



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^{1*}

¹ University of Leipzig, Germany

*Corresponding author: Annette Geßner 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, 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 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

Go

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

wollten. Noch so viele Freunde der Wahrheit mögen zusammenstehen, ihren Mitbürgern auf
Kanzel und Schaubühne Schule zu halten, der Pöbel hört nie auf, über sie zu schimpfen, und wenn
Sonne und Mond sich wandeln und Himmel und Erde verfallen wie ein Kleid. Vielleicht hätt
ich, den Schwacherzigen 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?

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.
Der Verrirte tritt wieder in das Geleise der Gesetze. Die Tugend geht siegend davon. 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.

Geschrieben in der Ostermesse. 1781.

Der Herausgeber.

Personen.

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

Erste Szene

Franken. Saal im Moorischen Schloß.
Franz. Der alte Moor.
FRANZ.
Aber ist Euch auch wohl, Vater? Ihr seht so blaß.
DER ALTE MOOR.
Ganz wohl, mein Sohn – was hattest du mir zu sagen?
FRANZ.
Die Post ist angekommen – ein Brief von unserm Korrespondenten in Leipzig –
DER ALTE MOOR
begierig.
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 Amteute, daß ich diese Worte vor [redacted] Ohren rede und [redacted] und [redacted] 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 zernüß sein [redacted] und sie sprengten [redacted] auf ihr Haupt gen [redacted]	candidate/automatic
Luther, Martin	Luther-Bibel 1912	3 und sage also: Daß eine ewige Gnade wird aufgehen, und du wirst deine [redacted] treulich [redacted] im [redacted]	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 verfallen, und die darauf wohnen, werden im Nu dahinstorben. 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 [redacted] und [redacted] Schein verlieren,	candidate/automatic
Luther, Martin	Luther-Bibel 1912	1 Und es erschien ein großes Zeichen im [redacted] ein Weib, mit der [redacted] bekleidet, und der [redacted] unter [redacted] Füßen und auf ihrem Heude eine Krone mit zwölf reidenden	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.

The interface is a web application for evaluating re-use results. It consists of two main panels. The left panel is a search and list interface, and the right panel is an evaluation form.

Left Panel (Search and List):

- Search Bar:** A text input field with a "Go" button.
- Author:** A dropdown menu with "Schiller, Friedrich" selected.
- Work:** A dropdown menu with "Die Räuber" selected.
- List of Text Passages:** A list of text passages with a search bar and a "Go" button. The list includes:
 - 1. wollen. Nach so viele Freunde der Wahrheit haben zusammengehört, ihren Mühen...
 - 2. Ich darf meiner Schrift zufolge ihrer merkwürdigen Katastrophe mit Recht einen Platz unter...
 - 3. Geschrieben in der Ostermesse. 1781.
 - 4. Der Herausgeber.
 - 5. **Personen.**
 - 6. Maximilian, regierender Graf von Moor. Karl, Franz, seine Söhne. Amalia von Edereich.
 - 7. **Erste Szene**
 - 8. Franken. Saal im Moorischen Schloß.
 - 9. Franz. Der alte Moor.
 - 10. FRANZ.
 - 11. Aber ist Euch auch wohl, Vater? Ihr seht so blaß.
 - 12. DER ALTE MOOR.
 - 13. Ganz wohl, mein Sohn – was hattest du mir zu sagen?
 - 14. FRANZ.
 - 15. Die Post ist angekommen – ein Brief von unserm Korrespondenten in Leipzig –
 - 16. DER ALTE MOOR
 - 17. begierig.
 - 18. Nachrichten von meinem Sohne Karl?

Right Panel (Evaluation Form):

- 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?** (Dropdown menu: Uncertain)
- How similar do these two text passages look like syntactically?** (Slider: 0 to 100)
- Does the connection between the two chosen text passages seem relevant to you?** (Buttons: Not Relevant, Relevant)
- How much do you think the author intended this textual relation (to be realized by the reader)?** (Slider: 0 to 100)
- ☐ 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?** (Dropdown menu: No, I didn't realize there could be textual relation.)
- 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 "w" in your comment

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.