

GSD 351D – *Identities / Patterns / Code: Digital Approaches to Culture*

Short title: Identity-Codes-Culture

Unique: 38055, 34450, 43760.

Instructor: Thorsten Ries <thorsten.ries@austin.utexas.edu>

Semester: Fall Term 2022

To be cross-listed with *Linguistics, Anthropology, Comparative Literature*.

Flags: Global Cultures, Quantitative Reasoning.

DESCRIPTION

“What if I could read patterns out of hundreds of texts, and gain new perspectives, create new knowledge about them using digital tools?” This course explores how to read identities as statistical patterns in literary texts, linguistic, cultural and historical corpora with digital methods, and how we can come to a deeper understanding of individual texts and textual phenomena. This course introduces digital research methods, tools and use cases. Students will work hands-on with literary, linguistic, cultural and historical sources – without requiring any previous programming knowledge.

We will start the course with an overview to concepts of “identity”, “patterns” and “models” in literature, culture, and humanities research. Our discussion will show how digital methods and tools transform the way literary, linguistic and culture studies research conceptualize identity, culture, as well as textual phenomena since the 20th century as data patterns and models. This course will consider digital research methods from the perspective of concepts of “identity” and “identities” as intersectional compounds of gender identity, sexual orientation, ethnicity, nationality, culture, religion, social and historical background, variation of physical and mental abilities, etc. - Quantitative-reasoning based digital methods and digital models offer the opportunity to analyse these components as more general patterns at scale, while preserving the complexity of their interconnection. At the same time, digital technology and digital research methods deserve our critical attention as well: do these methods contribute to equality, equity, or may their application introduce biases?

After the introduction, we will first move on to an introduction to basic concepts of statistics that will help us to understand the data and then to the hands-on work extracting and “distant reading” literary, cultural and linguistic patterns from textual sources, social science and online discourse datasets with digital methods and tools. In hands-on group projects (e.g. newspaper and online sources, literary texts, or sources in the language of your discipline, e.g. the Texas German project), you will learn how to think your own research digitally, ask new questions, extract, analyze and “read” literature and culture as data patterns: next to useful digital skills for your everyday work (e.g. corpus management, webscraping) you will learn hands-on to leverage text statistics, and methods like *named entity recognition*, *co-occurrence* and *collocation analysis*, *sentiment analysis*, and *topic modeling* to identify patterns and networks of identity features in literary texts, linguistic and historical cultural corpora as well as digital culture corpora – or the identity of a text’s author by its linguistic style features (*stylometry*). We will also learn digital social network analysis (SNA) skills and reflect on the methodology on the background of ANT (actor network theory).

Your practical experience with these digital methods and the results of your research will be the basis of our

(critical) discussion of these methods in comparison to the way you “read” literature, linguistic sources and culture before. These group projects – like the entire course - do NOT require any previous programming knowledge. Everything you need you will learn during this class (e.g. essential Python skills, ability to use different digital tools). All you need is willingness to hands-on explore digital methods and tools in class and in group projects.

A large portion of this course will be student-centered: throughout the semester, you will conduct three self-guided DH group projects of your own design and in your own interest area, which will help you learn to plan, manage and implement your project, and collaborate effectively. These projects will be supported by a combination of in-class instruction with available supervision and support, as well as student peer-review. The final section of the course will be dedicated to finishing these projects, documenting and reflecting on your results, in-class presentation.

This course is approved for credit in the Undergraduate Digital Humanities Undergraduate Certificate (<https://liberalarts.utexas.edu/ds/undergraduate/index.php>).

Please note

- No background in programming is required, only the willingness to explore digital tools.
- This course will be taught in English. Main readings will be available in English (or in other languages if projects require it, if applicable for chosen projects).
- All software will be open source and cross-platform (no special costs beyond having device access in class).
- This course is approved for credit in the Digital Humanities Undergraduate Certificate (<https://liberalarts.utexas.edu/ds/undergraduate/index.php>). Please contact certificate director after completion of the course.

Class format/ method of instruction:

- Class will be conducted in an in-person format. Hybrid options are available for students with specific needs. The in-class, hands-on format is complemented with flipped classroom, group project formats.
- Digital learning / teaching with Canvas, potentially Zoom, hypothes.is, other.
- Python IDE's suitable for beginners (PyCharm, Google Colab), other (Voyant Tools, AntConc, etc).

COURSE OBJECTIVES

The main objectives of the course are to

- offer an overview of and practical knowledge of DH methods and methodology, as well as the ability to critically evaluate methods and datasets.
- acquire essential digital skills they can build on and extend in future courses and their own projects.
- develop digital humanities research questions, project plans, implementation and evaluation.
- engage with existing DH research in a critical way.

COURSE OUTCOMES

By the end of the course, students will be able to

- apply a basic set of DH methods and critically evaluate methods, application and datasets;
- integrate literary and culture theory, text analysis with concepts of DH methods and theory;
- analyze cultural texts in a historical context with digital methods;
- work collaboratively in an interdisciplinary and comparative fashion.

GRADING

The grading is divided into three main rubrics:

- 30% Participation (10% in-class participation, 10% presentation of class material, 10% conference paper presentation),
- 60% Group Projects (three Digital Group Assignments with a short summary report: 20% each, consisting of 10% group project documentation (code, corpus, results, visualizations, interpretation, methodological description and reproducible description) and 10% individual reflection),
- 10% Research Paper.

Your group projects, the research paper and the in-class presentations (class material, conference paper) will be graded with the following criteria:

- 40% QR-methodology and documentation.
- 20% QR-related conceptual rigor, correctness.
- 40% general argument and formal aspects.

Preliminary reading list

General Perspectives: Introduction and Methodology in Digital Humanities

- Johanna Drucker. *The Digital Humanities Coursebook. An Introduction to Digital Methods for Research and Scholarship*. Routledge 2021.
- Andrew Piper, *Enumerations: Data and Literary Study*, (Chicago: University of Chicago Press, 2018)
- Melissa M. Terras, Julianne Nyhan, and Edward Vanhoutte. *Defining Digital Humanities: a Reader / Edited by Melissa Terras, Julianne Nyhan, Edward Vanhoutte*. Farnham, Surrey, England: Ashgate Publishing Limited, 2016.
- Ted Underwood. *Machine Learning and Human Perspective*, PMLA 135.1 (Jan 2020): 92-109.
- Ted Underwood. *Distant Horizons: Digital Evidence and Literary Change* (Chicago: University of Chicago Press, 2019).
- Annette Vee: *Coding Literacy: How Computer Programming Is Changing Writing* (Software Studies), MIT Press 2017.

Debates on DH and identity

- Kelly Baker Josephs, Roopika Risam (eds). *The Digital Black Atlantic. Debates in the Digital Humanities 6*, University of Minnesota Press 2021. URL: <https://dhdebates.gc.cuny.edu/projects/the-digital-black-atlantic>
- Matthew K. Gold, Lauren F. Klein (eds). *Debates in the Digital Humanities 2019. Debates in the Digital Humanities 5*, University of Minnesota Press 2019. URL: <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2019>.
- G. Griffin. *Intersectionalized Professional Identities and Gender in the Digital Humanities in the Nordic Countries*. *Work, Employment and Society*. 2019;33(6):966-982. doi:10.1177/0950017019856821
- Johnny Pitts: *Afropean: Notes from Black Europe*. Allen Lane 2019.
- Jacqueline Wernimont, Elizabeth Losh (eds). *Bodies of Information. Intersectional Feminism and Digital Humanities. Debates in the Digital Humanities 4*, University of Minnesota Press 2018. URL: <https://dhdebates.gc.cuny.edu/projects/bodies-of-information>.

Case Studies and Methods

- Nan Z. Da, *The Computational Case against Computational Literary Studies*. *Critical Inquiry* 45.3 (2019). pp 601-639.
- J. Berenike Herrmann, *In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics*, *DHQ* 11.4 (2017), URL: <http://www.digitalhumanities.org/dhq/vol/11/4/000341/000341.html>.

- Fotis Jannidis, Julia Flanders. Data modeling in a digital humanities context: an introduction. In: J.F./F.J.: The Shape of Data in Digital Humanities: Modeling Texts and Text-Based Resources. London: Routledge 2018, p. 3-25.
- Fotis Jannidis, Stefan Evert, Thomas Proisl, Isabella Reger, Steffen Pielström, Christof Schöch, Thorsten Vitt. Understanding and explaining Delta measures for authorship attribution. In: Digital Scholarship Humanities 2017. DOI: <https://doi.org/10.1093/llc/fqx023>.
- Folger Karsdorp, Mike Kestemont, Allen Riddell. Humanities Data Analysis. Case Studies with Python. Princeton University Press 2021.

Relevant Journals for Digital Scholarship

- DSH [Digital Scholarship in the Humanities](#) (former LLC)
- DHQ [Digital Humanities Quarterly](#)
- IJDH [International Journal of Digital Humanities](#).

Course Schedule (Draft)

Week 01 – Introduction / Research Questions			
1	Introduction About this course, formal introduction.	First quotes for discussion from research: “What is identity – as a concept, in numbers, in text?”, which impact has digital technology on our identity concepts, and intersectionality?	
2	The value of numbers: statistics, identity and pattern analysis in the Humanities What are identity concepts, and what happens if we try to see them through numbers, statistics?	Discussing excerpts from Underwood (2020), and DH Debates. Discussing statistical expressions of gender, religion, culture, orientation, ethnicity, ability and neurodiversity, social background. A short history and current use cases (excerpts) of quantitative and empirical reasoning in history, social sciences, linguistics and literary studies. Identity concepts: the Census and early DH history.	
Week 02 – Introduction to Statistical Concepts, Evaluation, and Research Design			
3	Introduction to concepts of textual statistics, text as data / data as narrative Statistical markers of identity concepts.	Analysis, in-class: Presidential speeches and tweets (with Voyant Tools, AntConc and a minimal Python example on Colab).	
4	Text as data / data narratives basics I Word frequencies, collocation, cooccurrence, words and texts as vectors. Conceptual and statistical basics, data interpretation, evaluation and narratives, and basic computational concepts.	In-class group exercise with AntConc and a ready-made Python example on Colab: “What changes if we change this parameter?”	
Week 03			
5	Statistics basics / data narratives II Correlation and regression Conceptual and statistical basics, data interpretation, evaluation and narratives about correlation, regression and identity (case-based discussion), and basic computational concepts (Python packages, code snippet, Colab).	In-class group exercise after the introduction into the stat. basics with a ready-made Python example on Colab: “Experiment with correlation and regression on two minimal examples” Online discussion, out-of class: Discussion of a small sample research paper: examine the research design, concepts, data and data narrative with respect to the encoding of identity concepts.	
6	Performance measures and bias Start group projects Introduction to statistical and ML performance measures. Basic concepts of bias,	In-class group exercise: Critique a sample study for QR biases, faults. Online discussion: What is a good research question?	<i>Start Group Project Phase I (research question development, corpus / data preparation)</i>

	and examples how bias is introduced into methods, datasets and data / results interpretation.		
Week 04 – Python basics 101.1: Syntax, basic functions			
7	Introduction to Python – Basics I Basic syntax and data operations	Switch from Colab to PyCharm to learn Python basics.	
8	Introduction to Python – Basics II Basic syntax and data operations		
Week 05 – Python basics 101.2: Working with datasets / dataframes			
9	Introduction to Python – Pandas I Working with data frames, structured data at scale, calculations and visualizations		
10	Introduction to Python – Pandas II Working with data frames, structured data at scale, calculations and visualizations		
Week 06 – Python basics 102.1: Text statistics			
11	Distribution, collocation, co-occurrence I Text by the numbers with Python	Week 6 small group project: statistical text analysis of a limited data / text sample for (gender, religion, culture, orientation, ethnicity, ability, social background) identity discourse markers of the group's choice. Goal: choose a method to investigate an identity aspect of a discourse / text, adapt the code to do what it has to do and evaluate the validity of the results	<i>Deadline Group Project Phase I</i>
12	Distribution, collocation, co-occurrence II Text by the numbers with Python		<i>Start Group Project Phase II (choice of methods, implementation plan, implementation, preliminary evaluation)</i>
Week 07 – QR perspectives and research designs			
13	Identity, Pattern, Data Quantitative research designs on identity, a question of concepts, methodology, and context.	Discussion, based on a (controversial) statistical research example (i.e. Ian Lancashire: Cybertextuality by the Numbers = flawed research on neurodiversity, text as symptom) vs. a qualitative, "humanities" example: diversified, intersectional identity concepts, how to translate into numbers, statistics, text analysis?	
14	Workshop Session for Workgroups Research design.	Self-guided work on the group projects, start of group project phase II: Choice of statistical methods, concepts of research questions, validity of results.	
Week 08 – Python basics 102.2: text statistics and topic modeling (patterns of identity)			
15	TF-IDF Term frequency-inverse document frequency or: the most salient words	TF-IDF is an important concept to start thinking about meaning / semantics as statistical pattern or word vectors. In-class explanation of the method and code that	

		analyses presidential speeches, then filtering for identity politics marker, and stylometric-thematic features.	
16	Sentiment Mining Calculating Polarity	Sentiment Mining can be based on TF-IDF, and measures semantic sentiment values. We will be using several forms of Sentiment-Mining to compute the opinions about the TF-IDF-"topics"	
Week 09			
17	Topic modeling I Gensim, word2vec, doc2vec	Next to learning about the QR-basis of different versions of topic modeling, students conduct the	
18	Topic modeling II LDA + critical evaluation	Week 9 short group project: Apply TF-IDF, word2vec and LDA to a small corpus, compare the results, change the parameters, document the changes of the results and explain! Consider the advantages and limitations of each method for your area of interest!	
Week 10 – Identification patterns: Authorship Verification via Stylometry			
19	Python: Authorship verification I	Examples with JGAAP, stylo / Delta (statistical methods that measuring different linguistic features).	
20	Python: Authorship verification II	Examples with Impostor Method (base: statistics). Week 10 short group project: identify an author, conduct style cluster analysis: which author verification method delivers the better results, explain why! If possible, come up with performance metrics and explain, interpret them!	
Week 11 – Parsing and measuring identity patterns – a Critical Code Studies example			
21	Python: NER and BookNLP	In-class experimentation with reproducing Underwood's results with different texts, critical code analysis of the gender-analytical functions of BookNLP. How may structural bias be introduced by the way BookNLP is implemented, how should this be improved?	<i>Deadline Group Project Phase II</i>
22	Python: BookNLP critical evaluation		<i>Start Group Project Phase III (evaluation, documentation, interpretation, write-up)</i>
Week 12 – Networked identity – Identity in Networks			
23	Basic theory of Social Network Analysis Network metrics and ANT	In-class introduction to SNA and network metrics, theoretical and mathematical background, ANT. In-class experiment with DraCor (Drama Corpus).	
24	Python: Web, social media and Internet Archive scraping		

Week 13			
25	Python: NetworkX, community metrics and how to read them		
26	Python: Reddit, social media network analysis	Visualize your favourite Reddit subreddit, and explain which contributors play which role, based on the metrics we learned!	
Week 14 – Group Project Sprint / Conference Papers			
27	Session for Workgroups		
28	Conference Paper Presentations		<i>Deadline Group Project Phase III</i>
Week 15			
29	Conference Paper Presentations		
30	Last Session		