



digital humanities 2012

Conference Abstracts

University of Hamburg, July 16–22

Hamburg University Press

The Alliance of Digital Humanities Organizations
The Association for Literary and Linguistic Computing
The Association for Computers and the Humanities
The Australasian Association for Digital Humanities
centerNet

The Society for Digital Humanities – Société pour l'étude des médias interactifs

Digital Humanities 2012

Conference Abstracts

University of Hamburg, Germany
July 16–22, 2012



Hamburg University Press
Publishing House of the Hamburg State and University Library
Carl von Ossietzky

Printed with the Support of the German Research Foundation

Editor

Jan Christoph Meister

Editorial Assistant

Katrin Schönert

Editorial Staff

Bastian Lomsché

Wilhelm Schernus

Lena Schüch

Meike Stegkemper

Technical Support

Benjamin W. Bohl

Daniel Röwenstrunk

Bibliographic information published by the *Deutsche Nationalbibliothek* (German National Library). The *Deutsche Nationalbibliothek* lists this publication in the *Deutsche Nationalbibliografie*; detailed bibliographic data are available on the Internet at <https://portal.dnb.de/>

The *Deutsche Nationalbibliothek* stores this online publication on its Archive Server. The Archive Server is part of the deposit system for long-term availability of digital publications.

Also available open access on the Internet at:

Hamburg University Press – <http://hup.sub.uni-hamburg.de>

PURL: http://hup.sub.uni-hamburg.de/HamburgUP/DH2012_Book_of_Abstracts

For an elaborated version of this Book of Abstracts with color photos and zoom function, please visit: www.dh2012.uni-hamburg.de

ISBN 978-3-937816-99-9 (printed version)

© 2012 Hamburg University Press, publishing house of the Hamburg State and University Library
Carl von Ossietzky, Germany

Cover design: Turan Usuk

Cover illustration: Dagmar Schwelle/laif

Printing house: Elbepartner, BuK! Breitschuh & Kock GmbH, Hamburg, Germany

The 24th Joint International Conference of
the Association for Literary and Linguistic Computing and
the Association for Computers and the Humanities

and

The 5th Joint International Conference of
the Association for Literary and Linguistic Computing,
the Association for Computers and the Humanities,
the Society for Digital Humanities – Société pour l'étude des médias interactifs,
for the first time organized together with
the Australasian Association for Digital Humanities and
centerNet

International Programme Committee

- Susan Brown (SDH/SEMI – Vice Chair)
- Arianna Ciula (ALLC)
- Tanya Clement (ACH)
- Michael Eberle-Sinatra (SDH/SEMI)
- Dot Porter (ACH)
- Jan Rybicki (ALLC)
- Jon Saklofske (SDH/SEMI)
- Paul Spence (ALLC – Chair)
- Tomoji Tabata (ALLC)
- Katherine Walter (ACH)

Local Organizing Committee

- Peer Bolten
- Imke Borchers
- Evelyn Gius
- Mareike Höckendorff
- Bastian Lomsché
- Jan Christoph Meister
- Marco Petris
- Wilhelm Schernus
- Lena Schüch
- Katrin Schönert
- Meike Stegkemper

Table of Contents

List of Reviewers	1
Plenary Sessions	
Dynamics and Diversity: Exploring European and Transnational Perspectives on Digital Humanities Research Infrastructures <i>Moulin, Claudine</i>	9
Embracing a Distant View of the Digital Humanities <i>Shimoda, Masahiro</i>	11
Pre-conference Workshops	
Digital Methods in Manuscript Studies <i>Brockmann, Christian; Wangchuk, Dorji</i>	15
Introduction to Stylomatic Analysis using R <i>Eder, Maciej; Rybicki, Jan</i>	16
NeDiMAH workshop on ontology based annotation <i>Eide, Øyvind; Ore, Christian-Emil; Rahtz, Sebastian</i>	18
Service-oriented Architectures (SOAs) for the Humanities: Solutions and Impacts <i>Hinrichs, Erhard; Neuroth, Heike; Wittenburg, Peter</i>	20
Here and There, Then and Now – Modelling Space and Time in the Humanities <i>Isaksen, Leif; Day, Shawn; Andresen, Jens; Hyvönen, Eero; Mäkelä, Eetu</i>	22
Crowdsourcing meaning: a hands-on introduction to CLÉA, the Collaborative Literature Exploration and Annotation Environment <i>Petris, Marco; Gius, Evelyn; Schüch, Lena; Meister, Jan Christoph</i>	24
Learning to play like a programmer: web mash-ups and scripting for beginners <i>Ridge, Mia</i>	25
Introduction to Distant Reading Techniques with Voyant Tools, Multilingual Edition <i>Sinclair, Stéfan; Rockwell, Geoffrey</i>	26
Towards a reference curriculum for the Digital Humanities <i>Thaller, Manfred</i>	27
Free your metadata: a practical approach towards metadata cleaning and vocabulary reconciliation <i>van Hooland, Seth; Verborgh, Ruben; De Wilde, Max</i>	28
Panels	
Text Analysis Meets Text Encoding <i>Bauman, Syd; Hoover, David; van Dalen-Oskam, Karina; Piez, Wendell</i>	33
Designing Interactive Reading Environments for the Online Scholarly Edition <i>Blandford, Ann; Brown, Susan; Dobson, Teresa; Faisal, Sarah; Fiorentino, Carlos; Frizzera, Luciano; Giacometti, Alejandro; Heller, Brooke; Ilovan, Mihaela; Michura, Piotr; Nelson, Brent; Radzikowska, Milena; Rockwell, Geoffrey; Ruecker, Stan; Sinclair, Stéfan; Sondheim, Daniel; Warwick, Claire; Windsor, Jennifer</i>	35

Developing the spatial humanities: Geo-spatial technologies as a platform for cross-disciplinary scholarship <i>Bodenhamer, David; Gregory, Ian; Ell, Paul; Hallam, Julia; Harris, Trevor; Schwartz, Robert</i>	41
Prosopographical Databases, Text-Mining, GIS and System Interoperability for Chinese History and Literature <i>Bol, Peter Kees; Hsiang, Jieh; Fong, Grace</i>	43
Future Developments for TEI ODD <i>Cummings, James; Rahtz, Sebastian; Burnard, Lou; Bauman, Syd; Gaiffe, Bertrand; Romary, Laurent; Bański, Piotr</i>	52
Compiling large historical reference corpora of German: Quality Assurance, Interoperability and Collaboration in the Process of Publication of Digitized Historical Prints <i>Geyken, Alexander; Gloning, Thomas; Stäcker, Thomas</i>	54
Computational models of narrative structure <i>Löwe, Benedikt; Físseni, Bernhard; León, Carlos; Bod, Rens</i>	57
Approaches to the Treatment of Primary Materials in Digital Lexicons: Examples of the New Generation of Digital Lexicons for Buddhist Studies <i>Nagasaki, Kiyonori; Tomabechi, Toru; Wangchuk, Dorji; Takahashi, Koichi; Wallman, Jeff; Muller, A. Charles</i>	61
Topic Modeling the Past <i>Nelson, Robert K.; Mimno, David; Brown, Travis</i>	64
Facilitating Research through Social-Document Networks <i>Pitti, Daniel; Simon, Agnès; Vitali, Stefano; Arnold, Kerstin</i>	70
Digital Humanities as a university degree: The status quo and beyond <i>Thaller, Manfred; Sahle, Patrick; Clavaud, Florence; Clement, Tanya; Fiormonte, Domenico; Pierazzo, Elena; Rehbein, Malte; Rockwell, Geoffrey; Schreibman, Susan; Sinclair, Stéfan</i>	72
Papers	
Exploring Originality in User-Generated Content with Network and Image Analysis Tools <i>Akdag Salah, Alkim Almila; Salah, Albert Ali; Douglass, Jeremy; Manovich, Lev</i>	79
Patchworks and Field-Boundaries: Visualizing the History of English <i>Alexander, Marc</i>	82
Developing Transcultural Competence in the Study of World Literatures: Golden Age Literature Glossary Online (GALGO) <i>Alonso Garcia, Nuria; Caplan, Alison</i>	84
Trees of Texts – Models and methods for an updated theory of medieval text stemmatology <i>Andrews, Tara Lee; Macé, Caroline</i>	85
Mapping the Information Science Domain <i>Arazy, Ofer; Ruecker, Stan; Rodriguez, Omar; Giacometti, Alejandro; Zhang, Lu; Chun, Su</i>	88

Words made Image. Towards a Language-Based Segmentation of Digitized Art Collections	
<i>Armasehu, Florentina</i>	91
HisDoc: Historical Document Analysis, Recognition, and Retrieval	
<i>Baechler, Micheal; Fischer, Andreas; Naji, Nada; Ingold, Rolf; Bunke, Horst; Savoy, Jacques</i>	94
Research infrastructures for Digital Humanities: The local perspective	
<i>Bärenfänger, Maja; Binder, Frank</i>	97
Pelagios: An Information Superhighway for the Ancient World	
<i>Barker, Elton; Simon, Rainer; Isaksen, Leif</i>	99
Putting TEI Tite to use – generating a database resource from a printed dictionary or reference type publication	
<i>Barner-Rasmussen, Michael</i>	102
Digital Humanities in the Classroom: Introducing a New Editing Platform for Source Documents in Classics	
<i>Beaulieu, Marie-Claire; Almas, Bridget</i>	105
DiaView: Visualise Cultural Change in Diachronic Corpora	
<i>Beavan, David</i>	107
Catch + Release: Research and Creation of a Digital New Media Exhibition in the Context of a Cultural and Heritage Museum	
<i>Beer, Ruth</i>	109
Opportunity and accountability in the ‘eResearch push’	
<i>Bellamy, Craig</i>	111
Connecting European Women Writers. The Selma Lagerlöf Archive and Women Writers Database	
<i>Bergenmar, Jenny; Olsson, Leif-Jöran</i>	113
Stylometric Analysis of Chinese Buddhist texts: Do different Chinese translations of the ‘Gandhavyūha’ reflect stylistic features that are typical for their age?	
<i>Bingenheimer, Marcus; Hung, Jen-Jou; Hsieh, Cheng-en</i>	115
Information Extraction on Noisy Texts for Historical Research	
<i>Blanke, Tobias; Bryant, Michael; Speck, Reto; Kristel, Conny</i>	117
Modeling Gender: The ‘Rise and Rise’ of the Australian Woman Novelist	
<i>Bode, Katherine</i>	119
Contextual factors in literary quality judgments: A quantitative analysis of an online writing community	
<i>Boot, Peter</i>	121
Violence and the Digital Humanities Text as Pharmakon	
<i>Bradley, Adam James</i>	123
Towards a bibliographic model of illustrations in the early modern illustrated book	
<i>Bradley, John; Pigney, Stephen</i>	124
Automatic Mining of Valence Compounds for German: A Corpus-Based Approach	
<i>Brock, Anne; Henrich, Verena; Hinrichs, Erhard; Versley, Yannick</i>	126

Networks of networks: a critical review of formal network methods in archaeology through citation network analysis and close reading <i>Brughmans, Tom</i>	129
On the dual nature of written texts and its implications for the encoding of genetic manuscripts <i>Brüning, Gerrit; Henzel, Katrin; Pravida, Dietmar</i>	131
Automatic recognition of speech, thought and writing representation in German narrative texts <i>Brunner, Annelen</i>	135
Bringing Modern Spell Checking Approaches to Ancient Texts – Automated Suggestions for Incomplete Words <i>Büchler, Marco; Kruse, Sebastian; Eckart, Thomas</i>	137
Designing a national ‘Virtual Laboratory’ for the humanities: the Australian HuNI project <i>Burrows, Toby Nicolas</i>	139
Beyond Embedded Markup <i>Buzzetti, Dino; Thaller, Manfred</i>	142
Myopia: A Visualization Tool in Support of Close Reading <i>Chaturvedi, Manish; Gannod, Gerald; Mandell, Laura; Armstrong, Helen; Hodgson, Eric</i>	148
Translation Arrays: Exploring Cultural Heritage Texts Across Languages <i>Cheesman, Tom; Thiel, Stephan; Flanagan, Kevin; Zhao, Geng; Ehrmann, Alison; Laramee, Robert S.; Hope, Jonathan; Berry, David M.</i>	151
Constructing a Chinese as Second Language Learner Corpus for Language Learning and Research <i>Chen, Howard</i>	154
Social Curation of large multimedia collections on the cloud <i>Chong, Dazhi; Coppage, Samuel; Gu, Xiangyi; Maly, Kurt; Wu, Harris; Zubair, Mohammad</i>	155
Sounding for Meaning: Analyzing Aural Patterns Across Large Digital Collections <i>Clement, Tanya; Auwil, Loretta; Tchong, David; Capitanu, Boris; Monroe, Megan; Goel, Ankita</i>	158
The Programming Historian 2: A Participatory Textbook <i>Crymble, Adam H.; MacEachern, Alan; Turkel, William J.</i>	162
Multilingual and Semantic Extension of Folk Tale Catalogues <i>Declerck, Thierry; Lendvai, Piroska; Darányi, Sándor</i>	163
Digital Language Archives and Less-Networked Speaker Communities <i>Dobrin, Lise M.</i>	167
Language Documentation and Digital Humanities: The (DoBeS) Language Archive <i>Drude, Sebastian; Trilsbeek, Paul; Broeder, Daan</i>	169
The potential of using crowd-sourced data to re-explore the demography of Victorian Britain <i>Duke-Williams, Oliver William</i>	173

Sharing Ancient Wisdoms: developing structures for tracking cultural dynamics by linking moral and philosophical anthologies with their source and recipient texts <i>Dunn, Stuart; Hedges, Mark; Jordanous, Anna; Lawrence, Faith; Roueche, Charlotte; Tupman, Charlotte; Wakelnig, Elvira</i>	176
Recovering the Recovered Text: Diversity, Canon Building, and Digital Studies <i>Earhart, Amy</i>	179
Mind your corpus: systematic errors in authorship attribution <i>Eder, Maciej</i>	181
Underspecified, Ambiguous or Formal. Problems in Creating Maps Based on Texts <i>Eide, Øyvind</i>	185
A Frequency Dictionary of Modern Written and Oral Media Arabic <i>Elmaz, Orhan</i>	188
Texts in Motion – Rethinking Reader Annotations in Online Literary Texts <i>Fendt, Kurt E.; Kelley, Wyn; Zhang, Jia; Della Costa, Dave</i>	190
May Humanists Learn from Artists a New Way to Interact with Digital Technology? <i>Franchi, Stefano</i>	192
A flexible model for the collaborative annotation of digitized literary works <i>Gayoso-Cabada, Joaquín; Ruiz, Cesar; Pablo-Nuñez, Luis; Sarasa-Cabezuelo, Antonio; Goicoechea-de-Jorge, Maria; Sanz-Cabrerizo, Amelia; Sierra-Rodriguez, Jose-Luis</i>	195
HyperMachiavel: a translation comparison tool <i>Gedzelman, Séverine; Zancarini, Jean-Claude</i>	198
Discrimination sémantique par la traduction automatique, expériences sur le dictionnaire français de Littré <i>Glorieux, Frédéric; Jolivet, Vincent</i>	202
The Myth of the New: Mass Digitization, Distant Reading and the Future of the Book <i>Gooding, Paul Matthew; Warwick, Claire; Terras, Melissa</i>	204
Designing Navigation Tools for an Environmental Humanities Portal: Considerations and Critical Assessments <i>Graf von Hardenberg, Wilko; Coulter, Kimberly</i>	206
Processing Email Archives in Special Collections <i>Hangal, Sudheendra; Chan, Peter; Lam, Monica S.; Heer, Jeffrey</i>	208
The Stylometry of Collaborative Translation <i>Heydel, Magda; Rybicki, Jan</i>	212
Focus on Users in the Open Development of the National Digital Library of Finland <i>Hirvonen, Ville; Kautonen, Heli Johanna</i>	215
The Rarer They Are, the More There Are, the Less They Matter <i>Hoover, David</i>	218

Experiments in Digital Philosophy – Putting new paradigms to the test in the Agora project <i>Hrachovec, Herbert; Carusi, Annamaria; Huentelmann, Raphael; Pichler, Alois; Antonio, Lamarra; Cristina, Marras; Alessio, Piccioli; Lou, Burnard</i>	221
Information Discovery in the Chinese Recorder Index <i>Hsiang, Jieh; Kong, Jung-Wei; Sung, Allan</i>	224
Complex Network Perspective on Graphic Form System of Hanzi <i>Hu, Jiajia; Wang, Ning</i>	228
A Computer-Based Approach for Predicting the Translation Time Period of Early Chinese Buddhism Translation <i>Hung, Jen-Jou; Bingenheimer, Marcus; Kwok, Jieli</i>	230
Bridging Multicultural Communities: Developing a Framework for a European Network of Museum, Libraries and Public Cultural Institutions <i>Innocenti, Perla; Richards, John; Wieber, Sabine</i>	232
Ptolemy’s Geography and the Birth of GIS <i>Isaksen, Leif</i>	236
Tracing the history of Noh texts by mathematical methods. Validating the application of phylogenetic methods to Noh texts <i>Iwata, Yoshimi</i>	239
Computing and Visualizing the 19th-Century Literary Genome <i>Jockers, Matthew</i>	242
Using the Google Ngram Corpus to Measure Cultural Complexity <i>Juola, Patrick</i>	245
‘All Rights Worth Recombination’: Post-Hacker Culture and ASCII Literature (1983-1993) <i>Katelnikoff, Joel</i>	247
Evaluating Unmasking for Cross-Genre Authorship Verification <i>Kestemont, Mike; Luyckx, Kim; Daelemans, Walter; Crombez, Thomas</i>	249
Literary Wikis: Crowd-sourcing the Analysis and Annotation of Pynchon, Eco and Others <i>Ketzan, Erik</i>	252
Social Network Analysis and Visualization in ‘The Papers of Thomas Jefferson’ <i>Klein, Lauren Frederica</i>	254
VariaLog: how to locate words in a French Renaissance Virtual Library <i>Lay, Marie H�el�ene</i>	256
DeRiK: A German Reference Corpus of Computer-Mediated Communication <i>Lemnitzer, Lothar; Beißwenger, Michael; Ermakova, Maria; Geyken, Alexander; Storrer, Angelika</i>	259
Estimating the Distinctiveness of Graphemes and Allographs in Palaeographic Classification <i>Levy, Noga; Wolf, Lior; Dershowitz, Nachum; Stokes, Peter</i>	264

Academic Research in the Blogosphere: Adapting to New Opportunities and Risks on the Internet <i>Littauer, Richard; Winters, James; Roberts, Sean; Little, Hannah; Pleyer, Michael; Benzon, Bill</i>	268
Feeling the View: Reading Affective Orientation of Tagged Images <i>Liu, Jyi-Shane; Peng, Sheng-Yang</i>	270
Characterizing Authorship Style Using Linguistic Features <i>Lucic, Ana; Blake, Catherine</i>	273
Investigating the genealogical relatedness of the endangered Dagon languages <i>Moran, Steven; Prokic, Jelena</i>	276
Landscapes, languages and data structures: Issues in building the Placenames Database of Ireland <i>Měchura, Michal Boleslav</i>	278
Interoperability of Language Documentation Tools and Materials for Local Communities <i>Nakhimovsky, Alexander; Good, Jeff; Myers, Tom</i>	280
Content Creation by Domain Experts in a Semantic GIS System <i>Nakhimovsky, Alexander; Myers, Tom</i>	283
From Preserving Language Resources to Serving Language Speakers: New Prospects for Endangered Languages Archives <i>Nathan, David John</i>	286
Retrieving Writing Patterns From Historical Manuscripts Using Local Descriptors <i>Neumann, Bernd; Herzog, Rainer; Solth, Arved; Bestmann, Oliver; Scheel, Julian</i>	288
Distractorless Authorship Verification <i>Noecker Jr., John; Ryan, Michael</i>	292
Cataloguing linguistic diversity: Glottolog/Langdoc <i>Nordhoff, Sebastian; Hammarström, Harald</i>	296
Geo-Temporal Interpretation of Archival Collections Using Neatline <i>Nowviskie, Bethany; Graham, Wayne; McClure, David; Boggs, Jeremy; Rochester, Eric</i>	299
Enriching Digital Libraries Contents with SemLib Semantic Annotation System <i>Nucci, Michele; Grassi, Marco; Morbidoni, Christian; Piazza, Francesco</i>	303
The VL3: A Project at the Crossroads between Linguistics and Computer Science <i>Nuñez, Camelia Gianina; Mavillard, Antonio Jiménez</i>	306
‘Eric, you do not humble well’: The Image of the Modern Vampire in Text and on Screen <i>Opas-Hänninen, Lisa Lena; Hettel, Jacqueline; Toljamo, Tuomo; Seppänen, Tapio</i>	308
Electronic Deconstruction of an argument using corpus linguistic analysis of its on-line discussion forum supplement <i>O’Halloran, Kieran Anthony</i>	310
Citygram One: Visualizing Urban Acoustic Ecology <i>Park, Tae Hong; Miller, Ben; Shrestha, Ayush; Lee, Sangmi; Turner, Jonathan; Marse, Alex</i>	313

Towards Wittgenstein on the Semantic Web <i>Pichler, Alois; Zöllner-Weber, Amélie</i>	318
Uncovering lost histories through GeoStoryteller: A digital GeoHumanities project <i>Rabina, Debbie L.; Cocciolo, Anthony</i>	322
Workflows as Structured Surfaces <i>Radzikowska, Milena; Ruecker, Stan; Rockwell, Geoffrey; Brown, Susan; Frizzera, Luciano; INKE Research Group</i>	324
Code-Generation Techniques for XML Collections Interoperability <i>Ramsay, Stephen; Pytlik-Zillig, Brian</i>	327
Uncertain Date, Uncertain Place: Interpreting the History of Jewish Communities in the Byzantine Empire using GIS <i>Rees, Gethin Powell</i>	329
Code sprints and Infrastructure <i>Reside, Doug; Fraistat, Neil; Vershbow, Ben; van Zundert, Joris Job</i>	331
Digital Genetic Criticism of RENT <i>Reside, Doug</i>	333
On the Internet, nobody knows you're a historian: exploring resistance to crowdsourced resources among historians <i>Ridge, Mia</i>	335
Formal Semantic Modeling for Human and Machine-based Decoding of Medieval Manuscripts <i>Ritsema van Eck, Marianne Petra; Schomaker, Lambert</i>	336
The Swallow Flies Swiftly Through: An Analysis of Humanist <i>Rockwell, Geoffrey; Sinclair, Stéfan</i>	339
The Digital Mellini Project: Exploring New Tools & Methods for Art-historical Research & Publication <i>Rodríguez, Nuria; Baca, Murtha; Albrezzi, Francesca; Longaker, Rachel</i>	342
Intertextuality and Influence in the Age of Enlightenment: Sequence Alignment Applications for Humanities Research <i>Roe, Glenn H.; The ARTFL Project</i>	345
Engaging the Museum Space: Mobilising Visitor Engagement with Digital Content Creation <i>Ross, Claire Stephanie; Gray, Steven; Warwick, Claire; Hudson Smith, Andrew; Terras, Melissa</i>	348
Aiding the Interpretation of Ancient Documents <i>Roued-Cunliffe, Henriette</i>	351
The Twelve Disputed 'Federalist' Papers: A Case for Collaboration <i>Rudman, Joseph</i>	353
Writing with Sound: Composing Multimodal, Long-Form Scholarship <i>Sayers, Jentery</i>	357
Intra-linking the Research Corpus: Using Semantic MediaWiki as a lightweight Virtual Research Environment <i>Schindler, Christoph; Ell, Basil; Rittberger, Marc</i>	359

Corpus Coranicum: A digital landscape for the study of the Qu'ran <i>Schnöpf, Markus</i>	362
The MayaArch3D Project: A 3D GIS Web System for Querying Ancient Architecture and Landscapes <i>Schwerin, Jennifer von; Richards-Rissetto, Heather; Agugiaro, Giorgio; Remondino, Fabio; Girardi, Gabrio</i>	365
Multi-dimensional audio-visual technology: Evidence from the endangered language documentation <i>Sharma, Narayan P.</i>	368
Contours of the Past: Computationally Exploring Civil Rights Histories <i>Shaw, Ryan Benjamin</i>	370
Notes from the Collaboratory: An Informal Study of an Academic DH Lab in Transition <i>Siemens, Lynne; Siemens, Raymond</i>	373
XML-Print: an Ergonomic Typesetting System for Complex Text Structures <i>Sievers, Martin; Burch, Thomas; Küster, Marc W.; Moulin, Claudine; Rapp, Andrea; Schwarz, Roland; Gan, Yu</i>	375
Federated Digital Archives and Disaster Recovery: The Role of the Digital Humanities in Post-earthquake Christchurch <i>Smithies, James Dakin</i>	380
Modeling Medieval Handwriting: A New Approach to Digital Palaeography <i>Stokes, Peter</i>	382
A Digital Geography of Hispanic Baroque Art <i>Suárez, Juan-Luis; Sancho-Caparrini, Fernando</i>	385
Approaching Dickens' Style through Random Forests <i>Tabata, Tomoji</i>	388
Interfacing Diachrony: Visualizing Linguistic Change on the Basis of Digital Editions of Serbian 18th-Century Texts <i>Tasovac, Toma; Ermolaev, Natalia</i>	392
Promise and Practice of Enhanced Publications to Complement Conventionally- Published Scholarly Monographs <i>Tatum, Clifford; Jankowski, Nicholas; Scharnhorst, Andrea</i>	394
Culpeper's legacy: How title pages sold books in the 17th century <i>Tyrkkö, Jukka Jyrki Juhani; Suhr, Carla Maria; Marttila, Ville</i>	396
The Differentiation of Genres in Eighteenth- and Nineteenth-Century English Literature <i>Underwood, Ted; Sellers, Jordan; Auwil, Loretta; Capitanu, Boris</i>	397
Digital editions with eLaborate: from practice to theory <i>van Dalen-Oskam, Karina; van Zundert, Joris Job</i>	400
Delta in 3D: Copyists Distinction by Scaling Burrows's Delta <i>van Zundert, Joris Job; van Dalen-Oskam, Karina</i>	402
Wiki Technologies for Semantic Publication of Old Russian Charters <i>Varfolomeyev, Aleksey; Ivanovs, Aleksandrs</i>	405

L'histoire de l'art à l'ère numérique – Pour une historiographie médiologique <i>Welger-Barboza, Corinne</i>	407
Benefits of tools and applications for a digitized analysis of Chinese Buddhist inscriptions <i>Wenzel, Claudia</i>	411
The ARTeFACT Movement Thesaurus: toward an open-source tool to mine movement-derived data <i>Wiesner, Susan L.; Bennett, Bradford; Stalnaker, Rommie L.</i>	413
The electronic 'Oxford English Dictionary', poetry, and intertextuality <i>Williams, David-Antoine</i>	415
Reasoning about Genesis or The Mechanical Philologist <i>Wissenbach, Moritz; Pravida, Dietmar; Middell, Gregor</i>	418
The Digital Daozang Jiyao – How to get the edition into the Scholar's labs <i>Wittern, Christian</i>	422
Posters	
A Digital Approach to Sound Symbolism in English: Evidence from the Historical Thesaurus <i>Alexander, Marc; Kay, Christian</i>	427
Collaborative Video and Image Annotation <i>Arnold, Matthias; Knab, Cornelia; Decker, Eric</i>	429
Le Système modulaire de gestion de l'information historique (SyMoGIH): une plateforme collaborative et cumulative de stockage et d'exploitation de l'information géo-historique <i>Beretta, Francesco; Vernus, Pierre; Hours, Bernard</i>	431
Realigning Digital Humanities Training: The Praxis Program at the Scholars' Lab <i>Boggs, Jeremy; Nowvieskie, Bethany; Gil, Alexander; Johnson, Eric; Lestock, Brooke; Storti, Sarah; Swafford, Joanna; Praxis Program Collaborators</i>	433
Supporting the emerging community of MEI: the current landscape of tools for note entry and digital editing <i>Bohl, Benjamin W.; Röwenstrunk, Daniel; Viglianti, Raffaele</i>	435
'The Past Is Never Dead. It's Not Even Past': The Challenge of Data Provenance in the e-Humanities <i>Clark, Ashley M.; Holloway, Steven W.</i>	438
The Social Edition: Scholarly Editing Across Communities <i>Crompton, Constance; Siemens, Raymond; The Devonshire MS Editorial Group</i>	441
Courting 'The World's Wife': Original Digital Humanities Research in the Undergraduate Classroom <i>Croxall, Brian</i>	443
The Academy's Digital Store of Knowledge <i>Czmiel, Alexander; Jürgens, Marco</i>	445
Building a TEI Archiving, Publishing, and Access Service: The TAPAS Project <i>Flanders, Julia; Hamlin, Scott; Alvarado, Rafael; Mylonas, Elli</i>	448
Author Consolidation across European National Bibliographies <i>Freire, Nuno</i>	450

Historical Events Versus Information Contents – A Preliminary Analysis of the National Geographic Magazine <i>Fujimoto, Yu</i>	453
‘Tejiendo la Red HD’ – A case study of building a DH network in Mexico <i>Galina, Isabel; Priani, Ernesto; López, José; Rivera, Eduardo; Cruz, Alejandro</i>	456
Adaptive Automatic Gesture Stroke Detection <i>Gebre, Binyam Gebrekidan; Wittenburg, Peter</i>	458
Towards a Transnational Multilingual Caribbean Digital Humanities Lab <i>Gil, Alexander</i>	462
NUScholar: Digital Methods for Educating New Humanities Scholars <i>Graff, Ann-Barbara; Lucas, Kristin; Blustein, James; Gibson, Robin; Woods, Sharon</i>	463
Latent Semantic Analysis Tools Available for All Digital Humanities Projects in Project Bamboo <i>Hooper, Wallace Edd; Cowan, Will; Jiao, David; Walsh, John A.</i>	465
Machine Learning for Automatic Annotation of References in DH scholarly papers <i>Kim, Young-Min; Bellot, Patrice; Faath, Elodie; Dacos, Marin</i>	467
An Ontology-Based Iterative Text Processing Strategy for Detecting and Recognizing Characters in Folktales <i>Koleva, Nikolina; Declerck, Thierry; Krieger, Hans-Ulrich</i>	470
Integrated multilingual access to diverse Japanese humanities digital archives by dynamically linking data <i>Kuyama, Takeo; Batjargal, Biligsaikhan; Kimura, Fuminori; Maeda, Akira</i>	473
Linguistic concepts described with Media Query Language for automated annotation <i>Lenkiewicz, Anna; Lis, Magdalena; Lenkiewicz, Przemyslaw</i>	477
Virtual Reproduction of Gion Festival Yamahoko Parade <i>Li, Liang; Choi, Woong; Nishiura, Takanobu; Yano, Keiji; Hachimura, Kozaburo</i>	480
Complex entity management through EATS: the case of the Gascon Rolls Project <i>Litta Modignani Picozzi, Eleonora; Norrish, Jamie; Monteiro Vieira, Jose Miguel</i>	483
TextGrid Repository – Supporting the Data Curation Needs of Humanities Researchers <i>Lohmeier, Felix; Veentjer, Ubbo; Smith, Kathleen M.; Söring, Sibylle</i>	486
RIgeo.net – A Lab for Spatial Exploration of Historical Data <i>Loos, Lukas; Zipf, Alexander</i>	488
Automatic Topic Hierarchy Generation Using Wordnet <i>Monteiro Vieira, Jose Miguel; Brey, Gerhard †</i>	491
Hypotheses.org, une infrastructure pour les Digital Humanities <i>Muscinesi, Frédérique</i>	494
TXSTEP – an integrated XML-based scripting language for scholarly text data processing <i>Ott, Wilhelm; Ott, Tobias; Gasperlin, Oliver</i>	497

Exploring Prosopographical Resources Through Novel Tools and Visualizations: a Preliminary Investigation <i>Pasin, Michele</i>	499
Heterogeneity and Multilingualism vs. Usability – Challenges of the Database User Interface ‘Archiv-Editor’ <i>Plutte, Christoph</i>	502
Medievalists’ Use of Digital Resources, 2002 and 2012 <i>Porter, Dot</i>	505
Cross-cultural Approaches to Digital Humanities – Funding and Implementation <i>Rhody, Jason; Kümmel, Christoph; Effinger, Maria; Freedman, Richard; Magier, David; Förtsch, Reinhard</i>	506
CWRC-Writer: An In-Browser XML Editor <i>Rockwell, Geoffrey; Brown, Susan; Chartrand, James; Hesemeier, Susan</i>	508
The Musici Database <i>Roeder, Torsten; Plutte, Christoph</i>	511
The TEICHI Framework: Bringing TEI Lite to Drupal <i>Schöch, Christof; Achler, Stefan</i>	514
What Has Digital Curation Got to Do With Digital Humanities? <i>Schreibman, Susan; McCadden, Katiet Theresa; Coyle, Barry</i>	516
Orbis Latinus Online (OLO) <i>Schultes, Kilian Peter; Geissler, Stefan</i>	518
Semantically connecting text fragments – Text-Text-Link-Editor <i>Selig, Thomas; Küster, Marc W.; Conner, Eric Sean</i>	520
The Melesina Trench Project: Markup Vocabularies, Poetics, and Undergraduate Pedagogy <i>Singer, Kate</i>	522
Digital Edition of Carl Maria von Weber’s Collected Works <i>Stadler, Peter</i>	525
Data sharing, virtual collaboration, and textual analysis: Working on ‘Women Writers In History’ <i>van Dijk, Suzan; Hoogenboom, Hilde; Sanz, Amelia; Bergenmar, Jenny; Olsson, Leif- Jöran</i>	527
Storage Infrastructure of the Virtual Scriptorium St. Matthias <i>Vanscheidt, Philipp; Rapp, Andrea; Tonne, Danah</i>	529
Digital Emblematics – Enabling Humanities Research of a Popular Early Modern Genre <i>Wade, Mara R.; Stäcker, Thomas; Stein, Regine; Brandhorst, Hans; Graham, David</i>	532
DTAQ – Quality Assurance in a Large Corpus of Historical Texts <i>Wiegand, Frank</i>	535
The Digital Averroes Research Environment – Semantic Relations in the Editorial Sciences <i>Willems, Florian; Gärtner, Mattias</i>	537

AV Processing in eHumanities – a paradigm shift

Wittenburg, Peter; Lenkiewicz, Przemyslaw; Auer, Erik; Lenkiewicz, Anna; Gebre, Binyam Gebrekidan; Drude, Sebastian 538

Plenary Sessions

connections. The entry for *honra* in the line ‘...*tuyo es mi corazón, tuya es mi vida, mi honra y mi hacienda*’ from *El abencerraje* reads:

Definition: Public esteem received in acknowledgement of one’s strength of character, specifically here, a woman’s chastity and/or marital fidelity.

Expansion of meaning: In this period, honor (good reputation) was based on the following factors: social status, material wealth, strength of character, achievements, and in the case of a woman, her chastity and/or marital fidelity. Here, in pledging *vida*, *honra*, and *hacienda*, the noblewoman was committing everything she was and had to the man.

By integrating qualitative analyses of literary texts with quantitative computer derived methods, *GALGO* promotes a novel form of active reading and engagement in language and culture. In the last two decades, the pedagogical approach to foreign language study has moved beyond a primary emphasis on developing functional communication skills to an equally compelling interest in acquainting students with representative literary and cultural texts in the target language. *The Modern Language Association’s 2007 report* places the acquisition of transcultural and translanguingual competencies, namely the ability to reflect on how realities are shaped differently across languages, historical periods and cultures, at the core of the foreign languages discipline. *GALGO* engages learners in the analysis of classic literary texts in a meaningful way, allowing them to test their hypotheses as to possible lexical and cultural preconceptions through immediate feedback.

References

Burgett, B., and G. Hendler (2007). *Keywords for American Cultural Studies*. New York: New York UP.

Hart, D. P., S. E. Jarvis, W. P. Jennings, and D. Smith-Howell (2005). *Political Keywords. Using Language that Uses Us*. New York: Oxford UP.

MLA Ad Hoc Committee on Foreign Languages (2007). *Foreign Languages and Higher Education: New Structures for a Changed World*. New York.

Williams, R. (1983). *Keywords: A Vocabulary of Culture and Society*. 2nded. New York: Oxford UP.

Trees of Texts – Models and methods for an updated theory of medieval text stemmatology

Andrews, Tara Lee

tara.andrews@arts.kuleuven.be
Katholieke Universiteit Leuven, Belgium

Macé, Caroline

caroline.mace@arts.kuleuven.be
Katholieke Universiteit Leuven, Belgium

The construction of a stemma codicum from the manuscript witnesses to a given text remains a somewhat divisive issue in philology. To a classical philologist, it is a necessary first step toward creating a critical edition of the text;¹ to a scholar who adheres to the principles of new philology, the assumptions inherent in the construction of a stemma are so fundamentally unsound that the exercise is pointless.²

The advent of digital philology demands a fresh look at the practice of stemmatology; indeed there has been a great deal of work in the past two decades on the subject, and in particular on the applicability of similar work from the field of evolutionary biology on the reconstruction of phylogenetic relationships between species. Among the numerous works on this subject are: ³ More recent work has begun to overcome the limitations of these techniques when applied to manuscript texts, such as the biological assumption that no extant species is descended from another, or the assumption that any ancestor species has precisely two descendants, or the converse assumption that any descendant species has precisely one ancestor.⁴

Despite this considerable methodological progress, the fundamental assumptions at the core of stemmatology have yet to be formalised, or even seriously questioned. The purpose of this paper is to revisit these assumptions and simplifications; to propose a model for the representation of texts in all their variations that is tractable to computational analysis; to set forth a method for the evaluation of stemma hypotheses with reference to text variants, and vice versa; and to discuss the means by which we can provide the field of stemmatology with a formal underpinning appropriate for the abilities that modern computational methods now provide us to handle vast quantities of data.

1. Modelling a text tradition

The first step to computational analysis of any text tradition is to model it programmatically. This act of object modelling is distinct from, and orthogonal to, text encoding – the component texts may reach us in a variety of forms, including TEI XML and its critical apparatus module, CSV-format alignment tables, a base text and a list of variants exported from a word processor, or even individual transcriptions that remain to be collated. All of these must be parsed into a uniform computational object model for further programmatic analysis.

Our model is relatively simple, taking into account only those features of the text that are computationally useful for our purposes. We define a tradition as the union of all the versions of a text that exist, represented by its witnesses. A collation is the heart of the tradition, representing in a unified way the entire text with all its known variations. We have adapted the variant graph model proposed by Schmidt and Colomb⁵ for this purpose. Whereas the Schmidt/Colomb model is a single graph, where readings are represented along edges, ours is a pair of graphs, in which readings are represented by vertices. The set of vertices is identical in both graphs; in the ‘sequence’ graph, the witnesses themselves are represented as a set of paths. Each witness path is directed and acyclic, taking a route through the reading vertices from beginning to end, according to the order of its text. The witness path may branch, if for example the text contains corrections by the original scribe(s) or by later hands; the branches are labeled appropriately. If transpositions are present, the graph as a whole is not directed and acyclic (DAG); we have therefore found it most useful to represent transposed readings as separate vertices, linked through a relationship as described below. A DAG may then be trivially transformed into an alignment table, with transpositions handled either as separate rows in the table or as variant characters within the same row.



Figure 1: A portion of a variant graph, with directed witness paths and undirected relationships

The second graph concerns relationships between readings, or types of variation that occur within the text. We may specify that two different readings are orthographic or syntactic variants of the same root word, or that they have some semantic relationship to each other. We may also specify transpositions in

this way. The result is an unconnected and undirected graph that adds extra information to the sequence graph without compromising its properties. The relationship information allows association and regularization of reading data, producing thereby more exact versions of alignment tables for our statistical procedures.

2. Modelling a stemma hypothesis

The goal of our research is to arrive at an empirical model of medieval text transmission, and set forth a formalized means of deducing copying relationships between texts. As such, we have two tasks before us. We must be able to evaluate a stemma hypothesis according to the variation present in a text tradition, and we must be able to evaluate variants in a tradition according to a given stemma hypothesis.

The problem becomes computationally tractable once we realize that a stemma hypothesis is also a DAG. No matter the complexity of the stemma, no matter the degree of ‘horizontal transmission’, a stemma can be represented as a connected and rooted DAG, where each extant or hypothesized manuscript is a vertex, and each edge describes a source relationship from one manuscript to another. (In our model, any stemma has a single archetype; where multiple archetypes exist in a text tradition, we must either consider a common theoretical origin for these archetypes, or speak of multiple stemmata.)

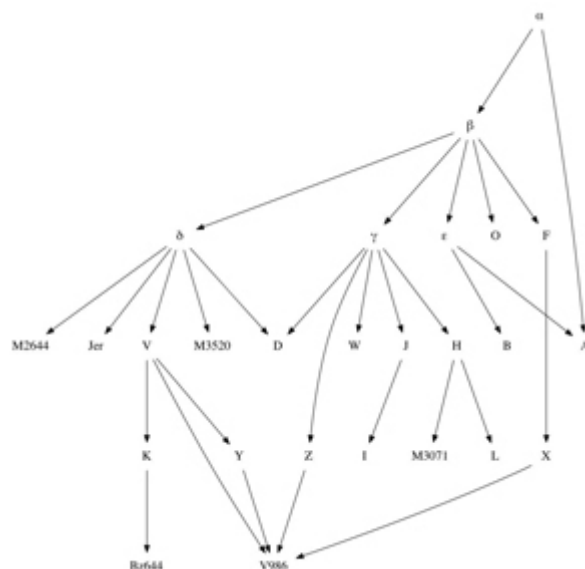


Figure 2: An arbitrarily complex stemma, represented as a DAG

In cooperation with the Knowledge Representation and Reasoning group within the KU Leuven Computer Science faculty,⁶ we have developed a method for evaluating sets of variants against a stemma hypothesis of arbitrary complexity, testing

whether any particular set of variants aligns with the stemma hypothesis, and measuring the relative stability of particular readings (defined in terms of how often they are copied vs. how often they are altered). In conjunction with the variant graphs of our text model, we are also able programmatically to draw some empirical conclusions about the relationships between readings as they are altered within the text.

3. An empirical approach to text variation

The idea of distinguishing ‘text-genealogical’ variants from the coincidental sort is not new;⁷ with the exception of U. Schmid (2004), however, none of the approaches taken have been empirical, in accordance with text relationships known from outside the variant tables. Given the means to quickly evaluate a stemma in detail according to the variants present in the text, distinguishing those variants that follow the stemma from those that do not, we should now easily be able to arrive at an empirical classification of genealogically ‘significant’ and ‘insignificant’ variants. This threatens to create a problem of circular logic: how to retrieve the genealogical variants in the real world? We build hypotheses on the basis of variation; we accept or reject variants as ‘genealogical’ based on the hypothesis. How do we break the cycle?⁸

We may adopt two approaches to anchor our methods in external results. First we test our methods on artificial traditions that have been created for the testing of stemmatological methods. These traditions are convenient in that the true stemmata are known, but they cannot reflect the true conditions under which medieval texts were copied. Thus we must also use ‘real’ text traditions, whose true stemmata are not fully known; in many medieval traditions, however, some paratextual information (e.g. colophon notations, physical damage or other physical characteristics of the texts) survives and can be used to establish or verify copying relationships. In our presentation we will discuss the results obtained on three artificial traditions, and three real medieval traditions with credible stemma hypotheses derived from external characteristics.

4. Conclusion

When examining a medieval tradition, we can rarely be certain what the scribe has copied in a text, and what has changed coincidentally or independently. To date, stemmatic reasoning has relied on the question of whether the scholar finds it likely that a given variant was preserved in copies. Given the models and methods for analysis proposed herein,

we have an opportunity to remove this limitation of ‘assumed likelihood’ and take all the evidence of a text into account, to build a statistical model independent of the constraints of evolutionary biology, using statistical probability rather than scholarly instinct alone.

References

- Blockeel, H., B. Bogaerts, M. Bruynooghe, B. de Cat, S. de Pooter, M. Denecker, A. Labarre, and J. Ramon** (2012). Modeling Machine Learning Problems with FO(\cdot). *Submitted to the 28th International Conference on Logic Programming*.
- Cameron, H. D.** (1987). The upside-down cladogram. Problems in manuscript affiliation. In H. M. Hoenigswald and L. F. Wiener (eds.), *Biological metaphor and cladistic classification : an interdisciplinary perspective*. Philadelphia: U of Pennsylvania P.
- Cerquiglini, B.** (1989). *Éloge de la variante : histoire critique de la philologie*, Paris: Éd. du Seuil.
- Howe, C. J., A. C. Barbrook, M. Spencer, P. Robinson, B. Bordalejo, and L. Mooney** (2001). Manuscript Evolution. *Trends in Genetics* 17: 147-52.
- Maas, P.** (1957). *Textkritik*. Leipzig: Teubner.
- Macé, C., and P. Baret** (2006). Why Phylogenetic Methods Work: The Theory of Evolution and Textual Criticism. In C. Macé, P. Baret, A. Bozzi, and L. Cignoni (eds.), *The Evolution of Texts: Confronting Stemmatological and Genetical Methods*. Pisa, Rome: Istituti Editoriali e Poligrafici Internazionali.
- Pasquali, G.** (1962). *Storia della tradizione e critica del testo*. Firenze: F. Le Monnier.
- Reeve, M. D.** (1986). Stemmatic method: ‘Qualcosa che non funziona?’ In P. F. Ganz (ed.), *The role of the book in medieval culture : proceedings of the Oxford International Symposium, 26 September-1 October 1982*. Turnhout: Brepols.
- Roos, T., and Y. Zou** (2011). Analysis of Textual Variation by Latent Tree Structures. *IEEE International Conference on Data Mining, 2011 Vancouver*.
- Salemans, B. J. P.** (2000). *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way: The Case of Fourteen Text Versions of Lanseloet van Denemerken*. Ph.D., Katholieke Universiteit Nijmegen.
- Schmid, U.** (2004). Genealogy by chance! On the significance of accidental variation (parallelisms). In P. T. van Reenen, A. den Hollander, and M.

van Mulken (eds.), *Studies in Stemmatology II*. Amsterdam: Benjamins.

Schmidt, D., and R. Colomb (2009). A data structure for representing multi-version texts online. *International Journal of Human-Computer Studies* 67: 497-514.

Spencer, M., and C. J. Howe (2001). Estimating distances between manuscripts based on copying errors. *Literary and Linguistic Computing* 16: 467-484.

P. T. van Reenen, A. den Hollander, and M. van Mulken, eds. (2004). *Studies in Stemmatology I*, Amsterdam: Benjamins.

West, M. L. (1973). *Textual Criticism and Editorial Technique: Applicable to Greek and Latin Texts*. Stuttgart: Teubner.

Notes

1. Maas 1957; Pasquali 1962; West 1973.
2. See, among many references on this question Cerquiglini 1989; Reeve 1986.
3. Cameron 1987; Howe et al. 2001; Macé & Baret 2006; Spencer & Howe 2001.
4. Roos & Zou 2011.
5. Schmidt & Colomb 2009.
6. Blockeel et al. 2012.
7. Salemans 2000. See also papers by Wachtel, van Mulken, Schmid, Smelik, Schøsler, and Spencer et al in van Reenen et al. 2004.
8. See the critique of concerning the work of Salemans.

Mapping the Information Science Domain

Arazy, Ofer

ofer.razy@ualberta.ca
University of Alberta, Canada

Ruecker, Stan

sruecker@ualberta.ca
Illinois Institute of Technology, USA

Rodriguez, Omar

omar.rodriguez@gmail.com
University of Alberta, Canada

Giacometti, Alejandro

alejandro.giacometti@gmail.com
University College London, UK

Zhang, Lu

lu.eva.zhang@gmail.com
University of Alberta, Canada

Chun, Su

schun@ualberta.ca
University of Alberta, Canada

In the scholarly community, there has been for the last fifteen years an increasing interest in interdisciplinary projects, where researchers with different disciplinary backgrounds form teams to collaborate in addressing problems that are either not clearly within a single research domain, or that can best be approached by combining a variety of approaches. However, there also remain a number of research areas where interdisciplinary approaches seem to be warranted, but scholars in single disciplines have continued to work largely in isolation from their colleagues in other domains. We focus on one specific scientific field – information science – which is investigated in various disciplines: library and archival studies, computer science, management (and specifically, management information systems), and engineering, to name just a few.

Prior works in the area have discussed the disconnect between various areas within the broad field of information science, relying primarily on co-citation techniques and content analysis. These studies concluded that although all domains within this area are focused on the concepts of information, people, and (information and communication) technologies, the disciplines remain disjunct in terms of their disciplinary recognitions and topic of research. Despite some similarities in research methodology ('research approach' and 'level of analysis') (Glass et al. 2004), there are substantial differences in



www.dh2012.uni-hamburg.de
ISBN 978-3-937816-99-9

