

# “The Bard meets the Doctor” – Computergestützte Identifikation intertextueller Shakespearebezüge in der Science Fiction-Serie Dr. Who

**Burghardt, Manuel**

burghardt@informatik.uni-leipzig.de  
Universität Leipzig, Deutschland

**Meyer, Selina**

Selina.Meyer@stud.uni-regensburg.de  
Universität Regensburg, Deutschland

**Schmidtbauer, Stephanie**

Stephanie.Schmidtbauer@stud.uni-regensburg.de  
Universität Regensburg, Deutschland

**Molz, Johannes**

johannes.molz@googlemail.com  
Ludwig-Maximilians-Universität München, Deutschland

## Einführung: Shakespeare und Intertextualität

In der jüngeren Literatur- und Kulturtheorie geht man davon aus, dass alle literarischen Texte immer auch durch eine reiche Tradition, ja ein ganzes Ökosystem, bestehender Literatur beeinflusst sind (Allen, 2000, S. 1). Dieser Einfluss, der sich im Text sowohl in impliziten als auch in expliziten Querverbindungen durch Zitate offenbart, wird gemeinhin als Intertextualität bezeichnet<sup>1</sup>. Als eine der zentralen Kulturtechniken der Postmoderne, lässt sich das Zitat an keinem Autor so gut studieren wie an William Shakespeare, dessen intertextuelles Nachwirken die gesamte westliche Kulturhemisphäre durchzieht (Maxwell & Rumbold, 2018, S. 1). Eine Untersuchung dieser allgegenwärtigen Spuren ist somit zwingend auch eine Geschichte unserer gesamten Kultur (Taylor, 1991) und beantwortet und eröffnet Fragen, die weit über Shakespeare hinausgehen. Dabei finden sich intertextuelle Bezüge auf das Werk Shakespeares nicht nur in unterschiedlichsten literarischen Genres, sondern in zunehmendem Maße auch im Massenmedium Film und Fernsehen (Malone, 2018).

Wenngleich Film in erster Linie ein visuelles Medium ist, so bietet sich über die Dialoge zusätzlich ein

verbaler Analysezugang (vgl. Kozloff, 2000). Klarer (1998, S. 54) spricht in dem Zusammenhang gar von Film als semi-textuelles Genre. Die den Filmen zugrunde liegenden Skripte lassen sich als Dramen lesen, da sie ausschließlich aus Figurennamen, Sprechakten und Bühnenanweisungen bestehen. Für die quantitative Analyse von Intertextualitätsphänomenen bei Filmen liegt ein großer Vorteil in der Verfügbarkeit vollständig transkribierter Dialoge die als Untertitel<sup>2</sup>, Drehbücher<sup>3</sup> oder Fan-Transkripte<sup>4</sup> über diverse Online-Portale angeboten werden. Die maschinenlesbaren Dialoge können sodann mit bestehenden Methoden und Algorithmen aus dem informatischen Anwendungsbereich der *text reuse detection*, also der Identifikation einer Wiederverwendung bestimmter Textfragmente in anderen Texten, eingesetzt werden. Verwandte Arbeiten zur computergestützten Intertextualitätsforschung finden sich vor allem im Bereich historischer Sprachen, insbesondere Latein und Griechisch (vgl. Berti et al., 2013; Büchler et al. 2014; Scheirer et al., 2014; Forstall et al., 2015; Bamman & Crane, 2018), weniger im Bereich der Anglistik und fast gar nicht in der Shakespeareforschung. Wenngleich das Thema umfangreich mit qualitativ-hermeneutischen Methoden untersucht wurde und wird (vgl. etwa den Sammelband von Maxwell & Rumbold, 2018), so finden sich nur wenige Arbeiten, die quantitative, digitale Methoden einsetzen, bspw. das Hyper-Hamlet-Projekt<sup>5</sup>, bei dem eine große elektronische Datenbank mit Hamlet-Referenzen aufgebaut wurde (Hohl Trillini & Quassdorf, 2010).

Vor diesem Hintergrund exploriert der vorliegende Beitrag intertextuelle Bezüge auf das Werk Shakespeares in der britischen Fernsehserie Dr. Who und setzt dabei auf computergestützte Analysemethoden, die bislang vor allem im Bereich klassischer Altertumswissenschaften und historischer Sprachen eingesetzt wurden. Dr. Who eignet sich dabei in besonderer Weise, da die TV-Serie fester Bestandteil der britischen Kultur ist, was automatisch eine größere Nähe zum Werk ihres Landsmannes Shakespeare bedeutet. Andererseits gibt es einige Folgen bei Dr. Who, bei denen bereits im Titel explizite Shakespeare-Bezüge deutlich werden<sup>6</sup>, was auf die Identifikation weiterer Zitate und Referenzen in anderen Folgen hoffen lässt. Unser Beitrag versteht sich im Wesentlichen als Fallstudie anhand derer überprüft werden soll, ob sich Standard-Methoden der *text reuse detection* prinzipiell auch für eine vergleichende Analyse von Shakespeare und TV-Serien eignen. Gleichzeitig soll am Beispiel von Dr. Who exemplarisch untersucht werden, welche Art und Menge von intertextuellen Shakespeare-Bezügen im Kontext von TV-Serien zu erwarten sind, um darauf aufbauend den computergestützten Ansatz zu optimieren und weitere Filme und Serien zu analysieren. Übergeordnetes und langfristiges Ziel ist es demnach, systematisch die Referenzierung Shakespeares in Film und Fernsehen mithilfe computergestützter Methoden zu erfassen und so die intertextuelle und intermediale Durchdringung von

Shakespeares Werk über die klassischen Literaturgenres hinaus zu quantifizieren und zu kartieren.

## Korpora: Shakespeare vs. Dr. Who

Im Rahmen unserer Fallstudie werden systematisch Referenzen aus zehn der bekanntesten Shakespeare-Stücke (*Hamlet*, *Macbeth*, *Merchant of Venice*, *A Midsummer Night's Dream*, *Much Ado About Nothing*, *Julius Caesar*, *The Tempest*, *Romeo and Juliet*, *King Lear*, *Othello*) sowie aus 154 kurzen Sonetten identifiziert werden. Die Transkripte dieser Werke stammen von der Website *Open Source Shakespeare*<sup>7</sup> und umfassen etwa 244.000 Tokens. Gesucht werden die Shakespearebezüge in einem Vergleichskorpus, welches aus den transkribierten Dialogen der TV-Serie Dr. Who besteht. Doctor Who ist eine der populärsten Science Fiction-Serien der britischen Geschichte und wird seit 1969 vom BBC produziert. Mittlerweile umfasst die Serie 840 Folgen, verteilt auf über 36 Staffeln und 13 verschiedene Doktoren. Für die vorliegende Studie wurden die Serientranskripte der ersten fünf Doktoren analysiert. Die Transkripte wurden von der Fan-Seite *Chakoteya*<sup>8</sup> heruntergeladen und umfassen insgesamt 141 Folgen mit einem Gesamtumfang von etwa zwei Millionen Tokens. Die Transkripte enthalten einerseits die Dialoge und andererseits Metadaten zu den jeweiligen Szenen, die vergleichbar mit Bühnenanweisungen bei Dramen sind.

## Methodik: Suche nach local alignments und keywords

Für die Identifikation von textuellen Übereinstimmungen und Textähnlichkeiten (*text reuse*) findet sich ein breites Spektrum an computergestützten Methoden, die im Überblickspapier von Bär et al. (2012) grundlegend hinsichtlich (1) inhaltlicher, (2) struktureller und (3) stilistischer Ähnlichkeit kategorisiert werden. In dieser Studie verwenden wir den Smith-Waterman-Algorithmus (Smith & Waterman, 1981), ein Standard-Verfahren zur Identifikation von *local alignments*, also textuellen Übereinstimmungen auf der Inhaltsebene, das u.a. auch im Bereich der Bioinformatik für die Analyse von Genomsequenzen eingesetzt wird. Dabei werden auch graduelle Unschärfen wie etwa *gaps*, also einzelne fehlende Worte in ansonsten größtenteils identischen Wortgruppen, berücksichtigt. Zur Analyse der Korpora wurden die R-Packages *TextMining*<sup>9</sup> für das *Preprocessing* der Daten und *TextReuse*<sup>10</sup> als Implementierung für den Smith-Waterman-Algorithmus verwendet. Beim *Preprocessing* wurde schnell deutlich, dass bspw. eine Stopwortentfernung nicht sinnvoll ist, da diese teilweise sehr charakteristisch für Shakespeare-Zitate sind (Bsp.: „to be or not to be“). Auch von Stemming und Lemmatisierung wurde abgesehen, da

dies zu sehr vielen falsch-positiven Treffern führt, die dann aufwendig durch manuelle Kontrolle ausgefiltert werden müssen.

Weiterhin wurde das Shakespeare-Korpus in jeweils überlappende 9-Gramme (*shingling*), also jeweils Gruppen von neun Wörtern zerlegt. Für jedes einzelne 9-Gramm wurde dann im Dr. Who-Korpus mithilfe des im TextReuse-Package implementierten Smith-Waterman-Algorithmus das optimale *local alignment*, also der beste Treffer ermittelt. Dabei ist erwähnenswert, dass für jedes einzelne Shakespeare-9-Gramm immer ein optimales *alignment* im Dr. Who-Vergleichskorpus berechnet wird, auch wenn dieses ggf. nur aus einem oder zwei gemeinsamen Wörtern besteht. Der Vorgang ist somit einerseits sehr rechenintensiv und liefert andererseits eine große Menge von *alignments*, deren ausschließlich manuelle Kontrolle enorm aufwendig wäre. Dabei ist offensichtlich, dass die Verwendung einzelner, noch dazu sehr allgemeiner Wörter (Artikel, Konjunktionen, etc.) längst keinen intertextuellen Bezug darstellt. Wir betrachten deshalb nur solche *alignments* näher, die aus mindestens drei oder mehr Wörtern bestehen.

Gleichzeitig sind allerdings sehr kurze Referenzen, bspw. die Nennung einer Figur wie „Othello“ oder „Macbeth“, nicht ausgeschlossen, gehen aber bei diesem Ansatz verloren. Wir ergänzen den *local alignment*-Ansatz deshalb um eine *keyword*-Suche, bei der wir eine manuell erstellte Liste mit etwa 140 Stichwörtern und sehr kurzen Phrasen verwendeten. Die Liste enthält alle *dramatis personae* mit eindeutigen Namen (bspw. Ophelia, Romeo, Caliban, etc.), die Titel aller Stücke und Apokryphen sowie allgemeine Referenzen zu Shakespeares Person und Biographie (bspw. Shakespeare, Shaksper, The Bard, Stratford upon Avon, Anne Hathaway, etc.).

## Ergebnisse und Diskussion

Insgesamt wurden mit der *keyword*-Suche 201 Treffer im Dr. Who-Korpus identifiziert, wobei ein Großteil der Treffer in einer einzelnen Folge vorkommen, da hier zwei der Hauptfiguren Troilus (108) und Cressida (37) heißen (vgl. Tabelle 1).

Beispielreferenz: “TROILUS: I'm sorry, **Cressida**, but I must obey orders.”

Bei genauerer Überprüfung der Treffer wird schnell klar, dass Figurennamen aus Shakespeares Historiendramen häufig auch in der ursprünglichen Bedeutung der zugrundeliegenden historischen Figur zitiert werden (bspw. *Horatio [Nelson]*, *King John*, *Julius Caesar*, *Pericles*) und somit keine genuine Shakespearereferenz vorliegt. Nach Entfernung dieser falsch-positiven Treffer verbleiben 185 echte Shakespearereferenzen. Hier muss allerdings angemerkt werden, dass die von uns verwendeten Transkripte neben den eigentlichen Dialogen auch umfangreiche Kommentare und Metadaten zu den Stücken enthalten, die häufig zusätzliche Shakespeare-Keywords enthalten. So kommt man ausschließlich über die Dialoge

auf 77 Treffer, in den Metadaten finden sich zusätzliche 108 Treffer.

Staffel	Keywords	Matches insgesamt	Treffer Dialoge	Treffer Metadaten	Alignments
Doctor 1	Troilus (108), Cressida (37), Shakespeare (13), Hamlet (5), Falstaff (2), Romeo and Juliet (1), Julius Caesar (1)	167	68	99	1
Doctor 2	Shakespeare (1), Richard III (1)	2	0	2	0
Doctor 3	<i>Horatio (2), King John (1)</i>	3	0	0	1
Doctor 4	Shakespeare (8), Hamlet (1), Henry V (1), Juliet (1), <i>Julius Caesar (1)</i> , Macbeth (1), Romeo (1), sonnets (1)	15	8	6	5
Doctor 5	<i>King John (11)</i> , Hamlet (1), Shakespeare (1), <i>Pericles (1)</i>	14	1	1	1
<b>Gesamt</b> (141 Einzelfolgen)		<b>201</b>	<b>77</b>	<b>108</b>	<b>8</b>

**Tabelle 1** : Überblick zu den gefundenen Keywords und *alignments* im Dr. Who-Korpus. Kursiv gesetzte Treffer sind falsch-Positive, d.h. nach manueller Prüfung stellte sich heraus, dass sie keine unmittelbare Shakespearereferenz beinhalten.

Auffällig niedrig scheint im Vergleich zu den *keyword*-Treffern die Anzahl längerer *alignments* zu sein: So finden sich insgesamt nur acht längere Shakespeare-Zitate in unserem Dr. Who-Korpus, bspw.:

(1) “Friends, Romans, countrymen! Lend me your ears. I come to bury Caesar, not to praise him.”

(Julius Caesar, Akt 3, Szene 2)

(2) “To be or not to be.” (Hamlet, Akt 3, Szene 1)

(3) “By the pricking of my thumbs, something wicked this way comes.” (Macbeth, Akt 4, Szene 1)

Gleichzeitig erhalten wir mit dem *local alignment*-Ansatz auch viele falsch-positive Treffer, die zwar längere, in beiden Texten vorkommende Sequenzen beschreiben, aber keine genuinen Shakespeare-Zitate sind, sondern eher hochfrequente Idiome, bspw.:

“You all know what you have to do.” (kommt gleichermaßen bei Shakespeare und Dr. Who vor)

## Methodenreflexion und Fazit

Der zweigleisige Ansatz einer *local alignment*-sowie einer gesonderten *keyword*-Suche bringt mit 193 intertextuellen Shakespearebezügen in unserem Dr. Who-Korpus eine beeindruckende Anzahl von Treffern, deren Identifikation manuell was Umfang und Arbeitsaufwand angeht, so nicht ohne Weiteres möglich gewesen wäre. Damit ergeben sich spannende Perspektiven für die Erweiterung des Ansatzes auf größere Korpora, mit weiteren Serien und Filmen. Gleichwohl muss angemerkt werden, dass der Löwenanteil der Treffer über den relativ trivialen Ansatz der *keyword*-Suche gefunden wurde. Von den insgesamt 193 Treffern stammen nur 8 aus dem *local alignment*-Ansatz. Bedenkt man den enormen Rechenaufwand (fast 200 Stunden mit Standard-Laptops) und den Aufwand der manuellen Filterung der *alignments*, so muss man den *local alignment*-Ansatz in dieser Form doch deutlich hinterfragen. Denkbar wären hier für künftige Experimente weitere Optimierungen auf

algorithmischer Seite, bspw. die Berücksichtigung von Wortarten-Ngrammen oder die Verwendung alternativer Algorithmen aus dem Bereich der *content similarity* (vgl. Bär et al., 2012). Weitere Verbesserungen sind auf Seite der Ergebnispräsentation (vgl. Burghardt & Wolff, 2015) möglich, welche die Beurteilung der Treffer deutlich erleichtern und damit beschleunigen könnte. Sinnvoll scheint auch eine Vorfilterung durch die *keyword*-Suche, d.h. nur diejenigen Episoden in denen bereits ein Shakespeare-*keyword* erkannt wurde werden anschließend auch noch mit dem *local alignment*-Ansatz auf komplexere Zitate hin untersucht. Als nächste Schritte sollen deshalb systematisch Korrelationen zwischen komplexen *alignments* und einzelnen *keywords* in einem größeren Korpus untersucht werden.

## Fußnoten

1. Für eine Einführung zu „Intertextualität“ siehe Allen (2000).
2. Untertitel online verfügbar unter <https://www.opensubtitles.org/de> (alle URLs in diesem Artikel wurden zuletzt am 14.10.2018 überprüft)
3. > Drehbücher online verfügbar unter <https://www.imsdb.com/>
4. Fan-Transkripte online verfügbar unter <http://transcripts.foreverdreaming.org/>, <https://www.springfieldspringfield.co.uk/>, <http://www.chakoteya.net/>
5. HyperHamlet online verfügbar unter <http://www.hyperhamlet.unibas.ch/>
6. Bspw. „The Shakespeare Code“, Dr. Who Folge 180, Transkription online verfügbar unter <http://www.chakoteya.net/DoctorWho/29-2.htm>
7. Open Source Shakespeare online verfügbar unter <https://www.opensourceshakespeare.org/>
8. Chakoteya Fan-Transkripte online verfügbar unter <http://www.chakoteya.net/DoctorWho/index.html>
9. Dokumentation online verfügbar unter <https://cran.r-project.org/web/packages/tm/index.html>
10. Dokumentation online verfügbar unter <https://cran.r-project.org/web/packages/textreue/index.html>

## Bibliographie

- Allen, G. (2000): *Intertextuality*, London, New York: Routledge.
- Bamman, D. / Crane, G. (2008) *The Logic and Discovery of Textual Allusion*, in: ACL Language Technology for Cultural Heritage, (1986).
- Bär, D. / Zesch, T. / Gurevych, I. (2012): *Text Reuse Detection using a Composition of Text Similarity Measures*, in: Proceedings of COLING 2012, 1 (December), 167–184.
- Berti, M. / Büchler, M. / Geßner, A. / Thomas, E. (2013): *Measuring the Influence of a Work by Text Reuse*, in: The Digital Classicist.

**Burghardt, M. / Wolff, C. (2015):** *Humanist-Computer Interaction: Herausforderungen für die Digital Humanities aus Perspektive der Medieninformatik*, in: Book of Abstracts Workshop “Informatik und die Digital Humanities”, Leipzig.

**Büchler, M. / Burns, P. R. / Müller, M. / Franzini, E. / Franzini, G. (2014):** *Towards a Historical Text Re-use Detection*, in **C. Biemann / A. Mehler (Eds.):** *Text Mining, Theory and Applications of Natural Language Processing*, Springer International Publishing.

**Forstall, C. / Coffee, N. / Buck, T. / Roache, K. / Jacobson, S. (2015):** *Modeling the scholars: Detecting intertextuality through enhanced word-level n-gram matching*, in: *Digital Scholarship in the Humanities*, 30 (4), 503–515.

**Hohl Trillini, R. / Quassdorf, S. (2010):** *A “key to all quotations”? A corpus-based parameter model of intertextuality*, in: *Literary and Linguistic Computing*, 25 (3), 269–286.

**Klarer, M. (1999):** *An Introduction to Literary Studies*, London, New York: Routledge.

**Kozloff, S. (2000):** *Overhearing Film Dialogue*, Berkeley et al.: University of California Press.

**Malone, T. (2018):** *Quoting Shakespeare in Twentieth-Century Film*, in: **Maxwell, J. / Rumbold, K. (Eds.):** *Shakespeare and Quotation* (pp. 194–208). Cambridge: Cambridge University Press.

**Maxwell, J. / Rumbold, K. (2018):** *Shakespeare and Quotation*, Cambridge: Cambridge University Press.

**Scheirer, W. / Forstall, C. / Coffee, N. (2016):** *The sense of a connection: Automatic tracing of intertextuality by meaning*, *Digital Scholarship in the Humanities*, 31 (1), 204–217.

**Smith, T. F. / Waterman, M. S. (1981):** *Identification of Common Molecular Subsequences*, *Journal of Molecular Biology*, 147, 195–197.

**Taylor, G. (1991):** *Reinventing Shakespeare: A Cultural History from the Restoration to the Present*, Oxford: Oxford University Press.