

# Computerphilologie: Überblick

---

Workshop Digitale Forschungsmethoden in der Romanistik  
Forum Junge Romanistik 2016, Universität Würzburg

---

Christof Schöch  
CLiGS - Universität Würzburg

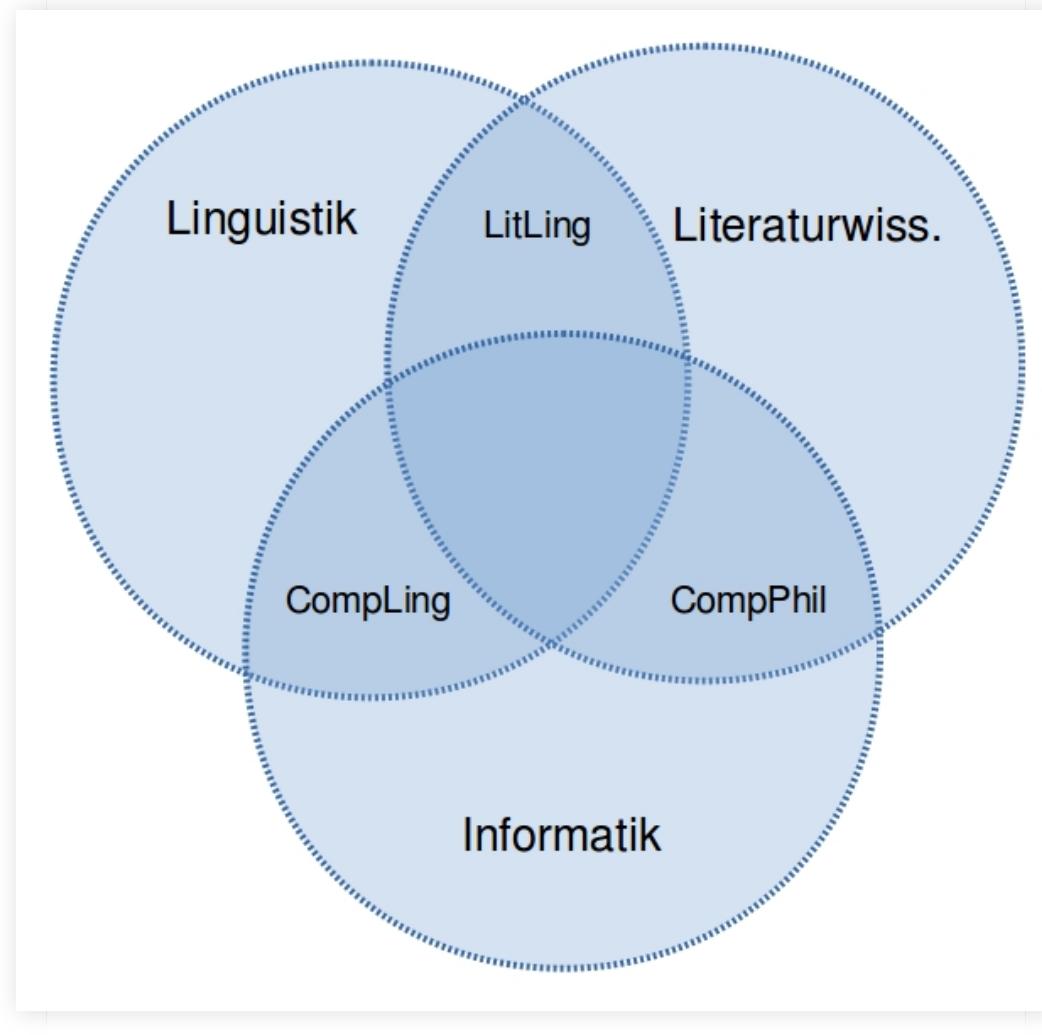


# Überblick

1. Computerphilologie
2. Texterzeugung
3. Textanalyse
4. Weiterführendes

# 1. Was ist Computerphilologie?

# Verortung der Computerphilologie



# Computerphilologie

- Literaturwissenschaft: Erfassen, Erzeugen, Anreichern, Analyse, Interpretation, Speicherung und Dissemination von (literarischen) Texten
- Computerphilologie: Literaturwissenschaft auf Grundlage digitaler Textdaten und mit Hilfe algorithmischer Werkzeuge und Methoden.

(Taxonomie: <https://github.com/dhtaxonomy/TaDiRAH>)

# Zwei zentrale Bereiche der Computerphilologie

- Erzeugen/Anreichern: Digitalisierung, (digitale) Editionen, Textsammlungen, Korpora
- Analysieren: Stil, Inhalt, Räumlichkeit, Figurennetzwerke, Gattungen, etc.

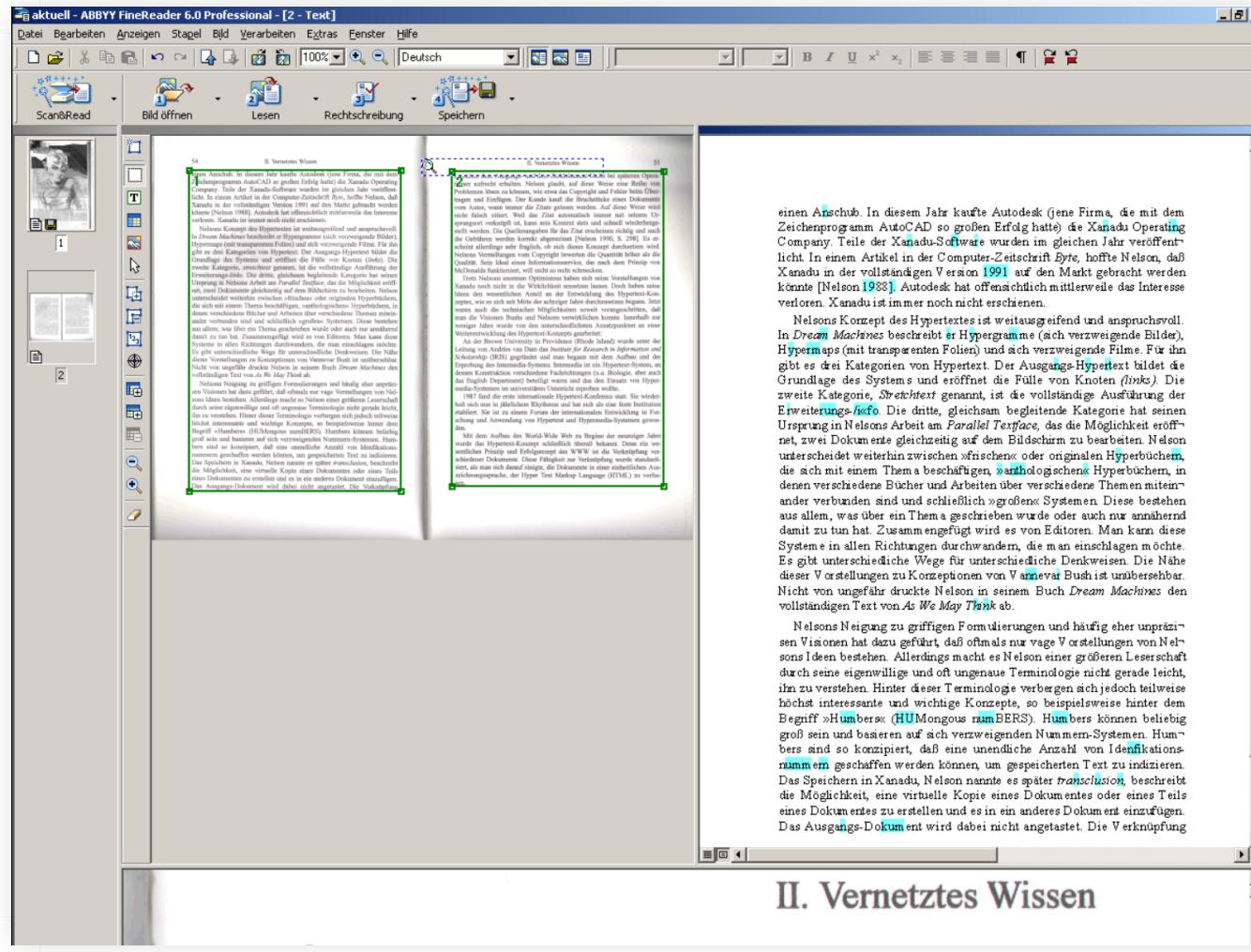
## 2. Erzeugen/Anreichern literarischer Texte

# Aspekte des Erzeugens/Anreichern von Text

- Digitalisierung (Scannen, OCR)
- Digitale Texteditionen und Textsammlungen
- Linguistisch annotierte Korpora
- Editionsphilologie: Stemmata, Kollationierung

# Werkzeuge

# Optical Character Recognition: FineReader



<http://www.abbyy.com/de-de/finereader>

# Text kodieren: TEI

The screenshot shows the homepage of the Text Encoding Initiative (TEI) at <https://www.tei-c.org>. The header features the TEI logo and navigation links for Home, Guidelines, Activities, Tools, Membership, Support, About, and News. A search bar is present. The main content area is titled "P5: Guidelines for Electronic Text Encoding and Interchange" and includes a note about the version being 2.9.1, last updated on 15th October 2015. Below this, there are language links for English, Deutsch, Español, Italiano, Français, 日本語, 한국어, and 中文, along with icons for PDF, XML, and TEI files.

**Front Matter**

- [Title](#)
  - i. [Releases of the TEI Guidelines](#)
  - ii. [Dedication](#)
  - iii. [Preface and Acknowledgments](#)
  - iv. [About These Guidelines](#)
  - v. [A Gentle Introduction to XML](#)
  - vi. [Languages and Character Sets](#)

**Back Matter**

- Appendix A [Model Classes](#)
- Appendix B [Attribute Classes](#)
- Appendix C [Elements](#)
- Appendix D [Attributes](#)
- Appendix E [Datatypes and Other Macros](#)
- Appendix F [Bibliography](#)
- Appendix G [Prefatory Notes](#)
- Appendix H [Colophon](#)

**Text Body**

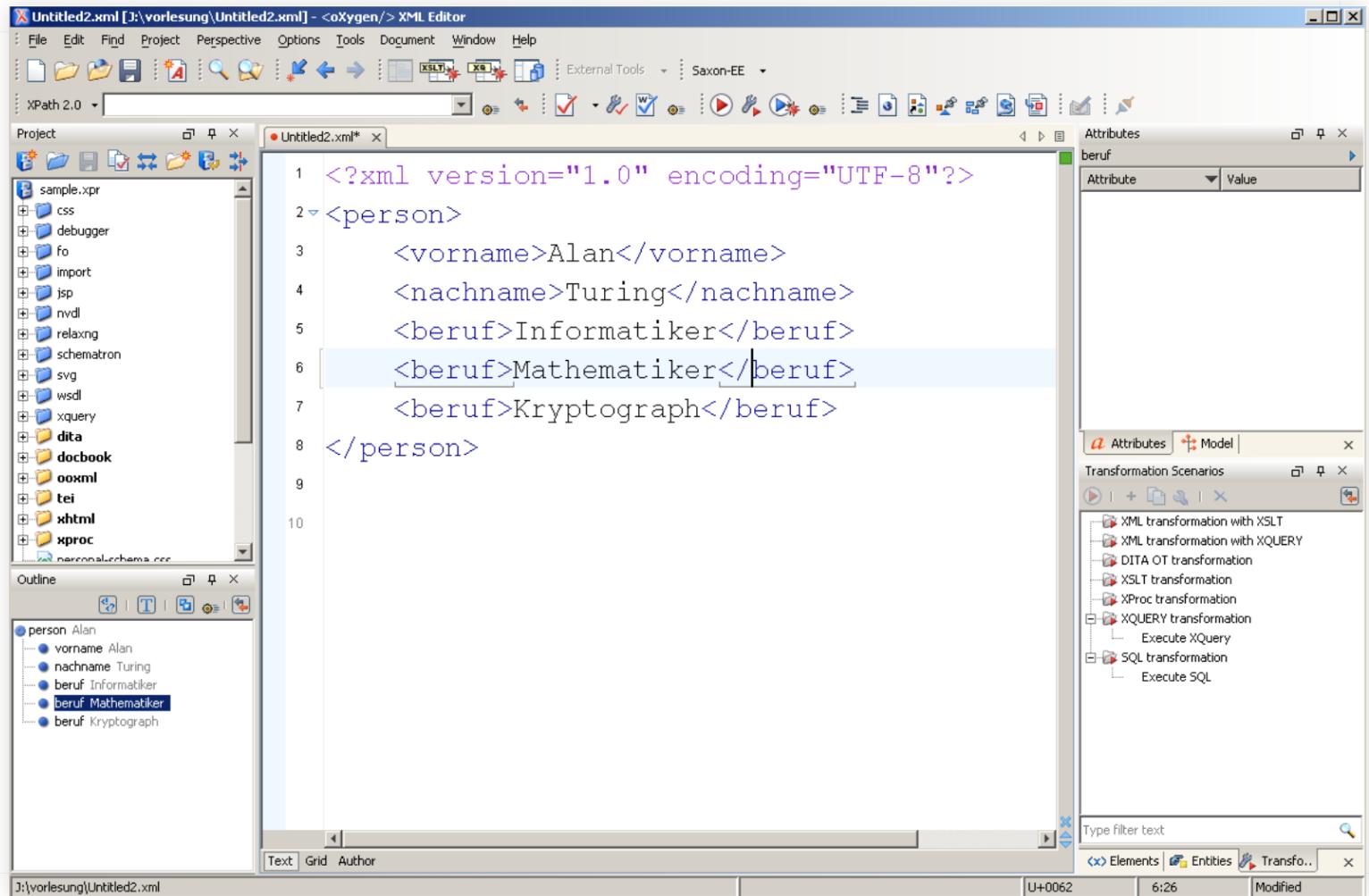
- 1 [The TEI Infrastructure](#)
- 2 [The TEI Header](#)
- 3 [Elements Available in All TEI Documents](#)
- 4 [Default Text Structure](#)
- 5 [Characters, Glyphs, and Writing Modes](#)
- 6 [Verse](#)
- 7 [Performance Texts](#)
- 8 [Transcriptions of Speech](#)
- 9 [Dictionaries](#)
- 10 [Manuscript Description](#)
- 11 [Representation of Primary Sources](#)
- 12 [Critical Apparatus](#)
- 13 [Names, Dates, People, and Places](#)
- 14 [Tables, Formulæ, Graphics and Notated Music](#)
- 15 [Language Corpora](#)
- 16 [Linking, Segmentation, and Alignment](#)
- 17 [Simple Analytic Mechanisms](#)
- 18 [Feature Structures](#)
- 19 [Graphs, Networks, and Trees](#)
- 20 [Non-hierarchical Structures](#)
- 21 [Certainty, Precision, and Responsibility](#)
- 22 [Documentation Elements](#)
- 23 [Using the TEI](#)

**TEI sourcecode**

- [Github Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

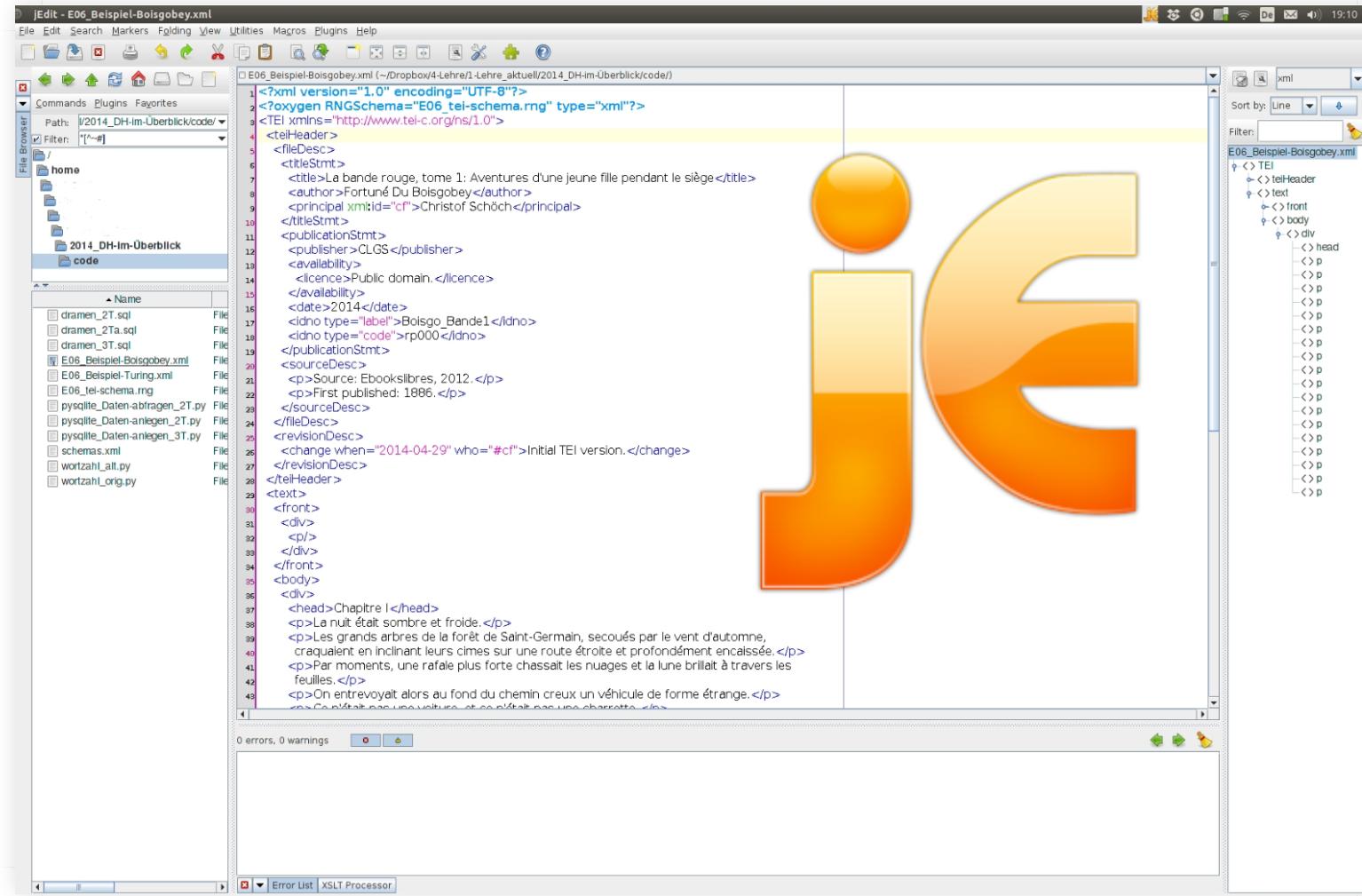
<https://www.tei-c.org>

# XML/TEI-Dateien bearbeiten: oXygen



<https://www.oxygenxml.com>

# XML/TEI-Dateien bearbeiten: jEdit



<http://www.jedit.org/>

# Annotation: CATMA

The screenshot shows the CATMA annotation interface. On the left, a main window displays a text document from Balzac's "La Peau de Chagrin". The text is annotated with various tags, some of which are highlighted with colored boxes. The right side of the interface features two management panes:

- Tag Manager (Top Right):** A table showing active tagsets and their properties. The "Visible" column contains checkboxes, and the "Writable" column contains checked checkboxes.

Markup Collections	Tag color	Visible	Writable
User Markup Collections			
Description			<input checked="" type="checkbox"/>
motivation		<input checked="" type="checkbox"/>	
miscellaneous		<input checked="" type="checkbox"/>	
description		<input checked="" type="checkbox"/>	
intérieur	yellow	<input checked="" type="checkbox"/>	
personnage	red	<input checked="" type="checkbox"/>	
animal	purple	<input checked="" type="checkbox"/>	
objet	green	<input checked="" type="checkbox"/>	

- Tag Manager (Bottom Right):** A detailed view of the "description" tagset, listing individual tags with their corresponding colors and visibility status.

Tags	Tag Color	Visible
miscellaneous		
motivation		
lexique		
description		
intérieur	yellow	
personnage	red	
animal	purple	
objet	green	
paysage état d'âme	light green	
tableau	pink	
bâtiment	dark blue	
paysage	teal	
métadiscours		

<http://www.catma.de/>

# Kollationierung: Juxta

juxta commons<sup>beta</sup>

Signed in as chirstof | Account Info | Sign Out

Select All Sources Sort by

- Bovary-ABU.txt
- Ed1872.txt
- Ed1560.txt
- Ed1512.txt
- Ed1506.txt
- Ed1494M.txt
- Ed1494B.txt

>> Add Source

Select All Witnesses Sort by

- Bovary-PG
- Bovary-WS
- Ed1872
- Ed1560
- Ed1512
- Ed1494B
- Ed1494M
- Ed1506

>> Create Set

Visible to Me Only | Share Description:

Narrenschiff

Witness List Sort by

- Ed1494B [base]
- Ed1494M
- Ed1506
- Ed1512
- Ed1560
- Ed1872

Difference from base

Ed1494B

Wer reden wil/so er nit sol  
Der fügt in narren orden wol  
Wer antwürt/ee man froget in  
Der zeigt sich selbs eyn narren syn

Need Help? Click here | Send us your feedback

The screenshot shows the Juxta Commons interface for the 'Narrenschiff' project. At the top, there are three lists: 'Sources' (Bovary-ABU.txt, Ed1872.txt, Ed1560.txt, Ed1512.txt, Ed1506.txt, Ed1494M.txt, Ed1494B.txt), 'Witnesses' (Bovary-PG, Bovary-WS, Ed1872, Ed1560, Ed1512, Ed1494B, Ed1494M, Ed1506), and 'Comparison Sets' (DRC, wordlist-comparison, Narrenschiff, 3Bovary, Bovary-WS, Bovary-PG, Bovary-ABU). Below these are sections for 'Visible to Me Only' sharing, help, and feedback. The main area displays a 'Witness List' for 'Ed1494B' with sub-sections for other witnesses like Ed1494M, Ed1506, etc., each showing a progress bar for 'Difference from base'. To the right, a detailed view of 'Ed1494B' shows text differences between the base and the witness.

<http://www.juxtasoftware.org>

# Editionen

# Walt Whitman Archive

**SUPPORT THE ARCHIVE**

**Search the Archive**

**Go**

**The Walt Whitman Archive**

Ed Folsom & Kenneth M. Price, Editors

**Published Works**  
Books by Whitman; Periodicals; Translations

**In Whitman's Hand**  
Manuscripts; Notebooks; Scribal Documents

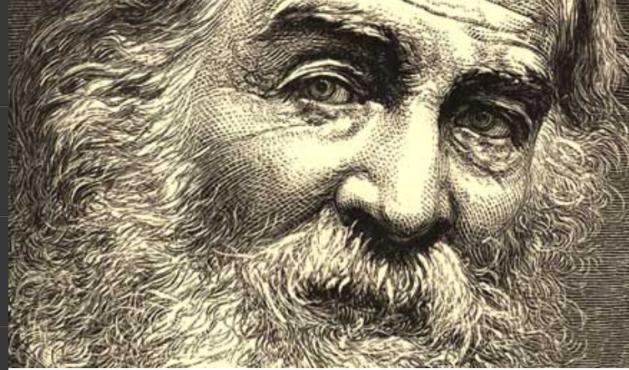
**Life & Letters**  
Biography; Chronology; Letters

**Commentary**  
Bibliography; Criticism; Disciples; Reviews

**Pictures & Sound**  
Audio; Portraits of Whitman

**Resources**  
Downloads; Manuscript Catalogs; Teaching; More...

**About the Archive**  
Follow Us; Tour; People; Policies; More...



**News & Updates**

February 2016  
[Applications and paper proposals](#) are currently being accepted for the Ninth Annual International Whitman Week Seminar and Symposium to be held in Exeter, England, in May/June 2016.

December 2015  
A new section of the Archive devoted to Whitman's [marginalia and annotations](#) is now available. Thanks to the National Endowment for the Humanities for support of this work.

November 2015  
With support from the National Historical Publications and Records Commission, we've published 74 additional [letters](#) from the post-Reconstruction years.

**More News**

Funding and support provided by the National Endowment for the Humanities, the US Department of Education, the Institute of

<http://www.whitmanarchive.org/>

# Devonshire MS Social Edition

Not logged in Discussion for this IP address Contributions Create account Log in

Book Discussion Read Latest draft Edit View history Search

## A Social Edition of the Devonshire MS (BL Add 17,492)

The latest reviewed version was checked on 2 February 2015. There are template/file changes awaiting review.

### Contents [edit]

Acknowledgments • Note on the Edition

▼ A Study of the Devonshire Manuscript

Introduction: The First Sustained Example of Men and Women Writing Together in the English Tradition

Overview • The Works of the Devonshire Manuscript • Public and Private, Personal and Communal • Current and Popular Contexts

Contributors to the Devonshire Manuscript

Biographies

Genealogical Tables

Family and Affairs of Henry VIII • The Howard Family • The Boleyn Family • Connections of the Boleyn, Knayet, Shelton, and Howard Families

Textual and Bibliographic Materials

Textual Introduction

Bibliographic Analysis • Provenance • Punctuation and Scribal Marks • Transcription

Manuscripts & Early Printed Books Associated with the Devonshire Manuscript

Sigla • Witness Descriptions, Manuscript • Witness Descriptions, Early Printed Books • Items in the Devonshire Manuscript Without Witnesses

Paleographic Information

► Detailed Hand Lists • Attribution, in Order of Appearance • Features

Witnesses

Bibliography

Bibliography A-M • Bibliography N-Z •

► The Devonshire Manuscript, BL ADD MS 17492

Search this book



[https://en.wikibooks.org/wiki/The\\_Devonshire\\_Manuscript](https://en.wikibooks.org/wiki/The_Devonshire_Manuscript)

# Van Gogh Letters

Van Gogh Museum huugens ing

**Vincent van Gogh**  
*The Letters*

by period  
by correspondent  
by place  
with sketches

Search keyword or number(s) Advanced search Search results

Van Gogh as a letter-writer  
Correspondents  
Biographical & historical context  
Publication history  
About this edition  
Chronology  
Concordance, lists, bibliography  
Book edition

038 | 040 »

To Theo van Gogh. Paris, Saturday, 24 July 1875.

original text + line endings facsimile translation notes artworks

039

Br. 1990.039 | CL: 32  
From: Vincent van Gogh  
To: Theo van Gogh  
Date: Paris, Saturday, 24 July 1875  
morgens

1 Goupil & Cie  
2 Editeurs Imprimeurs  
3 Estampes Françaises & Étrangères  
4 Tableaux Modernes  
5 Rue Chaptal, 9, PARIS.  
6 Succursales à la Haye, Londres, Berlin, New-York.

7 Paris, le 24 Juillet [1875]

8 Waarde Theo,  
9 Een paar dagen geleden kregen wij een Sch<sup>j</sup>  
van de Nittis, een gezicht in den ver-  
wonding, Westminster bridge & the house of  
Parliament. Ik ging elken morgen vannen  
over Westminsterbridge & weet hoe dat er uit  
siet als die zon achter Westminster abbey &  
the house of Parliament ondergaat & hoe het  
smorgens vroeg & 's winters met sneeuw & met  
mist.  
Toen ik dit sch<sup>j</sup> zag voelde ik hoe ik voor  
denken hoor.  
Toch geloof ik het goed voor mij is & daar-  
dien ben... Siet in ontwerp op ons vroeg-  
Siet op nuar denken gaaf geloof ik rebe-  
niet...

physical description

top ↴

original text + line endings facsimile translation notes artworks

[sketch A] ↴

[sketch A] ↴

[Letterhead: Goupil Paris]

Paris, 24 July 1875

My dear Theo,  
A couple of days ago we got a painting by De Nittis, a view of London on a rainy day, Westminster Bridge and the Houses of Parliament. I crossed Westminster Bridge every morning and evening and know what it looks like when the sun's setting behind Westminster Abbey and the Houses of Parliament, and what it's like in the morning, and in the winter with snow and fog.  
When I saw this painting I felt how much I love London.  
Yet I believe it's good for me to be away from it. This is answer to your question. I certainly don't think that you'll be going to London. ↴

Thanks for "Aus der Jugendzeit" and "Um Mitternacht" by Rückert. ↴ It's poignantly beautiful; the latter made me think of "La nuit de Décembre" by Musset. ↴ I wish I could send it to you, but don't have it.  
Yesterday we sent a crate to The Hague, what I had promised you was in it. ↴

I hear that Anna and Lies are at home; I'd like to see them again. I wish you the very best, and write again soon. With a handshake

Your loving brother  
Vincent

top ↴

Dank voor "Aus der Jugendzeit" & "Um Mitter-  
nacht" van Rückert. ↴ Is roerend schoon; 't laatst-  
genoemde deed mij denken aan "La nuit  
de Décembre" van Musset. ↴ Ik wilde ik U  
dat kon zenden, doch heb het niet.—  
Gisteren hebben wij een kist naar den Haag  
gezonden, waarin voor U, wat ik U beloofde. ↴  
Ik hoor Anna en Lies thuis zijn, ik zou hen  
wel eens willen zien. Heb het zoogood  
mogelijk & schrijf weer spoedig. En te  
serrant la main...

Uw liebh. broeder  
Vincent

1-7 Printed letterhead.

top ↴

1. Giuseppe de Nittis sold his work through Goupil. On the basis of the sketch in the upper left-hand corner of the letter, J.G. van Gelder concluded that the work Van Gogh is referring to here is *Victoria Embankment in London, Westminster Bridge (I)*, *Westminster Bridge (II)* and *Westminster Bridge (III)*. See Dini and Marini 1990, p. 397, cat. no. 542 (*Victoria Embankment in London*) and p. 405, cat. nos. 712-714 (*Westminster Bridge I-III*). Dini and Marini did not date the paintings; the date of completion is known only of *Westminster Bridge (I)* (1863). Moreover, *Victoria Embankment in London* clearly shows a sunny day, whereas Van Gogh speaks of a 'rainy day' (L.11). Cf. also Caroline Igra, 'Spatial engineer and social recorder: Giuseppe de Nittis and the development of nineteenth-century cityscape imagery', Van Gogh Museum Journal 1998. Amsterdam 1999, pp. 94-103; esp. pp. 96-97.

2. Rückert's "Aus der Jugendzeit" (From childhood) expresses a yearning for lost youth; "Mitternacht" (Midnight) is an expression of Weitschmerz, deliverance from which is sought in God. For these poems, see Rückert 1868-1869, vol. 5, pp. 29-30 and vol. 2, pp. 465-466.

3. Alfred de Musset's long poem, "La nuit de décembre" (A night in December) is about a man who continually sees a sorrowful ghost who resembles him like a brother. At the end the ghost reveals his identity: he is Loneliness and will be the man's companion for ever. See Musset 1957, pp. 310-315, 737-742. On p. [1-4] of letter 31, the eighteenth stanza of "La nuit de décembre" was written by another hand. See also letter 726, n. 16.

4. A French bible and *L'imitation de Jésus-Christ*; see letter 38, ll. 45-46.

top ↴

<http://vangoghletters.org/vg/>

# Moralische Wochenschriften

Allemand Anglais Français Italien Espagnol

*Periodical Press Spettatori Espectadores  
Spectateurs Moralische Wochenschriften*



ANGLETERRE FRANCE ALLEMAGNE ITALIE ESPAGNE Liste complète Recherche élargie Recherche Recherche

**Les « spectateurs » dans le contexte international**

Le genre journalistique des « spectateurs » (ou feuilles moralistes), qui trouve son origine au début du XVIII<sup>e</sup> siècle en Angleterre, s'est propagé rapidement dans toute l'Europe avant de devenir un indicateur important du discours des Lumières.

Le contenu des « spectateurs » s'adressait généralement à un large public de lecteurs urbains. Les idées et les valeurs du siècle des Lumières devaient y être propagées d'une manière divertissante. Ainsi les feuilles moralistes jouaient un rôle important dans la genèse de l'opinion publique.

Notre projet vise à créer une base de données intégrale pour toute la presse moraliste de ce genre. Actuellement notre base offre un grand nombre de représentations et d'analyses de textes espagnols, italiens et français. On peut y trouver également la visualisation de leurs niveaux de narration ainsi que de leurs formes narratives. D'autres textes sont en cours de préparation.

Klaus-Dieter Ertler, Alexandra Fuchs, Michaela Fischer, Elisabeth Hobisch Février 2011

• Description du projet  
• Bibliographie

  
Cliquez pour choisir un pays

Page principale Information éditoriale Cirilo



<http://gams.uni-graz.at/mws/>

# Faust-Edition

beta FAUST EDITION Archiv > Goethe- und Schiller-Archiv Genese > Faust I > Dom 1 H.9

39 Ich wäre nicht so arm an Witz  
Wär ich nicht gar so arm an Reimen.

33 Wie man nach Norden weiter kommt  
Da nehmen Rus und Hexen zu.

35 Ihr Leben ist ein blöder Zeitvertreib  
Zwey lange Beine kleinen Leib  
Sie kicken  
den Unfug den Sie jetzt in Deutfch  
angefangen Richet

34 Leuchtende Finger des Meph.  
Viel appetit und wenig Geschmac.  
Wir haben wie manche die Gelellen  
Viel appetit und wenig Geschmac.

44 Mußt dich nur her und wärs ein Duderlack  
Wie Vapek und Warts ein Duderlack  
Wir haben wie manche edle Getreide  
Wie Vapek nur her und wärs ein Duderlack

Impressum ?

Archiv Genese Text

39 Ich wäre nicht so arm an Witz  
Wär ich nicht gar so arm an Reimen.

33 Wie man nach Norden weiter kommt  
Da nehmen Rus und Hexen zu.

35 Ihr Leben ist ein blöder Zeitvertreib  
Zwey lange Beine kleinen Leib  
Sie kicken  
den Unfug den Sie jetzt in Deutfch  
angefangen Richet

34 Leuchtende Finger des Meph.  
Viel appetit und wenig Geschmac.  
Wir haben wie manche die Gelellen  
Viel appetit und wenig Geschmac.

44 Mußt dich nur her und wärs ein Duderlack  
Wie Vapek und Warts ein Duderlack  
Wir haben wie manche edle Getreide  
Wie Vapek nur her und wärs ein Duderlack

Goethe- und Schiller-Archiv X

+ - C C

« < 2 / 4 > »

≡ | | T T T

<http://beta.faustedition.net>

# 3. Analyse literarischer Texte

# Perspektiven der digitalen Textanalyse

- Quantitative vs. qualitative Verfahren
- Klassifikation vs. Clustering
- Informationsextraktion vs. Datenvizualisierung
- GUI vs. CLI

# Natural Language Processing: NLTK

## 1 Using a Tagger

A part-of-speech tagger, or **POS-tagger**, processes a sequence of words, and attaches a part of speech tag to each word (don't forget to `import nltk`):

```
>>> text = word_tokenize("And now for something completely different")
>>> nltk.pos_tag(text)
[('And', 'CC'), ('now', 'RB'), ('for', 'IN'), ('something', 'NN'),
('completely', 'RB'), ('different', 'JJ')]
```

Here we see that *and* is cc, a coordinating conjunction; *now* and *completely* are RB, or adverbs; *for* is IN, a preposition; *something* is NN, a noun; and *different* is JJ, an adjective.

### Note

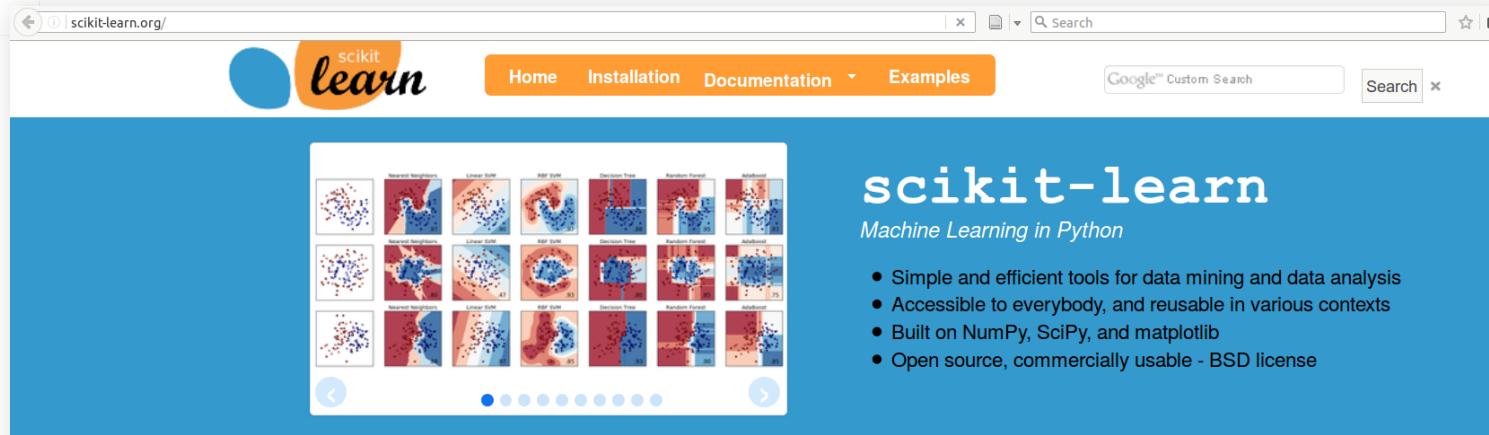
NLTK provides documentation for each tag, which can be queried using the tag, e.g. `nltk.help.upenn_tagset('RB')`, or a regular expression, e.g. `nltk.help.upenn_tagset('NN.*')`. Some corpora have README files with tagset documentation, see `nltk.corpus.???.readme()`, substituting in the name of the corpus.

Let's look at another example, this time including some homonyms:

```
>>> text = word_tokenize("They refuse to permit us to obtain the refuse permit")
>>> nltk.pos_tag(text)
[('They', 'PRP'), ('refuse', 'VBP'), ('to', 'TO'), ('permit', 'VB'), ('us', 'PRP'),
('to', 'TO'), ('obtain', 'VB'), ('the', 'DT'), ('refuse', 'NN'), ('permit', 'NN')]
```

Notice that *refuse* and *permit* both appear as a present tense verb (VBP) and a noun (NN). E.g. *refUSE* is a verb meaning "deny," while *REFuse* is a noun meaning "trash" (i.e. they are not homophones). Thus, we need to know which word is being used in order to pronounce the text correctly. (For this reason, text-to-speech systems usually perform POS-tagging.)

# Maschinelles Lernen: scikit-learn



The screenshot shows the official website for scikit-learn ([scikit-learn.org/](http://scikit-learn.org/)). The header features the scikit-learn logo and navigation links for Home, Installation, Documentation, Examples, and a search bar. Below the header is a banner with the text "scikit-learn" and "Machine Learning in Python", followed by a bulleted list of features: Simple and efficient tools for data mining and data analysis; Accessible to everybody, and reusable in various contexts; Built on NumPy, SciPy, and matplotlib; Open source, commercially usable - BSD license. A grid of nine small heatmaps illustrates different machine learning models like Nearest Neighbors, Linear SVM, RBF SVM, Decision Tree, Random Forest, and Adaboost. The main content area is divided into several sections: Classification, Regression, Clustering, Dimensionality reduction, Model selection, and Preprocessing, each with a brief description, applications, algorithms, and examples.

**Classification**  
Identifying to which category an object belongs to.  
**Applications:** Spam detection, Image recognition.  
**Algorithms:** SVM, nearest neighbors, random forest, ... — Examples

**Regression**  
Predicting a continuous-valued attribute associated with an object.  
**Applications:** Drug response, Stock prices.  
**Algorithms:** SVR, ridge regression, Lasso, ... — Examples

**Clustering**  
Automatic grouping of similar objects into sets.  
**Applications:** Customer segmentation, Grouping experiment outcomes.  
**Algorithms:** k-Means, spectral clustering, mean-shift, ... — Examples

**Dimensionality reduction**  
Reducing the number of random variables to consider.  
**Applications:** Visualization, Increased efficiency.  
**Algorithms:** PCA, feature selection, non-negative matrix factorization. — Examples

**Model selection**  
Comparing, validating and choosing parameters and models.  
**Goal:** Improved accuracy via parameter tuning  
**Modules:** grid search, cross validation, metrics. — Examples

**Preprocessing**  
Feature extraction and normalization.  
**Application:** Transforming input data such as text for use with machine learning algorithms.  
**Modules:** preprocessing, feature extraction. — Examples

<http://scikit-learn.org>

# Korpusanalyse: TXM

The screenshot displays the TXM (Text Mining) software interface, specifically the 'Cooccurrences' module. The left panel shows a tree view of the corpus structure, with 'ROMAN19' selected. The main panel contains two tabs: 'ROMAN19: [frlemma="musique"] (5, 5)' and 'ROMAN19:([frlemma="musique"] []\* [word="instruments|poésie"] ) | ([word="instruments|poésie"] []\* [frlem...'. The first tab shows a table of cooccurrence statistics for the word 'musique', with rows for 'instruments' and 'poésie' highlighted in orange. The second tab shows examples of context windows for these words.

**ROMAN19: [frlemma="musique"] (5, 5)**

Query: [frlemma="musique"]

Cooccurrences properties: word Edit Thresholds: Fmin ≥ 2 Cmin ≥ 2 Score ≥ 4

Context: word structure s use left Window use right Window include the keyword structure

From - 5 to - 0 and from 0 to 5

Cooccurrent	Frequency	Cofrequency	Score	Mean distance
la	99125	161	.34	.7
de	172450	162	.11	1.4
instruments	130	5	7	1.6
poésie	88	4	6	3.0
instrumentale	2	2	6	2.0
ophicléides	3	2	6	3.0
Théorie	3	2	6	2.0
militaire	138	4	5	.0
entendre	1136	7	5	3.1
copier	6	2	5	1.5
sons	58	3	5	2.3
musique	221	4	5	2.5
régiments	11	2	4	1.0
relier	12	2	4	1.0
pour	24246	30	4	2.2
maître	2145	8	4	1.6
des	32495	35	4	2.6
Nicétas	164	3	4	4.0
jouer	445	4	4	2.5

**ROMAN19:([frlemma="musique"] []\* [word="instruments|poésie"] ) | ([word="instruments|poésie"] []\* [frlem...)**

Query: [word="instruments|poésie"] []\* [frlemma="musique"] ) | ([word="instruments|poésie"] []\* [frlem...)

sort keys: #1 None #2 None #3 None #4 None Sort

Left context      Keyword      Right context

Left context	Keyword	Right context
principe vital Est une hypothèse gratuite ! La marchand de vin, par un facteur d'	poésie et la musique	parurent admirables à L et par un libraire qui ver
et des pincelettes, des manchons et des a des festivals pour tout : pour la le mur, étaient posés ou accrochés des des paysans couvrait entièrement parfois la chanson des	instruments de musique	, des dentelles, des mari , pour le vin blanc, pour l , des violons, des cornet déchirée par les voix dé
et, d'ailleurs, assez frottée de	instruments ; et la frêle musique	, depuis son enfance, po
sait toutes choses mieux que moi : la , vos livres, vos chasses, vos	musique, de peinture et de poésie poésie, la musique	, les langues, les science instruments de musique vous attendent, les être

Console

System Output  
Found 9 occurrences

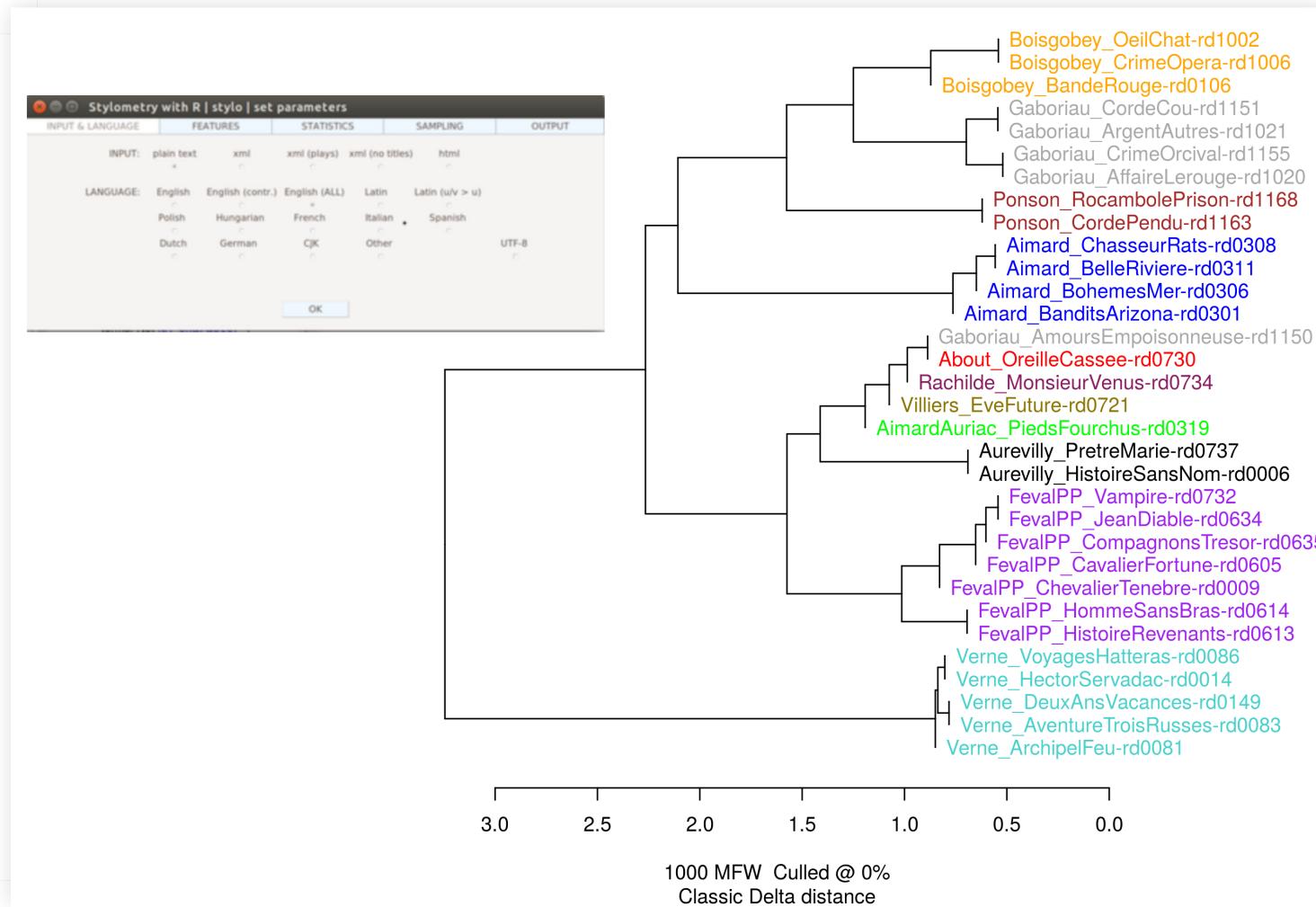
# Topic Modeling: MALLET

```
christof@DELL: ~/Programs/mallet
christof@DELL:~$ cd Programs/mallet
christof@DELL:~/Programs/mallet$ bin/mallet import-dir --input /home/christof/Re
pos/clgs/polar/txt10d --output polar.mallet --keep-sequence --token-regex '\p{L
}[\p{L}\p{P}]*\p{L}' --remove-stopwords TRUE --stoplist-file stoplists/fr3.txt
Labels =
    /home/christof/Repos/clgs/polar/txt10d
christof@DELL:~/Programs/mallet$ bin/mallet train-topics --input polar.mallet -
--num-topics 30 --optimize-interval 200 --num-iterations 4000 --num-top-words 30
--word-topic-counts-file results/polar10d_words-by-topics.txt --output-state top
ic-state.gz --output-topic-keys results/polar-10d_topics-with-words.txt --output
--doc-topics results/polar10d_topics-in-texts.txt --doc-topics-max 30
```

---

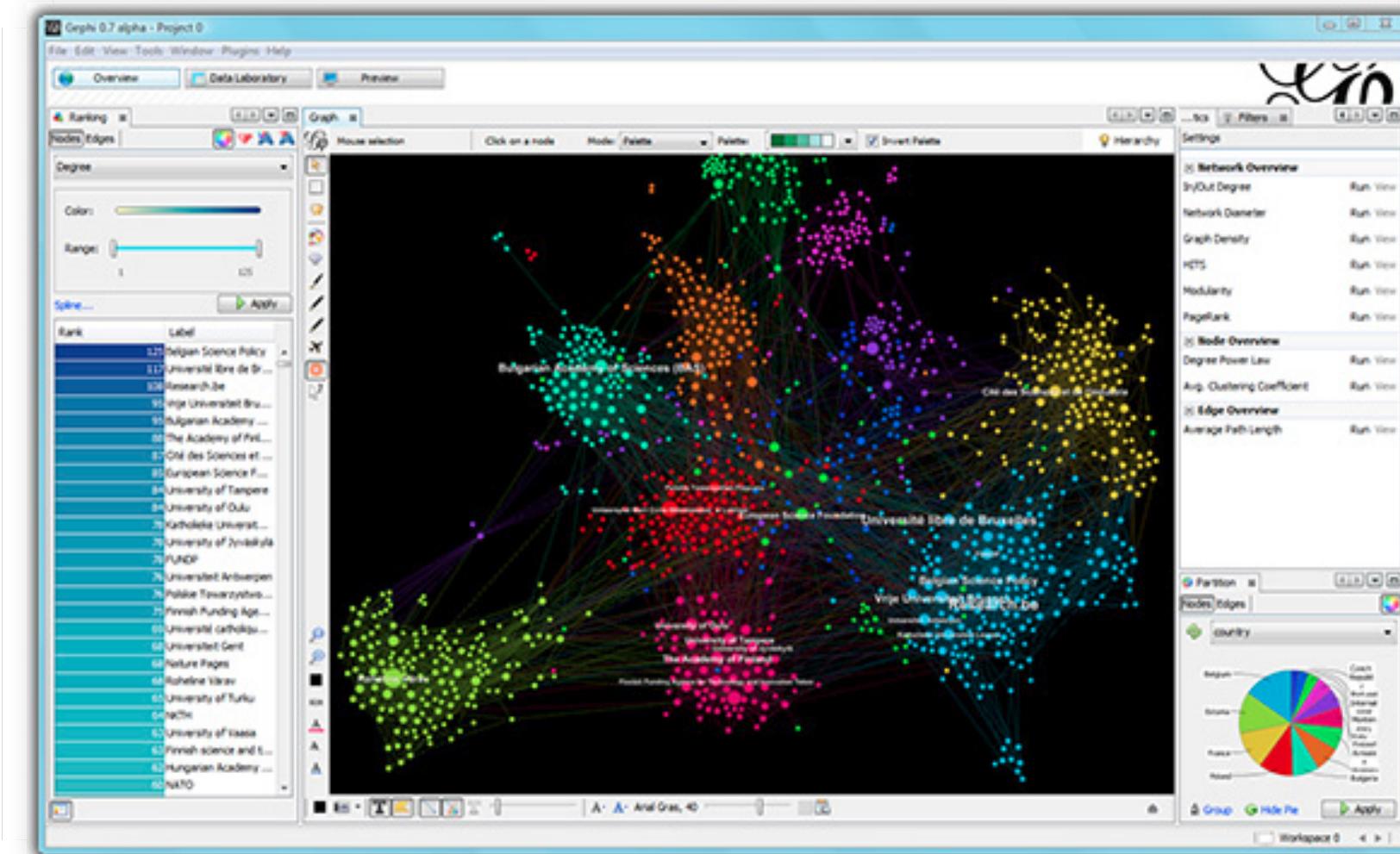
<http://mallet.cs.umass.edu/topics.php>

# Stilometrie: stylo for R



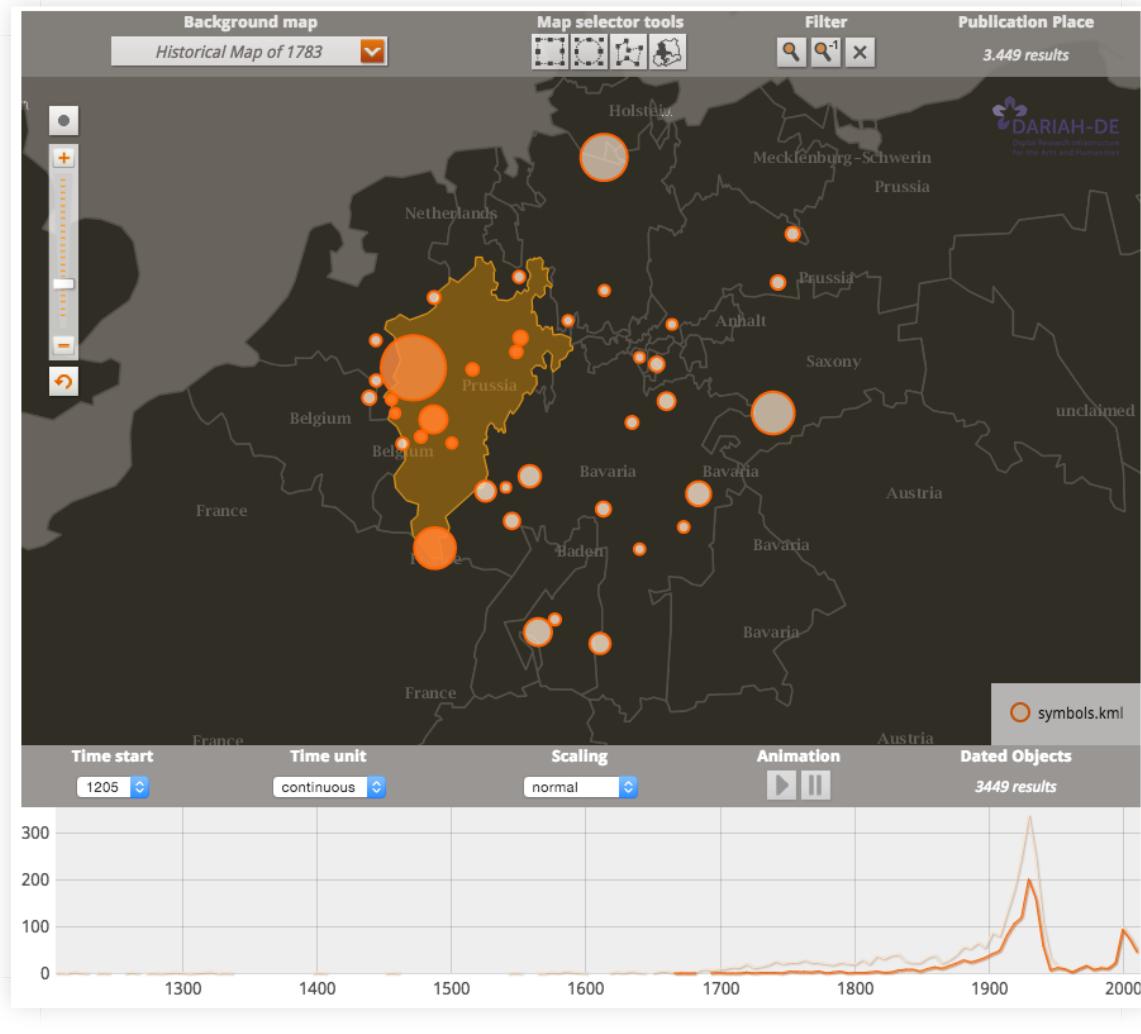
<https://sites.google.com/site/computationalstylistics/>

# Netzwerkanalyse: Gephi



<https://gephi.org/>

# Kartierung: DARIAH Geobrowser



<https://de.dariah.eu/geobrowser>

# 4. Weiterführendes

# Lektürehinweise (einführend)

- Jannidis, F. (2010). "Methoden der computergestützten Textanalyse". In Nünning, A. and Nünning, V. (eds), *Methoden der literatur- und kulturwissenschaftlichen Textanalyse*. Stuttgart & Weimar: Metzler, pp. 109–32.
- Jannidis, F. (2007). "Computerphilologie". In Anz, T. (ed), *Handbuch Literaturwissenschaft*, vol. 2 (Methoden und Theorien). Stuttgart: Metzler, pp. 27–40.
- Schöch, C. (im Erscheinen). "Ein digitales Textformat für die Literaturwissenschaften. Die Richtlinien der Text Encoding Initiative und ihr Nutzen für Textedition und Textanalyse". *Romanische Studien*.

# Lektürehinweise (vertiefend)

- Burnard, L. (2014). *What Is the Text Encoding Initiative?*. Marseille: OpenEdition Press.
- Shillingsburg, P. (2006). *From Gutenberg to Google. Electronic Representations of Literary Texts*. Cambridge: Cambridge Univ. Press.
- Alpaydin, E. (2010). *Introduction to Machine Learning*. 2nd ed. Cambridge, Mass: MIT Press.
- Ramsay, S. (2011). *Reading Machines : Toward an Algorithmic Criticism*. Urbana Ill.: University of Illinois Press.

# Online-Ressourcen

- Catalogue of Digital Scholarly Editions: <http://www.digitale-edition.de/index.html>
- Textdaten Romanistik Wiki: <http://wikis.fu-berlin.de/display/tdr/>
- Doing Digital Humanities Bibliography: [https://www.zotero.org/groups/doing\\_digital\\_humanities\\_-\\_a\\_dariah\\_bibliography](https://www.zotero.org/groups/doing_digital_humanities_-_a_dariah_bibliography)

Christof Schöch, 2016

<http://www.christof-schoech.de>

---

Lizenz: Creative Commons Attribution 4.0 International

<https://creativecommons.org/licenses/by/4.0/>