre), hat die Kafkaforschung früh bemerkt, was Kafka in verschiedenen Formulierungen auch selbst bekundet: dass es in diesem Falle eine besondere Schicht intertextueller Referenzen gibt, die offenbar auf ein durchgängiges auktoriales Kalkül zurückgeht. Mit anderen Worten, Kafkas Schreibweise ist mit einer ungewöhnlich systematischen Bewirtschaftung des intertextuellen Raums verbunden. Sie basiert auf einem Satz von literarischen Gesten, die zwischen den Polen der Tmesis und des Stereotyps oszillieren und so auf die methodisch grundlegende, wenn auch stets unscharfe Unterscheidung zwischen einem *auktorialen Feld* und einem *historischen Emergenzfeld* verweisen (Wagner/Reinhard 2007, 102f.).3

3 Zum Verhältnis von Tmesis, Stereotyp und intertextueller Katalysis vgl. Baßler 2005, 212ff. Als Beispiel kann hier die Strafmaschine in Kafkas *Strafkolonie* dienen. Sie schließt einerseits an einen hochgradig stereotypisierten Gebrauch der Maschinen-Metapher für die Organisation und Funktion von Staat und Verwaltung an, andererseits schneidet sie sich durch die Einführung des Menschen als Schreibfläche (des Verurteilten, der in die Strafmaschine hineingelegt wird, die ihm dann sein Urteil in die Haut ritzt) von diesem weiten Resonanzfeld ab. Dieses Spannungsverhältnis

Hier liegt nun die entscheidende Schnittstelle zwischen einer literaturwissenschaftlichen und einer texttechnologischen Untersuchung von Intertextualität. Das grundlegende (besser: bodenlose) analytische Problem besteht ja gerade darin, dass die topischen Einheiten, die wir hier *literarische Gesten* nennen, *keine Bestandteile des literarischen Matrix- Textes* sind, sondern dass sie aus der komplexen Differenz zwischen dem Matrix-Text und dem Feld der Echo-Texte allererst hervorgehen. Sie sind mithin *zugleich Ausgangspunkt und Ziel einer empirischen Intertextualitätsforschung*. Was immer von menschlichen Forschern – sei es aufgrund der literarischen Tradition, sei es aufgrund im Regelfalle vereinzelter und jedenfalls selektiver Text-Kontext-Hypothesen – als Ausgangspunkt einer oder mehrerer intertextueller Verweise betrachtet wird, kann mithilfe texttechnologischer Suchfunktionen auf sein intertextuelles Verweispotential zumindest jener Bestandteile des Archivs überprüft werden, der für einen solchen digitalen Zugriff zur Verfügung steht.4 Vor allem aber soll die von uns projektierte Arbeitsumgebung auch umgekehrt die Möglichkeit bieten, bestimmte Wortzusammenstellungen aus dem Syntagma des Matrix-Textes mittels der durch LiMeS bereitgestellten Text Mining-Funktionen den menschlichen Forschern als Kandidaten für literarische Gesten sichtbar zu machen (s.u., 4.1.).

4 Entwicklung eines *Literary Memory Information System* als eHumanities-Vorhaben

Ziel des Vorhabens, das den Brückenschlag zwischen Literatur- und Medienwissenschaft auf der einen Seite und geisteswissenschaftlicher Fachinformatik und Medieninformatik auf der anderen Seite anstrebt, ist die Entwicklung eines *Literary Memory Information System* (LiMeS), dessen Leistungen sich auf drei zentrale Desiderate gegenwärtiger und künftiger Literaturforschung richten:

1. Die informatische Rekonzeption der essentiell konnektiven Logik moderner Printliteratur, sowie die Entwicklung angemessener digitaler Techniken und Werkzeuge zu ihrer Erschließung und Nutzung im Rahmen eines dezentralen und kollaborativen Forschernetzwerks (mit differenzierbaren Zugangsoptionen für ein prinzipiell unbegrenzten Nutzerkreis),

---

zwischen einer technisch-deskriptiven oder metaphorisch-stereotypen Kookkurrenz von Zeichen auf der einen Seite und einer ‚unerhörten' Kookkurrenz auf der anderen wird nun durch eine begrenzte Serie ‚starker' Intertexte vermittelt, von denen man annehmen kann, dass sie Teil eines kalkulierten Spiels seitens des Autors sind – von Nietzsches *Genealogie der Moral* über einen Aufsatz zur Funktionsweise der elektrischen Lochkarten-Zählmaschine bis zu Texten über elektrotherapeutische Apparate, auf deren Ähnlichkeit die Geschichte sogar selbst verweist (dazu ausführlich Wagner 2006b; Dotzler 2006). Kafka selbst hat dieses grundlegende Verhältnis zwischen Autorschaft, intertextuellem Dialog und intertextuellem Rauschen einmal in einer Beschreibung seines Schreibtisches zusammengefasst: „Ich kenne beiläufig nur das, was obenauf liegt, unten ahne ich bloß Fürchterliches" (Kafka 1976, 153). Einer digitale Arbeitsumgebung wie LiMeS soll es erstmals ermöglichen, auch das ‚Unten' der Intertexts und insbesondere seine Beziehungen zum ‚Obenauf' als positiven Untersuchungsgegenstand zu behandeln.

4 Als kompaktes und forschungsstrategisch hoch relevantes Beispiel kann hier die historische Tagespresse dienen, deren umfassende Digitalisierung gegenwärtig große Fortschritte macht. So hat die germanistische Kafkaforschung in Zeitungen wie dem *Prager Tagblatt* oder der *Bohemia* vereinzelt prägnante intertextuelle Beziehungen zu Stellen in Kafkas literarischem Werk ‚entdeckt', konnte deren weiterreichende Signifikanz aber niemals durch eine umfassende Korpusanalyse verifizieren. „Sie hat keinen schlechten Blick, aber zu konzentriert, sie sieht nur den Kern; den Ausstrahlungen zu folgen, die ja eben den Kern fliehen, ist ihr zu mühsam", hat Kafka (1958, 162) in diesem Zusammenhang einmal hellsichtig als Manko erkannt, was hier als zentrale Aufgabenstellung einer digitalen Literaturforschung zu entwickeln ist.

2. die Erschließung und Aufbereitung literarischer Texte als dynamische Archive für kultur- und mediengeschichtliche Forschung, sowie als Folien für eine Zeitdiagnose unserer heutige Situation und
3. die Bereitstellung einer eHumanities-Forschungsumgebung, die auf die zunehmende Verfügbarkeit von Print-Literatur im digitalen Raum reagiert, indem sie der digitalen Erfassung „eine begleitende Theoriedebatte und [...] ein verfeinertes methodisches Rüstzeug" (so die Forderung von Embach/Rapp 2008) zur Seite stellt.

Das Ziel, innovative rechnergestützte Arbeitsplätze zu schaffen, ist dabei alles andere als neu und begleitet die geisteswissenschaftliche Fachinformatik (Texttechnologie, Computerlinguistik, automatische Sprachverarbeitung etc.) als zentrales Desiderat seit vielen Jahren (vgl. etwa Barrow 1997).

Qualitativ neu an dem hier vorgestellten Ansatz ist dagegen die Kombination der folgenden Merkmale eines solchen Arbeitsplatzes:

- weitestgehende Integration digitaler Ausgangsmaterialien (Primär- und Sekundärtexte, Integration anderer Medien, v.a. Karikatur, Photographie, Film, (vgl. Wolff 2008))
- Aufbau auf den heute etablierten texttechnologischen Standards (die XML-Familie als Basisstandard sowie ihre spezifischen Anwendungen in der Texttechnologie wie etwa TEI P5 (Lobin & Lemnitzer 2004)
- Nutzung von Text Mining-Diensten (Mehler & Wolff 2005), die auf den (hier: Kafkaschen) Matrix-Text ebenso bezogen sein können wie auf das zeitlich und material begrenzte Reservoir seiner Echotexte oder, noch weitergehend, einen beliebig konfigurierbaren Texthorizont als Vergleichsmaterial
- der Aufbau auf aktuellen Konzepten der Softwarearchitektur (Reussner & Hasselbring 2006)
- die Bereitstellung als webbasierte Oberfläche, die Kollaboration und damit auch die gemeinsame Texterstellung und -kommentierung unterstützt.

4.1 Texttechnologische Perspektive

Wir gehen davon aus, dass Text Mining-Verfahren (Mehler & Wolff 2005) mittlerweile einen Reifegrad erreicht haben, der sie jenseits rein experimenteller Kontexte (der Texttechnologie-Forschung) als für die Integration in den philologischen Arbeitskontext geeignet erscheinen lässt. Damit geht ein solcher Arbeitsplatz funktional erheblich über die reine Durchsuchbarkeit digitaler Textressourcen (z. B. mit Hilfe von KWIC-Darstellungen, Konkordanzen, Volltextsuche) hinaus.

Texttechnologische Informationssysteme wie der rein webbasierte eHumanities Desktop (Mehler, Gleim et al. 2009) verfügen mittlerweile über eine Vielzahl texttechnologischer Kernfunktionen, die weit über elementare Operationen der Korpuserstellung und -verwaltung hinausgehen. Dies betrifft grundlegende Funktionen der automatischen Annotation textueller Aggregate wie die automatische Spracherkennung, die Satzgrenzenerkennung, das Stemming, die Lemmatisierung, das Part-of-Speech-Tagging, die Eigennamenerkennung oder auch die automatische Segmentierung von Dokumentstrukturen und deren Abbildung auf geeignete Standards wie die TEI P5 (Burnard 2007) bzw. den CES (Ide & Priest-Dorman 1998). Aber auch *Text Mining*-Funktionen wie das *Lexical Chaining* und die hierauf basierende unüberwachte Themenklassifikation sowie die semantische Suche bilden bereits Kernbestandteile des eHumanities Desktops (Gleim, Waltinger, e.a. 2009; Mehler, Gleim, e.a. 2008). Im Folgenden skizzieren wir in Ergänzung zu diesen Kernfunktionalitäten texttechnologische Bausteine als *notwendige* Einheiten

eines *Literary Memory Information System*, welches die Aufdeckung, Verwaltung, Suchbarmachung und literaturwissenschaftliche Erforschung dieser Art von intertextuellen Beziehungen zwischen literarischen Matrixtexten und genreübergreifenden Echotexten ermöglichen soll. Dabei gehen wir auf den Umstand ein, dass das Instrumentarium der automatischen Textanalyse im Bereich literarischer Texte auf ein Datenbeschaffungsproblem stößt, für das es in der bisherigen *Text Mining*-Forschung kaum Lösungsansätze gibt. Den Ausgangspunkt der im Folgenden zu beschreibenden LiMeS-Komponenten bildet ein *Text Mining*-Modell basierend auf dem Intertextualitätsbegriff in Verbindung mit dem Begriff des *Semantic Web* (Mehler 2004; 2005a; 2005b), für den die folgenden Überlegungen maßgeblich sind:

- Anders als bei der klassischen Textkategorisierung (Sebastiani 2002) geht es nicht darum, Texte auf vorgegebene Mengen von Inhaltskategorien abzubilden. In dem anvisierten Informationssystem stellen diese Inhaltskategorien vielmehr jene Erkenntniseinheiten dar, welche seine Anwender *mittels* der LiMeS-seitig zu explorierenden Vernetzung von Matrix- und Echotexten herausarbeiten.
- Anders als bei klassischen Methoden der Textkonversion geht es ferner nicht darum, vorgefundene intertextuelle Beziehungen zwischen Texten hypertextuell zu explizieren. Vielmehr wird die Auffassung zugrundegelegt, dass das Netzwerk der intertextuellen Beziehungen ein semantisches Netzwerk referenzieller und typologischer Verweisbeziehungen von Inhaltskategorien widerspiegelt, welche die Texte manifestieren. Dieser Lesart gemäß kann ein Text A auch dann auf einen Text B intertextuell bezogen sein, wenn der Autor von Text A den Text B nicht kannte, dieser aber zur literaturwissenschaftlichen Klärung der Semantik von A beiträgt, und zwar vor dem Hintergrund einer beiden Texten gemeinsamen oder verwandten Topik.

Die hierfür erforderlichen texttechnologischen Komponenten werden nachfolgend beschrieben. LiMeS als Beispiel für eine Klasse von eHumanities-Informationssystemen erfordert demnach folgende drei Kernkomponenten:

- Ein System für die Verschränkung von literaturwissenschaftlicher Interpretation und maschinellem Lernen (siehe Sektion 4.1.1)
- Social Software für die literaturwissenschaftliche Fachinformatik (siehe Sektion 4.1.2) sowie
- ein System für die ereignisorientierte Exploration intermedialer Relationen.

Nachfolgend beschreiben wir die ersten beiden dieser Komponenten. Abbildung 1 zeigt schematisch die Bandbreite von Systemen und Ansätzen zur Annotation literarischer Daten wie sie in einem LiMeS benötigt werden. Sektion 1.2 greift den wichtigen Aspekt der HCI-Gestaltung auf, dem im Falle von LiMeS besondere Beachtung gilt, da dessen Nutzer gerade keine Texttechnologen oder Informatiker, sondern Literaturwissenschaftler sind.

Abbildung 1: Drei Quellen von Annotationen literarischer Basisdaten und Korpora: (1) klassische texttechnologische und Text Mining-Verfahren; (2) Verfahren, welche aus der Verschränkung von menschlicher Expertise von *Text Mining* hervorgehen sowie (3) Verfahren, die auf *Human Computation* beruhen.

4.1.1 Mining-Algorithmen für die Literaturwissenschaft

Die automatische inhaltsorientierte Annotation literarischer Daten kann nicht durch die einfache Übertragung herkömmlicher *Text Mining*-Ansätze gelingen. Die Interpretationsleistung eines Literaturwissenschaftlers lässt sich nicht in das Gerüst einer vorgegebenen Menge von Inhalts- oder Strukturkategorien zwängen. Auch ist hinsichtlich der Merkmalsselektion, die jedem *Text Mining* vorangeht, zu bedenken, dass das anvisierte LiMeS auf literarische Texte *und* auf Gebrauchstexte zielt. Das bedeutet nicht nur, dass Texte völlig verschiedener Genres zu verarbeiten sind. Vielmehr ist auch zu bedenken, dass *Text Mining*-Verfahren für Gebrauchstexte, nicht aber für literarische Texte konzipiert wurden, deren Sprache im Vergleich zu ersteren eine völlig andere Funktion hat. Zu dieser horizontalen, funktionalen Heterogenität — die auch das Problem der Multilingualität der Zielkorpora einschließt — tritt die vertikale, entstehungsgeschichtliche Vielfalt der Quellentexte hinzu: Denn Matrix- und Echotexte können zeitgeschichtlich verschiedenen Sprachstufen angehören. Vor dem Hintergrund dieses komplexen Gegenstands muss das *Text Mining* neue Wege beschreiten, wenn es im Bereich literarischer Texte vergleichbare Erfolge erzielen will wie im Bereich herkömmlicher Gebrauchstextsorten (Rolf 1993).

Ein solcher Ansatz soll im Folgenden am Beispiel der oben (3.1., 3.2.) beschriebenen *literarischen Gesten* skizziert werden. Seine Grundidee besteht in der Entwicklung eines Bootstrapping-Algorithmus, der das unüberwachte Lernen mit der teilüberwachten Interpretationsleistung von Literaturwissenschaftlern integriert:

1. Ausgehend von einem klassischen Lernszenario zur Identifikation von *literarischen Gesten* wird in einem ersten Schritt ein unüberwachter Klassifikator entwickelt, der potentielle *literarische Gesten* in Inputtexten identifiziert und dem beteiligten Literaturwissenschaftler vorlegt.
2. Der menschliche Experte hat dann die Gelegenheit, die ihm vorgelegten Kandidaten zu bewerten. Doch anders als bei klassischen Feedback-Algorithmen soll der Experte die Möglichkeit erhalten, auf der Basis wohldefinierter Operationen über dem operativen Repräsentationsmodell, gestaltcharakterisierende Transpositionen und vergleichbare Modifikationen der vorgelegten Kandidaten zu identifizieren. Beruht das Repräsentationsmodell beispielsweise auf

einem *bag of words*-Modell, so besteht eine Transposition etwa in dem Austausch oder dem Löschen merkmalbildender Wörter.

3. Die Transpositionen werden im nächsten Schritt dem Klassifikationsalgorithmus zugänglich gemacht, der sie für die neuerliche Identifikation von Relais-Instanzen nutzt. Durch die Iteration dieses Verfahrens besteht die Möglichkeit einer Konvergenz der Klassifikationsleistung bei einer feingliedrigen Abstimmung des zugrundeliegenden Repräsentationsmodells aufgrund der Interpretationsleistungen des Experten.

Dieses Szenario einer fortgesetzten Verschränkung von unüberwachtem maschinellem Lernen einerseits und Rückkoppelung an die Interpretationsleistung des Experten andererseits weicht insofern von klassischen Lernszenarien ab, als hier auch auf Seiten des Experten nicht die Kenntnis des Kategoriensystems als Bezugssystem für seine Transpositionsleistung vorausgesetzt wird. Dieses wird gewissermaßen erst im Zuge der Arbeit mit dem Algorithmus (Stichwort 'dynamische Kategorisierung') erschlossen und strukturiert. Im Extremfall koppeln sich zwei unüberwachte Agenten, ein künstlicher und ein menschlicher, wobei letzterer aufgrund seiner maschinell nicht einholbaren Intuition dem dynamischen Prozess der Exploration und Kategorisierung von Relais eine Richtung gibt. Zielgröße dieses Algorithmus sind Relais, über deren Existenz oder Tragweite der Wissenschaftler im Voraus keine (hinreichende) Kenntnis besitzt. Dabei handelt es sich um Relais, die erst im Zuge der Arbeit mit LiMeS identifiziert werden und daher möglicherweise einer enormen Fluktuation ausgesetzt sind. Der Schwerpunkt dieses Verfahrens liegt daher auf der Exploration zuvor unbekannter, überraschender, neuartiger literarischer Strukturen. In diesem Sinne ist von einem Literatur-Mining zu sprechen.

Auch wenn ein solcher Ansatz vielversprechend und wegen der Offenheit literarischer Interpretationsleistungen unabdingbar ist, basiert er letztlich auf der Koordination der Klassifikationsleistung eines menschlichen und eines künstlichen Agenten. Es ist also stets ein Literaturwissenschaftler, der in Kooperation mit dem Text-Mining-System tritt. Soll nun die Annotationsleistung einer solchen Paarung vergrößert werden, müssen weitere Literaturwissenschaftler hinzutreten. Das bedeutet aber, dass für ein Mehr an annotierten Daten der Einsatz menschlicher Experten linear zu vergrößern ist. Gerade für den vorliegenden Gegenstandsbereich mit seinen großen Mengen an multimedialen Daten, auf die LiMeS fokussiert, ist dies ein schwieriges Szenario. Zwar bedeutet diese Einschätzung keine prinzipielle Abkehr von Text-Mining-Algorithmen. Wenn es jedoch darum geht, eine Vielzahl hochwertiger Annotationen zu erzeugen, welche letztlich den Mehrwert eines LiMeS ausmachen, dann sind in Ergänzung hierzu alternative Wege zu beschreiten. Eine solche Alternative skizziert die folgende Sektion unter dem Stichwort der *Social Software*.

4.1.2 Exploration literarischer Daten im großen Maßstab: Social Software für die literaturwissenschaftliche Fachinformatik

Eine Bezugsgröße für die texttechnologische Vorverarbeitung literarischer Daten besteht in der Nutzbarmachung von *Social Software* (Bächle 2006) im Rahmen des so genannten *crowdsourcing* von Annotationsleistungen. Es geht dabei um die computergestützte Delegation von Annotationsaufgaben an das *Human Computation* (von Ahn 2008), an Gruppen von Annotatoren also, die auf der Basis eines *game with a purpose* (von Ahn & Dabbish, 2008; von Ahn, Liu, e.a., 2006) eine Aufgabe lösen, deren Lösung kaum oder gar nicht automatisierbar ist. Im vorliegenden Fall handelt es

sich dabei um die Annotation von Interpretationen literarischer Daten – im Hinblick auf die Annotation einzelner oder Gruppen semiotischer Aggregate, und zwar bezogen auf deren syntaktische, semantische oder pragmatische interne oder externe Strukturierung. Für solche Ansätze existieren mittlerweile erfolgreiche webbasierte Plattformen wie *mechanical turk* (http://www.mturk.com), auf denen Freiwillige – gegen symbolisches Honorar – als *human intelligence task* (HITs) bezeichnete Aufgaben wie Bildannotation, Textproduktion oder Befragungen – erledigen.

Unter dem Blickwinkel inhaltsbasierter Intertextualität und Intermedialität ist es die externe Strukturierung semiotischer Aggregate die im Fokus der anvisierten LiMeS-Instanz steht. Hier steht man unter anderem vor der Aufgabe der Annotation intermedialer Relationen, von Relationen zwischen Bildzeichen einerseits und Textzeichen andererseits. Es geht beispielsweise um die Frage, in welchem Segment eines Matrixtexts welches Bild beschrieben oder auch nur indirekt thematisiert wird. Offenbar handelt es sich bei der Beantwortung solcher Fragen um Aufgaben, die noch lange einer Automatisierung harren werden. Damit steht die literaturwissenschaftliche Fachinformatik jedoch vor dem Problem, einerseits große Datenmengen (etwa in Form von Trainingsdaten) für die Exploration intermedialer Relationen zu benötigen, diese aber nur über den aufwendigen Weg der intellektuellen Annotation beschaffen zu können. So werden beispielsweise im großen Maßstab textuelle Annotationen von Bildern benötigt, um deren (teil-)automatische Relationierung mit anderen Bild- oder Textmedien zu betreiben. Unter der Voraussetzung, dass sämtliche der zu verarbeitenden Bildmedien textuell annotiert sind, kann diese Aufgabe im Prinzip mittels etablierter Verfahren des *Text Mining* angegangen werden, so z.B. mit einer latenten semantischen Analyse (*latent semantic analysis*) (Berry, Drmač, e.a. 1999), die neben Texten auch Bilder, und zwar mittels ihrer textuellen Repräsentationen verarbeitet. *Wie aber gelangt man zu solchen großmaßstäblichen und zugleich qualitativ hochwertigen Annotationen?* Genau diese Frage führt unter dem vorliegenden Szenario auf das Konzept des *game with a purpose* (von Ahn & Dabbish 2008). In dieser Sektion skizzieren wir kurz die Kernbestandteile einer solchen Art von "Spiel mit literaturwissenschaftlichem Zweck", das zu einem wesentlichen Baustein jeder Art von LiMeS zählen darf, und zwar zur Gewährleistung von Annotationen literarischer Daten, die qualitative Tiefe mit quantitativer Breite verbinden und derzeit nicht anders für die literaturwissenschaftliche Fachinformatik zu beschaffen sind.

Allgemein gesprochen kann ein zweckorientiertes Annotationsspiel durch Implementierung zweier Arten von Informationsteilsystemen umgesetzt werden. Dies betrifft zum einen das Kernspiel selbst, das zwei oder mehr Spielpartner in einem Annotationspiel vereinigt und deren Spielresultate in Form von Annotationen an einer Serie von Zielobjekten (etwa Bilder, Texte oder Bild-Textbeziehungen) verwaltet. Zum anderen betrifft dies ein übergeordnetes Verwaltungsprogramm, dass Spieler und Zielobjekte auswählt, um deren Spielresultate schließlich je Zielobjekt über viele Spielabläufe hinweg in eine Art *Medianannotation* zu überführen, die als repräsentative Annotation dieses Objekts gilt. Diese beiden Informationsteilsysteme werden im Folgenden kurz beschrieben:

- *Annotationsspiele mit literaturwissenschaftlichem Zweck:* Den Ausgangspunkt bildet die Gestaltung von Spielen mit dem Zweck der Annotation literarischer Daten. Ein solches Spiel bezeichnen wir enger als literarisches Annotationsspiel bzw. als *LitErary Data Annotation* (LEDA) *game*. Anders als im Falle der von von Ahn (2006) und von Ahn, Liu, e.a. (2006) betrachteten Spiele zielt LEDA nicht auf die bloße Annotation von Aggregaten als Ganzes oder auf deren Segmentierung (wie im Falle der Bildsegmentierung). Vielmehr geht es darum, das *Human*

Computation für die gleichzeitige Segmentierung *und* Relationierung nicht explizit verknüpfter Aggregate nutzbar zu machen. Dabei wird der Annotationsprozess unter mindestens zwei Spielpartnern aufgeteilt, und zwar so, dass ein Spieler in der Rolle des Spielleiters bzw. Überwachers sein Gegenüber in der Rolle des Annotators so anleitet bzw. führt, dass dieser möglichst gute Annotationsleistungen erbringt. Im Falle eines literarischen Annotationsspiels steht das Spieldesign vor der besonderen Aufgabe, zugleich ikonographische und symbolische, textuelle Informationen den Spielpartnern darzubieten. Als Seiteneffekt solcher Annotationsspiele wird die korrekte Segmentierung und Relationierung einer Vielzahl von Text- und Bildmedien erwartet. Wegen des grundlegenden Spielcharakters des gesamten Annotationsprozesses wird weiterhin eine besondere Motivation und damit Qualitätssteigerung der Annotationen erwartet. Von Ahn, Liu, e.a. (2006) zeigen, dass *games with a purpose* es erlauben, hochkomplexe Annotations- und Suchaufgaben an Gruppen von Annotatoren zu delegieren, insbesondere in solchen Bereichen, in denen nur geringe Mengen annotierter Daten zur Verfügung stehen.

- *Medianannotationen:* Ein zweites Informationsteilsystem des Social Taggings als Bestandteil des anvisierten LiMeS thematisiert die Verwaltung des *community building* unter den LEDA-Spielern und ihrer Spielergebnisse. Bei einer gegebenen Zahl von Spielern besteht eine Vielzahl von Möglichkeiten der Auswahl von Paarungen, die – graphentheoretisch gesprochen – unterschiedliche Graphen aufspannen, von denen je nach Aufgabe, Gruppenzusammensetzung und tatsächlicher Spielerverfügbarkeit jene Paarungen auszuwählen sind, die einen effizienten Spielablauf garantieren. Es geht also um die Implementierung eines *Annotation Game Management System* (AGMS), das ferner die Speicherung, Verwaltung und das Retrieval von Spielergebnissen ermöglicht. Das AGMS hat unter anderem die Aufgabe, je Spielrunde Spielpartner und Zielobjekte zu selektieren, deren Annotationen zu verwalten und hierauf basierend zielobjektbezogene Medianannotationen zu berechnen. Es geht ferner um die Aufgabe der Nachverarbeitung von Annotationen, um die Kontrolle und Bewertung von Spielen gegebenenfalls mit dem Ziel, defizient annotierte Zielobjekte zum Gegenstand weiterer Spielrunden unter Wahl neuer Paarungen zu machen. Aus informatorischer Sicht bildet die Berechnung und Verwaltung von Medianannotationen die größte Herausforderung im Zuge der Implementierung eines AGMS. Zur Berechnung von Medianannotationen steht im Prinzip das Instrumentarium der Graphähnlichkeitsmessung zur Verfügung. Das mag zwar naheliegen – da Annotationen Graphen aufspannen –, jedoch auch paradox erscheinen, wenn man bedenkt, dass die Berechnung von Graphisomorphien NP-hart ist (Dehmer 2005). Es steht jedoch nunmehr eine Reihe von Graphähnlichkeitsalgorithmen zur Verfügung, welche in geschickter Weise an bestimmte Graphklassen angepasst sind und Näherungen für Ähnlichkeitswerte von Graphen zu berechnen erlauben, deren Komplexität weitaus geringer ist (Bunke & Günter 2001; Dehmer & Mehler 2007; Mehler, Geibel e.a. 2007).

Das AGMS mit seinem integrierten Annotationsspiel soll zur Segmentierung und Relationierung von Zeichnungen, Karikaturen, Bildern, Photographien, literarischen und Gebrauchstexten eingesetzt werden. Unter dem Aspekt der *Social Software* geht es also um die Entwicklung eines Bausteins der literaturwissenschaftlichen Fachinformatik, der gerade nicht auf die ausschließlich maschinelle Exploration intermedialer oder bloß intertextueller Relationen zielt – etwa unter der Entwicklung und Anwendung bekannter, im vorliegenden Anwendungsbereich jedoch hochwahrscheinlich hochgradig fehleranfälligen Algorithmus des maschinellen Lernens.

Vielmehr geht es um die Implementierung eines Algorithmus des *human computation* für das *crowdsourcing* von semantischen Annotationen, welche die Interpretationsleistungen von Literaturwissenschaftlern bzw. Textrezipienten abbilden. Dieser Weg erscheint uns unumgänglich zur Exploration der algorithmisch noch immer als uneinholbar geltenden menschlichen Kompetenz in Bezug auf das Verstehen, Explorieren und Verknüpfen semiotischer Aggregate. Die Implementierung einer solchen Software verspricht, einen Bereich der Intermedialität von Zeichen in einem sehr viel größeren Umfang zu erschließen als es bisher gelungen ist, einen Bereich, der seiner Komplexität wegen allen bisherigen Projekten im Dunstkreis literaturwissenschaftlichen Fachinformatik verschlossen geblieben ist.

4.2 Softwarearchitektur

Das *Literary Memory Information System* ist modular konzipiert. Seine Bausteine sollen als *notwendige* Einheiten die Aufdeckung, Verwaltung, Suchbarmachung und literaturwissenschaftliche Erforschung dieser Art von intertextuellen Beziehungen zwischen literarischen Matrixtexten und genreübergreifenden Echotexten ermöglichen. Dabei gehen wir auf den Umstand ein, dass das Instrumentarium der automatischen Textanalyse insbesondere im Bereich der Merkmalsselektion am Beispiel literarischer Texte auf Probleme stößt, für die es in der bisherigen Forschung zum *Text Mining* kaum Lösungsangebote gibt. Zur Bewältigung dieser Aufgabenlast beschreiben wir LiMeS als ein Informationssystem, das dem heute gebräuchlichen Mehrebenenmodell der Softwarearchitektur (Wolff 2004; Reusner & Hasselbring 2006) folgt und dabei auch Elemente serviceorientierter Architekturen realisiert. Abbildung 1 oben zeigt bereits die wesentlichen Schichten:

1. Die *Persistenzebene* der Datenverwaltung, die in LiMeS die Aufgaben der Corpus- und Medienverwaltung übernimmt. Neben den in LiMeS selbst verwalteten Texten und Medien müssen auch externe Ressourcen (Textarchive, Medienbibliotheken etc.) dynamisch eingebunden werden können.
2. Die in den Abschnitten 4.1.1 und 4.1.2 beschriebenen Komponenten der automatischen, teilautomatischen und intellektuellen Analyse, Relationierung und Annotation machen einerseits wesentliche Teile der *Applikationslogik* von LiMeS aus und bilden gleichzeitig die Brücke in die Interaktionsschicht (Benutzerschnittstelle), da intellektuelle Annotation Systeminteraktion erforderlich macht.
3. In der *Interaktionsschicht* müssen geeignete Sichten für die für LiMeS vorgesehenen vielfältigen Benutzerrollen geschaffen werden: Die Spanne reicht hier von der Text- und Medienverwaltung.

Da eine Nutzung der mit LiMeS erreichten Ergebnisse nicht nur über die unmittelbare Interaktion mit dem System, sondern auch auf dem Weg der Diensteintegration in andere Systeme denkbar erscheint, kommt mittelfristig auch die Erweiterung um Dienste im Sinne *der Service Oriented Architecture* (SOA) in Betracht (Wolff 2003).

Es ist geplant, die Software-Architektur von LiMeS so zu implementieren, dass sie als Aufsatz auf dem eHumanities-Desktop (Mehler, Gleim e.a. 2009) ruht. Das bedeutet insbesondere, dass LiMeS den Desktop als Korpusmanagementsystem wie auch als System zur Verwaltung von Nutzungsrechten verwendet. Auch ist damit die Möglichkeit gegeben, das umfangreiche System an text-

technologischen Methoden der Korpusvorverarbeitung und Korpusanalyse, das der Desktop bietet, unmittelbar zu nutzen. Die modulare und objektorientierte Architektur des rein webbasierten eHumanities Desktop unterstützt gerade solche Erweiterungen, wie sie mit der Implementierung von LiMeS am Beispiel des Werkes von Kafka geplant sind. Diesen Weg zu beschreiten wird in diesem Zusammenhang eine der kommenden texttechnologischen Hauptaufgaben sein.

5 Ausblick

In diesem Beitrag fokussieren wir auf die literatur- und medientheoretische Konzeption von LiMeS und die Möglichkeiten ihrer Umsetzung mit Hilfe von Verfahren aus der Texttechnologie und des sich schnell entwickelnden Feldes sozialer webbasierter Anwendungen. Im Hintergrund steht dabei die Annahme, dass Hypertextrelationierung, Medieneinbindung, automatische Erzeugung und intellektuelle Annotation in transbiblionomen Räumen zu materiellen Ergebnissen führen, die einen erheblichen qualitativen Unterschied zu den Ergebnissen traditioneller Editionsphilologie und deren Fortführung mit digitalen Mitteln aufweisen. Dies führt schließlich auch zu neuen Nutzungsmöglichkeiten für den forschenden und lernenden Anwender. Für die Operationalisierung dieses Modells wird daher die Bezugnahme auf nutzerseitige Anforderungen (vgl. Toms & O'Brien 2008) an e-Humanities-Anwendungen ein wesentlicher Erfolgsfaktor sein.

6 Literatur

Bächle, M. (2006), Social software. In: Informatik Spektrum, 29(2):121-124.

Barrow, J. (1997), A Writing Support Tool with Multiple Views. Computing and the Humanities, 31(1), 13-30.

Baßler, M. (2005), Die kulturpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie, Tübingen: Francke.

Benjamin, W. (1981), Franz Kafka. Zur zehnten Wiederkehr seines Todestages, in: Benjamin über Kafka. Texte, Brieferzeugnisse, Aufzeichnungen, ed. Hermann Schweppenhäuser, Frankfurt a.M.: Suhrkamp, 9-38.

Berry, M.W., Z. Drmač, & E. R. Jessup (1999), Matrices, vector spaces, and information retrieval. SIAM Review, 41(2):335-362.

Biemann, C.; Bordag, S., Quasthoff, U., & Wolff, C. (2004, Mai 2004), Web Services for Language Resources and Language Technology Applications. Paper presented at the Proceedings Fourth International Conference on Language Resources and Evaluation [LREC 2004], Lissabon.

Bolter, J. D. (2001), Writing Space. Computers, Hypertext and the Remediation of Print, Mahwah, NJ/London: Erlbaum.

Bunke, H. & Günter, S. (2001), Weighted mean of a pair of graphs. Computing, 67(3), 209-224.

Burnard, L. (2007), New tricks from an old dog: An overview of TEI P5. In L. Burnard, M. Dobreva, N. Fuhr, and A. Lüdeling, editors, Digital Historical Corpora- Architecture, Annotation, and Retrieval, number 06491 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Dehmer, M. (2005). Strukturelle Analyse Web-basierter Dokumente. Multimedia und Telekooperation. Berlin: DUV.

Dehmer, M. & A. Mehler, A (2007), A new method of measuring the similarity for a special class of directed graphs. Tatra Mountains Mathematical Publications, 36:39-59.

Deleuze, G.; Guattari, F. (1976), Kafka. Für eine kleine Literatur, Frankfurt a.M.: Suhrkamp.

Derrida, J. (1967), Grammatologie, Frankfurt a.M.: Suhrkamp.

Derrida, J. (1984), Two words for Joyce, in: Attridge, D. & Ferrer, D., eds., Poststructuralist Joyce: Essays from the French, Cambridge: CUP, 145-159.

Dotzler, B. J. (2006), Pervasive Bureaucracy: The Case of Herman Hollerith., in: Austriaca. Cahiers universitaires d'information sur l'Autriche, no. 60, 45-52.

Dotzler, B. J. (2008), Kafka zwischen den Medien, in: J. Paech; J. Schröter, eds., Intermedialität – Analog/Digital, Paderborn: Fink, 181-192.

Embach, M., Rapp, A. (2008), Rekonstruktion und Erschließung mittelalterlicher Bibliotheken Neue Formen der Handschriftenpräsentation. Berlin: Akademie-Verlag [=Beiträge zu den Historischen Kulturwissenschaften, Bd. 1].

Gleim, R., Waltinger, U., Ernst, A. Mehler, A., Esch, D., & Feith, T. (2009), The eHumanities desktop—an online system for corpus management and analysis in support of computing in the humanities. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics EACL 2009, 30 March - 3 April, Athens

Hofmannsthal, H. von (1907), Der Dichter und diese Zeit, in: Die Neue Rundschau. XVIIIer Jg. Der freien Bühne, Bd. 1, H.3, 257-276.

Ide, N., & G. Priest-Dorman, G. (1998), Corpus encoding standard. http://www.cs.vassar.edu/CES/.

Jenny, L. (1979), La stratégie de la forme, in: Poétique, no. 27, 257-281.

Kafka, F. (1958), Briefe, hrsg. v. Brod, M., Frankfurt a.M.: S. Fischer.

Kafka, F. (1976), Briefe an Felice, hrsg. v. Born, J., Frankfurt a.M.: S. Fischer.

Lachmann, R. (1984), Ebenen des Intertextualitätsbegriffs, in: Stierle, K.H.; Warning, R., Das Gespräch, München: Fink, 133-138.

Lobin, H. & Lemnitzer, L., eds., (2004). Texttechnologie. Perspektiven und Anwendungen, Tübingen: Stauffenburg.

Mehler, A. (2004a): Textmining. In: Lobin, Henning und Lothar Lemnitzer (eds.): Texttechnologie. Perspektiven und Anwendungen, Seiten 329-352. Stauffenburg, Tübingen.

Mehler, A. (2004b), Textmodellierung: Mehrstufige Modellierung generischer Bausteine der Textähnlichkeitsmessung. In Mehler, A. and Lobin, H., editors, Automatische Textanalyse: Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte, 101-120. Wiesbaden: Verlag für Sozialwissenschaften.

Mehler, A. (2005a), Lexical chaining as a source of text chaining. In Patrick, J. and Matthiessen, C. (eds.), Proceedings of the 1st Computational Systemic Functional Grammar Conference, University of Sydney, Australia, pages 12-21.

Mehler, A. (2005b), Zur textlinguistischen Fundierung der Text- und Korpuskonversion. Sprache und Datenverarbeitung, 1:29-53.

Mehler, A., & Wolff, C. (2005), Einleitung: Perspektiven und Positionen des Text Mining. LDV-Forum, 20(1), Einführung in das Themenheft Text Mining, 1-18.

Mehler, A., Geibel, P., & Pustylnikov, O. (2007), Structural classifiers of text types: Towards a novel model of text representation. LDV Forum – Zeitschrift für Computerlinguistik und Sprachtechnologie, 22(2):51-66.

Mehler, A., Gleim, R., Ernst, A., & 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., Ernst, A., Waltinger, U, Ernst, A., Esch, D., & T. Feith (2009), eHumanities Desktop — eine webbasierte Arbeitsumgebung für die geisteswissenschaftliche Fachinformatik. In Proceedings of the Symposium "Sprachtechnologie und eHumanities", 26. Und 27. Februar, Duisburg-Essen University.

Meister, J. C. (2005), Projekt Computerphilologie Über Geschichte, Verfahren und Theorie rechnergestützter Literaturwissenschaft, in: Segeberg, H. & Simone Winko, S., eds., Literarität und Digitalität. Zur Zukunft der Literatur, München: Fink, 315-341.

Pêcheux, M. (1983), Über die Rolle des Gedächtnisses als interdiskursives Material, in: Geier, M., Woetzel, H., eds. (1983), Das Subjekt des Diskurses. Beiträge zur sprachlichen Bildung von Subjektivität, Berlin: Argument-Verlag, 50-58.

Porombka, S. (2001), Hypertext. Zur Kritik eines digitalen Mythos, München: Wilhelm Fink.

Rayward, W. B. (2008), European modernism and the information society: informing the present, understanding the past, Aldershot: Ashgate.

Reussner, R., & Hasselbring, W. (eds.). (2006), Handbuch der Software-Architektur. Heidelberg: dpunkt.

Rolf, E. (1993), Die Funktionen der Gebrauchstextsorten. Berlin/New York: De Gruyter.

Schmidt, T., & Wolff, C. (2004), Dokumentbezogenes Wissensmanagement in dynamischen Arbeitsgruppen. Text Mining, Clustering und Visualisierung. In B. Bekavac, J. Herget & M. Rittberger (Eds.), 9. Internationales Symposium für Informationswissenschaft (ISI 2004) (Vol. Information zwischen Kultur und Marktwirtschaft, pp. 317-336). Chur (CH): UVK.

Sebastiani, F. (2002), Machine learning in automated text categorization. ACM Computing Surveys, 34(1):1-47.

Toms, E. G., & O'Brien, H. L. (2008). Understanding the information and communication technology needs of the e-humanist. Journal of Documentation, 64(1).

Topia, A. (1984), The Matrix and the Echo: Intertextuality in Ulysses, in: Post-Structuralist Joyce: Essays From the French, eds. D. Attridge/D. Ferrer, Cambridge: Cambridge UP, 103-125.

von Ahn, L. (2006), Games with a purpose. IEEE Computer Magazine, 39(6):92-94.

von Ahn, L. (2008), Human computation. In ICDE, pages 1-2. IEEE, 2008.

von Ahn, L. & L. Dabbish (2008), Designing games with a purpose. Commun. ACM, 51(8):58-67.

von Ahn, L., Liu, R, & Blum, M. (2006), Peekaboom: a game for locating objects in images. In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 55-64, New York: ACM Press.

Wagner, B. (2006a), Kafkas phantastisches Büro, in: K. Scherpe/E. Wagner, eds., Kontinent Kafka. Mosse Lectures an der Humboldt-Universität zu Berlin, Berlin: Vorwerk 8, 104-118.

Wagner, B. (2006b), Connecting Cultures. Heinrich Rauchberg, Franz Kafka, and the Hollerith Machine, in: Austriaca. Cahiers universitaires d'information sur l'Autriche, no. 60, 53-68.

Wagner, B./Reinhard, T. (2007), Das Virtuelle Kafka-Bureau, in: I. Jonas, ed., Sinn und Nutzen von Datenbanken in den Geisteswissenschaften, Frankfurt: Peter Lang, 95-114.

Wolff, C. (2003), Web Services im e-Learning und e-Publishing. In: K.-P. Fähnrich & H. Herre, eds., Content- und Wissensmanagement. Leipzig: Leipziger Informatik-Verbund / Universität Leipzig, 123-132.

Wolff, C. (2004), Systemarchitekturen. Aufbau texttechnologischer Anwendungen. In L. Lemnitzer & H. Lobin (Eds.), Texttechnologie. Perspektiven und Anwendungen (pp. 165-192). Tübingen: Stauffenburg.

Wolff, C. (2005), Generierung ontologischer Konzepte und Relationen durch Text Mining-Verfahren. Paper presented at the Knowledge eXtended. Die Zusammenarbeit von Wissenschaftlern, Bibliothekaren und IT-Spezialisten, Jülich.

Wolff, C. (2008), Veränderte Arbeits- und Publikationsformen in der Wissenschaft und die Rolle der Bibliotheken. In E. Hutzler, A. Schröder & G. Schweikl (eds.), Strategien zum Aufbau digitaler Bibliotheken (pp. 157-172). Göttingen: Universitätsverlag Göttingen.