
DH_BUDAPEST_2022
Digital Humanities Conference
November 23-25.

Welcome to DH_Budapest_22!

DH_BUDAPEST_2022 & DARIAH DAYS
3rd International Digital Humanities Conference
Gábor Palkó
Head of Department of Digital Humanities, Eötvös Loránd University

Dear Guests,
Dear DARIAH,
Dear Conference Participants,

I am delighted to announce that after two years of absence, we will again host the DH_Budapest conference in 2022. The event is supported by the National Laboratory for Digital Heritage, a project initiated and lead by the Department of Digital Humanities of the Eötvös University since 2020. The aim of the Laboratory is to promote the processing, research usability and broad smart accessibility of digital cultural heritage using digital technologies, especially artificial intelligence, in Hungary and in Hungarian speaking communities. Increasing Hungarian presence in the world of digital cultural heritage is one of the main tasks of the Laboratory since from the very beginning, we envisioned digital humanities research as an international collaboration.

DH_Budapest_2018, the first international conference we organized in Budapest, sought to provide a stimulating international forum to bring together researchers from Central Europe and beyond. It surveyed the current state of research in digital humanities in the hope of exposing further aspects of the role played by the digital medium in the present and the future of scholarly practices.

DH_Budapest_2019 was organized in cooperation with the COST Action '"Distant Reading for European Literary History" project and focused on the theories and practices of distant reading. Thanks to the successes of the previous year, we have already received support from sponsors such as the Ministry of Innovation and Technology, Springer Nature, GALE, and Qulto.

We are delighted to be organizing this year's conference jointly with Digital Research Infrastructure for the Arts and Humanities (DARIAH), a network of researchers and organisations devoted to digital humanities all over Europe. It is a great opportunity for us, as the only Hungarian member organization, to have the support of such an experienced and knowledgeable research alliance and infrastructure. The event incorporates a series of programs called DARIAH Days, which include roundtables, panel discussions, and workshops. DARIAH Days provide an opportunity for researchers and institutions to engage and build connections with others from the field of Digital Humanities, and serves as the initial step on the roadmap of the Hungarian full membership we work on for years now.

This year's theme is Network as a metaphor has been a tool for conceptualizing phenomena for a long time in the humanities, yet, with the emerging availability of digital and digitized datasets, it also has become a tool of quantitative analysis and data visualization. Today, the terms network and networking are widely used in humanities-related research from the investigation of semantic structures through the circulation of ideas, people, and artefacts to the collaboration patterns of institutions. Further, researchers have raised novel questions of self-reflection in the humanities regarding the epistemological grounds and consequences of the usage of tools and techniques developed in the natural sciences.

We invited speakers to discuss the terms networks and networking from multiple standpoints at an international research forum. The most innovative and prominent speakers of the DH_BUDAPEST_2022 conference will be inquired to submit their papers, which will be published as an issue in the International Journal of Digital Humanities published by the Department of Digital Humanities in association with Springer Nature.

DARIAH

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) aims to enhance and support digitally-enabled research and teaching across the arts and humanities. DARIAH is a network of people, expertise, information, knowledge, content, methods, tools and technologies from its member countries. It develops, maintains and operates an infrastructure in support of ICT-based research practices

and sustains researchers in using them to build, analyse and interpret digital resources. By working with communities of practice, DARIAH brings together individual state-of-the-art digital arts and humanities activities and scales their results to a European level. It preserves, provides access to and disseminates research that stems from these collaborations and ensures that best practices, methodological and technical standards are followed.

Venue:

We are delighted to introduce the location of DH_BUDAPEST_2022 and Dariah Days, the Trefort garden campus which is the home of the ELTE Faculty of Humanities. The campus is composed of several buildings designed by different architects and built in different architectural styles, such as Romanesque and Neo-Renaissance, making this complex a unique experience for visitors.

The Address: 1088 Budapest, Múzeum krt. 6-8.

Trefort garden is easy to approach since it is located next to Astoria, a busy place in Budapest city center, providing many options for public transport. Here are the most convenient options for international guests arriving from abroad:

- From Liszt Ferenc Airport (BUD): by the 100E Airport Shuttle, get off at „Astoria" station.
- From Keleti Train Station: by subway line M2 (red) in the direction of "Déli pályaudvar", get off at „Astoria" metro station.
- From Déli Train Station: by subway line M2 (red) in the direction of "Örs vezér tere", get off at "Astoria" metro station.
- From Nyugati Train Station: by metro replacement bus M3 in the direction of "Kálvin tér" metro station, get off at "Astoria".
- From Népliget Bus station: by subway line M3 (blue), get off at „Kálvin tér" metro station, the campus is about 7 minutes of walking distance.

Budapest metróhálózata / *Metro network in Budapest*

BUDAPEST

BKK BUDAPESTI KÖZLEKEDÉSI KÖZPONT

www.bkk.hu | bkk@bkk.hu | +36 1 3 255 255
bkkbudapest | bkkbudapest

23 NOVEMBER (WEDNESDAY)		
Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**	
9:00-10:30	**Registration**	
10:30-11:15	**Conference opening, Opening presentation (DARIAH)**	
11:15-11:35	Coffee break	
Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**	**Szekfű Gyula Könyvtár (Múzeum körút 6-8. / Main building, 1st floor)**
Chair	**Kees Teszelszky** (KB, National Library of the Netherlands, University of Groningen)	**Perczel Júlia** (Central European University, Department of Network and Data Science)
11:40-12:00	**Angel Abundis:** Use of Machine Learning Classification Models, both Image and Text, in the Network Graphing. Case Study: Community of Practice Among Graffiti Writers on Freight trains	**Akihiro Kawase, Junji Adachi, Kiichi Nakasu, Ayaka Kojima and Aoi Morikawa:** Structuring Melody in Traditional Japanese Music Using Network Theory
12:00-12:20	**Jun Ogawa, Satoru Nakamura, Asanobu Kitamoto:** Historical Knowledge Graph Creation with User-friendly Linked Data Editor	**Anna Matuszewska:** Interactive diagrammatic music analysis
12:20-12:40	**Mats Fridlund, Daniel Broden, Leif-Jöran Olsson, Victor Wåhlstrand Skärström, Magnus P. Ängsal and Patrik Öhberg:** Mapping the Domestic Politics of International Terror: An Actant Network Analysis of Swedish Parliamentary Debate on Terrorism, 1971–1978	**Chia-Ling Peng:** Invisible Network: Investigations of Graphic Notations through the Theory of Rationality and the Network Theory
12:40-13:00	**Yu Zhao, Zhaoyi Ma, Beijie He and Jie He:** Ontology-Based Image Knowledge Organization for Yangshi Lei Archives	**Christina Crowder and Clara Byom:** The The Klezmer Archive Project — Documenting Culture Bearers and Human Networks in Traditional Music Communities
13:00-13:30	**Discussion**	**Discussion**
13:30-14:30	Lunch	
Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**	
14:30-15:00	**Web archiving workshop – Márton Németh:** The theoretical and practical fundamental elements of web archiving – The first steps of institutional web archiving in Hungary	
15:00-15:30	**Web archiving workshop – Kees Teszelszky:** Archiving the Apocalypse: event harvests of crises in web archives for digital humanities research	
15:30-16:00	**Web archiving workshop – Zsófia Sárközi-Lindner, Balázs Indig, Mihály Nagy:** Opposing trends – Perspectives of using clean, small data with descriptive metadata in web archiving	
16:00-17:00	**Web Archiving Workshop** Panel Presentation: **Márton Németh** (Vera and Donald Blinken Open Society Archives) **Balázs Indig** (National Laboratory for Digital Heritage) **Gábor Palkó** (National Laboratory for Digital Heritage) **Mihály Nagy** (National Laboratory for Digital Heritage) **Zsófia Sárközi-Lindner** (National Laboratory for Digital Heritage) **Kees Teszelszky** (KB, National Library of the Netherlands)	
17:00-20:00	**Standing Reception – Kari Tanácsterem (Múzeum körút 4/A / Building A, Ground floor)**	

24 NOVEMBER (THURSDAY)

Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**		
9:00-10:00	**Keynote Lecture** **Mathieu Jacomy: Visual Network Analysis and Social Network Analysis for the humanities**		
10:00-10:30	Coffee Break		
Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**	Venue	**Központi olvasóterem (Múzeum körút 6-8. / Main building, ground floor)**
Chair	**Gábor Prószéky** (MorphoLogic, Hungarian Research Centre for Linguistics, Pazmany Peter Catholic University)	Chair	**Mathieu Jacomy** (Aalborg University, TANTLab, Gephi)
10:30-10:50	**Agoston Toth and Esra Abdelzaher:** BERT helps in sense delineation	10:30-10:50	**Jana-Katharina Mende:** Visualising Multilingual Literary Networks in 19th Century Monolingual Literary History
10:50-11:10	**Alejandro Napolitano Jawerbaum:** Of Manifestos and Mathematicians: A Case Study on Cross-Topic Identity Profiling Using Ted Kaczynski	10:50-11:10	**Gert Huskens, Christophe Verbruggen, Jan Vandersmissen and Julie M. Birkholz:** Lifting the veil of Levantine cosmopolitanism: diplomatic networking in Egypt, 1873-1914
11:10-11:30	**Begoña Altuna, Mikel Iruskieta, Ainara Estarrona, Aritz Farwell, Jose Maria Arriola, Jon Alkorta and Xabier Arregi:** CLARIAH-EUS: Building a Cross-border CLARIAH Node for the Basque Language	11:10-11:30	**Jan Lampaert:** Mapping the neo-avant-garde: visual network analysis of Flemish literary periodicals (1949-1970).
11:30-11:50	**Gábor Simon:** The network of personifications in online texts: case studies from the PerSE corpus	11:30-12:00	Discussion
11:50-12:10	**Mária Timári:** The edition's influence on the results of computer-based authorship attributoin of 19th century Hungarian novels	Chair	**Richárd Fejes** (National Laboratory for Digital Heritage)
		12:00-12:20	**So Miyagawa and Sophie Neutzler:** Digitization of a Japanese Christian Text from the Sixteenth Century
12:10-12:30	**Patrick Juola and Alejandro J. Napolitano Jawerbaum:** Stylometric Authorship Attribution in Seychellois Creole	12:20-12:40	**Tajana Jaklenec and Željka Tonković:** Creating knowledge of new architecture: Socio-semantic analysis of the magazine Arhitektura (1931-1934)
12:30-13:00	Discussion	12:40-13:00	Discussion
13:00-14:00	Lunch		
Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**		
14:00-14:30	**Panel Presentation (DARIAH)**		
Chair	**Tóth-Czifra Erzsébet** (DARIAH)		
14:30-14:50	**Adám Sebestyén:** Semantic networks in ELTEdata		
14:50-15:10	**Patryk Hubar, Nikodem Wołczuk, Dariusz Perliński, Róbert Péter and Vojtěch Malínek:** Literarybibliography.eu: harmonizing European bibliographical data on literature		
15:10-15:30	**Wachiraporn Klungthanaboon:** Thai Digital Humanities Researchers' Perspectives on Sharing Research Data		
15:30-15:50	**Kata Dobás:** The semantic pattern of Dezső Kosztolányi's bibliography		
15:50-16:20	Discussion		
16:20-17:00	Coffee Break		
17:00-18:00	**Panel Presentation (DARIAH)**		

25 NOVEMBER (FRIDAY)

Venue	**Nagy Laboratórium, Department of Digital Humanities (Múzeum körút 6-8. / Main building, 2nd floor)**
9:00-10:00	**Keynote Lecture**
	Molontay Roland: Introducing HSDSLab: How data and network science can help to answer research questions in human and social sciences?
10:00-10:30	Coffee Break
Chair	**László Bengi** (Eötvös Loránd University, Institute of Hungarian Literature and Cultural Studies)
10:30-10:50	**Indig Balázs, Palkó Gábor:** Automatic citation detection as a "distant reading" praxis: scrutinizing text similarity techniques on Hungarian texts
10:50-11:10	**Zsófia Fellegi, Anita Káli, Gábor Palkó, Zoltán Szénási**: "War as Network, Mihály Babits's poems about First World War "
11:10-11:30	**Claus-Michael Schlesinger and Pascal Hein:** From WARC to Graph: Link Extraction for Web Archive Analytics
11:30-12:00	Discussion
12:00-12:50	Coffee Break
Chair	**László Bengi** (Eötvös Loránd University, Institute of Hungarian Literature and Cultural Studies)
12:50-13:10	**Zsófia Fellegi:** Digital philology and the semantic web
13:10-13:30	**Andor Márton Horváth:** Critical editions in a database
13:30-13:50	**Levente Seláf and Anita Markó:** A Comprehensive Network Analysis of Early Hungarian Melodies and Poetical Forms
13:50-14:10	**Botond Szemes and Bence Vida:** Structural differences between tragedies and comedies
14:10-14:40	Discussion
14:40-15:10	**Conference Closing**

KEYNOTE SPEAKERS

Mathieu Jacomy:

Mathieu Jacomy is Doctor of Techno-Anthropology and post-doc at the Aalborg University TANT Lab. He was a research engineer for 10 years at the Sciences Po médialab in Paris, and is a co-founder of Gephi, a popular network visualization tool. He develops digital instruments involving data visualization and network analysis for the social science and humanities.

His current research focuses on visual network analysis, digital methods and issue mapping. He contributes to developing the web crawler Hyphe, the online network sharing platform MiniVan, and Gephi. He tweets at @jacomyma and blogs at reticular.hypotheses.org.

Roland Molontay:

Roland Molontay is an assistant professor of data science and network science at the Budapest University of Technology and Economics (BME), Budapest, Hungary. He is the founder and leader of the Human and Social Data Science Lab of BME, whose mission is to translate fundamental research in data and network science into a lasting impact in the social, human, policy, and business contexts. His research focuses on how one can use the tools of statistics, machine learning and network theory to solve problems arising in education, healthcare, and industrial settings.

He is the author of more than 30 scientific publications, a regular speaker at renowned international conferences, and the recipient of numerous professional awards, including the Gyula Farkas Memorial Prize.

Visual Network Analysis and Social Network Analysis for the humanities

Mathieu Jacomy
Aalborg UniversitetAalborg Universitet TANT-Lab
mathieuj@ikl.aau.dk

Keywords
Network, visualization, humanities, SNA, VNA

Abstract

We visualize networks to see something; but what? You will find yourself in one of two situations. If you have a precise idea of what you are looking for, then you know what to aim for. For instance, as a sociologist with a theory about *brokers*, you can operationalize it into network visualization. For you, the question is: how to manifest this concept visually? How to make brokers visible?

However, if you do *not* have a precise idea of what you are looking for, then the question becomes: what do I see? What do the visual patterns mean? What structures are manifested? You can still conceptualize network visualization as a test, but only insofar as you know what gets tested.

We find the first situation in social network analysis (SNA). That field studies the different ways people can be in relation with each other. It makes use of graph theory in various ways, and makes use of network visualization techniques, although those are not helpful all the time.

We find the second situation when one faces a relational phenomenon that one needs to describe or study. In the digital humanities, a corpus of letters; in sociology of science, academic publications; in the newsroom, e-mails. Visual network analysis (VNA) is a way to engage with such data without a precise idea of what to look for. It visualizes to mediate the empirical engagement with the data. But it does not mean that it is free of methodological commitments. VNA has its own goals and ways, that differ from practices in SNA, even though they largely use the same tools and techniques, which can be confusing.

I will expose the differences and commonalities between the two approaches, and show how a tool like Gephi can be used for each. I will notably focus on the layout algorithm, what is its purpose, how to read the patterns it produces, and understand them properly.

Introducing HSDSLab: How data and network science can help to answer research questions in human and social sciences?

Molontay Roland
Budapest University of Technology and Economics, HSDSLAB
molontay@math.bme.hu

Keywords
data science, network sciense, bibliographic analysis, co-authorship network, HSDSLAB

Abstract
In this talk, I will review some recent works of the Human and Social Data Science Lab (HSDSLab). HSDSLab is a newly established research group based in the Institute of Mathematics at the Budapest University of Technology and Economics. HSDSLab conducts both methodology-oriented basic research in data and network science and applied research with a human-centred and societal focus. The talk will revolve around two main topics: (1) Social network analysis and (2) Educational data science. As a tribute to the achievements of the network science community in the past 20 years, I will provide a bibliographic analysis and investigate the co-authorship network of network scientists to identify how the network science community has been evolving over time. Next, I will present our findings on the popularity of memes based on a content-based predictive analysis using memes from the Reddit social media site. Moreover, I will also sketch some data-driven research projects from the educational domain: including identifying students at risk of dropping out using explainable artificial intelligence; assessing the predictive validity of the admission system, and quantifying the impact of various interventions.


WEB ARCHIVING
WORKSHOP

The theoretical and practical fundamental elements of web archiving
The first steps of institutional web archiving in Hungary

Márton Németh
digital archivist
Vera and Donald Blinken Open Society Archive
nemethm@gmail.com

Keywords
web archiving, digital preservation, born-digital documents, National Széchényi Library, Hungary

Abstract

In this presentation an overview will be offered about the first Phd thesis had written in Hungarian about web archiving. In the thesis defended November 2021 at the Doctoral School of Informatics, University of Debrecen, a broad overview of theoretical and practical elements of web archiving had been offered, mainly in a Hungarian context, however in international perspective. The main aim is to facilitate further research and development activities in this field. The thesis offers an introduction to some basic conceptions and major challenges of web-archiving. A contextual analysis is being offered about web archiving related to the library, archive, and museum fields. A brief description of the international context is also being described. The core of the thesis has built on a detailed analysis about the initial conception, professional frameworks, and workflows of web-archiving in the National Széchényi Library, Hungary as the first organized project in that field in this country. The description of major workflow elements, the challenges in collection development and metadata management fields are being described together with the international context of the project. A brief outlook to the main elements of the software and hardware infrastructure, an introduction the challenges in long-term digital preservation context and a brief description of

the legal framework together with an outlook to the communication conception of the project are also essential elements of this analysis.

A whole chapter is focusing on the educational context of web archiving by international and Hungarian perspectives. An essential part of the thesis is describing the research perceptions in web archiving field by introducing various new sub-disciplines in this context and offering an overview to theoretical and practical research activities in the future related to the use of semantic microformats.

An overview will be offered about the major challenges could be met throughout the making of the thesis, and some ethical aspects of web archiving and research focusing on web archives will be also mentioned.

Bibliography

Drótos, L., & Németh, M. (2018). Web museum, web library, web archive The responsibility of public collections to preserve digital culture. In L. Petrovska, B. Ivane-Kronberga, & Z. Meldere (Eds.), The Power of Reading: Proceedings of the XXVI Bobcatsss Symposium, Riga, Latvia, January 2018 (pp. 124–126). Riga: The University of Latvia Press.
http://bobcatsss2018.lu.lv/files/2018/08/BOBCATSSS_2018_TheProceedings.pdf

Drótos, L., & Németh, M. (2019). A blended learning-based curriculum on Web archiving in the National Széchényi Library. Digital Library Perspectives 35(2), 97-114
https://www.emerald.com/insight/content/doi/10.1108/DLP-03-2019-0012/full/html
https://doi.org/10.1108/DLP-03-2019-0012

Németh M. (2019). Using semantic microformats for web archiving – an initial project conception. In Katarina Tomková (Eds.) LTP 2019 : Nové trendy a východiská pri budovaní LTP archívov: zborník príspevkov zo 4. medzinárodnej konferencie o dlhodobej archivácii Bratislava, 5. 11. 2019. (pp. 31-38). Bratislava: Univerzitná knižnica v Bratislave, 2019.

Geeraert F. & Németh M. (2020). Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary, WARCnet Papers, Aarhus, 2020.
https://cc.au.dk/fileadmin/user_upload/WARCnet/Geeraert_et_al_COVID-19_Hungary.pdf

Németh, M. (2021). The theoretical and practical fundamental elements of web archiving: The first steps of institutional web archiving in Hungary. Phd Thesis, University of Debrecen, Doctoral School of Informatics, 2021. http://hdl.handle.net/2437/310638 https://mek.oszk.hu/23400/23495/

Németh, M. (2021). Rákóczi thematic digital archive at the National Széchényi Library. ITlib, 2021/1-2, 42–45.
https://doi.org/10.52036/1335793X.2021.1-2.42-45

Archiving the Apocalypse: event harvests of crises in web archives for digital humanities research

Kees Teszelszky
KB, National Library of the Netherlands
kees.Teszelszky@kb.nl

Keywords: web archiving, Climate Change collection, War in Ukraine collection, digital humanities

Abstract

If we want to study world wide crises from a digital humanities perspective, we need to have broad international web collections containing carefully curated and preserved internet content. The International Internet Preservation Coalition (IIPC) does collaborative collecting of born digital web content about important international events. These efforts are done by the members of the IIPC Content Development Group, which are curators, collection specialists and web archivists from over 35 countries, including national, university and regional libraries and archives. The curated content is archived in cooperation with Archive-It, a part of the Internet Archive. IIPC works directly with researchers and research networks to promote use and analysis of archived Internet content. I my presentation, I want to introduce two built curated collaborative collections: the Climate Change collection (built in 2019) and the War in Ukraine collection (started this year). I want to show how this collection was built, what does it contain and what can be the use of these collections for digital humanities research.

Opposing trends – Perspectives of using clean, small data with descriptive metadata in web archiving

Balázs Indig
Department of Digital Humanities
indig.balazs@btk.elte.hu

Zsófia Sárközi-Lindner
Department of Digital Humanities, Doctoral School of History
lindner.zsofia@btk.elte.hu

Mihály Nagy
Department of Digital Humanities
nagy.mihaly@btk.elte.hu

Keywords
web archiving; small data; descriptive metadata; distant reading; trend viewer

Abstract

The need for institutionalised archiving of our born-digital cultural heritage is no longer an issue. However, who should archive what and in what form remains unclear. The large centralised archives, such as national libraries and other memory institutions (e.g. archive. org, Common Crawl), are present in the system as producers, while researchers and hobbyists are present as distinct groups of users or consumers. This imbalanced and arguably unhealthy dynamic forces the major actors to impose their will on the other party by means of brute force to avoid the fragmentation of the ecosystem. For example, the creation of goal-driven, small independent archives is deemed infavourable in the narrative set by centralised archives that are presumed to contain the same content, supposedly in a more standardised format. Access to the actual content of large archives may be limited either for copyright or technical reasons (e.g. overusing) or simply does not support distant reading.

While these archives are constantly evolving to enhance their services through modern approaches such as machine learning, there is still room for manual improvement. E.g. adapting feedback, additions and improvements by individuals. The changes imposed by such curation attempts are dwarfed by the size of the archive and add to its overall complexity resulting in their rejection.

In this presentation, we will introduce the results of our own independent long-term research, which approaches the issue of web archiving from the user's perspective. Our focus is on answering questions that scholars in different research fields and laymen have asked from us over the years. We will characterise a class of questions that can be answered sufficiently well with archives that are significantly smaller and cleaner than the data in central archives to encourage individual researchers to create their own archives or reuse others'.

While our workflow currently requires a lot of manual curation work to achieve our goals, those efforts are starting to pay off. We built our own service mimicking the Google n-gram viewer and other trend viewers to visualize the archived data in terms of the various types of manually curated descriptive metadata. With this service a number of questions can be answered based on the archive without writing a single line of code. While in our experience, the great majority of research questions involve descriptive metadata, it is currently not available in major archives, and require advanced technical skills to create. Besides we are working to simplify and automate the manual steps in our workflow, we created a pilot workflow as a demonstration to gain comparative results with only a few lines of code.

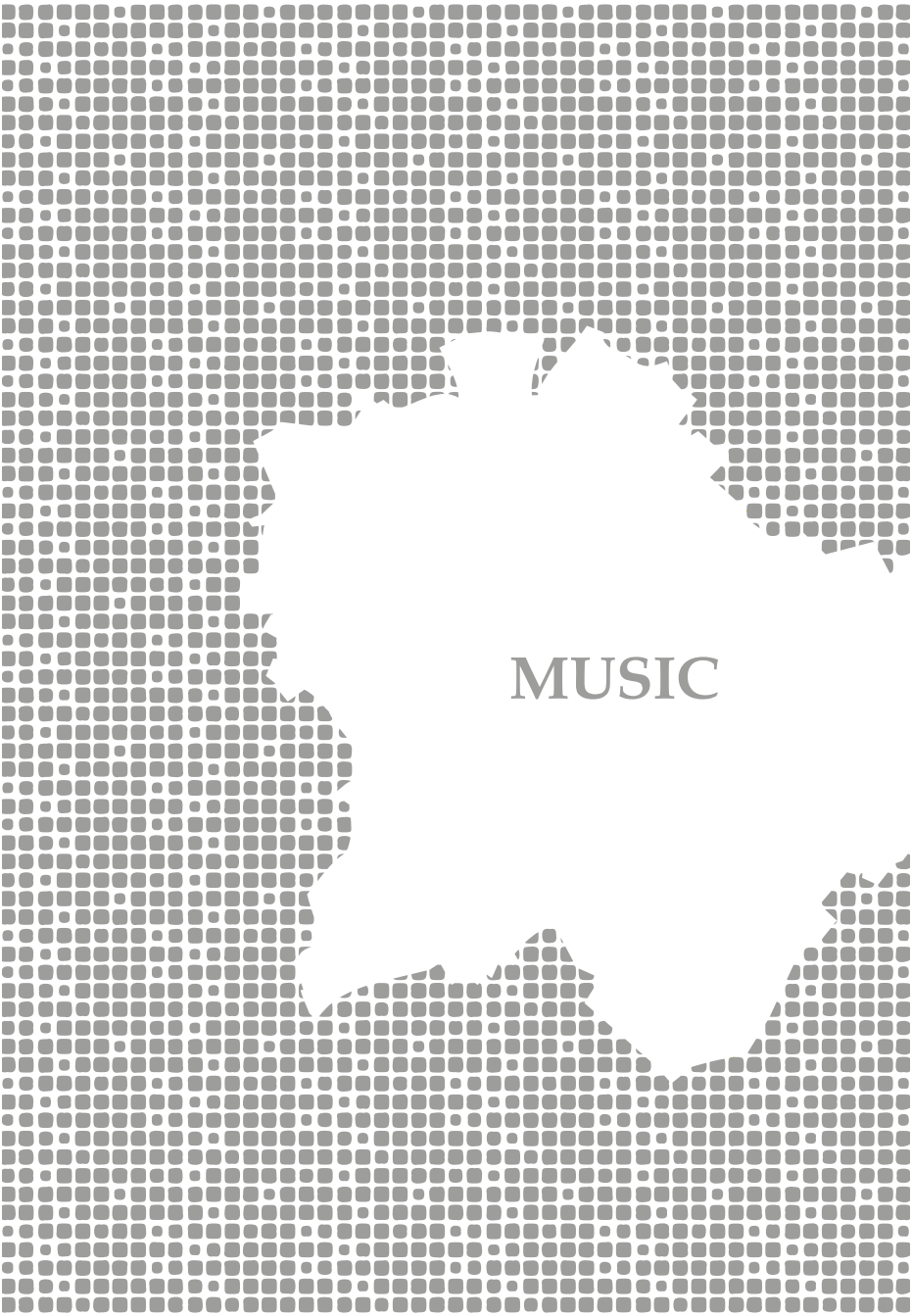
Finally, we lay out our future plans for the further refinement and enhancement of our service with advanced NLP and machine learning methods, which in our view can be used to easily add new layers to the data (e.g. sentiment, wikidata entites for semantic search). We exhibit further promising research initatives developed in our lab related to the application of web archives, which may serve to make further analytical goals more accessible.

Bibliography

Balázs Indig, Árpád Knap, Zsófia Sárközi-Lindner, Mária Timári, and Gábor Palkó. 2020. The ELTE.DH Pilot Corpus – Creating a Handcrafted Gigaword Web Corpus with Metadata. In Proceedings of the 12th Web as Corpus Workshop, pages 33–41, Marseille, France. European Language Resources Association.

Balázs Indig, Zsófia Sárközi-Lindner, Mihály Nagy. 2022. Use the Metadata, Luke! – An Experimental Joint Metadata Search and N-gram Trend Viewer for "Personal" Web Archives. In Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities, (accepted, in press), online.

Niels Brügger. 2018. The Archived Web: Doing History in the Digital Age. Cambridge, MA: The MIT Press.


MUSIC

Structuring Melody in Traditional Japanese Music Using Network Theory

Akihiro KAWASE
Faculty of Culture and Information Science, Doshisha University, Japan
kawase@dh.doshisha.ac.jp

Junji ADACHI
Faculty of Culture and Information Science, Doshisha University, Japan
adachi.junji@dh.doshisha.ac.jp

Kiichi NAKASU
Faculty of Culture and Information Science, Doshisha University, Japan
nakasu.kiichi@dh.doshisha.ac.jp

Ayaka KOJIMA
Faculty of Culture and Information Science, Doshisha University, Japan
kojima.ayaka@dh.doshisha.ac.jp

Aoi MORIKAWA
Faculty of Culture and Information Science, Doshisha University, Japan
morikawa.aoi@dh.doshisha.ac.jp

Keywords
tone system, musical structure, network centrality, children's songs, skeleton theory

Abstract

In musicology, the term "tone system" is used to refer to the laws inherent in melody. This abstract concept is based on the classics of Sachs (1943) and means "a system of music excluding the temporal component" or "the interrelationship of sounds in a certain music." It has been utilized to discuss the propagation, transformation, and universality of music in any given culture. Several qualitative studies have discussed the tone systems in traditional Japanese music, including the tetrachord theory proposed by musicologist Fumio Koizumi (1958) and the skeleton theory developed by composer Minao Sibata (1978). Koizumi (1958) analyzed the melodies of

Noh (classical Japanese dance-drama), popular music, folk songs, and traditional children's songs and discovered several final notes within an octave. He named these notes to distinguish them from the tonic in Western music theory. Furthermore, since in many cases, only one intermediate tone exists between two nuclear tones with a perfect fourth pitch interval, he considered this framework to be an important tone system in traditional Japanese music and named it "the tetrachord." In this regard, Sibata (1978) developed the skeleton theory in consideration of the limitations of the tetrachord theory's explanatory capabilities. In particular, he deconstructed "the tetrachord" into even smaller units, the perfect fourth and major and minor second pitch intervals. Skeleton theory abstractly representing the transitional relationships between pitches in a melody as a network model (directed graph). As described above, Koizumi proposed the tetrachord theory to understand the tone system of traditional Japanese music, and Sibata developed the skeleton theory, which is an abstract model of the tetrachord theory. Both these theories have been applied not only to traditional Japanese music, but also to the analysis of popular Japanese songs (e.g. Kawase 2018). They have been effectively utilized to visually grasp the primitive melodic structure of Japanese music.

However, the network's reproducibility could not be ensured during the analysis because the interpretation of nuclear tones varied depending on individual cases, including that of Sibata himself. Therefore, to ensure the reproducibility of the skeleton theory, in this study, we aimed to formulate a method for generating networks from melodies and to recapture it from a quantitative perspective. Specifically, we represented a group of Japanese children's songs as our network model. These songs have certain primitive characteristics associated with traditional Japanese music. Based on this model, we proposed a framework for classifying and comparing songs with similar melodies by applying various methods developed in the graph theory.

Our analysis revealed that the reproducibility of melodies' network representation can be guaranteed depending on how the nuclear tone is defined. Furthermore, by using multiple network centrality measures, we found that the different levels of importance and the functions of each nuclear tone can be revealed.

Acknowledgments
This work was supported in part by the Japanese Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research Number 18K18336, 21K12587, and the Area Informatics Project of Center for Information Resources of Area Studies (CIRAS), Kyoto University.

References
Sachs, C.: The Rise of Music in the Ancient World: East and West, Norton and Company. 1943.
Sibata, M.: Story of the Skeleton of Music (Ongaku no gaikotsu no hanashi). Ongaku no tomo-sha, 1978.
Kawase, A.: Comparisons of pitch intervals in Japanese popular songs from 1868 to 2010. In Proceedings of Japanese Association for Digital Humanities Conference 2018: JADH2018. 2018.
Koizumi, F.: Research on Japanese Traditional Music (Nihon dento ongaku no kenkyu). Ongaku no tomo-sha, 1958.
Game Songs of Japanese Children, Comparative Scores/Studies of Game Songs. (Warabeuta no Kenkyu, Gakufu-hen/Kenkyu-hen). Edited by Koizumi, Fumio. Warabeuta no kenkyu kanko-kai, 1969.

Bibliography
Akihiro Kawase is an Associate Professor at the Faculty of Culture and Information Science, Doshisha University, Japan. He is engaged in structuring and analyzing classical Japanese music. He won the ADHO Bursary Award at the Digital Humanities 2014 conference.

Interactive diagrammatic music analysis

Anna Maria Matuszewska
The Institute of Literary Research of the Polish Academy of Sciences
Institute for Music Informatics and Musicology, University of Music, Karlsruhe, Germany
an.ma.matuszewska@gmail.com

Keywords
digital musicology, data visualisation, interactive visual analysis, digital tools, diagrammatic reasoning

Abstract

One of the most popular trends in digital musicology is the creation of digital archives. On one hand it opens up new research opportunities; on the other hand, it forces researchers to develop a methodology for working with large data sets. The potential that digital analytical tools bring to both the analysis of single pieces of music and music corpora is very significant, but the lack of comprehensiveness of these tools, the difficulty of use and the unintuitive interface can discourage potential users. Although many solutions have already been proposed in this field and more and more software developers are designing tools with the needs of users in mind, it is often forgotten that also the analytical data obtained from computational analyses should be more explicit, understandable and presented in a way that allows its direct further transformation. Therefore, an interactive, visual exploration of the analytical results, as proposed in the project linked below, should be considered crucial. The purpose of this presentation is to demonstrate new ways of processing, displaying and analyzing musicological data using dashboards in Tableau Public - easily available data visualization software, and show how interactive diagrammatic music analysis provides a way to bridge the gap between computer science and musicology. The project is inspired by diagrammatic reasoning as proposed by Charles Sanders Peirce. The core of the project constitute the results of a computer-assisted analyses of a digital collection of Johann Sebastian Bach's fugues BWV 846-869 encoded in Humdrum **kern format, carried out in software such as Humdrum Toolkit

(running in unix-based computer systems), music21(Python-based) and Music Processing Suite (software for all operating systems). The results of the analyses were unified, translated into a relational database in Tableau Prep and visualized as a set of interactive coupled diagrams on analytical dashboards dedicated to melodics, rhythmics and harmony. The visualized data includes, but is not limited to, the formal structure of the composition, interval statistics, note durations and their placement in measures, dissonances and harmonic analysis. The project is adapted for both close reading, which facilitates detailed understanding of individual pieces, and for distance reading, which allows the discovery of relationships between subsets of the corpus. A crucial feature of the proposed environment for viewing analytical data is interactivity. The interactive solutions aimed to overcome some of the limitations of automated analysis and to make dashboard operations and diagram-based inference intuitive and easy to adopt in everyday research practice. The proposed analytical dashboards do not impose rigid rules of work on researchers, and thanks to many implemented filters, they allow for reconfiguration of the presented data and narrowing the scope of analysis of a piece of music to any combination of voices, motifs and measures, or for comparative analysis of individually defined combination of pieces of music from the collection. Because the diagrams on the dashboards are coupled to each other, as a result of an action on any of them, the information displayed is automatically recalculated on the entire set.

Bibliography

Cuthbert, Michael Scott and Christopher Ariza. "music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data". In J. Stephen Downie and Remco C. Veltkamp (Eds.). 11th International Society for Music Information Retrieval Conference (ISMIR 2010), August 9-13, 2010, Utrecht, Netherlands. pp. 637-642.

Giardino, Valeria. "3. Behind the Diagrams: Cognitive Issues and Open Problems". In Thinking with Diagrams. The Semiotic Basis of Human Cognition, ed. Sybille Krämer and Christina Ljungberg. Berlin, Boston: De Gruyter Mouton, 2016, 77–102, https://doi.org/10.1515/9781501503757-004.

Hofmann, David M. "Music Processing Suite: A Software System for Context-based Symbolic Music Representation, Visualization, Transformation, Analysis and Generation". University of Music, Karlsruhe, 2018. PhD Dissertation.

Krämer, Sybille, and Christina Ljungberg. "Thinking and Diagrams – An Introduction" in Thinking with Diagrams. The Semiotic Basis of Human Cognition, ed. Sybille Krämer and Christina Ljungberg. Berlin, Boston: De Gruyter Mouton, 2016, 1–20, https://doi.org/10.1515/9781501503757-001.

Peirce, Charles S. "Lectures on Pragmatism. Lecture VI: Three Types of Reasoning". In The Collected Papers of Charles Sanders Peirce [1866–1913], vol. 5: Pragmatism and Pragmaticism. Electronic edition reproducing vols. I–VI ed. Charles Hartshorne and Paul Weiss (Cambridge: Harvard University Press, 1931–1935); vols. VII–VIII ed. Arthur W. Burks (same publisher, 1958). Charlottesville: Intelex Corporation, 1994. https://web.archive.org/web/20210508045931/https://www.textlog.de/7658.html.

Sapp Craig Stuart. "Verovio Humdrum Viewer". https://verovio.humdrum.org/, accessed April 4, 2021.

Invisible Network: Investigations of Graphic Notations through the Theory of Rationality and the Network Theory

Chia-Ling Peng
C.Peng4@newcastle.ac.uk

Keywords
indeterminate music, Solo for Piano, theory of rationality, network theory

Abstract

In Die rationalen und soziologischen Grundlagen der Musik (1921 [Eng.] The Rational and Social Foundations of Music, 1958), Max Weber suggested Western music carries rational features, such as systematic, structural, and functional features. He manifested these characteristics through the arithmetical formation of intervals, harmonic progression, and dynamic movements of music (Weber 1958). In other words, systematic, structural, and functional features show relationships between musical materials, and they present networks within conventional music. However, when music entered its avant-garde phase, it became unpredictable, chaotic, and free-interpreted – does this mean that avant-garde music was no longer rational, and so we cannot discover networks? In this research, I take Cage's Solo for Piano (1957-58, the piano part of Concert for Piano and Orchestra) as an example to present a theoretical framework, which mingles the theory of rationality with the network theory to discover rational features and present networks of indeterminate music. I suggest that rational features and networks are hidden behind fragmental notations, we can discover them through compositional materials and composer's creating intentions. Considering uniqueness of indeterminate music, this paper puts composer's aspect at the first place, and applies the theory of rationality to reflect the essential foundation of Solo for Piano. The theory focuses on relations between individuals' behaviours and their intentions, purposes, and value concepts (Habermas 1984), it may construct a conceptual network by Cage's interpretation of Zen Buddhism, which believes in presenting the being of the world (Suzuki 1964). This concept motivated Cage to

invent a two-step graphic compositional system by using paper imperfections, symbols, and graphs as compositional material (Pritchett 1993). Along with the system, Cage offered performing guidance on conducting the notations, the guidance reinforces the connections between notations. In short, Cage's invention and guidance formed the systematic feature and connected compositional materials and notations. Following this, the network theory considers an aspect of compositional materials. The key concepts are actor, tie, dyad, relation, and network (Wasserman and Faust 2012), they represent compositional materials, connections between materials, a group of two materials and a link between them, individual systems, and inner structure, respectively. This means that I build inner structure (network) by different layer of connections; from the tiniest unit (actor), links between actors (ties), partial individual systems (dyad), the whole individual system (relation), to the inner structure (network). In other words, the network is built gradually, when putting all connections together, the network of Solo for Piano can be revealed. Finally, I utilise Gephi (a visualisation software) to present the result, which looks like a constellation consisting of several groups, including a trunk (common actors, ties, and dyads) and branches (ties, dyads, and relations). When compositional materials do not share tonal goals or tonal values, they present a neutral status, allow performers to interpret them freely; meanwhile, they cohere with one another through the composer's arrangement. This research combines the theory of rationality and the network theory to scrutinise the invisible network of Solo for Piano is constructed by rational features, including compositional materials and composer's intentions, purposes and value concepts.

Bibliography

Habermas, Jürgen. 1984. The Theory of Communicative Action. Boston: Beacon Press.

Pritchett, James. 1993. The Music of John Cage. Music in the Twentieth Century. Cambridge [England] ; New York: Cambridge University Press.

Suzuki, D. T. 1964. An Introduction to Zen Buddhism. Grove Press.

Wasserman, Stanley, and Katherine Faust. 2012. Social Network Analysis Methods and Applications. Cambridge University Press.

Weber, Max. 1958. The Rational and Social Foundations of Music. Southern Illinois University Press.

The Klezmer Archive Project: Documenting Culture Bearers and Human Networks in Traditional Music Communities

Christina Crowder
Klezmer Institute
christina@klezmerinstitute.org

Clara Byom
Klezmer Institute
clara@klezmerinstitute.org

Keywords
Public scholarship, digital archives, culture bearers, cultural knowledge, ontology

Abstract

Klezmer, the instrumental music of Ashkenazic Jews of Eastern Europe, was and continues to be a transnational music based in oral tradition. For decades members of the klezmer community have dreamt of a centralized repository for klezmer tunes and their historical/ethnographic context, but creating such a resource within current archival structures leaves out a critical source of knowledge—klezmer culture bearers. These individuals have a deep understanding of repertoire, history, and folklore that is highly valued within the international klezmer community, but it is only available to the whole community when it is collected and organized. With this in mind, the Klezmer Archive project (KA) aims to create a universally accessible, useful resource for interaction, discovery, and research on available information about klezmer music. Like many folk/ethnic music communities, the international network of klezmer specialists includes musicians with a range of expertise but centers around culture-bearers who collect and share information through performance, teaching, documentation/field work, recordings, and publishing. One of the core goals of the KA project is documenting the networks of human contact through which cultural (musical) information is exchanged in ways that are not reflected by traditional metrics, such as quantity/location of performances,

books published, or recording catalog (all of which have many well-worn paths for documentation).

The Klezmer Archive project is adapting and extending the DoReMus ontology (itself a harmonization of CIDOC CRM and FRBRoo, extending classes and properties specific to musical data, and a set of shared multilingual vocabularies) to document human relationships. The "human relationship" concept allows people to be connected in the data in a way that more accurately reflects how cultural knowledge is exchanged—through mentorships, within families, between bandmates, via private lessons, etc. Human relationship concepts not only allow us to see how tunes migrate throughout the contemporary klezmer and Yiddish song community, but also recognize the importance of certain individuals as culture bearers and teachers that might otherwise be difficult to see in a graph. For example, many of the most renowned klezmer teachers have a relatively small recorded catalog. If we document only the "artifacts" (i.e. recordings, tune books, concerts, etc.) of Ashkenazic expressive culture, we may not recognize that a teacher has taught hundreds of students dozens of tunes, leading to certain tunes becoming part of the contemporary klezmer canon. Developing the capacity to identify these kinds of relationships—of varying duration and intensity, and from the past to the present—will help surface the ways that culture bearers influence generations of musicians within traditional music communities and make this data available for novel data visualization and analysis.

While the Klezmer Archive Project is in its research and development phase, this paper will present the preliminary ontology being developed to describe networks of culture bearers in this community, and will examine the ways that this approach can be used to document networks in many areas of humanities research.

Bibliography

Achichi, Manel, Pasquale Lisena, Konstantin Todorov, Raphaël Troncy, and Jean Delahousse. "DOREMUS: A Graph of Linked Musical Works." In The Semantic Web – ISWC 2018, edited by Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, and Elena Simperl, 11137:3–19. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-030-00668-6_1.

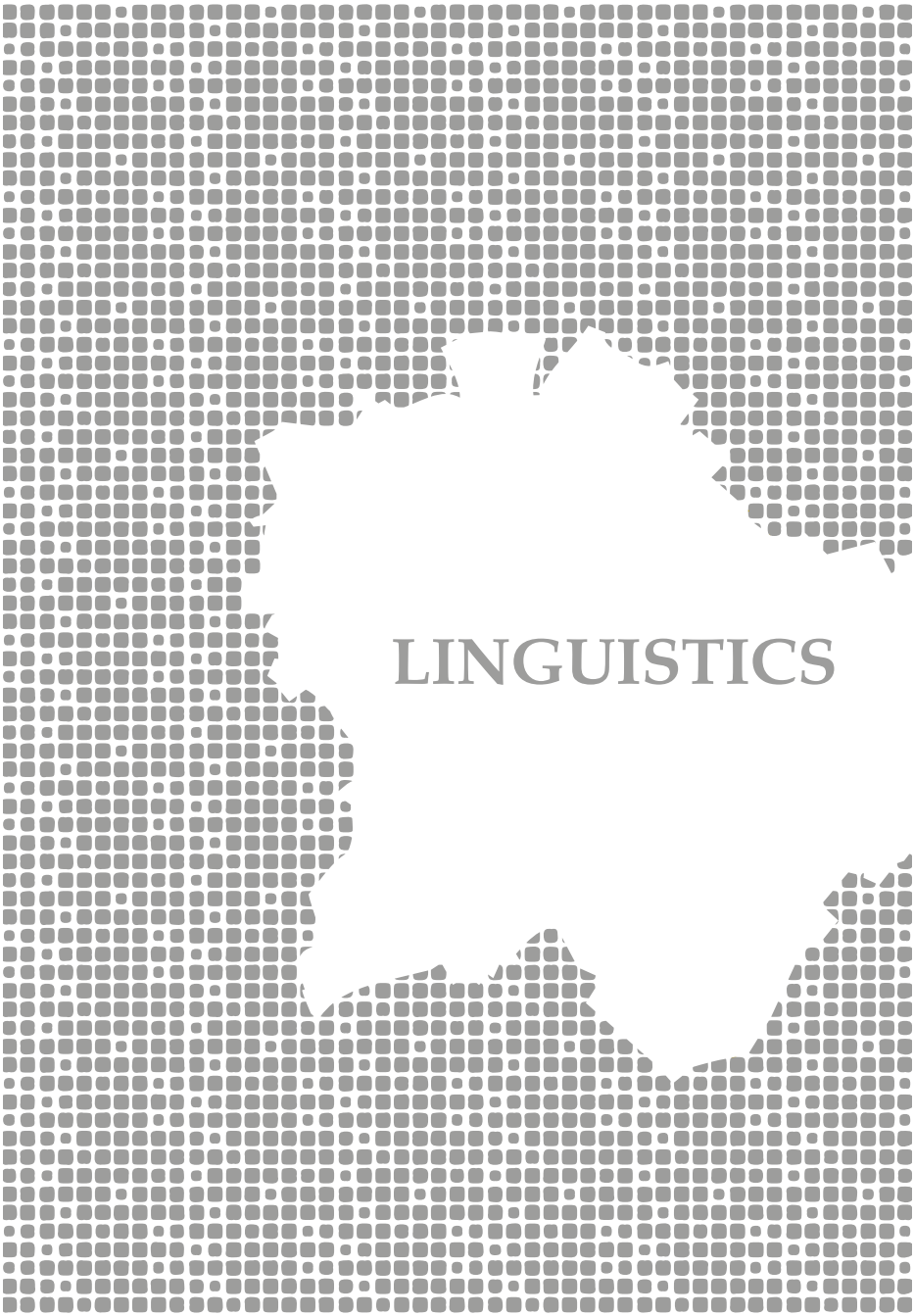
Bakka, Egil. "Revisiting Typology and Classification in the Era of Digital Humanities." ARV Nordic Yearbook of Folklore 75 (2019): 153–79.

Burrows, Toby, and Deb Verhoeven. "Linking and Sharing Data in the Humanities and Creative Arts: Building the HuNI Virtual Laboratory," n.d., 11.

Ilyefalvi, Emese. "The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics." Acta Ethnographica Hungarica 63, no. 1 (June 2018): 209–58. https://doi.org/10.1556/022.2018.63.1.11.

Kranenburg, Peter van, Martine de Bruin, and Anja Volk. "Documenting a Song Culture: The Dutch Song Database as a Resource for Musicological Research." International Journal on Digital Libraries 20, no. 1 (March 1, 2019): 13–23. https://doi.org/10.1007/s00799-017-0228-4.

Weissenberger, Lynnsey K. "Linked Data in Music and Its Potential for ITMA/Traditional Music Web Resources." Brio 55, no. 1 (2018): 52–57. https://doi.org/10.5281/ZENODO.1002056.


LINGUISTICS

BERT helps in sense delineation

Ágoston Tóth
University of Debrecen, Institute of English and American Studies, Department of English Linguistics
toth.agoston@arts.unideb.hu

Esra Abdelzaher
University of Debrecen, Institute of English and American Studies, Doctoral School of Linguistics
esra.abdelzaher@gmail.com

Keywords
BERT, computational linguistics, sense delineation

Abstract
Powerful neural networks are being used in Natural Language Processing (NLP) systems to learn patterns from large corpora automatically and utilize them in solving linguistic tasks. At the heart of these systems, word embeddings store information about the distributional properties of words, the typical contexts in which they appear. Crucially, when two words have similar meanings, their word embeddings will be similar, too. Contextualized word embeddings, which are neural network representations of word tokens in given sentences, can also circumvent the problem of lexical ambiguity.
This study addresses the challenge of sense delineation, which is one of the most challenging tasks for lexicographers (Kilgarriff, 1998) who need to abstract senses from corpus citations (Kilgarriff, 2007). There is evidence that contextualized embeddings (including BERT word representations, Devlin et al., 2019) form distinct clusters corresponding to different word senses (Wiedemann et al., 2019; Schmidt and Hofmann, 2020), making BERT successful at the word sense disambiguation task. This study further examines this evidence in detail. The experiment cites dictionary examples (from Oxford Learners' Dictionary at www.oxfordlearnersdictionaries.com: OLD) for a sample of words (e.g., risk and face). Contextualized embeddings were created to represent each sentence using BERT, and clusters were visualized in 2 dimensions.

Initial results revealed that the senses of the verb and noun forms of the same word appear in different clusters with a large distance separating them. No cases of overlap between the noun and verb senses were detected. Sentences representing the same word sense clustered together with different degrees of similarity to the original OLD senses. Some cases showed 100% match (e.g., the third sense of risk.n formed a single cluster). Other senses formed multiple but close clusters revealing some syntactic variations. Overall, clusters of verb senses were more similar to the original categorization of sentences in OLD than noun clusters. The distribution of noun senses, especially the first sense, was usually scattered (e.g., only about 20% of the sentences representing the first sense of risk.n appear in clusters).

Word embeddings are recent inventions, but they are spreading very quickly in NLP and affect our everyday lives; high-profile applications, including web search and online machine translation are also using them. Linguistic disciplines, both in theory and application, seem to be slow adaptors of deep-learning techniques, lexicography has also been left largely intact. We investigate ways of exploiting information that emerge in artificial neural networks for our lexicographic needs. The results of this study have implications for linguists and lexicographers facing the challenge of sense delineation. In the proposed way, lexicographers get a visual tool to assess the level of similarity between different uses of the headword in selected sentences, and we also get a pool of corpus sentences that (potentially) have the headword in the required sense. Arguments for splitting or lumping senses can also be made using distributional clues.

Bibliography

Devlin, J., Chang, M., Lee, K. & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL, 4171-4186. Minneapolis, Minnesota: Association for Computational Linguistics.

Kilgarriff, A. (1998). The hard parts of lexicography. International Journal of Lexicography, 11(1), 51–54. https://doi.org/10.1093/ijl/11.1.51

Kilgarriff, A. (2007). Word sense disambiguation. In P. Agirre, E., & Edmonds (Eds.), Word sense disambiguation: Algorithms and applications. Springer.

Schmidt, F. & Hofmann, T. (2020). BERT as a Teacher: Contextual Embeddings for Sequence-Level Reward. arXiv preprint arXiv:2003.02738.

Wiedemann, G., Remus, S., Chawla, A. & Biemann, C. (2019). Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings. Proceedings of the 15th Conference on Natural Language Processing, KONVENS 2019, Erlangen, Germany, October 9-11, 2019.

Of Manifestos and Mathematicians: A Case Study on Cross-Topic Identity Profiling Using Ted Kaczynski

Alejandro Jorge Napolitano Jawerbaum
Duquesne University, USA
napolitanojawea@duq.edu

Keywords
Stylometry, Profiling, Forensic linguistics, Cross-topic language analysis, Text analysis, Disciplinary science.

Abstract
This paper focuses on the task of profession detection in terrorist manifestos. The Unabomber is an ideal test case because he was a lone terrorist, had no co-authors, and was a mathematician. We investigate the question of whether the Unabomber's manifesto is more similar to mathematical writings than are other terrorist manifestos not written by mathematicians.
We selected seven manifestos by different terrorists and terrorist organizations, the principal one being the Unabomber. We also included right wing extremists, Hezbollah, the Weather Underground, and several other manifestos from people of diverse professions and backgrounds —with Kaczynski being the only mathematician— as well as the work of twelve different mathematicians unrelated to Kaczynski, with ten papers each, in order to see if the Unabomber could reliably be identified as a mathematician. More precisely, we tested to see if it was possible to identify the Unabomber as the author whose style is closest to mathematical papers, meaning to identify him as an author with a mathematical background. This is why we have excluded manifestos by mathematicians and related academic professions such as physics.
The way we structured the corpus was with the mathematics papers as unknown variables, training the model on the manifestos and using the mathematical papers as a test set while dividing the mathematical papers into two categories, both cleaned of mathematical symbols: Verbose, and non-verbose —separated by word-to-symbol ratio— in order to have a sample representative of the mathematical community. We used JGAAP, an authorship attribution program, to conduct 68 analyses across four features,

word trigrams, character decagrams, punctuation trigrams, and Part-of-Speech pentagrams. Of these analyses, 25 returned consistent results, with no ties for first place and a majority of sample documents associated with the same manifesto. Of these 25 consistent analyses, all of them identified the Unabomber as the author most similar to a mathematician with probability of at least 50%.

When conducting validation in Stylo, the mathematical papers themselves did not cluster, neither by verboseness nor topic, whereas all the manifestos formed a clear cluster far from the mathematical papers. Methods that can be used with different distance metrics gave consistently similar results across consistent metrics; these analyses are not independent of each other. The most successful distance metric was alt intersection.

This is an atypical problem formulation. A more intuitive formulation would involve classifying a manifesto into one of several profession related categories. To do this, we collected samples of philosophy, economics, and medical sciences. The Unabomber Manifesto was typically classified as a mathematical author, but no clear-cut pattern emerged for the other terrorist manifestos. This may be due to, for example, the Weather Underground manifesto being co-authored. Another key issue is the lack of a comprehensive corpus of terrorist manifestos, which hampers this kind of research.

In conclusion, this study provides solid forensic evidence that we can tell with a good probability whether the manifesto was written by a mathematician, which sets a precedent for being able to tell someone's profession from their writing.

Bibliography

Coyotl-Morales R.M, Pineda, L.V., Montes-y Gómez M., and Rosso P. (2006) Authorship Attribution using word sequences. In Proceedings of the 11th Iberoamerican Conference on Progress in Pattern Recognition, Image Analysis, and Applications, CIARP 2006, pp. 844-853, Berlin, Heidelberg. Springer-Verlag.

Johnson, R.C. (2013) Authorship Attribution with Function Word N-Grams. Doctoral dissertation. Nova Southeastern University. Retrieved from NSUWorks, Graduate School of Computer and Information Sciences. (188) https://nsuworks.nova.edu/gscis_etd/188.

Juola, P. (2009). "JGAAP: A System for Comparative Evaluation of Authorship Attribution." Journal of Digital Humanities and Computer Science 1(1)

Juola, P. (1997) 'What Can We Do With Small Corpora? Document Categorization Via Cross-Entropy'. In: Proceedings of an Interdisciplinary Workshop on Similarity and Categorization. Edinburgh, UK, Department of Artificial Intelligence, University of Edinburgh.

Stamatatos, E. (2013) On the robustness of authorship attribution based on character n-gram features. Journal of Law & Policy 21 (2013) 427–439

Wyner, A. J.: 1996, 'Entropy Estimation and Patterns'. In: Proceedings of the 1996 Workshop on Information Theory.

CLARIAH-EUS: Building a Cross-border CLARIAH Node for the Basque Language

Begoña Altuna
HiTZ Center - Ixa and Computer Languages and Systems department, University of the Basque Country UPV/EHU
begona.altuna@ehu.eus

Mikel Iruskieta
Euskara Institutua, HiTZ Center - Ixa and Didactics of Language and Literature department,
University of the Basque Country UPV/EHU.
mikel.iruskieta@ehu.eus

Ainara Estarrona
HiTZ Center - Ixa and Computer Languages and Systems department, University of the Basque Country UPV/EHU
ainara.estarrona@ehu.eus

Aritz Farwell
HiTZ Center - Ixa and Computer Languages and Systems department, University of the Basque Country UPV/EHU
aritz.farwell@ehu.eus

Jose Maria Arriola
HiTZ Center - Ixa and Basque Language and Communication department, University of the Basque Country UPV/EHU
josemaria.arriola@ehu.eus

Jon Alkorta
HiTZ Center - Ixa and Basque Language and Communication department, University of the Basque Country UPV/EHU
jon.alkorta@ehu.eus

Xabier Arregi
HiTZ Center - Ixa and Computer Languages and Systems department, University of the Basque Country UPV/EHU
xabier.arregi@ehu.eus

Keywords
Basque, CLARIAH-EUS infrastructure, research, networking, infrastructure set-up

The total population of Basque speakers is modest when compared to those of its neighbouring languages and, due to the administrative divisions within the Basque-speaking area and the fact that the language is not official across the entire region, there was no centralised infrastructure for the interaction of researchers interested in the language or in Basque culture and society. Gaining momentum in parallel to INTELE (INfraestructura de TEcnologías del LEnguaje), a strategic network to create the CLARIN-ERIC (Hinrichs and Krauwer, 2014) and DARIAH-ERIC (Edmond et al., 2017) nodes for Spain, CLARIAH-EUS was initially proposed by HiTZ, the Basque Center for Language Technology at the University of the Basque Country (UPV/EHU), to structure a response to the needs of humanists and social scientists, who were asking for help in linguistic data processing for Basque with ever greater frequency.

The CLARIAH-EUS project aims to create and develop the CLARIN-ERIC and DARIAH-ERIC infrastructures for the Basque language in the Basque Country and abroad in order to offer 21st-century digital data, resources, tools and services to researchers in the Humanities and Social Sciences (and beyond) (Bel et al., 2016). Accordingly, we are building a network of public and private stakeholders and have defined a roadmap for service creation and distribution for the coming years. In our presentation, we will address the CLARIAH-EUS creation process as an initiative born from the actual needs of Humanities and Social Sciences researchers.

The researcher net was consolidated in the CLARIAH-EUS project presentation and design workshop (2021), in which 30 researchers from 9 public and private universities and institutions in different countries listed and analysed the needs and interests that an efficient infrastructure should be addressing. Since then, it has received support from several public and private institutions and individuals in the areas of linguistics, sociolinguistics, language technologies, sociology, history and law. At the moment, CLARIAH-EUS is backed by a working group of eight institutions and over 130 people have signed its foundational manifesto.

CLARIAH-EUS is now in its nascent stage and the network is prioritising networking, fundraising and project definition tasks. As to project scheduling and organisation, As to project scheduling and organisation,

in the different areas of Humanities and Social Sciences (and beyond), either by creating them or by linking and integrating existing resources and tools for Basque to the European CLARIN and DARIAH repositories. On the tools side, the priority is to make the basic modular text (and speech) processing pipeline available to the entire community. On the data side, the schedule foresees awarding extra attention to each of the single Knowledge areas (linguistics, literature, education, journalism, sociology, history and law) in CLARIAH-EUS consecutively. More information on the CLARIAH-EUS node and its services is available at http://ixa2.si.ehu.eus/clariah-eus/.

Bibliography

Núria Bel, Elena Gonzalez-Blanco and Mikel Iruskieta. 2016. CLARIN Centro-K-español. Procesamiento del Lenguaje Natural 57: pages 151-154. http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5350

Jennifer Edmond, Frank Fischer, Michael Mertens and Laurent Romary. 2017. The DARIAH ERIC: Redefining Research Infrastructure for the Arts and Humanities in the Digital Age. ERCIM News, (111). https://ercim-news.ercim.eu/en111/special/the-dariah-eric-redefining-research-infrastructure-for-the-arts-and-humanities-in-the-digital-age

Erhard Hinrichs and Steven Krauwer. 2014. The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 1525–1531, Reykjavik, Iceland. European Language Resources Association (ELRA). https://aclanthology.org/L14-1356/

The network of personifications in online texts: case studies from the PerSE corpus

Gábor Simon
Eötvös Loránd University Budapest
simon.gabor@btk.elte.hu

Keywords
personification, cognitive linguistics, corpus, network of meaning

Abstract

The systematic, corpus-based investigation of personifying expressions is a relatively new direction in contemporary cognitive linguistics (Dorst 2011, Galac fc.). To enhance the improvements in this field, a new language resource has been developed for the corpus-driven analysis of personifications in Hungarian, following the methodological proposal of identifying personification in English (Dorst–Mulder–Steen 2011). The PerSE corpus consists of semi-automatically pre-processed online texts with manual annotation of personifying expressions (Simon fc.). This corpus also makes it possible to observe the textual organisation of personifying meaning both at the linguistic and the conceptual level of the discourse.

The aim of the paper is to focus on how individual personifications form a network structure in Hungarian online texts. Based on the annotated material of the PerSE corpus I explore (i) the micro-network of the multiword expressions identified as personifications, (ii) the macro-network of personifying expressions within the network structure of the mentally represented text (Givón 1993), and (iii) the conceptual network of personifications, i.e., the connections between the source and the target domains. An additional dimension of the analysis concerns the genre-specific patterns of these networks comparing online political articles and car reviews.

The study contributes to digital humanities in two ways. On the one hand, it sheds new light on the association of personifying meaning generation with specific parts of online articles. In other words, the potential textual

colligation (Hunston 2001) of personification can become observable. On the other hand, the results can be used to refine the semantic network models and computer-assisted text analysis.

Bibliography

Dorst, Aletta G. 2011. Personification in discourse: Linguistic forms, conceptual structures and communicative functions. Language and Literature 20 [2]: 113 – 135.

Dorst, Aletta G. – Mulder, Gerben – Steen, Gerard J. 2011. Recognition of personification in fiction by non-expert readers. Metaphor and the Social World 1 [2]: 174 – 201.

Galac, Ádám forthcoming. Megszemélyesítő konceptualizációk a látás, hallás és szaglás fogalmi tartományában: kontrasztív empirikus vizsgálat [Personifying conceptualizations in the conceptual domain of vision, hearing and olfaction: a contrastive empirical study]. Jelentés és Nyelvhasználat 2022.

Hunston, Susan 2001. Colligation, lexis, pattern, and text. In: Scott, Mike–Thompson, Geoff (eds.): Patterns of Text: in honour of Michael Hoey. Amsterdam, Philadelphia: John Benjamins, 13–34.

Simon, Gábor forthcoming. Identification and Analysis of Personification in Hungarian: The PerSECorp project. In: Proceedings of the LREC 2022 conference.

The edition's influence on the results of computer-based authorship attribution of 19th century Hungarian novels

Mária Regina Tímári
ELTE BTK, Departement of Digital Humanities
timari.maria@btk.elte.hu

Keywords
authorship attributoin, different editions, ortography, spelling

Abstract
While authorship studies in the field of computational stylistics are becoming increasingly widespread, it is still widely accepted that there are unique patterns of language use, so-called authorial "fingerprints", the metaphor of this term can falsely suggest that patterns specific to a particular author can be objectively extracted from texts. However, the construction of this authorial fingerprint is a much more complex, creative digital humanities task, which constructs a 'pattern', always interpretable only in comparison with other authorial texts, on the basis of a selection and combination of linguistic markers that can be interpreted statistically in the text and then of various similarity calculations based on these markers.
Given the size of the corpus of the studied texts and the complexity of the linguistic markers and similarity calculations, this is now unthinkable without the use of computer-based algorithms. However, despite the fact that stylistometric and computer-based authorship analyses are becoming more and more widespread, and that the last two decades have witnessed intensive technical and methodological changes in these fields, they are still not widely used on Hungarian texts, so further studies are needed to determine which methods, distance measures and stylistometric tools would be most accurate for Hungarian language. It is also unclear, for example, how the linguistic condition of the texts under study influences the results of the research. In my research, therefore, I am trying to establish whether there is any difference in the authorship attribution of older editions of the same texts and the grammatically normalised, ortographicly consolidated versions of more recent editions. In my research, I will present the results of authorship analyses of older and more recent editions of 19th century Hungarian novels.

Bibliography

Coyotl-Morales R.M, Pineda, L.V., Montes-y Gómez M., and Rosso P. (2006) Authorship Attribu-
Chaski, C.E. (2005). „Who's at the keyboard? Authorship attribution in digital evidence investigations" International Journal of Digital Evidence, 4(1).
Harald Baayen, Hans van Halteren, Anneke Nejit, Fiona Tweedie, „An experiment in authorship attribution", JADT 2002 : 6es Journ´ees internationales d'Analyse statistique des Donn´ees Textuelles. Conference Paper. 2002.
Timári, M. R., Bajzát, T. B., & Palkó, G. (2021). 19. századi magyar regényeken végzett kísérletek a magyar nyelvű szerzőazonosítás leghatékonyabb távolságméréseinek megtalálására.

Stylometric Authorship Attribution in Seychellois Creole

Patrick Juola
Duquesne University, USA
juola@mathcs.duq.edu

Alejandro J. Napolitano Jawerbaum
Duquesne University, USA
napolitanojawea@duq.edu

Keywords
Stylometry, authorship attribution, creole linguistics, individual variation.

Abstract

Stylometry (Ainsworth and Juola 2018), the computational study of writing style, has proven itself to be a practical method of answering questions of authorship in a wide variety of languages. However, previous research has focused on non-creole languages. Creole languages are "new languages that develop out of a need for communication among people who do not share a common language" (Siegel 2008) In contrast to older languages, they are often considered to be unusual, in having, for example, smaller vocabularies (Robinson, 2008) and less complex structure (McWhorter 1998). More specifically, they show (1) minimal use of inflection, (2) lack of tone used to contrast monosyllables or make grammatical distinctions, and (3) semantically regular derivation (McWhorter 1998). It is not clear whether and how these linguistic properties would affect the statistical methods used to infer authorship.

In particular, a widely accepted theory of authorship analysis (Coulthard, 2004) states that authors can be distinguished by "cumulatively unique rule-governed choices" among the available options in the language. With fewer lexical and syntactic options, are the opportunities for choice reduced? And is this reduced choice reflected in the performance of standard stylometric methods? This paper presents an analysis of a newly collected corpus of Seychellois Creole, an official language of and the most used language in the Republic of Seychelles.

We extracted 100 speeches from Verbatim, the official minutes of the sittings of the National Assembly of Seychelles, representing ten samples each of ten different speakers/authors. These speeches ranged from 814 to 8079 words. We then analyzed them in five different but typical ways using the JGAAP software package).

Random guessing would be expected to identify ten samples correctly. Our five analyses scored 33, 57, 62, 74, and 79 correct samples, all a highly significant ($p < 0.0001$) improvement on random guessing.

This shows that, despite the differences between older languages and creoles, differences that might be expected to mask the authorial signal, authorship attribution is still easily possible with high accuracy in creole languages. In fact, due to a proposed greater interspeaker variability (the "creole continuum", Bickerton 1973) in such languages, it may in fact be easier to do AA in creole languages. In addition to developing a better, larger, and more balanced corpus, planned future work includes studying other creoles (such as Haitian creole), other genres within Seychellois Creole, and comparing authorship attribution performance between the lexifying source language (in this case, French) and the creoles.

Bibliography

Ainsworth, Janet, and Patrick Juola. "Who wrote this: Modern forensic authorship analysis as a model for valid forensic science." Wash. UL Rev. 96 (2018): 1159.

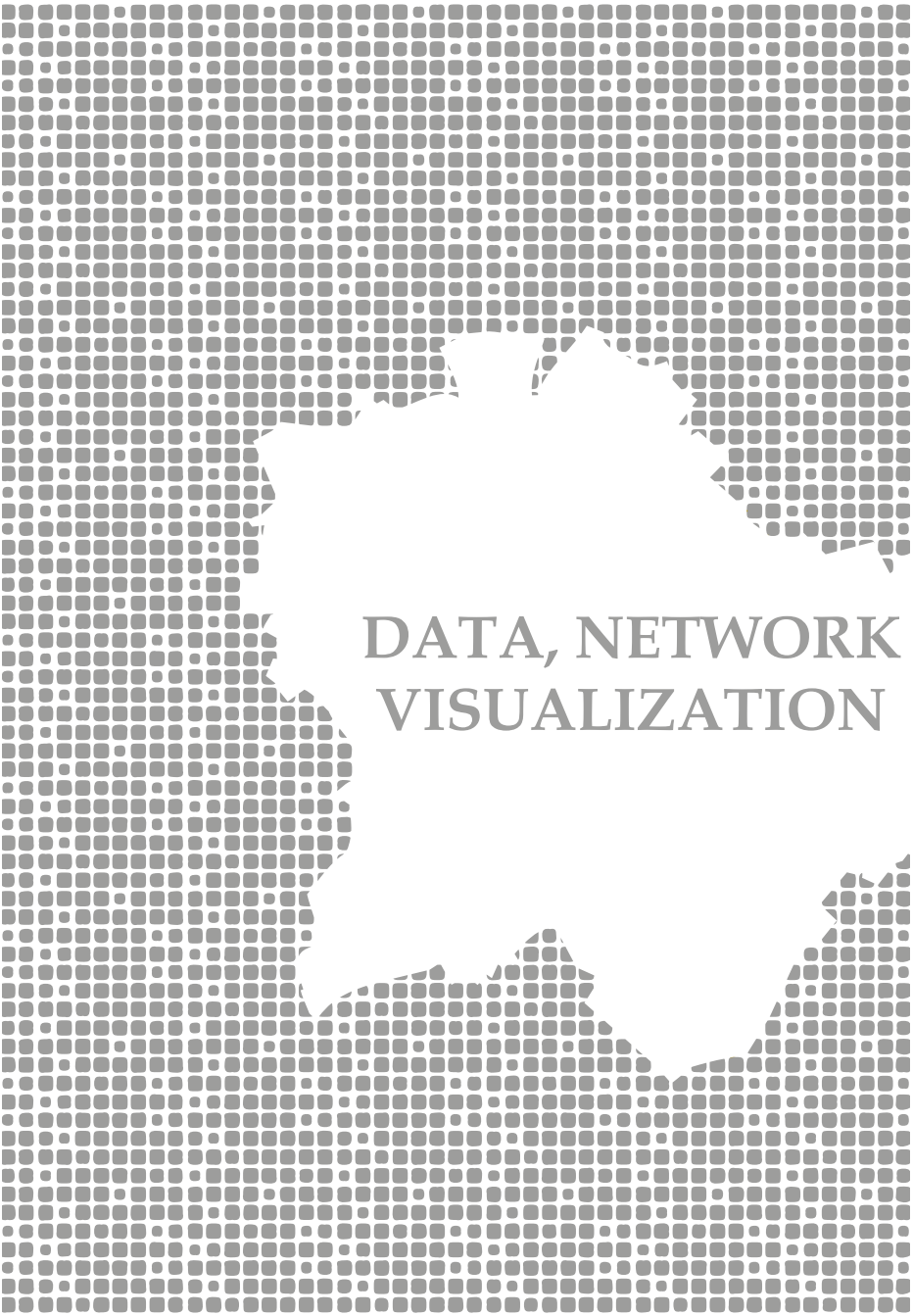
Bickerton, Derek. "The nature of a creole continuum." Language (1973): 640-669.

Coulthard, Malcolm. "Author identification, idiolect, and linguistic uniqueness." Applied linguistics 25, no. 4 (2004): 431-447.

McWhorter, John H. "Identifying the creole prototype: Vindicating a typological class." Language (1998): 788-818.

Robinson, Stuart. "Why pidgin and creole linguistics needs the statistician: Vocabulary size in a Tok Pisin corpus." Journal of Pidgin and Creole Languages 23, no. 1 (2008): 141-146.

Siegel, Jeff. The emergence of pidgin and creole languages. Oxford University Press, 2008.


DATA, NETWORK
VISUALIZATION

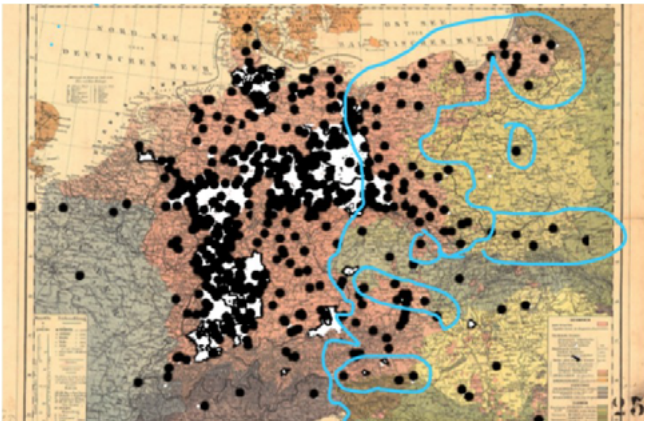
Visualizing Multilingual Literary Networks in Monolingual 19th Century Literary History: Reflections and Comparisons

Jana Mende
Martin-Luther-Universität Halle-Wittenberg
Jana-katharina.mende@germanistik.uni-halle.de

Keywords
multilingualism, literary history, actor-networks, literary geography, 19th century literature, European literature

Abstract
Literary history and literary history writing in Europe in the 19th century „is organized along national lines" (Casanova 2004, XI) – lines defined by culture and, above all, language. Mining 19th century sources – literary historiographies by Gervinus (1844), Gottschall (1855) or Scherer (1883) or literary dictionaries by Pataky (1898) and Brümmer (1913) – for multilingual authors and texts is mostly unsuccessful because those volumes tell a monolingual story of German literature.

However, even those monolingual documents contain data that can provide information about 'hidden' multilingual authors within national literatures. Usually, biographical literary dictionaries like Pataky's Frauen deutscher Feder only serve to give biographical information about often forgotten (female) authors. At the same time, those dictionaries contain the places of residence and travels of authors which give insights into their linguistic surroundings. By mapping places of residence and combining geographical data



with linguistic data about multilingual regions multilingual literary neighborhoods become visible – like those in Bratislava (Preßburg/Pozsony) during the 19th century (Figure 1). Different tools and visualization platforms like Recogito or nodegoat (Bree/Kessel 2013) allow for the creation of different networks, geographical clusters, social relations, and actor-networks which include entities like persons, publications, and languages. Visualizations with nodegoat (Bree/Kessels 2013) allow further investigations of those neighborhoods as networks, collaborating through publications, letters, and texts. Instead of focusing on individual authors, the linguistic and literary communities are on the foreground of that analysis. Relations within those communities could be maintained through shared linguistic resources and are visible through joined publications, anthologies, correspondence. Different types of networks aim to show whether and how language use, proximity, and multilingualism overlap.

Conceptually, certain forms of language contact like translations, translators, multilingual actors, multilingual publications, should be part of these networks. One part of my contribution discusses the different possibilities to visualize those forms within multilingual literary networks with different tools. The second part reflects on the categories and decisions made to construct different types of networks and their usefulness to analyze historical multilingual literature.

Bibliography

Bree, P. van, Kessels, G., (2013). nodegoat: a web-based data management, network analysis & visualisation environment, http://nodegoat.net from LAB1100, http://lab1100.com

Brümmer, Franz (1913): Lexikon der deutschen Dichter und Prosaisten vom Beginn des 19. Jahrhunderts bis zur Gegenwart. Eight volumess. Leipzig.

Casanova, Pascale (2004): The World Republic of Letters. Cambridge, Massachusetts, London, England. (= Convergences inventories of the present).

Pataky, Sophie (1898): Lexikon deutscher Frauen der Feder. Eine Zusammenstellung der seit dem Jahre 1840 erschienenen Werke weiblicher Autoren, nebst Biographieen [sic!] der lebenden und einem Verzeichnis der Pseudonyme. Two volumes. A-L; M-Z. 1. Aufl. Berlin.

Recogito, an initiative of Pelagios Commons, http://recogito.pelagios.org/ (accessed 15 June 2022)

Scherer, Wilhelm (1883): Geschichte der deutschen Literatur. Berlin.

Lifting the veil of Levantine cosmopolitanism: diplomatic networking in Egypt, 1873-1914

Gert Husken
Ghent University & Université libre de Bruxelles, Belgium
Gert.Huskens@UGent.be

Christophe Verbruggen
Ghent University, Belgium
Christophe.Verbruggen@UGent.be

Jan Vandersmissen
Ghent University, Belgium
Jan.Vandersmissen@UGent.be

Julie M. Birkholz
Ghent University, & KBR- Royal Library of Belgium, Belgium
Julie.Birkholz@UGent.be

Keywords
diplomatic networks - data visualization - cosmopolitanism - socio-cultural history of diplomacy - network analysis

Abstract
In this paper we explore through a combination of network visualizations, network analysis and historical research the individual longitudinal trajectories of regional elites in consular and diplomatic networks. The dataset, integrated into a online database powered by nodegoat, comprises a total of around 1300 actors which were collected from five categories of sources, namely 1) the diplomatic and consular sections of the Almanach de Gotha, 2) national annuaries and directories of countries with consular and diplomatic representation in Egypt, 3) Egyptian annuaries and directories, 4) consular and diplomatic archives 5) a selection of travelogs, each systematically analyzed in terms of career changes and promotions, organizational memberships and institutional engagements of the relevant diplomatic and consular actors, from 1873, year of the earliest useful Egyptian annuary to the pivot date of the beginning of WWI.

We conceptualize the careers of actors holding an office in the diplomatic and consular corps through a continuously shifting web of relations or field. We investigate a selection of individual trajectories (e.g. considering an ego's position over time), and relate these to the general changes of diplomatic life in Egypt consider both qualitative knowledge - the engagement of diplomats in local socio-cultural, philanthropic urban sociability of Alexandria and Cairo and companies; and quantitative network structure measures. By focusing on the Belgian case, a small state that combined the integration of local elites into its consular apparatus with the development of a career diplomacy-based system, a unique look is given on the history of diplomatic presence in this era. This combination of these approaches allows us to make contributions to our understanding of both this time period but also to theories on (diplomatic) networking. Specifically we work to refine views on the practice of exclusion and inclusion as defined in early 21th-century descriptions of Middle Eastern cosmopolitanism, among others by Sami Zabaida, Will Hanley and Edhem Eldem, Ulrike Freitag and others, and contribute to the debate on cosmopolitanism and the socio-cultural history of diplomacy. We also reflect on the epistemological grounds and consequences of both the bias in the sources that are integrated into the dataset and the usage of network analysis, visualizations and historical research of the diplomats' trajectories, asymmetrical relationships and inequalities.

Bibliography

Balázs Indig, Árpád Knap, Zsófia Sárközi-Lindner, Mária Timári, and Gábor Palkó. 2020. The Eldem, Edhem. 'Istanbul as a Cosmopolitan City: Myths and Realities'. In Istanbul as a Cosmopolitan City Myths and Realities, edited by Ato Quayson and Girish Daswani, 212–30. Wiley-Blackwell, 2013.

Dumoulin, Michel, and Catherine Lanneau. La biographie individuelle et collective dans le champ des relations internationales. Bruxelles: Presses Interuniversitaires, 2016.

Hanley, Will. 'Grieving Cosmopolitanism in Middle East Studies'. History Compass 6, no. 5 (2008): 1346–67. https://doi.org/10.1111/j.1478-0542.2008.00545.x.

Hanley, Will. Identifying with Nationality: Europeans, Ottomans, and Egyptians in Alexandria. New York: Columbia University Press, 2017.

Mösslang, Markus, Torsten Riotte, and German Historical Institute in London, eds. The Diplomats' World: A Cultural History of Diplomacy, 1815-1914. Studies of the German Historical Institute London. Oxford ; New York: Oxford University Press : German Historical Institute London, 2008Hanley, Will. 'Grieving Cosmopolitanism in Middle East Studies'. History Compass 6, no. 5 (2008): 1346–67. https://doi.org/10.1111/j.1478-0542.2008.00545.x.

Zubaida, Sami. 'Middle Eastern Experiences of Cosmopolitanism'. In Conceiving Cosmopolitanism: Theory, Context and Practice, edited by Steven Vertovec and Robin Cohen, 32–41. New York: Oxford University Press, 2003.

Mapping the neo-avant-garde: visual network analysis of Flemish periodicals (1949-1970)

Jan Lampaert
Ghent University
Jan.Lampaert@UGent.be

Keywords
Network visualization, visual network analysis, literary studies, periodical studies

Abstract

The 1950s and 60s constitute one of the most exciting and turbulent episodes in Flemish literary history. A radically new poetry and poetics emerged and flourished within a myriad of neo-avant-garde periodicals. The experimental poetry they published can be labelled as neo-avant-garde as it entailed the re-appropriation of literary techniques and strategies of the historical avant-garde (e.g. Dadaism, Surrealism, Expressionism). This presentation argues for visual network analysis (VNA) (Jacomy 2021, Venturini et al. 2021) as a method for charting and tracing literary dynamics in terms of centres and peripheries. As the analysis is applied to the full range of literary magazines published in Flanders between 1949 and 1970, it becomes possible to attain a finer grasp of the intricate and rapidly evolving post-war literary landscape. The global affiliation network consists of Flemish literary periodicals and their contributing poets (128 periodicals, 3.000 authors, and 17.500 contributions). The network visualization is generated in Gephi, an open source network exploration tool. In this paper, I will discuss the decisions made during the construction of the network as well as the specific methodology of VNA. VNA is the practice of analyzing networks by visual means, rather than mathematical metrics (Jacomy 2021, 5). The analysis is informed by the literary historical understanding of the way postwar experimental poetry develops in Flanders: rise (1949-1954), breakthrough and consolidation (1955-1960) and decline (1961-1970). VNA allows me to complement, refine and correct this commonly assumed development. Furthermore, the increasing associations with the international neo-avant-garde during the 60s are visualized through an additional geospatial network.

Finally, little research has been conducted on the post-experimental periodicals of the late 60s. However, the corresponding network slices reveal a sudden boom of these "marginal" mimeographed magazines. The subsequent formation of a significant and dense cluster alongside the established periodicals suggests a vibrant underground scene.

This paper demonstrates the distinctive heuristic force of VNA (Venturini et al. 2021, 13). It facilitates the exploration of the landscape of post-war Flemish literary periodicals by means of the intuitive spatial metaphors of centre and periphery. The relatively large scale of the analysis accounts for the complexities and rapid evolutions of the periodical network. Moreover, VNA not only reveals global patterns and notable nodes, but also brings into focus local configurations and marginal periodicals (Venturini 2012). Consequently, it offers possibilities for refining existing research and provides new perspectives.

Bibliography

Jacomy, Mathieu. Situating Visual Network Analysis. 2021. Aalborg University, PhD thesis. https://vbn.aau.dk/en/publications/situating-visual-network-analysis. Accessed 2022-08-13.

Venturini Tommaso. "Great expectations: méthodes qualiquantitative et analyse des réseaux sociaux. in L'Ere Post-Media. Humanités Digitales et Cultures Numériques, edited by J-P Fourmentraux, 2012, pp. 39–51.

Venturini Tommaso, Mathieu Jacomy, and Pablo Jensen. "What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts" in: Big Data & Society, 1, 2021.


ONLINE

Digitization of a Japanese Christian Text from the Sixteenth Century

So Miyagawa
National Institute for Japanese Language and Linguistics
runa.uei@gmail.com

Sophie Neutzler
Ruhr University Bochum (Mie University)
anna.neutzler@rub.de

Keywords
digital transcription, HTR, international collaboration, missionary linguistics, Jesuits

Abstract

At times there are still bibliographical rarities brought to light that must be made accessible and added to the written cultural heritage. Only a few copies of the Jesuit prints in 16th – 17th century Japan (so-called Kirishitanban) are preserved today, owing to harsh persecution of the works of the Jesuit mission press in Japan. Thus, the first discovery at the Herzog August Bibliothek Wolfenbüttel in Germany of such a Japanese Jesuit print by Katja Triplett (2018), a Japanese translation of Contemptus Mundi (1596), captured the attention of the public. Yet, a diplomatic transcription was not commenced until our application of Transkribus (Muehlberger et al. 2019, Colutto et al. 2019), a graphical user interface program for handwritten text recognition (henceforth HTR). This software utilizes its artificial neural network model based on the LSTM (long short-term memory) model. We trained several HTR Models via manually transcribed data to recognize the Japanese text written in Latin script (Romaji) in a Japanese-German research group uniting the expertise in Missionary Linguistics and Digital Humanities. The working cycle in Transkribus (e.g., Jaillant 2022, Fischer et al. 2020) begins with applying and correcting an automated layout analysis with text region, baselines, and polygons. Next, the HTR model is trained with manually transcribed data, the so-called ground truth. Resulting models can be re-trained by more data or used as a base model for new HTR models. Via the iterative application of this working cycle

(creating ground truth data – training model – applying new recognition – creating new ground truth data via error-modification), we achieved an HTR model with 1.00% CER (character error rate). This model significantly accelerated the work process compared to a manual transcription of such a comprehensive work as the Contemptus Mundi with its over 450 pages. The result of our digital transcription in Japanese-German collaboration is intended to be published by the National Institute for Japanese Language and Linguistics (NINJAL) in its open access digital archive repository. This repository has already published the facsimile images and digital transcripts of the romanized publications of Heike Monogatari (The Tale of the Heike), Esopo no Fabulas (Aesop's Tales), and Kinkushu (Proverbs) by the Jesuits in 17th century Japan. We aim to improve this repository by adding the Contemptus Mundi in a parallel-text edition showing facsimile and transcription alongside the visualized user interface. Additionally, the TEI XML format of the diplomatic edition of Contemptus Mundi will be provided in the repository to promote FAIR (Findable, Accessible, Interoperable, Reusable) standards. The Latin script used in Contemptus Mundi can serve to illuminate the nature of borrowing in the first phase of Japanese loanword history initiated by the Jesuits (Irwin 2011). While many modern editions of these Jesuit works fail to preserve the original Latin script presenting the text solely in a transliteration in Japanese script, it is our aim to add this bibliographical rarity, which must not be lost in the course of time, in its original form to the written cultural heritage of our global community.

Bibliography

Colutto, Sebastian / Kahle, Philip / Guenter, Hackl / Muehlberger, Guenter (2019): "Transkribus. A Platform for Automated Text Recognition and Searching of Historical Documents." In: Proceedings of 15th International Conference on eScience: 463–466. doi: 10.1109/eScience.2019.00060.

Fischer, Andreas / Liwicki, Marcus / Ingold, Rolf (2020): Handwritten Historical Document Analysis, Recognition, And Retrieval – State Of The Art And Future Trends. Singapore: World Scientific Publishing Company.

Irwin, Mark (2011): Loanwords in Japanese. Studies in Language Companion Series 125. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Jaillant, Lise (2022): Archives, Access and Artificial Intelligence. Working with Born-Digital and Digitized Archival Collections. Bielefeld: transcript Verlag.

Muehlberger, Guenter, et al. (2019): "Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study." In: Journal of Documentation 75.5: 954–976.

Triplett, Katja (2018): "The Japanese Jesuit Contemptus Mundi (1596) of the Bibliotheca Augusta: A Brief Remark on a New Discovery." In: Journal of Jesuit Studies 5.1: 123–127.

Creating knowledge of new architecture: Socio-semantic analysis of the magazine Arhitektura (1931-1934)

Tajana Jaklenec
University of Zagreb, Faculty of Architecture
tjaklenec@arhitekt.hr

Željka Tonković
University of Zadar, Department of Sociology
ztonkovi@unizd.hr

Keywords
Arhitektura; social network; semantic network; new architecture

Abstract

During 20th century the main media that represented actual architectural discourse were, besides exhibitions and lectures, architectural magazines. Architectural magazines are media that simultaneously shape and reflect architectural and planning production, they promote technical knowledge and profession. In a broader sociological sense, they represent a network in which complex personal, social, and temporal relations are cross-linked. Starting from the assumption of the magazine as a netdom (Harrison C. White, 2008), the aim of the research is to see how semantic network i.e., expressions in architectural texts, and social network i.e., actors' social ties of magazine Arhitektura jointly create the new architecture of the former Yugoslavia. The magazine Arhitektura (1931-1934) was published in Ljubljana ten years after Le Corbusier reviled five points of new architecture, firstly published in the artistic magazine L'Esprit Nouveau (1921), and later in the collection of essays Vers une architecture (1923). The aim of the magazine Arhitektura and its founders was "combining all available mental and material resources that will create the preconditions for the great construction culture of the Yugoslavia, whose main role will be played by the magazine itself and direct a diverse image of the architecture of the Yugoslavia." The main idea of this presentation is to show a methodological approach that combines quantitative data analysis and social network analyses which can be later used in research with bigger corpora of magazines. The first part of the presentation

focuses on the process of data sources, data structuring, data modeling, and data visualization, which are all in the service of digital humanities. It relies on the selected collection of problem texts from the magazine Arhitektura. This corpus is especially challenging because the data sources are written in Slovenian, Croatian, and Serbian, and the fact that is hard to find an open-source tool that works in the Croatian language. After collecting and pre-processing the data, the network analysis begins. Based on the corpus of texts from the magazine Arhitektura transformed into a database, three types of networks were mapped. The first type is a social network that reveals the structure of the actors involved in the magazine Arhitektura and the production of the new architecture. The second type is a semantic network that reveals the concepts used in the magazine, that is, it reveals the cultural structure. This analysis was done at two levels: the individual semantic networks of the authors, and the union semantic network that represents the semantic network of the magazine Arhitektura. The third network is the network of concepts that reveals the connection between authors and certain concepts, that is, cultural constructs that they share among themselves. This last type of network is in fact a socio-semantic network because it superimposes social and cultural structure. Visual presentation of the network will show used concepts of new architecture as they tend to concentrate in communities that share the same scope and ideas, and that textual materials play important role in the production, display, and reception of architecture.

Bibliography

Basov N., Lee J.-S., Antoniuk A. (2016). Social Networks and Construction of Culture: A Socio-Semantic Analysis of Art Groups. Complex Networks & Their Applications V, 693: 785-796. Springer International Publishing.

Basov, N., Breiger, R. & Hellsten, I., 2020. Socio-semantic and other dualities. Poetics, 78(Discourse, Meaning, and Networks: Advances in Socio-Semantic Analysis), pp. 1-12.

Mische, A., 2011. Relational Sociology, Culture, and Agency. U: J. Scott & P. J. Carrington, ur. The SAGE Handbook of Social Network. London–Thousand Oaks–New Delhi: Sage Publications, pp. 80-97.

White H. C. (2008). Identity and Control. How social formations emerge. Princeton University Press.


DATABASE,
RESEARCH DATA

Semantic networks in ELTEdata

Ádám Sebestyén
Department of Digital Humanities, Eötvös Loránd University
akkon88@gmail.com

Keywords
semantic networks, semantic database, data visualization

Abstract

ELTEdata, as a semantic database, developed by the Centre for Digital Humanities at the Eötvös Loránd University, organizes the sources of prosopographical, bibliographical, and other historical research groups into a semantic data network, and publishes them. Its structure and operation resemble other semantic prosopographies, such as the biographical database of the Austrian Academy of Sciences (APIS), or the FactGrid-platform, which stores and computes historical data. Like this last one, ELTEdata follows Wikidata's data structure, however, ELTEdata is connected to Wikidata based on the semantic statements and the entities, too. The database consists of items and properties with a unique identifier. Every semantic statement can be described as key-value pairs, which match a property with one or more entity values.

The SPARQL semantic query language enables complex search and visualization on maps or timelines. This feature makes it possible to structure a large data set and seems to be useful during the description of different networks. The fusion with the external databases plays an important role in the formation of a semantic bibliography, therefore it is relevant to match some properties to their variants, contained in the interfaces of structured metadata. My presentation will focus basically on data visualization. Because the uploading of data is continuous (currently, the database contains more than 10000 items), it is possible to create and analyze networks in this quite large data set. First of all, I will briefly present the available research groups in the database, with particular attention to the represented networks. I will also present the hierarchies between the created entities. Then, I will demonstrate the various kinds

of visualizations, focusing on graphs and other tools, in order to represent familial and professional relationships. My purpose is to describe those features, which can enlighten new aspects of the data set from a semantic point of view.

Bibliography

Christophe VERBRUGGEN, Hans BLOMME, Thomas D'HAENINCK, Mobility and movements in intellectual history. A social network approach. In: The Power of Networks – Prospects of Historical Network Research, ed. Florian KERSCHBAUMER, Linda von KEYSERLINGK-REHBEIN, Martin STARK, Marten DÜRING, Routledge, 2020, 125-150.

Olaf SIMONS, Imagine a Graph Query Helper for Graph Databases, Blog-article, date of access: 2022.09.05. https://blog.factgrid.de/archives/2636

Olaf SIMONS, How to map itineraries on FactGrid – and Robinson Crusoe's eight voyages, Blog-article, date of access: 2022.09.05. https://blog.factgrid.de/archives/2475

Literarybibliography.eu: harmonizing European bibliographical data on literature

Patryk Hubar
Institute of Literary Research, Polish Academy of Sciences, Warsaw
patryk.hubar@ibl.waw.pl

Nikodem Wołczuk
Institute of Literary Research, Polish Academy of Sciences, Warsaw
nikodem.wolczuk@ibl.waw.pl

Dariusz Perliński
Institute of Literary Research, Polish Academy of Sciences, Warsaw
dariusz.perlinski@ibl.waw.pl

Róbert Péter
University of Szeged
robert.peter@ieas-szeged.hu

Vojtěch Malínek
Institute of Czech Literature, Czech Academy of Sciences, Prague
malinek@ucl.cas.cz

Keywords
bibliography, data science, databases, linked open data, literary science

Abstract
The main objective of the paper is to present Literary Bibliography Research Infrastructure (LiBRI; literarybibliography.eu) as an aggregator of various literary bibliographic data and to show the main assumptions when binding and unifying diverse data sets with different structures. LiBRI is a joint initiative of Czech and Polish Literary Bibliography, both being long-term existing bibliographical infrastructures operating within the National Academies of Sciences. The aim of the service is to collect, merge and provide one interface to present multilingual bibliographical metadata for literary studies stored in MARC21 format. Currently, available collections are the Czech Literary Bibliography, the Polish Literary Bibliography,

and the literary content extracted from the National Library of Finland. The LIBRI data is presented within the VuFind discovery system. Within the LIBRI project framework, various software updates of VuFind adopted to the presentation of the literary bibliographical data were developed.

The main challenge was to implement a common data model and data processing workflow for records from different resources to create a highly interoperable and interconnected dataset. It was primarily necessary to unify the data on authors, source documents, and subject and genre description systems. Metadata was also enriched with identifiers from external controlled services, e.g. VIAF, ISSN portal. LIBRI team developed a custom mechanism for creating and updating authority records based on data from interoperable data sources as well. Each record is created automatically based on Wikidata and meets the predefined structure of MARC21 format for authority data. Further enrichment of authority data by Wikidata content enables to provide dataset not only for introduction of specific literary figures but allows to provide dataset for advanced data-driven research on the given literary field. The next step is to further harmonize and enrich the data according to the Semantic Web and Linked Open Data standards. We will present an RDF-based data model which will give users not only a basic overview of the query but also the possibility to explore external databases that significantly increase research potentials.

The workflow for dataset aggregation is designed in such a way that it can accommodate more data sources. We will showcase this capacity by providing an analysis and assessment of Hungarian bibliographical data on literature, hosted by at least five data providers such as the MOKKA (Hungarian National Shared Catalogue) Association. It investigates the current state of these databases (e.g. number of records, type, format of bibliographical data, international identifiers) as well as the legal and technical aspects of access to these resources. It also offers a pilot study drawing on Hungarian literary data to demonstrate the potentials of such a project.

Bibliography

Hatop, G. (2013). Integrating Linked Data into Discovery. Code4Lib journal, 21. Online: https://journal.code4lib.org/articles/8526

Kaltenbrunner, W. (2015) Scholarly Labour and Digital Collaboration in Literary Studies, Social Epistemology, 29:2, 207-233, DOI: 10.1080/02691728.2014.907834

Umerle, T. et al. (2022). An Analysis of the Current Bibliographical Data Landscape in the Humanities. A Case for the Joint Bibliodata Agendas of Public Stakeholders. Zenodo. https://doi.org/10.5281/zenodo.6559857

Thai Digital Humanities Researchers' Perspectives on Sharing Research Data

Wachiraporn Klungthanaboon
Department of Library Science, Faculty of Arts,
Chulalongkorn University
Wachiraporn.k@chula.ac.th

Keywords
research data, data sharing, digital humanities

Abstract

Open access to research data has been increasingly demanded from various funding agencies, research-related institutions, and publishers at the international level. This has presented challenges to our Thai scholarly community, particularly in the emerging field of Digital Humanities. Therefore, the research project explores the current state and the perspectives of Digital Humanities researchers in Thailand on research data sharing including the factors and barriers to data sharing. This mixed-methods research employs semi-structured interviews with 15 Thai Digital Humanities researchers from various disciplines, followed by online surveys to collect a broader range of quantitative data from Thai Digital Humanities researchers.

However, the findings of the first stage of data collection will be shared in this presentation, which may shed light on the understanding of research data sharing practices among Digital Humanities in Thailand. The 15 Digital Humanities researchers were recruited through purposive sampling by contacting the researchers in Digital Humanities Labs/Research Units and through snowball sampling. The one-on-one semi-structured interviews, held January to April 2022, were video-recorded via Zoom, a video meeting program due to the COVID-19 pandemic situation. The recorded interviews were transcribed, and qualitative analyses were conducted using ATLAS.ti 9 with thematic analysis technique.

The findings reveal that Thai DH researchers employ both non-computational and computational research methods. That results in the broad range of DH research data types, but increasingly in digital format.

Research data are mostly managed, stored and and archived in several ways, including external harddrives, serviced cloud storage, and institution-led cloud storage. All informants perceive the values of their research data and tend to keep it for as long as possible. Most of them recognize the benefits of sharing and reusing research data, but only a few have experience with data sharing and publishing. Some researchers in certain fields have a data sharing culture whereas others are unfamiliar. Considering the national and institutional factors influencing data sharing in Digital Humanities, no universities, research funders, and local journal publishers' policies on research data sharing is reported. However, a few informants have noticed that international journals in their fields have data sharing policies and publish research data as supplementary data with journal articles. This may have an impact on their research data sharing practices. The informants mostly share their research data with colleagues in the same research labs/projects, with familiar researchers, and upon request with other academics. Academic and professional networking, both formal and informal, may help DH researchers reach an agreement to share their research data. If national research funders impose mandatory data sharing policies, the policies should be clearly explained and well communicated to the researchers. However, concerns about publishing opportunities and ethical issues are mostly cited as barriers to research data sharing.

Bibliography

Ayris, P. (2017). Challenges and Opportunities for Research Data Management in the Arts, Humanities and Social Sciences: a practitioner's viewpoint (Issue February). http://discovery.ucl.ac.uk/1546540/1/_Learn_CS9_45-56.pdf

Borgman, C. L. (2012). The Conundrum of Sharing Research Data. Journal of the American Society for Information Science and Technology, 63(6), 1059–1078. https://doi.org/10.1002/asi.22634

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. Library and Information Science Research, 41(2), 109–122. https://doi.org/10.1016/j.lisr.2019.04.004

Funari, M. (2014). Research data and humanities: a European context. Italian Journal of Library & Information Science, 5(January 2014), 209. https://doi.org/10.4403/jlis.it-8927

Joint Information Systems Committee. (2020). Research data in arts, humanities, and social sciences. https://rdmtoolkit.jisc.ac.uk/plan-and-design/research-data-in-arts-humanities-and-social-sciences/

Kelli, A., Mets, T., Vider, K., Värv, A., Jonsson, L., Lindén, K., & Birštonas, R. (2019). Challenges of transformation of research data into open data: The perspective of social sciences and humanities. International Journal of Technology Management & Sustainable Development, 17(3), 227–251. https://doi.org/10.1386/tmsd.17.3.227_1

The semantic pattern of Dezső Kosztolányi's bibliography

Kata Dobás
Bölcsészettudományi Kutatóközpont Irodalomtudományi Intézet
kata.dobas@gmail.com

Keywords
wikibase, semantic network, Kosztolányi, ITIdata, bibliographies, patterns

Abstract

The first project that began to store its data in ITIdata is concerned with the works of Dezső Kosztolányi. Work on a critical edition of Kosztolányi's texts began in 2008. As the fundraising work progressed, it became clear that a significant part of Kosztolányi's œuvre could be found in printing press. We currently know of 250 periodicals in which writings from Kosztolányi were published, the final issue is probably much larger. From the results so far, we have published the first six volumes of the Kosztolányi's bibliography, which contains 11,000 items. It was characteristic of Kosztolányi's publishing practice that the same work was republished in another periodical, with changes and possibly a new title, but it was not uncommon for the same text to be published again years later. We needed a database to make the publishing network of a work visible. From 2018 we worked in the Koha library system and from 2022 in ITIdata. With the increase in the number of emerging data, the known pattern also became more nuanced. In several cases we had to use unique solutions in ITIdata, which led us to new literary and press history informations.

In the current state of data entry, it seems worthwhile to carry out a comparative study. The amount of data allows us to compare our existing pattern with other semantic networks. From this point of view, datas of historians can be just as interesting to us as the examination of a manuscript heritage or another wikibase-based database dealing with press bibliographies. In my presentation, I will attempt to

summarize the results of my comparative indicator work. By the end of the conference talk, it may be possible to answer the question in which direction the Kosztolányi database should be further expanded.

Bibliography
Prebor, Gila, From authority data, to linked open data and Wikidata: The case study of a Hebrew manuscript catalogue, https://www.ideals.illinois.edu/items/114151
Harangozó, Ádám, FactGrid: adatbázis történészeknek, Digitális Bölcsészet, 2020/3., 29–38.
Klijn, Edwin (2020). From Paper to Digital Trail: Collections on the Semantic Web. Historical Social Research, 45(4), 244–262, https://doi.org/10.12759/hsr.45.2020.4.244-262
Fazekas, Júlia – Dobás, Kata, ITIdata – Egy irodalmi adatbázis fejlesztése Wikibase alapon és ennek hasznosítása Kosztolányi Dezső forrásjegyzékénél, Networkshop 2022.

HISTORY, VISUAL ARTS

Ontology-Based Image Knowledge Organization for Yangshi Lei Archives

Zhao Yu
Tianjin University
2871688124@qq.com

Ma Zhaoyi
Nangyang Technological University
zhaoyi067@hotmail.com

He Jie
Harbin Institute of Technology (Shenzhen)
hejie2021@hit.edu.cn

He Beijie
Tianjin University
hebeijie@tju.edu.cn

Keywords
Ontology, Knowledge Organization, Yangshi Lei Archives, CIDOC CRM

Abstract
Yangshi Lei Archives include a collection of architectural drawings and other related documents designed, drawn and written by the Lei family during the Qing Dynasty. The archives contain a great deal of knowledge related to the mechanisms of ancient Chinese architectural practice and are of great value in preserving the architectural and cultural knowledge of the Qing Dynasty. At present, the study of Yangshi Lei Archives mainly uses the archives as historical materials to reveal the design procedures and methods of royal architecture in the Qing Dynasty, aiming at the systematic identification and overall understanding of the archives. The studies of Yangshi Lei Archives has generated a large amount of knowledge related to ancient Chinese architectural practices. However, the knowledge is presented as heterogeneous data from multiple sources that are difficult to integrate and organize, and the semantic relationships between the knowledge are complex. These problems hinder the in-depth understanding and study of Yangshi Lei Archives.

To address these issues, the study constructs an ontology of Yangshilei Archives with a combination of top-down and bottom-up approaches. Firstly, through the reuse and extension of the CIDOC CRM ontology (ISO 21127:2014), an ontology framework of Yangshi Lei Archives is constructed that associates architectural events with people, places, and physical objects at those events. With this ontology framework, Yangshi Lei Archives are returned to their original production, use, and circulation context. Then, based on the knowledge extraction from the existing studies, we extend the classes and relations of the existing ontology framework to form the ontology of Yangshi Lei Archives. Finally, the study takes the Archives of Dingling as an example to build the knowledge graph and conduct experimental validation.

As a result, Yangshi Lei Archives ontology was able to describe the complex semantic relationships of Yangshi Lei Archives and effectively support knowledge organization, thereby supporting a new digital paradigm for studying the material culture of Yangshi Lei Archives.

Bibliography

Beijie He, Qiheng Wang, History of the Qing Dynasty Style Lei Family and its Architectural Drawing File Research, China Construction Industry Press, 2017.

Signore, O. (2009), "Representing knowledge in archaeology: from cataloguing cards to semantic web", Archeologia e calcolatori, Vol. 20,111-128.

Wang, X., Song, N., Liu, X., & Xu, L. (2021). Data Modeling and Evaluation of Deep Semantic Annotation for Cultural Heritage Images. Journal of Documentation.

Historical Knowledge Graph Creation with User-friendly Linked Data Editor

Jun Ogawa
Center for Open Data in the Humanities, National Institute of Informatics
htjk6513khbk@gmail.com

Satoru Nakamura
Historiographical Institute, University of Tokyo
nakamura@hi.u-tokyo.ac.jp

Asanobu Kitamoto
Center for Open Data in the Humanities, National Institute of Informatics
kitamoto@nii.ac.jp

Keywords
knowledge graph, linked data, annotation, historical documents

Abstract

Up to now, several attempts to design a knowledge graph (KG) representation for semantic contents of historical documents, such as detailed information about social relationships, actions, or contacts mentioned in sources, have been made. One being the most famous and applied of those would be the Factoid Prosopography Ontology (FPO) [1]. This ontology represents various historical phenomena as a Factoid, which is "a spot in a source that says something about a person or persons" and all the related information and actors are linked to it [2]. The other example is the Chinese Text Project (CTP) organized by Donald Sturgeon at Durham University [3]. This project proposes a concept of 'knowledge claim' consisting of three basic elements: the subject, the object, and the predicative verb or relation, for structuring the historical information given by the sources in natural language. Despite these challenging projects, and other similar projects, the attempt to deal with the semantics of historical documents is still quite limited, and thus the sum of the data created in this way so far is relatively small. One of the reasons for this might be the cost of data construction, and this is indeed the very issue that we try to deal with in this study.

Based on the data model that we previously proposed for the historical knowledge description [4], we now proceed to the data construction process.

Since the data structure we adopt is complex enough to hesitate to enter all the entities and their connections by hand, we surely need a user-friendly interface. If the data input process is optimized, the pace of KG creation would be accelerated. Although CTP tackles this task by developing an interface and employing the crowdsourcing method, it is limited to Chinese materials and targets relatively formulaic texts. Thus, while consulting the CTP, we propose a comprehensive system for entity markup and KG creation for the texts written in Latin. The basic workflow of our system is 1) marking up entities, 2) creating a knowledge graph with Linked Data Editor, 3) transforming to RDF (turtle) data, and 4) visualizing.

The data behind the first two phases are provided in TEI/XML format and, as all the words have their proper ID, each of them can be included in the KG as an RDF resource. The fact that the KG created with this system is always connected to the source description is very important as it guarantees the reference to the original texts. On the other hand, the Linked Data editor enables us to create linked data flexibly and easily because the entities mentioned in sources are once separated from the strict textual structure and connected to each other with several clicks on the interface.

Our system, which helps users to create historical KG as effectively as possible while keeping the connection between specific source descriptions and a more conceptual knowledge sphere, will contribute to the further progress of the use of KG and linked data in the historical discipline.

Bibliography

M. Pasin and J. Bradley, 'Factoid-based prosopography and computer ontologies: towards an integrated approach', Digital Scholarship in the Humanities, 30-1, 2015, pp. 86-97.

[2] 'What is Factoid Prosopography all about?', Factoid Prosopography, King's College London, accessed 15 June 2022: https://www.kcl.ac.uk/factoid-prosopography/about.

[3] 'Semantic annotation', Chinese Text Project, accessed 15 June 2022: https://ctext.org/instructions/annotation.

[4] J. Ogawa, K. Nagasaki and I. Ohmukai, 'Modelling and Structuring Narrative Historical Sources with Temporal Context', International Journal of Humanities and Arts Computing, 16-1, 2022, pp. 17-32.

Mapping the Domestic Politics of International Terror: An Actant Network Analysis of Swedish Parliamentary Debate on Terrorism, 1971–1978

Mats Fridlund
University of Gothenburg, Sweden
mats.fridlund@gu.se

Daniel Brodén
University of Gothenburg, Sweden
daniel.broden@gu.se

Leif-Jöran Olsson
University of Gothenburg, Sweden
leif-joran.olsson@gu.se

Victor Wåhlstrand Skärström
University of Gothenburg, Sweden
victor.wahlstrand.skarstrom@gu.se

Magnus P. Ängsal
University of Gothenburg, Sweden
magnus.pettersson.angsal@gu.se

Patrik Öhberg
University of Gothenburg, Sweden
patrik.ohberg@gu.se

Keywords
Digital history, parliamentary data, actor-network theory, framing, terrorism studies

Abstract
This paper contributes to methodological development in digital humanities through a data-driven mapping of the discourse on terrorism in the Swedish Parliament (the Riksdag). Drawing on state of the art language technology and Social Network Analysis (SNA), we use and develop Actant Network Analysis (ANA) to map the socio-conceptual network of the political debate about the phenomenon of 'international terrorism' during a formative period (Stampnitzky 2013). We pursue two interconnected research questions: First, which Members of Parliament (MPs) and political parties participated in

the formation of the understanding of the notion of international terrorism during this critical period? Here, we analytically use Actant Network Analysis (ANA), drawing upon SNT and Actor-Network Theory (Callon 1986) and Controversy Mapping (Venturini & Munk 2022), in order to investigate different socio-conceptual clusters of reference enveloping the debates. Second, we ask what discursive elements – incidents, organizations, countries, ideologies, etc. – were referred to and integrated in the 'framing' of terrorism in the parliamentary debate, based, first, on frequencies of concepts indicating frames and, second, on co-occurrences on the level of word units. Here, we draw on the linguistic concept of framing in order to focus on the process of selection of salient elements and the highlighting or downplaying of different aspects (Entman 1993, Wehling 2016). In particular, our socio-conceptual ANA is developed to capture the discursive elements that connect different and partly antagonistic actors in the discourse network as well as distinctly separated clusters.

The primary source material is Parliamentary 'Speeches', edited transcriptions of deliberations by MPs in debates, which are extracted and automatically generated from the Riksdag's Open Data and further processed using language technology and social network analysis tools.

The study is part of the SweTerror project, a major multidisciplinary mixed methods investigation of the development of the Swedish parliamentary terrorism discourse 1968–2018 (Edlund et al., 2022). In extension, the paper presents possibilities for future research on the politics of terrorism provided by an interdisciplinary mixed method research design and the affordances of machine-learning and big data analysis.

Bibliography

Callon, M. (1986): "The Sociology of an Actor-Network: The Case of the Electric Vehicle". in: Michel Callon, John Law & Arie Rip (Eds.) Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World, Palgrave Macmillan, London, 19-34.

Edlund, J, D. Brodén, M. Fridlund, C. Lindhé, L-J. Olsson, M.P. Ängsal, P. Öhberg, (2022): "A Multimodal Digital Humanities Study of Terrorism in Swedish Politics: An Interdisciplinary Mixed Methods Project on the Configuration of Terrorism in Parliamentary Debates, Legislation, and Policy Networks 1968–2018", in: Kohei Arai (Ed.), Intelligent Systems and Applications. IntelliSys 2021. Lecture Notes in Networks and Systems, vol 295, Cham, Springer, 435–449.

Entman, R.M. (1993): "Framing: Toward Clarification of a Fractured Program", Journal of Communication, 43(4):51–58.

Stampnitzky, L. (2013): Discipling Terror: How Experts Invented 'Terrorism'. Cambridge University Press, Cambridge.

Venturini, T., A.K. Munk (202): Controversy Mapping: A Field Guide, Polity Press, Cambrideg.

Wehling, E. (2016): Politisches Framing: Wie eine Nation sich ihr Denken einredet - und daraus Politik macht, Herbert von Harlem Verlag, Köln.

Use of Machine Learning Classification Models, both Image and Text, in the Network Graphing. Case Study: Community of Practice Among Graffiti Writers on Freight trains

Angel R. Abundis
DECS / CUCSH / UDG
abundiscomunicacion@gmail.com

Keywords
keywords, help, identifying, suitable, reviewers, readers

Abstract
This work has as a phenomenon the graffiti that decorates the freight trains that circulate in a transnational railway network. Graffiti writers participate in a complex symbolic competition in which they seek to become known by marking the largest number of trains, with the best executed interventions and with the largest size, as well as registering and publishing them on social networks, particularly Instagram.

The Getting Up as principle of inner symbolic competition was documented by Castleman (1984) and has established itself as an object of study in works related to New York graffiti, answers questions such as: What shared values are found among graffiti writers? And, what is the engine that reproduces the practice of announcing the presence of an individual? It is from the conceptual framework of Jenkins' participatory culture and communities of practice (Jenkins et al., 2009) that we approach the graffiti writers of the New York tradition on freight trains. The main question of this work is to answer how to objectify a network of individuals who share an identity transnationally, without face-to-face communication and whom through constant participation, seniority and peer evaluation consolidate a community.

This paper aims to show the use of two machine learning models in the calculation of the weight of nodes and edges when graphing a network, whose data source is mined from Instagram posts of graffiti writers on freight trains in the North America region. The machine learning models allows listing attributes of both the content of a publication or the profile itselft

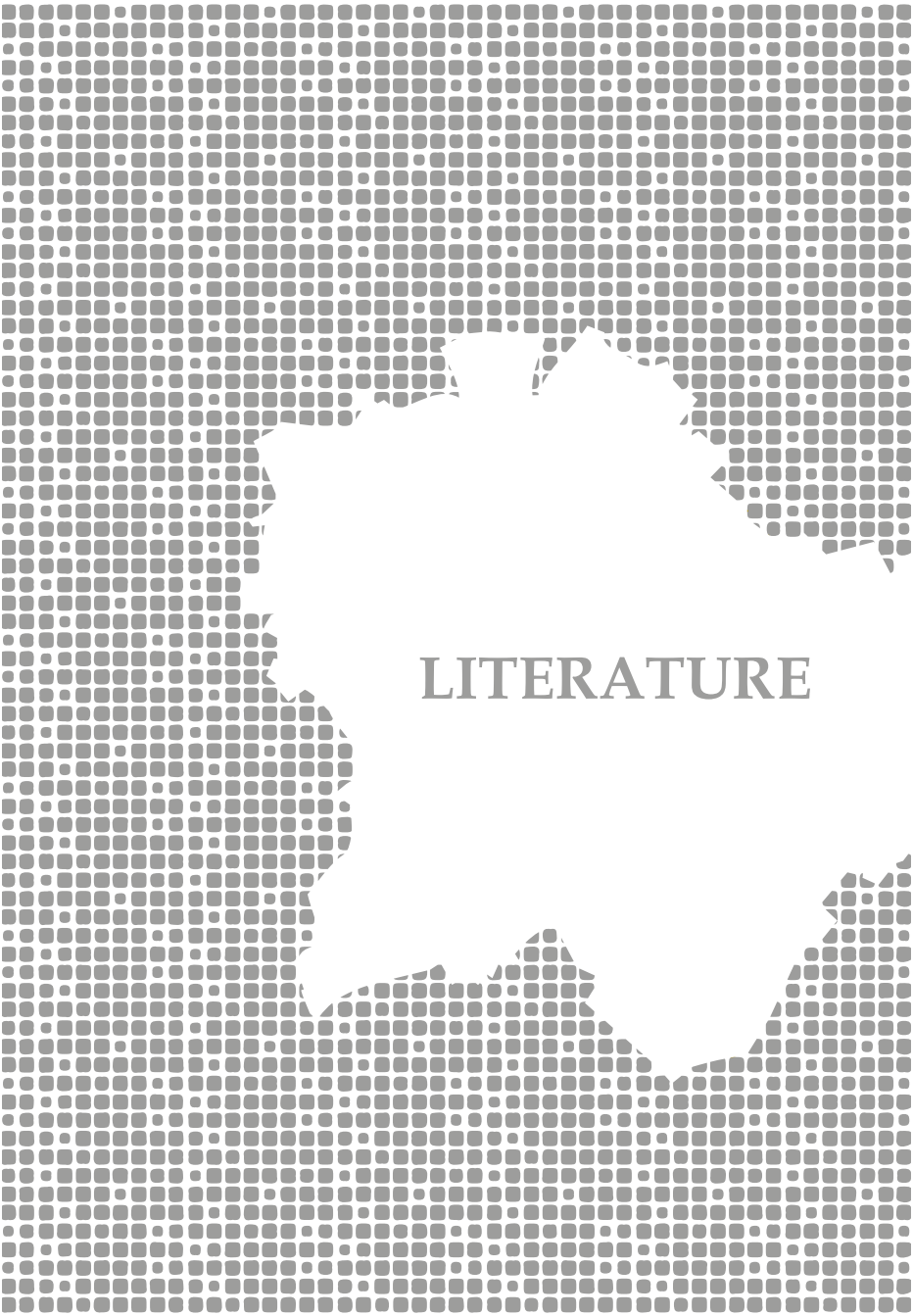
to add weight to the node, as well as the type of communication that exists between these points. For the image analysis, two convolutional neural networks are used: First, a pre-trained model with the ImageNet data set allows finding common elements that will help understand and describe the context that was documented. Second, a custom-trained object detection model distinguishes the types of graffiti with which freight trains are marked. To approach the captions, a classification model is used that allows hashtags or fragments to be grouped into four referential categories such as geographic, identity, self-representation and emotional texts.

This method makes it possible to ensure that there is a geographically dispersed group that shares symbolic elements in their publications, that document their interventions on the physical plane and that interact on the socio-digital plane, forming a circulation network of applied meaning. It is also possible to measure, and group, the frequency of references used by writers, the types of graffiti, the characteristics of the interventions and how they are evaluated by peers.

Bibliography

Castleman, C. (1982). Getting Up: Subway graffiti in NewYork. MIT Press

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). Confronting the Challenges of Participatory Culture: Media Education for the 21st Century. MIT Press


LITERATURE

Automatic citation detection as a "distant reading" praxis: scrutinizing text similarity techniques on Hungarian texts

Balázs Indig
Eötvös Loránd University, Department of Digital Humanities
indigbalazs@btk.elte.hu

Gábor Palkó
Eötvös Loránd University, Department of Digital Humanities
palko.gabor@btk.elte.hu

Keywords
intertextuality; distant reading, text similarity, MinHash

Abstract

Studies on the concept of intertextuality fill a library in the humanities. The network of texts that quote each other forms the fabric of literature, and of written culture in general, according to many theories. By understanding how this network is constructed, how texts enter into dialogue with each other and how the recipient is involved in this dialogue, the structure of culture becomes visible.

Both the greatest opportunity and the greatest challenge of intertextuality research is the historicity of the phenomenon. What constitutes the citational relationship between two texts, what the recipient identifies as intertextual, and how the citational relationship is judged by the recipient: all these show highly variable patterns in time and space. From the Renaissance cult of quoting ancient patterns to the Romanticism's originality principle, a wide variety of constructions are possible. How citation as a cultural practice works is not merely an internal matter of literature. Factors such as the changing social perception of plagiarism, the development of the institution of copyright, or the various institutional forms of physical reproduction of texts must also be taken into account in the analysis. However, these are not the focus of our argument. The way in which the concept of iof perspectiventertextuality

is approached today and in the future will certainly be influenced by two cultural practices. Firstly, the radical gesture of postmodern literature and theories: the global cult of quotation, which in Hungarian literature reaches its peak in the works of Péter Esterházy. Secondly, digital humanities, or, more narrowly, the machine technologies of distant reading, which relocate the act of citation, always generated in the practice of reading, into a hybrid medial space that presupposes the cooperation of machine and human actors. The fact that ready-made, out-of-the-box tools have emerged that promise the automatic exploration of intertextual relationships is a huge opportunity, to transcend the limitations of human close reading and human memory. However, this new hybrid practice is dangerous for several reasons. Dangerous, on the one hand, because it may imply a naïve belief that intertextuality is an objectively existing phenomenon independent of the temporality of texts and the recipient's individual and, in the meantime, culturally conditioned relationship to them. It is also dangerous because out-of-the-box tools act as a black box, hiding the machinery that generates lists of citation relations from the eyes of the researcher. In this talk, we will attempt to present the state of the art of a recent research project. We aim to gain and provide insight into the black box of tools such as the Intertext software published by the Yale DH Lab. The operation of the software, the algorithms hiding under the hood, the role of hyperparameters, software code errors and their consequences will be analysed. We discuss the role of Jaccard similarity and MinHash technology and other text similarity measures in identifying citation relations. Finally, we place the above analysis in the context of two specific corpora and research questions: the intratextuality of Péter Esterházy's works and the ELTE-DH web source article corpus, searching for relevant and irrelevant citation relations.

Bibliography

1. Jure Leskovec, Anand Rajaraman, Jeff Ullman: Mining of Massive Datasets, Cambridge UP., 2020(3) ;
2. Broder, Andrei Z. (1997), "On the resemblance and containment of documents", Compression and Complexity of Sequences: Proceedings, Positano, Amalfitan Coast, Salerno, Italy, June 11-13, 1997;
3. https://dhlab.yale.edu/projects/intertext/

War as Network
Mihály Babits's poems about First World War

Zsófia Fellegi
Research Centre for the Humanities, Institute for Literary Studies
fellegi.zsofia@abtk.hu

Anita Káli
Research Centre for the Humanities, Institute for Literary Studies
kali.anita@abtk.hu

Gábor Palkó
Eötvös Loránd University, Department of Digital Humanities
palko.gabor@btk.elte.hu

Zoltán Szénási
Research Centre for the Humanities, Institute for Literary Studies
szenasi.zoltan@abtk.hu

Keywords
World War, poem, Mihály Babits, intratextuality, intertextuality

Abstract

The First World War is usually interpreted as the first mass war, and its impact is not only historical and political, but also significant from the point of view of literary history. In the years following the outbreak of the war, thousands of Hungarian poems and prose works are born that chose the World War as their theme. Though the renowned Hungarian poet Mihály Babits (1883–1941) published his Young Soldier and Our Father in war-anthologies his criticism of war propaganda and war rhetoric is, in many ways, an essential element of his poetry of the time. Thus, the First World War is not merely the theme of Babits's poems, but can also be interpreted as a network of forms and modes of speech. The First World War as a network can therefore be understood from several perspectives in Babits's poetry: on the one hand, we can find this theme in terms of the poetic modes of speech and forms evoked in the war poems, and on the other hand, we can speak about poet's traditional roles. As a result, the question arises, to what extent

the long 19th century, understood from the perspective of Hungarian national traditionalism, is opposed to literary modernity in connection with the war. How the war and the Hungarian war poetry shapes the characteristics of Babits poems, how Babits's poetry changes continuously until the Fortissimo (1917). To answer these questions, the paper examines Babits's war poems Before Easter (1916), War Anthologies (1916) and Fortissimo, and also presents some manuscript written between 1916 and 1919. The paper introduces the related names and the network of the literary field, examines the presumed relationship of Babits's published war poems and unpublished manuscripts, as well as demonstrates the possible relationship between the poems in war anthologies and Babits's poems. Due to the nature of the oeuvre, there are many rough drafts, pre-studies and fair copies available, and a digital critical edition of these is currently being prepared. The digital representation of the manuscript network in XML format, as recommended by the Text Encoding Initiative, is non-trivial, and the possibilities of doing so will be presented, and the possibilities of publishing primary and secondary literature and metadata of manuscripts on the semantic web will be explored.

Bibliography

Zoltán SZÉNÁSI, "A Húsvét előtt előtt: Babits Mihály legismertebb háborúellenes versének keletkezéstörténete és a kronológia elve a kritikai kiadásban", Irodalomtörténet 51, no. 4 (2020): 471–483.

From WARC to Graph: Link Extraction for Web Archive Analytics

Claus-Michael Schlesinger
University of Stuttgart
cms@ilw.uni-stuttgart.de

Pascal Hein
University of Stuttgart
pascal.hein@ilw.uni-stuttgart.de

Keywords
net literature, web archive analytics, graph data

Abstract

In hypertext fiction, the text, before reading, exists only in a state of potential. The text of a reading will only be realized through a series of decisions by the reader. For example, in Susanne Berkenheger's Zeit für die Bombe (Time for the Bomb) (1997), the text is split into several pages. A page can contain hyperlinks and other reference types that are used to link to other pages or embed content like images into the site. Zeit für die Bombe also makes use of automatic forwards placed in the HTML header. Pages with one or multiple clickable links require the reader to take action in order to go to the next page. As each link leads to a different page, clicking a link means that the text will be constructed page by page based on the reader's decision about which link to follow. Pages can be modeled as nodes of a network, hyperlinks, automatic forwards and other types of references as connections between nodes. Pages with multiple references are thus represented as nodes with an outgoing degree greater than 1. The pages and hyperlinks will thus form a specific network that represents the condition of all possible texts and that represents the site as a network of its constituting elements. As hypertext fiction is, in the space-time continuum of the world wide web, an old genre, chances are that works don't exist on the live web any more, but only in web archives. Some web archiving software packages write hyperlinks found on a page into the web archive metadata, others do not. In Berkenheger's Zeit für die Bombe, some pages use HTML header functions to implement an automatic forwarding after

a specified amount of time. Also, links generated dynamically through javascript code can be difficult to extract for crawlers. While approaches working with large web archives prioritize scalability and focus on WARC metadata (Eldakar and Alsabbagh 2020, Lin et al. 2017), our approach prioritizes completeness of extracted reference information and applies several extraction methods including analysis of WARC payloads. In our presentation we describe the data model developed for deep extraction of reference data in web archives and our software implementation . Departing from our initial usecase - analyzing link structure in net literature - we describe how the link extraction procedure implies a transformation of WARC objects into graph data. We show that this transformation can be understood as an early step in web archive analytics pipelines not only for net literature, but more broadly for historical research which might be interested in the structure of single objects or small to medium web archive corpora.

Bibliography

Berkenheger, Susanne. 1997. "Zeit Für Die Bombe." Hyperfiction. 1997. http://www.berkenheger.netzliteratur.net/ouargla/wargla/zeit.htm.

Eldakar, Youssef, and Lana Alsabbagh. 2020. "LinkGate: Let's Build a Scalable Visualization Tool for Web Archive Research." April 23, 2020. https://netpreserveblog.wordpress.com/2020/04/23/linkgate-update/.

Hein, Pascal, Mona Ulrich, Claus-Michael Schlesinger, and Andre Blessing. 2022. "Warc2graph." https://github.com/dla-marbach/warc2graph.

Lin, Jimmy, Ian Milligan, Jeremy Wiebe, and Alice Zhou. 2017. "Warcbase: Scalable Analytics Infrastructure for Exploring Web Archives." J. Comput. Cult. Herit. 10 (4): 22:1-22:30. https://doi.org/10.1145/3097570.

Digital philology and the semantic web

Zsófia Fellegi
Research Centre for the Humanities, Institute for Literary Studies
fellegi.zsofia@abtk.hu

Keywords
digital philology, semantic web, graph data model, digital scholarly editions

Abstract

The much-debated thesis of Barbara Bordalejo's 2018 paper is that "there is no such thing as digital scholarly editing". In her view, digital technologies have brought no revolutionary innovations to the practice of academic text editing. Initially glance, Bordalejo's argument may indeed seem valid, since the digital editions she cites do not exploit the potential of technology. As Joris van Zundert has pointed out, as long as publishers continue to publish editions that are digital metaphors of the book, no significant change can be expected. However, in the field of digital philology and digital scholarly publishing, a paradigm shift has begun in recent years, both theoretically and technologically. While Zundert has moved from representing text as document to represent text as work or text as process, Jeffrey C. Witt is now looking at digital scholarly editions from the paradigm of text as network. Andreas Kuczera, however, points out that although Witt actually conceives digital scholarly editions as a multi-actor network, which can easily be described as a graph, he does not consider the actors involved in the edition's production as part of the network, whereas Kuczera's idea of digital scholarly edition is that all the individual decisions of the contributors to the edition are part of the network, as part of the work, and therefore he no longer imagines a publication as a multi-actor graph, but as a provenance knowledge graph that includes transcripts, the critical apparatus, the relationships between apparatuses and the provenance of each statement. Recently, we have seen several examples of mapping scholarly text edition data and critical apparatus using semantic web technology or graph data models, some projects attempting to replace the TEI (Text Encoding Initiative) format, others experimenting with a combination of TEI and graph data models.

In my presentation, I will explore the potential applications of the semantic web and graph data models in the field of digital scholarly text publishing through an analysis of individual projects, keeping in mind issues of standardisation, sustainability and the needs of the research community. I discuss the advantages and disadvantages of the XML format, the RDF data model and the graph data model. Finally, I will present the latest developments of DigiPhil, the largest digital philology platform in Hungary. In the new infrastructure, the TEI XML representation is still the basis for critical editions, but we have integrated semantic web technology.

Bibliography

ZUNDERT Joris VAN, „Barely Beyond the Book?", in Digital Scholarly Editing: Theories, Models and Methods, ed. DRISCOLL Matthew James és PIERAZZO Elena (Cambridge: Open Book Publishers, 2016), 83–106, 105.

WITT Jeffrey C., „Digital Scholarly Editions and API Consuming Applications", ed. BLEIER Roman, BÜRGERMEISTER Martina, KLUG Helmut W., NEUBER Frederike és SCHNEIDER Gerlinde, 12 (Norderstedt: BoD, 2018), 219–247, 222, http://www.uni-koeln.de/.

SPADINI Elena, TOMASI Francesca és VOGELER Georg, ed., Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing (Norderstedt: Books on Demand, 2021).

BORDALEJO Barbara, „Digital versus Analogue Textual Scholarship or The Revolution is Just in the Title", Digital Philology: A Journal of Medieval Cultures 7, 1 (2018): 7–28, 24, https://doi.org/10.1353/dph.2018.0001.

KUCZERA Andreas, „TEI Beyond XML - Digital Scholarly Editions as Provenance Knowledge Graphs", in Graph Technologies in the Humanities - Proceedings 2020, ed. ANDREWS Tara, DIEHR Franziska, EFER Thomas, KUCZERA Andreas és ZUNDERT Joris van, 3110, CEUR Workshop Proceedings (Graph Technologies in the Humanities 2020, Vienna, Austria: CEUR, 2020), 101, 102, http://ceur-ws.org/Vol-3110/#paper6.

Critical editions in a database

Andor Horváth
Eötvös Loránd University, Budapest
andor@sublot.org

Keywords
digital humanities, semantic networks, network theory, database

Abstract

The world's first digitally published poetical database, the Repertory of Old Hungarian Poetry (Répertoire de la poésie hongroise ancienne, RPHA), originally contained only metadata. Over the past year, we have managed to upload more than half of the database's text corpus from various sources using a variety of methods. The database, which has been continuously developed since 1976, still stands out from most poetry repertories in that it contains complete entries for all the contemporary sources of the poems. Thus, the database distinguishes between the different physical examples of the poems and the ideal version of each poem as established by literary historians. Each poem has one main data sheet, which lists the characteristics that apply to the work in general, but in addition, the variant data sheets list all the characteristics in which every variant differs. This structure is very similar to that of the critical editions, where all the textual variants of the relevant sources are included. Therefore, the two critical editions currently under preparation in our workshop are published within the RPHA system. The critical text with the annotation apparatus belongs to the main datasheet of the poems, while the text variants belong to the variant datasheets which also the often include scanned images of the sources. There are also difficulties in finding the right text format. For critical editions, the TEI XML is the usual choice, but there are also arguments for a TEI-compatible JSON format. Many full-text databases available on the Internet use TEI XML encoding, but these are primarily designed to be used through the publishing institution's interface. The freedom of TEI encoding allows for a very detailed text capture but does not facilitate the reconciliation of editions produced in different workshops. It is because of this, that the POSTDATA research at the Digital Humanities Laboratory of the Universidad Nacional

de Educación a Distancia in Spain is converting TEI XML texts into a machine-processable, uniform JSON format, although many details are lost in the conversion. Is it possible to define a non-destructive, yet detailed text format? What should a text database look like, if its main goal is that projects outside your workshop can use the data as easily as possible?

Bibliography

Répertoire de la poésie hongroise ancienne https://f-book.com/rpha/v7/index.php

Horváth, Andor Márton; Szarka, Judit: Mit mesél a film hangja? A hang szerepe a filmes történetmesélésben. In: Golden, Dániel (ed.) Filmhang, filmzene a tanórán. Budapest: Színház- és Filmművészeti Egyetem (2019) pp. 10-53.

Horváth, Iván (ed.); H. Hubert, Gabriella (ed.); Font, Zsuzsa; H., Hubert Gabriella; Herner, János; Horváth, Iván; Szőnyi, Etelka; Vadai, István; Horváth, Andor Márton (technical editor) Répertoire de la poésie hongroise ancienne (v7). Budapest: Gépeskönyv (2021) Web: https://f-book.com/rpha/v7

Horváth, Iván; Balázs, Mihály; Bartók, Zsófia Ágnes; Margócsy, István (eds.); Horváth, Andor (technical editor): Magyar irodalomtörténet. Budapest: Gépeskönyv (2021) Web: https://f-book.com/mi

A Comprehensive Network Analysis of Early Hungarian Melodies and Poetical Forms

Levente Seláf
ELTE
levente.selaf@gmail.com

Anita Markó
ELTE
anitamarko89@gmail.com

Keywords
early modern literature, network analysis, Hungarian poetry, Music and Poetry

Abstract

There are 1523 Hungarian poems conserved from before 1600; its waste majority from the 16th century. Most of these poems were sung, and some of the melodies are still conserved. The metrical uniformity of the corpus allowed to sing many compositions to the same melodical lines. In 16-17th century sources, when there was no possibility to print the musical sheets, so rather often, indications of tunes (in general the incipit of a well-know song written to a specific melody) were given to allow the oral, sung performance of the poems. The Répertoire de la poésie hongroise ancienne contains all relevant information of this poetical tradition. Thanks to the data contained in it in this paper we present the first attempt to visualise and to interpret the very complex network of poems, melodies, indications of tunes, authors, sources, and confessions, in case of religious poems or songbooks. We show how the visualization of the network allow us to discover specific clusters, authorial groups, and to shad a new light on the generic system of the Hungarian poetry of that time.

Bibliography
Répertoire de la poésie hongroise ancienne https://f-book.com/rpha/v7/index.php
Szilvia Maróthy – Petr Plecháč – Levente Seláf: Rhyming in Sixteenth-Century Hungarian Historical Songs: A Pilot Study, In: Anne-Sophie Bories, Robert Kolár, Petr Plecháč (éd.), Plotting Poetry 4. Tackling the Toolkit, Prague, Institute of Czech Literature of the Czech Academy of Sciences, 2021, 43-58.
Levente Seláf: Chanter plus haut. La chanson religieuse en langues vernaculaires, Paris, Champion, 2008.

Structural differences between tragedies and comedies

Botond Szemes
Eötvös Loránd University, Department of Digital Humanities
szemes.botond@btk.elte.hu

Bence Vida
Eötvös Loránd University, Department of Digital Humanities
vida.bence@btk.elte.hu

Keywords
Social Network Analysis, Digital Literary Studies, Drama History, Character Networks

Abstract

In this presentation, we would like to introduce a method that allows us to identify dramatic genres based on their dramaturgical properties converted into quantitative data. These data above all relate to the metrics of character networks of the plays treated as social networks. These values are complemented by features that may also characterize the structure of a play (e.g. the distribution of words among characters; distribution of characters in scenes; average length of speech acts). Our goal is to create a method that is independent of the number of characters in the plays (i.e. the size of the networks), but at the same time gives meaningful results that can be traced back to real, understandable characteristics of the dramas. In this way, we can offer literary scholars a more useful method than ever before for this classification task. After finding the best performing features, we use the metric of cosine distance on the normalized scores of the texts and the Ward method to cluster them; and also SVM classification in binary cases (e.g. comedy vs. tragedy). The aim of all this research is to draw conclusions from the quantification of character networks and character speech in drama, about how network theory can be used to find new ways in drama analysis. The model we offer also goes beyond previous categorisation methods in the digital humanities because it seeks to identify prototypical works that embody a particular genre category, but defines other works

spectrum'. This approach is more effective in dealing with exceptions and allows for a more stratified analysis than the models before. In addition to the prototypical approach, the impact of individual scenes on the character web of the drama as a whole will be examined in terms of the structural dynamics of the character networks. An attempt will be made to identify the dramaturgical points that are decisive for the tragic or comic development of the plot.

Bibliography

Algee-Hewitt, Mark. „Distributed Character: Quantitative Models of the English Stage, 1550–1900". New Literary History 48 (2017): 751–782.

Granovetter, Mark. „The Strength of Weak Ties: A Network Theory Revisited". Sociological Theory 1 (1983): 201–33. https://doi.org/10.2307/202051.

Moretti, Franco. „Network theory, plot analysis". New Left Review (2011): 80–102.

Stiller, James, Nettle, Daniel and Dunbar, Robin. „The small world of Shakespeare's plays". Human Nature 14 (2003): 397–408. https://doi.org/10.1007/s12110-003-1013-1.

Trilcke, Peer, Fischer, Frank, Göbel, Mathias and Kampkaspar, Dario. „Comedy vs. Tragedy: Network Values by Genre". https://dlina.github.io/Network-Values-by-Genre/. (Downloaded: 12.09.2022.)