



Biblical intertextuality in a digital world - The tool GERTRUDE

Göttingen Centre for Digital Humanities (GCDH)
& Leipzig University, Natural Language Processing

Annette Geßner, Christian Kötteritzsch, Gerhard Lauer

DH-CASE Workshop, Florence
2013-09-10

The Project

eTRACES - sub-project Göttingen

BMBF-sponsored

Humanists and computer scientists working
together

Text-Mining used for Text Re-Use-Detection
(automatic approach to find textual relations)

The Tool GERTRUDE:
Göttingen E-Research Text Re-Use for Digital
Editions

The Idea

Text-Re-Use-Detection



- How helpful can an approach to identify textual re-use automatically (text-mining) be?
 - How well does it help to find forms of intertextuality and other textual relations?
 - Is crowd-sourcing useful to determine this?
-



The Text Corpus

- German textcorpus from zeno.org, “Literatur” (dramas, novels, poems, ... ca. 1500-1900)
 - Very heterogeneous genres and orthography
 - Also German translations from other languages
- Under a Creative Commons License, sponsored by the BMBF
- Use-case: The Luther-Bible of 1912
 - very influential work
 - interesting for different fields of research



Preprocessing of the corpus

- Lemmatisation
 - Lots of different spelling variants
 - Making list out of the corpus
 - Check against the „Wortschatzportal“ word list
 - Look at words that don't appear in the list
 - rules to transform words in modern German
 - Check transformed words against the Zeno-corpus
 - Heldenthat → Heldentat
 - schwüle → Schwule



etraces.e-
humanities.net

Textmining

- Creating an inverted list
- Deconstruct sentences into words
- Deselecting most frequent features
- Looking where words appear in the corpus
- Including lemmas, cohyponyms synonyms
- Score results with quadratic distance measure

1.) Re-use strategy

2.) Found re-uses

Show text re-use from author and work:

Author:
Luther, Martin

Work:
Luther-Bibel 1912

Remove

Add another author and work

Select a strategy to find text re-use

Preprocessing

☒ Lemmatisation

☒ To Lower Case

☒ Cohyponym

☒ Synonyms

Please select how many features you want to use to find text re-use

Selecting

70%

Finish

The Tool GERTRUDE:

Göttingen E-Research Text Re-Use for Digital Editions

You are logged in as: ckoetteritzsch

Logout

View/Change User informations

Author:

Schiller, Friedrich

Work:

Die Räuber

search

Go

1 wollten. Noch so viele Freunde der Wahrheit mögen zusammenstehen, ihren Mitbürgern auf
2 Kanzel und Schaubühne Schule zu halten, der Pöbel hört nie auf, Pöbel zu sein, und wenn
3 Sonne und Mond sich wandeln und Himmel und Erde veralten wie ein Kleid. Vielleicht hätte
4 ich, den Schwachherzigen zu frommen, der Natur minder getreu sein sollen; aber wenn jener
5 Käfer, den wir alle kennen, auch den Mist aus den Perlen stört, wenn man Exempel hat, daß
6 Feuer verbrannt und Wasser ersäuft habe, soll darum Perle – Feuer – und Wasser konfisziert
7 werden?

8 Ich darf meiner Schrift zufolge ihrer merkwürdigen Katastrophe mit Recht einen Platz unter
9 den moralischen Büchern versprechen; das Laster nimmt den Ausgang, der seiner würdig ist.
10 Der Verirrte tritt wieder in das Geleise der Gesetze. Die Tugend geht siegend davon. Wer nur
11 so billig gegen mich handelt, mich ganz zu lesen, mich verstehen zu wollen, von dem kann ich
12 erwarten, daß er – nicht den Dichter bewundere, aber den rechtschaffenen Mann in mir
13 hochschätze.

14 Geschrieben in der Ostermesse. 1781.

15 Der Herausgeber.

16 **Personen.**

17 *Maximilian, regierender Graf von Moor. Karl, Franz, seine Söhne. Amalia von Edlreich.*
18 *Spiegelberg, Schweizer, Grimm, Razmann, Schuferte, Roller, Kosinsky, Schwarz, Libertiner,*
19 *nachher Banditen. Hermann, Bastard von einem Edelmann. Daniel, Hausknecht des Grafen von*
20 *Moor. Pastor Moser. Ein Pater. Räuberbande. Nebenpersonen.*

21 **Erste Szene**

22 *Franken. Saal im Moorischen Schloß.*
23 *Franz. Der alte Moor.*
24 **FRANZ.**
25 *Aber ist Euch auch wohl, Vater? Ihr seht so blaß.*
26 **DER ALTE MOOR.**
27 *Ganz wohl, mein Sohn – was hattest du mir zu sagen?*
28 **FRANZ.**
29 *Die Post ist angekommen – ein Brief von unserm Korrespondenten in Leipzig –*
30 **DER ALTE MOOR**
31 *begierig.*
32 *Nachrichten von meinem Sohne Karl?*

1.) Re-use strategy 2.) Found re-uses

Please select a page (Navigate on the left):

Process this page

Currently Processed processCurrent

Re-use found on this page:

Wenn Ihr krank seid – nur die leiseste Ahndung habt, es zu werden, so laßt mich – ich will zu gelegener Zeit zu Euch reden	11	rejected(0) candidates(11)
Ganz wohl, mein Sohn – was hattest du mir zu sagen	8	confirmed(0) rejected(0) candidates(8)
Aber ist Euch auch wohl, Vater	1	confirmed(0) rejected(0) candidates(1)
Wer nur so billig gegen mich handelt, mich ganz zu lesen, mich verstehen zu wollen, von dem kann ich erwarten, daß er – nicht den Dichter bewundere, aber den rechtschaffenen Mann in mir hochschätze	92	confirmed(0) rejected(0) candidates(92)
Ich darf meiner Schrift zufolge ihrer merkwürdigen Katastrophe mit Recht einen Platz unter den moralischen Büchern versprechen; das Laster nimmt den Ausgang, der seiner würdig ist	1	confirmed(0) rejected(0) candidates(1)
Vielleicht hätte ich, den Schwachherzigen zu frommen, der Natur minder getreu sein sollen; aber wenn jener Käfer, den wir alle kennen, auch den Mist aus den Perlen stört, wenn man Exempel hat, daß Feuer verbrannt und Wasser ersäuft habe, soll darum Perle – Feuer – und Wasser konfisziert werden	78	confirmed(0) rejected(0) candidates(78)
Noch so viele Freunde der Wahrheit mögen zusammenstehen, ihren Mitbürgern auf Kanzel und Schaubühne Schule zu halten, der Pöbel hört nie auf, Pöbel zu sein, und wenn Sonne und Mond sich wandeln und Himmel und Erde veralten wie ein Kleid	7	confirmed(0) rejected(0) candidates(7)

Select

You are logged in as: **ckoetteritzsch**

Logout

View/Change User Informations

Author:

Schiller, Friedrich

Work:

Die Räuber

search

Go

1 wollten. Noch so viele Freunde der Wahrheit mögen zusammenstehen, ihren Mitbürgern auf
2 Kanzel und Schaubühne Schule zu halten, der Pöbel hört nie auf, Pöbel zu sein, und wenn
3 Sonne und Mond sich wandeln und Himmel und Erde veralten wie ein Kleid. Vielleicht hätt
4 ich, den Schwacherzigen zu frommen, der Natur minder getreu sein sollen; aber wenn jener
5 Käfer, den wir alle kennen, auch den Mist aus den Perlen stört, wenn man Exempel hat, daß
6 Feuer verbrannt und Wasser ersäuft habe, soll darum Perle – Feuer – und Wasser konfisziert
7 werden?

8 Ich darf meiner Schrift zufolge ihrer merkwürdigen Katastrophe mit Recht einen Platz unter
9 den moralischen Büchern versprechen; das Laster nimmt den Ausgang, der seiner würdig ist.
10 Der Verirrte tritt wieder in das Geleise der Gesetze. Die Tugend geht siegend davon. Wer nur
11 so billig gegen mich handelt, mich ganz zu lesen, mich verstehen zu wollen, von dem kann ich
12 erwarten, daß er – nicht den Dichter bewundere, aber den rechtschaffenen Mann in mir
13 hochschätze.

14 Geschrieben in der Ostermesse. 1781.

15 Der Herausgeber.

16 Personen.

17 Maximilian, regierender Graf von Moor. Karl, Franz, seine Söhne. Amalia von Edelfreich.
18 Spiegelberg, Schweizer, Grimm, Razmann, Schusterle, Roller, Kosinsky, Schwarz, Libertiner,
19 nachher Banditen. Hermann, Bastard von einem Edelmann. Daniel, Hausknecht des Grafen von
20 Moor. Pastor Moser. Ein Pater. Räuberbande. Nebenpersonen.

21 Erste Szene

22 Franken. Saal im Moorischen Schloß.

23 Franz. Der alte Moor.

24 FRANZ.

25 Aber ist Euch auch wohl, Vater? Ihr seht so blaß.

26 DER ALTE MOOR.

27 Ganz wohl, mein Sohn – was hattest du mir zu sagen?

28 FRANZ.

29 Die Post ist angekommen – ein Brief von unserm Korrespondenten in Leipzig –

30 DER ALTE MOOR

31 begierig.

32 Nachrichten von meinem Sohne Karl?

1.) Re-use strategy

2.) Found re-uses

3.) All edges

Automatically found connections

Author Name	Work Name	Text passage	Status
Luther, Martin	Luther-Bibel 1912	28 So versammelt nun vor mich alle Ältesten eurer Stämme und eure Amtleute, daß ich diese Worte vor ihnen Ohren rede und Himmel und Erde wider sie zu Zeugen nehme.	candidate/automatic
Luther, Martin	Luther-Bibel 1912	12 Und da sie ihre Augen aufhoben von ferne, kannten sie ihn nicht und hoben auf ihre Stimme und weinten, und ein jeglicher zerriß sein Kleid und sie sprengten Erde auf ihr Haupt gen Himmel.	candidate/automatic
Luther, Martin	Luther-Bibel 1912	3 und sage also: Daß eine ewige Gnade wird aufgehen, und du wirst deine Wahrheit treulich halten im Himmel.	candidate/automatic
Luther, Martin	Luther-Bibel 1912	6 Hebet eure Augen auf gen Himmel und schauet unten auf die Erde, Denn der Himmel wird wie ein Rauch vergehen und die Erde wie ein Kleid veralten, und die darauf wohnen, werden im Nu dahinsterven. Aber mein Heil bleibt ewig und meine Gerechtigkeit wird kein Ende haben.	candidate/automatic
Luther, Martin	Luther-Bibel 1912	24 Aber zu der Zeit, nach dieser Trübsal, werden Sonne und Mond ihren Schein verlieren,	candidate/automatic
Luther, Martin	Luther-Bibel 1912	1 Und es erschien ein großes Zeichen im Himmel, ein Weib, mit der Sonne bekleidet, und der Mond unter ihren Füßen und auf ihrem Haupt eine Krone mit zwölf goldenen	candidate/automatic

Select this re-use

or

Add connection manually

or

remove connection

Logout
View/Change User Informations

Author:
Schiller, Friedrich

Work:
Die Räuber

search

Go

1 wollten. Noch so viele Freunde der Wahrheit mögen zusammenstehen, ihren Mitbürgern an
2 Kanzel und Schaubühne Schule zu halten, der Pöbel hört nie auf, Pöbel zu sein, und wenn
3 Sonne und Mond sich wandeln und Himmel und Erde veralten wie ein Kleid. Vielleicht hätte
4 ich, den Schwacherzigen zu frommen, der Natur minder getreu sein sollen; aber wenn jene
5 Käfer, den wir alle kennen, auch den Mist aus den Perlen stört, wenn man Exempel hat, daß
6 Feuer verbrannt und Wasser ersäuft habe, soll darum Perle – Feuer – und Wasser konfisziert
7 werden?

8 Ich darf meiner Schrift zufolge ihrer merkwürdigen Katastrophe mit Recht einen Platz unter
9 den moralischen Büchern versprechen; das Laster nimmt den Ausgang, der seiner würdig ist.
10 Der Verirrte tritt wieder in das Geleise der Gesetze. Die Tugend geht siegend davon. Wer nun
11 so billig gegen mich handelt, mich ganz zu lesen, mich verstehen zu wollen, von dem kann ich
12 erwarten, daß er – nicht den Dichter bewundere, aber den rechtschaffenen Mann in mir
13 hochschätze.

14 Geschrieben in der Ostermesse. 1781.

15 Der Herausgeber.

16 **Personen.**

17 *Maximilian, regierender Graf von Moor. Karl, Franz, seine Söhne. Amalia von Edelfreich.*
18 *Spiegelberg, Schweizer, Grimm, Razmann, Schusterle, Roller, Kosinsky, Schwarz, Libertiner,*
19 *nachher Banditen. Hermann, Bastard von einem Edelmann. Daniel, Hausknecht des Grafen von*
20 *Moor. Pastor Moser. Ein Pater. Räuberbande. Nebenpersonen.*

21 **Erste Szene**

22 *Franken. Saal im Moorischen Schloß.*
23 *Franz. Der alte Moor.*
24 **FRANZ.**
25 *Aber ist Euch auch wohl, Vater? Ihr seht so blaß.*
26 **DER ALTE MOOR.**
27 *Ganz wohl, mein Sohn – was hattest du mir zu sagen?*
28 **FRANZ.**
29 *Die Post ist angekommen – ein Brief von unserm Korrespondenten in Leipzig –*
30 **DER ALTE MOOR**
31 *begierig.*
32 *Nachrichten von meinem Sohne Karl?*

1

Genesis 1

1 Am Anfang schuf Gott Himmel und Erde.

2 Und die Erde war wüst und leer, und es war finster auf der Tiefe; und der Geist Gottes
schwebte auf dem Wasser.

3 Und Gott sprach: Es werde Licht! und es ward Licht.

4 Und Gott sah, daß das Licht gut war. Da schied Gott das Licht von der Finsternis

5 und nannte das Licht Tag und die Finsternis Nacht. Da ward aus Abend und Morgen der
erste Tag.

6 Und Gott sprach: Es werde eine Feste zwischen den Wassern, und die sei ein Unterschied
zwischen den Wassern.

Which of those two works is the older one?

How similar do these two text passages look like syntactically?

0 100

Does the connection between the two chosen text passages seem relevant to you?

How much do you think the author intended this textual relation (to be realized by the reader)?

☐ I believe the older text passage to be the (only) source of the later text passage.

Did you as a reader realize there was some relation to another text?

Commentary:

Please comment on:

- form of re-use (quote, paraphrase, allusion, ...)
- function of re-use (affirmative, negative, well-known, ...)
- how can this re-use be recognized (marked, interpreted, ...)
- used secondary literature
- other

Please do not use the '#' in your comment

Goal: Creating a good „prototype“

- Let users evaluate automatically found results and add their own results manually via Crowd-Sourcing
 - Free access for any user
 - University courses
 - Compare those results to:
 - Determine serendipity effect
 - Determine usability and user-behaviour
 - Improve Text-Mining-Algorithms
-

Further Work

- Integration of the Text-Mining-functions into existing annotation environments / digital editions / ...
 - Possible Enhancements/Addenda/Add:
 - Other texts
 - Other languages (one day maybe interlingual)
 - Visualisations
 - Additional tools (i.e. debugger to explain Text-Mining-steps to the user)
-



etraces.e-
humanities.net



Thank you for your attention!

Web-Debugger

- Screenshot



▣ Step 0: Searching ?

▣ Step 1: Preprocessing ?

▣ Step 2: Featuring ?

▣ Step 3: Selecting ?

Please select a selecting strategy: 02-01-01-01-GlobalMaxFeatureFrequencySelectorImpl:FeatDens=0.8

Agenda

word = This word belongs to the fingerprint

word = This word originally doesn't belong to the fingerprint but was selected by the user to belong to the fingerprint

word = This word doesn't belong to the fingerprint

word = This word originally belonged to the fingerprint but was selected by the user to not belong to the fingerprint

initial configuration: in the beginning god created the heavens and the earth

current configuration: in the beginning god created the heavens and the earth

selected features

in
beginning
god
created
heavens
earth

<-->

not selected features

the
the
and
the

Other users preference

Feature	users selected	users not selected
in	0	4
the	0	4
beginning	4	0
god	4	0
created	4	0
the	0	4
heavens	4	0
and	0	4
the	0	4
earth	4	0

Statistics

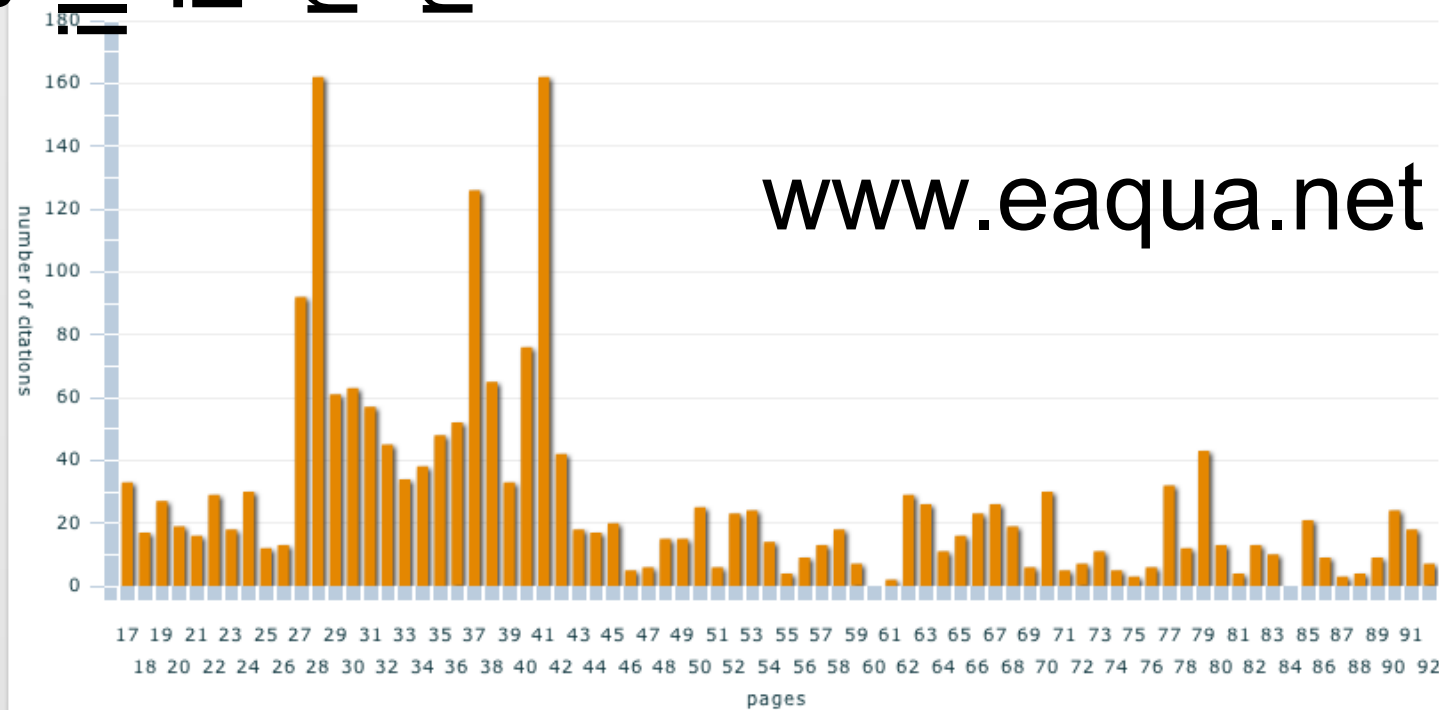
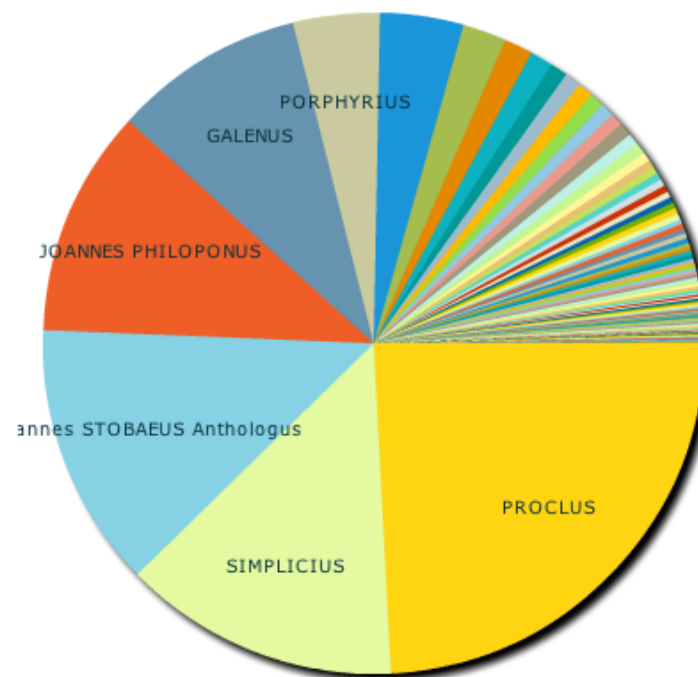
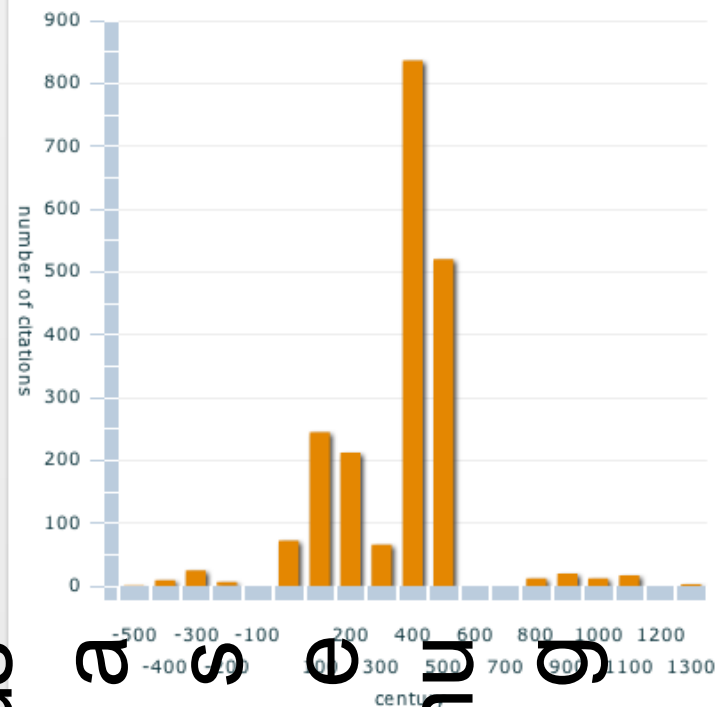
Feature	Selected Features	Total number of features
in	216	240
the	0	1822
beginning	6	6
god	320	382
created	25	35
the	0	1822
heavens	69	75
and	0	1706
the	0	1822
earth	159	215

▣ Step 4: Linking ?

▣ Step 5: Scoring ?

▣ Step 6: Postprocessing(Not available yet) ?

century range:



www.eaqua.net