

Luger, Daniel; Atzenhofer-Baumgartner, Florian

De alto et basso : ein Experiment aus dem Bereich der "Distant Diplomats" zum Formular von Urkunden aus dem Schiedsgerichtswesen

Studia historica Brunensia. 2024, vol. 71, iss. 2, pp. 237-245

ISSN 1803-7429 (print); ISSN 2336-4513 (online)

Stable URL (DOI): <https://doi.org/10.5817/SHB2024-2-13>

Stable URL (handle): <https://hdl.handle.net/11222.digilib/digilib.80895>

License: [CC BY-NC 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/)

Access Date: 03. 01. 2025

Version: 20250102

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

***De alto et basso*: Ein Experiment aus dem Bereich der „Distant Diplomatics“ zum Formular von Urkunden aus dem Schiedsgerichtswesen**

***De alto et basso*: An Experiment in the Field of ‘Distant Diplomatics’ on the Formula of Arbitral Deeds**

Daniel Luger / daniel.luger@uni-graz.at /
<https://orcid.org/0000-0003-1778-4908>

Florian Atzenhofer-Baumgartner / atzenhofer@acm.org /
<https://orcid.org/0000-0001-8157-8629>

Institut für Digitale Geisteswissenschaften, Geisteswissenschaftliche Fakultät,
Karl-Franzens-Universität Graz, Graz, Österreich

Abstract

In this paper, we demonstrate a computer-aided analysis of charter material on arbitration. We apply a soft classification and analyze feature importances in order to quantitatively identify potentially significant formulas of arbitration charters. More specifically, we identify the previously little-noticed formula *de alto et basso* as particularly characteristic and qualitatively trace the path of its inclusion in late medieval charters on arbitration. This exemplary experiment lays the foundation for future research in this area, which could incorporate more diverse and multimodal approaches, including the (semi-)automated analysis of visual elements such as seals, writing materials, or fonts.

Keywords

distant diplomatics – formulas – arbitration – natural language processing – Monasterium.net

Bei einer Untersuchung von Urkunden zum spätmittelalterlichen Schiedsgerichtswesen mithilfe des klassischen Instrumentariums der Diplomatie sind zunächst mehrere grundlegende Problemfelder zu benennen. Es handelt sich dabei naturgemäß um eine Urkundengruppe, die eine große Bandbreite an möglichen Austellern aufweist, eben jene Fülle an Personen, die in diversen Rollen an Schiedsgerichten in unterschiedlichen Stadien teilnahmen. Hinzu kommt eine große Fülle an möglichen Inhalten dieser Urkunden, von Schiedsvereinbarungen und -verträgen über Ladungsschreiben bis hin zu Schiedssprüchen und anderen prozessualen Dokumenten. Daraus folgt eine große Bandbreite an möglichen externen und internen Eigenschaften dieser Urkunden, und es ist daher durchaus nicht überraschend, dass in der Diplomatie Urkunden zum spätmittelalterlichen Schiedsgerichtswesen im Gesamten bislang nicht als eigene Gattung oder Analyse-Kategorie in den Blick genommen wurden.

Im folgenden Beitrag wird eine Möglichkeit dargelegt, das urkundliche Material zum Schiedsgerichtswesen auf breiter Basis zu untersuchen, und dies auf einer breiten überregionalen Quellengrundlage, im Idealfall auf einer europäischen Ebene. Denn schließlich ist eine derartige Untersuchung wohl von Nöten, um einige wichtige, bislang offene Forschungsfragen zu dieser Thematik in den Blick zu nehmen, wie etwa bezüglich regionaler Unterschiede in der Terminologie bzw. dem Formular von Urkunden aus dem Bereich des Schiedsgerichtswesens oder die Ausbreitung und Durchsetzung dieses Verfahrens in europäischer Perspektive.

Eine Möglichkeit, derartigen Fragestellungen systematisch nachzugehen, ist die im Folgenden vorzustellende Methode der „Distant Diplomatics“, deren Art zurzeit an der Universität Graz im Rahmen des ERC Advanced-Grant-Projekts „From Digital to Distant Diplomatics“ (DiDip) unter der Leitung von Georg Vogeler entwickelt und erprobt wird¹. Im Rahmen dieses Vorhabens werden erstmals moderne computergestützte Methoden zur überregionalen Erforschung spätmittelalterlicher Urkunden eingesetzt, um damit etwa Fragen nach der Verbreitung und Entwicklung der gesamteuropäischen Urkundenpraxis und -kultur im Spätmittelalter (ca. 1300–1500) beantworten zu können. Die Neuartigkeit des Projekts liegt zum einen in seinem Umfang, denn das Vorhaben stützt sich auf eine Datenbank von über einer halben Million digitalisierter Urkunden aus dem mittelalterlichen und frühneuzeitlichen Europa – Monasterium.net –, wobei diese Plattform im Laufe der nächsten Jahre sowohl quantitativ als auch qualitativ erweitert wird. Zum anderen können durch die Anwendung modernster Methoden der Makroanalyse auf einen gesamteuropäischen Datensatz „aus der Distanz“² neue Erkenntnisse gewonnen werden, die über den regionalen und institutionenbezogenen Rahmen von traditionellen Forschungsprojekten der Diplomatie hinausgehen.

Eben diese pragmatische Limitierung der Urkundenforschung des Spätmittelalters auf kleinere Einheiten, d.h. auf spezielle Kanzleien, Sammlungen einzelner Institutionen oder die Gewohnheiten eines Landes, einer Sprachgruppe oder einer Region³ ist

1 ERC Advanced Grant Project – Nummer 101019327.

2 Siehe allgemein JOCKERS (2013).

3 Dazu grundlegend JARRETT (2013).

der einschlägigen Forschung bereits seit langem als grundlegendes Problem bewusst⁴ und stellt den Ausgangspunkt des Projekts DiDip dar. Denn dieses hat sich eine integrierte Studie der europäischen Diplomatik im Spätmittelalter zum Ziel gesetzt, wobei dies insbesondere durch den Aufbau einer virtuellen Forschungsumgebung gelingen soll, um die Qualität der vorhandenen Forschungsdaten sowie die verfügbaren digitalen Methoden zu verbessern und damit konkrete diplomatische Fragestellungen auf einer Makroebene zu untersuchen und auf diesem Weg zu einem besseren Verständnis der Beziehungen zwischen den verschiedenen lokalen bzw. regionalen diplomatischen Praktiken beizutragen.⁵ Welche neuen Möglichkeiten sich dadurch konkret ergeben, soll im Folgenden anhand eines Experiments zu Urkunden aus dem Umfeld der Schiedsgerichtsbarkeit präsentiert werden.

Die Quellenbasis für diese Untersuchung stellt – wie oben angeführt – die Urkundensammlung von Monasterium.net dar. Ihre Plattform wurde im Jahr 2002 von niederösterreichischen Archivaren und Mönchen zunächst als regionale Plattform für mittelalterliche und frühneuzeitliche Urkunden geschaffen. Monasterium verbindet dabei digitale Bilder von Urkunden, archivische Erschließung und retrodigitalisierte Regesten und Editionen, bringt dieses bislang verstreute Quellenmaterial in einer digitalen Plattform zusammen und macht sie jedem jederzeit zugänglich. Von dieser Grundidee ließen sich im Laufe der folgenden Jahre auch zahlreiche Archive in umliegenden Ländern überzeugen, und so ist die Zahl der Urkunden in Monasterium.net in den letzten 15 Jahren auf über 650.000 angewachsen. Die Datenbank enthält heute Urkunden von Archiven aus ganz Österreich und seinen Nachbarländern (Tschechien, Deutschland, Italien, Ungarn, Schweiz, Slowenien, Slowakei), aber auch aus Estland, aus Polen, aus anderen Teilen Deutschlands und aus Spanien.⁶ Etwa die Hälfte der zurzeit verfügbaren Urkunden stammt aus dem 14. und 15. Jahrhundert, eine Fülle, die für den einzelnen Forscher keineswegs zu überblicken ist. Aber auch für die gezielte Suche nach einzelnen Urkunden in diesem Bestand müssen aktuell mehrere Problemfelder konstatiert werden. So existieren von über 80 % aller Urkunden auf Monasterium.net Metadaten in Form von Regesten, jedoch ist der Anteil jener mit Transkriptionen des Urkundentexts mit etwa 16 % deutlich geringer, was insbesondere für das Suchen nach bestimmten Urkundenformeln oder Begriffen höchst problematisch ist. Bei der Untersuchung von Urkunden zum Schiedsgerichtswesen auf Monasterium.net ist überdies zu beachten, dass von den über 500.000 Regesten annähernd 60 % in Deutsch verfasst sind, allerdings befindet sich in den vorhandenen Metadaten auch eine beträchtliche Zahl an Regesten in ungarischer (69.610 Stück bzw. 14 %), tschechischer (26.497 Stück bzw. 5,5 %) und italienischer (4.301 Stück bzw. 1 %) Sprache, was die Verwendung von Suchbegriffen in mehreren Sprachen notwendig macht.

4 Z.B. HLAVÁČEK (2006).

5 Zu weiteren grundlegenden Informationen und Neuigkeiten siehe die Projekt-Homepage <https://didip.hypotheses.org/>

6 Zu diesem Projekt siehe jüngst etwa VOGELER (2019).

Überdies schwankt die Qualität der auf Monasterium.net vorhandenen Informationen von Archiv zu Archiv, teilweise auch von Bestand zu Bestand beträchtlich. So finden wir eine große Bandbreite an Metadaten, von ausführlichen Vollregesten nach dem Muster der Regesta Imperii über knappe Kurzregesten, die häufig größere Teile des Urkundentextes vermissen lassen, bis hin zu (Teil-)Beständen mit Urkundendigitalisaten ohne jegliche Regesten oder andere Metadaten, was eine gezielte Durchsuchung über die Eingabemaske mithilfe von zuvor gewählten Suchbegriffen unmöglich macht.⁷

Angeichts dieser schwerwiegenden Probleme bei der aktuellen Online-Suche auf Monasterium.net kann jedes Ergebnis, etwa bei der Suche nach der Formel *Arbiter arbitrator seu amicabilem compositor*, die in der diplomatischen Forschung bereits seit längerem mit Urkunden zum mittelalterlichen Schiedsgerichtswesen in Verbindung gebracht wird⁸, nur höchst unvollständig und zufällig sein. Welche Möglichkeiten auf diesem Feld quantitative Analyseverfahren als brauchbare Alternative bieten können, soll im Folgenden anhand eines konkreten Experiments ausgeführt werden.

Ausgangspunkt unseres Experiments ist die (semi-)automatisierte Analyse der Urkundenformulare in einem ausgewählten Korpus, um dadurch die Formelhaftigkeit in Urkunden zur Schiedsgerichtsbarkeit im Allgemeinen und bislang nicht beachtete einzelne Formeln untersuchen zu können. Zunächst wird aus sämtlichen Dokumenten auf Monasterium.net eine kuratierte Liste von Urkunden aus verschiedenen europäischen Archiven und Sammlungen zusammengestellt, für die sowohl die Transkription des Urkundentextes als auch ein Regest verfügbar sind. Das Ergebnis ist eine Liste von ca. 41.000 Urkunden, davon ca. 25.000 in Latein und der Rest in deutscher Sprache. Die erschlossenen Regesten sind überwiegend in (modernem) Deutsch, Englisch, Latein und Ungarisch verfasst.⁹ Die durchschnittliche Länge der Transkriptionen beträgt 393 Wörter, die der Regesten 33. Abgesehen von einer automatisch durchgeführten konsequenten Kleinschreibung werden die Texte vor den weiteren Analyseschritten nicht weiter vorverarbeitet.

Unsere Methode führt dabei einen neuen Ansatz für die quantitative Analyse von Urkunden ein, indem wir ihre Identifizierung als eine weiche binäre Klassifizierungsaufgabe konzipieren. Diese besteht darin, sich auf die Kategorisierung von Daten auf der Grundlage des Vorhandenseins von vordefinierten historischen oder zeitgenössischen Begriffen zu konzentrieren, die in unserem Anwendungsfall einen Bezug zum historischen Schiedsgerichtswesen aufweisen, nämlich die Begriffe *arbiter*, *arbitrator*, *compromissum*, *spruchman*, *hindergang*, *hintergang* bzw. Schiedsgericht, -richter, -spruch bzw. -mann.¹⁰

7 Siehe beispielsweise die umfangreichen spätmittelalterlichen Urkundenbestände des Nadškofija Arhiv in Ljubljana ab dem Jahr 1415: <https://www.monasterium.net/mom/SI-NSAL/archive>, abgerufen am 18.10.2024.

8 Vgl. etwa die vorbildhafte Studie von BADER (1960).

9 Nicht bereits über Metadaten ausgewiesene Sprachen von Texten und Regesten wurden computergestützt erkannt und ergänzt; <https://huggingface.co/ERCDDip/langdetect>, abgerufen am 18.10.2024 für die Erfassung von Urkundentexten, fastText für Regesten.

10 Um Schiedsgerichtsurkunden von ähnlichen noch besser abgrenzen zu können, bilden diese Begriffe

Diese Methode der Verwendung einzelner Wörter und Sequenzen als erste Unterscheidungsmerkmale ist sowohl eine Maßnahme angesichts der arbeitsintensiven Natur der traditionellen Methoden der Diplomatik als auch eine strategische Entscheidung, die allgemein auf Ressourceneffizienz¹¹ und Reproduzierbarkeit abzielt.

Quantitative Analyse

Nach der Vorbereitung des Datensatzes wird die Anzahl des Vorkommens dieser Begriffe in Transkriptionen und Regesten ermittelt. Im ersten Set werden 732 Einträge zum Thema Schiedsgerichtsbarkeit gezählt, im zweiten 617. Nach der Reduktion der 15 wichtigsten Skip-Gram-Merkmale¹² zu formelhaften Tri-Grammen,¹³ deren Bestätigung bzgl. Zweckdienlichkeit sowie anschließender Zusammenführung auf der Grundlage des gemeinsamen Auftretens in der Wichtigkeitsmatrix können insgesamt 2.977 Urkunden als relevant für das Thema Schiedsgerichtsbarkeit identifiziert werden.

Um die Unterschiede in den Ergebnissen zu bewerten, werden für jedes Set zwei Teilstichproben im Verhältnis 1:1 bzw. 1:5 erstellt. Die Liste der kuratierten Tri-Gramme wird entsprechend getestet. Um Kandidaten für mögliche Formeln um Zusammenhang mit dem Schiedsgerichtswesen zu generieren (und zu bewerten) sowie insbesondere ihre diskriminierenden¹⁴ Eigenschaften zu berücksichtigen, wenden wir ein ausgewähltes Gradientenverfahren¹⁵ auf die Skip-Gramme ($n=3$, $k=2$) der Urkunden an, die auf Dokument-Term-Matrizen als Vektorisierer basieren, und untersuchen ihre Metriken entsprechend.

Aufgrund ihrer erwiesenen Zuverlässigkeit¹⁶ und ihres Einsatzes für Klassifikationsaufgaben in sogenannten ‚low-resource settings‘ ziehen wir zuzüglich eine einfache nicht-optimierte Stützvektormethode (SVM) hinzu. Weiters werden unterschiedliche Merkmalsextraktionen getestet, darunter die auf Termini bezogene Vorkommenshäufigkeit/inverse Dokumenthäufigkeit (TF/IDF) sowie Dokument-Term-Matrizen ($n=1,2,3$; $n=2,3/k=1,2$). Für alle Tests werden außerdem die Klassenverhältnisse berücksichtigt.

Wie erwartet stellen wir fest, dass Unigramme bei den weniger ausgewogenen Datensätzen schlecht abschneiden, was sich in einem Recall von $\leq 50\%$ widerspiegelt. Ihre ausgewogenen Varianten weisen allerdings einen Makro-Durchschnitt von etwa 87% auf. Bi-Gramme folgen diesem Muster klar. Bei Unigrammen schneidet die ursprüngliche

ein Sub-Set eines größeren, welches zudem die Worte Vermittlung, Beilegung und Abkunft im Regest beinhaltet. Ihre Differenz im höherdimensionalen Raum kann somit zur Kontrolle bei der Evaluierung der Klassifikation sowie deren Merkmalgewichtung herangezogen werden.

11 Siehe dazu allgemein NICOLAOU-LUGER-DECKER-RENET-CHRISTLEIN-VOGELER (2023), S. 284–295.

12 Skip-Gramme sind Wortketten mit einer Länge n , die k Auslassungen erlauben.

13 Tri-Gramme sind Wortketten mit einer Länge von $n=3$.

14 Eine Diskriminanzanalyse bezieht sich hier vor allem auf die negative Klassifizierung, also das Feststellen, dass und in der Folge warum es sich nicht um eine Schiedsgerichtsurkunde handeln soll.

15 Genauer: Extreme Gradient Boosting mit XGBoost; siehe CHEN-GUESTRIN (2016), S. 785–794.

16 WANG-MANNING (2012), S. 90–94.

Begriffsliste besser ab als die davon indirekt abgeleiteten Formeln, bei Bi-Grammen jedoch weist die Formelliste einen Makro-Durchschnitt von 93 % auf und offenbart überwiegend Begriffe mit domänenrelevanter Formelhaftigkeit, wie etwa die Wortsegmente *una parte*, *andern tail* oder *bono pacis*.

Wie erwartet, verstärkt sich dieser (gegensätzliche) Trend mit Tri-Grammen noch, erreicht man mit diesen doch Makro-Werte von 97 % für die Wiedererkennung und erhält signifikante Merkmale (z.B. *dem andern tail*), während die begriffsbasierte Klassifizierung vor allem unter einer Übergeneralisierung leidet, was sich in beträchtlichen Merkmalsbewertungen für sehr allgemeine Wortfolgen, z.B. *mit allen den*, widerspiegelt.

Ein Leistungsanstieg setzt sich fort, wenn einfache minimale Skip-Gramme verwendet werden ($n=2$, $k=1$). Für alle Teilmengen gilt jedoch, dass längere Skip-Gramme die Gesamtergebnisse erheblich verschlechtern – eine Vorahnung darauf, welches Level der Gramme am ehesten statistisch robust ist. Die besten Ergebnisse finden sich mit den kuratierten Tri-Grammen und einer begrenzten Term-Dokument-Matrix, nämlich mit Höchstwerten von 99 %.

Bei praktisch allen Durchläufen liegen Recall und Präzision für das Klassifizieren der vorliegenden Urkunden *ohne* Bezug zum Schiedsgerichtswesen durchwegs um oder über 90 %. Das bedeutet, dass sie insgesamt sogar leichter als Schiedsgerichtsurkunde ausgeschrieben als nominiert werden.¹⁷

Diskussion

Bei den Wortfolgen oder Formeln, die auf dem eben beschriebenen Weg zu Tage treten, handelt es sich im Kern überwiegend um lateinische und deutschsprachige Tri-Gramme, die im Prinzip in drei Kategorien eingeteilt werden können:

- (1) allgemein gebräuchliche Wortfolgen in Urkunden
- (2) Formeln aus dem Bereich des mittelalterlichen Gerichts- bzw. Rechtswesens
- (3) spezifische Formeln für Urkunden aus dem Bereich der Schiedsgerichtsbarkeit

Bei der Interpretation dieser Liste, deren Einträge im Folgenden nach der Wichtigkeit ihres Auftretens für einen Klassifikationsdurchgang sortiert sind, ist zu sehen, dass mehrere der höchstrangigen Treffer inhaltlich offensichtlich mit der Schiedsgerichtsbarkeit im Zusammenhang stehen und damit diese Urkundengruppe insgesamt deutlich von anderen abgrenzen. Darunter befinden sich auch Varianten der bereits oben angeführten Formel *Arbiter arbitrator seu amicus compositor*, die in Tri-Gramme gespalten zwei Mal auf den vorderen Plätzen der Liste auftaucht.

¹⁷ Im Übrigen zeigt sich auch bei der Analyse der Regesten anstatt der Urkundentexte eine ausgesprochen hohe Rate an korrekter Klassifikation, nämlich über 90% über alle Sets hinweg, bis hin zu Höchstwerten von 98%. Wir betrachten das als Zugeständnis an den Wert von Regesten und anderen ‚Metadaten‘, welche das Potenzial haben, einerseits die Genauigkeit einer Klassifizierung zu verbessern, andererseits eine solche etwa über Sprachen hinweg zu ermöglichen.

- *pro bono pacis*
- *arbitratores seu amicabile*
- *appellatione remota fine*
- *alto et basso*
- *dem andern tail*
- *et amicabile compositores*
- *ex una parte*
- *ex parte una*
- *omnibus et singulis*
- *cum inter nos*
- *quod si non*
- *baidet tail red*
- *ex altera super*
- *ab utraque parte*
- *inter nos et*
- *pacis et concordie*
- *red vnd widerred*

Besonders hingewiesen sei an dieser Stelle lediglich auf die an vierter Stelle aufgeführte Formel [*de*] *alto et basso*, die im untersuchten Material tatsächlich in erster Linie in Urkunden aus dem Bereich des spätmittelalterlichen Schiedsgerichtswesens Verwendung findet, allerdings von der diplomatischen Forschung bislang kaum in den Blick genommen wurde. Rechtlich sollte durch diese für den spätmittelalterlichen lateinischen Sprachgebrauch ungewöhnliche Formel wohl zum Ausdruck gebracht werden, dass sämtliche Streitigkeiten, gleich welcher Art, dem einzusetzenden bzw. eingesetzten Schiedsgericht unterworfen werden bzw. wurden.¹⁸ Diese Formel geht offensichtlich direkt auf die Rezeption des *Speculum iudicale* von Durantis zurück, wo im beispielhaften Abschnitt „De forma compromissi et arbitrii“ ebenfalls diese Wortfolge Verwendung fand: *concorditer compromittunt et consentiunt de alto et basso in Caium electum*.¹⁹

18 Siehe beispielsweise: ... *rectores, arbitros, arbitratores, amicabile compositores communes, amicos et bonos viros de alto et basso compromiserunt* ... (Praha, Národní archiv, Archivy českých klášterů zrušených za Josefa II., ŘC Zl. Koruna 1405 III 21 / Monasterium.net/mom/CZ-NA/AZK%7CKoruna/1405_III_21/charter, abgerufen am 18.10.2024); ... *non coactus uel compulsus sed bona et libera voluntate de alto et basso in honorabilem ac nobilem dominum Petrum de Rosemberch capelle regie omnium sanctorum in castro Pragensi prepositum tamquam arbitrum arbitratorem et amicabile compositorem beniuole compromisi* ... (Vyšší Brod, OCist, Urkunden (1173-1846) 1371 III 31 / Monasterium.net/mom/CZ-KAVB/VYBRO/1371_III_31/charter, abgerufen am 18.10.2024); ... *ac insuper in huiusmodi compromisso facere disponere ordinare et etiam promulgare de alto et basso super premissis ac quolibet eorumdem ita, quod* ... (Göttweig, Stiftsarchiv, Urkunden 1369 I 21 / Monasterium.net/mom/AT-StiAG/GoettweigOSB/1369_I_21/charter, abgerufen am 18.10.2024).

19 DURANTIS (1592), S. 104.

Bibliography

Unedited sources

Göttweig, Stiftsarchiv, Urkunden 1369 I 21 / [Monasterium.net/mom/AT-StiAG/GoettweigOSB/1369_I_21/charter](https://monasterium.net/mom/AT-StiAG/GoettweigOSB/1369_I_21/charter), abgerufen am 18.10.2024.

Praha, Národní archiv, Archivy českých klášterů zrušených za Josefa II., ŘC Zl. Koruna 1405 III 21 / [Monasterium.net/mom/CZ-NA/AZK%7CKoruna/1405_III_21/charter](https://monasterium.net/mom/CZ-NA/AZK%7CKoruna/1405_III_21/charter), abgerufen am 18.10.2024.

Vyšší Brod, OCist, Urkunden (1173-1846) 1371 III 31, [Monasterium.net/mom/CZ-KAVB/VY-BRO/1371_III_31/charter](https://monasterium.net/mom/CZ-KAVB/VY-BRO/1371_III_31/charter), abgerufen am 18.10.2024.

Edited sources

DURANTIS, Guilelmus (1592): *Speculum Iuris*, 2 Bde., Frankfurt.

Literature

BADER, Karl Siegfried (1960): *Arbiter arbitrator seu amicus compositor*, in: Zeitschrift der Savigny-Stiftung für Rechtsgeschichte: Kanonistische Abteilung 46, S. 239–276.

CHEN, Tianqi – GUESTRIN, Carlos (2016): *XGBoost: A Scalable Tree Boosting System*, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (Association for Computing Machinery), New York, S. 785–794.

HLAVÁČEK, Ivan (2006): *Das Problem der Masse: das Spätmittelalter*, in: Archiv für Diplomatik 52, S. 371–394.

JARRETT, Jonathan (2013): *Introduction: Problems and Possibilities of Early Medieval Charters*, in: Jonathan JARRETT – Alan SCOTT MCKINLEY (eds.): *Problems and Possibilities of Early Medieval Charters*, Turnhout, S. 1–12.

JOCKERS, Matthew (2013): *Macroanalysis: Digital Methods and Literary History*, University of Illinois Press.

NICOLAOU, Angelos – LUGER, Daniel – DECKER, Franziska – RENET, Nicolas – CHRISTLEIN, Vincent – VOGELER, Georg (2023): *Efficient Annotation of Medieval Charters*, in: Mickael COUSTATY – Alicia FORNÉS (eds.): *Document Analysis and Recognition – ICDAR 2023 Workshops*, Berlin, S. 284–295.

VOGELER, Georg (2019): *„Monasterium.net“ – eine Infrastruktur für diplomatische Forschung*, in: Roman BLEIER – Franz FISCHER – Torsten HILTMANN – Gabriel VIEHHAUSER – Georg VOGELER (eds.): *Digitale Mediävistik und der deutschsprachige Raum (Das Mittelalter 24, 1)*, Berlin, S. 247–252.

WANG, Sida – MANNING, Christopher D. (2012): *Baselines and Bigrams: Simple, Good Sentiment and Topic Classification*, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers – Volume 2 (Association for Computational Linguistics), S. 90–94.

De alto et basso: Experiment z oblasti „distanční diplomatiky“ k formuláři arbitrážních listin

Tento experiment nejprve ukazuje výhody využití formulářových obrátů namísto jednotlivých termínů při měkké klasifikaci textů. Dále se ukazuje, že význam znaků během trénování modelu i ručního vyhodnocování formulářových obrátů může těžit z toho, že se zohlední jazykové segmenty jako discriminatory typu text. Měkkou klasifikaci jsme použili k identifikaci potenciálně významných, dříve nepozorovaných formulářových obrátů jako markerů, poukazujících na arbitrážní listiny. Mimo jiné se podařilo identifikovat jako obzvláště charakteristickou formuli *de alto et basso*, které byla dosud věnována malá pozornost, a zhruba vysledovat cestu jejího začleňování do pozdně středověkých listin týkajících se arbitráží. Tento ukázkový experiment z oblasti distanční diplomatiky pokládá základy pro budoucí výzkum v této oblasti, který by mohl zahrnovat rozmanitější a víceúrovňové přístupy, včetně automatizované analýzy vizuálních prvků, jako jsou pečeti, psací látka nebo písmo. Kromě toho lze budoucí zlepšení a inovace metodiky dosáhnout začleněním dalších proměnných, jako je prostor, čas, instituce nebo jazyk, a díky metodám distanční diplomatiky tak získat další poznatky o rozšíření a specifických vlastnostech určitých typů listin.



This work can be used in accordance with the Creative Commons BY-NC-ND 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.

