

Introduction to Digital Literary Studies

Session 1 History & Theory: Introduction

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

Introduction Round

Name; Affiliation; “Why this course?”

What does literature let us analyse?

More broadly: what does a text “do”?

What does the transmission of literature tell us about the idea of literature itself?

What does the transmission of literature tell us
about the idea of literature itself?

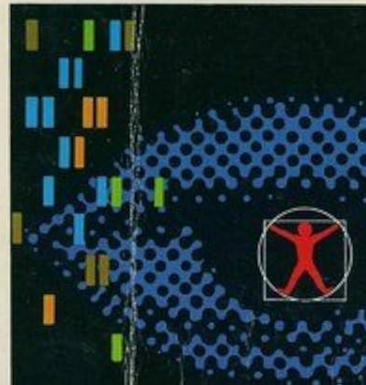
What does the medium tell us about the message?

A MENTOR BOOK  451-ME1986 • \$2.95

MARSHALL McLUHAN

UNDERSTANDING MEDIA

THE EXTENSIONS
OF MAN



“What we are considering here, however, are the psychic and social consequences of the designs or patterns as they amplify or accelerate existing processes. For the “message” of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs. The railway did not introduce movement or transportation or wheel or road into human society, but it accelerated and enlarged the scale of previous human functions, creating totally new kinds of cities and new kinds of work and leisure.”

“During long periods of history, the mode of human sense perception changes with humanity’s entire mode of existence. The manner in which human sense perception is organized, the medium in which it is accomplished, is determined not only by nature but by historical circumstances as well.”

- Walter Benjamin, “The Work of Art in the Age of Mechanical Reproduction”



Key Points

- literature as culture depends upon transmission of knowledge as we process it
- the transmission of knowledge depends upon the medium upon which data exists as portals of access to us
- studying and understanding this as a “mode of perception” can lead to reconfiguring how we can combine the different technical aspects with our critical perspective
- finding ways in which our technical tools allow us to understand data or not understand it is also a part of the critical perspective that we need

Group Work/Reflection Questions

1. What aspects of literary texts interest you? Can you find a quantifiable way of looking at a text in a critical manner?
2. What literary text or aspects of a literary text do you think cannot be quantified or studied in a quantifiable or visualisable study? Why do you think so?
3. What is the most dominant form of “perception” that you could use in your study of culture and literature? Think of a digital tool that you think is a useful way of reading literature today.

Introduction to Digital Literary Studies

Session 2 History & Theory: Research Paradigm

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

In the last episode...

1. literature as culture depends upon transmission of knowledge as we process it
2. the transmission of knowledge depends upon the medium upon which data exists as portals of access to us
3. studying and understanding this as a “mode of perception” can lead to reconfiguring how we can combine the different technical aspects with our critical perspective
4. finding ways in which our technical tools allow us to understand data or not understand it is also a part of the critical perspective that we need

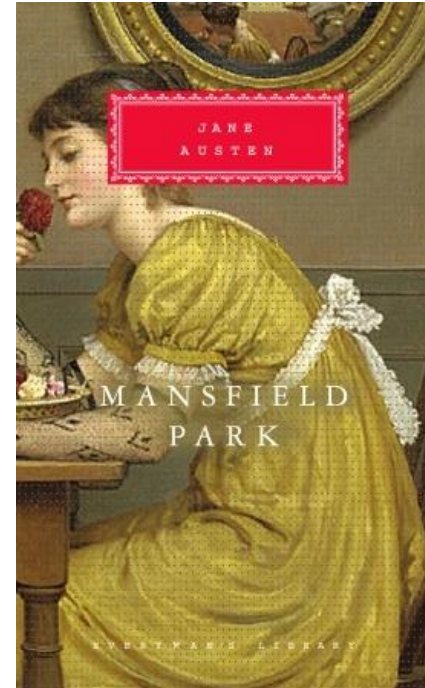
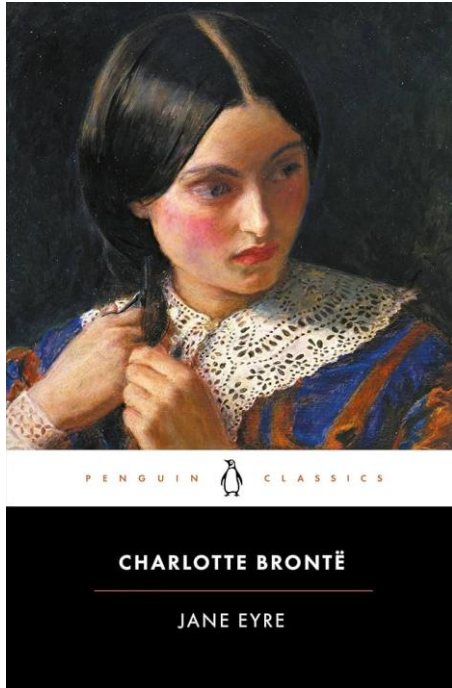
What makes a novel a “classic?”

What are the classical novels that you can think of and what did you “learn” from it?

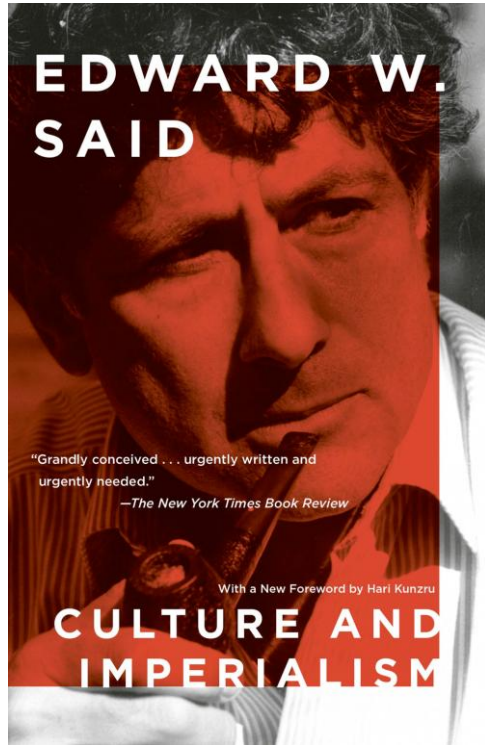
What makes a novel a “classic?”

- question of value that can be transferred from a reading experience to ... what exactly?
- question of pedagogical value that is inherent in a good work of literature
- question of the socio-cultural value of such a work

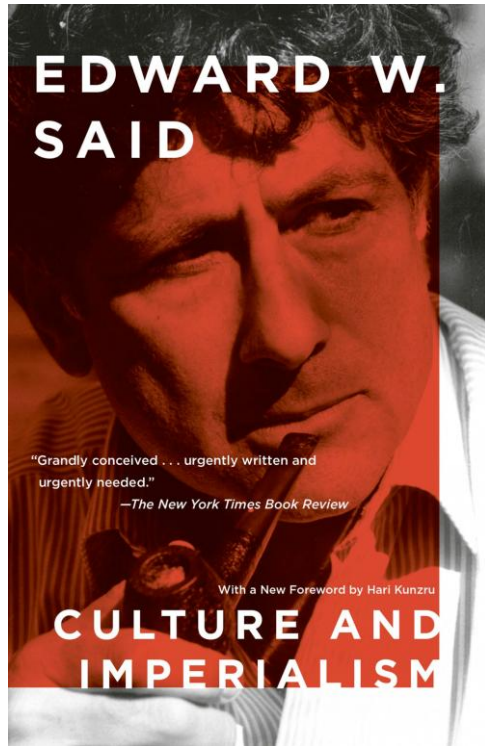
Let's look at two classics



- 19th century as the temporal site of the novel's setting
- the moral economy of the novel is steeped in the colonial policies of the time
- the background of the novel has the representation of the overseas territories of the empire where men like Sir Thomas (*Mansfield Park*) and Edward Rochester (*Jane Eyre*) made their fortunes
- allows us as researchers to look at the texts as “archives” of the historical times that we belonged to



“part of our difficulty today in any connection [between colonial art and criticism] is that we tend to reduce this complicated matter to an apparently causal one, which in turn produces a rhetoric of blame and defensiveness”



to explore and understand
“the signifying power of the
references to Antigua in *Mansfield
Park*; how do they occupy the
place they do, *what are they doing
there?*”

- texts are embedded within cultures that are determined by power relations <think Foucault's idea of discourse here>
- these backgrounds are references to historical sites that influence our “modes of perception” <insert Benjamin here>
- these questions of representation and perception, being always in flux, are also embedded in what is the “normative” at a point of time <insert Said; Raymond Williams; Frederic Jameson>

What is mappable if we were, in fact, able to map these questions?

RE-MAPPING MANSFIELD PARK: DECOLONIZING NINETEENTH-CENTURY BRITISH
CANONICAL FICTION THROUGH DIGITAL VISUALIZATION TOOLS

By
NAZUA IDRIS



Group Work

- Find one academic argument/thesis statement that you can make based on the information that you find on the visualisation website
- Find one way in which you could improve this project; ask Henny and Luca if that is a possibility.
- Discuss between/amongst yourselves which book would you use for a project of visualisation if you could do so.

Introduction to Digital Literary Studies

Session 3

History & Theory: DH Workflows

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje

luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

The typical workflow of a Humanities scholar

What do we do when working on a research question in the humanities/literary studies?

The typical workflow of a DH project

What do we do when working on a research question in DH?

What are the similarities and differences between the two kinds of research projects?

DH workflows

- Materials → Processing → Presentation
- Fundamental activities of digital humanities research:
 - mediation/remediation
 - datafication/modelling
 - processing/analytics
 - presentation/display
 - sustainability/preservation

The three fundamental quality criteria in empirical research

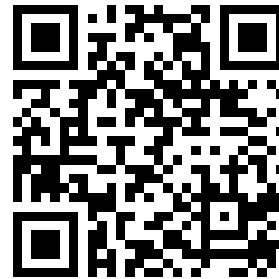
- **Objectivity**
 - Degree to which findings are independent of researcher bias
- **Reliability** (Reproducibility)
 - Consistency of measurement across time and conditions
- **Validity**
 - Whether the research truly measures what it claims to

Let's take a closer look at a DH project

Forgotten Books

Mike Kestemont, Folgert Karsdorp, Elisabeth de Bruijn, Matthew Driscoll, Katarzyna A. Kapitan, Pádraig Ó Macháin, Daniel Sawyer, Remco Sleiderink, Anne Chao (2022): Forgotten books: The application of unseen species models to the survival of culture. *Science* 375(6582): 765-769.

<https://doi.org/10.1126/science.abl7655>



Presented on this website:

<https://forgotten-books.netlify.app/>



DH workflows: *Forgotten Books*

- What are the materials?
- Were they remediated?
- Are they modelled and how?
- How are the materials processed?
- How is the project presented?

Group work, 2 people, 15 minutes

Let's try to design a DH workflow

1. Development of gendered representation of the bodily appearance of characters in literature in the last 300 years
2. Dominance of urban and rural places in 19th century British literature
3. Evaluation of “Otherness” in 21st century literature

Group work, 3-4 people, 20 minutes

Introduction to Digital Literary Studies

Session 4

History & Theory: Operationalisation & Annotation

What happened last time?

DH Workflows

DH workflows

- Materials → Processing → Presentation
- Fundamental activities of digital humanities research:
 - mediation/remediation
 - datafication/modelling
 - processing/analytics
 - presentation/display
 - sustainability/preservation

Canon of world literature

- What is a “canon”?
- What is “world literature”?

Canon of world literature

- How did you go about creating your list?
- What titles are in your list of the canon of world literature?
- Which factors influenced your decision to include or exclude a title from your list?
- Which of the factors that influenced your decision can be determined **objectively**?

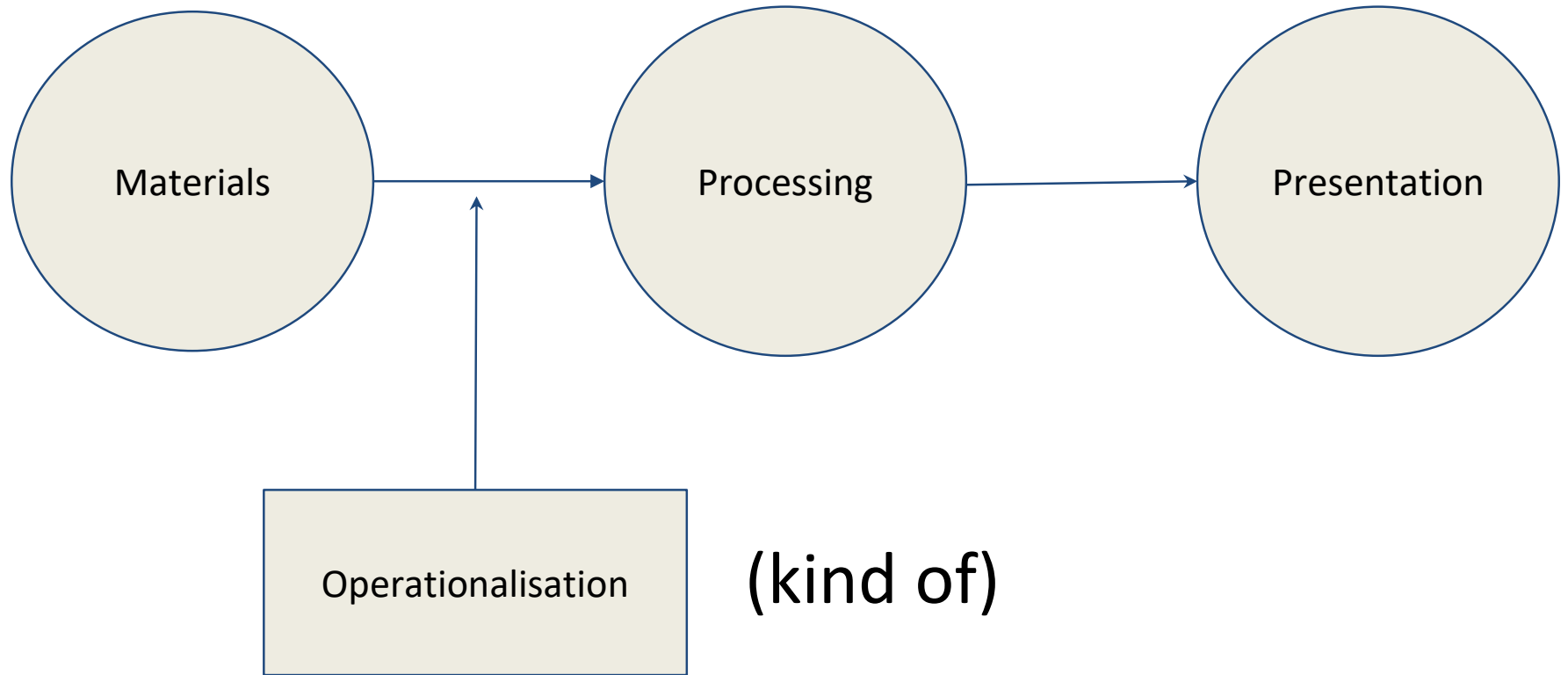
Operationalisation

Process via which concepts of humanistic inquiry are operationalised, that is, made **observable, measurable, formalised into rules** (from algorithms to software systems and applications) hence creating empirical objects of study which bear **theoretical consequences for the discipline** to which they are applied.

([Ciula et al. 2023: 13](#))

An operationalisation of canon

“A text is more likely to be canonized if (1) there is an edition of the complete or collected works of the author, (2) student editions of the text are available, and if it is mentioned in (3) exclusive narrative literary histories and anthologies and (4) other academic literature.” ([Brottrager et al. 2021: 197](#))



Back to our example (canonicity)

- Materials
 - What is the material?
 - How do we get to the material?
- *Processing*
- *Presentation*

Annotation

“Annotation means the (semantic) enrichment of research objects”

- in a narrower sense: “information which refer to a research object as a whole (meta information)”

👉 Example: metadata we collected about titles in our canon

(cf. [Seltmann & Horstmann 2024](#), German)

Operationalisation

Operationalization is the threefold task of (1) developing an adequate intensional definition, (2) identifying the steps necessary to decide whether a concept applies to a text or text passage and (3) carrying out this decision on a text or text corpus ([Jacke 2025: 4](#))

Operationalisation workflow

1. Develop definition
2. Identifying steps necessary to find out if it applies to the text
3. Carrying out the decision on the text (e.g. via **annotation**)

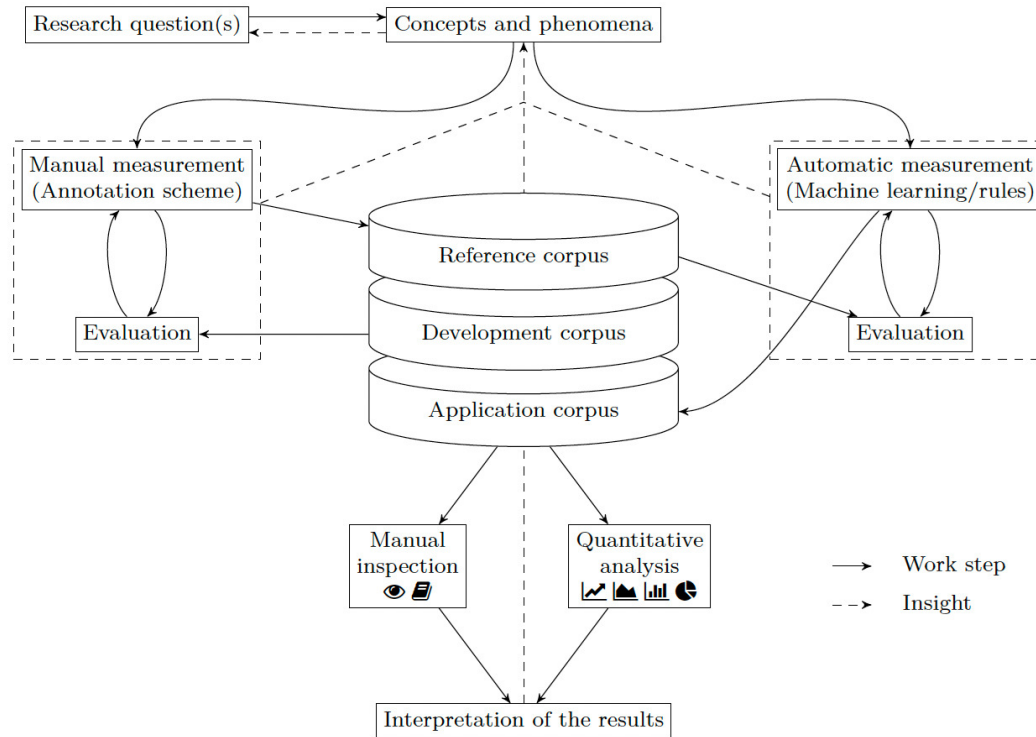
Example:



***Fear** in the gothic novel*

- *What needs to be defined?*
- *What steps do we need to take?*
- *How can we carry them out?*

Operationalisation workflow



Carnapian quality criteria of explications

- **Fruitfulness:** The operational definition must help researchers *discover new insights* about the text or corpus. This means preserving enough conceptual richness so that the operationalization remains relevant and interesting.
- **Simplicity:** one should strive for *conceptual clarity* and a reasonable level of simplicity — so long as it does not erase the complexity needed for the concept to remain faithful to scholars' intentions.
- **Faithfulness:** The operational definition must remain *close enough to the original meaning* of the concept so that researchers (especially those in “traditional” humanities) still recognize it.
- **Exactness:** An operationalized concept should be spelled out in a clear, intensional definition (i.e. specifying necessary and sufficient conditions). *A rigorous definition* is needed to ensure the concept can actually be applied and discussed consistently.

based on: Jacke (2025): 9

Operationalising literary concepts

Choose your literary concept or take one of these:

1. Suspense in the detective fiction in the 19th century
2. Tragic hero in classical drama
3. Flashback in films
4. Domestic space in the 19th century novel

and think about possible operationalisation for these concepts –
focus on the operationalisation, not the whole workflow!

Group work, 3-4 people, 20 minutes

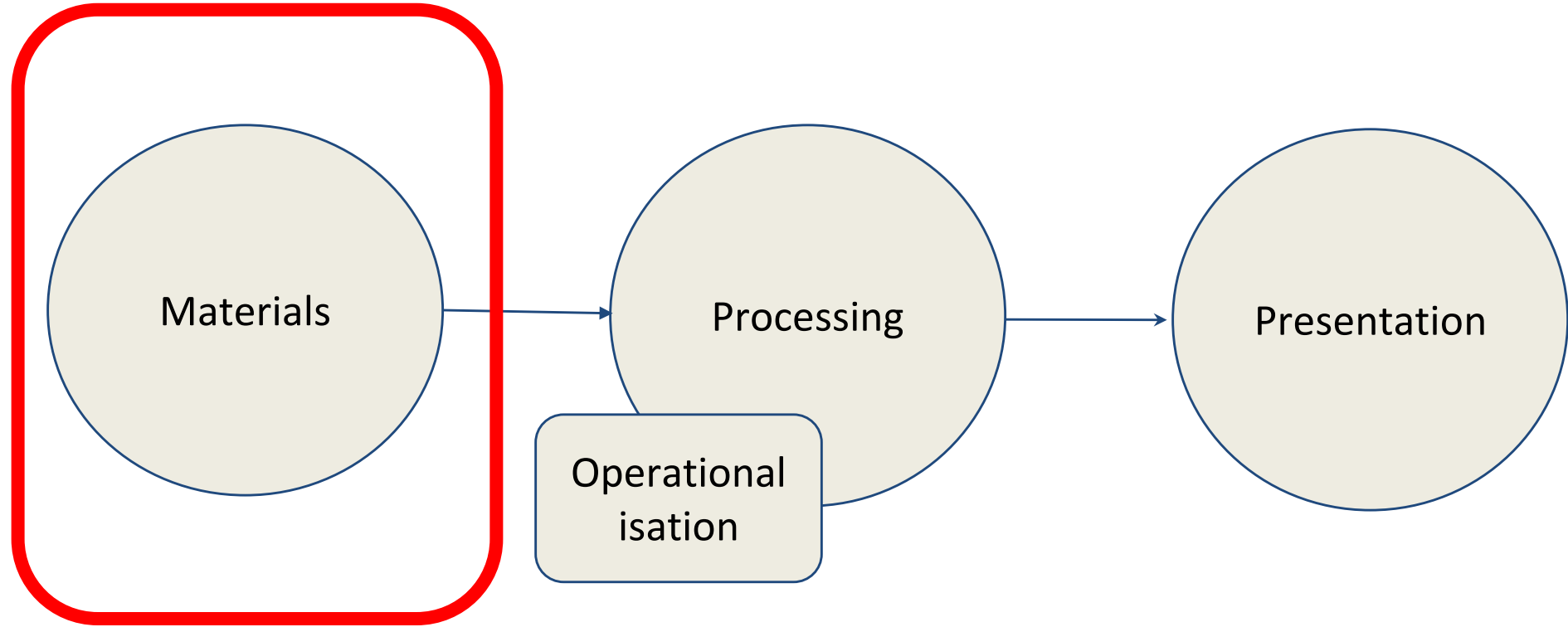
Introduction to Digital Literary Studies

Session 5 History & Theory: Corpus Building

Master-Seminar, WiSe 2025/26

Universität Potsdam

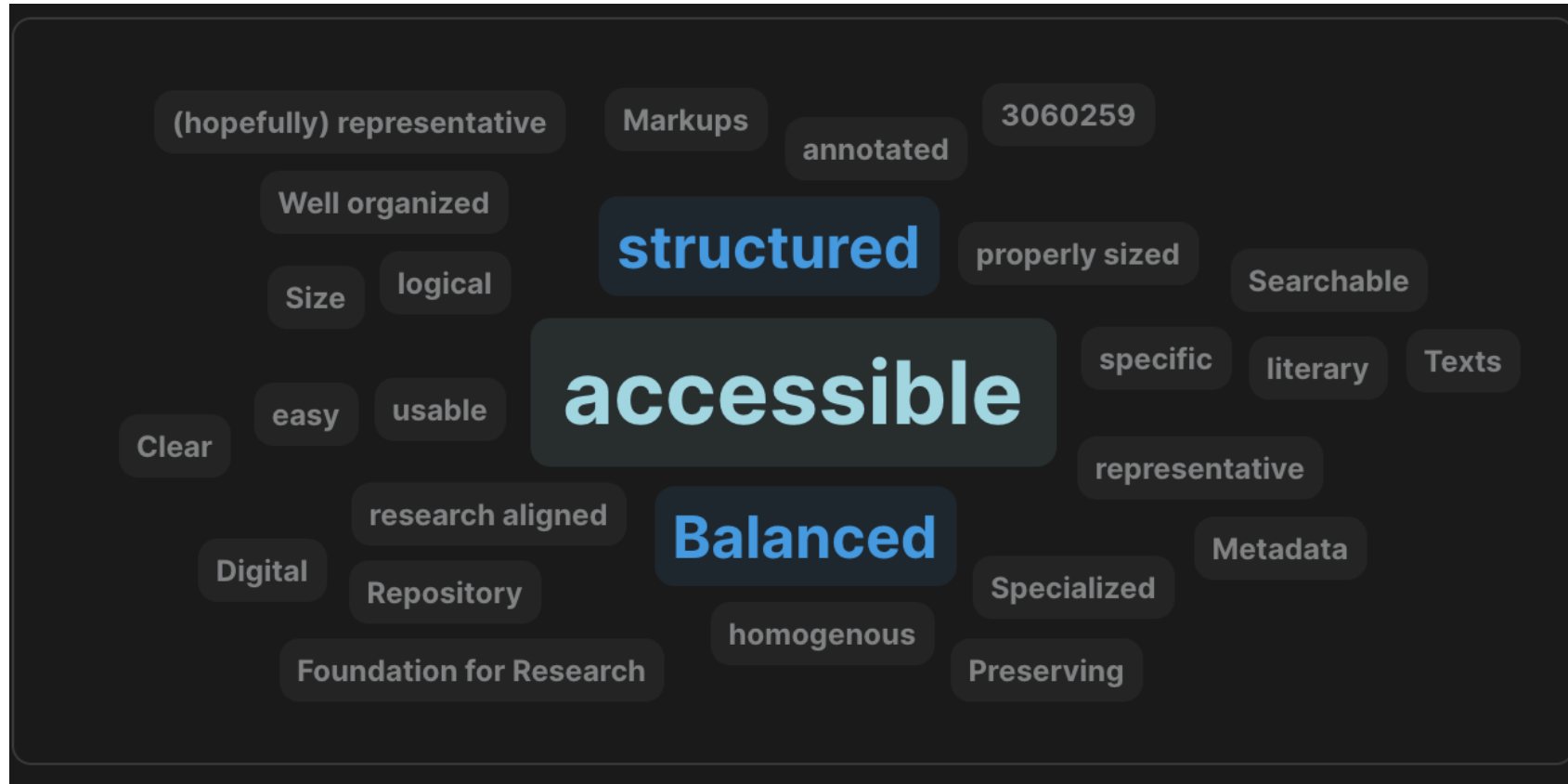
Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de



How should a (literary) corpus be?



Join at
slido.com
#3060 259



What is a (literary) corpus?

A **corpus** is a principled collection of texts assembled according to explicit criteria, typically annotated, and designed to be representative of a language, genre, or phenomenon.



Key concepts for corpus building

after [Herrman and Lauer](#) (2018: 136 ff.), [Biber 1993](#), et al.

**Golden rule: Corpus building almost always involve compromises
– transparency and documentation are essential!**



REPRESENTATIVENESS

- i.e. the extent to which a sample (= *corpus*) includes the full range of variability in a population
- Challenge: determining the total population requires bibliographies, catalogs, archives
- Published ≠ Preserved ≠ Digitized ≠ Actually read (reception history)



ACCURACY

- A corpus is "balanced" when it accurately reflects the dimensions that matter for your specific research question.
 - Maybe you want to assemble a collection of texts for general tasks (curation-driven corpus)
 - Maybe you want to collect texts for studying a specific linguistic or literary phenomenon (research-driven corpus)

Precise representation of the variables relevant for your research: genre definitions, authorship, period, region, social/cultural context

Corpus building strategies

This is our population of texts:



- **REPRESENTATIVE SAMPLE**

- Random, unbiased selection from known population
- Allows valid inferences about entire population but has some prerequisites (e.g. population must be finite and known)
- Suitable as reference corpus for comparative studies



(reflects the population share)

- **BALANCED COLLECTION**

- Targeted construction according to research-relevant criteria
- Goal: Minimum representation for all feature combinations



(each type is well represented)

- **OPPORTUNISTIC SELECTION**

- Availability-driven: digital accessibility determines corpus content
- Pragmatic approach for poorly documented domains, resource-constrained projects, etc.
- Ultimately unsatisfactory, but sometimes necessary



(that's all what we could find)

Exercise: Corpus critique

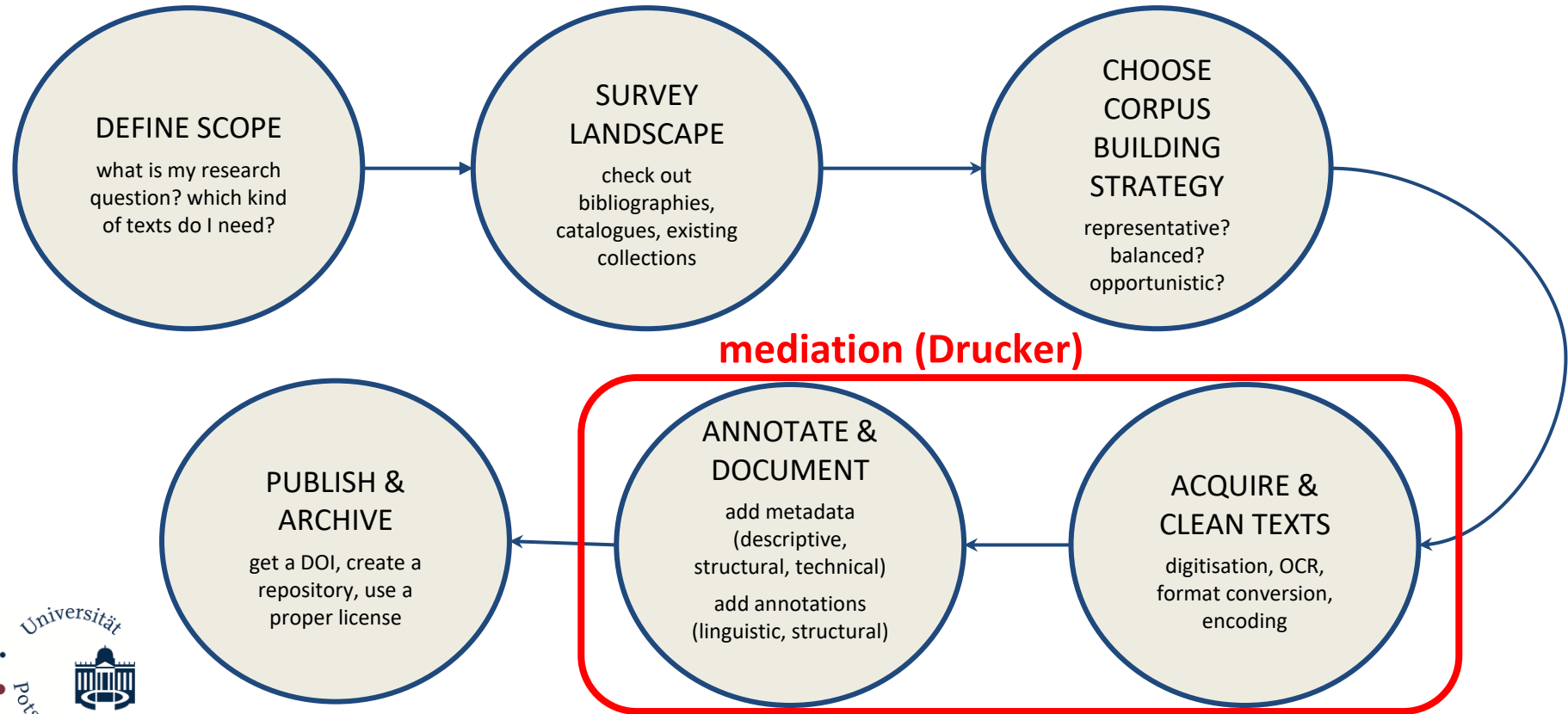
You will be assigned one of these corpora:

- **Group 1:** *DraCor English Drama Corpus* (dracor.org/eng)
- **Group 2:** *ELTeC (European Literary Text Collection)* - English subcorpus (<https://distantreading.github.io/ELTeC/>), navigate to the English corpus “eng”
- **Group 3:** *Early English Books Online (EEBO)*: <https://proquest.libguides.com/eebopqp/home>
- **Group 4:** *Common Library 1.0: A Corpus of Victorian Novels*: <https://arxiv.org/pdf/1909.02602>

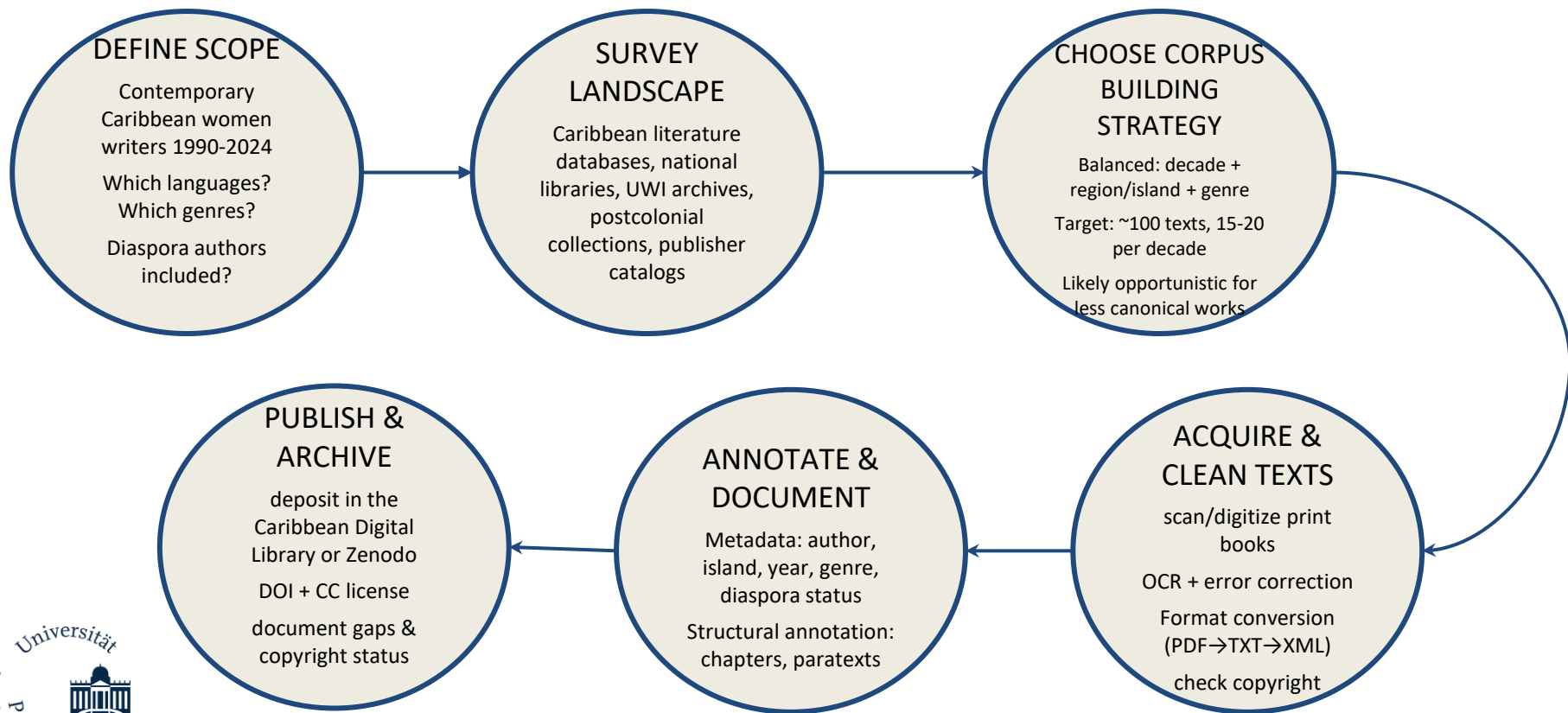
Collect all relevant information about your assigned corpus and evaluate it according to this rubric:

- Representativeness: Does it claim to represent a population? How?
- Balance: Are texts distributed across key variables? Which ones?
- Fidelity: What encoding/quality standards are used? (mediation/remediation)
- Documentation: Can you find clear selection criteria?

Corpus building workflow



An example: contemporary women writers in the Caribbean



Exercise: build your corpus!

Scenario: You need to assemble a corpus to investigate one of this topics:

- **Group 1**: Climate change in contemporary English literature
- **Group 2**: stylistic differences between male and female authors in English early modern drama.
- **Group 3**: Shakespeare's translations into German
- **Group 4**: narrative techniques in English-language web serials and online fiction

Brainstorm in your group and answer the following questions:

- Scope: What's your population?
- Selection strategy: Representative, balanced, or opportunistic? Why?
- Key variables: What 2-3 criteria would you balance?
- Size: How many texts? Why?

Introduction to Digital Literary Studies

Session 6 History & Theory: Data Modelling

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

What do you study? What are you interested in?

“If you ask [my colleagues] what it is they are studying, they would rather speak of books, paintings and movies; of drama and crime fiction, of still lives and action painting; of German expressionist movies and romantic comedy. They would mention Denis Diderot or Toni Morrison, Chardin or Jackson Pollock, Fritz Lang or Diane Keaton” ([Schöch 2013](#))

What are data?

Data in the humanities could be considered a (1) digital, (2) selectively constructed, (3) machine-actionable (4) abstraction representing some aspects of a given object of humanistic inquiry ([Schöch 2013](#))

Datasets are not just any collections of some information, they are objects structured in ways that allows them to exist within a computational medium. ([Manovich 2020](#))

“Data” are actually “capta”: not given, but produced through abstraction and interpretation ([Drucker 2021](#))

Data modelling

- Process of selecting and structuring information from sources, i.e., determining what will be identified as a feature, how it will be made explicit, and what format it will have.
- All data models embody values that carry implicit or explicit judgment and therefore often include biases.
- Every stage in data modelling raises ethical issues about privacy, data ownership and usage.

Why are humanities data special?

([Lorenz 2025](#))

Heterogeneity

High diversity among cultural objects = high diversity among data types

Textual, visual, audio, audiovisual, spatial, material – each with multiple storage formats

Multimodality

Might requires interdisciplinary collaboration and expert interpretation from multiple fields

Complexity

Multiple layers of meaning examined from different perspectives, varying across contexts/cultures/periods

Ownership

Researchers often don't generate primary data, but often work with representations owned by institutions

Researchers generate interpretive data (tags, annotations, categories)



“Eine virtuelle Reise
zu den Potsdams dieser Welt”
(<https://kultursegler.de/eine-virtuelle-reise-zu-den-potsdams-dieser-welt/>)

Unstructured data

- Ambiguous and context-dependent (“Potsdam”)
- Missing some sort of structure (*data model*) imposed upon it
- Examples: natural language, analog materials



Structured data

- Explicit, discrete, unambiguous
- Has an explicit form that has been deliberately designed, generally for a specific purpose
- Examples: encoded texts, tables, spreadsheets

Tokenisation and parametrisation

1. **Tokenisation:** Determining which units of representation can be identified in a corpus or artifact
2. **Parametrisation:** Determining what can be actually measured (according to quantity, scale, size, direction, frequency...)

Happy families are all alike; every unhappy family is unhappy in its own way. Everything was in confusion in the Oblonskys' house. The wife had discovered that the husband was carrying on an intrigue with a French girl...

Tokenisation: what is a sentence?

Parametrisation: what the prevailing sentiment within a sentence?

Checklist for creating a data model (Drucker 2021: 23-24)

1. Identify information that can be described unambiguously in your sources.
2. Create a set of labeled categories that are clear and distinct.
3. Decide which categories will be specified numerically, by true/false values, or by descriptive terms.
4. Assess the difficulty of keeping your entries consistent.
5. Standardize the format for each data type (dates, places, and names).
6. Limit the vocabulary for description except in the comments.
7. Read the categories you have established to see what the data model will take from the original sources.

Exercise: Build your own data model!

We have an issue from a 18th century Caribbean newspaper, the *Barbados Mercury and Bridge-Town Gazette* and we are interested in **individuals** and their **depiction**. To research this, we need to transform it into data.

[The issue is available as PDF in Moodle.](#)

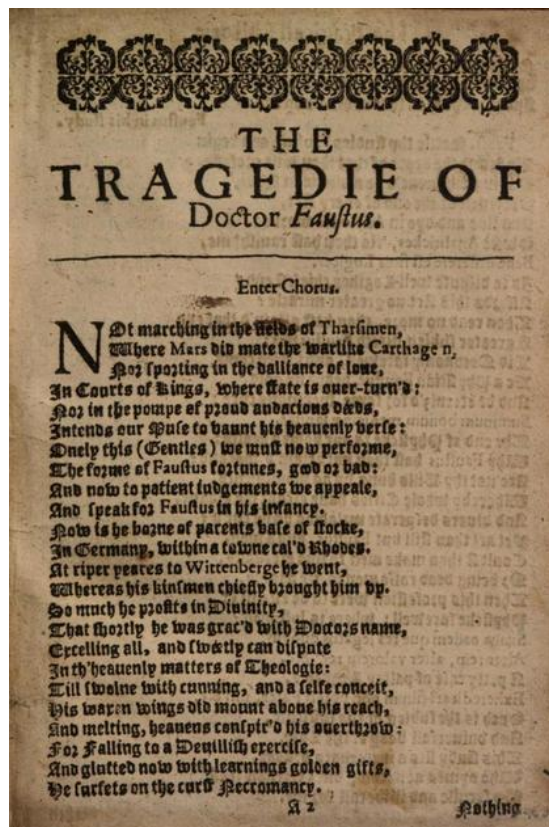
Try to address the following questions:

1. What information should be kept?
2. How should the information be structured?
3. Does the data allow for consistent structuring?
4. What vocabulary can be used to describe the structure (open vs. closed)?
5. How does the modelling reduce the original?
6. What biases could be part of the structuring process?



Group work, 20 minutes, 3-4 people

After this conceptual introduction,
let's see some concrete examples of
data modelling



```

<TEI xml:id="eng000126" xml:lang="eng">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title type="main">Dr. Faustus</title>
      <author>
        <persName>
          <forename>Christopher</forename>
          <surname>Marlowe</surname>
        </persName>
        <idno type="wikidata" xml:base="http://www.wikidata.org/entity/">Q28975</idno>
      </author>
    </titleStmt>
    <publicationStmt>
      <publisher xml:id="dracor">DraCor</publisher>
      <idno type="URL">https://dracor.org/</idno>
    </publicationStmt>
    <availability>
      <licence>
        <ab>CC0 1.0</ab>
        <ref target="https://creativecommons.org/publicdomain/zero/1.0/">Licence</ref>
      </licence>
    </availability>
    <idno type="wikidata" xml:base="http://www.wikidata.org/entity/" />
  </publicationStmt>
  <sourceDesc>
    <bibl type="digitalSource">
      <name>EarlyPrint Project</name>
      <idno type="URL">https://texts.earlyprint.org/works/A07009.xml</idno>
    </bibl>
    <availability>
      <licence target="http://creativecommons.org/licenses/by-nc/3.0/deed.en_US">
        Distributed under a Creative Commons Attribution-NonCommercial 3.0 Unported License
      </licence>
    </availability>
  </bibl>
  <biblFull n="TCP source">
    <title>

```

A text

A text *modelled* in TEI-XML

Popular structured data formats in digital humanities research

title	author	year	pages	genre
Middlemarch	George Eliot	1871	904	realist novel
Bleak House	Charles Dickens	1853	989	social novel
Jane Eyre	Charlotte Brontë	1847	532	bildungsroman

CSV/TSV

(modelling tables with
commas/tabs)

title, author, year, pages, genre
Middlemarch, George Eliot, 1871, 904, realist novel
Jane Eyre, Charlotte Brontë, 1847, 532, bildungsroman
Bleak House, Charles Dickens, 1853, 989, social novel

XML-TEI

(annotating texts with angled brackets and HTML-like syntax)

```
<corpus>
  <novel>
    <title>Middlemarch</title>
    <author>George Eliot</author>
    <year>1871</year>
    <pages>904</pages>
    <genre>realist novel</genre>
  </novel>
  <novel>
    ...
  </novel>
  <novel>
    ...
  </novel>
</corpus>
```

JSON

(storing data with curly brackets in a dictionary-like structure)

```
[
  { "title": "Middlemarch",
    "author": "George
    Eliot",
    "year": 1871,
    "pages": 904,
    "genre": "realist novel"
  },
  { next item },
  { next item }
]
```

How to create modelled data? ...annotation!

- adding structural information (e.g. paragraphs, chapters)
- marking literary concepts (e.g. fear, perspective)

Which modelling criteria could be difficult to achieve when annotating literary concepts?

I.

Als Gregor Samsa eines Morgens aus unruhigen Träumen erwachte, fand er sich in seinem Bett zu einem ungeheueren

Ungeziefer verwandelt. Er lag auf seinem panzerartig harten Rücken und sah, wenn er den Kopf ein wenig hob,

seinen gewölbten, braunen, von bogenförmigen Versteifungen geteilten Bauch, auf dessen Höhe sich die Bettdecke,

zum gänzlichen Niedergleiten bereit, kaum noch erhalten konnte. Seine vielen, im Vergleich zu seinem sonstigen

Umfang kläglich dünnen Beine flimmerten ihm hilflos vor den Augen.

»Was ist mit mir geschehen?« dachte er. Es war kein Traum, sein Zimmer, ein richtiges, nur etwas zu kleines

Menschenzimmer, lag ruhig zwischen den vier wohlbekannten Wänden, über dem Tisch, auf dem eine

auseinandergepackte Musterkollektion von Tuchwaren ausgebreitet war - Samsa war Reisender -, hing das Bild, das

er vor kurzem aus einer illustrierten Zeitschrift ausgeschnitten und in einem hübschen, vergoldeten Rahmen

untergebracht hatte. Es stellte eine Dame dar, die, mit einem Pelzhut und einer Pelzboa versehen, aufrecht

dasaß und einen schweren Pelzmuff, in dem ihr ganzer Unterarm verschwunden war, dem Beschauer entgegenhob.

Gregors Blick richtete sich dann zum Fenster, und das trübe Wetter - man hörte Regentropfen auf das

Fensterblech aufschlagen - machte ihn ganz melancholisch. »Wie wäre es, wenn ich noch ein wenig weiterschliefe

und alle Narrheiten vergaße,« dachte er, aber das war gänzlich undurchführbar, denn er war gewöhnt, auf der

rechten Seite zu schlafen, konnte sich aber in seinem gegenwärtigen Zustand nicht in diese Lage bringen. Mit

welcher Kraft er sich auch auf die rechte Seite warf, immer wieder Schaukelte er in die Rückenlage

zurück. Er versuchte es wohl hundertmal, schloß die Augen, um die zappelnden Beine nicht sehen zu müssen und ließ erst ab, als er in der Seite einen noch nie gefühlten, leichten, dumpfen Schmerz zu fühlen begann.

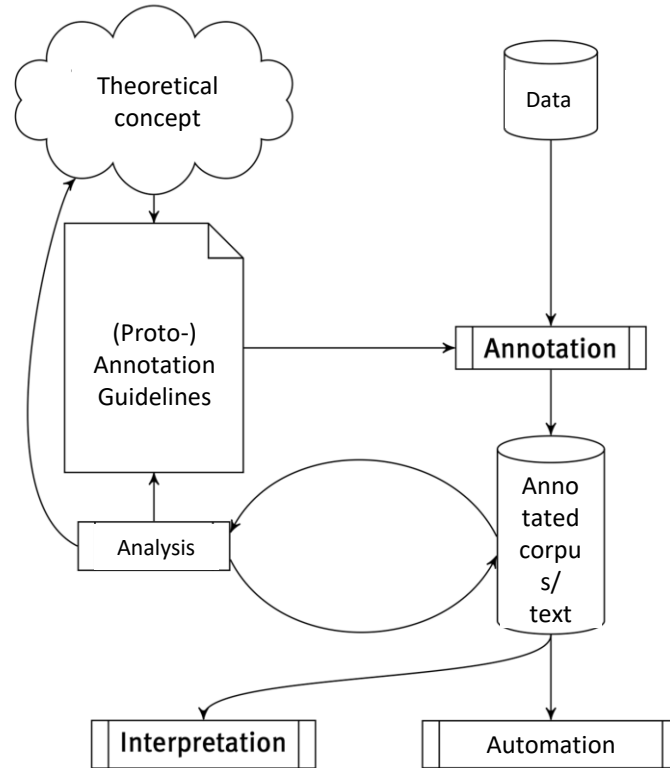
HSG:
Anno Runde stark-
schwach 2

HSG:
figuraler Artikel

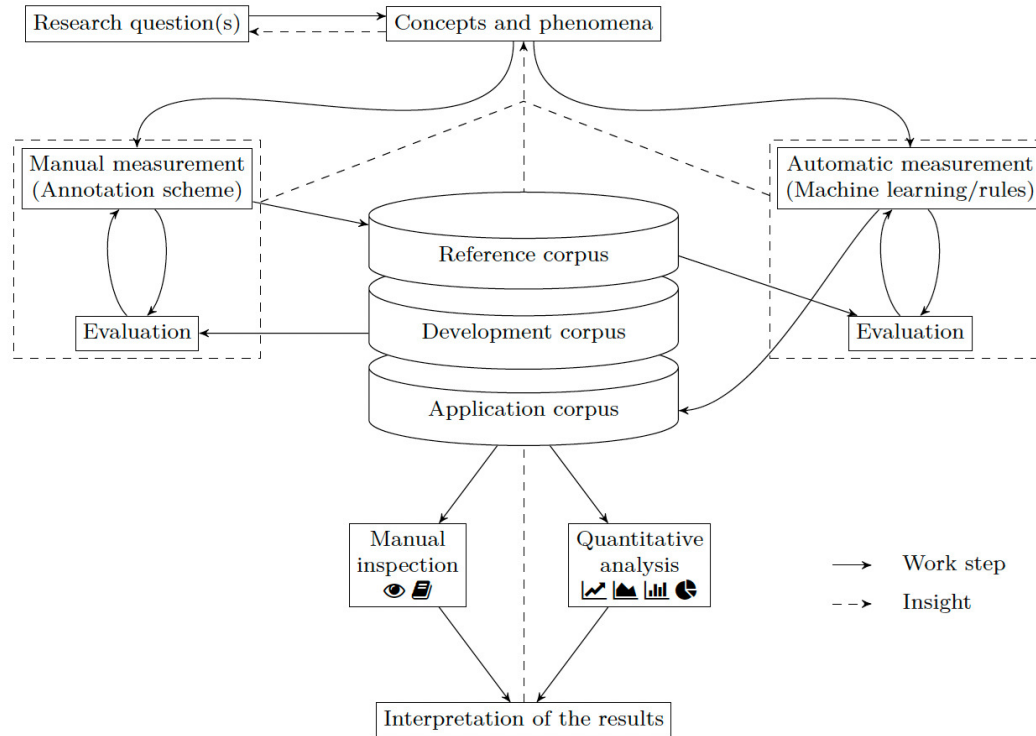
HSG:
Wertung? wg. Artikel?

HSG:
Partikel figural

Annotation Workflow



Recap: Operationalisation workflow



Introduction to Digital Literary Studies

Session 7 Methods: Network Analysis

Master-Seminar, WiSe 2025/26

Universität Potsdam

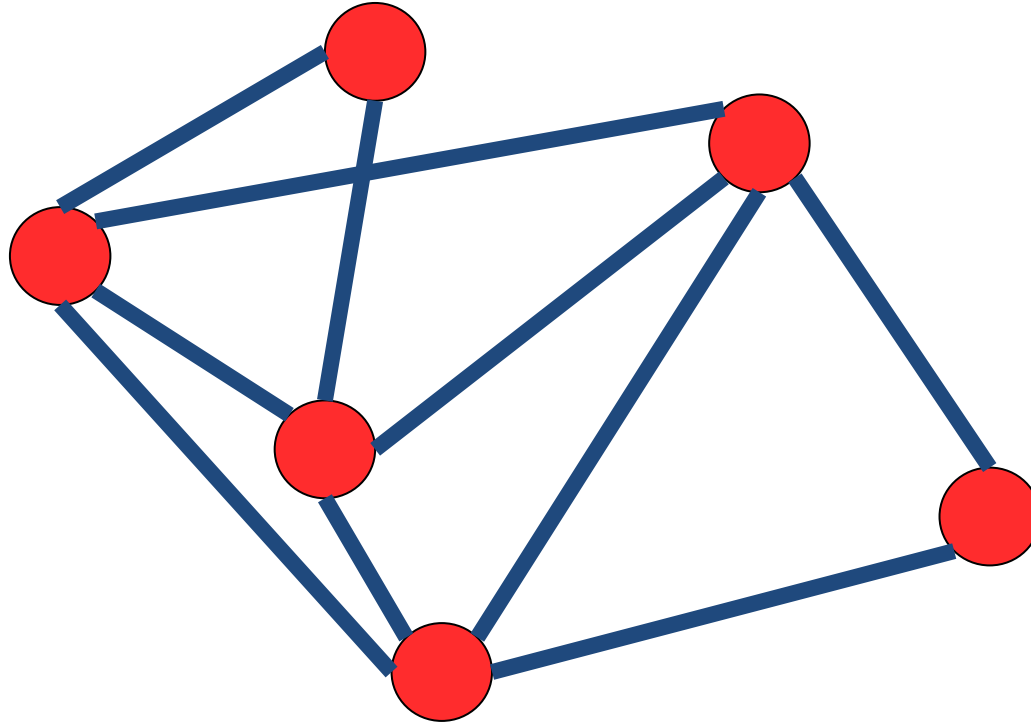
Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

The plan for today

- Learning basic terminology of network theory
- Modelling a literary text as a social network
- Reflecting critically on network visualisation

Basics

NODES



EDGES

Hands-on!



Let's build our Shrek network!

1. Open ezlinavis (<https://ezlinavis.dracor.org/>)
2. Enter character connections: NODES = characters / EDGES = they appear together in one scene

What we know (simplified, following the fabula):

1. Shrek saves Donkey from the peasants.
2. Shrek and Donkey travel to Duloc.
3. Shrek and Donkey defeat the knights and talk to Farquaad.
4. Shrek and Donkey rescue Fiona from the Dragon.
5. Dragon falls in love with Donkey.
6. Shrek, Fiona, and Donkey are ambushed by Robin Hood and the Merry Men.
7. Farquaad proposes to Fiona.
8. Shrek interrupts Fiona's wedding and Dragon devours Farquaad.
9. Shrek and Fiona marry.

Rules for Ezlinavis

SCENE 1.

Donkey

Shrek

Peasants

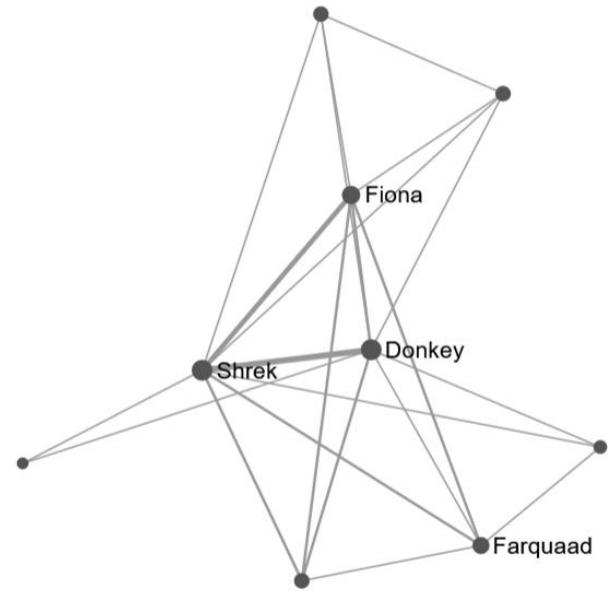
SCENE 2.

Shrek

Donkey

...

# SCENE 1	# SCENE 4	# SCENE 7
Shrek	Shrek	Farquaad
Donkey	Donkey	Fiona
Peasants	Fiona	# SCENE 8
# SCENE 2	Dragon	Shrek
Shrek	# SCENE 5	Fiona
Donkey	Donkey	Farquaad
# SCENE 3	Dragon	Dragon
Shrek	# SCENE 6	# SCENE 9
Donkey	Shrek	Shrek
Farquaad	Fiona	Fiona
Knights	Donkey	
	Robin-Hood	
	Merry Men	

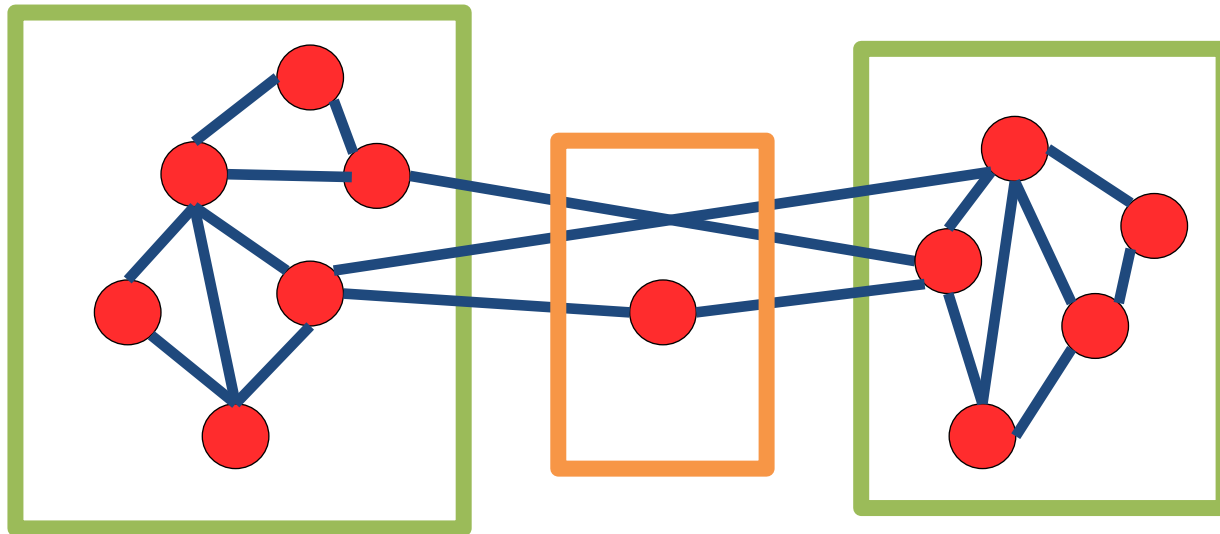


- Which type of network?
 - co-occurrence
- Which segmentation mode?
 - scene

Key concepts in network analysis

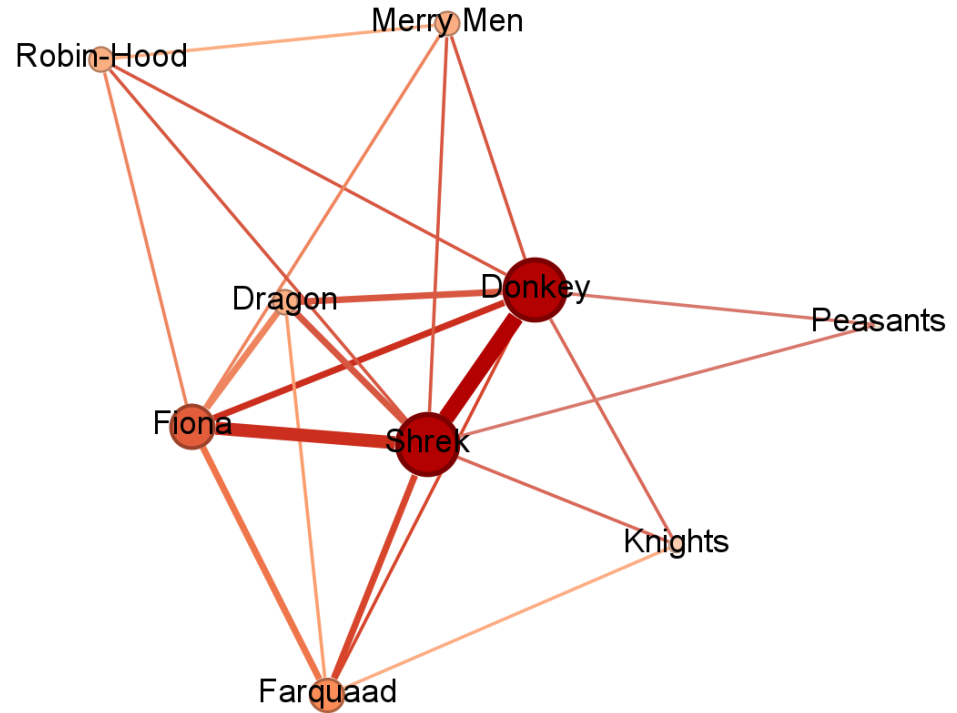
Centrality (role of the individual nodes)

Clustering (overall network structure)



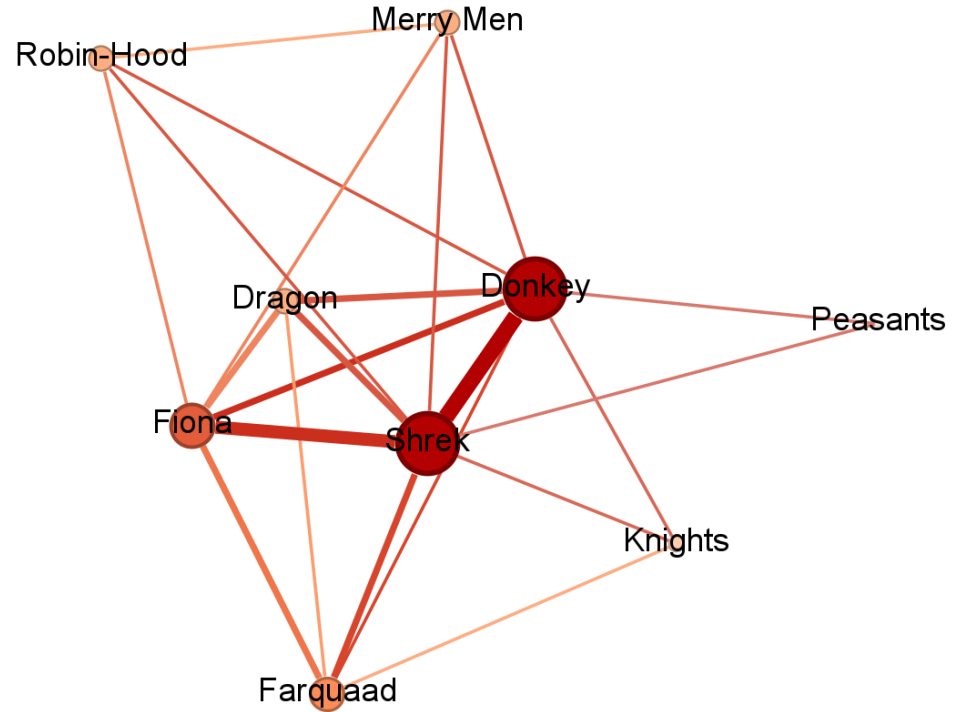
Degree

- Number of connections a node has
- Helps identify central or important characters
- Highest degree: Donkey and Shrek



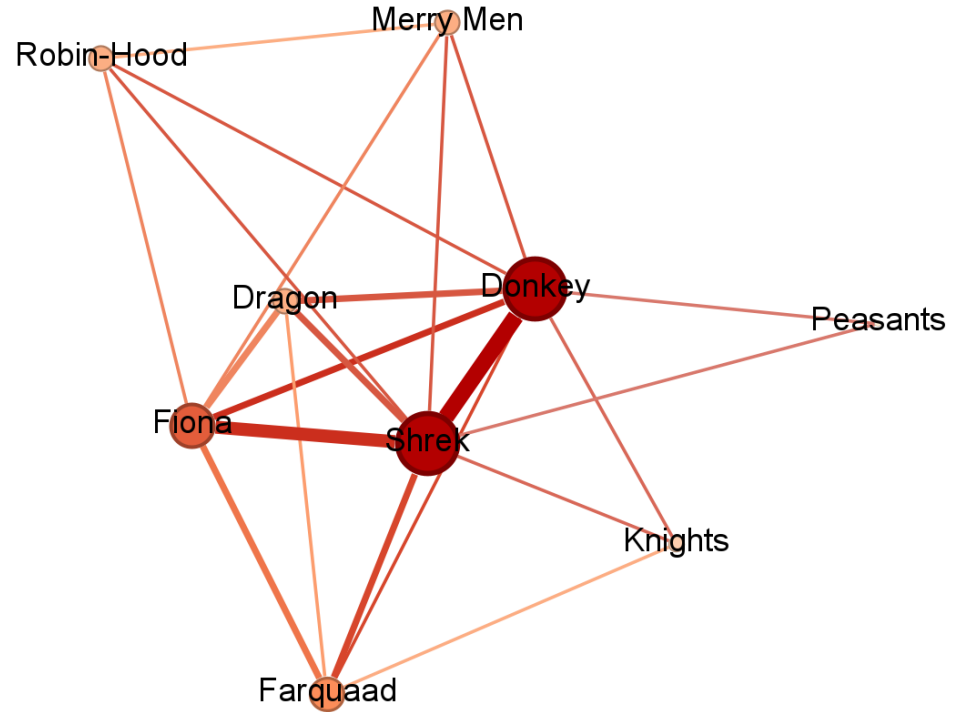
Path length

- Number of steps between any two nodes in the network
- Indicates how "spread out" or compact a network is



Betweenness

- Measures how often a character lies on the shortest path between other pairs of characters
- Donkey and Shrek are the best bridges across characters



Clustering

Walpurgis
Night's Dream

Prologue in
the Theatre

People from Faust's city
(Before the Gate)

Faust

Mephistopheles

Walpurgis Night

Prologue
in Heaven

Auerbach's Cellar/
Witch Kitchen

Goethe, *Faust I*
([DraCor data](#))

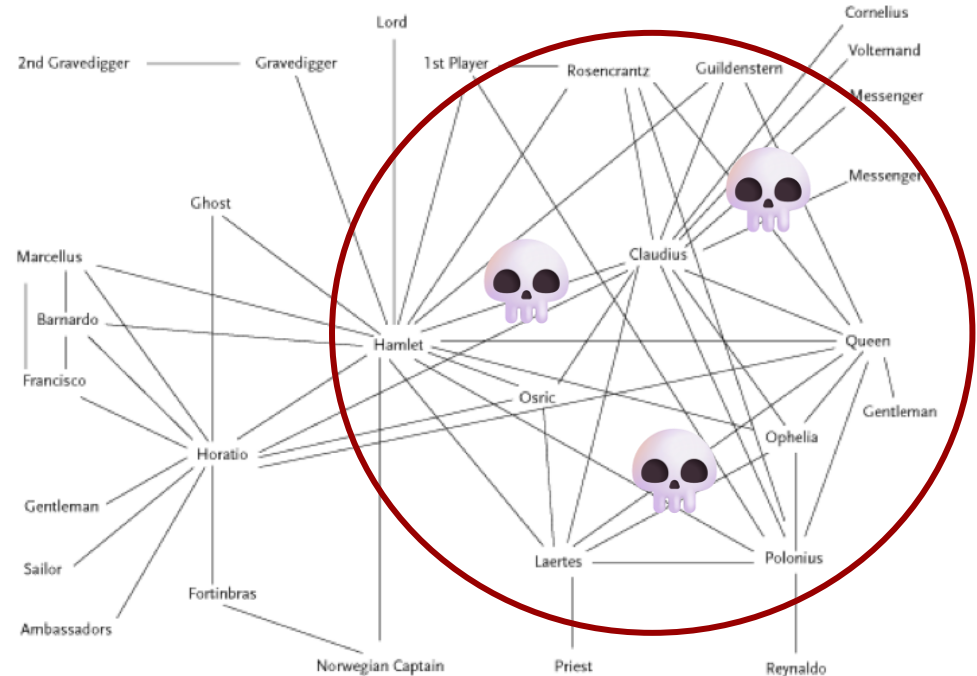
How could you apply network analysis to literary studies?

Which information do you need to be able to create literary networks?

Example 1: Network analysis for plot study

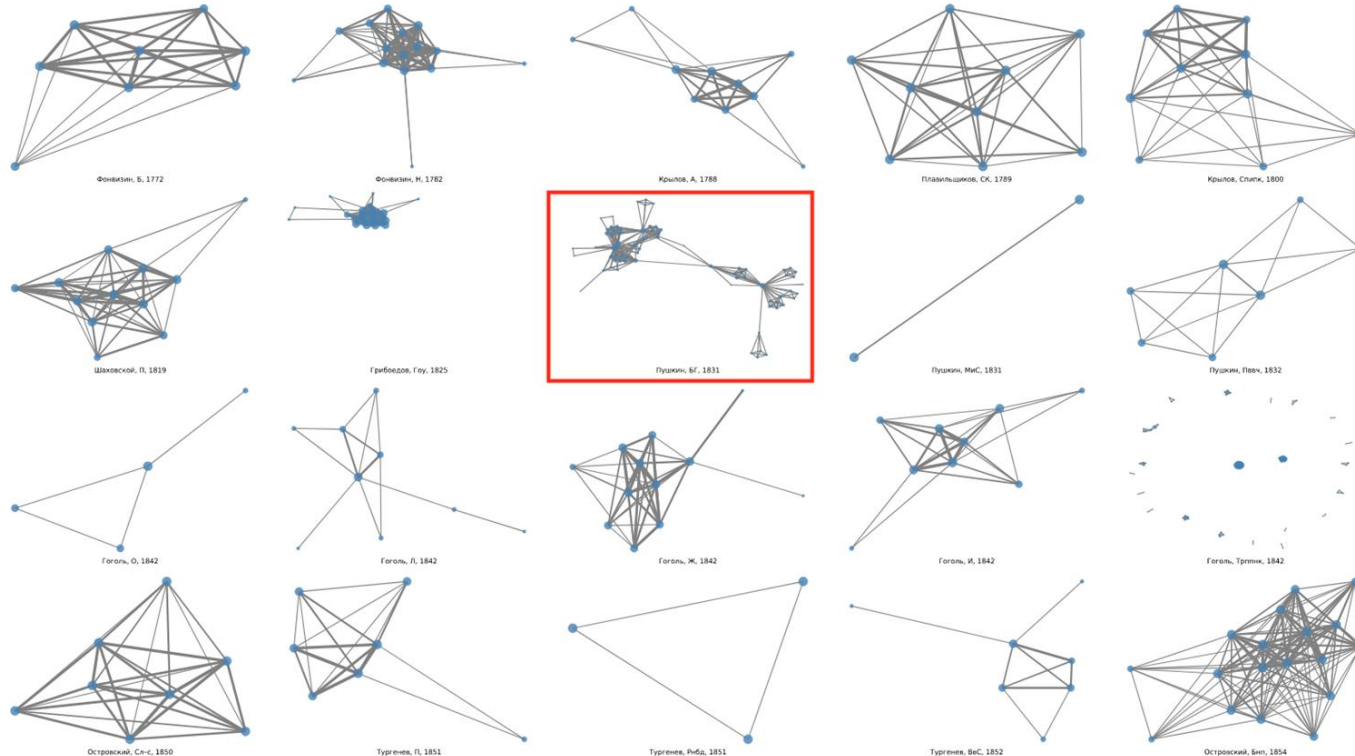
"**You reduce the text to characters and interactions**, abstract them from everything else, and this process of reduction and abstraction makes the model obviously much less than the original object [...] but also, in another sense, much more than it, because **a model allows you to see the underlying structures of a complex object**. It's like an X-ray: suddenly, you see the region of death* [...] which is otherwise hidden by the very richness of the play".

* "Outside of that bold region [= *character linked to Claudius and Hamlet*], [almost] no one dies in [the play]. The tragedy is all there".



[Franco Moretti, "Network Theory, Plot Analysis" \(2011\)](#)

Example 2: Exploring literary history via networks



What about visualisation itself?

- Go to the DraCor website (dracor.org).
- Choose one corpus.
- Choose one play.
- Go to the tab “Downloads”.
- Download the GEXF network file.





grants you a **network map** in one click.

- Go to *Gephisto*
(<https://jacomyma.github.io/gephisto/>)
and upload the
GEXF file there.
- Feel free to retry
more than once!

individual work, 10 minutes



I grant you **one great network visualization** in one click. Not knowing what you want, I will make *arbitrary* calls when I need to. Don't worry, my calls are always great, as far as anyone knows...

Gephisto

(else you can always use [Gephi](#))

OK (*sign in blood*)

Discussion

- Which algorithms was used?
- What happened to nodes/edges?
- How are node properties rendered (e.g. size)?
- How is the background colour?

☞ Epistemic value of network representation:
Does it help me understand something about the play itself or is it just some fancy graph?

Introduction to Digital Literary Studies

Session 8 Methods: Cartographic Analysis

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

Brainstorming: What can we do with maps in the humanities ?

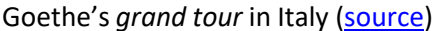
Visited?

- Yes
- No
- Unclear
- Start point (with day)
- STC's Tour

0 4 8 16 Kilometers

The map shows the Lake District region with a network of red lines representing the STC's Tour. Red dots indicate visited locations, while white circles indicate unvisited ones. The tour starts at a point marked with a red dot and a day number. A scale bar and a north arrow are also present.

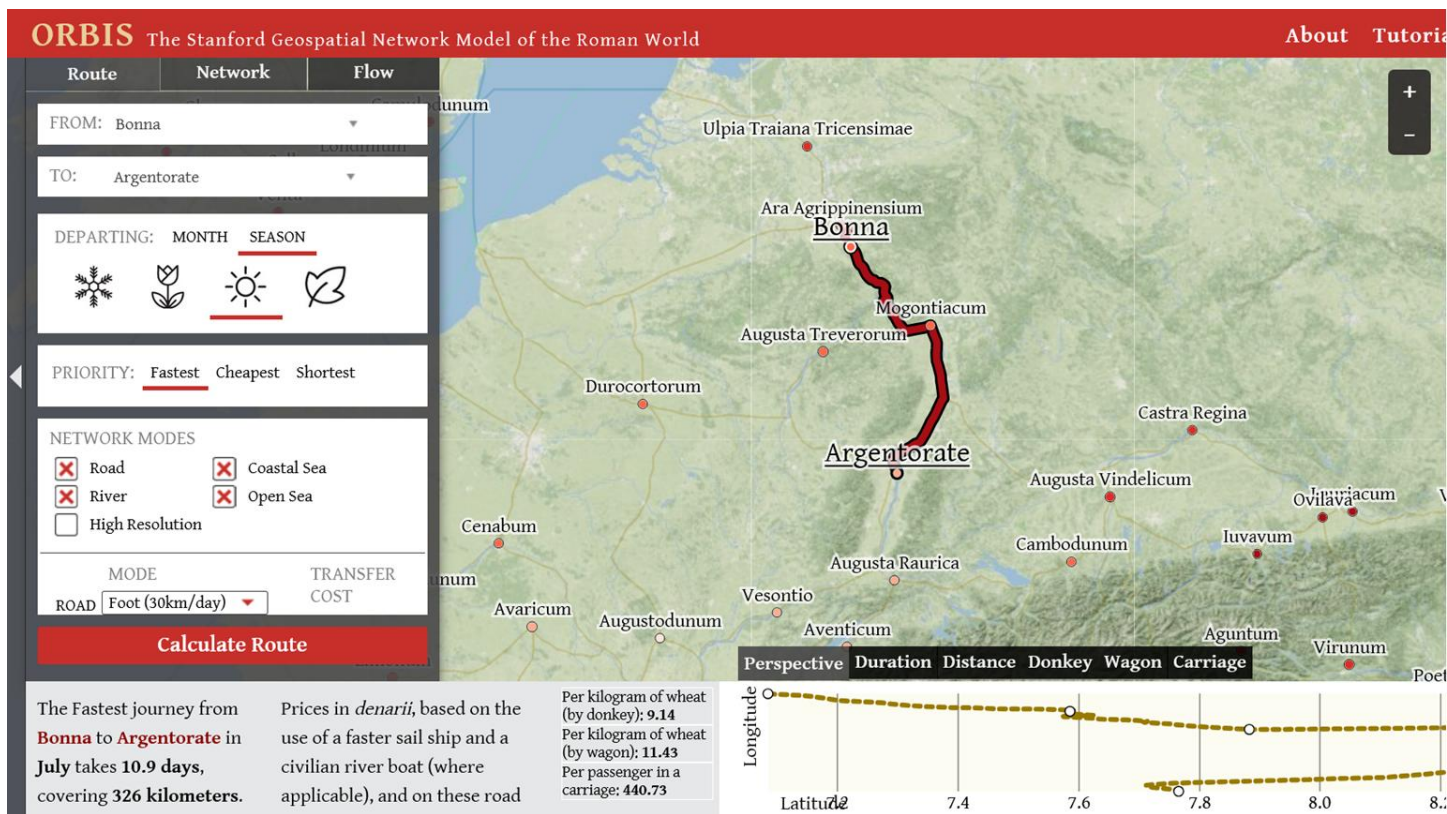
1. Sun 1st Aug Greta Hall, Keswick to Long Moor, Ennerdale
2. Mon 2nd Aug Long Moor, Ennerdale to St Bees
3. Tues 3rd Aug St Bees to Egremont
4. Wed 4th Aug Egremont to Burnthwaite, Wasdale Head
5. Thurs 5th Aug Burnthwaite, Wasdale Head to Taw House Farm, Eskdale
6. Fri 6th Aug Taw House Farm, Eskdale to Ulpha
7. Sat 7th Aug Ulpha to Black Bull Hotel, Coniston
8. Sun 8th Aug Black Bull Hotel, Coniston to Brathay Hall, Clappersgate
9. Mon 9th Aug Brathay Hall, Clappersgate to Greta Hall, Keswick



Analyze an aspect of spatial experience

ORBIS: The Stanford Geospatial Network Model of the Roman World

orbis.stanford.edu

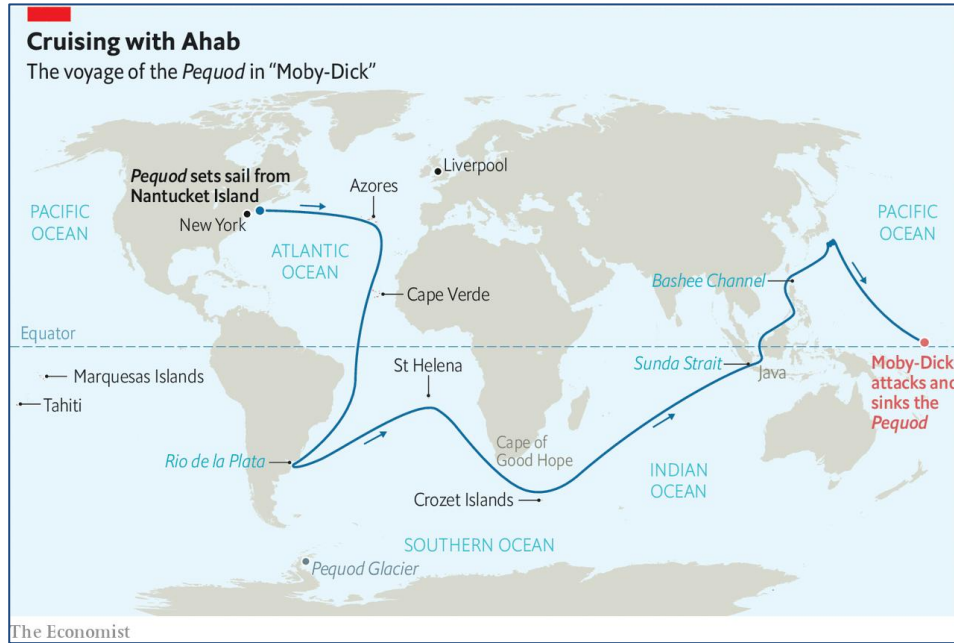


Interpret a map as a historical and critical form



Olaus Magnus, *Carta Marina et descriptio septentrionalium terrarum* (1539), detail ([info](#))

Create a map from place-based references



[an item](#) from the [David Rumsey Map Collection](#) – check it out!

Finding ways to express knowledge embodied in a text allows literary and literal aspects of geography to complement each other. (Drucker 2021: 133)

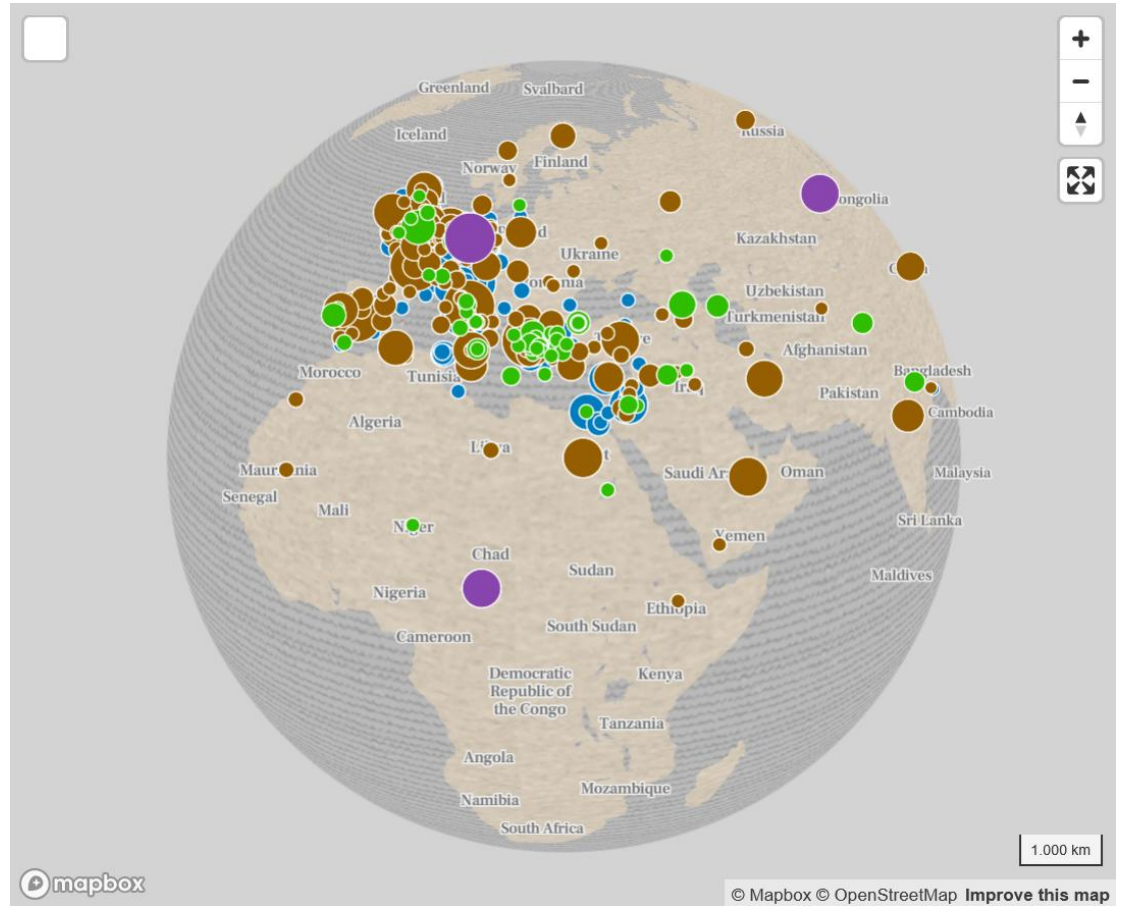
Epistemic significance of mapping

A map can be made in minutes in Google Maps without prior experience, but **the challenge is to understand why a map is useful.**

- What kind of map should it be?
- What aspects of the research are geographical, spatial, or related to the maps themselves?
- Is information in the map essential to the research, or just a presentation format, or is the map itself an object of study?
- Are you trying to answer a research question based on what the map shows or geospatial data you can generate from its features?

(Drucker 2021: 131)

Let's take a
closer look at
a recent
mapping
project



A Little Atlas of Early Modern English Drama (atlas-emed.github.io)

Explore the project website and try to answer the following questions:

- Which kind of map it is presented?
- Which research question is tackled?
- What aspects of the research are geographical/spatial?
- Is the information in the map essential to the research, or just a presentation format, or is the map itself an object of study?

HAVE YOU SEEN THIS MAN?

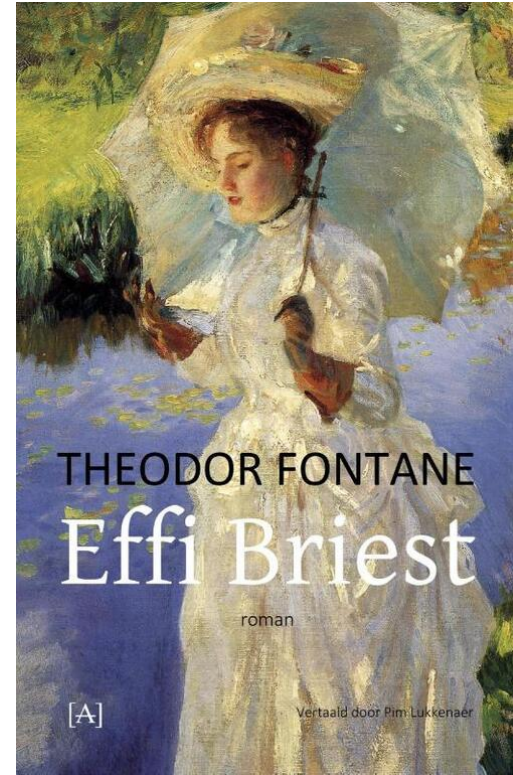


"THEODOR"

IF YOU HAVE SEEN THIS MAN PLEASE CONTACT:

LANDESPOLIZEI BRANDENBURG
POTSDAM

Giovannini, Goswami Choudhury, Sluyter-Gäthje | Introduction to Digital Literary Studies



Correspondence Database

Recommended citation style: *Theodor Fontane Correspondence Database*. In consideration of *Die Briefe Theodor Fontane: Verzeichnis und Register* (Munich: Hanser 1988) edited by the Theodor Fontane Archive. Potsdam 2021ff. URL: <https://www.fontanearchiv.de/fontane-briefdatenbank>

The *Correspondence Database* is in a beta stage and is currently still being optimized. It will also be updated on an ongoing basis. We gratefully accept [additions and corrections](#).

SEARCH WORD

Search

Detail search +

Your search returned **6215 results**

Letter metadata

Letter: Theodor Fontane to Emilie (geb. Labry) Fontane | Neuruppin, 20.06.[1833]

Type: Letter

HBV: 33/1

Sender (person, institution): [Theodor Fontane](#)

Recipient (person, institution): [Emilie \(geb. Labry\) Fontane](#)

Date: 20.06.[1833]

Place of origin: Neuruppin

Place of receipt: Neuruppin

Mentioned persons / corporate bodies: [Fontane, August](#) – [Bientz, Johann Leberecht \(Fam.\)](#) – [Gutike \(Syndikus\)](#) – [Jumptow \(Hauderer\)](#) – [Kolbe, Herr](#) – [Sommerfeldt, Jenny \(geb. Fontane\)](#)

Owner: Staatsbibliothek zu Berlin Preußischer Kulturbesitz

Print: [Prop I, 13 \(Nr.2\)](#) – [HAB I, 7 \(Nr. 1\)](#) – [Dieterle, "Liebe Mutter" in: Fontanes Briefe im Kontext \(2019\), 191-211](#)

Download `literary_mapping.ipynb` from Moodle to your device!

individual/group work, 30 minutes

Working in the JupyterHub

1. Navigate to <https://test.jupyterhub.uni-potsdam.de/>
2. Log in with your UP credentials
3. Once logged in, choose the first option “minimal notebook”

Server Options

Minimal Notebook

Command-line tools useful when working in Jupyter applications: Common useful utilities like curl, git, nano (actually nano-tiny), tzdata, unzip, and vi (actually vim-tiny), TeX Live for notebook document conversion.

R Notebook

Popular packages from the R ecosystem: The R interpreter and base environment, IRKernel to support R code in Jupyter notebooks, tidyverse packages from conda-forge, caret, crayon, devtools, forecast, hexbin, htmltools, htmlwidgets, nycflights13, randomforest, rcurl, rmarkdown, rodbc, rsqLite, shiny, tidymodels, unixodbc packages from conda-forge.

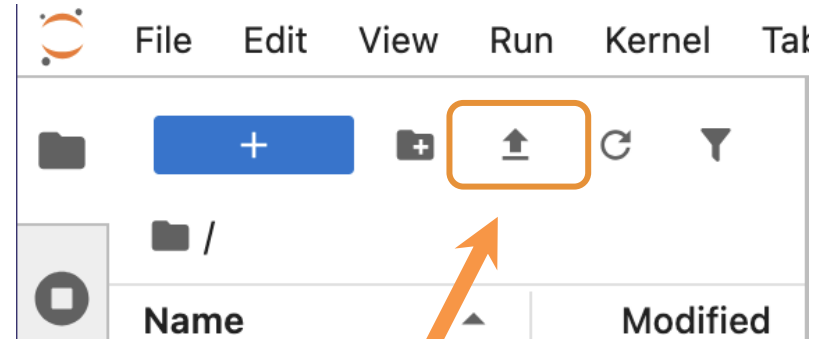
Julia Notebook

Popular packages from the Julia ecosystem: The Julia compiler and base environment, IJulia to support Julia code in Jupyter notebook, Pluto.jl reactive Julia notebook interface, made accessible with jupyter-pluto-proxy, HDF5 package.

SciPy Notebook

Working in the JupyterHub

4. Once you see the “launcher”, you can upload the notebook from your computer (orange box)



Introduction to Digital Literary Studies

Session 9 Methods: Semantic Field Analysis

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

Preliminary: Preparation for the analysis

Corpora for the analyses (more info here: <https://zenodo.org/records/7997242>):

- Canon Corpus of American English Fiction
- Canon Corpus of British Fiction

Two formats:

- Text
- Annotations (Python-specific)

Two version:

- Full corpus: British 86 & US 168 texts
- Minimal corpus: British 14 & US 19 texts

Decide for **one corpus** and download (at least) the **annotations** and the **metadata** file from moodle.

Upload the corpus and the metadata file to the JupyterHub or Colab (open notebook first).

Semantic Fields...

what's that?

Semantic Field

Two scenarios:

1. Starting off from the interest in the depiction of nature
2. Starting off from the observation that a text builds on the dichotomy of urban and rural areas

→ How do they differ?

Semantic Field Approaches

Approaches

- top-down/deductive (topic)
- bottom-up/inductive (from text)

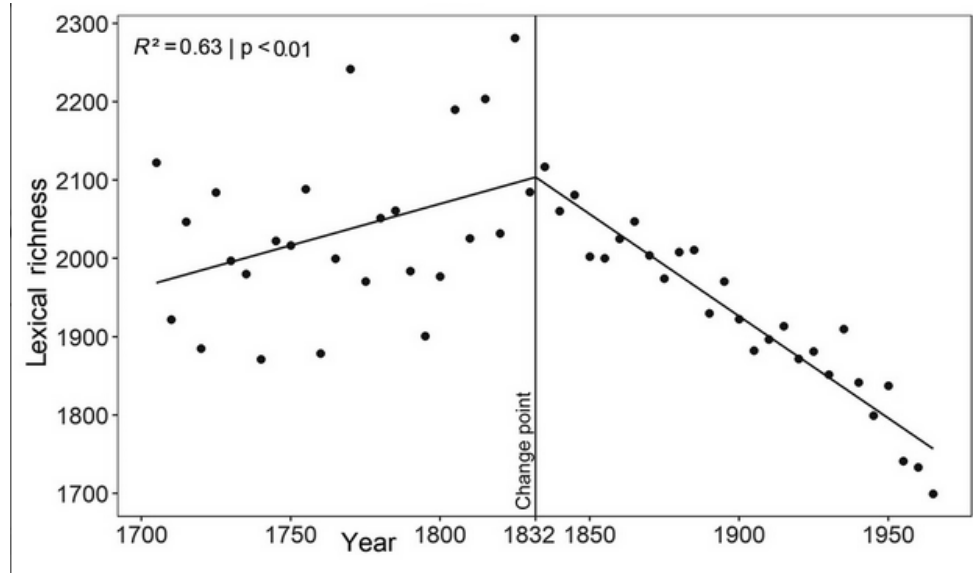
Semantic Fields: A Top-Down Example

The rise and fall of biodiversity in literature
([Langer et al. 2021](#)):

- 240,000 English biological taxon labels
- Calculate Biodiversity indices for 16,000 books



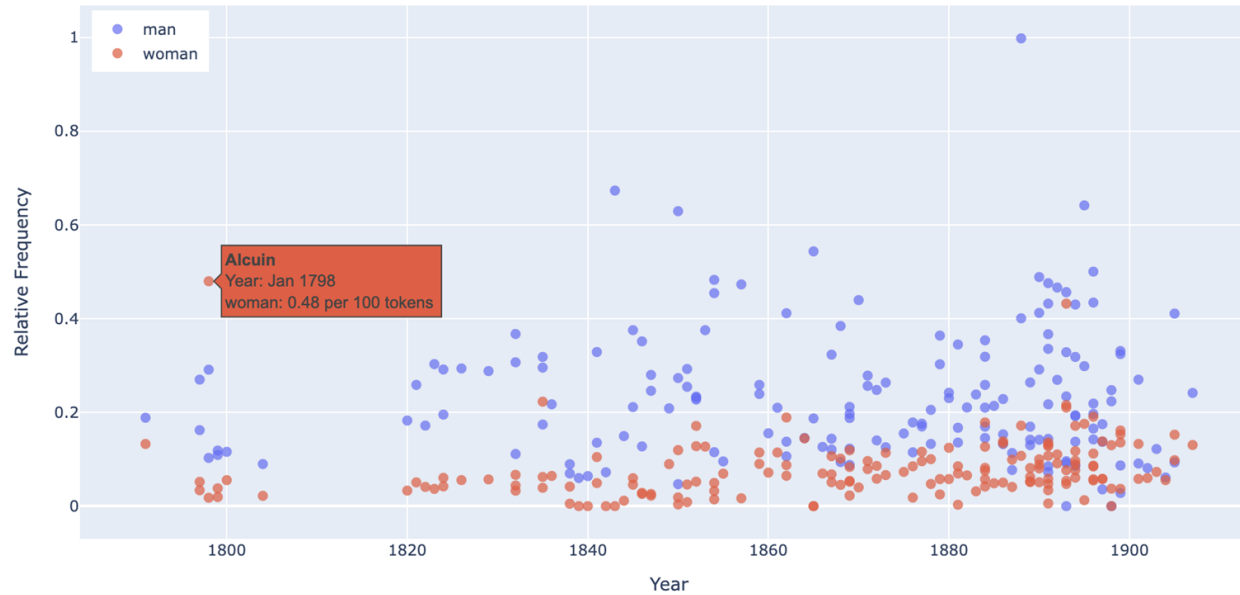
Decline as an
effect of
industrialisation?



[illegible]

Semantic Fields: Bottom-up & Top-Down

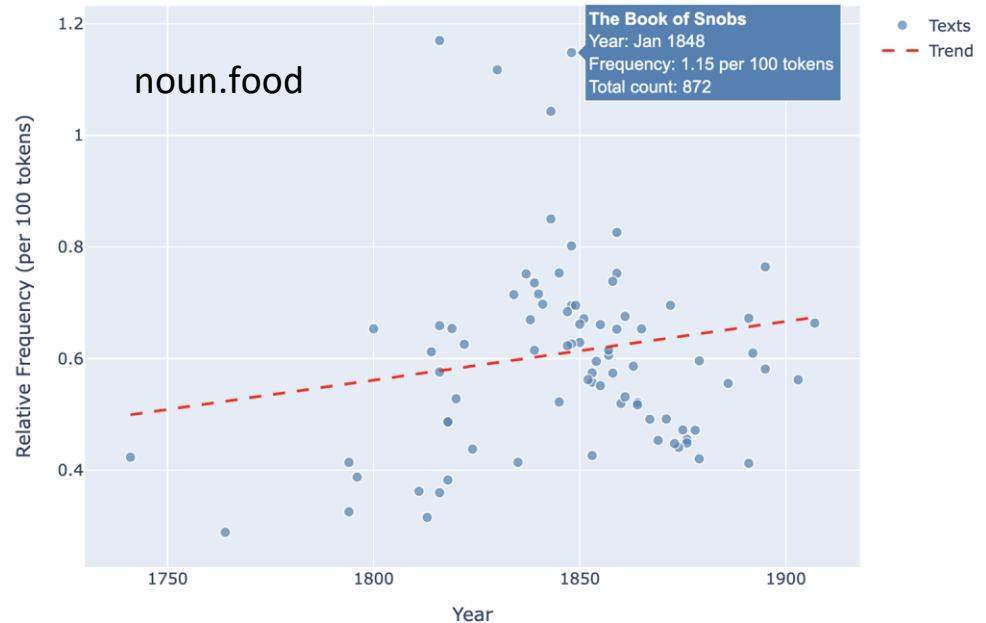
Relative Frequency of Selected Noun Lemmata Over Time



Semantic Fields: Top-Down

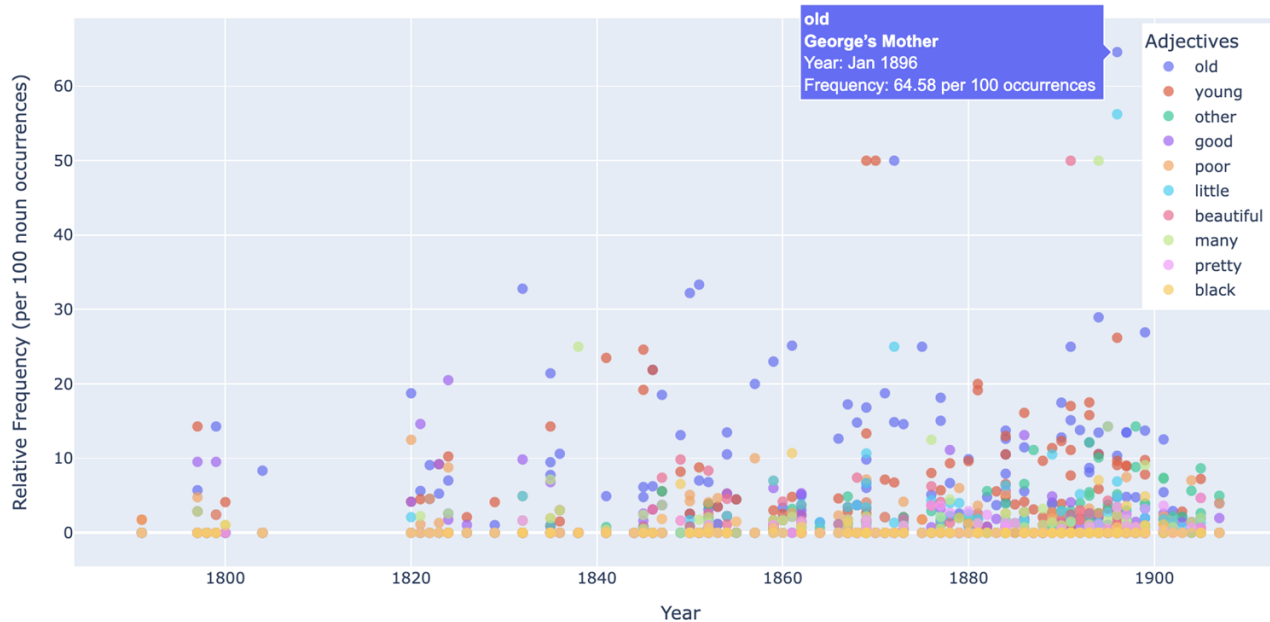
- Create a word list by yourself
- Create a word list from a lexical semantic resource (like wordnet or wikidata)
- **Wordnet**: groups nouns, adjectives, verbs into so-called synsets

<https://wordnet.princeton.edu/>



Attributes of the semantic field

Adjective Modifiers of "woman" Over Time



Comparison of two semantic fields



Analysis of Semantic Fields with Python



- Download the *Notebook for the Analysis of Semantic Fields* from moodle
 - Upload it to the JupyterHub or open Colab Notebook
 - Get an overview over the notebook by reading the first cell carefully
 - Execute Part 0: Install and load libraries (press Play )
 - Choose the corpus name in Part 1
 - Load the corpus in Part 2a (don't execute 2b) and wait  (takes about one to three minutes)
 - Load the metadata in Part 3
 - Inspect the corpus and the metadata in Part 4
- Decide which analysis you are interested in and jump there

Information on Group Project and Poster Presentation

Procedure:

1. Form groups of 2–4 students.
2. Develop a research question that fits thematically into the seminar, employing either one of the four proposed methods (method-based project) or focusing on the creation of a research corpus (corpus-based project). Look at the checklists for the projects.
3. Enter your question and the group members in this moodle forum until the **7th of January**: <https://moodle2.uni-potsdam.de/mod/groupselect/view.php?id=2638400>
4. Investigate the research question by conducting experiments and documenting the results or by collecting corpus material and documenting the criteria and the collection process.
5. Create a scientific poster and upload it in Moodle until the **23rd of January** (only one group member should upload the poster for the entire group): <https://moodle2.uni-potsdam.de/mod/publication/view.php?id=2638394>
We will print out your posters.
6. Present the poster on the **28th of January**.

Alles einklappen

Algemeines

Welcome to the Moodle course for Introduction to Digital Literary Studies
This course will take place every **Wednesday 12:00-14:00** in 1.09.1.15.

Ankündigungen

Information on the course (last updated 12/03/2025)
Please read this information carefully before the first class. There will be time set aside for questions in the first session.

General Information for Anglophone Modernities Students
This is a general information for students from Anglophone Modernities: Please note that you can take two classes with Off-University that we (Institut für Anglistik und Amerikanistik) can then transfer the credits for. All the information is also available in this document.

Reflection Paper Submission
Hey everyone, as you know this course requires you to write reflection papers to pass the course. Please find the mandatory requirements for the reflection in this document below. It has all the information about deadlines, submission prerequisites etc.

Group Project Information

Information on the group projects and poster presentations

PowerPoint Poster Template from the UP

Project Groups

Helpdesk for the poster projects
In this forum you can post questions concerning the seminar projects. These could for example be of:

- technical nature (like "I am not able to run the notebook for the semantic field analysis, the error I get is...")
- conceptual nature (like "The concept of centrality in network analysis is unclear to me, because...")
- organizational nature (like "When do we need to upload our posters?")

Poster Upload

Introduction to Digital Literary Studies

Session 10 Methods: Sentiment Analysis

Master-Seminar, WiSe 2025/26

Universität Potsdam

Luca Giovannini, Priyam Goswami Choudhury, Henny Sluyter-Gäthje
luca.giovannini@uni-potsdam.de, goswamichoudhury@uni-potsdam.de, henny.sluyter-gaethje@uni-potsdam.de

Sentiment Analysis

What do you remember from our sessions before about Sentiment Analysis?

Sentiment Analysis

I love the winter. The snow outside is terribly beautiful but I don't like to ski.

Sentiment Analysis

I love the winter. The snow outside is terribly beautiful but I don't like to ski.

love: 0.5

terribly: -1

beautiful: 0.85

don't like: 0.0

Sentiment Analysis

I love the winter. The snow outside is terribly beautiful but I don't like to ski.

love: 0.5	}	0.45
terribly beautiful: 0.85		
don't like: 0.0		

Emotion Analysis

From psychology:

- Plutchik's wheel of emotions
- [Ekman \(1990\)](#) Basic Emotions

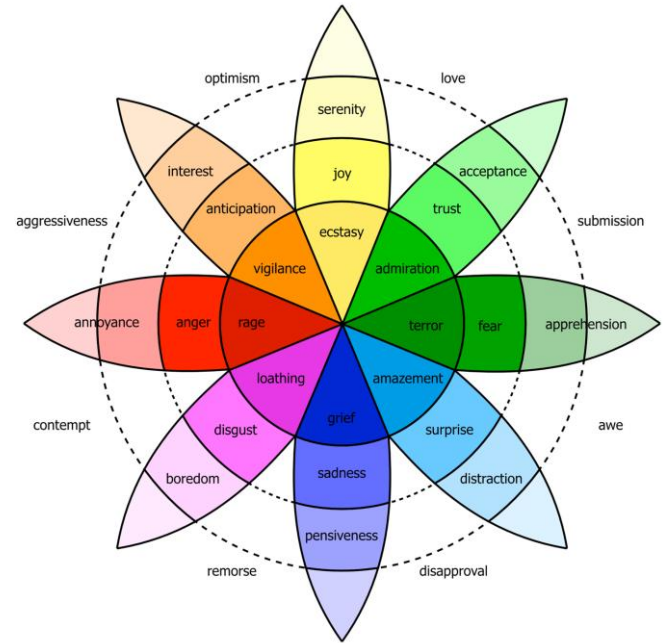
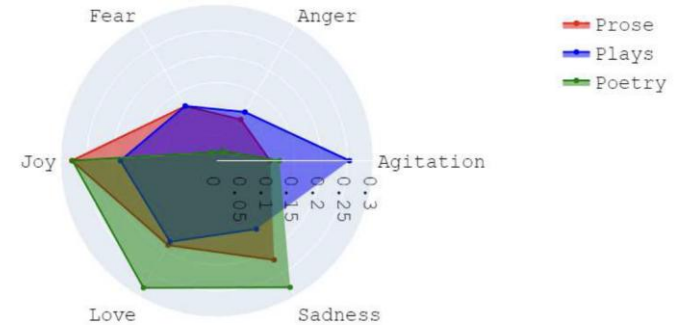
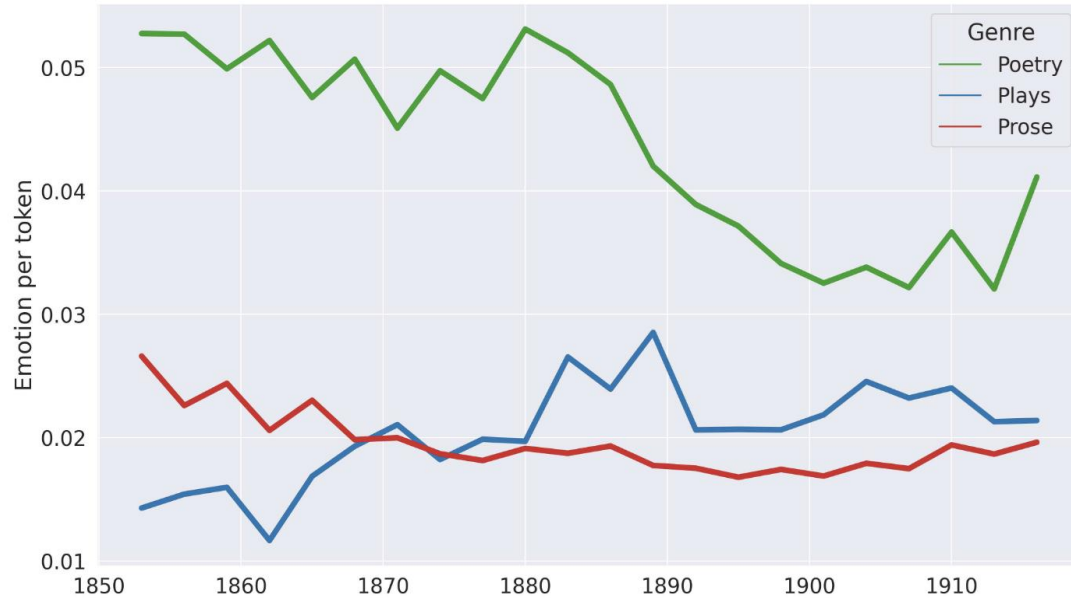


Image source:

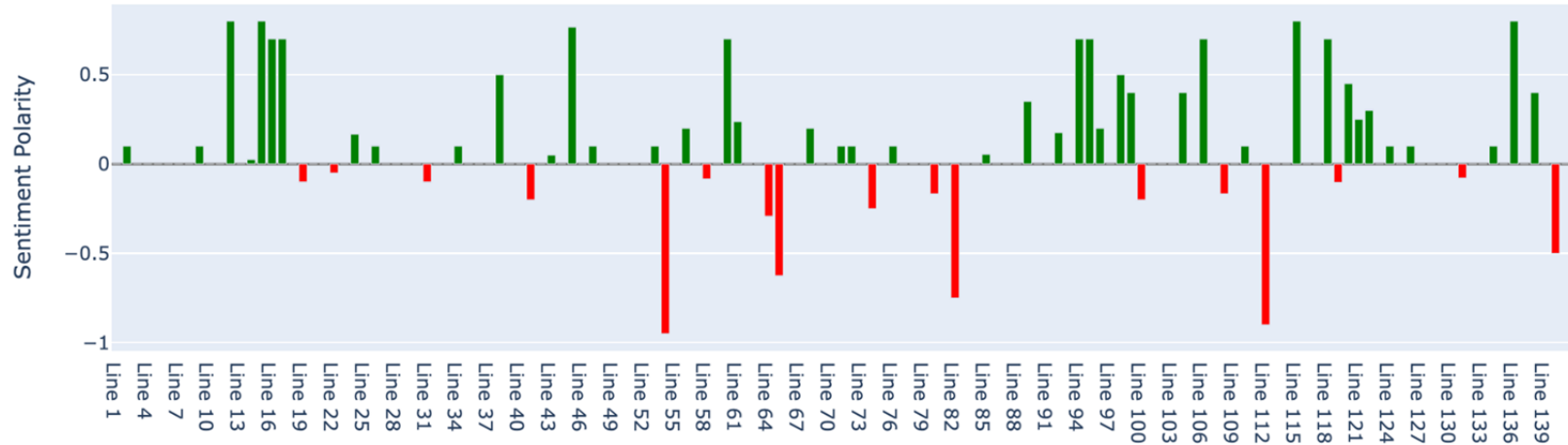
https://en.wikipedia.org/wiki/Robert_Plutchik#/media/File:Plutchik-wheel.svg

Emotion Analysis Example: “On the Unity of Literary Change”



[Konle et al. \(2024\)](#)

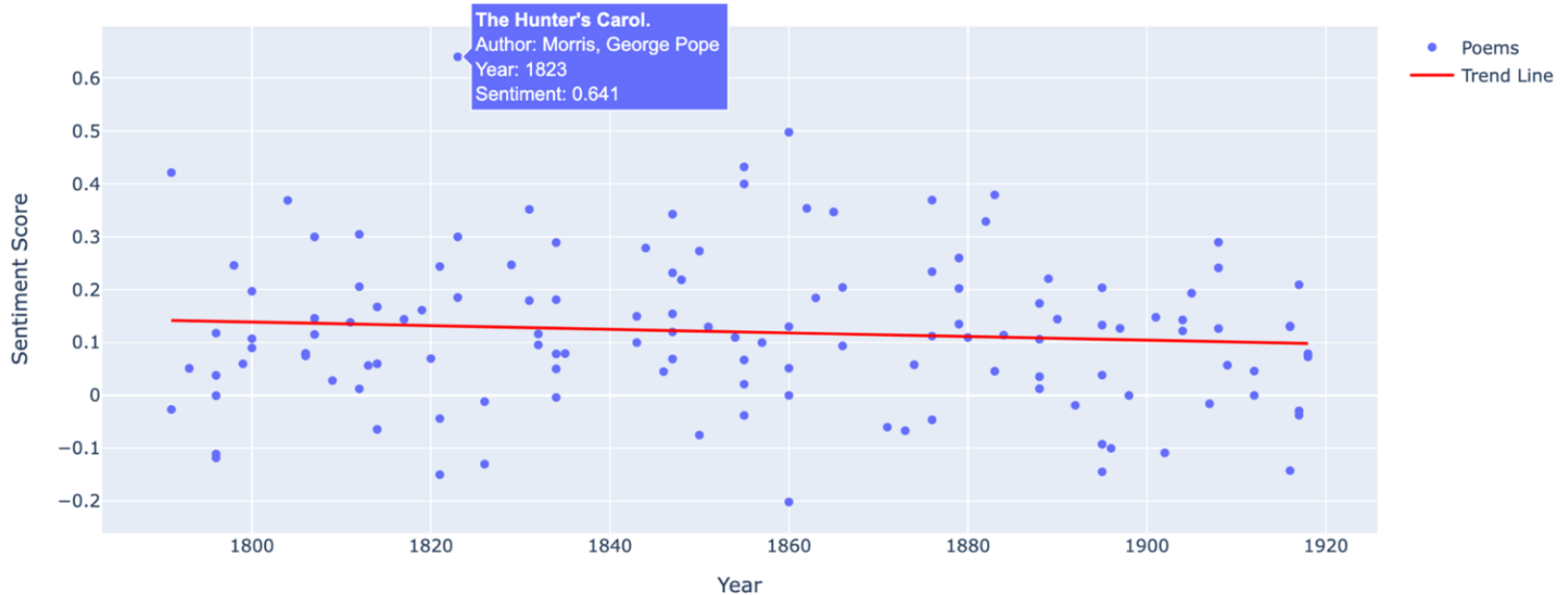
Sentiment Analysis of a single poem



Samuel G. Goodrich: The Mississippi. (1851)

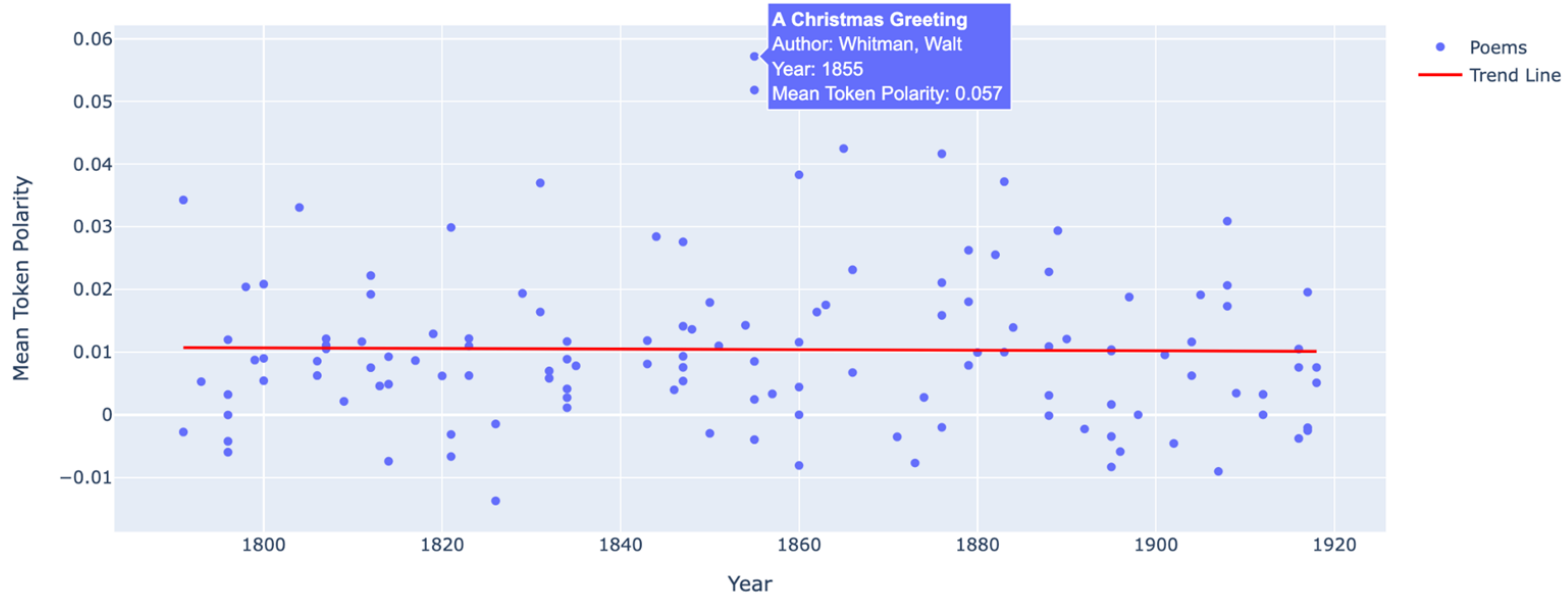
Sentiment Analysis of a corpus of poems

Poetry Sentiment Analysis Over Time



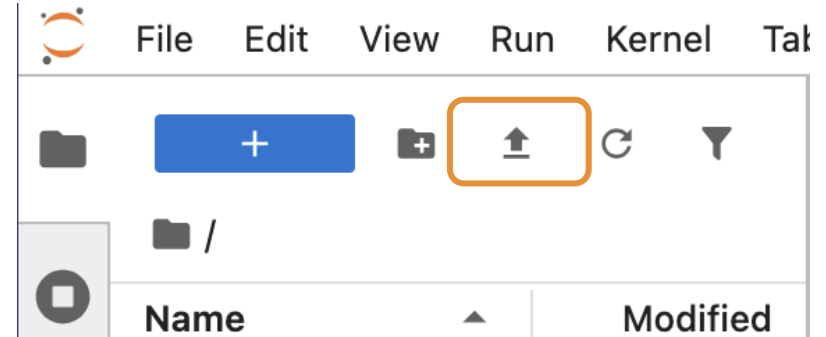
Sentiment richness of a corpus of poems

Mean Token Polarity Over Time



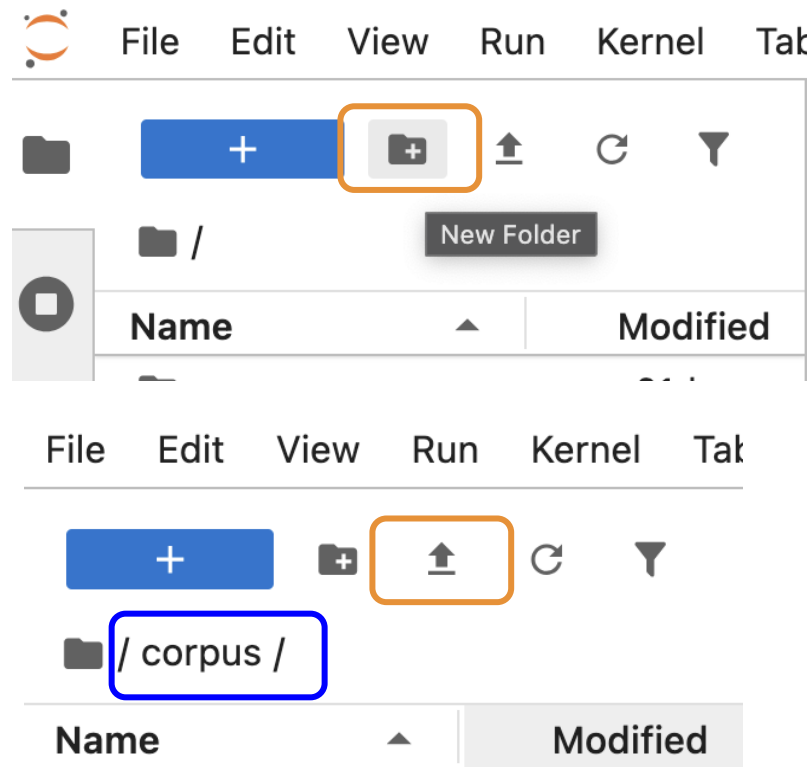
Working on the JupyterHub

1. Upload the Jupyter Notebook



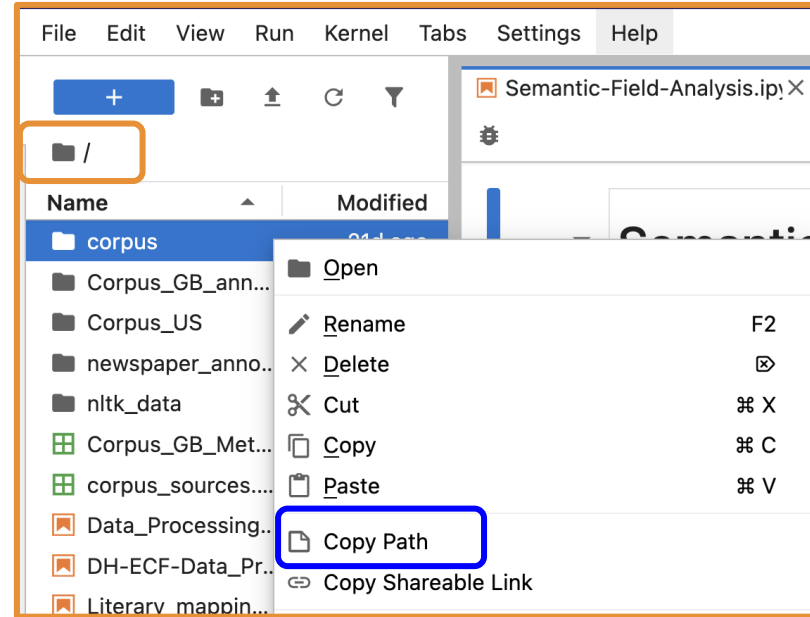
Working on the JupyterHub

1. Create a new folder and give it a telling name **without spaces**
2. Double click on the folder (you should see the name like in the blue rectangle)
3. Upload the corpus files



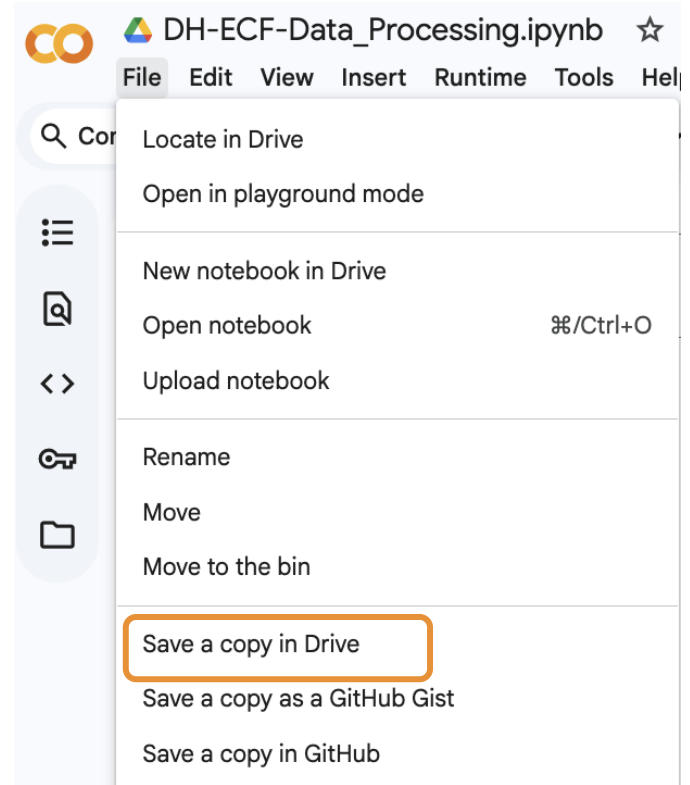
Working on the JupyterHub

1. Navigate back by clicking on the folder symbol (orange rectangle)
2. Right click on the corpus folder and choose “Copy Path” blue rectangle)
3. Paste it in the notebook, where the path to the corpus is set.



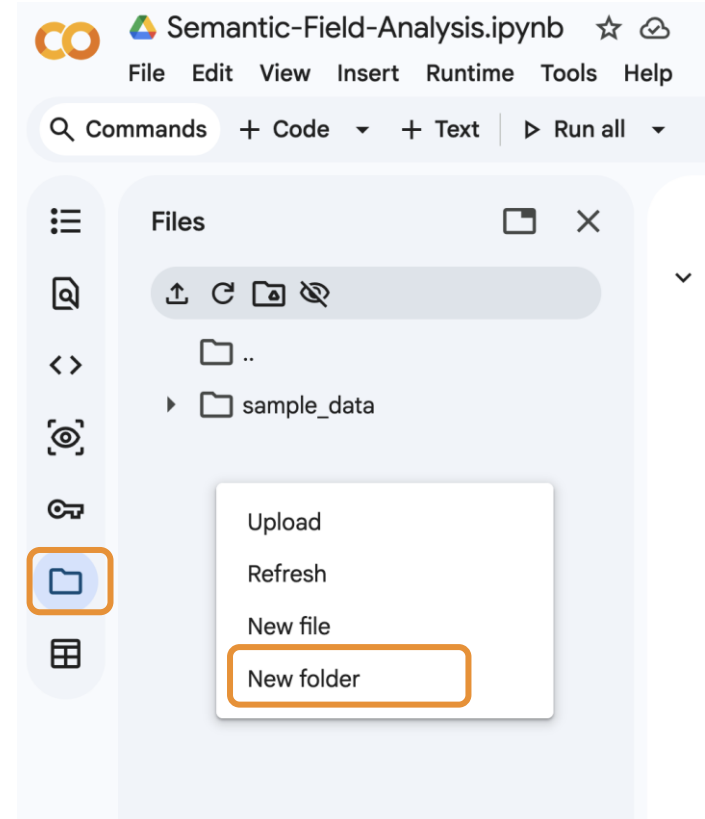
Working in Colab

1. Open Colab Link
2. Save a copy in your drive and work with your own copy



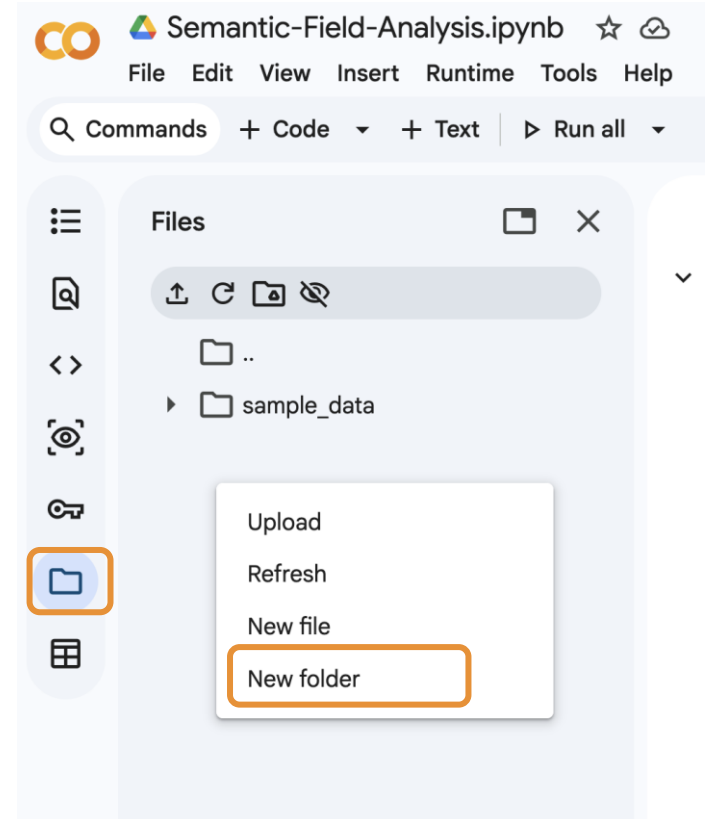
Working in Colab

1. Click on the to the folder on the left
2. Wait until you see the folder “sample data”
3. Right click -> New folder
4. Give the folder a telling name **without spaces**



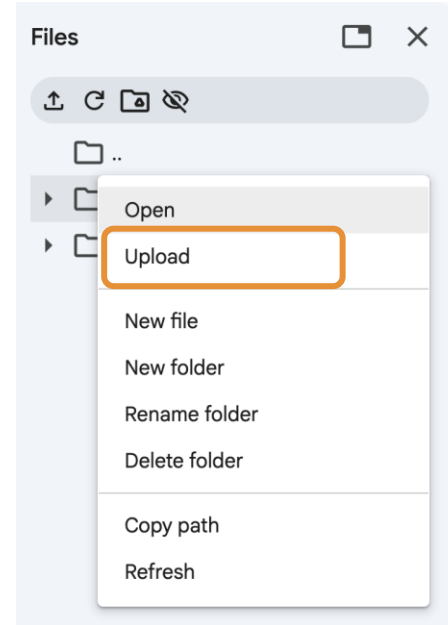
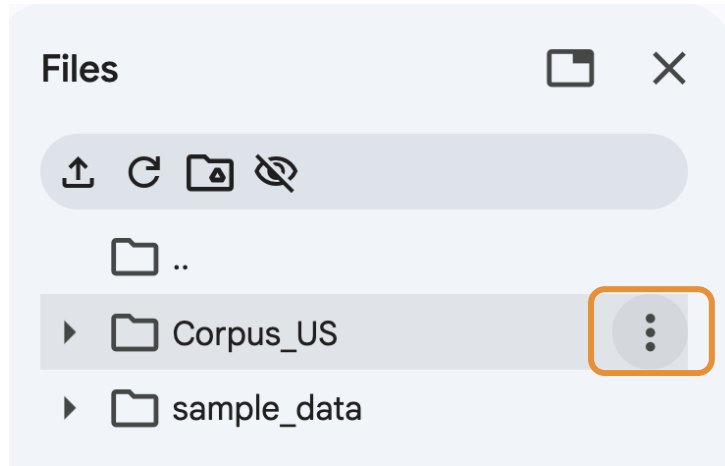
Working in Colab

1. Click on the to the folder on the left
2. Wait until you see the folder “sample data”
3. Right click -> New folder
4. Give the folder a telling name **without spaces**



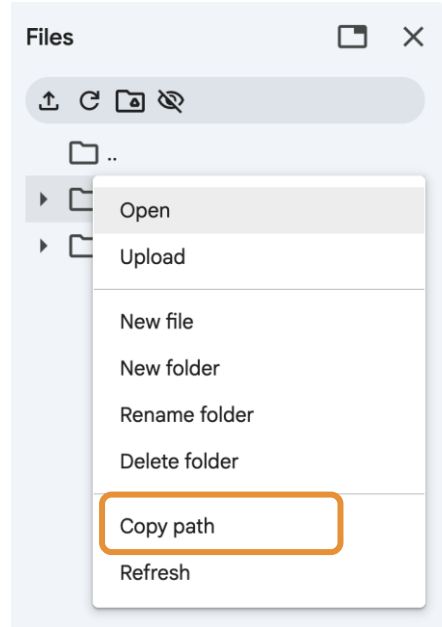
Working in Colab

Click the three dots and select “Upload”. Choose the files you want to upload from your computer.



Working in Colab

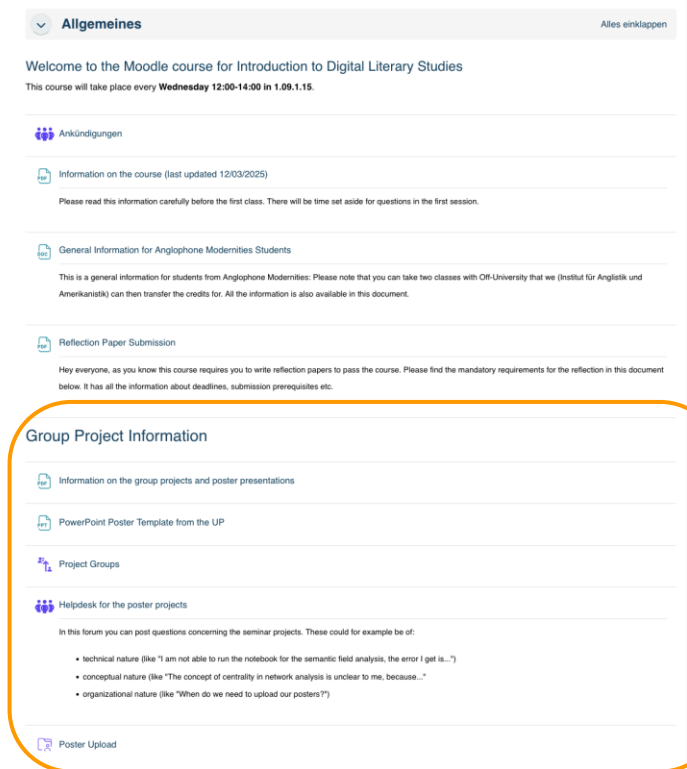
Click “Copy path” and paste it in the notebook, where the path to the corpus is set.



Information on Group Project and Poster Presentation

Procedure:

1. Form groups of 2–4 students.
2. Develop a research question that fits thematically into the seminar, employing either one of the four proposed methods (method-based project) or focusing on the creation of a research corpus (corpus-based project). Look at the checklists for the projects.
3. Enter your question and the group members in this moodle forum until the **7th of January**: <https://moodle2.uni-potsdam.de/mod/forum/view.php?id=2648749>
4. Investigate the research question by conducting experiments and documenting the results or by collecting corpus material and documenting the criteria and the collection process.
5. Create a scientific poster and upload it in Moodle until the **23rd of January** (only one group member should upload the poster for the entire group): <https://moodle2.uni-potsdam.de/mod/publication/view.php?id=2638394>
We will print out your posters.
6. Present the poster on the **28th of January**.



Alles einklappen

Welcome to the Moodle course for Introduction to Digital Literary Studies
This course will take place every **Wednesday 12:00-14:00** in 1.09.1.15.

Ankündigungen

Information on the course (last updated 12/03/2025)
Please read this information carefully before the first class. There will be time set aside for questions in the first session.

General Information for Anglophone Modernities Students
This is a general information for students from Anglophone Modernities: Please note that you can take two classes with Off-University that we (Institut für Anglistik und Amerikanistik) can then transfer the credits for. All the information is also available in this document.

Reflection Paper Submission
Hey everyone, as you know this course requires you to write reflection papers to pass the course. Please find the mandatory requirements for the reflection in this document below. It has all the information about deadlines, submission prerequisites etc.

Group Project Information

Information on the group projects and poster presentations

PowerPoint Poster Template from the UP

Project Groups

Helpdesk for the poster projects
In this forum you can post questions concerning the seminar projects. These could for example be of:

- technical nature (like "I am not able to run the notebook for the semantic field analysis, the error I get is...")
- conceptual nature (like "The concept of centrality in network analysis is unclear to me, because...")
- organizational nature (like "When do we need to upload our posters?")

Poster Upload

Introduction to Digital Literary Studies

Happy Winter Holiday 🧑‍❄️ ❄️ 🙌