

Explicitly Notated Citations. Von digitalen zu algorithmischen Editionen

Enns, Sebastian

sebastian.enns@mni.thm.de
TH Mittelhessen, University of Applied Sciences,
Deutschland
ORCID: 0009-0007-6336-8266

Kuczera, Andreas

andreas.kuczera@mni.thm.de
TH Mittelhessen, University of Applied Sciences,
Deutschland
ORCID: 0000-0003-1020-507X

Problemstellung

Das Zitieren ist eine uralte menschliche Kulturtechnik, die bereits in den frühesten schriftlichen Aufzeichnungen wie den alten ägyptischen Hieroglyphen und sumerischen Keilschriften verwendet wurde. Es dient dazu, den Ursprung von Wissen nachzuweisen und die Weitergabe von Weisheit über Generationen hinweg zu ermöglichen. Auch in den heute immer digitaler werdenden Zeiten ist es wichtig, digitale Ressourcen zuverlässig zitierbar zu gestalten (Bleier 2021, 1). Eine der größten Herausforderungen ist hierbei die *flexible Granularität*, die aufgrund fehlender Kapitel-, Paragraphen- oder Zeilenangaben beim Zitieren nicht adäquat identifiziert werden kann (Bernhart/Hahn 2014, 227; Ralle 2016, 155). In analogen Editionen bezieht sich das Zitieren meist auf die genannten Angaben und erlaubt ebenfalls keine genaue und flexible Referenzierung bis auf das einzelne Zeichen (Bleier 2021, 6). Ebenso problematisch sind die *Beständigkeit* und *Zugänglichkeit* digitaler Ressourcen. Generelle URLs sind oft unzuverlässig, da sie sich ändern oder die hinterlegten Ressourcen nicht mehr verfügbar sein könnten (Bleier 2021, 4; Stronks/Boot 2007, 169). Daher gewinnen die Verwendung von Permalinks und persistenten Identifikatoren wie DOIs zunehmend an Bedeutung (Bleier 2021, 4). Allerdings erfolgt die Referenzierung der Permalinks meist auf statische Inhalte und berücksichtigt nicht die verschiedenen parametrisierten und dynamischen Ansichten eines referenzierten Textes in einer digitalen Edition (Stäcker 2020, 9). Darüber hinaus hängt die Zuverlässigkeit dieser Links stark von der verantwortlichen Institution ab, die sie pflegen und verwalten muss (Bleier 2021, 4).

Bleier (2021, 5) hebt hervor, dass die Versionierung digitaler Inhalte, die sich jederzeit ändern können, eine weitere zentrale Herausforderung für die Zitierfähigkeit digitaler Editionen darstellt. Aus seiner Sicht ist es daher von entscheidender Bedeutung, genau festzulegen, welche Version eines Textes zitiert wird, um Transparenz und Nachvollziehbarkeit in wissenschaftlichen Arbeiten zu gewährleisten. Borgman (2010) betont ebenfalls, dass Persistenz in einer digitalen, verteilten Welt problematisch ist, jedoch unabdingbar bleibt, um Publikationen, Daten und andere Quellen zuverlässig zitieren zu können (Bernhart/Hahn 2014, 227). Broyles (2020, 42) argumentiert, dass die Versionierung verschiedener Komponenten einer Edition separat behandelt werden sollte, wobei die Inhalte unabhängig von der Plattform, auf der sie dargestellt werden, zu betrachten sind. Dies kann durch Stäckers Vorschlag ergänzt werden, Exportformate zu verwenden, die es ermöglichen, instanziierte Ansichten eines Textes dokumentier- und reproduzierbar zu machen (Bernhart/Hahn 2014, 227). So könnte die Versionierung digitaler Inhalte unabhängig von den Fähigkeiten der betreuenden Institutionen umgesetzt werden.

Daraus ergeben sich klare Herausforderungen hinsichtlich der Zitierfähigkeit granularer Textabschnitte und der Erstellung persistenter Zitierlinks unter der Verwendung von Exportmethoden instanziierten Ansichten eines Textes zum Zeitpunkt des Zitierens.

Stand der Forschung

Ralle (2016, 155) argumentiert, dass die Eigenschaften *flexibel* und *zitierbar* gegensätzlich sind. Empfohlen wird hier, dass der Zitierende lediglich die digitale Edition angibt, ohne einen direkten Link oder den genauen Ort des Zitats zu nennen. Stattdessen findet der Suchende die Textstelle mittels Volltextsuche in dieser Edition und folgt der dort angegebenen Zitierempfehlung. Entsprechend dieser Vorgehensweise bietet die digitale Edition „*Die Tagebücher (1621–1656) von Fürst Christian II. von Anhalt-Bernburg*“ einen Permalink zur gesamten Edition sowie eine spezifische Zitierempfehlung an. Galka (2023, 22) kritisiert diese Vorgehensweise und verweist auf die fehlende Akzeptanz der Nutzenden aus der historischen Fachwissenschaft.

Das Konzept der digitalen Edition „*Johann Wolfgang von Goethe, Briefwechsel mit Friedrich Wilhelm Riemer*“ verdeutlicht eine tiefergehende Herangehensweise. Dort werden Zitiervorschläge generiert, die auf spezifische Briefwechsel innerhalb der Edition verweisen (Eckle/Nadezhda 2013, 121). Die Zitierlinks bestehen aus einem einzigartigen Identifikator für den jeweiligen Briefwechsel und Parametern, die die ausgewählte Ansicht festlegen. Diese Methode ermöglicht es zwar, auf verschiedene Briefwechsel der Edition zu verweisen, erlaubt aber nicht die Referenzierung von exakten Textstellen innerhalb der Briefe.

Stäcker (2020, 9) empfiehlt die Einführung von festen Bezugspunkten durch die Kennzeichnung von Textelementen mit Identifikatoren in XML. Diese Technik erlaubt das ge-

zielte Zitieren spezifischer Textbereiche mittels eines Permalinks. Auf diese Weise wird nicht nur das zitierte Objekt einer digitalen Edition adressierbar, sondern es können auch spezielle Bestandteile innerhalb des Objekts referenziert werden. Ein Nachteil dieser Methode ist jedoch, dass die zu adressierenden Bereiche vorab in XML definiert werden müssen und somit keine dynamische Auswahl möglich ist. Die einzigartigen Identifikatoren beziehen sich nur auf festgelegte Geltungsbereiche, die dann insgesamt zitierfähig sind.

Dynamische Markup Zitierlinks

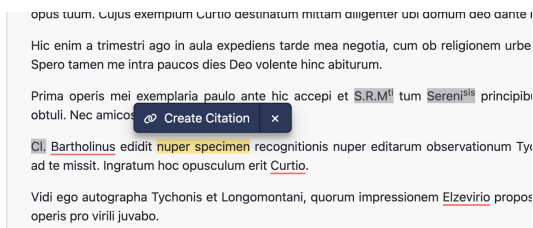


Abbildung 1: Screenshot der Funktion „Zitierlink“ in der digitalen Edition „Die sozinianischen Briefwechsel“, der das Erstellen von dauerhaften Verweisen auf spezifische Textstellen zeigt.

In der digitalen Edition „Die sozinianischen Briefwechsel“ führen wir die dynamische Platzierung von Markierungen innerhalb eines Textes sowie die Erstellung eines Zitierlinks ein, wie in Abbildung 1 illustriert. Der resultierende permanente Link ermöglicht den Zugriff sowohl auf den allgemeinen Kontext als auch auf die speziell zitierte Textstelle. Bei Nutzung des Links öffnet sich die entsprechende Seite der digitalen Edition und navigiert automatisch zur markierten Stelle. Diese Technik ist auch auf Annotationen und andere Objekte innerhalb der Edition anwendbar, da sie über dauerhafte Identifikatoren verfügen. Ein wesentlicher Aspekt dabei ist, dass der Zitierlink von den Start- und End-Indizes der markierten Stelle abhängt; Änderungen am Text können daher zu inkonsistenten Verweisen führen. Um solche Inkonsistenzen zu minimieren, könnten Exportmethoden implementiert werden, die das markierte Zitat und die zum Zeitpunkt des Zitierens angezeigte Textansicht dauerhaft festhalten. Zusätzlich könnte die Einführung von festen Bezugspunkten, durch Technologien wie *Labeled Property Graphen (LPG)* und *Applied Text as Graph (ATAG)*, den Gebrauch von Start- und End-Indizes durch stabile, einzigartige Identifikatoren ersetzen.

Applied Text as Graph (ATAG)

LPG erfassen komplexe Verbindungen und mehrdimensionale Aspekte (Kuczera 2022, 103). Diese Graphen setzen sich aus Knoten, Beziehungen und Eigenschaften zusammen und bieten eine flexible Darstellung von Textelementen sowie deren Verknüpfungen. Der Ansatz ATAG erweitert diese Vorteile, indem er einen Text in kleinere,

handhabbare Einheiten zerlegt, die in einer Graphstruktur zu einer Kette verbunden sind. Wie Kuczera (2024) zeigt, wird dabei jedes Zeichen zu einem Knoten mit einzigartigem Identifikator, der über Kanten verbunden ist, die deren Beziehungen und Abfolge definieren.

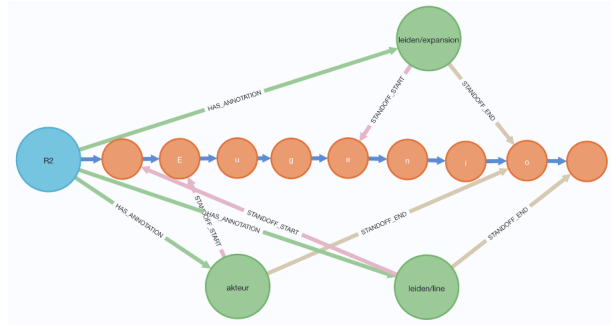


Abbildung 2: Ein LPG, der den Text „Eugenio“ visualisiert, wobei die Buchstaben als orange Knoten dargestellt und durch Kanten mit den grün repräsentierten Annotationen verbunden sind. Diese sind zusätzlich mit dem Textknoten R2 verknüpft (Kuczera 2024, 2).

Abbildung 2 zeigt einen Graphausschnitt, der den Text „Eugenio“ samt zugehörigen Annotationen abbildet. In diesem Modell wird der Text in Zeichen zerlegt, die als orange Knoten dargestellt und sequenziell zu einer Kette verbunden sind. Grüne Knoten, die an diese Textkette angebunden sind, repräsentieren die Annotationen und verdeutlichen sowohl deren Beziehungen untereinander als auch ihren jeweiligen Umfang. Ein zentraler blauer Textknoten fasst alle Elemente in einer einheitlichen Struktur zusammen, die nicht nur die Reihenfolge der Buchstaben, sondern auch den vollständigen Text in seinen Eigenschaften widerspiegelt. Jeder Knoten in diesem Graphen ist durch einen einzigartigen Identifikator (UUID) individuell referenzierbar. Dieses Verfahren ermöglicht verschiedene Bezugspunkte in einem Geflecht aus vernetzten Strukturen und Informationen zu schaffen.

Explicitly Notated Citations

Wir schlagen die Methode der Explicitly Notated Citations (ENC) vor, eine Weiterentwicklung der aktuellen Praxis zur Erstellung dynamischer Zitierlinks in der digitalen Edition „Die sozinianischen Briefwechsel“. Die Markierung des referenzierten Textabschnitts bleibt dabei gleich, allerdings erfolgt eine Anpassung der verwendeten Parameter. Anstelle der bisher genutzten Start- und End-Indizes werden die Identifikatoren der Zeichen aus der ATAG-Kette verwendet. Für einfache Referenzen reicht es aus, lediglich die Identifikatoren der Buchstaben am Anfang und am Ende des markierten Abschnitts zu nutzen. Zur Sicherstellung der Integrität der Referenz wird ein zusätzlicher Hash-Parameter eingeführt, der die UUIDs der gesamten betreffenden ATAG-Kette enthält. Sollte der Hash-Wert nicht übereinstimmen oder die Identifikatoren für Start und Ende nicht gefunden werden, so ist von einer Veränderung des referen-

zierten Textes auszugehen. Demnach kann auch ein Hinweis eingeblendet werden, der den Nutzenden der digitalen Edition darüber informiert.

	Base	Communication Identifier	Selected Text Identifier	Start Reference	End Reference	Reference Hash
Citation Link (Indices)	https://...de/id/	MAIN_ed_gbd_1hw_5tb	?guid=ed_gbd_1hw_5tb	&s= 1893	&e= 1932	
ENC-Link	https://...de/id/	MAIN_ed_gbd_1hw_5tb	?guid=ed_gbd_1hw_5tb	&s= UUID	&e= UUID	&h= Hash

Abbildung 3: Übersicht der Struktur eines Zitierlinks¹ mit Indizes und eines ENC-Links mit entsprechenden einzigartigen Identifikatoren anstelle der Indizes und einem Hash.

Diese Lösung adressiert die zuvor identifizierten Probleme der Zitierfähigkeit digitaler Editionen. Durch die Nutzung fester Bezugspunkte innerhalb einer Zeichenkette, wobei jedes Zeichen durch einen präzisen Identifikator definiert ist, ermöglicht ENC die Erstellung dynamischer, flexibler und granularer Zitierlinks. Diese ENC-Links verweisen nicht nur auf die zitierte ATAG-Kette, sondern auch auf deren Kontext. Die Anwendung von Identifikatoren und Parametern, die sich ausschließlich mit der Integrität des zitierten Textes auseinandersetzen, gewährleistet zudem die Persistenz dieser Links. Abbildung 3 zeigt die Struktur von Zitierlinks mit Indizes und ENC-Links. Zitierlinks verwenden feste Start- und Endpunkte des markierten Textes (&s=1893 und &e=1932), während ENC-Links eindeutige Identifikatoren (UUIDs) und einen Hash (&h=Hash) nutzen. Zuvor wird der Kontext mit weiteren Identifikatoren definiert, darunter der Briefwechsel (MAIN_ed_gbd_1hw_5tb) und der ausgewählte Brief (?guid=ed_gbd_1hw_5tb).

Änderungen am ursprünglich zitierten Text werden den Nutzenden der digitalen Edition beim Aufrufen des ENC-Links mitgeteilt. Allerdings wird nur darauf hingewiesen, dass Änderungen stattgefunden haben, nicht welche Veränderungen vorgenommen wurden. Zur Erfassung von Änderungen können die von Stäcker beschriebenen Exportformate von instanziierten Ansichten eingesetzt werden (Bernhart/Hahn 2014, 227). Ein Beispiel hierfür ist der Export der aktuellen Textansicht samt Annotationen und der markierten Textstelle im PDF-Format. Darüber hinaus schlagen wir vor, die Textknoten einschließlich Annotationen im JSON-Format zu exportieren, mit Git zu versionieren und den ENC-Link um einen Commit-Hash (&c=Hash) zu ergänzen. Dieser Ansatz ermöglicht die Speicherung der exportierten Kontextinformationen in einem abrufbaren Repository, wie beispielsweise GitHub. Die digitale Edition kann diese Daten bei Bedarf abrufen, um eine frühere Version des Textes in ihrer Ansicht wiederherzustellen. Dadurch wird sichergestellt, dass der Verweis weiterhin auf die ursprünglich zitierte Version zugreift und diese für die Nutzenden zugänglich bleibt. Der Commit-Hash kann dabei entweder dem klassischen Git Commit-Hash entsprechen oder in anderen Versionierungssystemen als Versionsnummer genutzt werden. Dieses Verfahren ergänzt bestehende Versionierungskonzepte, wie sie unter anderem von Bleier

(2021) und Broyles (2020) diskutiert wurden. Diese Konzepte bieten unterschiedliche Ansätze, um Änderungen in digitalen Editionen transparent und nachvollziehbar zu machen.

Abschließend ist zu betonen, dass ENC-Links die Auslieferung des zitierten Textes mitsamt den zugehörigen Annotationen in Formaten wie dem oben genannten JSON ermöglichen. Dadurch eröffnen sich zusätzliche, maschinenlesbare Auswertungsmöglichkeiten für digitale Editionen – ganz im Sinne einer algorithmischen Edition.

Zusammenfassung

Mit ENC wird eine innovative Lösung für die Herausforderungen des Zitierens in digitalen Editionen und Texten vorgeschlagen. Diese Methode ermöglicht es, spezifische Textabschnitte inklusive zugehöriger Annotationen durch einzigartige Identifikatoren präzise zu referenzieren, wodurch dynamische, flexible und granulare Zitierlinks erstellt werden können. ENC adressiert Probleme wie die Unbeständigkeit von URLs und die Schwierigkeit, granulare Textsegmente in digitalen Umgebungen zuverlässig zu zitieren. Dabei verweisen die Zitierlinks nicht nur auf die markierte ATAG-Kette, sondern auch auf deren Kontext, unterstützt durch Parameter, die ausschließlich mit der Integrität des zitierten Textes interagieren. Diese Ansätze tragen zur Persistenz der Zitierlinks bei und verbessern die wissenschaftliche Genauigkeit, indem sie sicherstellen, dass Veränderungen im ursprünglich zitierten Text den Nutzenden bei Aufrufen des Links mitgeteilt werden.

Ein besonderer Vorteil ist die Erweiterung des ENC-Links um einen Commit-Hash, der als digitale Versionsnummer dient. In Kombination mit dem Export der Textknoten und Annotationen im JSON-Format können frühere Versionen aus einem Repository wie GitHub abgerufen und in der digitalen Edition wiederhergestellt werden. So bleibt die ursprüngliche Zitierbarkeit trotz späterer Textänderungen erhalten. Dieses Verfahren erweitert bestehende Versionierungskonzepte und trägt zur langfristigen Transparenz und Nutzbarkeit digitaler Editionen bei.

Abschließend eröffnet ENC durch das Bereitstellen des zitierten Textes mitsamt Annotationen in Formaten wie JSON neue Perspektiven für maschinelle Analysen und unterstützt den Übergang von einer digitalen zu einer algorithmischen Edition.

Fußnoten

1. Zitierlink: https://sozinianer.de/id/MAIN_ed_gbd_1hw_5tb?guid=ed_gbd_1hw_5tb&s=1893&e=1932

Bibliographie

Bernhart, Toni, und Carolin Hahn. 2014. "Datenmodellierung in digitalen Briefeditionen und

ihre interpretatorische Leistung. Ontologien, Textgenetik und Visualisierungsstrategien. Workshop im Jacob-und-Wilhelm-Grimm-Zentrum der Humboldt-Universität zu Berlin, 15./16. Mai 2014." *Editio* 28, Nr. 1: 225-229. Berlin: De Gruyter.

Bleier, Roman. 2021. "How to cite this digital edition?" *DHQ: Digital Humanities Quarterly* 15, Nr. 3.

Borgman, Christine L. 2010. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.

Broyles, Paul A. 2020. "Digital Editions and Version Numbering." *Digital Humanities Quarterly (DHQ)* 14, Nr. 2. Alliance of Digital Humanities Organizations.

Eckle, Jutta, und Nadezhda Petrova. 2013. "„Wenigstens erlauben Ew Excellenz daß ich bey zu hoffender baldiger Zurückkunft auf dieses Thema das Gespräch lenken darf.“ Zur Konzeption der neuen Edition Johann Wolfgang von Goethe. Briefwechsel mit Friedrich Wilhelm Riemer (Hybrid Ausgabe)." *Internationales Jahrbuch für Editionswissenschaft* 27, Nr. 1: 112-123. Berlin: De Gruyter.

Galka, Selina. 2023. "Review of 'Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg (1599–1656)'." *RIDE* 16. doi: 10.18716/ride.a.16.2.

Kuczera, Andreas. 2022. "TEI Beyond XML – Digital Scholarly Editions as Provenance Knowledge Graphs." In: Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera and Joris van Zundert (eds.): *Graph Technologies in the Humanities - Proceedings 2020*, published at <http://ceur-ws.org/Vol-3110>, 2022.

Kuczera, Andreas. 2024. "Applied Text as Graph (ATAG)". *DHd 2024 Quo Vadis (DHd2024)*, Zenodo. doi: 10.5281/zenodo.10698323.

Ralle, Inga Hanna. 2016. "Maschinenlesbar--menschenslesbar: Über die grundlegende Ausrichtung der Edition." *Editio* 30, Nr. 1: 144-156. Berlin: De Gruyter.

Stäcker, Thomas. 2020. ">A Digital Edition is Not Visible<--Some Thoughts on the Nature and Persistence of Digital Editions." *ZfdG - Zeitschrift für digitale Geisteswissenschaften*, Nr. 5. Forschungsverbund Marbach Weimar Wolfenbüttel.

Stronks, Els, und Peter Boot. 2007. "Learned Love. Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)." *Data Archiving and Networked Services (DANS)*.