

Digital Humanities 2019: Complexities

July 8-12, 2019
Utrecht, The Netherlands

a conference report prepared by
Tyng-Ruey Chuang

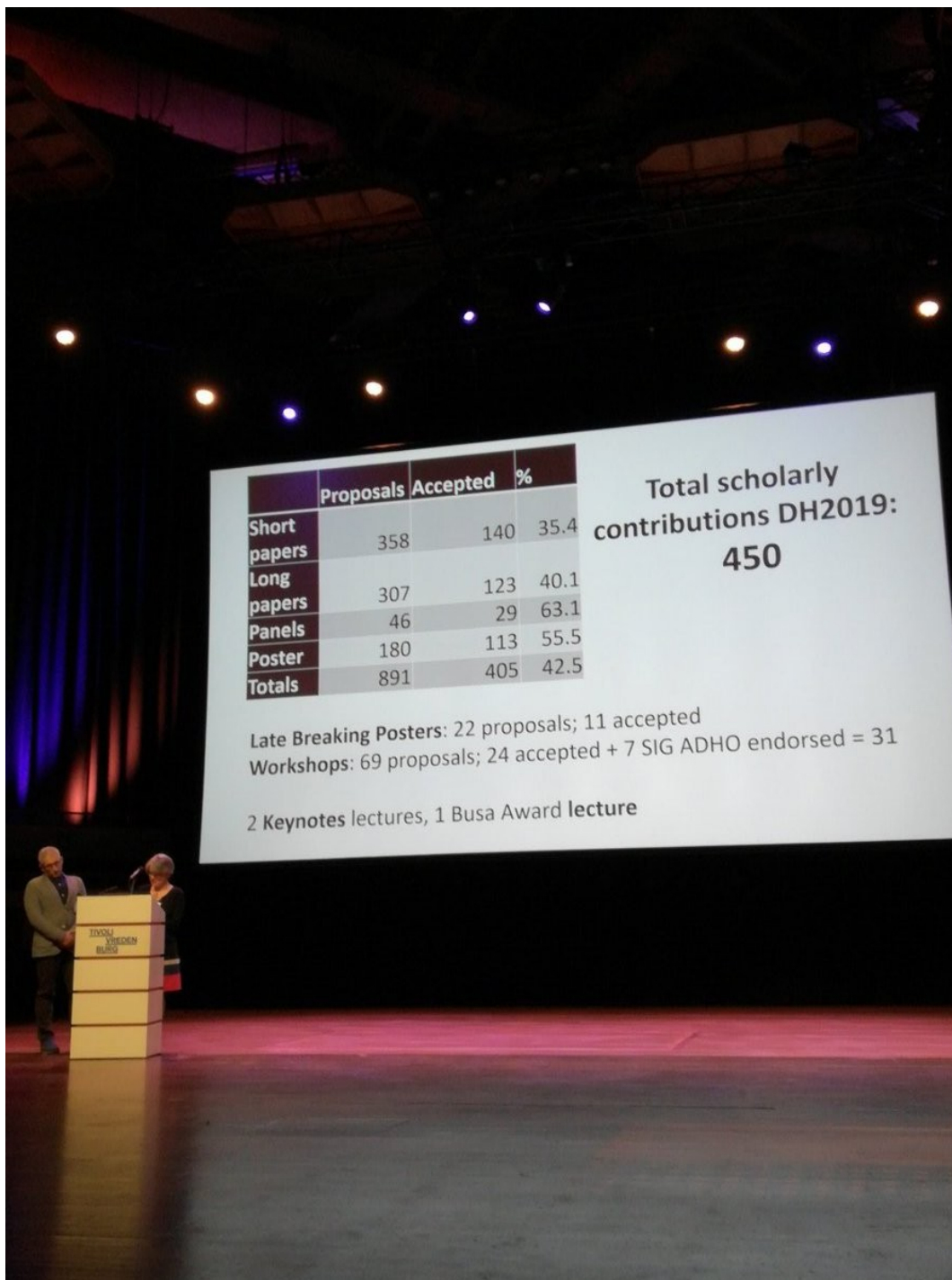
2019-08-06

Organization

- A three-day conference preceded by two days of workshops
 - 31 workshops (10 full-day, 21 half-day)
 - 2 keynotes, 1 Busa lecture
 - 41 long-paper sessions, 28 short-paper sessions, 29 panels, 1 poster session (~100 posters)
 - up to 10 parallel sessions at any given time
 - other co-located meetings
- Organized by the Alliance of Digital Humanities Organizations (ADHO) and the Utrecht University.

- 700 reviewers
- 3600 reviews
- over 900 proposals
- 1000+ participants

A sampler of three papers and three panels will be discussed today, with some of my thoughts and photos.





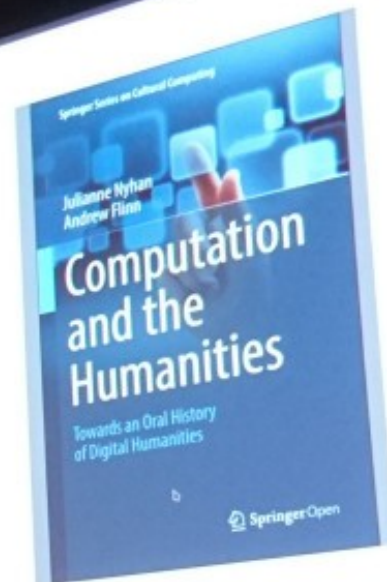
Why should GLAMs and data scientists collaborate?

GLAMs (galleries, libraries, archives and museums) have vast collections

- The British Library has up to 200 million items, including: 14 million books; 8 million stamps; 310,000 manuscript volumes; 4 million maps; Pamphlets, magazines, newspapers, sheet music; Television and radio recordings; Websites, e-books, e-journals. Over 3 million new items are added every year

But – that scale means cataloguing is often minimal / focused on





Springer Series on Cultural Computing

Julianne Nyhan
Andrew Flinn

Computation and the Humanities

Towards an Oral History
of Digital Humanities

Springer Open



Conference Materials and Discussions

- Website
 - <https://dh2019.adho.org/>
- Program
 - <https://www.conftool.pro/dh2019/index.php?page=browseSessions>
- Abstracts
 - <https://dh2019.adho.org/programme/book-of-abstracts/>
- Twitter
 - #DH2019
 - @DH2019_NL

Themes

- Cultural Heritage, Art/facts and Institutions
- Cultures Literatures and Texts
- Space, Territory, GeoHumanities
- Scholarly Communities, Communication, Pedagogy
- Tools Interfaces and Infrastructures
- Digital Humanities Theory and Methodology
- Society, Media, Politics, Engagement
- Language, Languages
- History and Historiographies

Papers from Taiwan

- Maps Re-imagined: Digital, Informational, and Perceptual Experimentations in Progress (Academia Sinica & NCKU)
- Toward Building Chronicles from Biographies in Local Gazetteers: An Application of Syntactic and Dependency Parsing (NCCU)
- Exploring Intertextuality in the Mahoyoga Section of the *Rin chen gter mdzod* (Dharma Drum)
- Onto Word Segmentation of the Complete Tang Poems (NCCU)
- Climate Event Classification Based on Historical Meteorological Records and Its Presentation on A Spatio-Temporal Research Platform (NCU & Academia Sinica & Dharma Drum)
- Some GIS-Based Analysis of the Complete Taiwan Poems (NCCU & Academia Sinica)





Paper: Repopulating Paris: massive extraction of 4 Million addresses from city directories between 1839 and 1922 (EPFL, Switzerland)

- <https://dev.clariah.nl/files/dh2019/boa/0878.html>
- [Quote] Extraction and analysis of the data contained in “Didot-Bottin” covering the period 1839-1922 for Paris, digitized by the Bibliotheque nationale de France.
- [Quote] Process more than 27,500 pages to create a database of **4.2 million entries linking addresses, person mention and activities.**
- [Quote] Report on the overall statistics of the distribution of professions in Paris and their evolution during more than 80 years, as well a general overview of the diversity of family names through time.

ROYAL;
ANNÉE M. DCC. LIII.
CONTENANT

Les Naissances des Princes & Princesses de l'Europe.
Les Archevêques, Evêques, Cardinaux, Abbés, Commandataires.
Les Maréchaux de France, les Lieutenans Généraux, Maré-
chaux de Camp, & Brigadiers des Armées, les Lieutenans
Généraux des Armées Navales, Chefs d'Escadres, les Che-
valiers, Commandeurs & Officiers des Ordres du Roy, les
Gouverneurs & Lieutenans Généraux des Provinces, &c.
Les Officiers du Roy, les Députés des Secrétaires d'Etat,
de des Intendants des Finances, les Conseillers d'Etat, les
Bureaux du Conseil, les Maîtres des Requêtes, les Intendants
des Provinces, la Grande Chancellerie, le Grand Conseil,
le Parlement, la Chambre des Comptes, la Cour des Aides,
l'Université, les Académies, les Bibliothèques publiques, &c.
Les Prévôts, les Juges, les Fermiers Généraux des Finances,
les Trésoriers des Deniers Royaux, les Payeurs des Rentiers,
& les Contrôleurs, la Compagnie des Indes, &c.

ET DE TOUTES LES ADRESSES DE L'ÉTAT
DE PARIS,
POUR L'AN VII.

(Un volume in-8.° de 700 pages, divisé en deux parties):

Classés par ordre d'États, Professions, Arts et Métiers, savoir: Agents, Baugiers, Commissionnaires, Courtiers, Négocians, etc., etc., etc., auxquelles on a joint les Noms et Demeures des Représentans du Peuple et leurs Départemens, le Directoire, les Ministres, les Tribunaux de Justice, Civils, de Commerce, Criminels et de Police-Correctionnelle, les Administrations Municipales et leurs Arrondissemens, les Bureaux-Centraux des Contributions, les Bureaux du Droit d'Enregistrement, des Patentes et des Droits de Consommation, l'Administration des Postes aux Lettres, le Départ et l'Arrivée des Courriers, la Distribution des Bureaux de ladite Poste dans Paris, les Bureaux de Poste-Timbres, de la Loterie Nationale, les Rentiers-Généralistes, &c.

ANNUAIRE ET ALMANACH
DU
COMMERCE,
DE L'INDUSTRIE,
DE LA MANUFACTURE ET DE L'AGRICULTURE,
ALMANACH DES 200,000 ADRESSES
DE PARIS, DES DÉPARTEMENTS ET DES PAYS ÉTRANGERS,
FERNIN DIDOT ET BOTTIN RÉUNIS

1857

ANNUAIRE-ALMANACH
DU
COMMERCE
DE L'INDUSTRIE
DE LA MANUFACTURE ET DE L'AGRICULTURE
ALMANACH 1,500,000 ADRESSES
DIDOT-BOTTIN
99^e ANNÉE DE LA PUBLICATION

1896

1896

PREMIÈRE PARTIE

ANNUAIRE
DU
COMMERCE
DIDOT-BOTTIN
100^e ANNÉE DE PUBLICATION
1922
PARIS - TOME I
AVEC, EN ANNEXE, L'ALPHABÉTIQUE DES PROFESSIONS
LE TOME II

1922



210

EUG

Etard, vins, Four-St-Honoré, 21.

Etat (ministère d'), place du Carrousel.

Etat - major (Ecole impériale d'application du corps de l'), Grenelle-St-Germ., 138.

name activity place street number

EYD

FAB

Eugénie (Mlle) modes Coq-St-Honoré, 3.

Eumont, architecte, Lalayette, 4.

Eurin et Cie, pastilleurs, Quincampoix, 31.

Eustache et Cie, bois et charbons de terre,

Eydt, bottier, Pélican, 11.

Eymard (N. et F. ✱) frères, bouchons, Temple, 14.

Eymard (J.-A.), conserves alimentaires, Tem-

Rasmussen (Vald), commiss. en librairie, quai St-Michel, 9. (5^e). **T.Gob. 24.33. Adr.†**
 VALDRAS-PARIS.
 Rasmussen et Kasembeck, ébénistes, r. de Charenton, 167. (12^e).
 Raspail (Sté an^{me} des Etabl^{ts}), produits hygié-

Rasnauss^n (Vald*, commiss. en librairie, l'laigt-llichel, 9. (51). MGob. 24.33. :

^LÛIUS-Pahis. ^ Adr,<r

R::mussen et Kasembeck, ébénistes, r. de Raspail -charenton, 167. (12.).

Alavoine (P.), fournitures de bureaux, fabr. de carnets, maroquinerie, r. St-Sabin, 52, imp. des Primevères, 5. (11^e). **TÉLÉPH 930.02.**
 Alavoine (L.) ✱ **NC** et Cie, tapissiers-ébénistes, r. Caumartin, 9, 9 bis et 11. (9^e). **TÉLÉPH 244.71.**

Alavoine (P.), fournitures de bureaux, fakr de carnets, maroquinerie, r. St-Sabm. __52. imp.desPrimevères, 5. Uf"). Klépr]930.0s:
 Alavoine (L.) !!ml et Cie, tapissiers-é bénis' tes, r. Caumartin,9,9 bis et 11.(Se).fglgigj

Ailleret jeune, cristaux et verreries décorés, Hauteville, 72.
 Ailleret, propriétaire, Vertbois, 38.
 Aillerot, vins, Ste-Opportune, 5.
 Aillet (Emile), vins, Bouloi, 5.
 Ailloud, boulanger, Lafayette, 233.

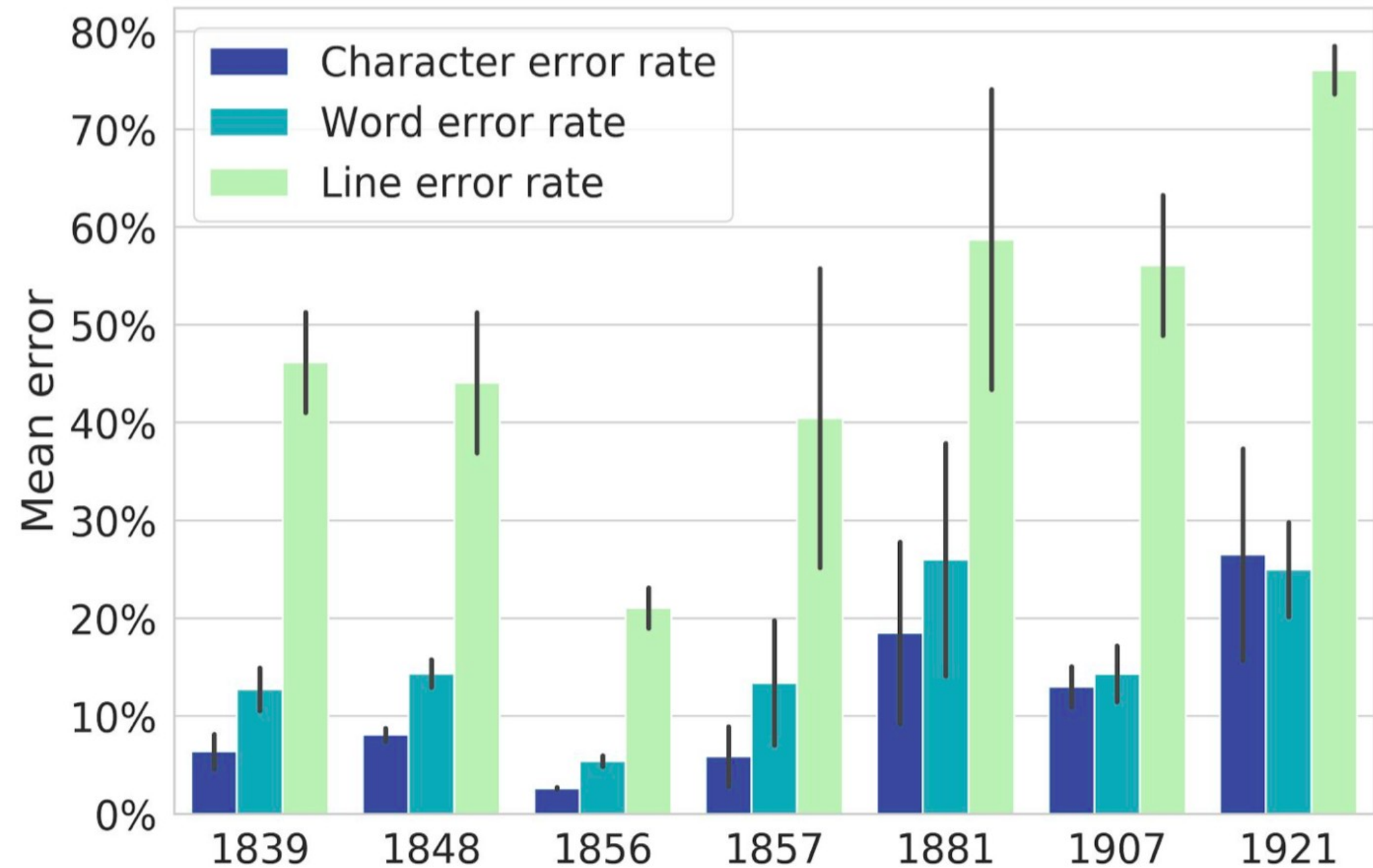
Ailleret jeune, cristaux et verreries décorés llautevill&, *72.-. . '

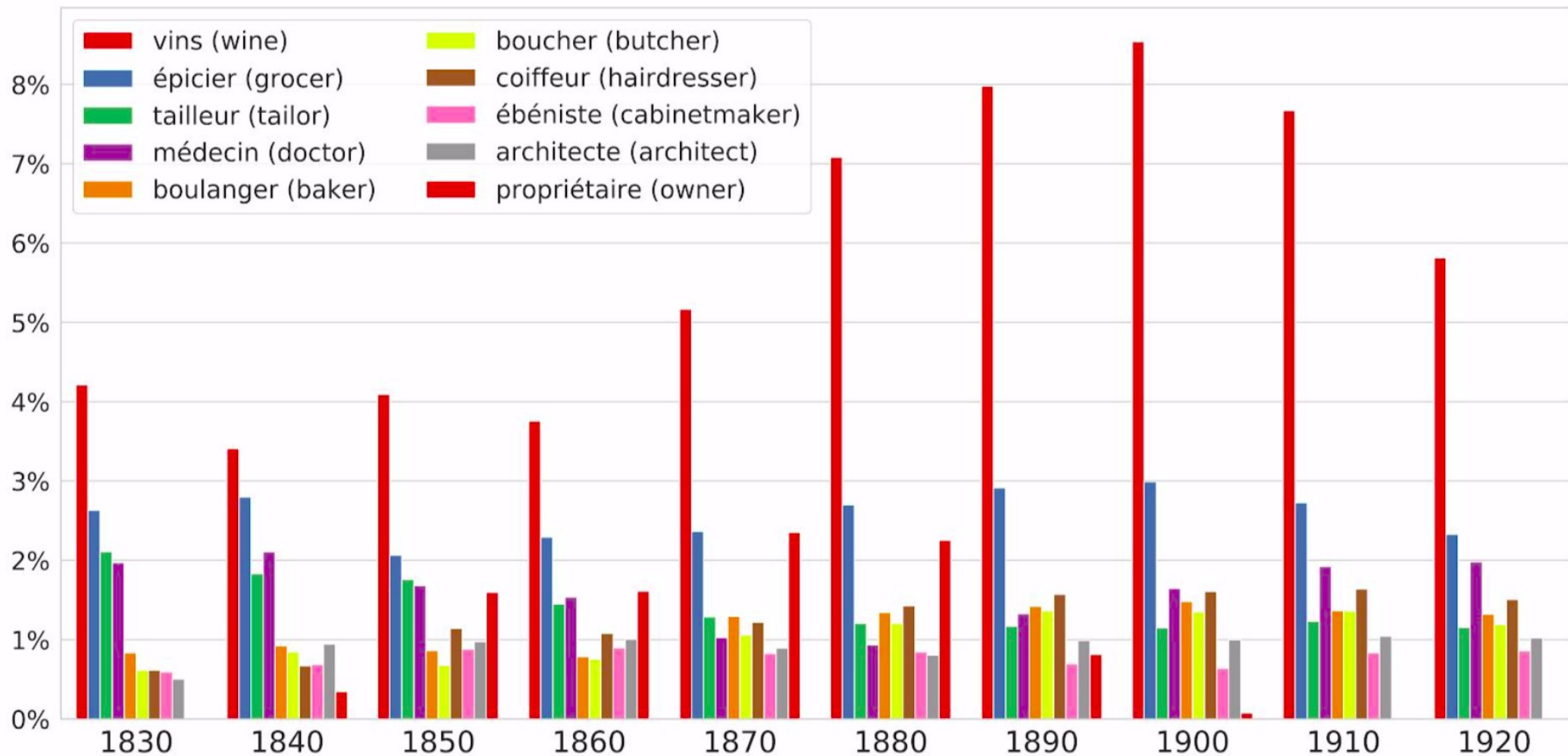
Ailleret, propriétaire, VcrUBois", 38.

4iîfert>t, Ste-Gppàrthrie, 5i

Aiilet (E 1 vlle, CotUoi. 5.

Ailloud, •boutiStde^, Eaftf.éwej;29a* "



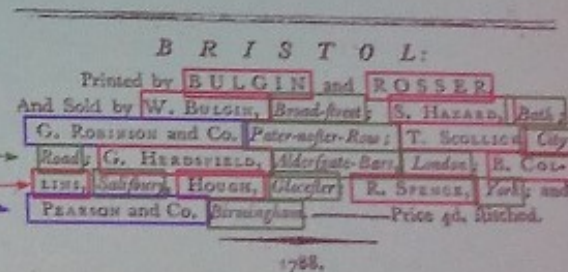


Paper: Publishers, Printers and Booksellers - Implications of Properly Structured Metadata for Digital History (University of Helsinki, Finland)

- <https://dev.clariah.nl/files/dh2019/boa/0650.html>
- [Quote] The English Short-Title Catalogue (ESTC, <http://estc.bl.uk>) is a standard source for analytical bibliographic research holding close to half a million titles published in English from 1473 to 1800.
- [Quote] A data matching process was applied to test if names detected in the **publisher statements** could be linked with **distinct entities found in external databases**.
- [Quote] The end result of the unification process is a bibliographic database of 900,000 non-unique records harmonized into fewer than 200,000 actors, of which **30,000 are identified as being part of the book trade**. Between these actors we have been able to make roughly **800,000 individual connections**.

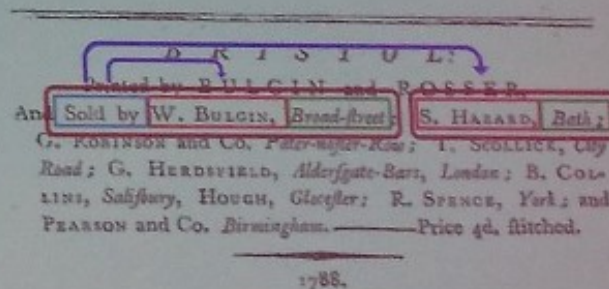
Step 1: Named Entity Recognition

- (Minor pre-cleanup before this step)
- Stanford NER parser
- Custom trained model:
 - LOCations
 - ORGANizations
 - PERsons
 - ROLES
- ~ 14k lines (1 token / line) of training data
- Works quite well for Latin, Welsh and French too



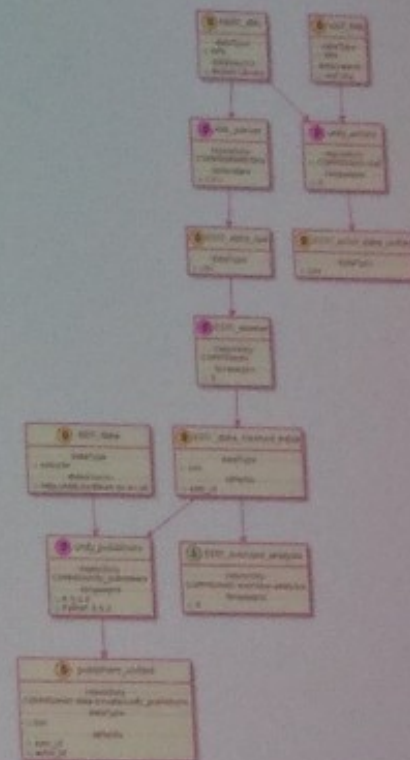
Step 2: Roles and relations

- Identify roles of actors in the statement (publisher, printer, bookseller), usually by a verb, eg.: "Sold by W. Bulgin"
- Identify groupings (which verbs refer to which actors, which locations and actors are connected)
- Rules based script - off the shelf NLP did not work that well with non-standard sentences in publisher statements.

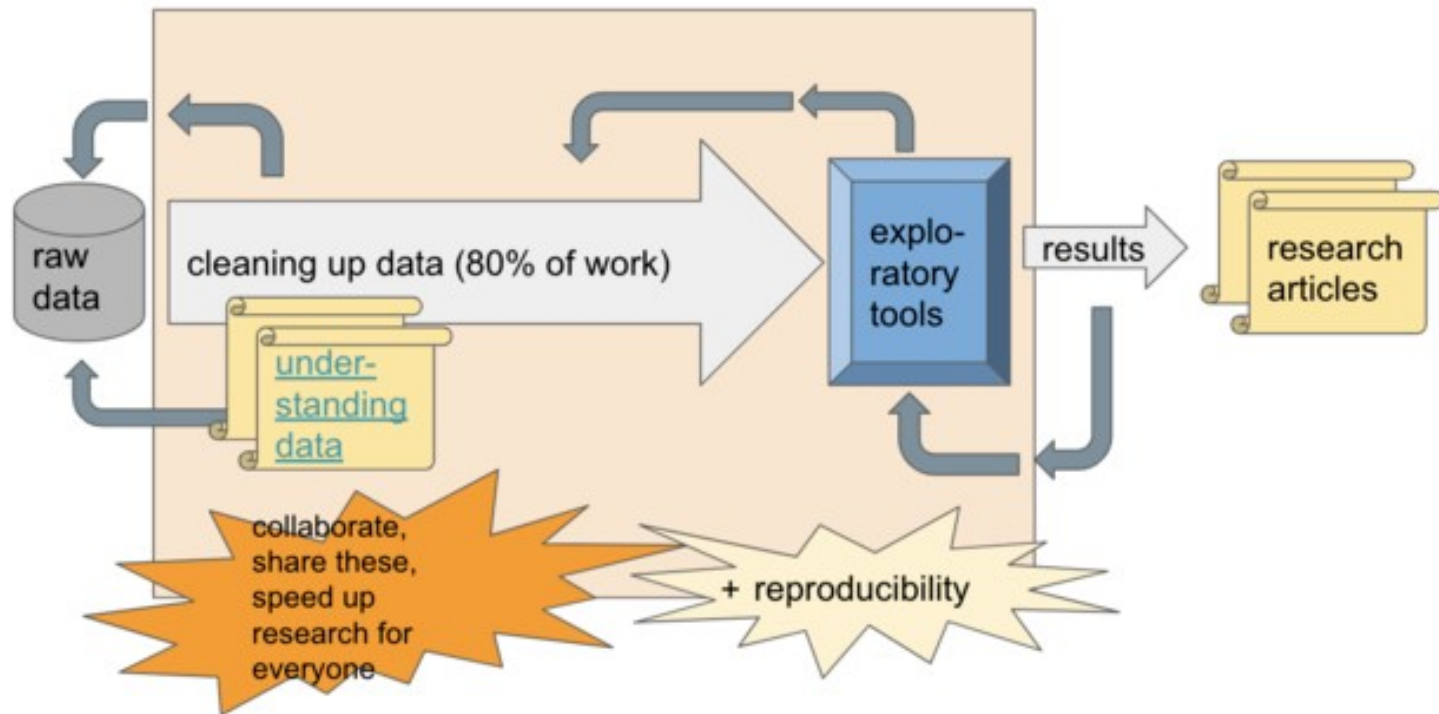


Format and logic behind it

- Relational model
- Distinct entities
 - actors and their connections
- Links with other data from processed, cleaned and unified version of ESTC
- Output in .csv files stored in a git repository
 - Deliberately not a web service, complicated linked data database, or something similar:
 - Robust: aims for **clarity, ease of use and sustainability**
 - Enables easy further development



Open science workflows to reduce individual workload



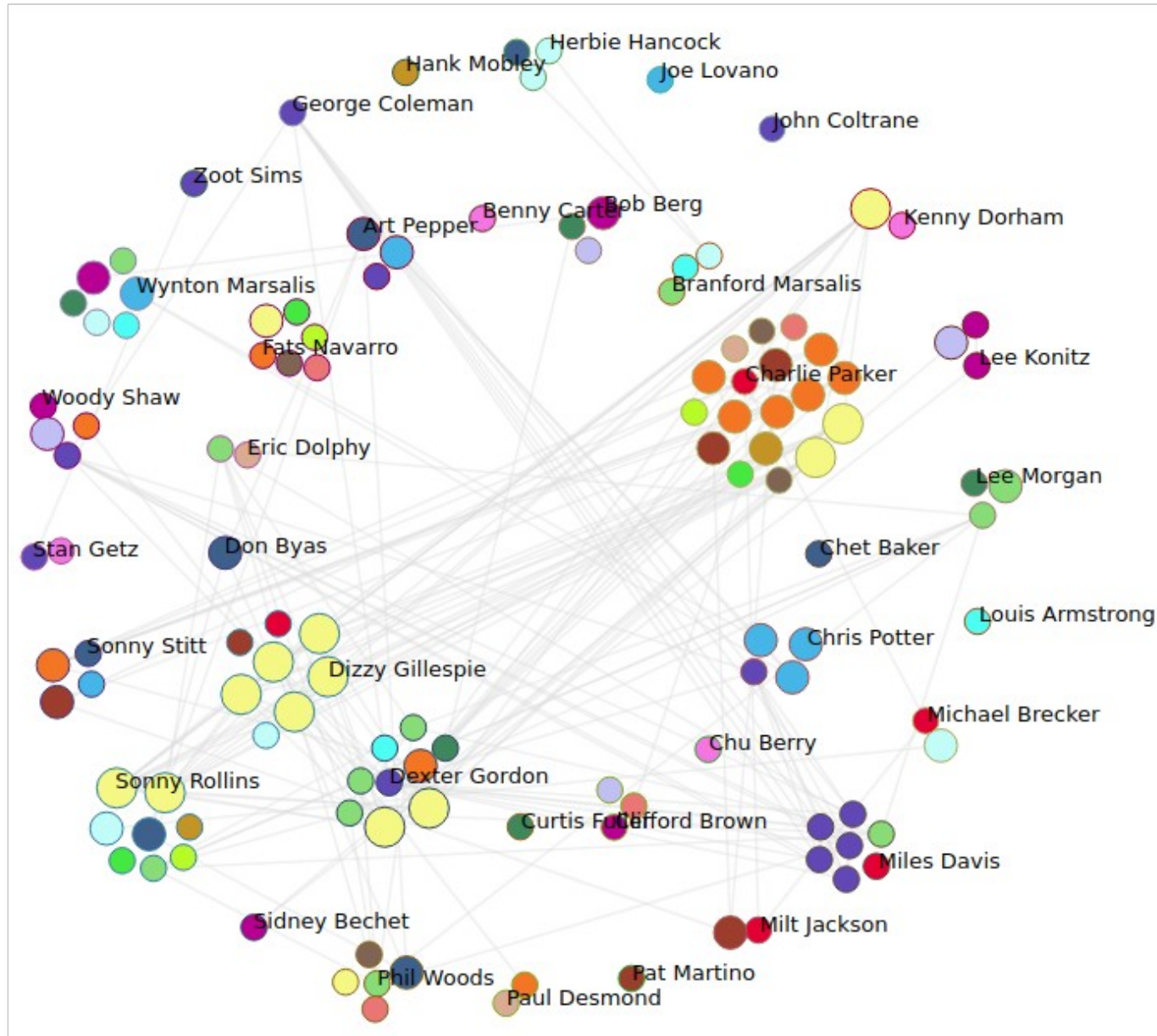
graph: Eetu Mäkelä & Mikko Tolonen

“Exploring ECCO: Key moments in 18th-century philosophical Literature”, Comhis-collective, Helsinki. <http://j.mp/comhis-nvm>

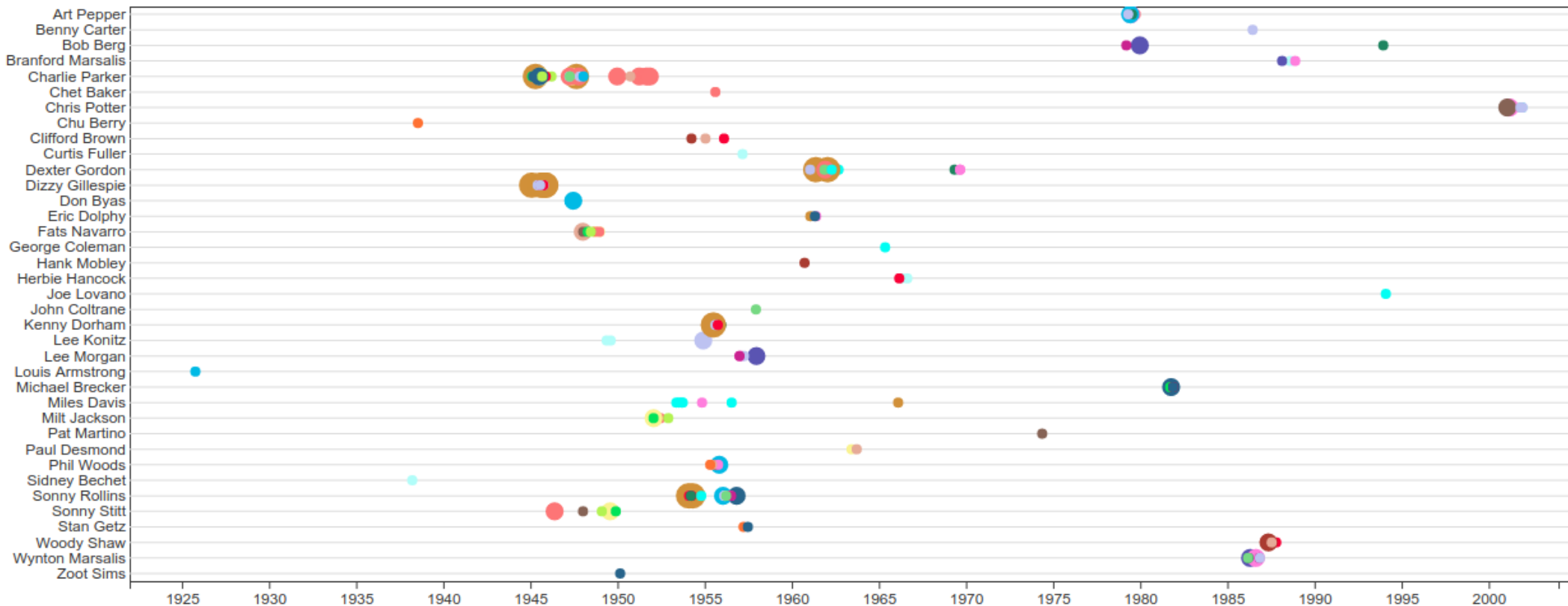
Paper: Digging Into Pattern Usage Within Jazz Improvisation
(Pattern History Explorer, Pattern Search and Similarity Search)
(University of Music “Franz Liszt” Weimar, Germany)

- <https://dev.clariah.nl/files/dh2019/boa/0723.html>
- [Quote] The main goal of the Pattern History Explorer is to enable the exploration of interval patterns in jazz solos by providing information from many different angles.
- [Quote] It maintains an overview of interval patterns frequently used in the Weimar Jazz Database, allowing for the discovery of cross-artist and cross-temporal relationships. Currently, **653 interval patterns with 11,630 instances are included.**
- [Quote] The similarity search operates on a database of the complete set of pitch and **interval n-grams of up to 20 elements** that were previously extracted from the Weimar Jazz Database.

[Quote] Network graph for interval pattern -1,-2,-1,3,3,3,-1,-2 grouped by performer; the biggest bunches stand for those patterns played by Charlie Parker, Dizzy Gillespie, Sonny Rollins and Dexter Gordon. Node colours denote identical patterns which are connected by edges. Node size represents the degree of similarity, where bigger means more similar.



[Quote] Timeline chart for interval pattern -1,-2,-1,3,3,3,-1,-2.
Node colours denote identical patterns. Node size represents
the degree of similarity, where bigger means more similar.









Panel: The Past, Present and Future of Digital Scholarship with Newspaper Collections

- <https://dev.clariah.nl/files/dh2019/boa/0377.html>
- Mia Ridge (BL)
 - Living with Machines <https://www.bl.uk/projects/living-with-machines>
 - Slides:
<https://www.slideshare.net/miaridge/living-with-machines-at-the-past-present-and-future-of-digital-scholarship-with-newspaper-collections-154700888>
- Maud Ehrmann (EPFL)
 - Impresso <https://impresso-project.ch/>
- Jean-Philippe Moreux (BnF)
 - News Eye <https://www.newseye.eu/>
- Laurel Brake (Birkbeck, Univ of London)
 - Nineteenth-century serial edition <https://ncse.ac.uk>
- Andrew Prescott (University of Glasgow)

Panel: Data Science (DS) & Digital Humanities (DH): new collaborations, new opportunities and new complexities

- <https://dev.clariah.nl/files/dh2019/boa/0364.html>
- Eirini Goudarouli (The National Archives) – Computational archive science
- Leonardo Impett (Bibliotheca Hertziana – Max Planck Institute) and Anne Alexander (University of Cambridge) – Way of machine seeing
- Mia Ridge (BL)
 - Infrastructure at the intersection of culture heritage and data science
 - Shared infrastructure as a platform for collaboration?
 - Slides:
<https://www.slideshare.net/miaridge/in-search-of-the-sweet-spot-infrastructure-at-the-intersection-of-cultural-heritage-and-data-science>
- Nora McGregor (BL)
 - Training GLAM staff to support digital scholarship
 - <https://blogs.bl.uk/digital-scholarship/>
- Beatrice Alex (Univ Edinburgh)
 - Interdisciplinary data science and digital humanity research
 - Plague.TXT – <https://www.ltg.ed.ac.uk/projects/plague-txt/>
- Barbara McGillivray (University of Cambridge) – A manifesto on DS & DH

Panel: Clearing the Air for Maintenance and Repair: Strategies, Experiences, Full Disclosure

- <https://dev.clariah.nl/files/dh2019/boa/0648.html>
- <https://drive.google.com/file/d/1DiA8X9aD9W8MzxJLWHHdYF38RY9gCSOm/view>
- James Smithies & Arianna Ciula (King's College London)
 - <http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html>
 - “Managing 100 Digital Humanities Projects...”
- Jessica Otis & Faolan Cheslack-Postava (George Mason University)
 - Roy Rosenzweig Center for History and New Media
 - [Quote] Near-zero hard funding; Unsustainable “free labor”; Continuous loss of institutional knowledge; Unfounded mandates for infinite lifetimes
- Martin Holmes & Stewart Arneil & Greg Newton (University of Victoria)
 - <https://hcmc.uvic.ca/endings/questionnaire.htm>
 - “How sustainable is my DH project?”
- Jasmine Mulliken (Stanford University Press)
 - “What parts of an interactive scholarship work should be preserved?”
 - “How long should an interactive scholarly work last?”
 - “Who is responsible for a work's longevity?”

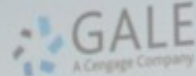
DH2019

DIGITAL HUMANITIES CONFERENCE 2019

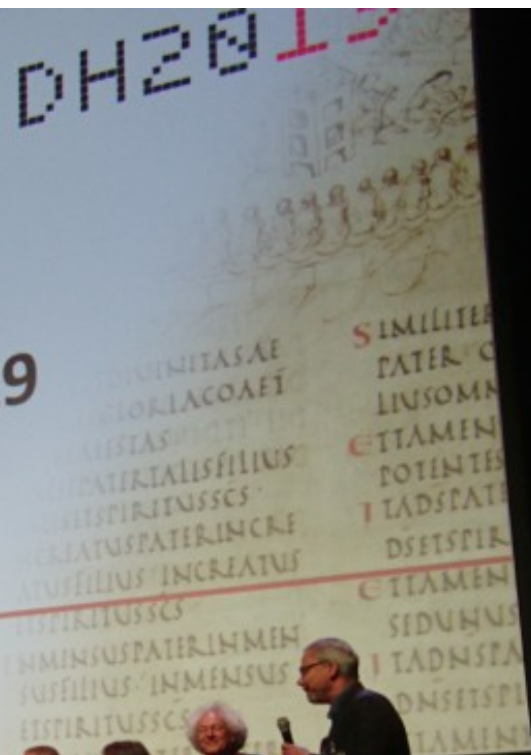
Local Organisers DH2019



Utrecht University



A Cengage Company



Conference Reports Found Online

- Mikko Tolonen, University of Helsinki

<https://www.newseye.eu/blog/news/a-perspective-on-research-on-digitised-newspapers-at-dh20190/>

- Mia Ridge, British Library

<http://www.openobjects.org.uk/2019/07/notes-from-digital-humanities-2019-dh2019-utrecht/>


- Albert Meroño Peñuela, Vrije Universiteit Amsterdam

<https://www.albertmeronyo.org/trip-report-dh2019/>

- ... and many more.

DH2020, DH2021, and DH2022

- DH2020 theme: carrefours/intersections (a place where paths cross)
 - July 22-24, 2020, Ottawa, Canada
 - Call for reviewers, deadline: August 15, 2019
 - <https://dh2020.adho.org/cfps/reviewers/>
 - @DH2020_Ottawa
- DH2021
 - July 26-30, 2021, Tokyo, Japan
 - @DH2021_Tokyo
- DH2022
 - July 11-15, Graz, Austria

09:36 Intercity 
3^a Enschede
via Amsterdam Z. Duivendrecht
Hilversum Amersfoort Apeldoorn
Deventer Almelo en Hengelo
Hilversum
Hilversum: 09:42 SPR Hilversum

