



**HAL**  
open science

# Identifying forms of intertextuality combining Text-Mining and Crowd-Sourcing with the tool GERTRUDE in the eTRACES project

Annette Gessner

► **To cite this version:**

Annette Gessner. Identifying forms of intertextuality combining Text-Mining and Crowd-Sourcing with the tool GERTRUDE in the eTRACES project. 2016. hal-01283673

**HAL Id: hal-01283673**

**<https://hal.science/hal-01283673>**

Preprint submitted on 5 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Identifying forms of intertextuality combining Text-Mining and Crowd-Sourcing with the tool GERTRUDE in the eTRACES project

Annette Geßner<sup>1\*</sup>

1 University of Leipzig, Germany

\*Corresponding author: Annette Geßner [agessner@e-humanities.net](mailto:agessner@e-humanities.net)

## Abstract

As part of the eTRACES project, the Göttingen sub-project focused on detecting various forms of intertextuality in a German text corpus, the texts found at [zeno.org](http://zeno.org), and creating a user interface to apply crowd-sourcing methods to evaluate the results.

## keywords

Text Re-Use, Visualization

## INTRODUCTION

Intertextuality is mostly understood as a way of referencing a former text in many various ways like citing, quoting, alluding and many more and has always been an important part of literature and literary studies.

## I THE eTRACES PROJECT

The BMBF (Bundesministerium für Bildung und Wissenschaft) funded the eTRACES project from 2011 to 2014 to further support close collaboration between humanists and computer scientists in working with text corpora in the field of Digital Humanities or e-humanities.

Following the concept of the eAQUA project, eTRACES had three sub-projects all centred around text re-use and the cooperation of humanities and computer science.

The Göttingen sub-project focused on detecting textual re-use in the text corpus of German literature which can be found on the web-page [zeno.org](http://zeno.org) and explored the possibilities and limitations of determining as many different kinds of intertextuality as possible. Doing so it also implemented a crowd-sourcing approach to improve the results and prove the diversity of the reception of different forms of intertextuality.

The use-case was to find biblical intertextuality in Weltliteratur, namely Goethe's "Die Leiden des jungen Werther" and Schiller's "Die Räuber", since those two authors are known to use various forms of intertextuality to refer to the Bible. Since intertextuality is very heterogeneous and thus detecting and describing all possible forms of it is a very difficult task even for a trained humanist, the sub project focused on the combination of two aspects: On the one hand, different Text Mining approaches were developed to detect pattern-based text re-use on the grounds of text similarity. To do so the corpus was preprocessed, segmented, an inverted list and the possibility to reference single passages was created before looking for similarities of words in sentences of a text, which could be refined by the user, as further explained in [Geßner et al. 2013].

On the other hand, a virtual environment called GERTRUDE (Göttingen E-Research Text Re-Use in Digital Editions) was created, allowing the user to change refine those algorithms, show the found results and evaluate them independently from other users, so that the individual results could be compared afterwards. Thus a fruitful symbiosis of Text-Mining and crowd-sourcing was created.

## II Visualisation and crowd-sourcing

Exploring the many different facets of intertextuality, the focus went to creating an online interface, where different factors about found re-use could be documented, and letting as many people as possible do this, not influenced by other users results (crowd-sourcing), to compare different results and explore the understanding of intertextuality.

In this interface every user can log in and individually choose a work, refine the re-use algorithms, look at the results found automatically, evaluate and correct them as well as add new results. Those user results can be compared afterwards, to see how differently forms of intertextuality can be perceived. A visualization for comparison has unfortunately not been implemented.

You are logged in as: **ckoetteritzsch**

Logout

View/Change User Informations

Author:  
Schiller, Friedrich

Work:  
Die Räuber

search

1 wollen. Noch so viele Fremde der Welt hat mögen zusammenstehen, ihren Mühsert auf  
2 sangen und Schaubühne Schule zu halten, der Pöbel hat nie a 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 Edle reich.  
18 Spiegelberg, Schweizer, Grimm, Razmann, Schufertle, 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 Aeltesten eurer Stämme und eure Antheute, daß ich diese Worte vor Ohren rede und 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 zeriß sein Haupt gen auf ihr Haupt gen	candidate/automatic
Luther, Martin	Luther-Bibel 1912	3 und sage also: Daß eine ewige Gnade wird aufgehen, und du wirst deine treulich im	candidate/automatic
Luther, Martin	Luther-Bibel 1912	6 Hebet eure Augen auf gen Himmel und schauet unten auf die Erde. Denn der Himmel wie ein Rauch vergehet und die Erde wie ein Kleid veralten, und die darauf wohnen, werden im Nu dahinstreben. Aber mein Heil bleibt ewiglich und meine Gerechtigkeit wird kein Ende haben.	candidate/automatic
Luther, Martin	Luther-Bibel 1912	24 Aber zu der Zeit, nach dieser Trübsal, werden und Schein verlieren,	candidate/automatic
Luther, Martin	Luther-Bibel 101?	1 Und es erschien ein großes Zeichen im ein Weib, mit der bekleidet, und der unter Füssen und auf ihrem Haupt eine Krone mit zwölf	candidate/automatic

Select this re-use or Add connection manually or remove connection

Figure 1. GERTRUDE interface, with found re-use passages (right) to highlighted passage (left).

Crucial part the interface was the approach to create a user-friendly yet helping retrieve meaningful results evaluation form to track differences in different forms of intertextuality, covering information from syntactic similarity of the text passages, to the different question of the author's intention according to the users impression, to the question, if the user realized this being a form of intertextuality without help.

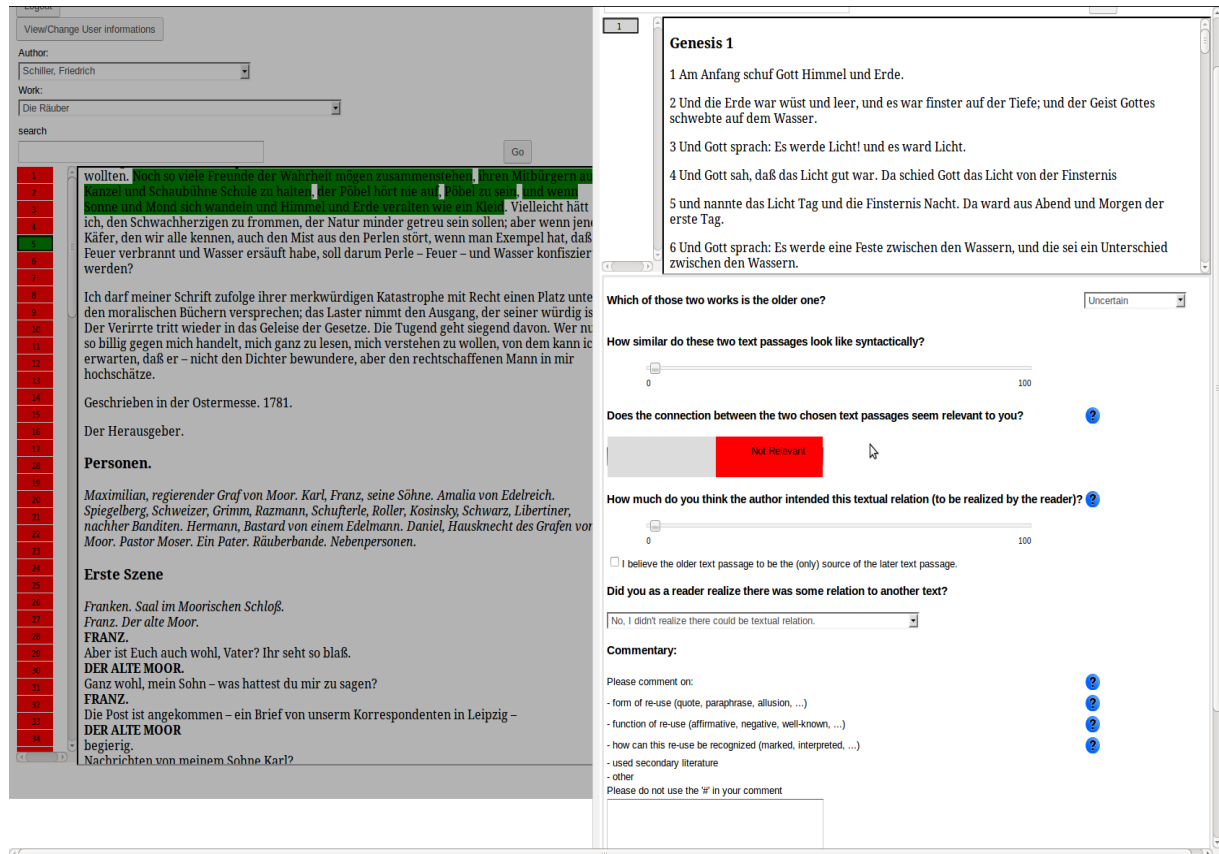


Figure 2. Evaluation of a re-use result.

## IV DISCUSSION AND CONCLUSION

The text mining approach broadened the possibilities that were implemented in the eAQUA project. By adding a crowd-sourcing element it enabled to refine the algorithms and further investigate the nature of different forms of intertextuality by comparing different users evaluations of re-use results found automatically and manually.

Including the results of as many users as possible for different research questions is one of the big goals in Digital Humanities by getting different people interested in participating in projects like this. We may not be there yet, it was rather difficult finding students interested in this kind of activity during the project, but there's hope approaches like this will become more often and more well known in the future.

The question remains, whether ALL possible forms of intertextuality can one day be determined automatically (or even manually). If yes, probably only by including all kinds of user-generated results and as many technical approaches and possibilities to refine as possible or useful. This will be a long way, probably improved by working on more mining-approaches, a good tutorial for the users, a visualization for comparison of user results and more.

## References

- Geßner, A., Kötteritzsch, C., Lauer, G.: Biblical intertextuality in a digital world: the tool GERTRUDE: In: *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities (DH-CASE '13)*. ACM, New York, NY, USA, 2013.
- Geßner, A.: Überlegungen zur computergestützten Auffindung und Bewertung von textuellen Übereinstimmungen in den Projekten eAQUA und eTRACES. In: Valente, Stefano; Koch, Lutz; Deckers, Daniel; Herausgeber: Brockmann, Christian: *Handschriften- und Textforschung heute - Zur Überlieferung der griechischen Literatur. Festschrift für Dieter Harlfinger aus Anlass seines 70. Geburtstages*, Wiesbaden, 2014.