kafkabureau.net

A Literary Memory Exploration System (LiMES) for ICT-based Literary and Media Studies

(Bernhard Dotzler, Regensburg / Alexander Mehler, Frankfurt / Benno Wagner, Taipei / Christian Wolff, Regensburg)

This interdisciplinary (cross-domain) project opens groundbreaking perspectives for literary scholarship. It sets out to develop a novel architecture of information systems for exploring multilingual, multimedia and multilevel (M^3) universes of literary data. The work of Franz Kafka and its multilingual and multicultural field of emergence (the Habsburg Monarchy and the early Czech Republic, the Central European situation before, during and after WWI) will serve (1) as a paradigmatic case and (2) as an object of its own right. Building on a European author of global impact, kafkabureau.net will be the first application and instantiation of the LiMES environment for scholarly work. It goes far beyond current approaches to literary scholarship platforms (e.g. goethezeitportal.de, or hypernietzsche.org). Combining text analytics and semantic web structures with the cooperative potential of social software and offering multiple modes of presentation and interaction, kafkabureau.net will provide the first digital workspace opening up the specific (i.e. multivocal) informational potential of *modern literary text as a tool of cultural memory* to researchers, students, and public users of all levels. The successful implementation of this project will have far-reaching consequences for future scholarly and, more generally, cultural modes of making use of literary print text in a post-Gutenberg culture. Its potential impact reaches beyond the academic sphere, e.g. by offering a new technology to implement multicontextual readings of literature in secondary education.

Challenges

Our project takes off from three observations:

1) *the proto-informatic potential of modern literary text.*

Challenged by emerging technical and electronic media (photography and film, telegraph and phone, electric counting machines), the most sophisticated layers of modern literature anticipate in the medium of print some basic features and functions of today's digital media. Authors like Joyce, Proust, or Kafka redesign the intertextual dimension of literature by making it function beyond the realm of the aesthetic. The poetic word is no longer restricted to quoting and rewriting literary and cultural tradition. Instead, it is now devised to *trigger complex series of echoes in a synchronic field of statements and voices*, reaching across the entire spectrum of discourses and media. By virtue of such "complex multiphonic allusiveness", printed text may be regarded as a "precursor of hypermedia" (Theall 2003-4), as it succeeds in implementing a monomedial intermediality (Dotzler 2008) capable of 'storing' and 'processing' information of any kind and origin. In 1984, Jacques Derrida elaborated this analogy by maintaining that Joyce was "in advance, decades in advance, to compute you, control you, forbid you the slightest inaugural syllable be-

cause you can say nothing that is not programmed on this 1000th generation computer [*Ulysses*, *Finnegans Wake*] beside which the current technology of our computers and micro-computerfied archives and translating machines remain a bricolage of a prehistoric child's toys" (Derrida 1984, 147). The last part of this statement points to the crucial challenge of our project.

2) *the (medial and conceptual) inability of print based literary scholarship to conceive of, explore and make use of this transbiblionomic potential of modern literature*.

Obviously, print-based research has limited access to the informational potential described above. These limits include (a) analytical access to context data (the cultural 'archive', to which literary texts are related), (b) display of results, and (c) feedback of research results to the scholarly community. In brief: although theorists like Kristeva or Barthes thought of literary text as a set of semiotic nodes emerging from the wider network of pre-existing cultural text, this concept of open intertextuality was eventually discarded, more or less explicitly, for lack of informational manageability. As the cultural 'archive' – to which literature was imagined to be related as an "echo chamber" (Barthes 1978, 81) – exceeded the digestive power of any scholarly authority, intertextuality was to be "tamed" (Lachmann 1984, 137) by connecting it to the consciousness, intentions, and preferences of an author subject, and/or the more or less 'adequate' operations of a reader subject in retracing the intertextual connections intended by the author (→ unjustified limitation of primary data). In this framework, literary texts are questioned for their *representative* potential as acts of communication. Depending on the reader's level of sophistication, this communication would result in one or several meanings or the insight into the impossibility of meaning, while literature's full *connective* potential of data storage and processing is necessarily ignored. At the same time, under these conditions any new reading is only loosely and stochastically informed by previous readings (→ unjustified limitation of secondary data).

3) *the failure so far of digitally based literary scholarship to address successfully the challenge of proto-informatic writing*.

With these observations (1, 2) in mind, we can look at the transcription (Jäger e.a. 2008) or remediation (Bolter 2001) of literary texts into digital media and observe a symptomatic bifurcation. On the one hand, restricting its operations to "traditional objects of philology" (Meister 2005), the new discipline of computer philology gave birth to a huge variety of digital projects that successfully enhance the biblionomic dimensions of literary texts. On the other hand, it was precisely the enthusiastic reception of hypertext and hypermedia in the philologies that stood in the way of tackling successfully the transbiblionomic, intertextual potential of literary texts. Too rashly influential authors such as Landow (1997, 2f.) 'recognized' the *inter*textual dynamics of printed text as an anticipation of digital *hyper*textuality. This quick analogy only served to increase the complexity gap between the proto-informatic logic of polyphonic print texts and the current potential of computer software that Derrida had pointed out in such a poignant way.

Goals

Against this problematic background our project sets out to develop an innovative digital research environment. Using the work of Franz Kafka as both a paradigmatic and culturally highly significant case, we aim to build a *Literary Memory Information System* (LiMES) to systematically address, for the first time, *four crucial desiderata of contemporary and future literary studies*:

1. to unfold and display the proto-informational, essentially connective textual logic of modern literature, and enable the development of adequate techniques of reading/using such text

2. to make modern literary text accessible as a unique memory tool for the historical study of media and culture

3. to use step (2) as a point of departure for a media theoretical reassessment of the present-day cultural situation

4. to enable an up-to-date philological research logic, based on the transcription resp. remediation of printed text in the digital medium, thereby responding to the overall increased availability and usage of texts in the digital space.

Kafka as a Paradigmatic Case

The work of Kafka provides an ideal test case for our project. Benjamin's famous remark – that "Kafka's entire work constitutes a codex of gestures" (Benjamin 1981, 18) – may be well transferred from Kafka's protagonists to his text itself: to the operation of connecting to other texts and media that is constantly being executed by certain elements of this text. According to L. Jenny, the secondary language (J. Lotman) of literature consists of a set of "gestes littéraires" connected to a background of archetypes that they keep reproducing, transforming, transcending (Jenny 1976, 257). However, Kafka's *poetic gestures* go beyond this definition – they are based on the insight that in the age of hyperliterarization and emerging electronic media this background has already disintegrated. First of all, it is the polycentric, mass-media based nationalism of the Habsburg Empire that exploits archetypical heritage as a quarry of fragmented "archaisms" to which one may "impart a contemporary sense" (Deleuze/Guattari 1986, 24), according to occasion and point of view. Therefore, the intertextual resonance of Kafka's poetic gestures – such as the wall building Chinese, imitating apes, investigating dogs, and the like – runs through and hence connects occidental and oriental mythology, historical narratives, contemporary science and philosophy, and the stereotypes of the mass media (foremost: the printed press) with their immeasurable field of utterances. But Kafka's poetic gestures do not only consist of storing and, as recent research has shown, processing (rearranging) the cultural context of his time (Neumann 1996; Wagner 2006); they also imply an ongoing self-

reflection of the medial and institutional foundations of the aforementioned performance. Whenever Kafka writes about telephone switchboards (*America*), punching machines (*In the Penal Colony*), or the bureaucratic circulation of files (*The Castle*), he transcribes the media dispositives of his time into an autoreferential protocol of his own mode of connective, intermedial writing. Due to the immense richness of its significant intertextual resources, and to the enormous poetological and media theoretical awareness implied in his poetic text, Kafka's work is an ideal challenge for the development of our novel research tool, and, last not least, a cultural treasure in its own right, justifying any effort to open it up for public access.

From high risk to high gain: mutual challenge and recursive interaction as principles for cross-domain cooperation

"In language processing the use of computers is not aimed toward less human effort, or for doing things faster and with less labor, but for more human work, more mental effort; we must strive to know more systematically, deeper, and better, what is in our mouth at every moment, the mysterious world of our words" (Roberto Bursa). Approaching the specific informational potential of modern literature (its intermedial and interdiscursive multivocality) bears as many risks as it has inspired phantasies of cutural and even politic gains. The philosophy and architecture of this project are designed to avoid the two main roads to underachievement, or even complete failure, in this respect: (1) adapting the level of analytical complexity of philological research frameworks to the limited potential of artificial intelligence in understanding complex, multivocal texts (underachievement); (2) to overtax the potential of the digital writing space, esp. of 'hypertext', in displaying the semiotic structure of such texts. Instead, our cooperation across the humanites/information divide is based on repeatedly asking three questions: (a) in what ways can existing information technology by modified or further developed to better suit the specific needs of high level literary scholarship? (b) in what way can the theory and methodology of literary scholarship be modified or developped to make best possible use of existing information technology? (c) where precisely in the process of comptuer based philology do we have to locate the interface between human and artificial intelligence (problem solving) potential? In the framework of kafkabureau.net/LiMES we have designed four paradigmatic subprojects that will enable us (1) to test our theoretical hypotheses and our practical guidelines of cooperation in the framework of four cutting edge research perspectives within Kafka scholarship; and (2) to build up a critical mass and complexity of data for the take-off of kafkabureau.net as a permanently expanding digitital workspace. These subprojects – Kafka's Intermedia Network/Media Archaeology *(Bernhard Dotzler);* Kafka's War Files – Literature as Cryptology *(Benno Wagner);* Kafka's Czech Context – an Intermedial Exploration *(Marek Nekula);* Negotiations with Kafka – The Language of

Zionism *(Andreas Kilcher)* – will thus function as practical relay units between the philological work packages (1 and 2, seeproposal), and the informational work packages connecting to them.

Exploring Intertextual Semiotic Space: New Developments in Information Systems

The planned project aims at developing a novel architecture of information systems for exploring multilingual, multimedia and multilevel (M^3) universes of literary data whose focal point is given by a single literary work (Franz Kafka). This data is *multilingual* as it comprises primary and secondary text in different languages (primarily German and Czech). It is *multimedia* as it includes various kinds of non-textual data (ranging from images and videos to sound tracks). Finally, it is *stratified into multiple levels* as it concerns experts (scholars from divergent disciplines) as well as non-experts (readers, students etc.). In order to master this multiplicity the project develops a framework for a L*iterary* Me*mory Information* S*ystem* (LiMES) for archiving, processing and exploring a challenging kind of data which has been neglected by the range of existing information systems in spite of its outstanding cultural significance. This relates to *literary data* concerning the production and reception of literary works.

The targeted LiMES architecture provides a new type of scholarly digital work environment that is designed to preserve and, at the same time, to enhance our cultural heritage in the area of modern literature by transcribing into the digital not only the literary text and its (intermedial) context, but also the referential dynamics between them. LiMES will be developed as an information system by analogy to the newest architecture of information systems. It will focus on integrating, maintaining, analyzing, sharing and displaying *literarily referenced information*, in other words: it will offer, for the first time, a systematic (soundly theorized) and operational approach to the specific knowledge space opened up by modern literature. Thus, in an interdisciplinary effort LiMES will integrate two areas of research – *computer science* and *literary studies* – at a new, transbiblionomic level of complexity. This integration is novel and unique since it aims at a groundbreaking combination of literary studies, social software and machine learning. This special integration is explained in the following sections.

Exploring the Hidden Order of Poetic Gestures: A major focus of the planned project is to explore relations of textual and non-textual data. On the one hand, this relates to topic detection, labelling & tracking of recurrent topic units in literary oeuvres and related matrix texts. On the other hand, the project aims at reconstructing literary notions from the point of view of semantic web-technologies. That is, literary categories (as, e.g., poetic gestures) are reconstructed as building blocks (top mode) of a semantic web whose law-like networking constrains the linkage of the underlying media base (based on textual units, images and other media). This is the point where computer science is merged with the humanities: instead of applying classical machine learning, the project allows for controlled interventions by literary scholars,

media scientists and other researchers (*bottom mode*) at any stage of text mining. This integration requires *two additional innovations*: (1) a novel class of machine learning algorithms based on the common expertise of computer scientists and literary scholars and (2) a novel kind of *scientific desktop* which integrates the versatility of web-based information systems with the capabilities of *semantic information retrieval* to be applied in the humanities. Both innovations will be provided by LiMES. They are described in the following sections.

Exploring Literary Data on a Large Scale: Social Software in Service of Literary Studies: A characteristic of LiMES is its coverage of multimedia beyond traditionally text-based data. Here, LiMES will break new ground in the integration of data mining, social software and feedback based on the expertise of humanists. This will be done in order to close the gap between high-quality mining of textual and non-textual data. More specifically, a seamless integration of multimedia will be offered in order to contextualize literary studies on a much broader level than provided by state of the art approaches to computing in the humanities. For this task, novel methods of data mining will be developed which guarantee *high-level* semantic annotations by covering *large* corpora of *multimedia* data. As the complexity of this task goes beyond existing approaches to machine learning, the project plans to integrate *social software* with text and multimedia mining: while social software is used to enable thousands of scholars to collaboratively generate high-level semantic annotations of literary data, machine learning is used to guarantee their consistency. Computer science has shown that this sort of computer-based distributed cognition has a high potential in problem solving. Machine learning can take an invaluable benefit from this kind of *crowdsourcing* in order to solve tasks which have been seen to be hardly tractable so far. Following this approach, the project will develop social software for annotating literary data. That is, LiMES will integrate a *Social-Software-based Annotation Management System* for the storage, management, maintenance and retrieval of annotations of multimedia data and their relations. The output annotations will be used for automatically interrelating visual and textual media in order to close the gap between mining textual and non-textual data mentioned above. This integration of human and machine learning is highly innovative as it further develops social software subject to the demands of computer-based literary studies. As a result, LiMES-based social software will explore areas of intermediality which have been impassable to literary studies thus far. In summary, instead of trying to automatically explore intermedial relations (by a machine learning algorithm which is likely error-prone) the project performs *human computation* by implementing software for *crowdsourcing* semantic annotations. However, other than existing approaches to human computation the planned project will integrate the expertise of humanists who guarantee a higher level of semantic consistence of annotations of literary data.

Semi-Automatic Literary Studies: a Novel Paradigm of Machine Learning: Because of the complexity of literary data and the novelty of its computer-based exploration, standard methods in machine learning cannot be straightforwardly applied to it. Rather, exploring this data requires a close coupling of human and machine learning. Facing this challenge, the planned project offers significant progress in computing in the humanities by elaborating this coupling. It focuses on methods which seamlessly integrate the expertise of still unrelated areas, that is, literary interpretation (as being completely human-based so far) and text mining (as a computer-based method rarely based on human feedback). This interdisciplinary integration will be groundbreaking as it does not simply bridge two unrelated areas, but develops a new paradigm of machine learning. Its integral part is a bootstrapping mechanism which combines unsupervised machine learning with human-based text understanding in a crowdsourcing scenario.

The LiMES Desktop – Interfacing Scholars and Readers

The basic presentation interface of this project is the LiMES desktop and data management system. Its development comprises three major tasks: The software architecture for the LiMES desktop, the data management layer and the interface itself with its various possible views for different media, users, and approaches to the text.

The software architecture of LiMES will be based on a three-tier model (persistent storage – application logic – interface level). An open source content management framework that supports Web 2.0 technologies for interface modelling and which is based on standard database technology will be used so that project resources can focus on the innovative functions of the LiMES desktop (hyperlinking, different interface views). An additional layer is given by the integration of Web 2.0-based collaboration features (text / media integration and annotation). In terms of software architecture, this means adding a Web 2.0 mashup layer on top of the "traditional" three tier core architecture.

Data Management for the LiMES desktop (and for the project as a whole) starts from a "focus and context" viewpoint with a corpus of 'primary' (matrix) text, i.e. Kafka's work in this case, with all other data and media layers having relations (typed hyperlinks) with that primary layer. Technically, data management will be done using standard RDBMS technology. The additional data layers comprise secondary textual material (e.g. books or articles Kafka has read) as well as project, group and user (reader) specific texts and media. At the same time, results from text mining analyses and cooperatively generated knowledge structures have to be integrated. Library and information services (search in external resources) can be added generically as well as in a user or project specific way. Additionally, usage-oriented metadata has to be stored so that each user or group may easily load their specific view, media configuration, and (information) service portfolio into their version of the LiMES desktop. Project, group or reader specific

additions can be loaded dynamically into the users' workspace. Thus, each group, project or reader can create, use, and publish an individual cooperative workspace.

The complexity of hypertextual relations as well as of the different data and media layers integrated in the LiMES desktop calls for different presentation modes. Different text-centric views allow for presenting primary text along with interrelated (textual) material (cross references, secondary texts) in micro- and macroscopic views. Search interfaces provide search facilities in the different data layers as well as in external information and library services. "Analytic views" integrate results from text mining analyses (see above) and give users the possibility of triggering, controlling and visualising such analyses themselves. Finally, the integration of various additional media (photography, maps, scanned documents, digital movie material) calls for an additional presentation mode. A map-oriented interface giving annotation and media integration functions (comparable to those used in geo tagging platforms) allows for an innovative way of presenting literary texts in multi media contexts.

The innovative software architecture, multiple modes of presentation and interaction as well as dynamic service integration promise a working environment that goes far beyond the state of the art in computer-based philology. The coupling with recent research in text mining, ontology management and integration of social software approaches (tagging systems, Wikis) will result in a new paradigm of scholarly work and interaction with a focus on cooperation and flexible dynamic views on complex networks of literary text and media.

The Innovative Potential of LiMES

The rich innovative potential of the LiMES platform is based on an informational logic situated beyond the opposition between a mere enhancement of a biblionomic space by digital means (the rationale of computer philology) and a post-biblionomic space resulting in an empty culture of interrelated text snippets – as outlined in Maurer's dystopic study (Maurer e.a., 2007). In the *transbiblionomic space* opened up by LiMES, a digital workspace is used to present print text as an integral part of the intermedia network 'around 1900'. While in this transcription to the digital medium the biblionomic features of print literature are well preserved, its potential to operate beyond the framework of the book – a major stake of aesthetic experimentation in literary modernity, and, as argued above, a first priority challenge for its digital remediation – is opened up on the level of the object as well as of the research process. Departing from here we see four major trajectories for innovation:

1. *A new cultural technique of accessing, using, and enriching literary text*. In the LiMES environment, literary texts are no longer reduced to objects of interpretation, but can be used as tools to explore knowledge and reflect on it, i.e. as a medium, and not just as an object, of cultural memory. In turn, well acquainted habits of reading can be observed and reassessed from this new point of view, resulting in an

increased consciousness of the cultural uses of text in general. In his way, LiMES will enable users to make the best possible use from the fact that in increasing amounts of texts, and acts of text usage, are already available/taking place in the digital medium. LiMES therefore is the first adequate response to the challenge of the new medial ontology of text.

2. *A new informational logic of literary scholarship*: the LiMES desktop will enable literary scholars, for the first time in the history of the field, to explore empirically and manage computationally of the complex and open intertextual space of a literary text. On this basis, exchange between literary text and cultural context must no longer be either repressed (the poststructuralist option) or underscored (the cultural studies option), but can be reflected at the level of a theory of intermediality. At the same time, enabling the decentralized collaboration of scholars with various disciplinary, cultural, and career backgrounds, LiMES will provide a technical basis to overcome the corresponding limitations of perspective and encourage an ongoing comparison of and reflection on the usage of theories, methods, and expertise in the field of literary studies.

3. Due to its well reflected design and deployment of human intelligence and artificial intelligence in the service of literary studies, this project will lead to *significant progress in the architecture of information systems and in software development in the humanities.* Three important fields of research in computer science and computational linguistics – text mining, declarative ontology modelling, and cooperative social software will converge in a innovative working environment that will have significant impact on future literary, media, and cultural studies. The networked interactive media space of LiMES shall serve as a model for numerous projects of similar complexity.

4. *Kafkabureau.net* will not only serve as a paradigmatic challenge for the development of the LiMES technology, but will, by the end of the five year project period, constitute *a groundbreaking achievement of its own right*. The first web portal of its kind world wide, kafkabureau.net will present the work of an author of German language and of global rank and outreach within its specifically Central European cultural context. As this context will be structured and made accessible on the basis of the poetological concept of intertextuality (and the core concept of poetic gestures), poetology, media theory, cultural studies, and digital research technology will be integrated to open a whole new view on Kafka's work: In an area of crucial significance for the political and cultural history of Europe, and in an era of the emergence of a new medial foundation of society, politics, and culture, we will learn to observe, and to use, this unique achievement in modern writing within the framework of an information theory of social and cultural diversity.

Literature

Antoniou, G. and F. v. Harmelen (2008). A Semantic Web Primer (2 ed.). Cambridge: The MIT Press.

Auer, S., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives (2007). DBpedia: a nucleus for a web of open data. In Proceedings of the 6th International and the 2nd Asian Semantic Web Conference, ISWC'07/ASWC'07, pp. 722–735. Berlin, Heidelberg: Springer.

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, Briefzeugnisse, 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.

Borges, J. L. (1995), Kafka and his Precursors, in: S. Burke, ed., Authorship. From Plato to the Postmodern, 335-337.

Bunke, H. and M. Neuhaus (2007). Graph matching: exact and error-tolerant methods and the automatic learning of edit costs. In D. J. Cook and L. B. Holder (Eds.), Mining Graph Data, pp. 17–34. Hoboken: Wiley.

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.

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

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.

Ebeling, K., Günzel, S., Archivologie. Theorien des Archivs in Wissenschaft, Medien und Künsten, Berlin: Kadmos 2009.

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

Emmert-Streib, F. and M. Dehmer (2010). Identifying critical financial networks of the djia: Toward(s) a network based index. Complexity 16, 24–33.

Engel, M., D. Lamping (2006, eds.), Franz Kafka und die Weltliteratur, Göttingen: Vandenhoeck & Ruprecht.

Feldman, R. and J. Sanger (2007). The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. Cambridge: Cambridge University Press.

Foucault, M. (2001), Nachwort (zu Gustave Flaubert, Die Versuchung des Heiligen Antonius), in: M.F., Dits et Ecrits. Schriften in vier Bänden, vol. 1, 1954-1969, Frankfurt: Suhrkamp, 397-443.

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.

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

Jussen, B., A. Mehler, and A. Ernst (2007). A corpus management system for historical semantics. Sprache und Datenverarbeitung. International Journal for Language Data Processing 31(1-2), 81–89.

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

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

Mehler, A. (2005b). Zur textlinguistischen Fundierung der Text- und Korpuskonversion. Sprache und Datenverarbeitung. International Journal for Language Data Processing 1, 29–53.

Mehler, A. (2008). Structural similarities of complex networks: A computational model by example of wiki graphs. Applied Artificial Intelligence 22(7&8), 619–683.

Mehler, A. (2009). Generalized shortest paths trees: A novel graph class applied to semiotic networks. In M. Dehmer and F. Emmert-Streib (Eds.), Analysis of Complex Networks: From Biology to Linguistics, pp. 175–220. Weinheim: Wiley-VCH.

Mehler, A. (2011). A quantitative graph model of social ontologies by example of Wikipedia. In M. Dehmer, F. Emmert-Streib, and A. Mehler (Eds.), Towards an Information Theory of Complex Networks: Statistical Methods and Applications. Boston/Basel: Birkhäuser.

Mehler, A., N. Diewald, U. Waltinger, R. Gleim, D. Esch, B. Job, T. Küchelmann, O. Pustylnikov, and P. Blanchard (2011). Evolution of Romance language in written communication: Network analysis of late Latin and early Romance corpora. Leonardo 44(3).

Mehler, A., A. Lücking, and P. Menke (2011). From neural activation to symbolic alignment: A network-based approach to the formation of dialogue lexica. In Proceedings of the International Joint Conference on Neural Networks (IJCNN 2011), San Jose, California, July 31 – August 5.

Mehler, A., A. Lücking, and P. Weiß (2010). A network model of interpersonal alignment. Entropy 12(6), 1440–1483.

Mehler, A., S. Schwandt, R. Gleim, and B. Jussen (2011). Der eHumanities Desktop als Werkzeug in der historischen Semantik: Funktionsspektrum und Einsatzszenarien. Journal for Language Technology and Computational Linguistics (JLCL). Appears.

Mehler, A., S. Sharoff, and M. Santini (Eds.) (2010). Genres on the Web: Computational Models and Empirical Studies. Dordrecht: Springer.

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

Nagel, B. (1983), Kafka und die Weltliteratur, München: Winkler.

Mielke, B., Wolff, C. (2011), Virtuelle Forschungsumgebungen im Bereich der Rechtswissenschaft: Ein Anforde-rungsprofil. Jusletter IT 24 [Beitrag zum internationalen Rechtsinformatik Symposium IRIS 2011, Salzburg].

Nekula, M. (2003), *Franz Kafkas Sprachen*, Tübingen: Niemeyer.

Nekula, Marek/Fleischmann, Ingrid/Greule, Albrecht (Hgg.) (2007): *Franz Kafka im sprachnationalen Kontext seiner Zeit. Sprache und nationale Identität in öffentlichen Institutionen der böhmischen Länder.* Weimar, Köln: Böhlau.

Neumann, G. (1996), Lektüre und Lebenswelt, in: G.N, ed., *Canetti als Leser*, Freiburg: Rombach, 7-32.

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

Rehm, G., O. Schonefeld, A. Witt, E. Hinrichs, and M. Reis (2009). Sustainability of annotated resources in linguistics: A web-platform for exploring, querying and distributing. Literary and Linguistic Computing 24(2), 193–210.

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. (2006), 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. (2011), 'Lightning no longer flashes'. Kafka's Chinese Voice and the Thunder of the Great War, in: Jakob Lothe, Beatrice Sandberg, Ronald Speirs, Hg., Franz Kafka: Narrative, History, Genre, Ohio State University Press, Columbus 2011, 58-80.

Wagner, B. (2013), Fürsprache – Widerstreit – Dialog. Karl Kraus, Franz Kafka und das Schreiben gegen den Krieg, in: Manfred Engel, Ritchie Robertson (Hg.), Kafka, Prague, and the First World War, Würzburg: Könighausen&Neumann 2013 (in Vorb.).

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.

Wagner, B., Mehler, A., et al. (2009), Bausteine eines Literary Memory Information System (LiMeS) am Beispiel der Kafka-Forschung", in: Wolfgang Hoeppner, Hg., GSCL-Symposium Sprachtechnologie und eHumanities. Technischer Bericht, Fakultät für Ingenieurwissenschaften, Universität Duisburg-Essen, , 119-134.

Waltinger, U. and A. Mehler (2008). Who is it? Context sensitive named entity and instance recognition by means of Wikipedia. In Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI-2008), pp. 381–384. IEEE Computer Society.

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 Evelinde Hutzler, Albert Schröder & Gabriele Schweikl (eds.), Strategien zum Aufbau digitaler Bibliotheken 157-172. Göttingen: Universitätsverlag Göttingen.