

Named Entity Uncertainty Mining: Von der intellektuellen zur computergestützten Untersuchung unsicherer Annotationen

Schneider, Sophie

sophie.schneider@sbb.spk-berlin.de
Staatsbibliothek zu Berlin, Deutschland
ORCID: 0000-0002-8303-1798

Bubula, Michal

michal.bubula@sbb.spk-berlin.de
Staatsbibliothek zu Berlin, Deutschland
ORCID: 0009-0004-6586-1319

Labusch, Kai

kai.labusch@sbb.spk-berlin.de
Staatsbibliothek zu Berlin, Deutschland
ORCID: 0000-0002-7275-5483

Lehmann, Jörg

joerg.lehmann@sbb.spk-berlin.de
Staatsbibliothek zu Berlin, Deutschland
ORCID: 0000-0003-1334-9693

Neudecker, Clemens

clemens.neudecker@sbb.spk-berlin.de
Staatsbibliothek zu Berlin, Deutschland
ORCID: 0000-0001-5293-8322

Die Nutzungsszenarien von Named Entities (NE) erstrecken sich in den Digital Humanities von der semantischen Anreicherung digitalisierter Texte (Düring et al., 2022; Pfanzelter et al., 2021; Ehrmann et al., 2019) über weiterführende Untersuchungen mittels maschineller Verfahren wie Opinion Mining oder Topic Modeling (Labusch et al., 2024) bis hin zur Historischen Netzwerkanalyse (Bludau et al., 2022; Bludau et al., 2020). Es handelt sich dabei um einzelne Token bzw. eine Abfolge von Token im Text, die im Sinne des "rigid designator" nach Kripke (1980) einen eindeutigen Namen tragen, beispielsweise Eigennamen. Named Entities werden mittels computergestützter Verfahren wie der Named Entity Recognition (NER, s.a. Nadeau/Sekine, 2007) erkannt und z.B. als Personen, Orte oder Organisationen klassifiziert. Beim Entity Linking (EL) werden die Entitäten disambiguiert und mit Wissensbasen wie

der Gemeinsamen Normdatei oder Wikidata verknüpft. Die Notwendigkeit einer Formalisierung für maschinelle Verfahren (vgl. Piotrowski, 2019; Rapp, 2017), etwa durch die Klassifizierung in vorab festgelegte Tagsets, führt jedoch zwangsweise dazu, dass im Falle einer Unsicherheit bei der Annotation eine unzulängliche Entscheidung zugunsten simulierter Eindeutigkeit getroffen werden muss. Auch Annotationsrichtlinien, die üblicherweise in einem iterativen Verfahren u.a. anhand der Besprechung solcher Unzulänglichkeiten fortlaufend verbessert werden (Reiter, 2020) und unter Umständen bereits sehr detaillierte Anweisungen enthalten, helfen bei dieser Problemstellung nur bedingt. Bei der Anwendung endlicher Regelwerke auf komplexe Gegenstände (z.B. historische Zeitungstexte) können immer wieder neue Fälle auftreten, die durch die Regeln nicht abgedeckt oder gar durch eindeutige Festlegungen entscheidbar sind. Dementsprechend kommt es in der Praxis zu Abweichungen und Unsicherheiten, die hinsichtlich der Vergleichbarkeit und Nachnutzbarkeit Probleme aufwerfen.

Es existiert im Kontext geisteswissenschaftlicher Daten keine allgemeingültige Definition von Unsicherheit, allerdings wurden in verschiedenen Anwendungsbereichen der Digital Humanities Versuche einer Annäherung an den Begriff unternommen (vgl. z.B. Windhager et al., 2019; Edmond, 2018; Boukhelifa et al., 2017). Im Kontext von Annotationen besteht Unsicherheit dort, wo den Annotierenden keine eindeutige Zuordnung möglich ist. Unsicherheit entspricht in diesem Sinne einem höchst subjektiven Konzept, und die erzeugten Annotationen sind als konstruierte "capta" statt gegebener Fakten zu verstehen (Drucker, 2011).

Dieser Beitrag führt in die Problematik der Nutzung unsicherer historischer Daten für das Training maschineller Verfahren ein, diskutiert Ursachen und Arten von Unsicherheiten in den für NER/EL zu annotierenden Daten und zeigt die Möglichkeiten computergestützter Unsicherheitsmessung auf. Auf der Grundlage einer umfassenden Literaturrecherche unternehmen wir eine Theoriereflexion, um die folgende Frage zu beantworten: Wie lassen sich Annotationsrichtlinien für historische NER/EL-Verfahren optimieren und die bei der Annotation auftretenden Unsicherheiten minimieren?

Entscheidungen bei der Annotation

Die Annotation von Named Entities erweist sich aufgrund kontextabhängiger Bedeutungen oder der Verwendung stilistischer Mittel (z.B. Synonyme, Metonymien, Eponomastien) als problematisch. Unsicherheiten bezüglich der Identifizierung als Named Entity und der Klassifizierung in NE-Tags ergeben sich beispielsweise bei der Unterscheidung zwischen realen und fiktiven Entitäten, Familiennamen und Geburtsorten sowie Orten und Organisationen (Provatorova et al., 2024; Álvarez-Mellado et al., 2021; Heino et al., 2017). Ist der *Päpstliche Stuhl* eine Named Entity, eine Organisation? Taggt man das *Weißes Haus* oder den *Dreibund*

als Organisation oder als Ort? Lässt sich *Hegel'sche* als Person auszeichnen?

Eine in Umfang und Grenzen korrekte Segmentierung von Entitäten stellt ebenfalls eine Herausforderung dar. Gehört bei *F. Butzke & Co., Aktiengesellschaft für Metallindustrie* die *Aktiengesellschaft für Metallindustrie* noch zum Namen der Entität? Wie wird bei Funktionen und Titeln wie *Fürst* oder *Dr.* vorgegangen? Abseits der Urteilkraft einzelner Annotierender können auch vorangegangene Prozessierungsschritte wie die Layout- und Texterkennung sowie Token-/Satzsegmentierung die Annotation hierbei erschweren. Sehen die Annotationsrichtlinien eine Verschachtelung (Nesting) von Entitäten vor, steigt die Granularität von Annotationen – dies kann sich im Hinblick auf das Auftreten von Unsicherheiten sowohl positiv als auch negativ auswirken (Rücker/Akbik, 2023; Monroc et al., 2022; Álvarez-Mellado et al., 2021).

Sowohl die Disambiguierung als auch das daran anschließende Linking von Entitäten erfolgt unter Zuhilfenahme externer Datenquellen. Hierbei wird bestimmt, ob in der verwendeten Wissensbasis ein Eintrag zur jeweiligen Entität enthalten ist. Aufgrund von Ambiguitäten können mehrere Einträge infrage kommen, sodass ein Verfahren Wahrscheinlichkeiten für die Liste an Kandidaten für die Verknüpfung generieren muss (Labusch/Neudecker, 2020). Sind in der Datenbank Einträge vorhanden, die nur teilweise zutreffen (da sie zu spezifisch oder abstrakt sind), oder Duplikate, aus denen eine Auswahl getroffen werden muss (Ling et al., 2015), ist es für Annotierende nicht immer leicht zu entscheiden, welcher Eintrag in einer externen Ressource die gegebene Entität am besten beschreibt. Zum Teil liefert ein Text keine ausreichenden Hinweise darauf, welche Person innerhalb eines tief verzweigten Adelsgeschlechts oder welche Regierung (Auf welches Kabinett Deladier (I - V) bezieht sich *Regierung Daladier*?) genau gemeint sein könnte. Eine stilistische Verwendung von Sprache erschwert den Prozess der Zuordnung zusätzlich: Wenn von *diesem Saulus-Paulus* die Rede ist, kann zwar auf Paulus von Tarsus verlinkt werden, gemeint ist allerdings im Sinne der Redewendung *vom Saulus zum Paulus*¹ eine andere Entität.

Original und Bearbeitungen

Bereits im Original kann eine historische Quelle Fehler enthalten und/oder der natürlichen Materialzersetzung unterliegen (Blouin et al., 2024; Hamdi et al., 2020). Die spätere Edition und Transkription historischer bzw. handschriftlicher Texte ist an eine Interpretation gebunden, welche auf falschen Annahmen aufbauen oder letztlich zu solchen führen kann (Dereza et al., 2023; Cameron et al., 2022). Weitere Verarbeitungsschritte, die NER und EL primär ermöglichen, sich allerdings auch als fehleranfällig und damit ursächlich für die Entstehung von Unsicherheiten erweisen können, sind:

- der Digitalisierungs-/Scanprozess

- Optical Character Recognition (OCR) bzw. Handwritten Text Recognition (HTR) (Monroc et al., 2022; Hamdi et al., 2020)
- NER als Voraussetzung für EL

Kontext und Beschreibung

Texte sind Produkte ihres Kontexts, ihrer Zeit und Regionalität, und die enthaltenen Bezüge sind im Nachhinein oftmals nur schwer zu rekonstruieren (Ardanuy et al., 2023). Der Verlust entsprechenden Kontextwissens stellt eine Herausforderung für die Disambiguierung und Verknüpfung von Entitäten dar, z. B. wenn der Titel oder die Funktion einer Person ohne Namen, ein Ereignis ohne genaues Datum oder ein Familienname ohne zusätzliche Informationen gegeben wird (Munnelly/Lawless, 2018; Heino et al., 2017). Metadaten können Auskunft zur Provenienz eines Objekts sowie dessen Zugangs-/Nutzungsmöglichkeiten geben, fehlende oder unvollständige Metadaten hingegen können Unsicherheiten erzeugen oder verstärken.

Die Qualität und Abdeckung der Beschreibung von Entitäten in Wissensbasen (z.B. Wikidata, DBpedia), Normdateien (z. B. GND, s.a. Rettinghaus, 2024) oder domänenspezifischen Ontologien haben ebenfalls Einfluss auf die Entstehung von Unsicherheiten bei der Annotation. Die Ressourcen decken nicht alle Sprachen gleichermaßen ab und Kategorien passen teilweise nicht zu definierten NE-Klassen (Linhares Pontes et al., 2020). Gerade in historischen Texten erwähnte Entitäten sind in modernen Wissenssystemen nicht enthalten (sogenannte NIL oder OOKB (=out-of-knowledge-base) Entities) oder unvollständig bzw. fehlerhaft repräsentiert (Clematide et al., 2023; Ehrmann et al., 2022; Heino et al., 2017; Munnelly/Lawless, 2018; Ling et al., 2015). Die Dynamiken community-basierter Änderungen erschweren zudem langfristige Evaluierungsstudien (Menzel et al., 2021).

Diachroner Wandel

Ressourcen, die für NER und EL verwendet werden, sind nicht statisch und können sich im Laufe der Zeit wandeln. Vor allem im Rahmen digitaler Editionen führt jede Bearbeitung neue Ergänzungen, Streichungen oder andere Änderungen in ein Dokument ein. Grundsätzlich handelt es sich um etablierte Editionspraktiken, die allerdings einem diachronen Wandel sowie individuellem Einfluss unterliegen und selten in den Metadaten dokumentiert werden (Dereza et al., 2023).

Sprachliche Standards entwickeln sich im Laufe der Zeit. So weisen historische Texte unzählige sprachliche Varianten auf, auch bei der Schreibung von Entitätsnamen (Provatorova et al., 2024; Ehrmann et al., 2023; Cameron et al., 2022). Die Mehrsprachigkeit von Dokumenten erzeugt weitere Variation, z. B. wenn Entitätsnamen oder andere Wörter und Phrasen nicht mit der Hauptsprache des Tex-

tes übereinstimmen (Ehrmann et al., 2022; Hamdi et al., 2021; Linhares Pontes et al., 2020). Da sich orthografische Konventionen und die Verwendung oder Bedeutung von Wörtern diachron entwickeln, können zudem Diskrepanzen zwischen der historischen Sprache eines Textes und der in modernen Ressourcen wie Wissensdatenbanken oder Benchmark-Datensätzen verwendeten Sprache beobachtet werden (Dereza et al., 2023; Lai et al., 2021; Menzel et al., 2021).

Ein verwandter Fall ist der sogenannte “Entity Drift”, welcher das Verschwinden oder Auftauchen von Entitäten im Laufe der Zeit beschreibt (Ehrmann et al., 2023). Entitäten können beispielsweise ihren Namen ändern und sich dennoch in ihrer Bedeutung weiterhin auf dieselbe Entität beziehen, z. B. beim Wechsel des Familiennamens nach einer Heirat. Zugleich kann sich die Bedeutung oder der Umfang ein- und derselben Entität über die Zeit ändern, z. B. bei der Übertragung von Titeln, bei der Veränderung von Landesgrenzen oder bei Umstrukturierungen von Armeeeinheiten (Munnelly/Lawless, 2018; Heino et al., 2017).

Praktiken und Ressourcen

Gängige Verfahren und Werkzeuge können Unsicherheiten verstetigen, z.B. wenn keine Optionen zur expliziten Annotation und Einordnung von Unsicherheiten in entsprechenden Werkzeugen vorhanden sind (Hägele et al., 2022).

Allgemeingültige Standards für NER und EL auf der Grundlage historischer Texte bzw. Sprachen existieren nicht. Während die Annotation in der Regel von entsprechenden Annotationsrichtlinien begleitet wird, sind die darin definierten Anweisungen und Entitätstypen oder -attribute meist auf eine bestimmte Forschungsfrage oder Domäne zugeschnitten (González-Gallardo et al., 2023). Teilweise überschneiden sich Annotationsrichtlinien zwar, in anderen Fällen jedoch erweisen sich Annotationen als nicht kompatibel (Ehrmann et al., 2022). Um mehr Kompatibilität und Generalisierbarkeit zwischen verschiedenen annotierten NER/EL-Datensätzen zu ermöglichen (Ehrmann et al., 2023; Hamdi et al., 2021), können Mappings zwischen verschiedenen Tagsets erstellt werden, wobei auch hiermit Informationsverlust einhergehen kann (Palladino/Yousef, 2024).

Ressourcen für die Entwicklung und Evaluation von NER/EL-Systemen wie z. B. Benchmark-Datensätze sind für historische Daten nur spärlich vorhanden, nicht zuletzt aufgrund der hohen Variabilität in Domänen und Sprachen (Ehrmann et al., 2023; Hamdi et al., 2021; Lai et al., 2021). Es zeichnet sich ein Bias zugunsten bestimmter Jahrhunderte und Sprachen in Tools und Korpora ab (Blouin et al., 2024; Palladino/Yousef, 2024; Provatorova et al., 2024; Novotny et al., 2023). Die Verfügbarkeit von und der Zugang zu (teilweise digitalisierten) Sammlungen stellen weitere Herausforderungen dar (González-Gallardo et al., 2023).

Unsicherheiten computergestützt messen

Wie oben dargestellt, sind Annotationen von Named Entities nicht immer klar und eindeutig vorzunehmen. Bei der Annotation durch mehrere Annotierende ergeben sich daher schnell Unstimmigkeiten, und solche “Human Label Variations” (Peng et al., 2024; Plank, 2022) können ein praktikabler Indikator für Unsicherheit sein. Mithilfe des Inter Annotator Agreements und Metriken wie Cohen's Kappa lässt sich die Übereinstimmung zwischen mehreren Annotierenden quantifizieren und somit unsichere Annotationen identifizieren (Rapp, 2017). Während sich die Aussagekraft der Annotationen über zusätzliche Annotationen bzw. ein entsprechendes Re-Labeling steigern lässt (Rücker/Akbik, 2023; Hamdi et al., 2021), kann der damit verbundene Aufwand ein Hindernis bei der Optimierung solcher Datensätze darstellen.

Über die vorgestellten Analysetechniken hinaus existieren weitere computergestützte Ansätze zur Identifizierung von Unsicherheiten in entsprechenden Daten. Mittels Bayesscher Verfahren können differenzierte Metriken zur Abschätzung von Unsicherheit generiert werden (vgl. z.B. Zhang et al., 2024; Miok et al., 2020; Xiao/Wang, 2019). Ein verbreitetes Maß ist hierbei die Monte-Carlo-Dropout-Methode (Gal/Ghahramani, 2016; Srivastava et al., 2014). Bei dieser werden im Unterschied zur Dropout-Regularisierungstechnik die Bernoulli-verteilten Dropout-Masken eines Deep Neural Networks auch zur Testzeit angewandt. Nach mehrmaligen Durchläufen unter verschiedenen Masken erhält man eine Verteilung, deren Mittelwert als Klassifikationsvorhersage und deren Varianz als Unsicherheitsmaß der Klassifikation gedeutet werden können.

Im Anschluss an eine erfolgte Identifizierung von Unsicherheiten in den Referenzdaten können dann auf der Grundlage vorliegender Auswertungen intellektuelle Untersuchungen der Annotationssicherheit erfolgen: Auf welche der hier vorgestellten Arten und Ursachen von Unsicherheit verteilen sich die fraglichen Annotationen? Wann lassen sich Entitäten besonders schwer bestimmten Klassen zuordnen oder eindeutig verlinken? Gibt es wiederholt auftretende Fälle von Unsicherheit? Bezüglich der Annotationsrichtlinien lässt sich fragen: Können anhand der beobachteten Unsicherheiten die zugrunde liegenden Richtlinien identifiziert werden? Lassen sich aus den gemessenen Unsicherheiten schwer umzusetzende Anweisungen in den Annotationsrichtlinien ableiten? Können Untermengen identifiziert werden, die trotz unterschiedlicher Annotationsrichtlinien konsistent und zuverlässig klassifiziert werden?

Ausblick

In diesem Beitrag haben wir den aktuellen Forschungsstand zur Entstehung von Unsicherheiten in historischen Daten referiert. Wir planen, die vorgestellten Bayesschen Verfahren auf einem umfangreichen Datensatz zu erproben, die Ergebnisse in einer Publikation zu präsentieren und die gemessenen Unsicherheiten weiter zu diskutieren. Wir hoffen, auf dieser Grundlage Vorschläge für die Optimierung von Annotationsrichtlinien zu erarbeiten, so dass Unsicherheiten reduziert und die Kompatibilität entsprechender Datensätze verbessert werden können. Die hier ausgeführte Theoriereflexion dient dafür als Vorarbeit.

Danksagung

Diese Arbeit wurde teilweise unterstützt durch die Beauftragte der Bundesregierung für Kultur und Medien (BKM), Projektförderung "Mensch.Maschine.Kultur - Künstliche Intelligenz für das Digitale Kulturelle Erbe" (2522DIG002).

Fußnoten

1. S.a. <https://de.wikipedia.org/wiki/Damaskuserlebnis#Redewendung> (zugegriffen: 24.07.2024).

Bibliographie

Álvarez-Mellado, Elena, María Luisa Díez-Platas, Pablo Ruiz-Fabo, Helena Bermúdez, Salvador Ros und Elena González-Blanco. 2021. "TEI-friendly annotation scheme for medieval named entities: a case on a Spanish medieval corpus." *Language Resources and Evaluation* 55, Nr. 2: 525–549. <https://doi.org/10.1007/s10579-020-09516-2>.

Ardanuy, Mariona Coll, Federico Nanni, Kaspar Beelen und Luke Hare. 2023. "The Past is a Foreign Place: Improving Toponym Linking for Historical Newspapers." In *CHR 2023*. Paris, France. <https://ceur-ws.org/Vol-3558/paper4426.pdf> (zugegriffen: 15. Juli 2024).

Blouin, Baptiste, Cécile Armand und Christian Henriot. 2024. "A Dataset for Named Entity Recognition and Entity Linking in Chinese Historical Newspapers." In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, hg. von Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, und Nianwen Xue, 385–394. Torino, Italia: ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.35> (zugegriffen: 4. Juni 2024).

Bludau, Mark-Jan, Marian Dörk, Heiner Fangerau, Thorsten Halling, Elena Leitner, Sina Menzel, Gerhard Müller, Vivien Petras, Georg Rehm, Clemens

Neudecker, David Zellhoefer und Julian Moreno Schneider. 2020. "SoNAR (IDH): Datenschnittstellen für historische Netzwerkanalyse." In *DHd 2020*. <https://doi.org/10.5281/ZENODO.4621861>.

Bludau, Mark-Jan, Thorsten Halling, Eva Maria Holly, Jasmin Wieloch, Hannes Schnaitter, Sandra Balck, Melina Plakidis, Georg Rehm, Heiner Fangerau und Marian Dörk. 2022. "Was sehe ich? Visualisierungsstrategien für Datentransparenz in der Historischen Netzwerkanalyse." In *DHd 2022*. <https://doi.org/10.5281/ZENODO.6327930>.

Boukhelifa, Nadia, Marc-Emmanuel Perrin, Samuel Huron und James Eagan. 2017. "How Data Workers Cope with Uncertainty: A Task Characterisation Study." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3645–3656. Denver Colorado USA: ACM. <https://doi.org/10.1145/3025453.3025738>.

Cameron, Helena Freire, Fernanda Olival, Renata Vieira und Joaquim Santos. 2022. "Named entity annotation of an 18th-century transcribed corpus: problems and challenges." In *DHandNLP@PROPOR 2022*. <https://ceur-ws.org/Vol-3128/paper8.pdf> (zugegriffen: 12. Juli 2024).

Clematide, Simon, Mariona Coll-Ardanuy und Yves Maurer. 2023. "Tracking Discourses on Public and Hidden People in Historical Newspaper." In *Computational Approaches to Digitised Historical Newspapers (Dagstuhl Seminar 22292)*, hg. von Maud Ehrmann, Marten Düring, Clemens Neudecker und Antoine Doucet, 127–138. <https://doi.org/10.4230/DAGREP.12.7.112>.

Dereza, Oksana, Theodorus Fransen und John P. McCrae. 2023. "Do not Trust the Experts - How the Lack of Standard Complicates NLP for Historical Irish." In *Proceedings of the Fourth Workshop on Insights from Negative Results in NLP*, hg. von Shabnam Tafreshi, Arjun Akula, João Sedoc, Aleksandr Drozd, Anna Rogers, und Anna Rumshisky, 82–87. Dubrovnik, Croatia: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.insights-1.10>.

Drucker, Johanna. 2011. "Humanities approaches to graphical display." *Digital Humanities Quarterly* 5, Nr. 1: 1–21. <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (zugegriffen: 17. Juli 2024).

Düring, Marten, Estelle Bunout und Daniele Guido. 2022. *Transparent Generosity: Introducing the impresso interface for the exploration of semantically enriched historical newspapers*. <https://hal.science/hal-04154431> (zugegriffen: 12. Juli 2024).

Edmond, Jennifer. 2018. "Managing Uncertainty in the Humanities: Digital and Analogue Approaches." In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 840–844. Salamanca, Spain: ACM. <https://doi.org/10.1145/3284179.3284326>.

Ehrmann, Maud, Estelle Bunout und Düring. 2019. "Historical Newspaper User Interfaces: A Review." In *IFLA WLIC 2019*. <https://doi.org/10.5281/ZENODO.3404155>.

- Ehrmann, Maud, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello und Antoine Doucet.** 2023. "Named Entity Recognition and Classification in Historical Documents: A Survey." *ACM Computing Surveys* 56, Nr. 2: 1–47. <https://dl.acm.org/doi/10.1145/3604931>.
- Ehrmann, Maud, Matteo Romanello, Sven Najem-Meyer, Antoine Doucet und Simon Clematide.** 2022. "Extended Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents." In *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum*. <https://doi.org/10.5281/zenodo.6979577>.
- Gal, Yarin und Zoubin Ghahramani.** 2016. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning." In *Proceedings of The 33rd International Conference on Machine Learning*, 48: PMLR. <https://proceedings.mlr.press/v48/gal16.html> (zugegriffen: 4. Juni 2024).
- González-Gallardo, Carlos-Emiliano, Emanuela Boros, Nancy Girdhar, Ahmed Hamdi, Jose G. Moreno und Antoine Doucet.** 2023. "Yes but.. Can ChatGPT Identify Entities in Historical Documents?" In *2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 184–189. <https://doi.org/10.1109/JCDL57899.2023.00034>.
- Hägele, David, Christoph Schulz, Cedric Beschle, Hannah Booth, Miriam Butt, Andrea Barth, Oliver Deussen und Daniel Weiskopf.** 2022. "Uncertainty visualization: Fundamentals and recent developments." *IT - Information Technology* 64, Nr. 4–5: 121–132. <https://doi.org/10.1515/itit-2022-0033>.
- Hamdi, Ahmed, Axel Jean-Caurant, Nicolas Sidère, Mickaël Coustaty und Antoine Doucet.** 2020. "Assessing and Minimizing the Impact of OCR Quality on Named Entity Recognition." In *Digital Libraries for Open Knowledge*, hg. von Mark Hall, Tanja Merčun, Thomas Risse, und Fabien Duchateau, 87–101. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-54956-5_7.
- Hamdi, Ahmed, Elvys Linhares Pontes, Emanuela Boros, Thi Tuyet Hai Nguyen, Günter Hackl, Jose G. Moreno und Antoine Doucet.** 2021. "A Multilingual Dataset for Named Entity Recognition, Entity Linking and Stance Detection in Historical Newspapers." In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2328–2334. <https://doi.org/10.1145/3404835.3463255>.
- Heino, Erkki, Minna Tamper, Eetu Mäkelä, Petri Leskinen, Esko Ikkala, Jouni Tuominen, Mikko Koho und Eero Hyvönen.** 2017. "Named Entity Linking in a Complex Domain: Case Second World War History." In *Language, Data, and Knowledge*, hg. von Jorge Gracia, Francis Bond, John P. McCrae, Paul Buitelaar, Christian Chiarcos, und Sebastian Hellmann, 120–133. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59888-8_10.
- Kripke, Saul A.** 1980. *Naming and necessity*. Bd. 217. Oxford: Blackwell.
- Labusch, Kai und Clemens Neudecker.** 2020. "Named Entity Disambiguation and Linking on Historic Newspaper OCR with BERT." In *Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum*, hg. von Linda Cappellato, Carsten Eickhoff, Nicola Ferro, Aurélie Névoul, CEUR-WS Vol. 2696. https://ceur-ws.org/Vol-2696/paper_163.pdf (zugegriffen: 24. Juli 2024).
- Labusch, Kai, Sophie Schneider und Clemens Neudecker.** 2024. "Automatisierte semantische Anreicherung von historischen Texten. Erkennung und Verknüpfung von Entitäten mit Wikidata und Wikipedia." *B.I.T.online*, Nr. 3. <https://www.b-i-t-online.de/heft/2024-03-fachbeitrag-labusch.pdf> (zugegriffen: 12. Juli 2024).
- Lai, Viet Dac, Minh Van Nguyen, Heidi Kaufman und Thien Huu Nguyen.** 2021. "Event Extraction from Historical Texts: A New Dataset for Black Rebellions." In *Findings of the Association for Computational Linguistics*, 2390–2400. Association for Computational Linguistics. <https://aclanthology.org/2021.findings-acl.211.pdf> (zugegriffen: 15. Juli 2024).
- Ling, Xiao, Sameer Singh und Daniel S. Weld.** 2015. "Design Challenges for Entity Linking." *Transactions of the Association for Computational Linguistics* 3: 315–328. https://doi.org/10.1162/tac1_a_00141.
- Linhares Pontes, Elvys, Luis Adrián Cabrera-Diego, Jose G. Moreno, Emanuela Boros, Ahmed Hamdi, Nicolas Sidère, Mickaël Coustaty und Antoine Doucet.** 2020. "Entity Linking for Historical Documents: Challenges and Solutions." In *Digital Libraries at Times of Massive Societal Transition*, hg. von Emi Ishita, Natalie Lee San Pang, und Lihong Zhou, 215–231. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-64452-9_19.
- Menzel, Sina, Hannes Schnaitter, Josefine Zinck, Vivien Petras, Clemens Neudecker, Kai Labusch, Elena Leitner und Georg Rehm.** 2021. "Named Entity Linking mit Wikidata und GND – Das Potenzial handkuratierter und strukturierter Datenquellen für die semantische Anreicherung von Volltexten." In *Qualität in der Inhaltsschließung*, hg. von Michael Franke-Maier, Anna Kasprzik, Andreas Ledl, und Hans Schürmann, 229–258. De Gruyter. <https://doi.org/10.1515/9783110691597-012>.
- Miok, Kristian, Gregor Pirs und Marko Robnik-Sikonja.** 2020. "Bayesian Methods for Semi-supervised Text Annotation." In *Proceedings of the 14th Linguistic Annotation Workshop*, hg. von Stefanie Dipper und Amir Zeldes, 1–12. Barcelona, Spain: Association for Computational Linguistics. <https://aclanthology.org/2020.law-1.1> (zugegriffen: 4. Juni 2024).
- Monroc, Claire Bizon, Blanche Miret, Marie-Laurence Bonhomme und Christopher Kermorvant.** 2022. "A Comprehensive Study of Open-Source Libraries for Named Entity Recognition on Handwritten Historical Documents." In *Document Analysis Systems*, hg. von Seiichi Uchida, Elisa Barney, und Véronique

Eglin. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-06555-2_29.

Munnely, Gary und Seamus Lawless. 2018. "Investigating Entity Linking in Early English Legal Documents." In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2018*. Fort Worth, Texas, USA. <http://www.tara.tcd.ie/handle/2262/86854> (zugegriffen: 4. Juni 2024).

Nadeau, David und Satoshi Sekine. 2007. "A survey of named entity recognition and classification." *Linguisticae Investigationes* 30, Nr. 1 (15. August): 3–26. <https://doi.org/10.1075/li.30.1.03nad>.

Novotny, Vit, Kristina Luger, Michal Štefánik, Tereza Vrabцова und Ales Horak. 2023. People and Places of Historical Europe: "Bootstrapping Annotation Pipeline and a New Corpus of Named Entities in Late Medieval Texts." In *Findings of the Association for Computational Linguistics: ACL 2023*, 14104–14113. Toronto, Canada: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-acl.887>.

Palladino, Chiara und Tariq Yousef. 2024. "Development of Robust NER Models and Named Entity Tagsets for Ancient Greek." In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, hg. von Rachele Sprugnoli und Marco Passarotti, 89–97. Torino, Italia: ELRA and ICCL. <https://aclanthology.org/2024.lt4hala-1.11> (zugegriffen: 4. Juni 2024).

Peng, Siyao, Zihang Sun, Sebastian Loftus und Barbara Plank. 2024. *Different Tastes of Entities: Investigating Human Label Variation in Named Entity Annotations*. <https://doi.org/10.48550/ARXIV.2402.01423>.

Pfanzelter, Eva, Sarah Oberbichler, Jani Marjanen, Pierre-Carl Langlais und Stefan Hechl. 2021. "Digital interfaces of historical newspapers: opportunities, restrictions and recommendations." *Journal of Data Mining & Digital Humanities* 6: 6121. <https://doi.org/10.46298/jdmmdh.6121>.

Piotrowski, Michael. 2019. "Accepting and Modeling Uncertainty." In *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*, hg. von Andreas Kuczera, Thorsten Wübbena, und Thomas Kollatz. Wolfenbüttel. (= Zeitschrift für digitale Geisteswissenschaften / Sonderbände: 4). https://doi.org/10.17175/SB004_006A.

Plank, Barbara. 2022. "The "Problem" of Human Label Variation: On Ground Truth in Data, Modeling and Evaluation." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 10671–10682. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.emnlp-main.731>.

Provatorova, Vera, Marieke van Erp und Evangelos Kanoulas. 2024. "Too Young to NER: Improving Entity Recognition on Dutch Historical Documents."

In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, hg. von Rachele Sprugnoli und Marco Passarotti, 30–35. Torino, Italia: ELRA and ICCL. <https://aclanthology.org/2024.lt4hala-1.4> (zugegriffen: 4. Juni 2024).

Rapp, Andrea. 2017. "Manuelle und automatische Annotation." In *Digital Humanities*, hg. von Fotis Jannidis, Hubertus Kohle, und Malte Rehbein, 253–267. Stuttgart: J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_18.

Reiter, Nils. 2020. Anleitung zur Erstellung von Annotationsrichtlinien. In: *Reflektierte algorithmische Textanalyse*, hg. von Nils Reiter, Axel Pichler, und Jonas Kuhn, 193–202. De Gruyter. <https://doi.org/10.1515/9783110693973-009>.

Rettinghaus, Klaus. 2024. "Normdaten Quo Vadis." In *DHd 2024*. <https://doi.org/10.5281/ZENODO.10698462>.

Rücker, Susanna und Alan Akbik. 2023. "CleanCoNLL: A Nearly Noise-Free Named Entity Recognition Dataset." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 8628–8645. Singapore: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.533>.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever und Ruslan Salakhutdinov. 2014. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15, Nr. 1: 1929–1958. <https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf> (zugegriffen: 19. Juli 2024).

Windhager, Florian, Saminu Salisu und Eva Mayr. 2019. "Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections." *Informatics* 6, Nr. 3: 29. <https://doi.org/10.3390/informatics6030029>.

Xiao, Yijun und William Yang Wang. 2019. "Quantifying Uncertainties in Natural Language Processing Tasks." In *Proceedings of the AAAI Conference on Artificial Intelligence* 33, Nr. 01: 7322–7329. <https://doi.org/10.1609/aaai.v33i01.33017322>.

Zhang, Zhen, Yuhua Zhao, Hang Gao und Mengting Hu. 2024. "LinkNER: Linking Local Named Entity Recognition Models to Large Language Models using Uncertainty." In *Proceedings of the ACM on Web Conference 2024*, 4047–4058. Singapore: ACM. <https://doi.org/10.1145/3589334.3645414>.