The European Association for Digital Humanities (EADH) Association for Computers and the Humanities (ACH) Canadian Society for Digital Humanities / Société canadienne des humanités numériques (CSDH/SCHN) centerNet

Australasian Association for Digital Humanities (aaDH) Japanese Association for Digital Humanites (JADH)

Digital Humanities 2016

Conference Abstracts

Jagiellonian University め Pedagogical University

> Kraków 11–16 July 2016



Edited by

Maciej Eder Jan Rybicki

DHConvalidator service

Marco Petris

On-line abstracts

Michał Woźniak

Design and typesetting

Maciej Eder

Proof-reading

Aleksandra Ptasznik Karolina Eder Daria Wyka Karolina Wróbel Weronika Ślęczkowska Anita Uryga Sandra Romanowicz

Welcome to Digital Humanities 2016

The annual "Digital Humanities" conference, first held in 1989, gives a clear proof that the field of DH is, first and foremost, very well established, even if more and more subfields emerge almost every year. Secondly, the conference confirms the fact that the discipline is constantly growing. This year is no exception. Even more: the 27th joint conference of EADH (ALLC) and ACH, and the 8th conference under the auspices of ADHO, is by far the biggest event in the field, with its almost 450 accepted submissions in different categories: panels, long papers, short papers, posters, and pre-conference workshops. The number of registered participants exceeded 850 at the time of writing these words.

The conference takes place in Kraków; this is only the second time (after Debrecen 1998) that it comes to Central/Eastern Europe. The region's rich past and its recent rapid growth has inspired the conference theme, "Digital Identities: the Past and the Future". This theme aims at stressing the very strong connections between DH and its roots in the medieval idea of a university with a prominent role of the Liberal Arts. We strongly believe that the unique relation between the origins of the humanities' scholarship and the opportunities provided by computer algorithms and the enormous amount of resources (text collections, linguistic corpora, databases, virtual libraries) can lead to a new scientific revolution.

Kraków, the venue of the conference, has been a major center of learning and culture in this part of the world: the Jagiellonian, founded in 1364, is usually ranked first among Polish universities, and the same is true of the Kraków's Pedagogical University within Poland's quite extensive community of pedagogical universities. No wonder, then, that the conference is hosted jointly by those two institutions: the Jagiellonian University and the Pedagogical University of Kraków. Their collaboration is a manifestation of the vivid digital humanities scene emerging in Poland's major centre of learning and culture.

I would like to thank all those who submitted proposals this year and all those who agreed to act as reviewers. The work load for the group of scholars who undertook the demanding task of reviewing the submissions was significantly higher than in previous years – I would like to thank them all for their great job. I would like to give my thanks to the members of the Program Committee, who this year included: Diane Jakacki (CSDH/SCHN), Michael Eberle-Sinatra (CSDH/SCHN), Jennifer Guiliano (ACH), Brett D. Hirsch (aaDH), Leif Isaksen (EADH), Asanobu Kitamoto (JADH), Inna Kizhner (centerNet), Maurizio Lana (EADH), Kiyonori Nagasaki (JADH), Roopika Risam (ACH), Glenn Roe (aaDH), Sinai Rusinek (centerNet) and Deb Verhoeven (aaDH). My special and warmest thanks go to the Program Committee Chair Manfred Thaller, whose contribution to the conference was simply outstanding. I also want to mention the Local Organizers, who were brave enough to suggest Kraków as a potential host for the DH conference, and efficient enough to make it happen.

Karina van Dalen-Oskam ADHO Steering Committee Chair The 27th Joint International Conference of the Association for Literary and Linguistic Computing and Association for Computers and the Humanities,
The 8th Joint International Conference of the Alliance of Digital Humanities Organizations

Program Committee

Manfred Thaller (chair)
Diane Jakacki (vice-chair)
Michael Eberle-Sinatra
Jennifer Guiliano
Brett D. Hirsch
Leif Isaksen
Asanobu Kitamoto
Inna Kizhner
Maurizio Lana
Kiyonori Nagasaki
Roopika Risam
Glenn Roe
Sinai Rusinek
Deb Verhoeven

Local Organizers

Maciej Eder Jan Rybicki

Local Organizing Board

Katarzyna Bazarnik
Elżbieta Górska
Rafał Górski
Magda Heydel
Władysław Marek Kolasa
Krzysztof Nowak
Iwona Pietrzkiewicz
Maria Piotrowska
Bogusław Skowronek
Elżbieta Tabakowska
Ewa Willim
Grażyna Wrona

Gold Sponsor

GALE Cengage Learning

Organizers

Alliance of Digital Humanities Organizations (ADHO) Jagiellonian University, Faculty of Philology Pedagogical University in Kraków, Faculty of Philology

Table of Contents

Plenary lectures

Can CERN serve as a model for Digital Humanities?	35
Early Funding of Humanities Computing: A Personal History	35
Touching the interface: Bishop Cosin and unsolved problems in (digital) information design Claire Warwick	35
Panels	
Playable Books at Electronic Literature's Interface	39
Institutionalizing and implementing the Digital Yiddish Theatre Project	42
Linked Ancient World Data: Relating the Past	43
Creating Feminist Infrastructure in the Digital Humanities	47
The Scholarly Digital Edition: Best Practices, Guidelines, and Peer Evaluation	51
Are the Digital Humanists Prepared for Open Access to Research Data?	53
Developing Local Digital Humanities Communities: The Atlanta Studies Network	54
Digitally Mapping Romantic Literature and Culture	56
Quality Matters: Diversity and the Digital Humanities in 2016	61
Literary Concepts: The Past and the Future	63

Katherine Mary Faull; Diane Katherine Jakacki; James O'Sullivan; Amy Earhart; Micki Kaufman	00
Digital Data Sharing: Opportunities and Challenges of Opening Research	68
Recovering Shared Heritage via Spectral Imaging: Problems, Solutions, Interpretations Gregory Heyworth; Michael Phelps; Adrian Wisnicki; Kenneth Boydston; Roger Easton; Chet Adam Van Duzer	69
Diverse Digitalities: Targeted Models for Postcolonial Challenges in the Digital Discourse Nirmala Menon; Alex Gil; Rahul Gairola	72
Web Historiography - A New Challenge for Digital Humanities?	74
Boundary Land: Diversity as a defining feature of the Digital Humanities	76
Intersectional Scholarship in Electronic Literature and Digital Humanities	82
The Trace of Theory: Extracting Subsets from Large Collections	85
Digital Folkloristics: the Use of Computational Methods in Revealing the Characteristics of Folkloric Communication	92
APIs in Digital Humanities: The Infrastructural Turn	93
MEDEA (Modeling Semantically Enriched Digital Editions of Accounts)	96
A Model for International Cooperation: Emblematica Online and Linked Data in Research and Pedagogy	98
When DH Meets Law: Problems, Solutions, Perspectives	104
Long papers	
Representations Of Race: Mining Identity In American Fiction, 1789-1964	111

For Architectural Historians	113
Geocoding Thousands of Fiscal Records: Methodological Approach for a Study on Urban	
Retail Trade in the Belle Époque	110
Unlocking The Mayan Script With Unicode	119
Developing Competencies in Digital Scholarship Among Humanities Scholars	12
Prototypes as Thinking through Making. Decision Points and Evaluation in Prototyping a Visualisation Framework for Historical Documents	123
Combining Corpora and Statistics using Geographical Technologies: New Evidence on Nineteenth Century Infant Mortality Decline in England and Wales	127
vWise: Visual Workspace for Information Seeking and Exploration	130
From Index Cards to a Digital Information System: Teaching Data Modeling to Master's Students in History	132
Modelling Taxonomies of Text Reuse in the Deipnosophists of Athenaeus of Naucratis: Declarative Digital Scholarship	133
Two Centuries of Russian Roads – Diachronic Study of Polysemy in the Context of Cultural Change	137
Ancient Maya Writings as High-Dimensional Data: a Visualization Approach	139
The Evolution of Virtual Harlem: Bringing the Jazz Age to Life	143
Jonathan Edwards and Thomas Foxcroft: In Pursuit of Stylometric Traces of the Editor Michał Choiński; Jan Rybicki	147
What's in a Topic Modell fondamenti del text mining negli studi letterari	14
An OWL 2 Formal Ontology for the Text Encoding Initiative	15
Digital Palaeography: What is digital about it?	15

ARLO (Adaptive Recognition with Layered Optimization): a Prototype for High Performance Analysis of Sound Collections in the Humanities	158
Tanya Clement; Steve McLaughlin; David Tcheng; Loretta Auvil; Tony Borries	
Contextualized Integration of Digital Humanities Research: Using the NeMO Ontology of Digital Humanities Methods	161
Frequency and Meaning	163
Creating An EpiDoc Corpus for Ancient Sicily	165
Sustainable publishing – Standardization possibilities for Digital Scholarly Edition technology . Alexander Czmiel	167
Visual Patterns Discovery in Large Databases of Paintings	169
Urban Youth and Community Media: A Digital Place-Making Process in Vanuatu	172
Sequentiality in Genetic Digital Scholarly Editions. Models for Encoding the Dynamics of the Writing Process	174
correspSearch - A Web Service to Connect Diverse Scholarly Editions of Letters Stefan Dumont	175
Corpus Analyses of Multimodal Narrative: The Example of Graphic Novels	178
Déchiffrer Le Mythe De l'Amour	180
Historical Linguistics' New Toys, or Stylometry Applied to the Study of Language Change Maciej Eder; Rafał Górski	182
Go Set A Watchman while we Kill the Mockingbird in Cold Blood, with Cats and Other People Maciej Eder; Jan Rybicki	184
OVAL: A Virtual Ecosystem for Immersive Scholarship and Teaching	186
Outliers or Key Profiles? Understanding Distance Measures for Authorship Attribution Stefan Evert; Fotis Jannidis; Thomas Proisl; Thorsten Vitt; Christof Schöch; Steffen Pielström; Isabella Reger	188
Die digitale Modellierung experimenteller Druckgrafiken des 15. Jahrhunderts. Anforderungen und Chancen computerbasierter Dokumentationsverfahren	192

REDEN ONLINE: Disambiguation, Linking and Visualisation of References in TEI Digital Editions Francesca Frontini; Carmen Brando; Jean-Gabriel Ganascia	193
New DH Publishing Models and Geopolitical Diversity	197
Crossed Semantic Analysis of Literary Texts with DeSeRT	199
Pulp Science Fiction's Legacy to Women in Science	202
Building Blocks of Fiction: Lexical Bundles in Nineteenth-Century Novels	203
Performance, the Document, and the Digital: the Case of Lynn Hershman Leeson's 'Robertas' . Gabriella Giannachi	206
RICardo Project: Exploring 19th Century International Trade	208
An Augmented Reality Mobile Application for Intergenerational Learning and Critical Connection: Experiencing The Histories of Flushing Meadows Park	210
Bringing Migration Data Into Context Using Digital Computational Methods	213
Cross-Institutional Music Document Search	215
Significance Testing for the Classification of Literary Subgenres	218
Microanalyzing Parts of Texts	220
Textual Variation, Text-Randomization, and Microanalysis	223
Assessing a Shape Descriptor for Analysis of Mesoamerican Hieroglyphics: A View Towards Practice in Digital Humanities	225
Visualizing Mouvance: Towards an Alignment of Medieval Vernacular Text Traditions Stefan Jänicke; David Joseph Wrisley	230
Topic Modeling Literary Quality	233
Teaching Digital Humanities Through a Community-Engaged, Team-Based Pedagogy Andrew Jewell; Elizabeth Lorang	237

Geohumanities Projects	239
Authorship Attribution Using Different Languages	241
Regional Classification of Traditional Japanese Folk Songs from the Chugoku District	243
Authorship Verification with the Ruzicka Metric	246
GLAMorous! Edizione Digitale Di Beni Culturali Con Contenuto Testuale, Multidisciplinarietà Ed Epigrafia Digitale	249
Du Texte Antique À La Publication Scientifique : Outils D'Analyse Numériques Des Contenus Et Ponts Conceptuels	250
Harvesting History: Democratizing The Past Through The Digitization Of Community History Connie Lee Lester	254
If Paintings were Plants: Measuring Genre Diversity in Seventeenth-Century Dutch Painting and Printmaking	256
Quantitative Analyses of Chinese Poetry of Tang and Song Dynasties: Using Changing Colors and Innovative Terms as Examples	260
Evaluating Modal Use in News Corpus for Constructing Rhetorical Context of Historical Event . Jyi-Shane Liu; Ching-Ying Lee; Ke-Chih Ning	262
CORE - A Contextual Reader based on Linked Data	267
Khepri – a Modular View-Based Tool for Exploring (Historical Sociolinguistic) Data Eetu Mäkelä; Tanja Säily; Terttu Nevalainen	269
Where Close and Distant Readings Meet: Text Clustering Methods in Literary Analysis of Weblog Genres	273
Wikidition: Towards A Multi-layer Network Model of Intertextuality	276
Mapping the Bentham Corpus	279
Contextualizing Receptions of World Literature by Mining Multilingual Wikipedias	282

Exploring and Discovering Archive-It Collections with Warcbase	285
Representación De Otras Literaturas Mexicanas En Medios Digitales	288
Toward A Use-Value Paradigm For The Sustainability Of Digital Research	291
If You Build It Will They Come? Digital Infrastructure And Disciplinary Practice In Language Documentation	294
The Dialogic Turn and the Performance of Gender: the English Canon 1782–2011 Grace Muzny; Mark Algee-Hewitt; Dan Jurafsky	296
Digital Humanities in Cultural Areas Using Texts That Lack Word Spacing	300
Player-Driven Content: Analysing Textual Communications in Online Roleplay James O'Sullivan; Michelle Shade; Ben Rowles	303
Digital Annotation Tooling for Opera Performance Studies	306
Bretez: Conjugaison du passé au futur	309
Chronological corpora: Challenges and opportunities of sequential analysis. The example of ChronoPress corpus of Polish	311
Social Networks and Archival Context: People and Cultural Heritage	313
Early English Books in Context: Towards a History of the Technological Humanities Daniel James Powell	315
Projet Odysseus, Outil d'Etudes Comparatives Du Traductologue	317
The Lifecycle of a Digital African Studies Projects: Creating Sustainable, Equitable, and Ethical Projects	320
Addressing Torture in Iraq through Critical Digital Media Art—Hearts and Minds: The Interrogations Project	322
An Iterative 3DGIS Analysis of the Role of Visibility in Ancient Landscapes	324
The Trajectories Tool: Amplifying Network Visualization Complexity	328

Haute Figure Dialogique	330
Modelling Music Reception: An Ontology For Representing Interpretations of Richard Wagner's Leitmotifs	333
Digging into ECCO: Identifying Commonplaces and other Forms of Text Reuse at Scale Glenn Roe; Clovis Gladstone; Robert Morrissey; Mark Olsen	336
Implementation of a National Data Center for the Humanities (DaSCH)	339
Digicraft and 'Systemic' Thinking in Digital Humanities	341
Using Computer Numerical Control Techniques to Prototype Media History	344
Straight Talk! Automatic Recognition of Direct Speech in Nineteenth-Century French Novels Christof Schöch; Daniel Schlör; Stefanie Popp; Annelen Brunner; Ulrike Henny; José Calvo Tello	346
#dariahTeach: online teaching, MOOCs and beyond	354
The Computer Graphic Simulation of the Battle at Mount Street Bridge. Problems, Perspectives, and Challenges	356
The Mutual Relationship of Linguistic and Non-linguistic Elements in Breaking Down the Hierarchy of Language in Digital Narrative	358
What Do Boy Bands Tell Us About Disasters? The Social Media Response to the Nepal Earthquake	361
Full Stack DH: Building a Virtual Research Environment on a Raspberry PI	364
SpotiBot-Turing testing Spotify	366
Reconstruction of Labour Relations in the North Sea Region in the Late Middle Ages: Spatio-Temporal Analysis Using Historical GIS, Taxation Sources, and Coin Finds	366
Curating Just-In-Time Datasets from the Web	369
Read, Play, Build: Teaching Sherlock Holmes through Digital Humanities	371

The North Carolina Jukebox Project: Archives Alive and the Making of Digital Cultural Heritage	372
The Online Archive "Forced Labor 1939–1945. Memory and History". A digital Application for Research and Education	375
Enabling Complex Analysis of Large-Scale Digital Collections: Humanities Research, High Performance Computing, and transforming access to British Library Digital Collections Melissa Terras; James Baker; James Hetherington; David Beavan; Anne Welsh; Helen O'Neill; Will Finley; Oliver Duke-Williams; Adam Farquhar	376
From Order to Order Switch. Mediating between Complexity and Reproducibility in the Context of Automated Literary Annotation	379
Printing in a Periphery: a Quantitative Study of Finnish Knowledge Production, 1640–1828 Mikko Tolonen; Niko Ilomäki; Hege Roivainen; Leo Lahti	383
Theatre Plays as 'Small Worlds'? Network Data on the History and Typology of German Drama, 1730–1930	385
A literary rat race	388
Taalportaal: A New Tool For Linguistic Research	390
Stylochronometry and the Periodization of Samuel Beckett's Prose	393
Qu'est-ce qu'un texte numérique? A New Rationale for the Digital Representation of Text Joris J. Van Zundert; Tara L. Andrews	395
Music notation addressability	398
Introducing HistoGraph 2: Exploration of Cultural Heritage Documents Based on Co-Occurrence Graphs	400
Schrifttanz: Written Dance/Movement Poems	402
Flexible Community-driven Metadata with the Component Metadata Infrastructure	405
The Manuscripts of David Livingstone and New Frontiers for Spectral Imaging	407
Public and Private Views of Texts in Digital Editions – The Case of the Kanseki Repository Christian Wittern	409

Oulipian Stylometry	411
DIVAServices-Spotlight – Experimenting with Document Image Analysis Methods in the Web Marcel Würsch; Michael Bärtschi; Rolf Ingold; Marcus Liwicki	414
Visualising the Dynamics of Character Networks	417
Work In A Globalised World. Allocation Algorithm To Add Labour Relations To Digitised Census Data	419
Short papers	
Automatic Detection of Characters in Case Insensitive Text in Comics	425
Minimal Editions in the Classroom: A Pedagogical Proposal	426
Testing the Doctrine of Election: A Computational Approach to Karl Barth's Church Dogmatics . Christopher Scott Bailey; Eric Rochester	428
Concept Modeling the Advertising Chinese Modern Society	431
The Great War on the Web: the Making of Citing and Referencing by Amateurs Valérie Beaudouin; Zeynep Pehlivan	433
Countering Counter Mapping Methods: Constructing A Humanities GIS Methodology In the Age of Electracy	436
Does Character Speech Matter? A Quantitative Approach	438
Tool-based Identification of Melodic Patterns in MusicXML Documents	440
The Preparation of the Topic Model	442
Éditions Critiques Électroniques Et Structuration Du Contenu Sur Les Plateformes De Consultation Numériques: Normes Et Pratiques	445
An Islamic Manuscript Database as a Network of Objects	447

Deux Projets D'Edition Numérique Dans Le Cadre Du Projet SyMoGIH: Le Journal De Léonard Michon Et Les Actes Des Synodes Des Églises Réformées De Bourgogne	448
Constructing Evidence in the Photographic Archive: The Experience of Digital Humanists	450
Alexandra M Chassanoff	
Compiling a Database on Historical China from Local Records: The Local Gazetteers Project at MPIWG	452
Shih-Pei Chen; Zoe Hong; Dagmar Schäfer; Martina Siebert; Jorge Urzúa	
Who's Doing What?: Examining The Relationships Among Subjectivity, Agency, and Syntax In The 19th Century Novel	456
Etymology Meets Linked Data. A Case Study In Turkic	458
#ww1. The Great War on Twitter	461
A Method for Record Linkage with Sparse Historical Data	463
Metacanon.org: Digital Humanities and the Canon	466
Mining Leitmotif in James Joyce's 'Ulysses'	467
WordPress as a Framework for Automated Data Capture, Filtering and Structuring Processes: The New Order of the Authors	469
The Project Zeri Photo Archive: Towards a Model for Defining Authoritative Authorship Attributions	472
Materiality and Metadata of Digitised Photographs: A Theoretical Inquiry	474
Notes from the Transcription Desk: Modes of engagement between the community and the resource of the Letters of 1916	476
Automatisation Du Workflow Audiovisuel, Quel Impact Sur Le Spectateur?	478
Using Big Data Techniques For Searching Digital Archives: use cases in Digital Humanities Janet Delve; Sven Schlarb; Rainer Schmidt; Richard Healey	480
SAMEBibl: Sistema Automático de Migración a Europeana para Bibliotecas	482

"Parcours Numériques"	487
How IBM Watson Can Help Us Understand Character in Shakespeare: A Cognitive Computing Approach to the Plays	488
Visualisation Strategies for Comparing Political Ideas with the ORATIO Platform	492
Arduino Circuits and Javanese Puppets: 'Re-materializing' Digital Archives through Tangible Interfaces	495
Topical Diversification Over Time In The Royal Society Corpus	496
Writing Composition in the Close Reading Cycle: Developing The Annotation Studio Idea Space	500
Seeing Andalucia's Late Gothic heritage through GIS and Graphs	501
Choosing Words for Stylometric Authorship Attribution: Evaluating Most Distinguishing Words (MDWs) vs. Most Frequent Words (MFWs)	504
File Formats for Archiving: Stability and Persistence Issues	507
Approaches to Thematic Classification for Latin Epic	508
Toccata: Text-Oriented Computational Classifier Applicable To Authorship	510
Repairing William Playfair: Digital Fabrication, Design Theory, and the Long History of Data Visualization	513
Anonymity and Online Discussion: A New Framework for Analysis	516
Indigenous digital humanities. Participatory geo-referenced-mapping and visualization for digital data management platforms in digital anthropology	518
Standardized Digital workflow for Archiving Local Knowledge	519
Playing with French Drama: from Old Research Questions to New Research Tools	522

Big Data and the Study of Allusion: an Exploration of Tesserae's Multitext Capability James O'Brien Gawley; A. Caitlin Diddams	523
Biblissima - Following Medieval Manuscripts and Incunabula through their Existence via a Semantic Web Application	525
DH Poetry Modelling: a Quest for Philological and Technical Standardization	526
Translating Electronic Literature. Multicultural, Multilingual and Cross-Platform Encounters Monika Górska-Olesińska; Mariusz Pisarski	528
The journal al-Muqtabas between Shamela.ws, HathiTrust, and GitHub: producing open, collaborative, and fully-referencable digital editions of early Arabic periodicals – with almost no funds	529
Archives Distant Reading: Mapping the Activity of the League of Nations' Intellectual Cooperation	531
Scholarly Requirements for Large Scale Text Analysis: A User Needs Assessment by the HathiTrust Research Center	534
New Maps for the Lettered City: a Data Visualization Exploration of 19th Century Salons in Mexico	537
Project Dialogism: Toward a Computational History of Vocal Diversity in English-Language Literature	543
Measuring the Dynamics of Lexico-Semantic Change Since the German Romantic Period	545
Academic Pillow-Talk and Two Immersive Explorations of Linguistic Space	547
Tracing the Genesis of Pessoa's Envisioned Work: a Digital Edition of his Editorial Projects and Publications	549
KARREK: Building and Annotating a Kafka/Reference Corpus	552
Invisible Cities In Literature And History: Interfaces To Scalable Readings Of Textual And Visual Representations Of The Imaginary	553
Silva Portentosissima – Computer-Assisted Reflections on Bifurcativity in Stemmas	557

Susan Hogervorst	360
Stylometry on the Silver Screen: Authorial and Translatorial Signals in Film Dialogue Agata Hołobut; Jan Rybicki; Monika Woźniak	561
Identity Lenses in Analyzing Evolving Social Structures	565
Exploring the Rules of Rhyme: Operationalizing Historical Poetics	567
From Keyword Search To Discourse Mining – The Meaning Of Scientific Management In Dutch Vocabulary, 1900-1940	569
Publishing Second World War History as Linked Data Events on the Semantic Web Eero Hyvönen; Erkki Heino; Petri Leskinen; Esko Ikkala; Mikko Koho; Minna Tamper; Jouni Tuominen; Eetu Mäkelä	<i>57</i> 1
Contextualizing Historical Places in a Gazetteer by Using Historical Maps and Linked Data Esko Ikkala; Jouni Tuominen; Eero Hyvönen	573
Wired!: Collaborative Teaching & Critical Digital Making In An Art History Classroom Hannah L. Jacobs	577
Comparison of Methods for the Identification of Main Characters in German Novels Fotis Jannidis; Isabella Reger; Markus Krug; Lukas Weimer; Luisa Macharowsky; Frank Puppe	578
Diplomatic History by DataUnderstanding Cold War Foreign Policy Ideology using Networks and NLP	582
Exploring The History Of The Qur'an Digitally	585
Adding Value to a Research Infrastructure Through User-contributed ePublications	586
Implementing Canonical Text Services in the Croatiae Auctores Latini Digital Collection Neven Jovanović; Alexander Simrell	588
The Stanford Code Poetry Slam through Critical Code Studies	590
Trading Zones of Digital History	591
Converting the Liddell Scott Greek-English Lexicon into Linked Open Data using lemon Fahad Khan; Francesca Frontini; Federico Boschetti; Monica Monachini	593
Digital Criticism Platform for Evidence-based Digital Humanities with Applications to Historical Studies of Silk Road	596

Museum Digitization Practices Across Russia: Survey and Web Site Exploration Results Inna Kizhner; Melissa Terras; Maxim Rumyantsev	600
Tracking Online User Behaviour With A Multimethod Research Design	602
Researchers to your Driving Seats: Building a Graphical User Interface for Multilingual Topic-Modelling in R with Shiny	605
The First World War in Perm Provincial Periodicals	608
Language Attitudes of Twitter Users Toward New York City English	609
All in the Family: Testing Burrows' Delta on Robert Louis Stevenson's Collaboratively Authored Volumes The Dynamiter and The Wrecker	610
Morphology beyond inflection. Building a wordformation based dictionary for Latin Eleonora Litta; Marco Carlo Passarotti	612
Analyzing the 17th Century Theatre Critique Texts with a Semantic Annotation Tool Driven by a Dedicated Ontology	614
Méthodes computationnelles et analyse d'une langue de chancellerie: le logiciel d'analyse textuelle Machiato et la correspondance diplomatique et administrative de Machiavel	617
"El Atambor de Plata Suena como Cascaveles de Turquesa". Reconstrucción de la Experiencia Sonora de la Colonización Europea (c. 1480–1650) a Través de un Glosario y un Tesauro Digital	619
Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data	621
The Digital Scholarship Training Programme at British Library: Concluding Report & Future Developments	623
Building Large Persons' Networks to Explore Digital Corpora	625
Collaborative Translation Using FromThePage	629
Linguistic Variation In The Hebrew Bible: Digging Deeper Than The Word Level	630

Linking and Labeled LDA	632
Visualising Cultural Spheres – Virtual Tours and Epigraphical Data	635
Stefan George Digital: Exploring Typography In A Digital Scholarly Edition Frederike Neuber	637
A Comparative Analysis of Bibliographic Ontologies: Implications for Digital Humanities Terhi Nurmikko-Fuller; Jacob Jett; Timothy Cole; Chris Maden; Kevin R. Page; J. Stephen Downie	639
Digging the Aboveground: Visual Archeology of an Asylum	642
Data Praxis in the Digital Humanities: Use, Production, Access	644
Dépasser La Liste : Quand La Bibliothèque Entre Dans La Danse Des Corpus Web	646
RAPSCAPE – un'esplorazione dell'universo linguistico del rap attraverso il text-mining e la data-visualization	648
Metaphor Mining in Historical German Novels: Using Unsupervised Learning to Uncover Conceptual Systems in Literature	651
When Traditional Ontologies are not Enough: Modelling and Visualizing Dynamic Ontologies in Semantic-Based Access to Texts	653
Picture to Score: Driving Vector Animations with Music in the XML Ecosystem Stephen Ramsay; Brian Pytlik-Zillig	656
Git-Lit: an Application of Distributed Version Control Technology toward the Creation of 50,000 Digital Scholarly Editions	657
Researchers' perceptions of DH trends and topics in the English and Spanish-speaking community. DayofDH data as a case study	658
Mnemosyne: A Smartlibrary for Rare and Forgotten Texts	661
Some Problems in the Non-Traditional Authorship Attribution Studies of the Dramatic Canon of William Shakespeare: Are they Insurmountable	662

Climate Negotiation Analysis	663
Modeling as Discourse: The case for 3D	667
Mapping European Periodical Counterpublics: Building a Sustainable Collaborative Framework for European Periodical Studies	668
Erlebter Raum im Rom der späten Republik - eine digitale Forschungsumgebung Leif Scheuermann; Klaus P. Jantke; Walter Scheuermann	670
Linking Qualitative and Quantitative Research by Thin Descriptions through Semantic Graphs. Experiments in Apparatus Design with Semantic CorA	672
Making Sense of Illustrated Handwritten Archives	674
Digital History "From Below": a Call to Action	677
Change, Transition and Governance: Lessons from a Long-Term, Large Scale DH Collaboration Lynne Siemens	678
From Handwritten Text to Structured Data: Alternatives to Editing Large Archival Series Ronald Sluijter; Marielle Scherer; Sebastiaan Derks; Ida Nijenhuis; Walter Ravenek; Rik Hoekstra	680
SMTP: Stedelijk Museum Text Mining Project	683
Comparing Digital Scholarly Editions	686
Evaluating GitHub as a Platform of Knowledge for the Humanities	688
A Study of Knowledge Integration in Digital Humanities Based on Bibliographic Analysis Muh-Chyun Tang; Yun Jen Cheng; Kuang Hua Chen; Jieh Hsiang	690
Computers on Law and Order	692
Thresholds: Valuing the Creative Process in Digital Publishing	693
Adding Semantics To Comics Using A Crowdsourcing Approach	694
Mapping Multilingual Responses To Famine And Dearth In The Early Modern Landscapes Of India And Britain	697

A lesson in applied minimalism: adopting the TEI processing model	698
Exploring Networks Of Confidentiality And Secrecy In Early Modern Transconfessional Correspondences	699
Using Big Cultural Data To Understand Diversity And Reciprocity In The Global Flow Of Contemporary Cinema	<i>7</i> 01
"Digital" in practice: survey of Russian historians' research practices	703
The School of Salamanca on the Semantic Web	704
Preserving Ireland's Digital Cultural Identity towards 2116	706
Faceting and Mining Network Graphs	708
Tackling Terms In Furetiere's 'Dictionnaire Universel'	710
Digitale Tools und Methoden für die geisteswissenschaftliche Forschung praxisnah erklärt: Ein neues Format im Test	<i>7</i> 11
The Formation of Australia's Economic History Community, 1950–1970: A Multidimensional Network Analysis	713
A Management of Personal Name with Alternate Name and its Searching for Japanese Historical Study	716
Computation-Aided Analysis on Film Credits	717
From Language Revolution to Revolutionary Language	720
Posters	
Content Based Social Network Analysis of Reşat Nuri Güntekin's Letters	725
Engaging Students in Digital Literary Analysis: GALGO (Golden Age Literature Glossary Online), a Social Semiotic Platform	725

The DEFC-App: A Web-based Archaeological Data Management System for 'Digitizing Early Farming Cultures'	726
Peter Andorfer; Edeltraud Aspöck; Matej Ďurčo; Anja Masur; Ksenia Zaytseva	
Pundit. Semantic Annotation for Digital Humanities	728
Playing With Cultural Heritage Through Digital Gaming: The New Narrative of the ARK4 Project	730
ITAF: Rewiring the Italian 'Nation' of the Army of Flanders (1567-1714)	730
Documenting Material Culture: 3D Laser Scanning, Photogrammetry and Archaeological Ceramics from Lincoln Pottery Works	<i>7</i> 31
Extraction and Visualization of Toponyms in Diachronic Text Corpora	732
Named Entity Extraction from digitized texts of Mongolian Historical Documents in Traditional Mongolian Script	734
Interoperability: a new horizon for data sharing in Humanities and Social Sciences. The input of three digital services developed by Huma-Num	735
Seeing the Argument: Visualize Your Database with DAVILA	736
First We Feel Then We Fall – Multimedia Adaptation of Joyce's Finnegans Wake	738
Boutique Big Data: Reintegrating Close and Distant Reading of 19th-Century Newspapers $\rm M.\ H.\ Beals$	739
Scholarly Research Activities and Digital Tools: When NeMO met FLOSS	740
Reflecting On And Refracting User Needs Through Case Studies In The Light Of Europeana Research	742
The symogih.org Project: Towards an International Consortium	743
Crosslingual Textual Emigration Analysis	744

Mark Edward Boettcher; John Wallace	/43
Collaborative Annotation and Exploration of Literary Works in Learning Contexts	746
Verbal Identity of a Fictional Character: a Quantitative Study with a Machine Learning Experiment	747
Mobile Makerspaces: Te(a)chnology, Design and Digital Humanities	749
Studying Linguistic Changes on 200 Years of Newspapers	<i>7</i> 51
Beyond Shot Lengths – Using Language Data and Color Information as Additional Parameters for Quantitative Movie Analysis	753
From Ashes to Ashé: Digital Disaster Archives as Memorials	755
Poetry in Prose: automatic identification of verses in brazilian literature	756
Towards an XML Corpora Exposition as LOD with the Lightweight Xquery-Based Framework SynopsX	757
Enhancing Close Reading	758
Modelling between Digital and Humanities: Thinking in Practice	762
Using Digital Editions to Analyze Iliadic Text Reuse and its Poetic Tradition	763
Editing in a text-image-sound form: the eTalks	764
Discursive Constructions Of Culture: A Semantic Model For Historical Travel Guides Ulrike Czeitschner	766
Visualizing the Gradual Production of a Text	767
Bridging the Gap: the Digital Innovation Group	769

The Attribution of the Lazarillo de Tormes. Shedding Some Light into a Centuries Old Problem . Javier de la Rosa Pérez	770
Defining the Core Entities of an Environment for Textual Processing in Literary Computing Angelo Mario Del Grosso; Davide Albanesi; Emiliano Giovannetti; Simone Marchi	<i>77</i> 1
EVI-LINHD. A Virtual Research Environment for the Spanish-speaking Community	776
Diachronic Semantic Lexicon of Dutch (Diachroon semantisch lexicon van de Nederlandse Taal; DiaMaNT)	777
A Web-based Tool Called Gauntlet: From Iterative Design To Interactive Drawings Annotation . Gregory Dessart; Martial Sankar; Anastasia Chasapi; Guido Bologna; Zhargalma Dandarova Robert; Pierre-Yves Brandt	778
EVT 2.0: a new architecture for critical editions in digital form	780
Testing Delta on Chinese Texts	781
The Categories of Philosophy in the Digital Era	783
"Remapping Leigh Hunt's Circles": Voyant Tools and Hunt's Dramatic Criticism	785
The Digital Émigré: Russian Periodical Studies and DH in the Slavic Fields	786
Faraway, So Close!: Reading Adeline Mowbray Closely Using Topic Modelling Michael Gregory Falk	787
The Corpus of Revenge Tragedy (CoRT): Toward Interdisciplining Early Modern Digital Humanities and Genre Analysis	789
3D Scanning for Preservation: Difficulties and Dissemination	791
Linguistic DNA: Modelling Concepts and Semantic Change in English, 1500-1800 Susan Fitzmaurice; Marc Alexander; Michael Pidd; Justyna Robinson; Fraser Dallachy; Iona Hine; Seth Mehl; Brian Aitken; Matthew Groves; Katherine Rogers	792
The digital breadcrumb trail of Brothers Grimm	793
The Southern Netherlands and the Infrastructure of Early Modern Globalisation (1500-1800) . Ulrike Valeria Fuss; Christian Pistor; Werner Thomas; Cesar Esponda de la Campa; Lieve Behiels; Cesar Manrique Figueroa	795

A DH-Leavened Musicological Toolbox	796
The Garden: A 3D Adventure Puzzle Game Exploring Bosch's Garden of Earthly Delights Elizabeth Goins; Christopher Egert; Andrew Phelps	797
Authorial {X}: A Research and Teaching Platform for Literary Geography	798
Documenting the pain: Sharing Second World War survivors' stories to help meaning making and lessons learning through curating trans-European digital narrative trajectories Siegfried Handschuh; Simon Donig; Adamantios Koumpis; Hanna Diamond	800
The Ibn Darraj Project: SpatioTemporal Reasoning Engine Based on Evidence Combination Mohammad Hossein Haqiqat khah; Babak Nadjar Araabi	801
The Kuzushiji Project: Developing a Mobile Learning Application for Reading Early Modern Japanese Books	803
Blacks In American Medicine Archive: Exploring Forgotten Stories	805
Abbreviations In Manuscripts – Systematization And Crowdsourcing By Ad Fontes Tobias Hodel	806
Annotating and Georeferencing of Digitized Early Maps	807
The »Hidden Kosmos« of scientific instruments in Alexander von Humboldt's »Vorlesungen über physikalische Geographie« – or: What on earth is an »Ariometer«?	808
Data-First Digital Humanities: How Adopting a Data-First Strategy Fosters Research, Collaboration, Pedagogy, and Scholarly Communication in the Digital Humanities	811
Rebuilding Digital Harlem for Sustainability and Change	812
Digital Resources and Research Data in the Digital Humanities: The Digital Knowledge Store at the Berlin-Brandenburg Academy of Sciences and Humanities	813
Visual Forms of Information Presentation and Their Place in Formal Digital Scientific Communication	815
The Royal Society Corpus: Towards a high-quality corpus for studying diachronic variation in scientific writing	816
On the Distant Reading of Musicians' Biographies	818

Pedagogy and Scholarship	820
Moving Images and the Connection to other Media Types	821
Identifying the Same Ukiyo-e Prints from Databases in Dutch and Japanese	822
1 Million Dutch Newspaper Images available for researchers: The KBK-1M Dataset Martijn Kleppe; Elliott Desmond	824
Automatic Emotion Detection for Quantitative Literary Studies: A case study based on Franz Kafka's "Das Schloss" und "Amerika"	826
Creating A Management Plan For A Cultural Heritage Project – Best Practise	829
Geography Of Russian Poetry: Countries And Cities Inside The Poetic World Elizaveta Kuzmenko; Boris Orekhov	830
The Holy and the Godless – Cultural Stereotypes Featured in the Language of the Polish Medieval Hagiography. A Corpus-based Study	832
Taking VIVO into the Past: Adapting the VIVO Researcher Profile System to Historical Persons . The a Lindquist; Alex Viggio	833
Digital Humanities Pedagogy as Digital Liberal Arts: A Framework for Curriculum Development Brandon Locke	834
Conjuring up the Artist from the Archives: Ivar Arosenius. Digitization and Coordination of Archives for Enhanced Accessibility and Research	835
The Latin of Matthew of Cracow (c. 1345–1410) – a corpus based study of his language and style	836
Edit Histories and Literary Turf Wars: John Ashbery, Academic Criticism and Wikipedia $ \text{Jim McGrath} $	838
Free FannyPacks: A Model for the Easy Digital Publication of Archival Periodical Material $ \text{Kevin McMullen} $	838
Notoriously Toxic: The Language and Cost of Hate in the Chat Systems of Online Games Ben Miller; Antal van den Bosch; Cameron Kunzelman; Jennifer Olive; Wessel Stoop; Kishonna Gray; Cindy Berger; Shiraj Pokharel	840
El Uso de las Tecnologías Digitales para la Difusión del Patrimonio Cultural en México Ernesto Miranda Trigueros; Ernesto Priani Saisó; Isabel Galina Russell	842

Challenges in Setting up a Digital Humanities Centre in Romania	844
Collecting Judgments on Artworks Through a Similarity Game	846
A Workflow for Encoding and Publishing Inscriptions	847
UpCASE - A Web Application for Building and Maintaining Language Resources Claes Neuefeind; Francisco Mondaca; Mihail Atanassov	848
Medialatinitas.eu. Towards Meaningful Integration and Retrieval of Resources for Medieval Latin	850
Time Series Analysis Enhances Authorship Attribution	852
Adding Flexibility to Large-Scale Text Visualization with HathiTrust+Bookworm Peter Organisciak; Sayan Bhattacharyya; Loretta Auvil; Leena Unnikrishnan; Benjamin Schmidt; Muhammad Saad Shamim; Robert McDonald; J. Stephen Downie; Erez Lieberman Aiden	854
Cultivating Digital Humanities Biomes: A Collaborative Model	856
Curating Community: Building A Communications Strategy For The European Association For Digital Humanities	857
Integrative 3D Recording Methods of Historic Architecture: Burg Hohenecken from Southwest Germany	858
WebSty – an Open Web-based System for Exploring Stylometric Structures in Document Collections	859
A new approach to libraries in the Digital Humanities: the case of "Fonte-Gaia" Elena Pierazzo; Filippo Fonio; Claire Mouraby	862
Agréger le passé en ligne: Euchronie, le passé ici et maintenant!	864
Commens Digital Companion to Charles S. Peirce	865
Excerpta Constantiniana: From Palimpsest to a Digital Edition of a Medieval Encyclopaedia Dariya Rafiyenko	866
Mapping Imagined and Experienced Places: An Exploration of the Geography of Willa Cather's Writing	868

CTRaCE: Canonical Text Reader and Citation Exporter	869
A Tool for NLP-Preprocessing in Literary Text Analysis	871
Mapping Languages Performance by Performance	873
Sussex Humanities Lab – Emotion, Automation and Sonic Socialities	874
Textual Communities	876
Multivalent reuse of web data about temporary art exhibitions: the Exhibitium project Nuria Rodríguez-Ortega; José Pino Díaz; Juan Luis Suárez; Rafael Bailón Moreno	877
Critical Edition as Graph: The Chronicle of Matthew of Edessa Online	879
Development of a Support Tool for Categorizing Ukiyo-e's Pictorial Themes: A System to Deal with Visual Features and Similarities	880
eComparatio - Editionsvergleich	882
Authorship Attribution of Mediaeval German Text: Style and Contents in Apollonius von Tyrland	883
You Better Monetize! Monetization Strategies in Publishing and Disseminating Digital Scholarly Editions	885
Corpus of Ioannes Dantiscus' Texts and Correspondence dantiscus.al.uw.edu.pl	886
Linking Graph with Map for the Purpose of Historical Research	887
New Facets Of The Multimedia Annotation Tool ELAN	888
Research Data in the Humanities: Present, Preserve and Share	889
The Latin Batrachomyomachia Collection	890

Complex Nerworks-Based Approach to Russian Rhyme History Description: Linguostatistics and Database	891
Cantus Network – a Semantically Enriched Digital Edition of Libri Ordinarii of the Salzburg Metropolitan Province	893
EDM in Use: Collecting Metadata for a Regional Cultural Heritage Portal	894
Making George Washington's Financial Documents Accessible: Transcription, Data, and the Drupal Solution	895
The Models of Authority Project: Extending the DigiPal Framework for Script and Decoration Peter Anthony Stokes; Stewart J. Brookes; Geoffroy Noël; John Reuben Davies; Tessa Webber; Dauvit Broun; Alice Taylor; Joanna Tucker	896
Diachronic changes of the Russian Presidential Addresses to the Federal Assembly: From the perspective of archetype key words	899
Ruthenian Metrica: technological aspects of the electronic publication	900
Regional Digital Humanities Consortia: An Emerging Formalization of Informal Network Ties? . John Christopher Theibault	902
Photogrammar	903
Automatic quotation detection in Russian nonfiction texts	904
Project COLEM for CREATE (University of Amsterdam)Adapting NPL-Tools for Creating an Orthographic Layer for Early Modern Dutch Texts	906
Reconstructing Past Teaching	907
Beyond Digital Humanities? Furthering the Exploration of Language Diversity and Pan-European Culture by Means of Transdisciplinary Research Infrastructures: Introducing the new DARIAH CC Science Gateway	908
The GeoHumanities Special Interest Group: Fostering and facilitating the geospatial turn Katherine Hart Weimer; Karl Grossner	911
DH Bridge: Teaching Computational Thinking in the Humanities	912

Insightful Island	913
Mapping Kipnis	915
Towards a Cross-generation Social Network for Jewish Sages	916
Pre-conference workshops	
RDA/ADHO Workshop: Evaluating Research Data Infrastructure Components and Engaging in their Development	921
Bridget May Almas; Kim Fortun; Natalie Harrower; Eveline Wandl-Vogt	
A Place for Places: Current Trends and Challenges in the Development and Use of Geo-Historical Gazetteers	923
GAMS and Cirilo: research data preservation and presentation	925
Martina Bürgermeister; Zsófia Fellegi; Gábor Palkó; Gerlinde Schneider; Martina Scholger; Elisabeth Steiner; Gunter Vasold	
Biographical Data Workshop: modeling, sharing and analyzing people's lives	927
Digital Literary Stylistics Workshop	928
A Demonstration of Multispectral Imaging	929
Visual Network Analysis with Gephi WorkshopCollective Interpretation of DH Communities Through Twitter Networks	930
Web Communities Mapping With HypheHow to use a curation-oriented web crawler to map communities?	931
Mathieu Jacomy; Paul Girard; Benjamin Ooghe-Tabanou	
Data mining digital libraries	932
Audiovisual Data And Digital Scholarship: Towards Multimodal Literacy	934
Big Data: Complex Systems and Text Analysis	935

TEI Processing Model Toolbox: Power To The Editor	936
From Digitization to Knowledge: Resources and Methods for Semantic Processing of Digital Works/Texts	936
Mining Texts with the Extracted Features Dataset	938
Reflectance Transformation Imaging (RTI) for Cultural Heritage Artefacts	939
Building Capacity with Care: Graduate Students and DH work in the Library	940
Translation Hack-a-thon!: Applying the Translation Toolkit to a Global dh+lib	941
Minimal Computing: A Workshop	943
Working with WissKI - A Virtual Research Environment for Object Documentation and Object-Based Research	944
Digital Archiving and Storytelling in the Classroom with Omeka and CurateScape Victoria Szabo; Hannah Jacobs; Edward Triplett	945
View Source: Reading the Hidden Texts of the Web	947
Introduction to Natural Language Processing	947
Music Information Retrieval Algorithms for Oral History Collections	948

Animate, Innovate, Lyon, France: https://halshs.archives-ouvertes.fr/halshs-01251915v1.

Cellier, J. and Cocaud, M. (2001). *Traiter des données historiques: méthodes statistiques, techniques informatiques.* Rennes: Presses universitaires de Rennes.

Eide, Ø. (2014-2015). Ontologies, Data Modeling, and TEI. *Journal of the Text Encoding Initiative*, **8:** Selected Papers from the 2013 TEI Conference. http://jtei.revues.org/1191.

Erickson, A. T. (2013). Historical Research and the Problem of Categories. In: Dougherty, J. and Nawrotzki, K. (eds), *Writing History in the Digital Age.* Ann Arbor: University of Michigan Press, pp. 133-145.

Gast, H., Leugers, A. and Leugers-Scherzberg, A. H. (2010). Optimierung historischer Forschung durch Datenbanken. Die exemplarische Datenbank "Missionsschulen 1887-1940". Bad Heilbrunn: Verlag Julius Klinkhardt.

Jordanous, A., Stanley, A. and Tupman, C. (2012). Contemporary transformation of ancient documents for recording and retrieving maximum information: when one form of markup is not enough. Proceedings of Balisage, 8: The Markup Conference 2012, http://www.balisage.net/Proceedings/vol8/html/Jordanouso1/BalisageVol8-Jordanouso1.html.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. (2003). *WonderWeb Deliverable D18 Ontology Library (final)*. Trento: Laboratory For Applied Ontology. PDF version: http://wonderweb.man.ac.uk/deliverables.shtml .

Notes

- http://phn-wiki.ish-lyon.cnrs.fr/doku.php?id=td_histoire_ numerique:accueil(all websites were accessed on 30 October 2015).
- Documentation about the project and links to its the different components are provided on the project main website: http:// symogih.org/.
- ³ http://xml-portal.symogih.org/web-publications.html .
- ⁴ Cf. https://en.wikipedia.org/wiki/Generic_data_model. Databases are used in historical research both at individual and project level. See, e.g., Gast, Leugers and Leugers-Scherzberg, 2010; Alerini and Lamassé, 2011; Cellier and Cocaud, 2001; Erickson, 2013. The novelty of the method proposed by the symogih.org project, and in my course, is the abstraction level allowing to treat any kind of historical information.
- DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering: http://www.loa.istc.cnr.it/old/DOLCE.html.
- ⁶ CIDOC Conceptual Reference Model (CRM): http://www.cidoc-crm.org/.
- ⁷ http://www.postgresql.org/docs/9.4/static/plpgsql.html.
- 8 http://postgis.net/.
- 9 http://symogih.org/?q=rdf-publication .
- ¹⁰ Cf. Masolo et al., 2003: 14.
- http://phn-wiki.ish-lyon.cnrs.fr/doku.php?id=td_histoire_ numerique:installation_db_2013.
- ¹² Some pages of the symogih.org project's user manual provide the encoding specification for XML/TEI semantic annotated texts using the symogih.org ontology: https://groupes.renater. fr/wiki/symogih/symogih_manuel/edition_de_textes_en_ xml-tei. This method was presented at the TEI 2015 conference in Lyon, cf. Beretta 2015. The Special Interest Group Ontologies in the TEI Consortium is devoted to this approach. See

the GIS Ontologies wiki: http://wiki.tei-c.org/index.php/SIG:Ontologies and Eide, 2014-2015. A similar approach is represented by Jordanous, Stanley and Tupman, 2012.

http://phn-wiki.ish-lyon.cnrs.fr/doku.php?id=td_histoire_ numerique:exercice_2.

Modelling Taxonomies of Text Reuse in the Deipnosophists of Athenaeus of Naucratis: Declarative Digital Scholarship

Monica Berti monica.berti@uni-leipzig.de University of Leipzig, Germany

Mary Daniels ellie.daniel@furman.edu Furman University, USA

Samantha Strickland sami.strickland2@furman.edu Furman University, USA

Kimbell Vincent-Dobbins kimbell.vincent-dobbins@furman.edu Furman University, USA

This paper presents work on documenting text reuse of fragmentary authors and of extant works. By **fragmentary** we mean authors whose texts are lost and known through quotations and references by other authors. Within ancient Greek literature 60% of authors is preserved only in fragments, showing the challenge of working with innumerable pieces of reuse scattered in our textual heritage (Berti et al., 2009). This work is necessarily prior to any specific research questions. We cannot inquire into, e.g., the historical works of Istrus the Callimachean until we can comprehensively and precisely catalogue the surviving fragments of Istrus; nor can we ask "how did intellectuals in the 3rd century CE read epic poetry?", until we can comprehensively identify instances of Homeric text reuse and work with them in their context.

The term **fragment** is the result of print editorial practices, where chunks of text preserving traces of lost authors and works are extracted from their contexts and reprinted in separate collections. Even if such editorial workflow has produced invaluable results for reconstructing lost authors, the concept of **textual fragment** is problematic: It includes different kinds of text reuse and implies a certain degree of originality, which is difficult to assess and represent because the original text from which the reuse derives

is hidden by the **cover text**, i.e., by the intention of the quoting author and the characteristics of the preserving context (Most, 1997; Schepens, 2000; Berti, 2013).

Our data model defines taxonomies of text reuse for representing references to authors and works not as separate chunks of text but as contextualized annotations, expressing their nature of reuse of textual evidence. These annotations include not only the portion of text classifiable as a reuse, but also biographical and bibliographical data preserved in the source text.

Text reuse of fragmentary authors presents the challenge of documenting text aligned with no extant exemplar. Text reuse of extant works presents additional challenges of aligning as precisely as possible (but no more precisely than is possible) two or more extant passages of text that may differ in small ways or large. Our data model documents uniquely instances of text reuse and it is developed on the Canonical Text Services (CTS), which is a protocol for identifying and retrieving passages of text based on concise, machine-actionable canonical citation. It is founded on the assumption that a "text" can be modelled as "an ordered hierarchy of citation objects" (Smith and Weaver, 2009). CTS URNs can identify passages more grossly or more finely; they can identify a range of passages at various levels of specificity; by the addition of an indexed substring, a CTS URN can identify a particular string within a passage of text (Blackwell and Smith, 2012). CTS is one component of a larger digital library architecture, developed for the Homer Multitext project and called CITE (Collections, Indices, Texts, and Extensions): http:// www.homermultitext.org/hmt-doc/cite/.

In order to produce citable analyses of text reuse in their context, we have been working with the *Deipnosophists* of Athenaeus of Naucratis, which is the account of a banquet where learned men quote authors and works of Greek literature concerning a wide range of topics related to dining and food. The *Deipnosophists* is significant because it is a very rich collection of many different kinds of text reuse of fragmentary authors and of extant works (Braund and Wilkins, 2000; Lenfant, 2007; Jacob, 2013).

Our data model specifies four subjects of analyses:

- 1. **Authors**: enumerate and identify authors reused by Athenaeus;
- 2. **Works**: enumerate and identify works reused by Athenaeus;
- 3. **Mentions**: catalog every mention of authors and works in the text of Athenaeus, including his vocabulary for identifying them. For example, Athenaeus may mention that a work by Archestratus of Syracuse was known by four different names (i.e., *Gastronomy*, *Life of Pleasure*, *Science of Dining*, or *Art of Cooking*); this would generate five entries in this list: one mention of Archestratus, and four mentions of the same work.
- 4. **Reuses**: uniquely identify instances of text-reuse in the text of Athenaeus.

A fifth analysis will also include the twenty-two learned men who take part in the banquet described by Athenaeus and who are actually the **characters** who quote and reuse a huge amount of authors and works.

We need seven records to produce citable analyses of the above mentioned subjects:

- 1. **Analysis Record URN**. Every documented instance of text reuse (authors, works, mentions, reuses) has a CITE URN uniquely identifying this instance in a CITE collection
- 2. **Sequence Number**. The collection of instances of text reuse is an **ordered collection**; each item has a sequence number, reflecting the item's sequence in the text of Athenaeus. This value is programmatically generated by a CTS-aware script before publishing the collection.
- 3. **Analyzed Text**. A CTS URN defining, as precisely or imprecisely as necessary, the span of text in the *Deipnosophists* that is the subject of this analysis of text reuse. The scope of the **Analyzed Text** is determined by the nature of the text reuse. In the case of authors and works, this CTS URN identifies a passage in the *Deipnosophists* that serves to justify the inclusion in the respective list. When an author or a work is reused often, the passage should be a clear, unambiguous reference (e.g., "Homer says ...").
- 4. Reused Text. While the Analyzed Text identifies a coherent and contiguous span of text, as it appears in the edition being analyzed, the Reused Text is a string identifying only the text being reused. The Analyzed Text provides context and a basis for alignment, while the Reused Text gives us the flexibility to call out non-contiguous text, to normalize text, or even to promote morphological forms determined by indirected statement to those appropriate for direct speech, without doing violence to our source-edition. The Reused Text record allows us to represent different intepretations of the same text reuse, especially in the case of non-verbatim quotations.
- 5. **Alignment URN**. This collection documents reuse of extant authors and works, for which we have extant editions with canonical citation. The **Alignment URN** is a CTS URN pointing to the quoted extant author (identified with a CtsGroupUrn) or to one specific edition of the reused work (identified with a CtsWorkUrn) that (a) justifies our claim of text reuse, and (b) is the basis for attaching a citation of a still extant work to this analysis.
- 6. Analytical Edition URN. The collected instances of text reuse of extant work in the *Deipnosophists* represent a new edition of these works, whose text-content is based on our analysis of our project's edition of Athenaeus. The Analytical Edition URN is a CTS URN to an Athenaeus Edition of these works; the citation-value is based on that of the Alignment URN; the text-content of this edition is the Reused Text in Athenaeus. The Analytical Edition gives us an orthogonal view of the text reuse of extant authors in Athenaeus.

7. CITE Collection of Lost Works. For text reuse of lost authors and works, there is no citation scheme, nor any inherent order to the text. For these, we produce a collection of text-reuse. This Collection can be cited by CITE URNs.

Initial work on documenting text reuse has been focused on references to Homer's *Iliad* in the *Deipnosophists* (data available at http://digitalathenaeus.github.io/). The aim is to extend our data model including the categorization of different kinds of text reuse and further concrete examples of references to fragmentary authors and extant works in the *Deipnosophists* of Athenaeus of Naucratis.

Bibliography

- **Berti, M.** (2013). Collecting Quotations by Topic: Degrees of Preservation and Transtextual Relations among Genres. *Ancient Society* 43: 269–88. doi:10.1145/1555400.1555442.
- Berti, M., Romanello, M., Babeu, A. and Crane, G. (2009). Collecting Fragmentary Authors in a Digital library. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York: ACM, pp. 259–62. doi:10.1145/1555400.1555442.
- Blackwell, C. W. and Smith, D. N. (2012). Four URLs, Limitless Apps: Separation of Concerns in the Homer Multitext Architecture. In Muellner, L. (ed.), Donum Natalicium Digitaliter Confectum Gregorio Nagy Septuagenario a Discipulis Collegis Familiaribus Oblatum. Washington, DC: The Center for Hellenic Studies. http://chs.harvard.edu/wa/pageR?tn=ArticleWrapper&bdc=12&mn=4846.
- **Braund, D. and Wilkins, J.** (2000). *Athenaeus and His World. Reading Greek Culture in the Roman Empire.* Exeter: University of Exeter Press.
- **Jacob, C.** (2013). *The Web of Athenaeus*. Center for Hellenic Studies: Harvard University Press.
- Lenfant, D. (2007). Athénée et les fragments d'historiens. Actes Du Colloque de Strasbourg (16-18 Juin 2005). Paris: De Boccard.
- **Most, G. W.** (1997). *Collecting Fragments. Fragmente sammeln.* Göttingen: Vandenhoeck & Ruprecht.
- **Schepens, G.** (2000). Probleme der Fragmentedition. (Fragmente der Griechischen Historiker). In Reitz, C. (ed.), *Vom Text Zum Buch*. St. Katharinen, pp. 1-29.
- Smith, D. N. and Weaver, G. (2009). Applying Domain Knowledge from Structured Citation Formats to Text and Data Mining: Examples Using the CITE Architecture. *Text Mining Services* 129. http://katahdin.cs.dartmouth.edu/reports/TR2009-649.pdf

Two Centuries of Russian Roads – Diachronic Study of Polysemy in the Context of Cultural Change

Anastasia Bonch-Osmolovskaya abonch@gmail.com National Research University Higher School of Economics Moscow, Russian Federation

Preamble

The goal of the study is to show links between lexical and social diachronic change. The study is conducted in the culturomics framework (Michel et al., 2011). In contrast to the Big data approach the study promotes the idea of medium data, i.e. amount of data which allows both to make quantitative and qualitative analysis (Bonch-Osmolovskaya, 2015). The main characteristics of the medium data are:

- The reliability of sources, which metadata can be filtered manually
- The sufficiency of the data amount for reliable statistical measures
 - The possibility of additional semantic mark-up

The research is based on the data from Russian National Corpus (ruscorpora.ru) (see Plungian, Sitchinava, 2003). The study pursues changes of context frequencies for the lexeme road in the period from 1800 till 2000, and correlates the observations with social and economic progress as well as change in conceptual language space

Choice of concept

Russia is a big country, so transportation has been traditionally a critical problem. The choice of the word *road* for culturomics study is based on our expectations of the concept's centrality for the economy, society and culture in Russia of the 19th–20th centuries. *Road* appears to be a productive sign in terms of semiotics of art (Tchepanskaya, 2003,), that's why I expected to collect numerous relevant contexts both in fiction and nonfiction. At the same time road in Russian has several meanings, the nature of its polysemy has been treated a lot in previous works (Arutiunova, 1999). We can distinguish three basic meanings which are contrasted by the position of Observer (Paducheva, 2006) - the one that percepts the *road*. The first meaning is *road* as a physical object, a line on the ground the observer sees while standing on it. It can be characterized by the quality of its surface or surrounding landscapes (i.e. dirty road). The second meaning is *road* as a vector, a line on a map, that connects two points (i.e central road). The observer operates in this case with the abstract idea of the road's topology. The third meaning is metonymical and it stands for the travel-event the Observer experiences while moving