

Papers

| authors | title | keywords | topics | abstract |
|--|--|--|---|---|
| Scipione, Gabriella; Guidazzoli, Antonella; Imboden, Silvano; Trotta, Giuseppe; Montanari, Margherita; Liguori, Maria Chiara; Caraceni, Simona | I-Media-Cities: A Digital Ecosystem Enriching A Searchable Treasure Trove Of Audio Visual Assets | Multimedia and multimodal retrieval, Automatic tagging, Search interfaces, User interface management systems, Media arts | film and performing arts studies metadata crowdsourcing digital research infrastructures and virtual research environments linking and annotation English computer science and informatics | Cultural heritage and museum institutions across Europe are progressively holding vast digital collections – either digitized or born digital – that can significantly affect many research fields, both in Social Sciences and Humanities. The I-Media-Cities platform is an interactive ecosystem aiming to share, provide access to and use 9 European film archives AV digital contents to allow multidisciplinary research by means of manual and automatic annotations. This contribution describes the approach and the actions taken in order to perform an effective design and implementation for the visual interfaces required by the project in order to foster the User Experiences for different audiences, such as researchers and citizens. |
| Guidazzoli, Antonella; Bellavia, Giovanni; De Luca, Daniele; Delli Ponti, Francesca; Farroni, Federica; Liguori, Maria Chiara; Chiavarini, Beatrice; Imboden, Silvano | More Than Just CG: Storytelling And Mixed Techniques In An Educational Short Movie | Media arts, Education, Computer Graphics, Animation, 3D modelling | audio, video, multimedia history and historiography virtual and augmented reality GLAM: galleries, libraries, archives, museums modeling, simulation, 3D/4D modeling English computer science and informatics | <u>_Il piccolo Masaccio e le Terre Nuove_</u> is a short animated Computer Graphics educational movie explaining the origins and history of San Giovanni Valdarno, a city of foundation. Set in 1409, its main characters are a young Masaccio - the painter - and the Vicar, city governor of the time, who explains to Masaccio and to the public the birth and creation of the city, based on the rational rules that have informed its development. The video has been developed with manifold techniques: procedural modelling and real watercolours; photogrammetry; crowd, cloth and particle simulations and a varied vegetation library without overshadowing storytelling and characters credibility. The short movie aims primarily at the younger audience of the museum of San Giovanni Valdarno, keeping, at the same time, a high level of scientific and historical accuracy. An effort able to produce a wealth of reusable 3D assets for a future crossmedia exploitation. |

| authors | title | keywords | topics | abstract |
|---|--|---|---|---|
| Chavez Heras, Daniel Alberto (1); Blanke, Tobias (1); Cowlshaw, Tim (2); Fiala, Jakub (2); Herranz Donnan, Amaya (2); Man, David (2) | Seen by Machine: Computational Spectatorship in the BBC television archive | computational spectatorship, machine learning, television, computational creativity, cultural analytics | audio, video, multimedia film and performing arts studies cultural artifacts digitisation - theory, methods and technologies English computer science and informatics artificial intelligence and machine learning cultural analytics | This paper is a report and critical examination of research undertaken in the project <i>_Made by Machine: When AI Met the Archive_</i> . In this project we used three distinct computational approaches to analyse and automatically browse the BBC television archive; we then proposed a novel way of combining these approaches through machine learning by fitting their outputs to a recurrent neural network as a time-series. First we frame <i>_Made by Machine_</i> as cultural big data project. We then describe the technological approaches we used to traverse archive space, learn and extract features from video, and model their relations through time. And finally, we introduce the idea of <i>_computational spectatorship_</i> as a concept with which to understand the objects and practices of automated seeing/editing through machine learning. |
| Schweizer, Tobias Julius; Alassi, Sepideh; Mattmüller, Martin; Rosenthaler, Lukas; Harbrecht, Helmut | An Interactive, Multi-layer Edition Of Jacob Bernoulli's Scientific Notebook Meditationes As Part Of Bernoulli-Euler Online | digital edition, history of science, research platform | digital research infrastructures and virtual research environments linking and annotation English digital humanities (history, theory and methodology) history of science manuscripts description and representation | The <i>_Meditationes_</i> is the scientific notebook of the mathematician Jacob I Bernoulli (1654–1705), member of the Bernoulli dynasty from Basel, Switzerland. The notebook consists of 367 pages; more than 90% of the 287 entries deal with questions of mathematics and physics. Parts of the <i>_Meditationes_</i> have been edited in six volumes over the past decades according to rather varying standards. Our project will provide an integral edition of the manuscript for the first time, providing facsimiles, transcriptions (from diplomatic to normalized texts), translations, and comments. The edition is part of <i>_Bernoulli-Euler Online_</i> (BEOL), a platform for early modern mathematical texts. BEOL makes texts available to historians of science and will provide tools to work with the resources available on the platform. |

| authors | title | keywords | topics | abstract |
|------------------|--|--|---|---|
| Hoover, David L. | The Invisible Translator Revisited | translation, authorship attribution, computational stylistics | corpus and text analysis literary studies stylistics and stylometry authorship attribution / authority translation studies English | Examination of the extent to which the translator's "authorial" signal can be disentangled from that of the original author |
| Li, Weixuan | Visualizing Networks of Artistic Ideas in History Paintings in the Seventeenth-Century Netherlands | network of ideas, network analysis, paintings, iconography, Dutch Golden Age | art history and design studies spatial & spatio-temporal analysis, modeling and visualization network analysis and graphs theory English digital humanities (history, theory and methodology) | This research conceptualizes and visualizes art historical networks of ideas inscribed in iconography through Network Analysis. The research focuses on history paintings created in the Low Countries between 1575-1700, combining key digital resources on Netherlandish paintings. By placing the works of art at the center of the network analysis, the research bridges art historical research favoring both quantitative and qualitative approaches and revealing the changes in the structures of artistic interactions and the diffusion patterns of subject matters within the artist community in different cities. |

| authors | title | keywords | topics | abstract |
|--|--|--|---|--|
| Vitali-Rosati, Marcello | Toward a Theory of Editorialization | editorialization, digital space, knowledge production | philosophy scholarly editing digital textualities and hypertext English public humanities and community engaged scholarship digital humanities (history, theory and methodology) scholarly publishing, open content and open science | The notion of “editorialization” has a fundamental place within the francophone scientific community as a key-concept for understanding and interpreting the digital culture. The concept has been at the center of theoretical works of the academic community for the last ten years. It has been recently the subject of a book in English. In my communication I will present the result of ten years of work on this concept and explain how the theory of editorialization can help DH scholars to think about the architectural space of the digital environment, and how it impacts knowledge production, circulation and legitimation. More specifically, I will argue that this theory can be a very powerful theoretical framework to take into account the political implications of our practices as dh scholars. |
| Shepard, Dave (2); Hashimoto, Takako (1); Shin, Kilho (3); Uno, Takeaki (4); Kuboyama, Tetsuji (5) | The Narrow Scopes of Fake News: Detecting Fake News Using Topic Diversity Measures | Fake News, Social Media, Micro Clustering, Data Mining, Topic Extraction | content analysis social media data mining / text mining English computer science and informatics artificial intelligence and machine learning communication and media studies | This paper proposes a method for analyzing fake news topics on twitter using a micro-clustering algorithm. To form micro-clusters, we leverage a data polishing algorithm that one of the authors developed. Our method shows that micro-clustering can assist in differentiating fake news from real news: we observe that fake news tweets show low topic diversity. To show the effectiveness of our method, we present a case study of tweets targeting the rumor of the Cosmo Oil petrochemical complex explosion during the Great Japan East Earthquake in 2011. |

| authors | title | keywords | topics | abstract |
|--|--|--|--|---|
| Schumann, Anne-Kathrin (1); Beierle, Christoph (1); Blößner, Norbert (2) | A Finite-State Approach to Automatic Greek Hexameter Analysis | Greek, Homer, hexameter annotation, finite-state technology | corpus and text analysis data models and formal languages philology linking and annotation English computer science and informatics artificial intelligence and machine learning | This paper presents a fully automatic approach to the scansion of Ancient Greek hexameter verse. In particular, we describe how finite-state automata can be used to discriminate between the 32 variants of Ancient Greek hexameter. We evaluate the performance of our annotation algorithm against hand-annotated scansion data. The project code is available online. |
| Kil, Aleksandra Maria | Index Cards and the Analog Humanities. A Media Archaeology of Cultural Studies in Poland | index cards, knowledge apparatus, epistemic infrastructures, media archaeology, paper technologies | cultural studies English digital ecologies, digital communities and critical infrastructure studies media archaeology communication and media studies | This paper explores -- in a media archaeological vein -- a yet unexamined archive of paper index cards created and used by Stanisław Pietraszko, the pioneer of cultural studies in Poland, and propose a media-specific and infrastructure-oriented account of knowledge-making in the humanities, especially in the Polish postwar reflection on culture. In examining the artifact and its practices as revealed by interviews, narratives, and manuals, the following questions relating to technicality, instrumentality and creativity in the analog humanities come to light: How is knowledge crafted with a slip-box? What does this apparatus consist of? What is the genealogy of its format and standardization? Why were scholarly index cards resisted or criticized by some and, on the other hand, what made them so eagerly adopted by others? What kinds of humanistic work was recognized as valuable thinking, as opposed to mundane, auxiliary labor? Can index cards affect the style of thinking? |

| authors | title | keywords | topics | abstract |
|---|--|---|---|--|
| Stutzmann, Dominique (1); Currie, Jacob (1); Daille, Béatrice (2); Hazem, Amir (2); Kermorvant, Christopher (3) | Integrated DH. Rationale of the HORAE Research Project | Handwritten Text Recognition, text analysis and visualisation, Distant reading, Manuscript Studies, IIF | digital archives and digital libraries corpus and text analysis image processing medieval studies English digital humanities (history, theory and methodology) OCR and hand-written recognition | HORAE (Hours, Recognition, Analysis, Editions) is a cross-disciplinary research project studying the religious practices and experiences in the late Middle Ages and Renaissance as evidenced by the medieval bestseller, Books of Hours. Developing tools in artificial intelligence, computer vision and image analysis, in Natural Language Processing (NLP), and in manuscript studies, book history and religious practices, HORAE also implements diverse tools developed in the DH community (network analysis, data visualization, text mining). This paper presents HORAE as a research Gesamtkunstwerk that tackles a common challenge of complexity, uncertainty and granularity with different tools from different fields. By broadening the perspective in a genuinely cross-disciplinary research, we argue that DH is not a specific field of the humanities, but an integral (and to-be-integrated) part of current research methodologies. |
| Uslu, Tolga; Mehler, Alexander; Schulz, Clemens; Baumartz, Daniel | BigSense: a Word Sense Disambiguator for Big Data | word sense disambiguation, big data, wikipedia | natural language processing semantic analysis English computer science and informatics artificial intelligence and machine learning | We describe BigSense, a neural network based approach for efficient word sense disambiguation (WSD). BigSense uses the entire English Wikipedia disambiguation pages to train a model that achieves state-of-the-art results and is many times faster than its competitors. In this way, it is possible to disambiguate huge amounts of data using one of the largest disambiguation models freely available while maintaining high performance. Our approach paves the way for large-scale disambiguation in the text-based digital humanities. |

| authors | title | keywords | topics | abstract |
|--|---|--|--|--|
| Daquino, Marilena (1); Daga, Enrico (2); Tomasi, Francesca (1) | MAuth - Mining Authoritativeness In Art History | Information retrieval, art history, ontologies, authoritativeness, semantic crawling | authorship attribution / authority GLAM: galleries, libraries, archives, museums semantic web and linked data data mining / text mining English library & information science digital humanities (history, theory and methodology) | In this paper we present _mAuth_, a framework for harvesting information about paintings in the Web of data and ranking paintings attributions according to a documentary, evidence-based approach. In detail, we introduce (1) an ontology for representing questionable information related to attributionship, (2) an ontology-based ranking model resulted from art historical data analysis, and (3) mAuth, a proof-of-concept web application and API for consuming data about the history of attributions related to artworks of the Modern Age. |
| Bludau, Mark-Jan; Dörk, Marian; Heidmann, Frank | Relational Perspectives as Situated Visualizations of Art Collections | visualization, cultural collection, perspective, exploration, relational | digital archives and digital libraries art history and design studies interface, user experience design, gamification metadata GLAM: galleries, libraries, archives, museums English digital humanities (history, theory and methodology) | With relational perspectives we explore the potential of a new type of approach for the exploration of cultural collections. Cultural collections can contain thousands of artifacts, of which each typically possesses a diverse set of properties constituting a unique relationship to the rest of the collection. Therefore, to create an appropriate representation of the complex data of each underlying artifact, oftentimes it is not only interesting to get an overview about the entire collection from one perspective, but to explore the particular context and relations of individual items. To investigate the potential of relational perspectives we selected an art collection as a particularly promising case study. By following a collaborative and iterative design process with an art museum, we developed a web interface that contrasts a collection overview with three perspective-dependent views to examine the viability of this approach and to expose the diversity of each artifact. |

| authors | title | keywords | topics | abstract |
|--|--|---|---|--|
| Gooding, Paul Matthew (1); Terras, Melissa (2); Berube, Linda (3); Bennett, Mike (2); Hadden, Richard (2) | Subjectifying Library Users to the Macroscope Using Automatic Classification Matching | Classification, Library Datasets, User Studies | digital archives and digital libraries metadata GLAM: galleries, libraries, archives, museums interdisciplinary & community collaboration English library & information science digital humanities (history, theory and methodology) | Libraries are sources of large-scale data: both in terms of their collections and the information they collate on their spaces, users, and systems. These data provide opportunities to explore technical, methodological, and ethical questions from the valuable interdisciplinary perspective of Data Science and the Digital Humanities. In light of this, we will introduce our analysis of library datasets using Subjectify, an automatic classification matching tool developed to assist analysis of UK Non-Print Legal Deposit (NPLD) collections. We will address the following questions: what insights into users of digital library collections can be derived from automatic classification matching? What limitations are introduced by the use of existing classification schemes? And, in light of ongoing debates on responsible data curation in DH, how might DH and LIS scholars collaborate to inform ethical analysis of large-scale library datasets? |
| Kretzschmar, William; Stanley, Joey | Visualization of Big Data Phonetics | phonetics, Shiny, complexity science | digital archives and digital libraries databases & dbms spatial & spatio- temporal analysis, modeling and visualization linguistics modeling, simulation, 3D/4D modeling English computer science and informatics | We describe a new tool for visualization of our Big Data phonetic results, called the Gazetteer of Southern Vowels (GSV). The site was built in Shiny, a web application framework for R. GSV exploits the framework to supply traditional F1/F2 plots of our phonetic data, and also to supply point-pattern F1/F2 plots. A key feature of GSV is a range of user-selected display features, as applied to user-selected vowel types in specific environments, and used to display results from user-selected groups of speakers. GSV works with the c. 1.5 million vowel measurements. Our modifications of forced alignment and automatic formant extraction processes are described in XXXXX, but our methods are not the subject of this paper. Visualizations from GSV challenge traditional methods of understanding phonetic systems, and support complexity science as a description of human speech. |

| authors | title | keywords | topics | abstract |
|---|---|---|---|--|
| Spence, Paul Joseph; Brandao, Renata | Topographies of Digital Modern (Foreign) Languages research | 'Modern Foreign Languages', multilingual, linguistic diversity | multilingual / multicultural approaches spanish and spanish american studies English digital humanities (history, theory and methodology) | Languages (and their associated cultures) have been a key area of focus for the Digital Humanities from its origins as a field, and yet DH has rarely explored what this means in any real depth for its own research and pedagogical practices. Starting with a landscape study of topographies of Digital Modern (Foreign) Languages research, we examine the ways in which Modern Foreign Languages and DH interact, exploring how DH can achieve deeper linguistic engagement in its research ecologies and epistemic practices. |
| Kleymann, Rabea | Prototypes as Proto-Theory? A Plea for digital theory formation | theory, software prototype, research infrastructure, epistemology | literary studies digital research infrastructures and virtual research environments English digital ecologies, digital communities and critical infrastructure studies digital humanities (history, theory and methodology) | The presentation will show to what extent software prototypes can be understood as a "proto-theory" of digital humanities. Starting from the current theoretical debate in the Digital Humanities, the presentation will investigate the theory-building effects of software prototypes and discuss new digital forms of theory building. |

| authors | title | keywords | topics | abstract |
|--|---|---|---|---|
| Berti, Monica | Digital Fragmenta Historicorum Graecorum (DFHG) | text-reuse, fragmenta, historical corpora, philology, OCR | classical studies corpus and text analysis databases & dbms digital textualities and hypertext linguistics English scholarly publishing, open content and open science | The <i>Digital Fragmenta Historicorum Graecorum</i> (DFHG) is the digital version of the five volumes of the <i>Fragmenta Historicorum Graecorum</i> (FHG), which is the first big collection of ancient Greek historical fragments published by Karl Müller (1841-1873). The <i>FHG</i> is a corpus of quotations and text reuses (<i>fragmenta</i>) of 636 ancient Greek fragmentary historians preserved by Classical sources. The <i>DFHG</i> is not a new edition of ancient Greek fragmentary historians, but a new digital resource to provide textual, philological, and computational methods for representing fragmentary authors and works in a digital environment. This paper presents tools and services that have been developed by the project not only for accessing the entire collection of the <i>FHG</i> , but also for offering a new methodology based on digital and computational approaches to represent complex historical text reuse data. |
| Applegate, Matt; Evans, Sarah; Cohen, Jamie | Locating Absence with Narrative Digital Maps | Spatial Humanities, Locative Media, Absence, Humanities GIS, Politics | audio, video, multimedia spatial & spatio- temporal analysis, modeling and visualization authorship attribution / authority cultural studies English geography and geohumanities digital activism and hacker cultures | This paper outlines an interdisciplinary undergraduate digital humanities course, study abroad trip to Rome, Italy, and consultation with representatives from the non-profit organization Shoot 4 Change (http://www.shoot4change.eu/) focused on teaching students to visualize space critically. Utilizing the Spatial Humanities Kit (http://spatialhumanitieskit.org), deployed via both Molloy College and Hofstra University, we showcase narrative geospatial humanities work, media production, and a simple mix of HTML and GeoJSON as vehicles for our students' critical analysis. Our maps prioritize student experience, encapsulated in still images, written description, VR video, and vlogs embedded within them. Our course prioritizes methods for researching and unearthing embattled histories of public space, particularly within architecture, monuments, and urban design. Combined, our maps and critical framework result in a practice of teaching students to visualize cultural conflict that prefigures their experience of the space they inhabit--what is formative of, but currently absent or obscured from, the landscape they engage with. |

| authors | title | keywords | topics | abstract |
|--|--|--|---|---|
| Fischer, Frank (1); Börner, Ingo (2); Göbel, Mathias (3); Hechtl, Angelika (4); Kittel, Christopher (5); Milling, Carsten (1); Trilcke, Peer (6) | Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama | European drama, Research infrastructure, API, TEI | literary studies digital research infrastructures and virtual research environments English | This talk introduces DraCor, an infrastructure for the digital literary studies. It allows for multiple ways of access to European drama corpora encoded in TEI and is accessible at https://dracor.org/ or via GitHub repo or API. We understand DraCor as a *Programmable Corpus* – a new term that we would like to bring into play with this paper. |
| Ivanov, Lubomir | Comparing Assonance and Consonance for Authorship Attribution | Authorship attribution, consonance, assonance, machine learning, prosody | corpus and text analysis authorship attribution / authority English computer science and informatics digital humanities (history, theory and methodology) | This paper compares the performance of assonance and consonance as stylistic features for authorship attribution. We present results obtained from machine learning experiments using several corpora. |

| authors | title | keywords | topics | abstract |
|---|--|---|---|--|
| Odebrecht, Carolin | A Model-to-model Approach for Developing Corpus Metadata. An "Odd" TEI Customization for Encoding Metadata | Metadata, modeling, TEI, documentation, corpora | text encoding and markup languages metadata data models and formal languages information architecture and usability linguistics English library & information science | The TEI environment provides a generic document model for text encoding which provides modules for encoding text via mark up and modules for the metadata referring to the TEI document and the text represented within this document. The specialized ('odd') customization presented here follows an explicit metamodel for metadata representation and expands the range of applications of the TEI metadata modules. This metamodel can be defined as a a format-independent, concrete and extensive content model for metadata information. The application of the metamodel benefits from the TEI environment and its interoperability, and can make use of the modeling tool ODD and its validation mechanism. This approach enables use cases for metadata that were previously separated from the TEI universe and proves that the adaptability and flexibility of the TEI allows reuse scenarios which have not been initially intended |
| Manjavacas Arévalo, Enrique; Kestemont, Mike; Karsdorp, Folgert | A Robot's Street Credibility: Modeling authenticity judgments for artificially generated Hip-Hop lyrics | Neural Networks, Hip-Hop, Natural Language Generation | natural language processing data mining / text mining English computer science and informatics artificial intelligence and machine learning cognitive sciences and psychology digital art | This study aims to advance and enhance our understanding of the properties that contribute to the perceived authenticity of a specific art form: Hip-Hop lyrics. The basis of our study is an experiment carried out in the context of a large, mainstream contemporary music festival. We crowdsourced a large dataset of authenticity judgements for both authentic and neurally generated Hip-Hop lyrics, which enable us to quantitatively assess human biases toward artificially generated text as well as which linguistic characteristics are perceived as authenticity cues. Additionally, the dataset provides solid ground for evaluating different neural language generation systems with respect to their perceived credibility. We compare a variety of language models and techniques. Our experiments contribute equally to improving the credibility of generated text and enhancing our understanding of the cognitive processes at play in the perception of authentic and artificial art. |

| authors | title | keywords | topics | abstract |
|--|--|---|--|--|
| Roe, Glenn H (1); Gladstone, Clovis (2); Olsen, Mark (2); Morrissey, Robert (2) | Enlightenment Legacies: Sequence Alignment and Text-Reuse at Scale | Intertextuality, Sequence Alignment, Enlightenment, Big Data | corpus and text analysis literary studies french studies linking and annotation data mining / text mining English cultural analytics | This paper aims to examine the complexities of Enlightenment legacies in 19th-century literary and intellectual culture using new 'distant reading' approaches. In particular, we will describe the development and deployment of TextPAIR, a new sequence alignment system that detects reused passages in very large corpora. We use this software to compare several important collections of 18th-century texts to the Très Grande Bibliothèque (TGB) corpus of 19th-century printed materials made available by the Bibliothèque Nationale de France (BNF). While our interpretive work on this set of reuses is still in its initial phases, we have already been able to identify significant findings that challenge our understanding of the impact of the Lumières in the 19th century. Our full paper will expand upon these observations and begin the systematic exposition of the various complexities of identifying text reuse at such an unprecedented scale. |
| Liu, Chao-Lin | Onto Word Segmentation of the Complete Tang Poems | classical Chinese word segmentation, point-wise mutual information, Chinese poetry, Chinese culture | corpus and text analysis literary studies natural language processing content analysis English oriental and asian studies | We aim at segmenting words in the _Complete Tang Poems_ (CTP). Although it is possible to do some research about CTP without doing full-scale word segmentation, we must move forward to word-level analysis of CTP for advanced research topics. At this moment, we have collected 2433 poems that are segmented by domain experts, and use the segmented poems to evaluate the segmenter that considers domain knowledge of Chinese poetry. We trained probabilistic scores based on the CTP poems (excluding the 2433 poems, which were used exclusively only for testing) and the domain knowledge. The segmenter recovered 85.7% of words in the test poems. |
| Odijk, Jan | Faceted Search for Discovering Software | Metadata for software, faceted search | corpus and text analysis history and historiography metadata natural language processing speech processing linguistics English | We present a proposal for faceted search in metadata for software, which is based on a CLARIN Component Metadata Infrastructure (CMDI) profile for the description of software that enables discovery of the software and formal documentation of aspects of the software. We have tested the profile and the faceted search based on this profile by making metadata for over 80 pieces of software, and by creating an implementation of the faceted search for these metadata and some 280 more based on a different profile. It is available on the web on a URL that we will reveal after the paper has been accepted |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Benatti, Francesca (1); Gooding, Paul (2); Sillence, Matthew (3) | Developing a Community of Practice: The CHASE Model for Digital Humanities Researcher Training | pedagogy, methodology, training, collaboration, skills | teaching, pedagogy, and curriculum English digital humanities (history, theory and methodology) | This paper discusses the challenges of developing and delivering digital humanities training for graduate researchers. It presents a model of multi-institution, collaborative, methods-based training based on a four-year study of the Consortium of the Humanities and Arts South-East England Arts and Humanities in the Digital Age (CHASE AHDA) training programme for doctoral students. The paper debates the pedagogical approach of CHASE AHDA, which views DH as a community of practice based on methodological commons, and analyses the outcomes of the programme in terms of self-reflective learning, academic socialization and academic literacies. The paper will conclude by presenting the process of iterative course development based on student feedback and evaluate the results of the revised 2018/19 CHASE AHDA syllabus. |
| Bermeitinger, Bernhard (1); Gassner, Sebastian (2); Handschuh, Siegfried (1); Howanitz, Gernot (2); Radisch, Erik (2); Rehbein, Malte (2) | Deep Watching: Towards New Methods of Analyzing Visual Media in Cultural Studies | Deep learning, RCNN, visual media, cultural studies, quantitative methods | image processing audio, video, multimedia cultural studies data mining / text mining English computer science and informatics cultural analytics | Culture Studies often deal with non-textual media such as photographs, films and video clips. In order to broaden the methodological spectrum of this discipline, we propose "Deep Watching", i.e. applying deep learning and computer vision algorithms to large image and video corpora. By means of combining several state-of-the-art RCNN, we identify not only objects and symbols, but also persons and their mimics and postures in two sample corpora: the first contains 275 Youtube videos about Ukrainian nationalist Stepan Bandera, the second 1000 trading cards of American-German actress Marlene Dietrich. By means of investigating various questions connected to memory cultures and star cultures we show how our "Distant Watching" approach allows Cultural Studies to approach visual media in a quantitative manner. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Jander, Melina Leonie | Towards Tool Criticism: Complementing Manual with Computational Literary Analyses | Tool criticism; sentiment analysis; dystopian novels; manual and automatic literary analyses | literary studies multilingual / multicultural approaches digital textualities and hypertext standards and interoperability data mining / text mining English digital humanities (history, theory and methodology) | The term tool criticism found its way into the Digital Humanities. The necessity of the deliberated exposure to computational methods is discussed widely, aiming to understand the potential and limitations of the tools. Our current research provides a case study which illustrates the complementation of automated and manual analyses to bridge the gap between the two approaches. A self-designed corpus of dystopian novels serves as data for the application of different sentiment analysis tools. While the manual investigation of dystopian novels leads to an understanding of the genre as primarily delivering a pessimistic world view, the computational analyses do not only support this finding, but also challenge it. The latter points us in two directions: i) The tools need thorough investigation; ii) Textual sentiment analysis as such has to be put into question, since text does not carry intrinsic sentiments, but they are brought to the text through non-textual instances. |
| Van Hulle, Dirk; Schäuble, Joshua | Modelling Text-Genetic Relationships | text-genetic paths, modelling textual genesis, genetic editing, digital scholarly editing | corpus and text analysis literary studies scholarly editing philology digital textualities and hypertext English manuscripts description and representation | A traditional danger of manuscript research is that the researcher gets lost in the details of the archival material. Most digital archives and scholarly editions offer the traces of a work's genesis as digitized items, side by side. Instead of such a juxtaposition, this paper proposes a model that enables scholarly editors to discover and record the connections between these textual traces. The ability to record these connections facilitates a more comprehensive understanding of a work's genesis. This implies a dynamic model that allows users to turn the different genetic traces into the "motion picture" of the genesis. With the proposed model, zooming in on the smallest level of textual change no longer entails the danger of getting lost in the manuscripts thanks to the possibility to zoom out again at every stage in the enquiry and see the bigger picture. |

| authors | title | keywords | topics | abstract |
|---|---|--|---|--|
| Escobar Varela, Miguel (1); Arps, Bernard (2) | Scholarly Multimedia Editions for Theatre Studies | Theatre studies, digital philology, multimedia editions, Indonesia, wayang kulit | audio, video, multimedia film and performing arts studies scholarly editing English oriental and asian studies scholarly publishing, open content and open science | A scholarly infrastructure for the critical annotation of audiovisual documents has yet to emerge in theatres studies, even though relevant resources and technologies exist. To sketch a prototype for scholarly, multimedia editions for theatre studies, the present authors embarked on a collaborative journey of creativity and discussion. Both authors have an interest in the Javanese tradition of <i>_wayang kulit_</i> (shadow puppet theatre). We are currently collaborating on an interactive, multimedia version of a previously published print translation of an influential performance. The development of a digital portal is not just a matter of ‘adding’ audiovisual materials but a dialogical experimentation with the format and possibilities of a digital philology of performance, which we undertake with reference to previous work on digital scholarly editions in other disciplines. We are also developing a shareable, and reusable software package which we hope will be of general interest to theatre scholars. |
| Van Zundert, Joris J. (1); Neill, Iian D. (2); Andrews, Tara L. (3); Andreas, Kuczera (2) | Is There Anything It Is Like To Be a Text? | digital textuality, ontological status, text-as-graph | text encoding and markup languages scholarly editing digital textualities and hypertext linking and annotation English digital humanities (history, theory and methodology) | This paper takes up the problem of the ontological status of text in the digital environment to which some attention has been called before (e.g. Van Zundert and Andrews, 2017). We propose an analytical metaphor for the current prevalent perception of text in the article “Is There Anything It Is Like To Be a Bat?” (Hacker, 2003). We explain the congruence we have found between debates on consciousness and the currently-dominant understanding within scholarship of digital textuality. Based on three use cases, we then contend that understanding digital text as an emergent property of interconnected knowledge, rather than as a purely semiotic description, creates new affordances for textual scholarship that are complementary to the current status quo of digitally mediated text as mimesis of print culture. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Wuttke, Ulrike (1); Neuroth, Heike (1); Laura, Rothfritz (1); Jennifer, Edmond (2); Vicky, Garnett (2); Frank, Uiterwaal (3); Marie, Annisius (4) | The PARTHENOS Training Suite: Empowering eHumanities and eHeritage Research(ers) with essential Knowledge and Skills | digital Humanities, Cultural Heritage Studies, pedagogy, webinars | teaching, pedagogy, and curriculum GLAM: galleries, libraries, archives, museums digital research infrastructures and virtual research environments English digital humanities (history, theory and methodology) | In this long paper the main principles, insights, and results from the development of the PARTHENOS Training Suite will be presented to the international Digital Humanities community in order to stimulate the discussion of Digital Humanities pedagogy and to stipulate the further uptake of these resources. PARTHENOS develops educational resources that focus on fundamental, interdisciplinary knowledge and skills that facilitate successful engagement and use of digital research infrastructures for digital humanities and cultural heritage scholarship in an increasingly complex, networked, and open environment. These resources are brought to the target groups and audiences in appropriate ways based on didactic and practical insights, using up-to-date means of communicating knowledge and information. |
| Mellet, Margot Lise (1); Vitali-Rosati, Marcello (1); Bouchard, Elsa (2); Monjour, Servanne (3); Agostini-Marchese, Enrico (1); Kearney, Beth (4) | Palatine Anthology. Complexity for a digital research project | anthology, collaborative edition, digital classics | classical studies corpus and text analysis information architecture and usability digital textualities and hypertext semantic web and linked data English digital humanities (history, theory and methodology) | Contemporary technologies allow scholars to explore the boundaries of such research questions by fostering the development of new critical approaches to classical texts. In order to investigate these questions, the Canada Research Chair on digital textualities, directed by Marcello Vitali-Rosati, initiates a project in digital classics, selecting a corpus particularly suited for such research — the Palatine Anthology (PA), a collection of Greek epigrams. We created an open database searchable via an API that allows one to transcribe the manuscript, propose translations, align translations, transcribe scholia, and link epigrams both to each other, and to external literary and artistic references. In particular, we are looking to foreground a body of work by the main participatory re-translation and the design of an open platform. Our project aims to demonstrate how a digital and collaborative edition of the PA was designed to demonstrate its own philological, editorial and pedagogic challenges. |

| authors | title | keywords | topics | abstract |
|--|---|---|--|--|
| Burghardt, Manuel (1); Fuchs, Florian (2) | A Computational Approach to Analyzing Musical Complexity of the Beatles | musical complexity, computational musicology, Beatles | musicology content analysis data mining / text mining English digital humanities (history, theory and methodology) | In this paper we present a computational approach to measure the musical complexity in a corpus of user-generated transcriptions of Beatles songs. We show that musical complexity of the Beatles did not consistently increase over time (only a weak correlation was measured), as high musical complexity seems to be a situational phenomenon that can occur for single songs rather than for a complete album. The approach presented in this paper can be considered as a case study for further computational studies on musical complexity, thus adding to the branch of computational musicology as part of the Digital Humanities. |
| Edmond, Jennifer (1); Garnett, Vicky (1); Goulis, Helen (2); Schuster, Kristen (3); Wuttke, Ulrike (4) | Developing and Integrating Training Materials on Research Infrastructures into Higher Education course design: The PARTHENOS Experience | PARTHENOS, Higher Education Curricula, Research Infrastructures, Training Materials, pedagogical design | teaching, pedagogy, and curriculum English digital humanities (history, theory and methodology) | This long paper will set a context for training around Research Infrastructures, and their inclusion in formal Higher Education courses. Using PARTHENOS materials as a case study, we shall discuss the rationale for and process of designing the course to incorporate training materials around Research Infrastructures following extensive research via a survey, desk research and workshops into the current state of training provision in the Digital Humanities within Higher Education Institutions. It will then discuss the design of the resulting module in more detail, and the results of the evaluation process we have undertaken. Finally we offer the lessons learned throughout this process and recommendations for the incorporation of training materials from Digital Research Infrastructures. |

| authors | title | keywords | topics | abstract |
|--|---|--|---|--|
| Hanneken, Todd R. | Deep Digitization: Considerations and Tools for Imaging Cultural Heritage Beyond the Basics | Cultural Heritage; Digital Imaging; Spectral Imaging; Texture Imaging; Reflectance Transformation Imaging (RTI); Spectral RTI; | archaeology digital archives and digital libraries classical studies image processing cultural artifacts digitisation - theory, methods and technologies English manuscripts description and representation | Just as human perception is full of complexities, the construction of a digital facsimile is full of complexities. The project of “digitization” requires reflection on the nature of scholarly perception of the artifact and the digital tools suited to capture and represent the pertinent information. This presentation considers first the modes of perception in scholarly investigation of artifacts such as manuscripts including the importance of texture and interactivity for humanities questions. Second, we will consider the ability of “spectral imaging” to meet and surpass the capabilities of the human eye on first-hand inspection. Third, we will consider the tools for capturing and representing texture and interactivity. Finally, we will present the results of a recent project, funded by the U.S. National Endowment for the Humanities, to integrate Spectral Imaging and Reflectance Transformation Imaging (Spectral RTI). The software and documentation created by the project are freely available online. |
| Freedman, Richard (1); Fiala, David (2); Walter, Micah (3) | The Quotable Musical Text in a Digital Age: Modeling Complexity in the Renaissance and Today | Music, Annotation, Analysis, Citation, Quotation | musicology ontologies and knowledge representation renaissance studies linking and annotation English | We will report on recent developments in the encoding, representation, and citation of musical scores in a digital environment. Our work centers on an important but neglected corpus of sixteenth-century music (the Imitation Mass) in which composers adapted, quoted, and transformed pre-existing works as the basis of long cyclic settings of the Catholic Mass. The aural and textual resonances of these works are of intrinsic interest from a cultural standpoint. But of primary importance in the context of the Digital Humanities is the challenges of representing the complex musical connections in a durable, collaborative environment. We do this through various open-source technologies, above all the Music Encoding Initiative (which brings to musical scores what the TEI does for literary and historical documents). The CRIM Project (Citations: The Renaissance Imitation Mass) inaugurates a new kind of quotable text for a digital age, with wide application in musicology, and beyond. |

| authors | title | keywords | topics | abstract |
|---------------------------|--|---|---|--|
| Meister, Jan Christoph | Digital Synesthesia, Digital Parekbasis or the Point Of the Digital for the Humanities | Digital Synesthetics, Hermeneutics, Methodological reflection | cultural artifacts digitisation - theory, methods and technologies linking and annotation data mining / text mining English digital humanities (history, theory and methodology) | <p>The paper introduces _digital synesthesia_ and _digital parekbasis_ as methodological principles of DH– and it does so with the express strategic aim to make a philosophical argument for DH's humanistic nature and relevance.</p> <p>I argue that the digital mode of representation and analysis of symbolic artefacts is 'synesthetic' in that it (a) is able to represent artefacts irrespective of their medial modality and (b) is at base level semiotically impartial, i.e. not invested with ontological, cultural or conceptual preferences. Moreover, (c) its foundation in formalization and abstraction and the alienating effect of these principles in the modeling of symbolic artefacts trigger 'parekbasis', i.e. self-reflection.</p> <p>Against this background DH can, in a history of ideas perspective, be contextualized as methodology resonating with Romantic ideals.</p> |
| Cheng, Jonathan | Enumerating Gendered Bodies In Two Centuries of English-Language Fiction | Text Analysis, Gender Studies, Body Studies, Cultural Analytics | corpus and text analysis gender studies english studies data mining / text mining English cultural analytics | <p>In this project, I use text mining methodologies in order to model the physical dimensions of characterization. My project computationally addresses several questions: How much of characterization is comprised of bodily description? To what extent is character embodiment marked by gendered divisions? If gendered patterns do emerge, how do they behave across two-hundred years of English-language fiction? How does such analysis potentially change our understanding of gender representation both in earlier periods and our current moment?</p> <p>I analyze a collection of roughly 10,000 works of English-language fiction spread over 210 years (1780 to 1990) and compiled from two corpora. Roughly 1,000 novels (1780-1900) are from Chadwyck-Healy's nineteenth-century fiction corpus. Another 9,000 novels (1880-1990) were curated by the Chicago Text Lab. I tabulate the frequency of bodily features in character description as well as any verbs, adjectives, and other parts of speech elaborating them.</p> |

| authors | title | keywords | topics | abstract |
|---|---|---|---|--|
| O'Donnell, Daniel Paul (1); Singh, Gurpreet (1); Porter, Dot (2); Rosselli Del Turco, Roberto (3); Callieri, Marco (4); Dellepiane, Matteo (4); Scopigno, Roberto (4) | Publishing (and Forgetting) the Small or Medium-sized Scholarly Edition or Cultural Heritage Collection as Linked Open Data: Using Zenodo and Github to Publish the Visionary Cross Project | Zenodo, GitHub, Open Research Methods, Digital Libraries, Digital Scholarly Editing | digital archives and digital libraries scholarly editing GLAM: galleries, libraries, archives, museums English library & information science digital humanities (history, theory and methodology) scholarly publishing, open content and open science | We discuss an approach to publishing heterogeneous file data and long-form humanities research as both linked open data and a (human readable) digital scholarly edition using Zenodo and Github. This approach is broadly generalisable and answers a number of long-standing issues surrounding the publication of data and results in DH: 1 It promotes the discovery and long-term survival of published data and results with no requirement for future maintenance; 2 It conforms to archival standards and principles; 3 It is fully available for future extension, addition, excerption, reuse, repurposing, or reanalysis by others without negotiation; 4 It ensures that data and contextual analysis are linked bi-directionally meaning that users are always able both to access the discrete data points from which a Humanities-focused analysis and commentary is build and understand each data point in the context of these larger synthetic research products. |
| Hyvönen, Eero (1,2); Rantala, Heikki (1,2) | Relational Search in Cultural Heritage Linked Data: A Knowledge-based Approach | linked data, knowledge dicoverly, relational search, semantic web | information retrieval and query languages metadata ontologies and knowledge representation semantic web and linked data English library & information science digital humanities (history, theory and methodology) | This paper presents a new knowledge-based approach for finding serendipitous semantic relations between resources in a knowledge graph. The idea is to characterize the notion of ``interesting connection" in terms of generic ontological explanation patterns that are applied to an underlying linked data repository to instantiate connections. In this way, 1) semantically uninteresting connections can be ruled out effectively, and 2) natural language explanations about the connections can be created for the end-user. The idea has been implemented and tested based on a knowledge graph of biographical data, enriched by data linking to collection databases of museums, libraries, and archives. The demonstrator is in use as part of the BiographySampo portal of 13 100 interlinked biographies. |

| authors | title | keywords | topics | abstract |
|--|--|---|---|---|
| Googasian, Victoria; Heuser, Ryan James | Digital Animal Studies: Modeling Anthropomorphism in Animal Writing, 1870-1930 | animal studies, eco-criticism, anthropomorphism, cultural analytics, machine learning | literary studies natural language processing english studies English artificial intelligence and machine learning eco-criticism cultural analytics | What do animals do in the pages of fiction? Do they make decisions, have feelings, express interiority? Are they more similar to human characters, or to things, objects, and machines? In this paper, we approach these questions with computational methods in one of the first attempts to apply digital methods to animal studies. In a variety of corpora—from popular natural history to scientific writing about animal behavior to animal-driven fictions historically accused of anthropomorphism—we compare the semantic and syntactic footprints left behind by animals and humans. We discover that, from a computational standpoint, animals in fiction are indeed recognizable as characters, albeit characters who register intentionality through physical movement over speech and display a mental paradigm delimited by instinct and associative learning. Natural history writing, on the other hand, narrates animals in ways that seem surprisingly human-like when compared to animal representations in fiction more broadly. |
| Katsma, Holst | The Novel And The Quotation Mark | novel, quotation marks, evolution, free indirect discourse | digital archives and digital libraries corpus and text analysis literary studies bibliographic methods / textual studies english studies English cultural evolution | Although it is commonly assumed that the novel emerged during the 18th century, it has proven notoriously difficult to say what exactly was new about this new literary form. My work addresses this problem by starting with the novel's new visual features: chapter headings and quotation marks. This paper presents research on the latter. Over the course of the 18th century, the look and purpose of the quotation mark transformed dramatically within the English novel. A survey of genres shows that the novel played a principal role in this development (more than scientific journals, trial transcripts, poetry, and drama). It's the gradual emergence of a new literary device within a new literary genre—a consequence of the novel's "new and important role of dialogue" (Lukács). What's more, the quotation mark led to the conceptual clarification and prominence of "direct discourse," a necessary step for the emergence of free indirect discourse. |

| authors | title | keywords | topics | abstract |
|--|--|--|--|--|
| Wang, Qian (1); Nurmikko-Fuller, Terhi (1); Swift, Ben (2) | Analysis and Visualization of Narrative in Shanhaijing Using Linked Data | ontology, Chinese literature, mythology, linked data | multilingual / multicultural approaches ontologies and knowledge representation ethnography and folklore semantic web and linked data English oriental and asian studies | This paper seeks to bridge that gap by creating, analyzing and publishing a case study example - the Chinese classic 'Shanhaijing' (Classic of Mountains and Seas, 山海经) - using Linked Data methods. We recount the complexities of representing ancient Chinese literary narratives, captured through a close reading of the narrative (unstructured) data in both English and Classical Chinese. We evaluate the challenges in using tools developed from Western perspectives and for complete and largely homogeneous, highly-structured data for the capture of the characteristics and related information about the monsters and mythical creatures described within the 'Shanhaijing'. We present a purpose-built user-interface, which allows users to explore this data both with and without needing to write SPARQL queries. |
| Kizhner, Inna (1); Terras, Melissa (2); Rumyantsev, Maxim (1); Khokhlova, Valentina (1); Demeshkova, Elizaveta (1) | Measuring Bias in Aggregated Digitised Content: a Case Study on Google Arts and Cultur | Google Arts and Culture, diversity, bias, GLAMs, museums | digital archives and digital libraries art history and design studies GLAM: galleries, libraries, archives, museums cultural artifacts digitisation - theory, methods and technologies English diversity digital humanities (history, theory and methodology) | We use an example of Google Arts and Culture project and we demonstrate that this digital aggregator of art history images reinforces biases of art history canons that are prevalent in western culture. Although Google Arts and Culture represents almost all countries from the United Nations list, the number of images for a selected group of countries and museums is incomparably larger than that for most countries. This is important because patterns found in big data and the outcomes of the research in Digital Humanities will be skewed towards those presented by a small group of western countries and museums if data are taken at scale. |

| authors | title | keywords | topics | abstract |
|--|--|--|--|---|
| Andrist, Patrick (1,2) | Mapping the Complexity of Ancient Codices in Databases: the Syntactic Model | manuscripts, codex, graphic representation, manuscript descriptions and structure, manuscript complexity | corpus and text analysis databases & dbms medieval studies data models and formal languages English library & information science manuscripts description and representation | <p>Most ancient codices reveal one or more layers of complexity tied to their history. However, online databases usually do not represent these historical layers, so that searches with multiple criteria often result in inaccurate answers.</p> <p>This paper presents a language-independent model for describing manuscripts, called the syntactic model, which is already in use in several MySQL as well as XML/TEI databases, and was the object of a RDF exploratory prototype. A few examples will illustrate how it operates on three levels: the first is the codex as it is today; the second relates to its historical parts; and the third to their pieces of content.</p> <p>The next section will show how the structure of the codex can be visualised using a new tool, which generates graphic representations on demand from a manuscript description based on the syntactical model.</p> <p>The final part discusses some of the main current and upcoming challenges</p> |
| Endres, Bill; Cook, Matthew; Grime, John | Sublime Complexities and Extensive Possibilities: Strategies for Building an Academic Virtual Reality System | virtual reality, platform, sensory experience, 3D, tools | software design and development virtual and augmented reality digital research infrastructures and virtual research environments modeling, simulation, 3D/4D modeling English library & information science digital humanities (history, theory and methodology) | <p>Virtual reality (VR) opens profound possibilities for teaching and research. In the past, cost has hindered its uses by individual researchers and for public engagement. Fortunately, prices have fallen as technical solutions have evolved. However, reduced prices, like VR systems, have developed to facilitate gaming. This paper presents strategies for building an interdisciplinary VR system. As its case study, it uses the Oklahoma Virtual Academic Laboratory (OVAL), developed at the University of Oklahoma through an interdisciplinary and multi-university approach. OVAL is a scholarly-oriented, human-centered VR system. It is free to download and use. Strategies include cultivating partnerships with other universities (facilitated by presenting early work at DH 2016), recruiting researchers from across campus, and building a 2-person traveling VR workstation (to engage international scholars). These strategies have been essential to address the complexities of human experience and develop a flexible VR system to meet diverse disciplinary needs.</p> |

| authors | title | keywords | topics | abstract |
|--|---|---|---|--|
| Anderson, Clifford B.; Ramey, Lynn T. | Thinking Computationally in the Digital Humanities: Toward Block-Based Programming for Humanists | computational thinking, visual programming, block-based computing | teaching, pedagogy, and curriculum English computer science and informatics digital humanities (history, theory and methodology) | We review the potential of visual or “block-based” programming environments for teaching computational thinking in the humanities. Educators have created friendly, inclusive, and generative visual programming languages for teaching the principles of computing without presupposing that students who use them will become computer programmers. We argue that digital humanists should learn from these tools’ emphasis on the ludic over the pragmatic. We also suggest how these tools could evolve beyond their origins in the K-12 arena to provide a shared platform for teaching computational thinking in the digital humanities. By developing a block-based environment for humanists, we hope to advance computational thinking in our discipline while also providing educators with resources for introducing digital humanities into secondary and postsecondary courses on computational thinking. |
| Busch, Anna (1); Bludau, Mark-Jan (2); Brüggemann, Viktoria (2); Genzel, Kristina (1); Möller, Klaus-Peter (1); Seifert, Sabine (1); Trilcke, Peer (1) | Scalable Exploration. Prototype Study For The Visualization Of An Author’s Library On The Example Of 'Theodor Fontane’s Library'. | Author's Libraries, Visualization, Prototype | digital archives and digital libraries spatial & spatio-temporal analysis, modeling and visualization interface, user experience design, gamification german studies English library & information science manuscripts description and representation | This article presents a prototype for the explorable and scalable visualization of an author's library and its reading traces. It aims to combine design-oriented approaches to the visualization of cultural collections with philological, archival and library research questions. The visualization concept developed in the project places a special focus on continuous navigation within an author's library that can be zoomed and filtered at several granularity levels and that allows the exploration of individual objects as well as their comparison. It focuses on the experience and comprehensibility of the collection as a whole and the reading and usage patterns reflected in it, as well as provides access to individual phenomena. |

| authors | title | keywords | topics | abstract |
|--|--|---|---|--|
| Zaugg, Isabelle Alice (1,2) | Global Language Justice in the Digital Sphere: The Ethiopic Case | digital extinction, language diversity, critical digital studies, Ethiopia, Unicode | content analysis linguistics English diversity public humanities and community engaged scholarship globalization & digital divides communication and media studies | We currently face unprecedented rates of extinction of minority and indigenous languages and scripts, and digital technologies appear to be contributing to their decline. Scholars predict 50-90% of languages will become extinct this century, while only 5% of languages will attain digital vitality. This paper investigates what can be done to close this digital divide through an instrumental case study of Unicode inclusion and the development of supports for the Ethiopic script and its languages, including the national language of Ethiopia. Mixed methods include observation of digital governance institutions, archival research, a content analysis of script and language choices on social media, and interviews with Ethiopic digital pioneers. This paper concludes with recommendations to strengthen supports for digitally-disadvantaged languages, from inclusion in the Unicode Standard, to grassroots coding within and on behalf of digitally-disadvantaged language communities, to advancing the idea that supporting linguistic diversity is Silicon Valley's corporate social responsibility. |
| Jannidis, Fotis (1); Konle, Leonard (1); Leinen, Peter (2) | Thematic complexity | complexity, pulp fiction, genre | corpus and text analysis GLAM: galleries, libraries, archives, museums german studies English digital humanities (history, theory and methodology) | Our attempt in this paper to describe a useful approach to thematic complexity in fiction is part of our ongoing research on the complexity of fiction. This study now looks at thematic complexity. Obviously, there is no limit to the themes and topics a novel can deal with, but an infinite amount is difficult to measure. So we use the mixture of genres in documents as a proxy for thematic complexity, and we measure this mixture using topic modeling and Zeta to describe the genres. |

| authors | title | keywords | topics | abstract |
|---|--|--|---|--|
| Nishioka, Chifumi (1); Nagasaki, Kiyonori (2) | Log Analysis Method towards Understanding Detailed IIF Image Usage | IIF, log analysis, visualization, images, digital archives | digital archives and digital libraries image processing interface, user experience design, gamification English computer science and informatics library & information science | In these years, a lot of libraries and museums have adopted IIF (International Image Interoperability Framework), which promotes mutual use of images among institutions. In IIF-compatible digital collections, images are retrieved via IIF Image API, where regions of images are specified. Thus, it is possible to investigate the detailed image usage by examining which regions of images have been requested. In this paper, we propose a log analysis method to measure the image usage and to visualize the analysis result. Specifically, we employ the number of accesses to each pixel as a metric and visualize the result by heat maps. Since a pixel is the smallest unit that composes an image, we enable a fine-grained analysis. For instance, the analysis method can be applied to research platform, in which heat maps show which regions have been already examined by collaborators. |
| Southall, Humphrey; Aucott, Paula | Capturing the Geography of 1900s Britain as Text: Findings from the GB1900 Crowd-Sourced Gazetteer Project | citizen science, gazetteers, place engagement, visualisation | spatial & spatio-temporal analysis, modeling and visualization history and historiography crowdsourcing digital research infrastructures and virtual research environments English geography and geohumanities | Between September 2016 and January 2018 1,200 volunteers transcribed 2.6m. place names and other text strings from large scale historical mapping of Great Britain. This paper begins by summarising the history of the project, then describes the final outputs it has made available under Creative Commons licenses. It then explores two issues in greater detail. Firstly, it draws on a questionnaire survey and a series of depth interviews to investigate the volunteer experience, arguing that they were motivated not by an abstract desire to extend knowledge but by an engagement with specific places within the maps which had particular meaning for them. Secondly, it presents a series of visualisations not of the place names gathered by the project but the diverse other types of transcribed text, arguing that the maps, which contain no symbology, only text and the outlines of features, form an unusual but compellingly descriptive textual corpus. |

| authors | title | keywords | topics | abstract |
|--|---|--|--|---|
| Chuang, Tyng-Ruey (1); Hsu, Chih-Chuan (1,2); Syu, Huang-Sin (1,2) | Maps Re-imagined: Digital, Informational, and Perceptual Experimentations in Progress | Historical Maps, OpenStreetMap, Cartography, Taiwan, Japan | spatial & spatio-temporal analysis, modeling and visualization interface, user experience design, gamification software design and development cultural artifacts digitisation - theory, methods and technologies English computer science and informatics geography and geohumanities | A map can be thought as the fixture of certain places in a time to an artifact of indefinite longevity. The fixture was completed at the moment when the artifact had been produced. The places depicted on the map, however, can always be re-interpreted. We are interested in technical arrangements about historical maps upon which new possibilities can be experimented. We report on our experience in re-basing and rendering of early 20th Century land survey maps upon the OpenStreetMap technical infrastructure. Successive layers of digitalization and transformation are performed so as to bring out new representations from the old maps. Populated places, road networks, administrative centers and boundaries, water systems and diverse landscaping areas -- surveyed and mapped long ago on paper -- become digital objects waiting to be examined and navigated online. These places, connecting us to the past as well as to the present, can now be perceived anew. |
| Dunst, Alexander; Hartel, Rita | Quantifying Complexity in Multimodal Media: Alan Moore and the "Density" of the Graphic Novel | Complexity, comics, multimodality | corpus and text analysis image processing audio, video, multimedia cultural studies English communication and media studies cultural analytics | In an interview he gave in 2000, the well-known comics author Alan Moore remarked that canonical titles such as Art Spiegelman's Maus and his own Watchmen might legitimately be described as novelistic on the basis of their higher "density". Moore thus implicitly, and perhaps intuitively, hypothesized that the critical appreciation of graphic novels and memoirs may have a formal basis. In this paper, we describe the operationalization of Moore's concept of density with the help of a total of six textual and visual measures. We then present a pilot study of 40 graphic narratives drawn from a larger corpus, six of which can be described as canonical examples of this genre given their frequent discussion in academic scholarship. Our pilot study provides empirical evidence that clearly supports Moore's hypothesis that critically esteemed, or canonical, graphic novels and memoirs are characterized by higher density. |

| authors | title | keywords | topics | abstract |
|---|--|--|---|--|
| Hernández Lorenzo, Laura (1); Byszuk, Joanna (2) | Challenging Stylometry: The Authorship of the Baroque Play La Segunda Celestina | Stylometry, Baroque, Sor Juana Inés de la Cruz | gender studies stylistics and stylometry spanish and spanish american studies authorship attribution / authority English | The aim of this study was to verify the possibility of Sor Juana Inés de la Cruz authoring the anonymous part of the baroque play <i>La Segunda Celestina</i> , commissioned to Agustín de Salazar, and left unfinished after his death. This is a first systematic stylometric study on this problem and a baroque hispanoamerican text. In our study, we faced building a balanced corpus from few available resources, and took extensive evaluation measures to deal with unclear stylometric signals. We use a variety of established attribution and verification methods, and introduce a novel evaluation procedure of examining historic texts with scarce corpora. The results support Sor Juana's authorship, and unravel new connections between her and other authors of the time, showing, still undermined, powerful impact of her works on the epoch. The solutions adopted in solving methodological problems of such a complex task show how stylometry can overcome similar challenges. |
| Meyer-Sickendiek, Burkhard (1); Hussein, Hussein (1); Baumann, Timo (2) | From Fluency To Disfluency: Ranking Prosodic Features Of Poetry By Using Neural Networks | free verse prosody | audio, video, multimedia literary studies natural language processing speech processing English artificial intelligence and machine learning digital humanities (history, theory and methodology) | Our paper offers a method to detect the degree of fluency and disfluency used by a poet when reading his/her poems. In a first step, we will offer a precise framework to use it for estimating a spectrum of (dis-)fluency by using two important theories for analyzing poetry: The grammatical ranking and the rhythmic phrasing. In a second step, we will establish a gradual one-dimensional continuum, whose two poles are denoted by the terms fluent and dis-fluent. We illustrate this prosodic spectrum by ranking nine different poetic styles respectively prosodic patterns within a corpus of free verse poetry. We collected German poems available on the website of our partner (www.lyrikline.org). We present a model for the automatic classification of rhythmical patterns in the free verse poetry by using deep hierarchical attention networks. |

| authors | title | keywords | topics | abstract |
|--|---|--|---|--|
| Schmidt, Thomas (1); Burghardt, Manuel (2); Dennerlein, Katrin (3); Wolff, Christian (1) | Katharsis – A Tool for Computational Drametrics | Drametrics, Drama Mining, Sentiment Analysis, Computational Literary Studies | corpus and text analysis literary studies natural language processing content analysis german studies data mining / text mining English | We present Katharsis, a tool for "computational drametrics" that implements Solomon Marcus' (1973) theory of mathematical drama analysis. The tool computes and visualizes character configurations and speech statistics for different levels of analysis and allows users to compare different collections of plays. We illustrate the usefulness of the tool for literary studies via several use cases. The tool is freely available online for a test corpus of approximately 100 German plays: http://lauchblatt.github.io/Katharsis/index.html |
| Heße, Sascha | Clean Separation Of Overlapping Components In Line Segmentation Of Historic Handwritten Documents | segmentation, handwritten, deep learning, recognition | image processing medieval studies English computer science and informatics artificial intelligence and machine learning OCR and handwritten recognition | Most state of the art methods for handwritten text recognition work on segmented lines of text. The quality of the line segmentation directly influences the performance of the recognition system. In documents with extensive touching/overlapping parts, it is therefore important to separate the text lines precisely. We present two methods in this paper, how to achieve a high quality separation of touching/overlapping components and evaluate them on a medieval manuscript with extensive overlapping components. |
| Grandjean, Martin (1); Jacomy, Mathieu (2) | Translating Networks | Networks, Visualisation, Methodology, Metrics | spatial & spatio-temporal analysis, modeling and visualization network analysis and graphs theory English computer science and informatics digital humanities (history, theory and methodology) | Network interpretation is a flexible exercise in the digital humanities. While there is now a wide variety of uses in different fields, many projects highlight the difficulty of bringing graph theory and their discipline into dialogue. Fortunately, the development of accessible software has been accompanied in recent years by a critical reflection on our practices. Yet, it often focuses on technical aspects. In this paper, we propose to shift this emphasis and address the question of the researcher's interpretative journey from visualisation to metrics resulting from the network structure. Often addressed in relation to graphical representation, the subjectivity of _translation_ is all the more important when it comes to interpreting structural metrics. This is why this communication will propose to examine the practice of visual reading and metrics-based analysis in a correspondence table that clarifies the translation while presenting possible and generic interpretation scenarios. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Wang, Sean (1); Belouin, Pascal (1); Ho, Hou leong (2); Chen, Shih-Pei (1) | RISE and SHINE: A Modular and Decentralized Approach for Interoperability between Textual Collections and Digital Research Tools | e-infrastructure, interoperability, API, text mining, license | software design and development open access, copyright, licensing digital research infrastructures and virtual research environments standards and interoperability English digital humanities (history, theory and methodology) oriental and asian studies | Digital humanities as a field has been grappling with the significant issue of interoperability. In response, many have proposed that DH needs basic infrastructures behind research projects to ensure its long-term success. Yet existing infrastructures often suffer from too generic coverage and focus on only open-access resources. How can we, as scholars in DH and Asian studies, design a digital research infrastructure fit for our specific needs, taking past experiences with these large-scale infrastructural projects into consideration? In this paper, we present our technical answers to this question. "RISE" stands for Research Infrastructure for the Study of Eurasia. It is a pioneering approach for resource dissemination and emerging data analytics (such as text mining and other fair-use but consumptive research techniques) in the humanities, developed by the Max Planck Institute for the History of Science. |
| Hall, Elliott; Smithies, James | Misremembering Machines: A Creative Collaboration on Memory in AI-driven Storytelling | Artificial Intelligence, Research Software Engineer, Collaboration, Games, Digital Storytelling | interface, user experience design, gamification software design and development interdisciplinary & community collaboration literacy and creative writing English artificial intelligence and machine learning digital humanities (history, theory and methodology) | Misremembering Machines details a collaboration between an SME game company and an embedded Digital Humanities (DH) Research Software Engineer (RSE) to study the effect of memory on AI driven storytelling. The RSE is both a developer and a writer, uniquely suited to providing technical and creative feedback, and providing the connective tissue between the development and critical strands of the project. The project centred around translating an immersive theatre show into a video game, where AI characters replace the interactive audience elements. The AI characters dissemble, misremember and outright lie about the player's actions and choices. This translation raised a series of interrelated technical, social, and creative issues related to agency, bias and representation in AI driven storytelling, as well as the challenge of effective collaboration between different academic disciplines and the creative industries. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Hill, Mark J.; Säily, Tanja | Patterns of Early Modern Authorship: Using Metadata as Historical Record | digital history; quantitative analysis; social network analysis; metadata; book history; bibliographic data; intellectual history; library catalogues; literary studies | corpus and text analysis history and historiography metadata bibliographic methods / textual studies network analysis and graphs theory English digital humanities (history, theory and methodology) | This paper makes use of complex bibliographic metadata – the English Short Title Catalogue (ESTC) – to create a dataset which is analysed with quantitative tools in a way allowing for novel insights into historical perceptions of authorship and the structural backdrop for them. In doing this it demonstrates the relevance of both these tools and datasets for humanistic research. |
| Frieler, Klaus; Frank, Höger; Martin, Pfleiderer | Anatomy of a Lick: Structure & Variants, History & Transmission | jazz, oral transmission, patterns, similarity | corpus and text analysis audio, video, multimedia musicology data mining / text mining English | Oral transmission, not only of entire songs but also of smaller musical units, so-called licks, formulas or patterns, plays an important role in jazz. These patterns greatly facilitate the construction of very fast lines typical for jazz. Here, we present a case study of a common bebop lick. Using a newly developed similarity search system, we found 100 instances of the search pattern and the variants in the 456 jazz solo transcriptions from the Weimar Jazz Database. Despite a great deal of variation, certain structural invariants regarding metrical and harmonic placements can be observed and a prototypical version can be reconstructed. The lick was popularized by Charlie Parker and Dizzy Gillespie, probably based on earlier models from the swing era, and then spread to other bebop and hard bop players. Even modern post-bop players are still using this pattern, presumably as a direct reference to the bebop masters. |

| authors | title | keywords | topics | abstract |
|--|--|---|---|---|
| Esposito, Eleonora | Towards a Critical Approach to Digitally-Mediated Discursive Practices of Gender-Based Hostility | social media, online hostility, misogyny, critique, discourse | corpus and text analysis gender studies linguistics social media English digital ecologies, digital communities and critical infrastructure studies | <p>This paper presents a number of epistemological considerations in relation to digital media, discourses of hostility and critique, grounded in the results of a multi-lingual pilot study conducted in the context of a H2020-funded project.</p> <p>The study maps the multimodal discursive strategies of online hate against women in the public sphere by collecting and analysing a corpus of user-generated comments on Social Networking Sites from three different linguistic landscapes and political cultures in Europe, namely Italy, Spain, and the U.K. This paper envisions both audacious interdisciplinarity and substantial intersectionality as the epistemic way forward to make fresh contributions both for the fast-developing scholarship in Social Media Discourse Studies as well as for studies on digital hate and harassment.</p> <p>The domain of online misogyny as a digital discursive practice would be, therefore, conceptualized and investigated at the intersection of digital media scholarship, discourse theorization and critical feminist explication.</p> |
| Willkomm, Jens (1); Schmidt-Petri, Christoph (2); Schäler, Martin (1); Schefczyk, Michael (2); Böhm, Klemens (1) | The CHQL Query Language for Conceptual History Relying on Google Books | Temporal text corpora, query algebra, conceptual history | corpus and text analysis information retrieval and query languages philosophy data mining / text mining English computer science and informatics cultural evolution | <p>In this paper, we present a query algebra for empirical analyses of temporal text corpora, the Conceptual History Query Language (CHQL). A *temporal text corpus* in our sense is a set of words and word chains, i.e., ngrams, together with their usage frequency at various points of time, like the Google Books Ngram Corpus. Our query language is meant to be useful for conceptual historians, i.e., be descriptive and complete (match all actual and potential hypotheses of conceptual history), and bear optimization potential to allow fast query processing on large data sets. We focus on an algebra inspired by the German tradition of *Begriffsgeschichte* (conceptual history), as exemplified by the work of Reinhart Koselleck. We also show first results, namely, the change of the words "East" and "West" from parallel concepts in the geographical sphere to counter concepts in the political sphere.</p> |

| authors | title | keywords | topics | abstract |
|---|---|--|---|---|
| Edmond, Jennifer C (1); Lehmann, Jörg (2); Priddy, Mike (3) | Digital Humanities, Knowledge Complexity and the Six 'Aporias' of Digital Research | Interdisciplinarity, Collaboration, Big Data, epistemic cultures | project design, organization, management content analysis interdisciplinary & community collaboration English digital ecologies, digital communities and critical infrastructure studies digital humanities (history, theory and methodology) | This paper will introduce a frame of reference for understanding fundamental challenges that inform digital humanities as an interdisciplinary research area between arts, humanities, information and computer science. Its conclusions are based upon the evidence found in a corpus of 38 linked interviews about big data research, developed within an EU-funded collaboration known as Knowledge Complexity, or KPLEX for short (www.kplex-project.eu). When viewed from the perspective of the digital humanities, six distinct points of 'aporia' with a significant impact on DH appear in this corpus, places where the interviewees explicitly or tacitly expose gulfs in epistemic culture that are clearly at the heart of tensions between disciplines as they seek to collaborate. This paper will explore these areas of apparent irreconcilability, and conclude with a series of reflections on how digital humanities researchers might develop to become uniquely able to negotiate some of these critical conversations. |
| Dillen, Wout | Web Accessibility in Digital Scholarly Editing: Considerations from a Survey on Inclusive Design and Dissemination. | web accessibility, digital scholarly editing, survey results, inclusive design | interface, user experience design, gamification literary studies scholarly editing philology digital textualities and hypertext English diversity | Starting from the results of a recent survey on Inclusive Design and Dissemination in Digital Scholarly Editions, this paper will zoom in on the issue of web accessibility in the field. After situating the general survey, the paper will explore the results relating to this issue in more detail by mapping its respondents' awareness of relevant accessibility guidelines, as well as their position towards implementing them; illustrating what kind of accessibility features are offered; delineating how web accessibility issues are tested, resolved, and incorporated in the edition's workflow; and discussing in which cases the survey data suggests a regional divide in the answers we received to these questions. Taking some of the possible biases in the survey's data into account, this paper will then draw its conclusions from the survey, review their implications for the field of digital textual scholarship, and suggest a way forward. |

| authors | title | keywords | topics | abstract |
|---|---|--|--|---|
| Leavy, Susan; Greene, Derek; Wade, Karen; Meaney, Gerardine | Migration and Biopolitics in Cultural Memory: Conceptual Modelling and Text Mining with Neural Word Embedding | migration, contagion, biopolitics, word embedding, text mining | cultural studies data mining / text mining English artificial intelligence and machine learning digital humanities (history, theory and methodology) cultural analytics | This paper explores the cultural representation of migration and the biopolitics of contagion and disease represented in a digital corpus of literary fiction from the British Library. This work is part of a project examining shifting representation of migration, ethnicity and contagion in cultural memory. A curated subset of the British Library Nineteenth Century Corpus was examined using techniques from artificial intelligence and text mining. Concept modelling with neural word embedding revealed complex relational dynamics between societal views of migration, ethnic identity and contagion that question prevailing theories. Lexical ontologies were generated with word embedding to mine the corpus for excerpts of text that capture these conceptual relationships and enable critical analysis. This bridging of digital analysis and close reading sets out a methodology whereby patterns identified in corpora with artificial intelligence techniques may be critically evaluated through close reading of the text. |
| Meroño Peñuela, Albert (1); van Berchum, Marnix (2); van den Hout, Bram (3) | The Oldest Song Score in the Newest Notation: The Hurrian Hymn to Nikkal as Linked Data | RDF, Linked Data, MIDI, Music History, Sumerian Music | metadata musicology ontologies and knowledge representation semantic web and linked data English computer science and informatics digital art | In this paper, and following established practice [12], we propose to use the Resource Description Framework (RDF), the Linked Data paradigm, and the SPARQL query language [4] to answer fundamental questions in music history about the Hurrian Hymn to Nikkal. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Ares Oliveira, Sofia; di Lenardo, Isabella; Tourenc, Bastien; Kaplan, Frederic | A deep learning approach to Cadastral Computing | cadastre, deep-learning, segmentation, transcription, gis | image processing English computer science and informatics artificial intelligence and machine learning OCR and hand-written recognition | This article presents a fully automatic pipeline to transform the Napoleonic Cadastres into an information system. The cadastres established during the first years of the 19th century cover a large part of Europe. For many cities they give one of the first geometrical surveys, linking precise parcels with identification numbers. These identification numbers points to registers where the names of the proprietary. As the Napoleonic cadastres include millions of parcels , it therefore offers a detailed snapshot of large part of Europe’s population at the beginning of the 19th century. As many kinds of computation can be done on such a large object, we use the neologism “cadastral computing” to refer to the operations performed on such datasets. This approach is the first fully automatic pipeline to transform the Napoleonic Cadastres into an information system. |
| Vitali, Giovanni Pietro | Mapping Fascist Repression, Following The Italian Resistance | Resistance, Mapping, War, Network analysis | spatial & spatio-temporal analysis, modeling and visualization italian studies cultural studies network analysis and graphs theory English | This paper is the first official report on a Marie Curie project entitled Last Letters from the World Wars: Forming Italian Language, Identity and Memory in Texts of Conflict, which started in September 2018. This project deals with a linguistic and thematic analysis of the last letters of people sentenced to death during the First and the Second World Wars, conducted with digital humanities tools. In this very first part of the project, I am preparing the lexicon analysis that will be the focus of his methodology. I am also creating a geographical representation of the corpus because this project is intrinsically geographical in its approach. I am developing some tools that are specific to the analysis of the Second World War. Notably, I have already collected and georeferenced four datasets (heretofore named as DS) regarding the Italian Resistance against fascism and the Nazi occupation. |
| Lassner, David (1); Baillot, Anne (2); Coburger, Julius (1) | Attributions Of Early German Shakespeare Translations | attribution of translation, collaborative translation, machine learning | stylistics and stylometry authorship attribution / authority translation studies german studies English | This paper investigates the contribution of the three actors Dorothea Tieck, Wolf Graf von Baudissin and Ludwig Tieck involved in the joint translation project of Shakespeare's plays edited by August Wilhelm Schlegel and Ludwig Tieck. Machine Learning methods are used to analyse the English plays and their corresponding German translations to identify features of a joint translation system and deviations individual to the translators. |

| authors | title | keywords | topics | abstract |
|--|---|---|---|---|
| Lang, Sabine; Ommer, Björn; Ufer, Nikolai | An Interactive Search Engine to Detect Objects in Artistic Images | interface, object retrieval, large data collections, form analysis, semantic analysis | art history and design studies corpus and text analysis interface, user experience design, gamification semantic analysis content analysis English digital humanities (history, theory and methodology) | To find objects in images is of paramount importance to art history. Digitization has produced large image corpora, but manual methods prove to be insufficient to analyze these collections; the collaboration between art history and computer vision provides methods and tools which enable a comprehensive evaluation. It is the objective of this paper to present a user-oriented interface and underlying search engine for object retrieval, which assists with art historical research. After presenting specific requirements for retrieval systems, the paper introduces the design and handling of the search engine as well as shows results of different queries in diverse datasets. Eventually, these validate the applicability and usability of the system. We also include critical remarks on existing tools and possible issues, which might arise when working with art data. |
| Höger, Frank; Frieler, Klaus; Pfleiderer, Martin | Digging Into Pattern Usage Within Jazz Improvisation (Pattern History Explorer, Pattern Search and Similarity Search) | jazz, music information retrieval, web development, patterns, similarity | audio, video, multimedia information retrieval and query languages musicology software design and development data mining / text mining English computer science and informatics | The paper presents three novel user interfaces for investigating melodic patterns in jazz improvisation. Patterns are of particular interest for the analysis of improvisation styles, the oral transmission of musical language, the practice of improvisation, and the psychology of creative processes. Currently, the interfaces work with 456 manually transcribed jazz solos of the Weimar Jazz Database. The Pattern History Explorer allows for the examination of a set of 653 of the most common patterns within this database. The Pattern Search interface is a general two-staged pattern search in the Weimar Jazz Database that allows for extracting audio and score snippets. The Similarity Search application allows for exploring variants of a certain pattern according to established similarity measures. All applications could be easily transferred to other music corpora, too. They aim both at jazz researchers and at a wider audience of jazz teachers, students, and fans. |

| authors | title | keywords | topics | abstract |
|---|--|---|--|--|
| Mäkelä, Eetu; Tolonen, Mikko; Kanner, Antti | Charting the Material Development of Newspapers | newspapers, materiality, metadata, digital history | digital archives and digital libraries history and historiography metadata content analysis data mining / text mining English communication and media studies | This paper argues for the importance of taking into account the material aspects of newspapers when subjecting them to computational analyses. Our approach here is twofold: First, to ground discussion, we describe in concrete terms how we've extracted materiality-related information such as page and column counts, font information, page sizes and publication frequencies from a newspaper dataset in the ALTO XML format commonly used for such data. Second, we'll argue how this information is interesting not only for the study of the development of newspapers themselves, but also for studies that target the textual content of the newspapers. |
| Mathiak, Brigitte; Metzmacher, Katja; Helling, Patrick; Blumtritt, Jonathan | "The Role Of Data Archives In The Humanities At The University Of Cologne" | Research, Research Data Management, Survey, Digital Humanities, Humanities | digital archives and digital libraries digital research infrastructures and virtual research environments English library & information science digital humanities (history, theory and methodology) | There are three groups of stakeholders, when it comes to research data: Those who make data, those who use data and those who build infrastructure to match those two. In the literature, we find a lot of research on how to build infrastructure and how to share data (often written by the same group of people), yet there is relatively little research (but see Caria and Mathiak, 2018, Kern and Mathiak, 2015, Porter, 2016, Tenopir et al., 2011,2015, Warwick et al., 2008) on what the third group, the users, or rather re-users, actually want and what they do. Most of these studies do also not focus on the Humanities. While for other area of studies, research data sharing and reuse through data archives or journals is far more institutionalised (for different subject culture, see Tenopir, 2011, 2015), this is not so in the subjects covered by Humanities. |

| authors | title | keywords | topics | abstract |
|--|---|--|--|--|
| Yeager, Sean A. | The Temporal Space of Narratives | time, fabula, syuzhet, narrative, visualization | corpus and text analysis text encoding and markup languages spatial & spatio-temporal analysis, modeling and visualization literary studies ontologies and knowledge representation English digital humanities (history, theory and methodology) | This essay introduces temporal space as a means for revealing narratological complexities. A narrative's temporal space is visualized by plotting its fabula (the timeline of events within the story) against its syuzhet (the order in which they're presented to readers). The resultant graph is unique to each text, with the vertical axis corresponding to the characters' timeline and the horizontal axis corresponding to the reader's chronology. This is a visualization of the framework developed by Gérard Genette in Narrative Discourse. Scenes will correspond to at least one line segment within the two-dimensional plane defined by these temporal axes, causing the narrative as a whole to be isomorphic to a particular subspace of the plane. After surveying the relevant theoretical landscape, this paper outlines a methodology for the analysis of temporal spaces, theorizes on the myriad forms they may take, and demonstrates their potential to enhance traditional interpretative techniques. |
| Pinche, Ariane (1,2); Camps, Jean-Baptiste (1); Clérice, Thibault (1) | Stylometry for Noisy Medieval Data: Evaluating Paul Meyer's Hagiographic Hypothesis | OCR, Meyer, Legendaries, Anonymous works, stylometry | medieval studies stylistics and stylometry philology authorship attribution / authority English OCR and hand-written recognition | Stylometric analysis of medieval vernacular texts is still much of a challenge: the importance of scribal variation, be it graphic or more substantial, as well as the variants and errors introduced in the tradition, complicates the task of the would-be stylometrist. Basing the analysis on the study of the copy from a single hand of several texts can partially be a way around this issue (Camps & Cafiero, 2012), but the limited availability of complete diplomatic transcription might make it difficult. In this paper, we use a workflow combining handwritten text recognition and stylometric analysis, and apply it to the case of the hagiographic works contained in MS BnF, fr. 412. We seek to evaluate Paul Meyer's hypothesis about the constitution of groups of hagiographic works, as well as to examine potential authorial groupings in a vastly anonym corpus. |

| authors | title | keywords | topics | abstract |
|--|--|---|--|---|
| Zaagsma, Gerben Zaagsma | Digital History and the Politics of Digitization | digital history, cultural heritage, politics, digitization, nationalism | digital archives and digital libraries history and historiography cultural artifacts digitisation - theory, methods and technologies English globalization & digital divides digital humanities (history, theory and methodology) sustainability and preservation | The paper explores the question of the politics of digitization by focusing on one specific dimension: the question of digitization and selection, and its implications for historical research. It combines a theoretical, critical-reflexive approach with concrete examples. |
| Erdmann, Alexander (1,4); Wrisley, David Joseph (2); Joyeux-Prunel, Béatrice (3) | Active Learning from Scratch in Diverse Humanities Textual Domains: Optimizing Annotation Efficiency for Language-Agnostic NER | active learning, named entity recognition, machine learning, multilingual corpora, spatial humanities | corpus and text analysis multilingual / multicultural approaches natural language processing semantic analysis English computer science and informatics geography and geohumanities | We explore language agnostic strategies for named entity recognition and propose an active learning pipeline for diverse, multilingual, multidomain humanities corpora. Our examples comes from research in spatial humanities approaches to corpora. |

| authors | title | keywords | topics | abstract |
|---|---|---|--|--|
| Tuters, Marc; de Keulenaar, Emillie; Kisjes, Ivan; Bach, Daniel; Beelen, Kaspar | The Intellectualisation of Online Hate Speech: Monitoring the Alt-Right Audience on Youtube | YouTube, computational linguistics, digital methods, visual network analysis, extreme speech, hate speech | corpus and text analysis social media English digital ecologies, digital communities and critical infrastructure studies communication and media studies | This paper looks at YouTube as a platform for the circulation of far-right pseudo-intellectual ideas, as well as a possible machine for radicalization. While a significant amount of attention has of late been focused on the general problem of 'fake news' within digital humanities research, this paper proposes to take an empirically focused approach to broader seemingly philosophical problem of epistemological relativism in landscape of contemporary social media. Specifically it proposes to consider how it is that social media platforms, in this case YouTube, seem to create conditions amenable to the recruitment and mobilization of far-right politics, particularly via the medium of pseudo-intellectual debate. Our paper looks at how hate speech patterns by commenters on Youtube can be obfuscated into extreme speech. Specifically we consider how this may occur via commenters presenting their arguments in a pseudo-academic format. |
| Khan, Anas Fahad (1); Bohbot, Hervé (2); Frontini, Francesca (2); Khemakhem, Mohamed (3,4,5); Romary, Laurent (3,4,6) | Historical Dictionaries as Digital Editions and Connected Graphs: the Example of Le Petit Larousse Illustré | e-lexicography, semantic web, digital editions, linguistic linked data, dictionaries | text encoding and markup languages lexicography ontologies and knowledge representation linguistics semantic web and linked data English computer science and informatics | In this article we describe the publication of several consecutive editions of Le Petit Larousse Illustré as an RDF dataset using the ontolex-lemon vocabulary. |

| authors | title | keywords | topics | abstract |
|---|---|--|--|--|
| Heiden, Serge | Coping With The Complexity Of The TXM Platform Annotation Services With A Unified TEI Encoding Framework | digital text encoding, annotation, XML, TEI, textometry, TXM, digital hermeneutics | corpus and text analysis text encoding and markup languages natural language processing standards and interoperability English computer science and informatics digital humanities (history, theory and methodology) | <p>This paper presents new developments in the TXM textual corpora analysis platform towards direct text annotation functionalities. Annotations are related to entities of an external historical ontology called SyMoGIH, to co-reference information between words or to word properties like part of speech or lemma.</p> <p>The paper discusses methodological issues related to the unification in a single framework of the production and analysis of these annotations with the traditional annotations already available in TXM corresponding to the XML encoding of the text sources before they are imported into the platform and to the linguistic annotations automatically added to texts by NLP tools when imported into the platform.</p> |
| Smeenk, Kim (1); Bilgin, Aysenur (2); Klaver, Tom (3); Tjong Kim Sang, Erik (3); Hollink, Laura (2); van Ossenbruggen, Jacco (2); Harbers, Frank (1); Broersma, Marcel (1) | Grounding Paradigmatic Shifts In Newspaper Reporting In Big Data. Analysing Journalism History By Using Transparent Automatic Genre Classification. | Genre, journalism history, machine learning, transparency | corpus and text analysis English artificial intelligence and machine learning communication and media studies | <p>This paper shows how the systematic and quantitative study of genre in large digitized newspaper collections sheds light on the development of journalism discourse. It adopts a machine learning approach to add genre labels to newspaper articles. In order to make an informed decision on the right pipeline and understand its inherent biases, we have developed a dashboard that allows the scholar to explore the underlying decision-making process of the machine learning pipeline.</p> <p>We perform two studies: the first one aims to gain insight in how the machine learning pipelines' output compares to the distribution of manually coded golden standard data. In the second study the most suitable and trustworthy machine learning pipeline, is applied to large-scale unlabelled data to gain insight into the development of genre distribution between 1950 and 1995.</p> |

| authors | title | keywords | topics | abstract |
|--------------------------------------|---|---|--|--|
| Domber, Gregory Frank; Bodwin, Kelly | Mapping the Indefinable: Designing a Social Network Analysis Shiny App to Explore the Influence of East-West Exchanges on Poland's Political Transformation | International Exchanges, Democratization, Longitudinal Network Analysis, Shiny App | spatial & spatio-temporal analysis, modeling and visualization history and historiography network analysis and graphs theory English digital humanities (history, theory and methodology) prosopography | Beginning in the 1950s Americans sponsored international exchange programs for Polish scientists and professionals, believing that exposure to the West would undermine Communism. In 1989, Poles underwent a negotiated revolution. Were these extensive American public diplomacy efforts successful? Can pathways of influence and shifts in perception within specific epistemic communities be measured, mapped, and visualized longitudinally to better understand exogenous influences on Eastern Europe's democratization process? Based on an approach to quantifying individuals' lives based on their "institutional affiliations," our interdisciplinary team has designed an interactive social network analysis visualization app, built in R Statistical Software using the Shiny package. The app allows users to interactively explore the overlapping networks of political revolution and international exchange, and illustrate how these connections shifted over time. This provides insights into Poland's specific experience, as well as a model for studies of other complex, longitudinal networks. |
| Tweheyo, Robert | Collaborative Decision Making and Food Security: Digitizing Indigenous Knowledge of Rural Farmers in Uganda. | Collaboration, Decision Making, Food Security, indigenous knowledge, rural farmers, digitizing. | interdisciplinary & community collaboration cultural artifacts digitisation - theory, methods and technologies digital research infrastructures and virtual research environments English sociology indigenous studies sustainability and preservation | It intends to establish how local farmers' innovations and practices can be collected and shared among rural farmers and stakeholders to ensure sustainable food security. The major objective is to examine how relevant indigenous knowledge and practices that enhanced food security can be harnessed and digitized for sustainable use. Following a design science research philosophy in an engaged scholarly research paradigm, a digital artifact for managing and sharing indigenous knowledge for enhancing food security is developed as an intervention base. The artifact was instantiated and evaluated by the users and domain experts in the field of food security and information systems. It was perceived as a usable and useful artifact for digitizing and sharing indigenous knowledge |

| authors | title | keywords | topics | abstract |
|--|--|--|---|--|
| Hengchen, Simon; Marjanen, Jani | A data-driven approach to the changing vocabulary of the 'nation' in English, Dutch, Swedish and Finnish newspapers, 1750-1950 | newspapers, nationhood, word embeddings, vector space | corpus and text analysis history and historiography multilingual / multicultural approaches natural language processing English digital humanities (history, theory and methodology) | This project aims to mine two centuries worth of digitised newspapers in four languages, and to propose a methodologically sound, reusable approach to carry out quality historical research on the changing vocabulary relating to nationhood. The newspapers stem from different sources and countries, and are available in different formats. Our contribution proposes a reusable, methodologically-sound, data-driven way of tackling this problem. |
| Moynihan, Bridget (1); Armoza, Jonathan (2); Lang, Anouk (1) | Database Aesthetics and Ergodic Ephemerality: Remediating the Scrapbooks of Edwin Morgan | scrapbooks, prototypes, databases, digitization, print culture | digital archives and digital libraries databases & dbms literary studies digital textualities and hypertext cultural artifacts digitisation - theory, methods and technologies English communication and media studies | This paper presents a digital prototype which remediates a selection of scrapbooks of the Scottish poet Edwin Morgan, in order to engage with current debates around the digitization of print culture artefacts. Building on Ellen Gruber Garvey's work on nineteenth-century scrapbooks and their resonance with contemporary information management practices, Espen Aarseth's concepts of scriptons and textons, and Lev Manovich's theorizing of the poetics, aesthetics and ethics of databases, it explores some of the challenges involved in "denaturalizing" an unruly material object such as a scrapbook by remediating it in digital form, when its imbricated clippings, its refusal to conform to the imperatives of machine-readable data structures, and its inability to be reproduced as a facsimile due to copyright constraints all present significant challenges. |

| authors | title | keywords | topics | abstract |
|---|---|--|---|--|
| Humbel, Marco (1); Nyhan, Julianne (2) | The Application of HTR to Early-modern Museum Collections: a Case Study of Sir Hans Sloane's Miscellanies Catalogue | HTR; Sloane; manuscript; catalogue; model | GLAM: galleries, libraries, archives, museums English library & information science OCR and hand-written recognition | Handwritten Text Recognition (HTR) is “the ability of a computer to transform handwritten input represented in its spatial form of graphical marks into an equivalent symbolic representation as ASCII text.” (Romero et al., 2012, p. 5) What is the state of the art of the application of HTR to early modern manuscripts? With what level of accuracy can HTR models automate their transcription? What is known about how HTR currently accommodates manuscript text that shows changing writing styles, hands and text in multiple languages? We will explore these questions with reference to the wider literature and a case study of the first HTR model to be created for the hand of Sir Hans Sloane (1660-1753). |
| Bonch-Osmolovskaya, Anastasia; Daniil, Skorinkin | The Complexity of Character-building: Speech, Portraits, Interactions in Leo Tolstoy's "War and Peace | Tolstoy, characters, stylometry. corpus analysis, network analysis | corpus and text analysis literary studies stylistics and stylometry network analysis and graphs theory data mining / text mining English | This paper aims to introduce a new approach to the task of capturing textual complexity. We use five methods to model the character system in a novel, each one is aimed to discover one of the layers of this system. The combination of these layers gives as a result a complex view on the novel's composition enriched by computationally obtained data, quantitative and statistical metrics and graphical schemes and networks We apply stylometric and alternative non-lexical analysis to characters' direct speech, two alternative methods of network analysis to model characters interactions and clustering method for comparison of portrait descriptions in Leo Tolstoy's “War and Peace”. We claim that with the help of *the complex layer analysis* we can reveal some new structural constituents of the novel composition, that could not be captured by traditional interpretations of Tolstoy's poetics. |

| authors | title | keywords | topics | abstract |
|--|---|---|--|---|
| Martin, Kim (1); Miya, Chelsea (2); Brown, Susan (1); Murphy, Shawn (3) | Humanities Centered Design Features: Emergent Serendipity with HuViz | information visualization, user-testing, serendipity | interface, user experience design, gamification literary studies ontologies and knowledge representation information architecture and usability semantic web and linked data English library & information science | Recent attempts at designing for serendipity (The Bohemian Bookshelf 2012, STAK 2017, FERASAT 2017) highlight the importance of the accidental acquisition of information to research discovery. Sometimes, however, the aspects of visualization that lend themselves to serendipity are themselves discovered by chance. This paper describes HuViz, the Humanities Visualizer, a digital tool for visualizing semantic relationships and ontologies represented using the Resource Description Framework (RDF). Though not conceived as an environment to foster serendipitous experiences, HuViz has proven through user studies conducted over the past two years to do just that. After providing background on its development, we demonstrate how HuViz aligns with specific design features from the library and information science (LIS) literature on fostering serendipity. Finally, we demonstrate how participant feedback and observation led to the development of three new features in HuViz, which we will briefly demonstrate and connect back to the literature on serendipity. |
| Frank, Ingo | Diagramming the Complexities of Historical Processes: From Ontology- based Modeling to Diagrammatic Reasoning | Ontology-based Modeling, Applied Ontology, Visual Contextualization, Diagrammatic Reasoning, Historical Explanation | history and historiography ontologies and knowledge representation semantic web and linked data modeling, simulation, 3D/4D modeling English digital humanities (history, theory and methodology) public and oral history | This paper presents ongoing foundational theoretical and practical work on the application of ontology-based modeling to represent and visualize the complexity of knowledge disseminated in historical narratives. In short, the new approach combines modeling informed by philosophical ontology and philosophy of history with semiotically founded visualization of historical processes in order to support historical understanding. |

| authors | title | keywords | topics | abstract |
|---|--|--|---|---|
| Foletto, Leonardo Feltrin (1); Dalla Chiesa, Carolina (2) | Towards An Epistemology Of "Gambiarra": Technical Resignification In Brazil | hacker culture, global south, epistemology, technical resignification | corpus and text analysis anthropology English globalization & digital divides digital humanities (history, theory and methodology) digital activism and hacker cultures cultural analytics | The practice we discuss in this paper is what we call "Gambiarra": a term applied to a myriad of improvisations, usually material and technical ones as a result of scarcities of all sorts. Gambiarras are normally the technical expression through which people overcome everyday obstacles from the most ordinary ones to the outmost complex environments. Gambiarra has a meaning in the daily lives of people that tactically adapt its (digital) apparatuses in order to resist to daily-life problems of all sorts. As such, the paper is supported in two examples: "Gambiarra Favela Tech" and "Hacker Clubs" activities. In the first case, digital apparatuses are used as a way to produce art in Brazilian favelas; in the second case, the institutionalization of hacker activities for education and digital activism is discussed. |
| Kubis, Marek | Quantitative Analysis of Character Networks in Polish XIX and XX Century Novels | character networks, quantitative literary methods, corpus based methods, natural language processing | corpus and text analysis natural language processing network analysis and graphs theory data mining / text mining English computer science and informatics | The paper presents a study on induction and quantitative analysis of character networks inferred from Polish novels. The corpus gathered for this study contains both XIX century and XX century literary works in the form of electronic editions of novels and OCR-ed texts. |
| Hess, Leopold; Bary, Corien | Quantifying narrative perspective in Ancient Greek: Narrator language and character language in Thucydides | narrative perspective, narratology, Ancient Greek, Thucydides, corpus methods | classical studies corpus and text analysis literary studies semantic analysis stylistics and stylometry English | The paper presents a quantitative study of the role of content words in construction of narrative perspective in the text of Ancient Greek historian Thucydides. With the use of statistical analysis of relative frequencies, we identify as "character language" words that are used by the author importantly more often in characters' speeches than in the narrator's text, and we show that their infrequent but salient occurrences in narrator text contribute to perspectival effects. Importantly, not all of the "character language" words are ones that at first sight encode perspective-sensitive meanings, and so they could not be identified without the quantitative approach. We discuss the consequences of our findings for the study of narratives and for the linguistics of perspective-sensitive items. |

| authors | title | keywords | topics | abstract |
|---|--|--|--|--|
| Reina-Navarro, Alicia; Romero-López, Dolores | Interactive Reading of the Silver Age: teaching and research promoted by the National Library of Spain | interactive digital edition, Spanish Silver Age, Biblioteca Nacional de España | digital archives and digital libraries history and historiography teaching, pedagogy, and curriculum GLAM: galleries, libraries, archives, museums literacy and creative writing English scholarly publishing, open content and open science | The National Library of Spain (BNE) and the research group on Spanish Silver Age Literature (LOEP) at the Complutense University of Madrid present "La Edad de Plata interactiva" ("The interactive Spanish Silver Age"), a collaborative project to enrich the Library's digital collections and their use in teaching and research, exploring the resources that new digital technologies offer to the edition of texts in the field of cultural dissemination. The results of this research are inserted within the framework of eLITE-CM's Project (H2015/HUM-3426). "La Edad de Plata interactiva" has carried out three collections of interactive books from the Silver Age linked to different topics, which will allow readers to access knowledge about this period through points of view not so widely considered by canonical historiography: 1) The Modern Woman in the Silver Age Literature; 2) Children's Literature in the Silver Age; 3) Madrid in the Silver Age Literature. |
| di Lenardo, Isabella (1); Barman, Raphaël (2); Descombes, Albane (2); Kaplan, Frédéric (1) | Repopulating Paris: massive extraction of 4 Million addresses from city directories between 1839 and 1922. | City Directories, Urban Analysis, Geohistory, Social History | spatial & spatio-temporal analysis, modeling and visualization data mining / text mining English geography and geohumanities OCR and hand-written recognition | This article reports on the extraction and analysis of the data contained in "Didot-Bottin" covering the period 1839-1922 for Paris, digitized by the Bibliotheque nationale de France. We process more than 27 500 pages to create a database of 4,2 Million entries linking addresses, person mention and activities. The quality of the document analysis process is assessed diachronically and a conservative strategy was chosen in order to populate the database with only information of high confidence. An initial analysis of the data is presented, reporting on the overall statistics of the distribution of professions in Paris and their evolution during more than 80 years, as well a general overview of the diversity of family names through time. Seven case studies corresponding to different streets are briefly compared, showing how information in city directories capture statistically the dynamics of segmentation of the city into functionality differentiated neighbourhoods. |

| authors | title | keywords | topics | abstract |
|---|---|---|---|--|
| Brown, Susan; Cummings, Joel; Drudge- Willson, Jasmine; Lemak, Abigel; Martin, Kim; Mo, Alliyya; Stacey, Deb | Linked Literary History, or An Ontology of One's Own: The Canadian Writing Research Collaboratory Ontology | ontology design; semantic web; feminist theory | ontologies and knowledge representation cultural studies semantic web and linked data standards and interoperability English feminist studies | This paper describes significant decisions involved in creating an ontology for feminist literary history with a view to its extensibility to other literary and cultural work. It outlines the pragmatic process of designing an ontology to mobilize the data of The Orlando Project and eventually other data from the Canadian Writing Research Collaboratory. The principles and decisions outlined are: antifoundationalism and representationality; deferred upper-level ontology; provenance and citation; annotation; ambiguity, diversity, and nuance; and linking to legacy terms. It concludes that there are major challenges in the tensions between complexity and nuance, on the one hand, and readability and processability on the other, as well as between standards and bespoke terms. The Orlando British Women's Writing Dataset will advance feminist digital literary history, and experiments with the CWRC ontology will help refine strategies for writing feminist literary history, among other complicated stories, into the Web. |
| Arneil, Stewart; Holmes, Martin; Newton, Greg | Project Endings: Early Impressions From Our Recent Survey On Project Longevity In DH | sustainability, project planning, best practices | digital archives and digital libraries project design, organization, management standards and interoperability English digital humanities (history, theory and methodology) sustainability and preservation | This paper will present some of the findings of a recent survey conducted by the University of Victoria Project Endings team on issues surrounding the long-term preservation of DH projects. We also introduce an interactive questionnaire that we hope will challenge users to consider how the issues of long-term sustainability apply to their research projects. |

| authors | title | keywords | topics | abstract |
|---|--|---|---|--|
| Porter, J.D. (1); Algee-Hewitt, Mark (1); Fredner, Erik (1); Bronstein, Michaela (1); Manshel, Alexander (1); Nomura, Nichole (1); Droge, Abigail (2) | Microgenres | literature, interdisciplinary, narrative, novel, disciplines | corpus and text analysis literary studies data mining / text mining English cultural analytics | In this project, we explore the discursive interdisciplinarity of novels, using machine learning to identify points at which authors incorporate the language and style of other contemporary disciplines into their narratives. How do authors signal the shift between novelistic narrative and, for example, history, philosophy or natural science? And how do these signaling practices change across historic, disciplinary, and narrative measures? These stylistic shifts not only indicate the historically contingent ways that novels are assembled from heterogeneous discourses, but also shed light on the practices of disciplinary knowledge itself. |
| Chiarcos, Christian; Ionov, Max | Linking the TEI: Approaches, Limitations, Use Cases | linked data, TEI | text encoding and markup languages semantic web and linked data linking and annotation standards and interoperability English digital humanities (history, theory and methodology) | As a highly popular community standard, the Text Encoding Initiative provides the most frequently adopted model for the semantic markup of text data in the Digital Humanities. For several years, its co-existence with other standards for representing data on the web lead to vastly different strategies how these could be aligned with each other. With this paper, we provide an overview of these approaches and their respective implications and areas of application. |

| authors | title | keywords | topics | abstract |
|--|--|--|---|--|
| Dick, Jonathan; Hammond, Adam | "The Mold That's Branded On M Soul": A Computational Approach to Racialized Voice in Jean Toomer's "Kabnis" | Vernacular English; Sound Studies; Nonsemantic Analysis; Racialized Voice; Literature | audio, video, multimedia multilingual / multicultural approaches speech processing linguistics english studies English diversity | The eponymous character in Jean Toomer's "Kabnis" resists easy racial classification: not only is he described in mixed racial epithets (metaphors, for instance, like "lemon face"), but his speech, while occasionally dialectal, toes the line between an African-American Vernacular English (AAVE) register and the register of a standardized English. Leaving his racial identity to the interpretation of readers, literary scholars need to establish new methods through which to analyze this text, particularly when it is performed. The developing field of sound studies provides one such method. Using pause and pitch trackers in combination with visualization software, we can begin to identify and analyze the nonsemantic qualities associated with racialized performance, with the resulting quantitative data being an asset to potential arguments about the politics of vernacular voice in audio text. |
| Calderón Campos, Miguel; Díaz Bravo, Rocío | An Online Corpus For The Study Of Historical Dialectology | Corpus linguistics, historical linguistics, Spanish, computational linguistics, digital textual scholarship | corpus and text analysis spanish and spanish american studies linguistics English | The proposed paper aims to present the development of a new digital resource for the study of historical dialectology, thanks to TEITOK, "a web-based framework for corpus creation, annotation, and distribution, that combines textual and linguistic annotation within a single TEI based XML document" (Janssen 2016). The new online corpus has successfully overcome the following difficulties: a) It combines digital textual scholarship (TEI) and computational linguistics (based on the EAGLES international standard for morphosyntactic annotation and lemmatisation). b) It allows working in a single edition that can be visualised in different formats by the end user in the digital resource. c) Furthermore, it permits independent management, since scholars can upload and edit their work, having control over their own research. |

| authors | title | keywords | topics | abstract |
|---|---|---|---|---|
| Dillen, Wout; Schäuble, Joshua | IIIFarm. Teaching Image Interoperability on a Raspberry Pi Network of IIIF-Compliant Image Servers. | IIIF, pedagogy, raspberry pi, summer school, workshop | teaching, pedagogy, and curriculum standards and interoperability English library & information science sustainability and preservation | This paper will discuss a two-and-a-half-day workshop on IIIF that the authors designed and tutored as part of a summer school on analyzing and processing images in Digital Humanities. After introducing the setting, concept, setup, and structure of the course, we will present key lessons we learned organizing and teaching this course. In addition, we will report on the feedback we received from our students which they submitted to us anonymously. We will then reflect on how a hands-on and in-depth treatment of a concept as complex and relevant as IIIF can be as rewarding for the teacher as it is for the student. In that spirit, we will end by presenting the tutorial we are currently developing on the basis of this workshop, and which will be available for reuse at the time of the conference. |
| Malínek, Vojtěch (2); Umerle, Tomasz (1); Wciślik, Piotr (1) | From a Reference Book to Research Data: Literary Bibliographies as Sources for the Data-driven Research | bibliography, literary studies, cultural datasets, data-driven research | literary studies bibliographic methods / textual studies interdisciplinary & community collaboration digital research infrastructures and virtual research environments English library & information science cultural analytics | Different methods of data research - like “distant reading”, or “macroanalysis” - have gained recognition in the scholarly field. However, recently the quality of datasets for such research have come under scrutiny. Researchers like Katherine Bode call for data-driven research based on datasets that are richer, better documented, curated, and more systematic. This paper is written from the perspective of creators of large bibliographic databases (Czech Literary Bibliography and Polish Literary Bibliography) that have assumed a systematic approach to registering literary culture. The paper focuses on the challenges facing existing literary bibliographies in the age of data-driven research. It assesses how they adapt to the need for advanced data uses, and how application of data-driven methods in literary research revolutionises the way bibliographies are prepared, standardised and published. In conclusion, the paper identifies the gap between data studies and data production, which should be bridged through the interdisciplinary cooperation within digital humanities. |

| authors | title | keywords | topics | abstract |
|--|--|--|---|---|
| Eder, Maciej; Byszuk, Joanna | Feature Selection in Authorship Attribution: Ordering the Wordlist | stylometry, authorship attribution, feature selection, most frequent words, TFIDF | literary studies stylistics and stylometry authorship attribution / authority linguistics data mining / text mining English | The aim of this study is to compare (i) three different ways of ordering the features (word frequencies) according to their assumed importance, and (ii) four different ways of weighing the features themselves. These are: (i) mean term frequency (equal to values as on list of most frequent words), tf-idf (term frequency-inverse document frequency) and coefficient of variance, and (ii) term frequencies, z-scores, tf-idf and z-scored tf-idf. The results shows that while mean term frequency is the best approach to ordering texts, it offers even better results when combined with coefficient of variance, leading to optimal selection of important features. Among the weighting methods, z-scores outperformed other techniques. |
| Reckziegel, Martin (1); Wrisley, David Joseph (2); Hixson, Taylor Wright (2); Jänicke, Stefan (1) | Using Visualization to Understand the Complex Spatiality of Mappae Mundi | visualization, map warping, distortion, mappae mundi, medieval studies | image processing spatial & spatio- temporal analysis, modeling and visualization medieval studies English geography and geohumanities | A common practice in spatial humanities is georeferencing historical maps to generate rasters for use in Geographic Information Systems. Algorithmic analysis of this process can help identify geometric inaccuracies of “old” maps and visualizing those allows one to analyze historical cartographic technique. In this paper we turn instead to very old maps--examples of so-called “complex” medieval maps that blend conventional T-O structure with pseudo-geographic detail. Georeferencing them can lead to problems of occultation, or said another way, to hidden “folds” in the distorted map. The aim of this paper is to employ those extreme distortions in order to understand better their pre-modern organizational structure through the use of visualization. In our research we have found that such a practice allows us to situate such maps on a spectrum between the more topological and the more symbolic. |

| authors | title | keywords | topics | abstract |
|---|--|---|---|--|
| Page, Kevin (1); Burrows, Toby (1); Hankinson, Andrew (1); Holford, Matthew (1); Morrison, Andrew (1); Lewis, David (1); Velios, Athanasios (1,2) | A Layered Digital Library for Cataloguing and Research: Practical Experiences with Medieval Manuscripts, from TEI to Linked Data | Linked Data, TEI, catalogues, manuscript provenance, CIDOC CRM | medieval studies ontologies and knowledge representation GLAM: galleries, libraries, archives, museums digital research infrastructures and virtual research environments semantic web and linked data English library & information science | In this paper we report our experiences developing and applying a set of digital infrastructure elements which, in combination, realise a layered digital library (Page et al 2017) for the investigation of manuscript provenance. We describe several related technical contributions: encoding of manuscript catalogue and local authority records as TEI; using Github for version control, issue tracking, and collaboration; automated production of catalogue user interfaces derived from the TEI; an XML processing workflow identifying, extracting, and processing TEI elements for reuse in research; mapping workflow output into a CIDOC-CRM RDF export; reconciliation of RDF entities with external authorities enabling the creation and use of Linked Data bridging multiple datasets. We contextualise the co-evolution of these components and exemplify their use in studies of the provenance of medieval manuscripts. We reflect on the flexibility and extensibility provided by our layered approach, and the independent benefits for catalogers and scholars. |
| El Khatib, Randa; Arbuckle, Alyssa; Siemens, Ray | An "Open Lab?" The Electronic Textual Cultures Lab in the Evolving Digital Humanities Landscape | open social scholarship, physical place of DH, virtual space of DH, lab as incubator, lab as tree | digital research infrastructures and virtual research environments English public humanities and community engaged scholarship digital ecologies, digital communities and critical infrastructure studies digital humanities (history, theory and methodology) | The evolution of the digital, and its intersection with the traditional role of the humanities, has impacted academic and non-academic modes of communication and research practices. As the scholarly landscape evolves, so does the nature of the places and spaces of research, including those of digital humanities. Engaging with these transformations in knowledge creation, but also continuously expanding and evolving with them, is the Electronic Textual Cultures Laboratory (ETCL). This paper is based on the premise that there is a correlation between the developing knowledge landscape and the structure of an intellectual center, especially when it is committed to 'open' values. Our engagement with DH lab infrastructure operates on two levels: 1) lab infrastructure in physical and virtual settings to correspond to and reflect the evolving knowledge landscape and 2) lab practices that engage open social scholarship in physical place and virtual space, across research, teaching, and service. |

| authors | title | keywords | topics | abstract |
|--|--|---|--|---|
| Sperberg-McQueen, C. M. (1); Huitfeldt, Claus (2) | Bootstrapping Project-specific Spell-checkers | spell-checking, quality assurance, transcription | data models and formal languages project design, organization, management scholarly editing philology cultural artifacts digitisation - theory, methods and technologies English computer science and informatics | Spell-checking is a relatively mature technology which can benefit many DH transcription and data-creation projects. Informal inquiries establish that many such projects don't use any spell-checking technology. Our pilot studies, however, suggest that even for projects involving under-resourced languages, older texts, or texts written in variable, non-standardized spelling, spell-checking techniques can find some transcription errors at lower cost than manual proofreading. |
| Turton, Alexander Robert | Operationalising Ambiguity; Mapping the Structural Forms of Comics | Comics, data creation, literature, boutique | corpus and text analysis literary studies semantic analysis concording and indexing English | In response to the conference's theme of 'Complexities' and the "humanist way of building complex models of complex realities", this paper will report the findings of my doctoral research into how different structural forces "meet, reroute and disrupt" one another (Levine, p.23) in Alison Bechdel's graphic memoir, Fun Home (2006). In addition to reporting these specific findings, this paper will communicate the investigative method by which I reached them in the belief that it has a broader application for Digital Humanities projects that must work with non-indexical or otherwise ambiguous sources. |
| Schlör, Daniel (1); Schöch, Christof (2); Hotho, Andreas (1) | Classification of Text-Types in German Novels | text-types, machine learning, dataset, annotation | corpus and text analysis literary studies natural language processing stylistics and stylometry English computer science and informatics artificial intelligence and machine learning | When working with literary texts, a problem for linguists, literary scholars and for machine-based text understanding is the classification of text-types "descriptive", "narrative" and "argumentative". Being able to automatically assign sentences to these text-types highly desirable when aiming to support quantitative literary studies. In this work we present our text-type dataset, a feature based machine-learning model and a deep-learning based model and show that both are able to classify text-types well. |

| authors | title | keywords | topics | abstract |
|--|--|--|---|---|
| Jeremi, Ochab (1); Joanna, Byszuk (2); Pielström, Steffen (3); Maciej, Eder (4) | Identifying Similarities in Text Analysis: Hierarchical Clustering (Linkage) versus Network Clustering (Community Detection) | Stylometry, Clustering, Network Analysis | stylistics and stylometry English digital humanities (history, theory and methodology) | The aim of this paper is to introduce to stylometry the methods allowing for evaluation of classification results obtained with (i) hierarchical clustering methods, with the distinction of performance of individual linkage methods, and (ii) network clustering, with the comparison of community detection techniques. We compare three recognized evaluation measures: AMI, ARI and NMI using 6 model datasets of known clustering, of which three constitute binary problems and three – corpora with a large number (25) of expected internal groups, which were designed for authorship attribution (or similar multiclass problems). Our results show (i) superiority of Ward linkage method as compared to 5 other, (ii) greater performance and stability of Cosine Delta for both hierarchical and network clustering, (iii) Louvain as the most reliable method of community detection, and (iv) usefulness of AMI method for hierarchical clustering, which we propose for general use making our scripts available. |
| Tsai, Richard Tzong-Han (1,5); Lu, Yi-Hsuan (2); Wang, Yu-Chun (3); Fan, I-Chun (4) | Event Extraction on Classical Chinese Historical Texts: A Case Study of Extracting Tributary Events from the Ming Shilu | Event Extraction, Classical Chinese Historical Texts, Tributary Events, Ming Shilu | corpus and text analysis history and historiography natural language processing data mining / text mining English digital humanities (history, theory and methodology) | For investigating Ming-Chinese historical events, imperial historical records are one of the most reliable sources. There are two main challenges of studying imperial historical records. The first is that the records are usually quite lengthy. To trace the evolution of an event or a subject, historians need to mark up all mentions in relevant paragraphs, which is a tedious, time-consuming job. The second obstacle is that there is no publicly available text-mining tool that can extract person, location, time, and event mentions from Ming dynasty historical records. The aim of this work is to begin development of text-mining tools that tackle these two challenges. We focus on identifying tributary events in the Ming Shilu automatically by using SRL techniques. This work represents the first machine-learning-based attempt to extract atomic historical events from Classical Chinese as well as the first effort to solve this problem using SRL. |

| authors | title | keywords | topics | abstract |
|---|--|--|---|---|
| Dorn, Amelie (1); Wandl-Vogt, Eveline (1); Palfinger, Thomas (1); Theron, Roberto (2); Way, Andy (3); Abgaz, Yalemisew (3); Benito, Alejandro (2); Losada, Antonio (2) | "A Project Review Under The Focus Of Complexities On The Example Of ExploreAT!" | Digital Humanities; cultural analysis; society; Open Innovation infrastructure | lexicography interdisciplinary & community collaboration german studies ethnography and folklore English public humanities and community engaged scholarship | This paper provides insights into dealing with complexities in the Digital Humanities project exploreAT!. Exploring a non-standard language collection for cultural insights, a threefold approach is presented looking into concrete realisations and solutions of tackling challenges in terms of Open Innovation infrastructure, technology and the topic of choice, food. Methods and processes applied and developed in the project are aimed to serve as examples for future projects with similar datasets. |
| Kamosiori, Christina (1); Warwick, Claire (2); Mahony, Simon (1) | Embedding Creativity Into Digital Resources: Improving Information Discovery For Art History | art history, digital libraries, information behaviour, resource design | digital archives and digital libraries art history and design studies interface, user experience design, gamification digital research infrastructures and virtual research environments English library & information science | Over the past decades, the increase in the use of digital resources and the growth of research conducted in digital environments has transformed academic scholarship. Yet, as the employment of digital resources increases, so does the necessity to understand user behaviour and provide digital infrastructure tailored to the needs of researchers. Through this paper, we aim to explore how the design of digital libraries and resources can be improved to better facilitate information discovery and use in art history; for this purpose, we will look at scholars' creative encounters with information and present the implications for resource design. |

| authors | title | keywords | topics | abstract |
|--|---|---|---|--|
| Beshero-Bondar, Elisa (1); Mulligan, Rikk (2); Viglianti, Raffaele (3) | The Frankenstein Variorum Challenge: Finding a Clearer View of Change Over Time | Frankenstein, variorum interface, accessible design, collation, digital edition | text encoding and markup languages interface, user experience design, gamification literary studies scholarly editing digital textualities and hypertext English digital humanities (history, theory and methodology) | This long paper investigates questions of how best to illuminate change over time in designing an interface for Variorum editions, whose typically rich data and metadata can be offputting and difficult to share with human readers. Work is underway on the Frankenstein Variorum project to design an inviting interface for exploring changes to Frankenstein over five editions from its manuscript notebooks in 1816 to the last revision by the author in 1831. The authors address the development of a browser-based interface designed for scholars, fans, and new readers. They discuss the incorporation of accessibility and responsive design features, inviting visitors to read specific editions, review variants, explore edition-specific and cross-edition annotations, and visualize the movements of characters, author(s) and co-editors in space and time. Optimally, they seek to design an interface that teaches nonspecialists how to read change over time without sacrificing the standards of textual scholarship. |
| Haider, Thomas Nikolaus (1,2) | Diachronic Topics in New High German Poetry | poetry, topics, diachronic, authorship, evolution | literary studies natural language processing authorship attribution / authority data mining / text mining English computer science and informatics cultural evolution | Statistical topic models are increasingly and popularly used by Digital Humanities scholars to perform distant reading tasks on literary data. It allows us to estimate what people talk about. Especially Latent Dirichlet Allocation (LDA) has shown its usefulness, as it is unsupervised, robust, easy to use, scalable, and it offers interpretable results. We apply LDA to a corpus of New High German poetry (textgrid, with 51k poems, 5M token) and interpret salient topics, their trend over time (1575--1925 A.D.), and use the distribution of topics over documents for a classification of poems into time periods and for authorship attribution. |

| authors | title | keywords | topics | abstract |
|--|---|---|---|--|
| Ilovan, Mihaela (1); Brown, Susan (2) | CWRC-Writer Design and Survival Strategies: Observations from the Post-Launch Trenches | editing; tool development; user uptake; sustainability | literary studies scholarly editing english studies English sustainability and preservation | We here outline the approach employed by the Canadian Writing Research Collaboratory (CWRC) to encourage uptake and promote the sustainability of the CWRC-Writer. This browser-based XML and RDF editor is the centerpiece of the CWRC platform for the production, hosting, dissemination, and re-use of digital humanities scholarship. In development since 2011 and launched with the rest of the platform in 2016, the CWRC-Writer editor has achieved maturity and stability. Yet well prior to the launch, the CWRC team had begun thinking seriously about sustainability and building that into development priorities. We note the reliance of sustainability on uptake, review major reasons for lack of uptake, and relate them to our strategies for CWRC-Writer. |
| Roberts-Smith, Jennifer | Performing Historical Place: Leveraging Theatre Historiography to Generate Presence in Virtual Reality Design for Restorative Justice | restorative justice; virtual reality; historical reconstruction; historiography | film and performing arts studies virtual and augmented reality English public humanities and community engaged scholarship digital humanities (history, theory and methodology) digital art public and oral history | This paper reports a new theoretical approach, with accompanying methodology, to the representation of place in historical digital humanities projects arising out of a new digital resource that will be used in secondary schools in Nova Scotia, Canada to teach restorative justice practices as a means of addressing the impacts of systemic historical racisms still active in participating school communities. The Digital Oral Histories for Reconciliation project (DOHR) VR experience fulfils the educational mandate of the Restorative Inquiry underway in the province of Nova Scotia, Canada. We have drawn on discourse-, memory-, and repertoire-based theatre-historiographical theory to render the historical NSHCC as a place that is performed by (brought into being, and hence ontologically subsequent and secondary to) the speech acts of oral historians, rather than occupied by them (pre-existing and hence ontologically prior to their voices). We will make the Beta build available at DH 2019. |