

Beitrag aus:

Sonderband 4 der ZfdG: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019. DOI: [10.17175/sb004](https://doi.org/10.17175/sb004)

Titel:

Modellierung von Zweifel – Vorbild TEI im Graphen

Autor/in:

Andreas Kuczera

Kontakt: andreas.kuczera@adwmainz.de

Institution: Akademie der Wissenschaften und der Literatur, Mainz

GND: [1167802993](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63864-p0001-9) ORCID: [0000-0003-1020-507X](https://orcid.org/0000-0003-1020-507X)

Autor/in:

Dominik Kasper

Kontakt: dominik.kasper@adwmainz.de

Institution: Akademie der Wissenschaften und der Literatur, Mainz

GND: [1018231137](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63864-p0001-9) ORCID: [0000-0002-6587-381X](https://orcid.org/0000-0002-6587-381X)

DOI des Artikels:

[10.17175/sb004_003](https://doi.org/10.17175/sb004_003)

Nachweis im OPAC der Herzog August Bibliothek:

[1037067967](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63864-p0001-9)

Erstveröffentlichung:

18.07.2019

Lizenz:

Sofern nicht anders angegeben



Medienlizenzen:

Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:

03.06.2019

GND-Verschlagwortung:

[Edition](#) | [Graphdatenbank](#) | [Semantische Modellierung](#) | [Text Encoding Initiative](#) | [Ungewissheit](#) |

Zitierweise:

Andreas Kuczera, Dominik Kasper: Modellierung von Zweifel – Vorbild TEI im Graphen. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4). PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/sb004_003](https://doi.org/10.17175/sb004_003).

Andreas Kuczera, Dominik Kasper
Modellierung von Zweifel – Vorbild TEI im Graphen

Abstracts

Im Fokus der hier ausgewerteten Integration von TEI-Dokumenten aus dem Deutschen Textarchiv (DTA) in eine Graphdatenbank (Neo4j) steht die Auszeichnung von unsicheren Lesarten und editorischen Ergänzungen in Handschriften. In diesem Zusammenhang legen wir auch die TEI-Richtlinien zum Umgang mit Zweifel und Unsicherheiten dar. In Editions- oder Transkriptionsvorhaben arbeiten zumeist mehrere Personen. Deshalb sehen wir responsibility-Angaben an zweifelhaften Stellen als zentral für die Interoperabilität der Daten und die intersubjektive Nachvollziehbarkeit von Einzelentscheidungen an. Dies gilt insbesondere dann, wenn zusätzlich Angaben zum Grad der Sicherheit einer Auflösung möglich sind. Graphtechnologien bieten hier Möglichkeiten zur Modellierung, Visualisierung und Analyse von Unsicherheit und Verantwortung. Bei einer ausreichend großen Datenmenge lassen sich beispielsweise persönliche Auszeichnungsprofile der jeweiligen Bearbeiter erstellen.

The focus of the imports of TEI documents from the German Text Archive (DTA) into a graph database (Neo4j) evaluated here is the marking of uncertain readings and editorial additions in manuscripts. In this context, we also briefly present the TEI guidelines for dealing with doubts and uncertainties. Since several people usually work in editing or transcription projects, we regard responsibility information at dubious points as central to the interoperability of data and the intersubjective traceability of individual decisions. This applies in particular if additional information on the degree of security of a resolution is possible. Graph technologies offer possibilities for modeling, visualization and analysis of uncertainty and responsibility. With a sufficiently large amount of data, personal labeling profiles of the respective editors can be created, for example.

1. Modellierung von Zweifel in der TEI

1.1 Codierung von Unsicherheiten in der Lesart und von Lücken im Text

Die Text Encoding Initiative (TEI) bietet bei der Transkription von Texten verschiedene Möglichkeiten, zweifelhafte Lesarten auszuzeichnen und diese sowie damit in Zusammenhang stehende Informationen umfangreich zu dokumentieren. In diesem Beitrag werden zunächst die Elemente und Attribute zur Modellierung von Unsicherheit und Zweifel in der TEI dargestellt. Zentral ist in diesem Zusammenhang unseres Erachtens das *Kapitel 11.3.3.2 Use of the gap, del, damage, unclear, and supplied Elements in Combination* der TEI-Richtlinien,¹ wo deren kombinierte Verwendung erklärt wird. Im Folgenden liefern wir hier eine Zusammenfassung dieser speziellen Codierungsmöglichkeiten. Die Links in den

¹ TEI-Guidelines 2018 Kapitel 11.3.3.2. *Use of the gap, del, damage, unclear, and supplied Elements in Combination*.

Fußnoten am Ende jeder Einzelerklärung führen zum Grundeintrag des Elements in der Online-Dokumentation der TEI, worin dessen allgemeine Bedeutung und Verwendung beschrieben wird. Im Einzelnen gehen wir hier ein auf:

- `<gap>` ist ein leeres Element, das eine Lücke im Text kennzeichnet. Unter einer Lücke wird eine Stelle verstanden, an der durch Tilgung oder Schaden alles komplett unlesbar ist.²
- `<supplied>` umschließt einen editorisch ergänzten Teil, bei einer durch Tilgung oder Schaden komplett unlesbaren Stelle im Text (siehe `<gap>`).³
- `<unclear>` umschließt einen transkribierten Buchstaben oder Textteil an einer Stelle, wo noch etwas unsicher lesbar ist, aber ein Teil durch Tilgung oder Schaden unlesbar wurde.⁴
- `<@cert>` ist ein Attribut, das den Grad an Sicherheit bzw. Gewissheit bei der unsicheren Lesung (als Attribut von `unclear`) oder der ergänzten Stelle (als Attribut von `<supplied>`) beinhaltet.⁵ In der Regel werden fixe Werte vorgegeben (`high`, `medium`, `low`, `unknown`).⁶
- `<@resp>` ist ein Attribut, das einen Verweis auf den oder die Edierende beinhaltet, der oder die für die Auflösung der unsicheren Lesart oder die Ergänzung verantwortlich zeichnet.⁷
- `<@reason>` ist ein Attribut, indem die Ursache für Schäden, Tilgungen oder Lücken dokumentiert werden kann.⁸ Auch hier werden in der Regel fixe Werte vorgegeben.⁹

Da die TEI-Richtlinien die Verwendung von `<gap>`, `<supplied>` und `<unclear>` für den hier betrachteten Modellierungszusammenhang in Kombination mit weiteren transkriptionstypischen Elementen erklären,¹⁰ seien diese hier ebenfalls kurz erläutert:

- `<damage>` umschließt einen beschädigten Teil, bei einer Stelle, an der noch etwas lesbar ist, aber ein Teil durch Schaden unleserlich wurde.¹¹
- `<subst>` umschließt einen Ersetzungsvorgang, der beispielsweise mit `add` und `del` näher beschrieben werden kann.¹²
- `` umschließt den getilgten Teil bei einer Stelle, an der noch etwas lesbar ist, aber ein Teil durch Tilgung unleserlich wurde.¹³
- `<add>` umschließt den ergänzten Teil bei einer Stelle, an der etwas verbessert bzw. ersetzt wurde.¹⁴

² Siehe auch TEI Guidelines 2018, Kapitel 3.4.3 Additions, Deletions, and Omissions `<gap>` .

³ Siehe auch TEI Guidelines 2018, Kapitel 11.3.3.1 Damage, Illegibility, and Supplied Text `<supplied>` .

⁴ Siehe auch TEI Guidelines 2018, Kapitel 11.3.3.1 Damage, Illegibility, and Supplied Text `<unclear>` .

⁵ Siehe auch TEI Guidelines 2018, `att.global.responsibility <@cert>` .

⁶ Siehe auch TEI Guidelines 2018, `teidata.certainty` .

⁷ Siehe auch TEI Guidelines 2018, `att.global.responsibility <@resp>` .

⁸ Siehe auch TEI Guidelines 2018, `teidata.certainty <@reason>` .

⁹ Siehe auch TEI Guidelines 2018, Kapitel 3.4.3 Additions, Deletions, and Omissions `<@reason>` .

¹⁰ Vgl. TEI Guidelines 2018, Kapitel 11.3.3.2 Use of the gap, del, damage, unclear, and supplied Elements in Combination .

¹¹ Siehe auch TEI Guidelines 2018, Kapitel 11.3.3.1 Damage, Illegibility, and Supplied Text `<damage>` .

¹² Siehe auch TEI Guidelines 2018, Kapitel 11.3.1.5 Substitutions `<subst>` .

¹³ Siehe auch TEI Guidelines 2018, Kapitel 3.4.3 Additions, Deletions, and Omissions `` .

¹⁴ Siehe auch TEI Guidelines 2018, Kapitel 3.4.3 Additions, Deletions, and Omissions `<add>` .

1.2 Codierung von allgemeiner Unsicherheit, Verantwortlichkeit und Genauigkeit

Neben der Möglichkeit, die Sicherheit einer editorischen Entscheidung, deren Begründung und Verantwortlichkeit zu dokumentieren, beinhaltet die TEI auch ein Modul, um Zweifel und Unsicherheit ob der richtigen Verwendung von TEI-Elementen selbst zu kodieren. Ebenfalls abgedeckt werden darin die Auszeichnung von Unsicherheiten bei der (vermeintlichen) Identifikation einer Entität, einer Textstruktur oder auch bei der Angabe und Auflösung numerischer Werte.

Erwähnt werden müssen hier daher auch die im Modul `certainty` auftretenden Elemente `<precision>`, `<respons>` und das mit dem Modul gleichnamige `<certainty>`:¹⁵

- `<certainty>` dient der Kodierung von Unsicherheiten bei der Verwendung von Elementen und Attributen bzw. damit ausgezeichneter Bereiche und zielt damit prinzipiell auf die Inhaltsebene. So können beispielsweise Zweifel daran dokumentiert werden, ob es sich um einen Orts- oder Personennamen handelt oder auch ob ein Absatz mit der Seite endet oder sich noch auf die nächste erstreckt.¹⁶

- `<precision>` ermöglicht die Exaktheit bzw. Genauigkeit von numerischen Aspekten des Markups (Datierungen, Einheiten, sonstige Zahlenwerte) graduell zu codieren bzw. näher zu beschreiben.¹⁷

- `<respons>` identifiziert den oder die Edierende, welche für bestimmte Aspekte von Inhalt und Auszeichnung verantwortlich zeichnet. Es ist gegenüber dem oben vorgestellten `resp`-Attribut angesichts verschiedener Attributionsmöglichkeiten deutlich feingranularer.¹⁸

Die genannten Elemente können mit zahlreichen Attributen versehen werden und bieten insgesamt sehr detaillierte Optionen, um Zweifel und Unsicherheit auf verschiedenen Ebenen in TEI-XML abzubilden. In der Praxis von mit TEI arbeitenden Projekten wie der Carl-Maria-von-Weber-Gesamtausgabe (WeGa),¹⁹ dem Deutschen Textarchiv (DTA) und der noch in der Entwicklung befindlichen PROPYLÄEN. Forschungsplattform für Goethes Biographica spielt das Modul `certainty` keine Rolle. Die Ergebnisse einer 2010 durchgeführten und 2012 veröffentlichten Studie zur TEI-Benutzung²⁰ bei der Handschriftenkodierung lassen annehmen, dass die scientific community der TEI-Benutzenden die in den Modulen `core` (`<unclear>`, `<gap>`, `<add>`, `core`) und `transcr` (`<supplied>`, `<damage>`, `<subst>`) befindlichen Elemente für ausreichend hält, um das gewünschte Maß an Dokumentation von Zweifel und Unsicherheit zu erreichen. Die Attribute `@resp` und `@cert` gehören zur TEI-Infrastruktur, im Basis-Modul `tei`.

¹⁵ Vgl. TEI Guidelines 2018, Kapitel 21 *Certainty, Precision, and Responsibility* und besonders TEI Guidelines 2018 den Abschnitt 21.1.2 *Structured Indications of Uncertainty* zur hier folgenden Zusammenfassung.

¹⁶ Siehe auch TEI Guidelines 2018, Kapitel 21.1.2 *Structured Indications of Uncertainty <certainty>*.

¹⁷ Siehe auch TEI Guidelines 2018, Kapitel 21.2 *Indications of Precision <precision>*.

¹⁸ Siehe auch TEI Guidelines 2018, Kapitel 21.3 *Attribution of Responsibility <respons>*.

¹⁹ Vgl. Carl-Maria-von-Weber-Gesamtausgabe 2018.

²⁰ Vgl. Burghart / Rehbein 2012.

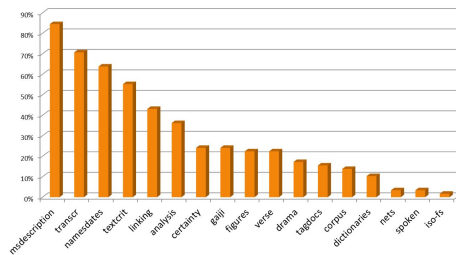


Abb. 1: TEI-Modules used in manuscript encoding projects (apart from the four basic ones: core, tei, header, textstructure). [Burghart / Rehbein 2012 , Fig. 11. CC BY-ND 3.0.]

Wie die obige Grafik zeigt, kommt das Modul certainty nur bei ca. 22 % der Vorhaben, die Handschriften auszeichnen, zum Einsatz. Gleichzeitig sehen nur sehr wenige TEI-Anwenderinnen und -Anwender Bedarf für eine Ausweitung oder Verbesserung der Richtlinien zur Verwendung der Elemente dieses Moduls, wie die folgende Grafik deutlich macht.

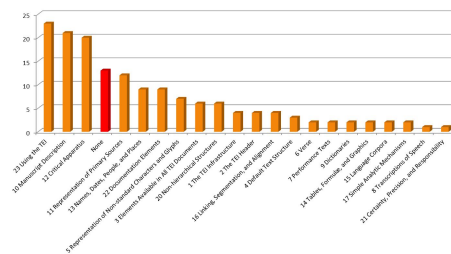


Abb. 2: In what areas do you wish the Guidelines to be improved? Figures are given in absolute numbers.
[Burghart / Rehbein 2012, Fig. 18. CC BY-ND 3.0.]

Das entsprechende Kapitel der Richtlinien steht an letzter Stelle. Damit muss dieser Bereich als gut dokumentiert gelten. Interpretiert man dieses Ergebnis aber in Kombination mit der Aussage von [Abbildung 1](#), so liegt die Annahme eines nur geringen Interesses der Fachwelt an einer sehr detaillierten Dokumentation von Zweifel und Unsicherheit nah. Zweifellos werden dabei arbeitsökonomische Aspekte eine Rolle spielen, aber auch Fragen nach dem Mehrwert solcher Informationen.

Nach dem kurzen Exkurs über die theoretische Tiefe der TEI hinsichtlich der Modellierung von Zweifel und Unsicherheit, ist der starke Eindruck entstanden, dass nur ein geringer Teil der Möglichkeiten genutzt wird. Im Folgenden richten wir den Blick auf zwei Beispiele aus der Praxis zum Umgang mit unsicheren Lesarten und Lücken im Text.

2. Zwei Beispiele aus der Praxis

2.1 DTA-Basisformat-Realisierung

Das Deutsche Textarchiv (DTA) hat ein eigenes, sehr umfangreich dokumentiertes Datenformat etabliert, das **DTA-Basisformat**, welches ein subset²¹ der TEI ist. Zu unsicheren Lesarten bzw. schwer lesbaren Zeichen heißt es dort:

»Ist die Leserlichkeit der Quelle eingeschränkt, sodass der Text rekonstruiert werden muss bzw. die Lesung des Editors nicht gesichert ist, kann dies durch die Elemente <unclear> und <supplied> wiedergegeben werden. Dabei wird <unclear> verwendet, wenn in der Quelle vorhandenes Material nur undeutlich lesbar ist. Der Grund für die Verwendung des <unclear>-Elements wird mit dem @reason-Attribut, der Grad der Sicherheit der Lesung kann im @cert-Attribut wiedergegeben werden. [...] Die Verwendung des Attributs @reason in <unclear> ist dabei obligatorisch, die Verwendung von @cert ist fakultativ.[...] Wenn in der Quelle wahrscheinlich oder möglicherweise vorhandenes Material rekonstruiert wird, so ist dies mit dem Element <supplied> wiederzugeben. Der Grund für die Unleserlichkeit wird im @reason-Attribut wiedergegeben, die Sicherheit der Rekonstruktion steht im @cert-Attribut.«²²

»Lassen sich die Zeichen nicht erkennen und nicht mehr rekonstruieren, wird das Tag <gap> gesetzt, um die Lücke anzuzeigen. Innerhalb des <gap>-Tags kann mittels der Attribute @unit, @quantity und @reason der Bezug angezeigt werden, wie viele Zeichen die Lücke umfasst, so wie der Grund der Fehlstelle[.]«²³

2.1.1 Code-Beispiel

Das folgende XML-Fragment stammt aus der Transkription von Gotthilf Patzigs Mitschriften von Humboldts Vorträgen über physische Geographie.²⁴

- 1) <xml>
- 2) [...] So wie die Geognoſie durch die Auf-
- 3) <lb /> findung u. nähere Beachtung der thieriſchen Ver-
- 4) <lb /> ſteinerungen aufgeklärt wurde: ſo hat
- 5) <lb /> der phyſiſche Theil der Aſtronomie durch
- 6) <lb /> die Entdeckungen im Gebiet der Optik gewoñ<supplied
- 7) reason="damage" resp="#BF">en;</supplied>
- 8) <lb /> u. die Cometen ſind beſonders näher

²¹ Der Begriff ist hier im engeren mathematischen Sinne als echte Teilmenge zu verstehen. Das heißt, das DTABf ist eine reduzierte Fassung der TEI-P5-Richtlinien ohne Erweiterungen durch eigene Elemente.

²² DTA-Basisformat 2011–2018, *Unsichere Lesarten*. Dort finden sich auch mögliche Werte für die genannten Attribute.

²³ DTA-Basisformat 2011–2018, *Schwer bzw. nicht entzifferbare Zeichen und Auslassungen*.

²⁴ Patzig 2007 (1827/1828), S. 13. Zur HTML-Ansicht des Beispiels: [Deutsches Textarchiv 2007–2019, Patzig](#).

9) <lb /> beobachtet werden. Jn beſtändiger
 10) <lb /> Bewegung kañ man dieſe eine perio-
 11) <lb /> diſch oſcilirende neñen. Dieſe kañ
 12) <lb /> gehem̃t, geſtöhr't werden auf viele Weiſe;
 13) <lb /> deñ welch ein geringer Stoß von auß<unclear
 14) reason="illegible" cert="high" resp="#CT">en</unclear>
 15) <lb /> dürfte dazu gehören ſie in Bewegung
 16) <lb /> zu ſetzen, da die Düñigkeit derſelben
 17) <lb /> Alles überſteigt was wir ſelbſt von
 18) <lb /> Gas-Arten auf der Erde keñen. Dieſe
 19) <lb /> Düñigkeit iſt 5000 mal geringer als die
 20) <lb /> Dichtigkeit der Erde. – Kom̃en wir jetzt
 21) <lb /><note place="left"><hi rendition="#u">Telluriſche
 22) Verhältniſſe</hi>
 23) <lb /></note><hi rendition="#u">zu den telluriſchen
 24) Verhältniſſen, <subst><del rendition="#erased">gap
 25) reason="illegible" /><add place="across">ſo
 26) werd</add></subst>en
 27) <lb /> wir die Form, Größe u. Dichtigkeit des
 28) <lb /> Planeten betrachten –</hi>[...]
 29) </xml>

In Zeile 6 und 7 ist eine editorische Textergänzung vorgenommen worden. Hier hat der oder die Bearbeitende mit dem Kürzel »BF« aufgrund einer Beschädigung der Vorlage die Zeichen »en;« ergänzt. In Zeile 13 und 14 dokumentiert der Bearbeitende »CT« die unsichere Lesart der Zeichen »en« am Ende einer Zeile. Die Auflösung erfolgte mit hoher Gewissheit. Der in den Zeilen 24 bis 26 in einem <subst> codierte Vorgang lässt sich mit Bezug auf die Dokumentation in natürlicher Sprache so formulieren: Eine Tilgung durch Radieren, Auskratzen o. ä. hat hier eine Lücke im Text entstehen lassen, eine nicht mehr lesbare Stelle. Direkt darüber wurde nun die Zeichen »fo werd« geschrieben. Wer diese Annotation vorgenommen hat, ist hier nicht ersichtlich.

2.2 Aus der Entwicklung – PROPYLÄEN. Goethes Biographica

Im Projekt PROPYLÄEN werden die textkritischen Anmerkungen nicht direkt im Text codiert, sondern in einem per Referenz verknüpften Apparatbereich (in TEI-Code ausgedrückt: <variantEncoding method="location-referenced" location="external" />). Im Apparat und konstituierten Text werden in der Handschrift *nicht eindeutig entzifferbare* Buchstaben oder Zahlen ebenfalls mit <unclear reason="illegible"> codiert. In der Handschrift *nicht entzifferbare* Zeichen werden mit <gap reason="illegible" extent="ANZAHL DER ZEICHEN SOFERN ERMITTELBAR"> ausgezeichnet.

2.2.1 Code-Beispiele

Die Codierung wurde zu Ansichtszwecken vereinfacht, alle Kommentarreferenzen wurden entfernt.

Unsichere Lesart (Codierung im konstituierten Text):²⁵

- 1) <div type="entry" xml:id="GT01_1782_007">
- 2) <head>
- 3) <origDate when="1782-01-07" rendition="#fraktur">
- 4) 7 Mont. Isidorus
- 5) </origDate>
- 6) </head>
- 7) <note ana="metadaten">
- 8) <placeName type="uebernachtungsort">Weimar</placeName>
- 9) </note>
- 10) <p>Ackten und verschiedene Besorgungen. Mittags Crone.
- 11) um halb 5 zur reg. H. dann zu Seckend. wo # war und über
- 12) Aufzüge gesprochen wurde p zur Waldner
- 13) war # dasel<unclear reason="illegible">b</unclear>st und
- 14) Stein. kam #. Ging mit ihm auf Zimmer, ihm die Erfindung
- 15) zu erzählen.</p>
- 16) </div>

In Zeile 13 wird die unsichere Lesart des Buchstabens »b« vermerkt.

Nicht mehr lesbare Stelle oder Lücke (Codierung im Apparat):²⁶

- 1) <app xml:id="app_05">
- 2) <rdg>
- 3) <subst>
- 4) <del rendition="#sofortkorrektur">
- 5) <gap reason="illegible" extent="1 char" />
- 6)
- 7) <add>den Fus.</add>
- 8) </subst>
- 9) </rdg>
- 10) </app>

Der Apparateintrag codiert, dass ein nicht mehr lesbares Zeichen zu »den Fus.« korrigiert wurde. Die Korrektur erfolgte sofort und nicht später, bspw. durch einen Schreiber.

²⁵ Bisher unveröffentlichte und in Modellierungs- und Auszeichnungsarbeit befindliche Daten aus der Retrodigitalisierung, s. Goethe Tagebücher 1998, T I,1, S. 130, Z. 1–5.

²⁶ Goethe Tagebücher 1998, T I,1, S. 130, Z. 14.

3. Modellierung im Graphen

3.1 Die Kosmos-Vorträge von Alexander von Humboldt

Im zweiten Teil des Beitrags werden die Elemente <unclear> und <supplied> aus dem DTA-Basisformat näher betrachtet. Datengrundlage sind hierbei die Kosmos-Vorträge von Alexander von Humboldt, die dieser 1827/28 in Berlin einmal an der Universität und einmal an der Sing-Akademie gehalten hat. Zu diesen Vorträgen liegen Mitschriften vor, die im Rahmen eines Forschungsprojekts im Deutschen Textarchiv transkribiert wurden.²⁷ Humboldts eigene Manuskripte zu den Vorträgen sind nicht erhalten. Es gibt aber für beide Vortragsreihen Mitschriften von Zuhörenden, die aber voneinander abweichen. Vereinfacht gefragt, geht es also darum, was Humboldt wirklich gesagt hat.

Im Folgenden werden fünf der Mitschriften gemeinsam in eine Graphdatenbank eingespielt, die Verwendung der Elemente <unclear> und <supplied> untersucht und schließlich in Relation zu den Edierenden gebracht, die sie in der Transkription verwendet haben (also jenen, die im @resp-Attribut genannt sind). Graphdatenbanken sind sehr gut für die Darstellung stark vernetzter Daten geeignet und in diesem Fall wäre es ein Versuch wert, die Rolle von Editorinnen und Editoren über die Grenzen der Dateien hinweg auszuwerten.

3.2 Import von TEI-XML in eine Graphdatenbank

Für die Analyse von TEI-Unsicherheitsannotationen im Graphen müssen die XML-Daten zunächst in die Graphdatenbank Neo4j importiert werden. Hierfür hat Stefan Armbruster²⁸ die apoc-Bibliothek von Neo4j um die procedure apoc.load.xml erweitert. Prinzipiell können XML-Dateien ohne größere Probleme in einen Graphen importiert werden, da sie einen geordneten, gerichteten azyklischen Graphen darstellen, der vielfache Elternbeziehungen verhindert. Damit stellen sie ein Ordered Hierarchy of Content Objects (OHCO) dar.

Der folgende query importiert eine der fünf Vorlesungsmitschriften:²⁹

call

```
apoc.xml.import('http://www.deutschestextarchiv.de/book/download_xml/  
parthey_msgermqu1711_1828',{createNextWordRelationships:true})
```

²⁷ Vgl. Thomas et al. 2016, S. 287–318.

²⁸ Stefan Armbruster ist bei neo4j tätig.

²⁹ Mit dem Befehl wird die Vorlesungsmitschrift von Gustav Parthey importiert. Zur Partheymitschrift vgl. *Deutsches Textarchiv 2007–2019, Parthey 1827/1828*. Informationen zu den weiteren Mitschriften sind zu finden unter *Deutsches Textarchiv 2007–2019, Titeldaten*. Der Query für den Import aller fünf Mitschriften steht als *Download* zur Verfügung. Der Import von Texten des Deutschen Textarchivs in die Graphdatenbank Neo4j wird ausführlich erklärt im Kapitel *XML Text im Graphen* in Kuczera 2018.

```
yield node return node;

// URL von Dokument auf alle Wort-Knoten kopieren:

match (d:XmlDocument)-[:NEXT_WORD*]->(w:XmlWord)

set w.url = d.url;

// Knoten durchzählen

MATCH p =

(start:XmlDocument)-[:NEXT*]->(end:XmlTag)

WHERE NOT (end)-[:NEXT]->() AND start.url =

'http://www.deutschestextarchiv.de/book/download_xml/parthey_msgermqu1711_1828'

WITH nodes(p) as nodes, range(0, size(nodes(p))) AS indexes

UNWIND indexes AS index

SET (nodes[index]).DtaID = index;
```

Für den Import der weiteren Mitschriften muss in dem Befehl die DTA-URL entsprechend geändert werden.

Beim Import werden die XML-Knoten in Graphknoten umgewandelt und verschiedene Arten von Kanten erstellt, welche die Baum-Hierarchie des XMLs im Graphen abbilden. Mit der Option `createNextWordRelationships:true` wird darüber hinaus festgelegt, dass die im XML vorhandenen Textknoten über `NEXT_WORD`-Kanten miteinander verknüpft werden. Zu beachten ist hierbei, dass es in TEI-XML zwei verschiedene Arten von Elementen gibt. Die eine Klasse dient der Klassifizierung von Text, die zweite Art bringt Varianten und zusätzlichen Text mit, der beim Import in seiner Serialität eingelesen und mit `NEXT_WORD`-Kanten verbunden wird. Dies kann dann zur Sinnentstellung des Textes führen.³⁰ Mit dem zweiten cypher-Befehl wird jedem `XmlWord`-Knoten die URL des XML-Dokuments als property mitgegeben. Damit behält man im Graphen beim Betrachten der Ergebnisse den Überblick und kann die `XmlWord`-Knoten einem XML-Dokument zuordnen. Der dritte query nummeriert die Knoten pro Datei durch und macht sie damit innerhalb des Dokuments eindeutig referenzierbar.

³⁰ Zum Import von XML-Text in die Graphdatenbank Neo4j vgl. insbesondere das Kapitel zu *XML-Text im Graphen* in Kuczera 2018.

3.3 Das XML-Element <unclear> im Graphen

Mit dem folgenden query wird eine Stelle im importierten XML aufgerufen, an der ein <unclear>-Element verwendet wurde:

```
// unclear-Beispiel
```

```
MATCH
```

```
(t1:XmlTag {_name:'lb'})<-[NEXT_SIBLING]-(t2:XmlTag {_name:'unclear'})<-[NEXT]-(w3:XmlWord {text:'auß'}),
```

```
(w1:XmlWord)-[:NEXT_WORD]->
```

```
(w2:XmlWord)-[:NEXT_WORD]->
```

```
(w3:XmlWord)-[:NEXT_WORD]->
```

```
(w4:XmlWord)-[:NEXT_WORD]->
```

```
(w5:XmlWord)
```

```
RETURN *;
```

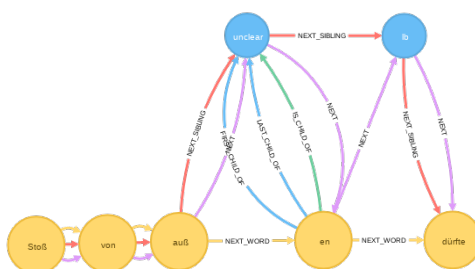


Abb. 3: Beispiel zur Graphmodellierung eines unclear-Elements. [Kasper / Kuczera 2019.]

Die entsprechende Stelle sieht in XML wie folgt aus:³¹

```
Wei&#x017F;e;<lb/>
den&#x0303;welch ein geringer Stoß von auß<unclear
reason="illegible" cert="high" resp="#ct">en</unclear>
<lb/>
dürfte dazu gehören &#x017F;ie in Bewegung<lb/>
zu &#x017F;etzen, da die Dün&#x0303;igkeit
```

Abb. 4: Das unclear-Beispiel in der XML-Ansicht des DTA. [Kasper / Kuczera 2019.]

³¹ Vgl. Deutsches Textarchiv 2007–2019, Patzig 1827/1828.

Das für den Grad des Zweifels maßgebliche @cert-Attribut befindet sich in den properties des unclear-Knotens.

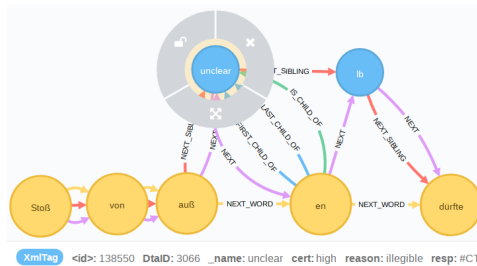


Abb. 5: Die properties des unclear-Knotens. [Kasper / Kuczera 2019.]

Wie im Beispiel aufgezeigt, können mit dem Import alle Informationen des XMLs verlustfrei in den Graph überführt und abgebildet werden.

3.4 Die Zweifel der Edierenden

Die objektive Gewichtung von Zweifeln im Hinblick auf die Interoperabilität ist schwierig. Ermöglicht man den Edierenden feingranularere Abstufungen, um Zweifel zum Ausdruck zu bringen (z. B. in 10er-Schritten von 0 % bis 100 %) führt das oft zu Verunsicherung. Gibt es nur zwei Stufen, wie im DTA-Basisformat mit *high* und *low*, bleibt die Gewichtung grob, Vergleiche fallen aber leichter. Der hier vorgestellte Ansatz verzichtet auf eine objektive Vergleichbarkeit und ordnet die von den Edierenden vergebenen gewichteten Zweifeln den Personen zu. Stehen genügend Daten zur Verfügung, könnte aus den Annotationen ein persönlicher Fingerabdruck des jeweiligen Edierenden erstellt werden.

Zunächst wird mit folgendem cypher query abgefragt, welcher Edierende in welchem Dokument welche XML-Elemente genutzt hat, wobei die XML-Elemente sowohl das @cert- als auch das @resp-Attribut haben müssen:

```
// Zweifelsattribute in der TEI pro Dokument
```

```
MATCH (n:XmlTag)
```

```
WHERE n.resp IS NOT NULL
```

```
AND n.cert IS NOT NULL
```

```
RETURN n.url, n._name AS Element, n.resp AS Person, n.reason, n.cert, count(n.resp) AS Anzahl
```

```
ORDER BY Anzahl DESC
```

Name	Element	Person	n.reason	n.cert	Anzahl
patzig_msgermfol841842_1828	unclear	#BF	illegible	high	137
patzig_msgermfol841842_1828	unclear	#BF	illegible	low	97
parthey_msgermq1711_1828	unclear	#CT	illegible	high	34
nn_msgermq2345_1827	unclear	#BF	illegible	high	27
parthey_msgermq1711_1828	unclear	#CT	illegible	low	25
parthey_msgermq1711_1828	unclear	#CT	covered	high	22
nn_msgermq2345_1827	unclear	#BF	illegible	low	19
patzig_msgermfol841842_1828	unclear	#BF	covered	low	19
patzig_msgermfol841842_1828	unclear	#BF	covered	high	16
parthey_msgermq1711_1828	unclear	#CT	covered	low	13

Abb. 6: Die gekürzt wiedergegebene Tabelle zeigt die häufigsten Ergebnisse des obigen Querys. Die Angaben in der ersten Spalte der Tabelle wurden aus Gründen der Übersichtlichkeit um den URL-Teil gekürzt, der bei allen Mitschriften gleich ist. [Kasper / Kuczera 2019.]

Mit Abstand am häufigsten wurde bei der Transkription der Vorlesungsmitschriften das <unclear>-Element verwendet, mit einigem Abstand gefolgt vom <supplied>-Element.

3.4.1 Die Identifizierung des Edierenden

Die Spalte Person in der Tabelle gibt den Inhalt der resp-property an, in der die Person des Edierenden mit einem Kürzel wiedergegeben wird. Im XML-Header werden diese Kürzel auf folgende Personen aufgelöst:

Kürzel	Name
CT	Christian Thomas
BF	Benjamin Fiechter
TK	Tina Krell

Abb. 7: Aufschlüsselung der Edierendenkürzel. [Kasper / Kuczera 2019.]

Es sind Personen, die mit der Transkription der Humboldt-Vorlesungsmitschriften befasst waren. Im nächsten Schritt werden nun alle in den resp-properties genannten Edierenden explizit als Personenknoten erstellt und mit jenen unclear- und supplied-Knoten verknüpft, für die sie verantwortlich sind. Mit folgendem query werden die Personen erstellt:

```
// Edierende Personenknoten erstellen
```

```
MATCH (n:XmlTag)
```

```
WHERE n.resp IS NOT NULL
```

```
AND n.cert IS NOT NULL
```

```
MERGE (p:Person {name:n.resp})
```

```
RETURN *;
```

// Alle XML-Elemente mit resp-Attribut den erstellten Personen zuordnen

```
MATCH (n:XmlTag), (p:Person {name:n.resp})
```

```
WHERE n.resp IS NOT NULL
```

```
AND n.cert IS NOT NULL
```

```
MERGE (p)-[:RESPONSIBLE {cert:n.cert}]->(n)
```

```
RETURN *;
```

und mit den entsprechenden unclear- und supplied-Knoten im Graphen über RESPONSIBLE-Kanten verknüpft.

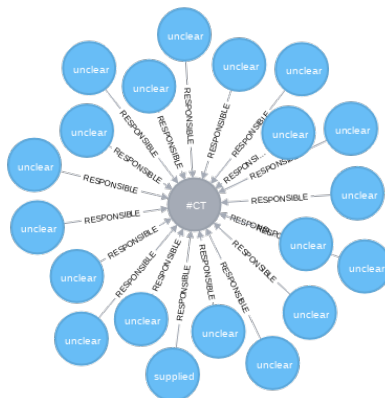


Abb. 8: unclear- und supplied-Knoten, die von #CT erstellt worden sind. [Kasper / Kuczera 2019.]

3.4.2 Statistik zur Zweifel im Graphen

Mit dem folgenden cypher query fragen wir die Häufigkeit der unclear- und supplied-Knoten im Graphen ab und ordnen sie den Edierenden zu:

```
// 1 Elementanzahl pro Person
```

```
MATCH (n:XmlTag)-[:RESPONSIBLE]->(p:Person)
```

```
RETURN n._name AS Elementname, p.name AS Editorname,
```

```
count(n._name) AS Elementanzahl ORDER BY Elementanzahl DESC;
```

Elementname	Editorname	Elementanzahl
"unclear"	"#BF"	337
"unclear"	"#CT"	134
"supplied"	"#CT"	7
"supplied"	"#TK"	5
"supplied"	"#BF"	2

Abb. 9: Häufigkeit der unclear- und supplied-Knoten der jeweiligen Edierenden. [Kasper / Kuczera 2019.]

Der Editor #BF hat in den fünf in der Graphdatenbank enthaltenen Dokumenten insgesamt 337 <unclear>- und nur zwei <supplied>-Elemente eingefügt, während der Editor #CT nur für 134 <unclear>- und für 7 <supplied>-Elemente verantwortlich ist.

Der folgende cypher query nimmt noch den Inhalt des @cert-Attributs hinzu.

```
// 2 Zweifel pro Person über alles
```

```
MATCH (n:XmlTag)-[:RESPONSIBLE]->(p:Person)
```

```
RETURN n._name AS Elementname, n.cert AS Zweifel, p.name AS Editorname, count(n._name)  
AS Elementanzahl ORDER BY Elementanzahl DESC;
```

Elementname	Zweifel	Editorname	Elementanzahl
"unclear"	"high"	"#BF"	190
"unclear"	"low"	"#BF"	147
"unclear"	"high"	"#CT"	76
"unclear"	"low"	"#CT"	58
"supplied"	"high"	"#CT"	7
"supplied"	"high"	"#TK"	5
"supplied"	"high"	"#BF"	2

Abb. 10: Häufigkeit der unclear- und supplied-Knoten mit Angabe des cert-Attributs. [Kasper / Kuczera 2019.]

Damit differenziert sich das Bild etwas, jedoch sind die Anteile von high- und low-Werten bei den jeweiligen Editoren im Durchschnitt gleich.³²

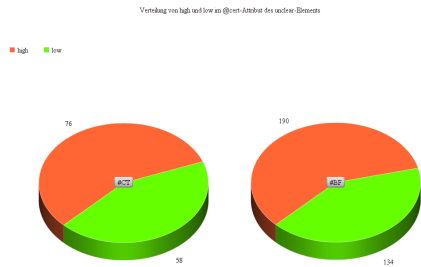


Abb. 11: Verteilung von high- und low-Werten sind beim cert-Attribut bei den Edierenden etwa gleich verteilt. [Kasper / Kuczera 2019.]

Mit dem nächsten query wird das Raster auf Dokumentenebene verfeinert:

```
// 5 Dokumente mit Bearbeitern, Elementen und Zweifeln
```

```
MATCH (n:XmlTag)-[:RESPONSIBLE]->(p:Person)
```

```
RETURN n.url AS Dokument, p.name AS Editorname, n._name AS Elementname, n.cert AS  
Zweifel, count(n._name) AS Elementanzahl ORDER BY Dokument, Elementname, Zweifel,  
Elementanzahl;
```

³² Die Grafik wurde erstellt mit dem Onlinetool [Diagrammwerkzeug](#).

Dokument	Elementname	Zweifel	Ediormame	Elementanzahl
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"unclear"	"high"	"BF"	153
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"unclear"	"low"	"BF"	116
"http://www.deutscher-testarchiv.de/books/download_ernstparthey_mngemq01711_1828"	"unclear"	"high"	"CT"	56
"http://www.deutscher-testarchiv.de/books/download_ernstparthey_mngemq01711_1828"	"unclear"	"low"	"CT"	38
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"high"	"BF"	36
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"low"	"BF"	31
"http://www.deutscher-testarchiv.de/books/download_ernsthubland_privatbestz_1829"	"unclear"	"high"	"CT"	9
"http://www.deutscher-testarchiv.de/books/download_ernsthubland_privatbestz_1829"	"unclear"	"low"	"CT"	8
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"unclear"	"low"	"CT"	7
"http://www.deutscher-testarchiv.de/books/download_ernsthubland_privatbestz_1829"	"supplied"	"high"	"CT"	5
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"high"	"CT"	5
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"high"	"CT"	4
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"supplied"	"high"	"CT"	3
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"low"	"CT"	3
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"supplied"	"high"	"CT"	2
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02345_1827"	"unclear"	"low"	"CT"	2
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"supplied"	"high"	"BF"	2
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq018442_1828"	"unclear"	"high"	"CT"	2
"http://www.deutscher-testarchiv.de/books/download_ernstparthey_mngemq01711_1828"	"supplied"	"high"	"CT"	1
"http://www.deutscher-testarchiv.de/books/download_ernstpatzig_mngemq01711_1828"	"unclear"	"high"	"BF"	1
"http://www.deutscher-testarchiv.de/books/download_ernst_mngemq02124_1827"	"supplied"	"high"	"CT"	1

Abb. 12: Verteilung der unclear- und supplied-Knoten mit Angaben zum Zweifel auf Dokumentenebene. [Kasper / Kuczera 2019.]

Es ist zu erkennen, dass #BF die Vorlesungsmitschrift von Patzig ediert hat und dabei die meisten <unclear>-Elemente vergeben hat. #CT hat in der Vorlesungsmitschrift von Parthey dagegen nur 94 <unclear>-Elemente verwendet. Dies könnte daran liegen, dass die Parthey-Vorlesungsmitschrift besser lesbar ist. Eine kurze Nachfrage beim Projekt Humboldt-Kosmos ergab aber, dass #BF wissenschaftliche Hilfskraft, #CT aber wissenschaftlicher Mitarbeiter ist. Die Ergebnisse der Tabelle könnten also auch zeigen, dass #CT die Handschriften besser lesen kann und deshalb weniger <unclear>-Elemente vergeben hat.

3.5 Verfeinerung des Profils

Im Folgenden wird die o. a. Auswertung in Kreisdiagrammen dargestellt. Die Farben der Legenden sind über alle Grafiken gleich. Im äußersten Ring werden die Anteile der Bearbeitenden an den in einem Dokument vergebenen <supplied>- und <unclear>-Elemente gezeigt.

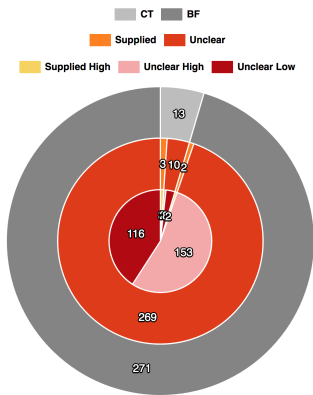


Abb. 13: Unsicherheitsverteilung auf Editorenebene bei Patzig. [Kasper / Kuczera 2019.]

Patzig

Bei der Transkription der Mitschrift von Patzig hat Benjamin Fiechter (BF) den größten Teil der <supplied>- und <unclear>-Elemente erstellt, ein kleinerer Teil wurde von Christian Thomas eingegeben. Im zweiten Ring ist zu erkennen, dass überwiegend <unclear>-Elemente vergeben wurden und dass Christian Thomas von den (wenigen) <supplied>-Elementen im Vergleich den größeren Teil eingetragen hat. Schließlich bleibt anzumerken, dass Benjamin Fiechter bei der Vergabe der <unclear>-Elemente im Verhältnis wesentlich öfter Zweifel hatte.

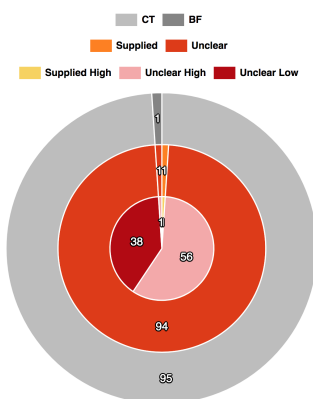


Abb. 14: Unsicherheitsverteilung auf Editorebene bei Parthey. [Kasper / Kuczera 2019.]

Parthey

In der Mitschrift von Parthey zeichnet ganz überwiegend Christian Thomas für die <supplied>- und <unclear>-Elemente verantwortlich, verwendet aber fast nur das <unclear>-Element. Die folgenden Kreisdiagramme zeigen noch die Verteilungen der Vorlesungsmitschriften NN1, NN2 und Hufeland.

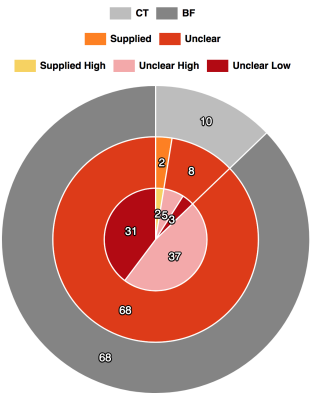


Abb. 15: Unsicherheitsverteilung auf Editorebene in NN2. [Kasper / Kuczera 2019.]

NN1

Die Mitschrift NN1 wurde überwiegend von Benjamin Fiechter ausgezeichnet, der kein unclear-Element verwendet. Christian Thomas annotiert dagegen zwei Stellen mit supplied-Elementen.

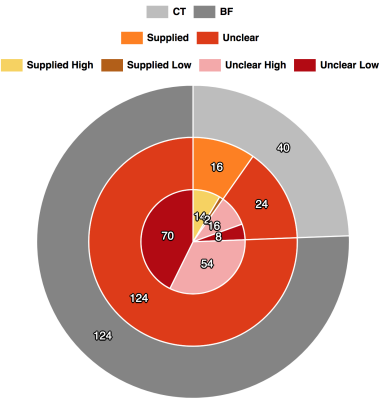


Abb. 16: Unsicherheitsverteilung auf Editorebene in NN2. [Kasper / Kuczera 2019.]

NN2

Ein ähnliches Bild ergibt sich für die Mitschrift NN2. Auch hier verwendet Benjamin Fiechter nur unclear-Elemente, während Christian Thomas auch einen Anteil an supplied-Elementen vergibt.

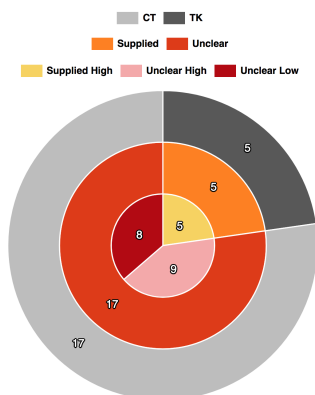


Abb. 17: Unsicherheitsverteilung auf Editorebene bei Hufeland. [Kasper / Kuczera 2019.]

Hufeland

Interessant scheint vor allem die Mitschrift Hufeland (Abbildung 17), bei der Christian Thomas fast alle <unclear>-Elemente eingetragen hat, während Benjamin Fiechter alle <supplied>-Elemente vergeben hat.³³ Eine sehr interessante Art von Arbeitsteilung, die sich so in keiner anderen Transkription findet. Ein kleiner Anteil der Bearbeitung wurde bei dieser Mitschrift auch von Tina Krell vorgenommen.

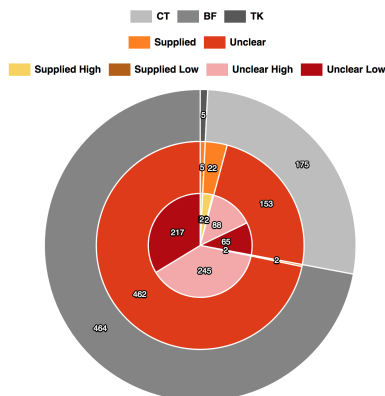


Abb. 18: Unsicherheitsverteilung auf Editorebene insgesamt. [Kasper / Kuczera 2019.]

³³ Zur Mitschrift von Otto Hufeland vgl. [Deutsches Textarchiv 2007–2019, Hufeland 1829](#).

Gesamt

In der letzten Grafik wurden alle Angaben noch einmal über alle Handschriften zusammengefasst. Es zeigt sich, dass Benjamin Fiechter einen großen Teil der <unclear>-Elemente vergeben hat, Christian Thomas einen kleineren, dafür aber fast alle <supplied>-Elemente.

Für die Erstellung eines persönlichen Auszeichnungsprofils von Edierenden wäre es am besten, verschiedene Edierende unabhängig voneinander die gleiche Quelle annotieren zu lassen und die Ergebnisse zu vergleichen. Liegen ausreichend Daten vor, wäre es denkbar, über Dokumentengrenzen hinweg persönliche Auszeichnungsprofile der Edierenden zu erstellen. Mit diesen Profilen könnten die verschiedenen, in der TEI möglichen Werte für die Attribute von Unsicherheit, näher bestimmt und möglicherweise auch vereinheitlicht werden.

4. Zusammenfassung

Die Nähe der Richtlinien von TEI einsetzenden Vorhaben wie DTA, PROPYLÄEN, und anderen, z. B. der Carl-Maria-von-Weber-Gesamtausgabe (WEGA), zu den TEI-Guidelines macht die Daten dieser Editionen im Bereich des allgemeinen Umgangs mit Textlücken und unsicheren Lesarten (auch mit XML-Mitteln) vergleichbar. Gleichzeitig erleichtert diese Nähe auch die Entwicklung von spezielleren TEI-Import-Routinen für Neo4j.

Schwieriger ist dies jedoch für den Vergleich von Gewichtungen in der Sicherheit (Attribut @cert) der Auflösung von unsicheren Lesarten oder Textergänzungen. Hier spielt die subjektive Entscheidung des Edierenden eine zentrale Rolle. Die Angaben im Attribut @resp lassen sich hier allerdings heranziehen, um einen Eindruck zu bekommen, wie Edierende in welchen Fällen gewichtet.³⁴ Stehen genügend Daten zur Verfügung, könnte aus den Annotationen ein persönliches Auszeichnungsprofil des Bearbeitenden erstellt werden.

³⁴ Zugleich wird damit auch eine Anforderung an digitale Editionen erfüllt, nämlich: »Every act of editing in a digital edition should be attributed explicitly to the person who did it.« Robinson 2013.

Bibliographische Angaben

Marjorie Burghart / Malte Rehbein: The Present and Future of the TEI Community for Manuscript Encoding. In: Journal of the Text Encoding Initiative (2012), H. 2. Artikel vom 03.02.2012. DOI: [10.4000/jtei.372](https://doi.org/10.4000/jtei.372)

Carl-Maria-von-Weber-Gesamtausgabe. Digitale Edition. Editionsrichtlinien zur Ausgabe der Briefe, Tagebücher und Dokumente Webers. Hg. von Gerhard Allroggen. Version 3.2.1 vom 08.01.2018. [\[online\]](#)

Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Hg. von der Berlin-Brandenburgischen Akademie der Wissenschaften, Berlin 2007–2019. [\[online\]](#)

DTA-Basisformat. Das von CLARIN-D und der DFG empfohlene TEI-Format für historische Texte. Hg. vom Zentrum Sprache an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), Berlin 2011–2018. [\[online\]](#), hier besonders DTA-Basisformat Manuskript: [\[online\]](#)

Johann Wolfgang Goethe. Tagebücher. Hg. von Wolfgang Albrecht / Andreas Döhler. Band I,1. 1775–1787. Bisher unveröffentlichte retrodigitalisierte Datenfassung. Druckfassung: Stuttgart 1998.

Andreas Kuczera: Graphentechnologien in den digitalen Geisteswissenschaften. Modellierung – Import – Analyse. Github Pages. August 2018–. [\[online\]](#)

Andreas Kuczera (2017a): Graphentechnologien in den Digitalen Geisteswissenschaften. In: ABI Technik 37 (2017) H. 3. 15.09.2017. DOI: [10.1515/abitech-2017-0042](https://doi.org/10.1515/abitech-2017-0042)

Andreas Kuczera (2017b): Das Deutsche Textarchiv in der Graphenwelt. In: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte. Blogbeitrag vom 04.04.2017, aktualisiert am 06.06.2017. [\[online\]](#)

Gotthilf Patzig: Vorträge über physische Geographie des Freiherrn Alexander von Humboldt: gehalten im großen Hörsaal des Universitäts-Gebäudes zu Berlin im Wintersemester 1827/28 vom 3ten Novbr. 1827. bis 26 April 1828. Aus schriftlichen Notizen nach jedem Vortrage zusammengestellt vom Rechnungsrath Gotthilf Friedrich Patzig. Berlin 1827/28 (= Nachschrift der „Kosmos-Vorträge“ Alexander von Humboldts in der Berliner Universität, 3.11.1827–26.4.1828), S. 9. In: Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Hg. von der Berlin-Brandenburgischen Akademie der Wissenschaften. Berlin 2007–2019. [\[online\]](#)

Peter Robinson: Five Desiderata for Scholarly Editions in Digital Form. In: Proceedings of Digital Humanities (University of Nebraska–Lincoln, 16.–19.07.2013). Long Paper vom 19.07.2013. [\[online\]](#)

TEI Guidelines. P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.4.0. Revision 1fa0b54 vom 23.07.2018. [\[online\]](#)

Christian Thomas / Benjamin Fiechter / Marius Hug: Methoden und Ziele der Erschließung handschriftlicher Quellen zu Alexander von Humboldts Kosmos-Vorträgen. Das Projekt Hidden Kosmos der Humboldt-Universität zu Berlin. In: Horizonte der Humboldtforchung: Natur, Kultur, Schreiben. Hg. von Ottmar Ette / Julian Drews. Hildesheim u. a. 2016, S. 287–318. (= Potsdamer inter- und transkulturelle Texte (Pointe), 16). [\[Nachweis im GBV\]](#) Siehe auch Preprint PDF [\[online\]](#)

Abbildungsverzeichnis

Abb. 1: TEI-Modules used in manuscript encoding projects (apart from the four basic ones: core, tei, header, textstructure). [Burghart / Rehbein 2012, Fig. 11. CC BY-ND 3.0.]

Abb. 2: In what areas do you wish the Guidelines to be improved? Figures are given in absolute numbers. [Burghart / Rehbein 2012, Fig. 18. CC BY-ND 3.0.]

Abb. 3: Beispiel zur Graphmodellierung eines unclear-Elements. [Kasper / Kuczera 2019.]

Abb. 4: Das unclear-Beispiel in der XML-Ansicht des DTA. [Kasper / Kuczera 2019.]

Abb. 5: Die properties des unclear-Knotens. [Kasper / Kuczera 2019.]

Abb. 6: Die gekürzt wiedergegebene Tabelle zeigt die häufigsten Ergebnisse. Die Angaben in der ersten Spalte der Tabelle wurden aus Gründen der Übersichtlichkeit um den URL-Teil gekürzt, der bei allen Mitschriften gleich ist. [Kasper / Kuczera 2019.]

Abb. 7: Aufschlüsselung der Edierendenkürzel. [Kasper / Kuczera 2019.]

Abb. 8: unclear- und supplied-Knoten, die von #CT erstellt worden sind. [Kasper / Kuczera 2019.]

Abb. 9: Häufigkeit der unclear- und supplied-Knoten der jeweiligen Edierenden. [Kasper / Kuczera 2019.]

Abb. 10: Häufigkeit der unclear- und supplied-Knoten mit Angabe des cert-Attributs. [Kasper / Kuczera 2019.]

Abb. 11: Verteilung von high- und low-Werten sind beim cert-Attribut bei den Edierenden etwa gleich verteilt. [Kasper / Kuczera 2019.]

Abb. 12: Verteilung der unclear- und supplied-Knoten mit Angaben zum Zweifel auf Dokumentebene. [Kasper / Kuczera 2019.]

Abb. 13: Unsicherheitsverteilung auf Editorenebene bei Patzig. [Kasper / Kuczera 2019.]

Abb. 14: Unsicherheitsverteilung auf Editorenebene bei Parthey. [Kasper / Kuczera 2019.]

Abb. 15: Unsicherheitsverteilung auf Editorenebene in NN1. [Kasper / Kuczera 2019.]

Abb. 16: Unsicherheitsverteilung auf Editorenebene in NN2. [Kasper / Kuczera 2019.]

Abb. 17: Unsicherheitsverteilung auf Editorenebene bei Hufeland. [Kasper / Kuczera 2019.]

Abb. 18: Unsicherheitsverteilung auf Editorenebene insgesamt. [Kasper / Kuczera 2019.]