

Automatisierte Kategorisierung mittellateinischer Bittbriefe an den Papst

Nico Faltin, FU Berlin (Lateinische Philologie) / HU Berlin (DFG-Projekt Daidalos)

Konstantin Schulz, HU Berlin (DFG-Projekt Daidalos)

Kontext

- Promotionsprojekt (FU Berlin): spätmittelalterliche Bittbriefe an den Papst
- DFG-Projekt Daidalos (HU Berlin): NLP-Forschungsinfrastruktur für die Klassische Philologie
 - NLP = Natural Language Processing = maschinelle Sprachverarbeitung
 - Internetseite mit NLP-Forschungswerkzeugen
 - Community of Practice (z.B. Workshops, Lernmaterialien, weitere NLP-Ressourcen)
 - Forschungsstandems = Zusammenarbeit zwischen Forschenden und Daidalos

Was erwartet Sie?

1. Was sind mittellateinische Bittbriefe an Papst?
2. Warum sollte man sie maschinell kategorisieren?
3. Wie funktioniert die automatisierte Kategorisierung?
4. Wie gut sind die Ergebnisse?

1. Was sind mittellateinische Bittbriefe an Papst?

Bittbriefe (= Bittschrift = Supplik = Gnadengesuch)

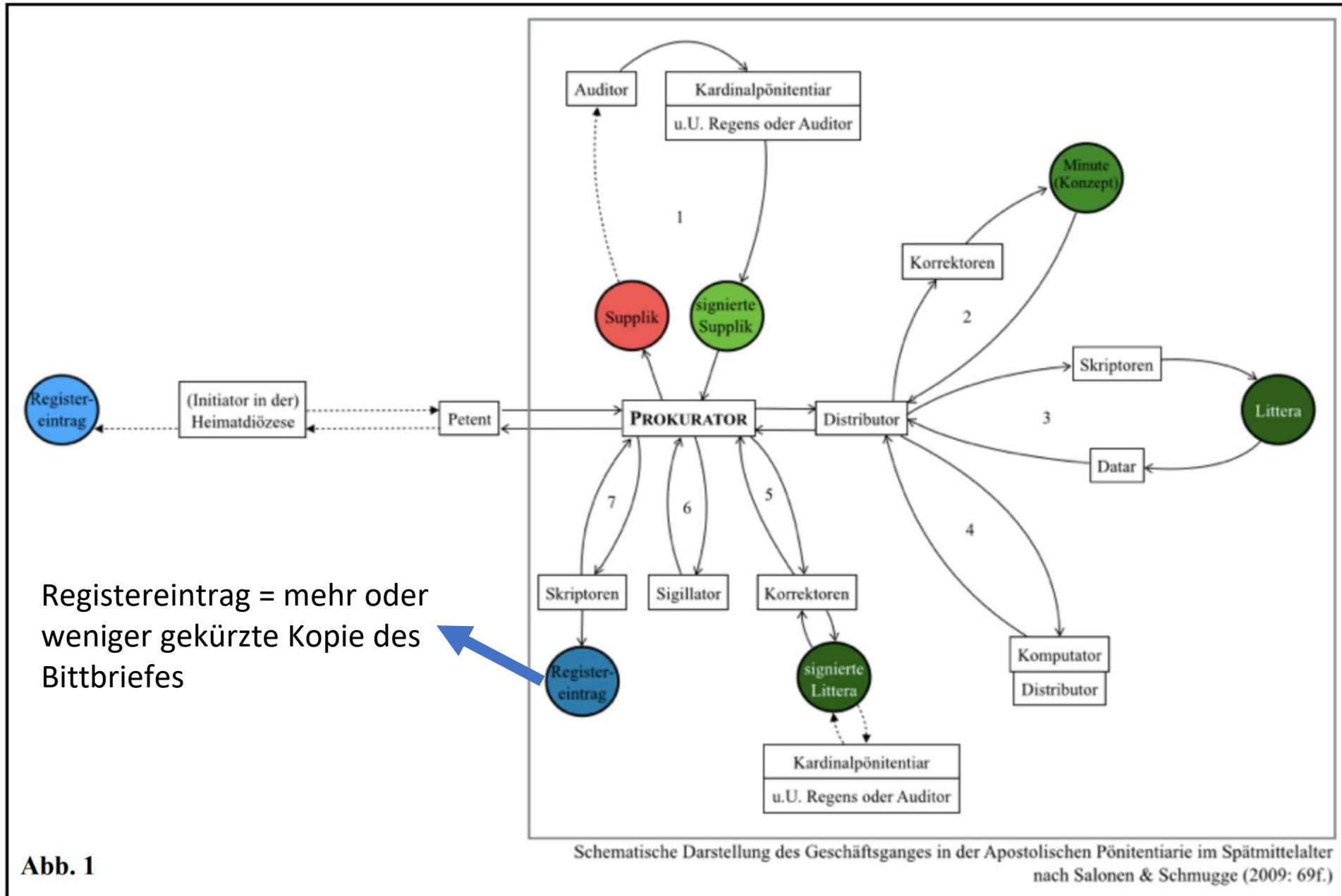
- Medium zur Kontaktaufnahme mit den Päpsten
- Person, die den Bittbrief einreicht = Petent / Bittsteller (lat. *exponens*)
- Gründe für Bittbriefe: Sorge um Seelenheil, weltliche Gründe
- Warum nicht beim Dorfpriester? → *casus papales* / Reservatsfälle
- in lateinischer Sprache
- i.d.R. über eine vermittelnde Instanz (Prokuratoren) → Einhaltung sprachlicher und struktureller Formalia (*stilus curiae*)

1. Was sind mittellateinische Bittbriefe an Papst?

- steigende Fallzahlen → Delegation päpstlicher Vollmachten (plenitudo potestatis) an ...
 - Legaten
 - Konzilien
 - bestimmte Personen, die einer Institution vorstehen

1. Was sind mittellateinische Bittbriefe an Papst?

- Kardinalgroßpönitentiar
- **Apostolische Pönitentiarie** (Sacra Poenitentiaria Apostolica)
= höchstes päpstliches Gnadenamt
- Gewährung verschiedener Gnaden (z.B. Absolution von der Exkommunikation)
- Bittbrief → Antwortbrief: viele Arbeitsschritte



1. Was sind mittellateinische Bittbriefe an Papst?

Registriert ...

- wurden nur die Bittbriefe, nicht die Antwortschreiben.
- wurden nur positiv beschiedene Bittbriefe.
- wurde nur gegen Aufpreis.
- 1431-1523: > 40.000 registrierte Bittbriefe aus dem Gebiet des ehem. Dt. Reiches
 - *Repertorium Poenitentiariae Germanicum* (RPG), 11 Bd., 1992-2018, teilweise digital verfügbar

1. Was sind mittellateinische Bittbriefe an Papst?

- Thematische Gliederung der Registerbände:

1. Ehesachen (De matrimonialibus)
2. Geburtsmakel (De defectu natalium et de uberiori)
3. Klerikale Karrieren (De promotis et promovendis)
4. Beichtprivilegien (De confessionalibus)
5. **Diverses (De diversis formis et de declaratoriis)**

→ Welche Themen tauchen in diesen Bittbriefen auf?

2. Was spricht für die maschinelle Kategorisierung dieser „diversen“ Bittbriefe?

1. Bittbriefe der diffusen Kategorie *De diversis formis et de declaratoriis*:
 - „hoher recht-, kultur-, sozial- und sprachgeschichtlicher Wert“ (RPG IV, 1996, S. XXIV)
 - bis auf wenige Fallstudien aus dem RPG-Kreis kein großes Echo in der Forschung
 - u.a. wegen Unübersichtlichkeit? → Welche Themen werden in den Bittbriefen verhandelt?
2. viele Tausend Bittbriefe, in weiten Teilen formelhaft, (teilweise) digital verfügbar → prädestiniert für maschinelle Auswertung

3. Wie funktioniert die automatisierte Kategorisierung?

Erstgespräch mit Konstantin Schulz

- „Lassen sich die Registereinträge, die in den Registern der Pönitentiarie unter Überschrift *De diversis formis et de declaratoriis* verzeichnet sind, mithilfe von NLP-Tools automatisiert in vorher festgelegte thematische Kategorien einordnen?“
- „automatisiert in Kategorien einordnen“ → NLP: Textklassifikation
- Eingrenzung auf das Gebiet d. ehem. dt. Reiches → Transkripte vorhanden
- RPG, Bd. 1-8 → Texte in digitaler Form vorhanden = 9.070 Bittbriefe (Diverses)

3. Wie funktioniert die automatisierte Kategorisierung?

Erstgespräch mit Konstantin Schulz

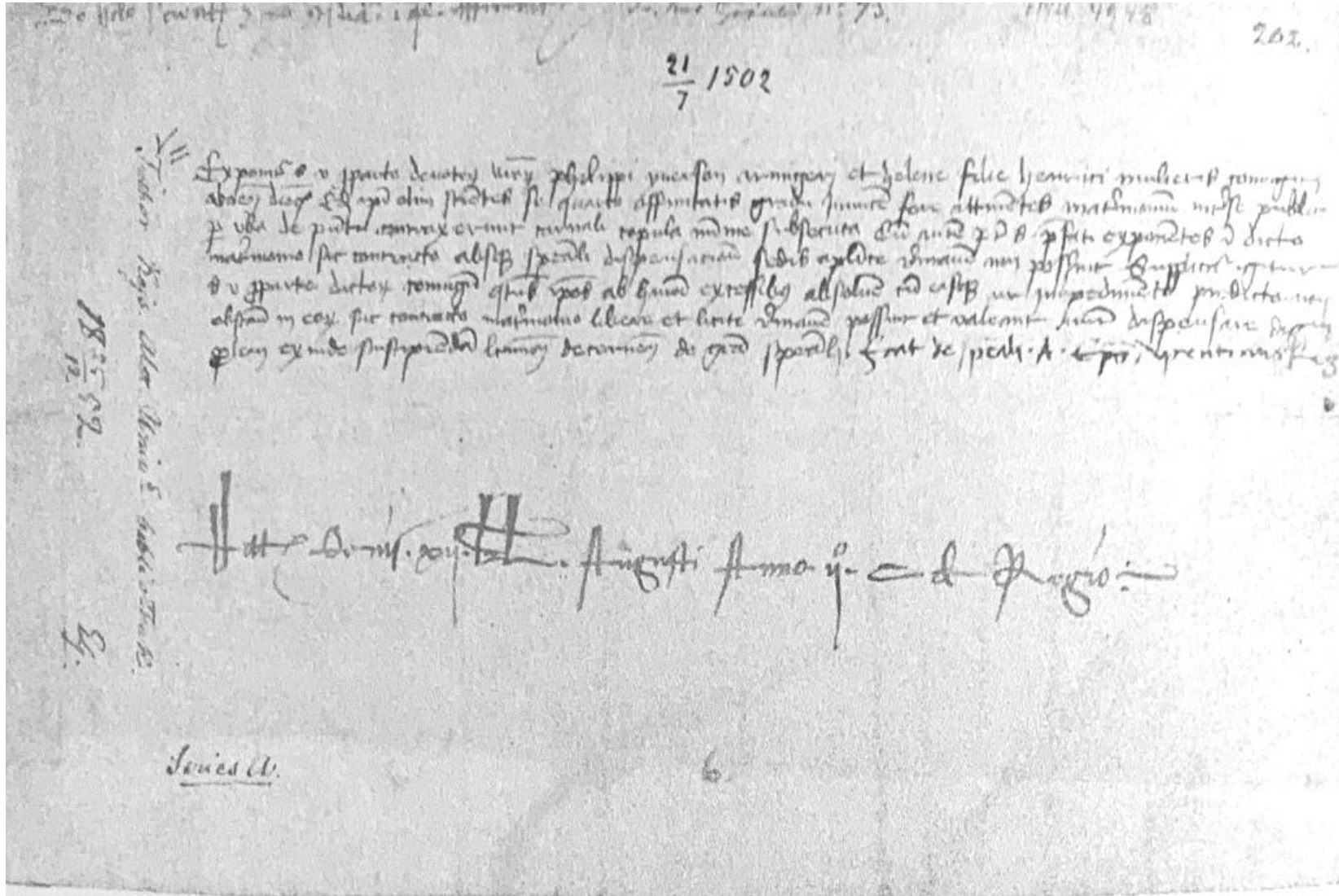
„Ich brauche:“

- Texte der 9.070 Bittbriefe in digitaler Form, z.B. XML-Format
- händische Kategorisierung in Excel-Dateien (Trainingdaten)
 - 260 Bittbriefe (Diverses) aus der Erzdiözese Magdeburg
 - 17 thematische Kategorien
 - Modifikation der Kategorien nach Schmugge/Hersperger/Wiggenhauser (1996)

3. Wie funktioniert die automatisierte Kategorisierung?

Tötungsdelikte	Raubzüge/Kriegsbeteiligung	Klosterangelegenheiten
Körperverletzung	Lockerung von Fastenvorschriften	Gottesdienst trotz Exkommunikation/Interdikt
Lockerung/Umwandlung von Gelübden	Besuchserlaubnis für Pilgerstätten	Beichtprivileg
Körperlicher Defekt	Absolution von Sentenzen	Kirchliche Karriere
Simonie (verbotener Ämterkauf)	Bestattung	Übriges
Sex	Sonderregeln für die Messe	

Beispiel: Bittbrief, 1460 (s. Handout)



Anrede; Name, soziale und lokale Herkunft der Bittsuchenden	Exponitur sanctitati vestre pro parte devotorum vestrorum Philippi Yverson armigeri et Helene filie Henrici mulieris coniugum Aboensis diocesis	It is set forth to your Holiness by your devotees squire Philippus Yverson and wife Helena, daughter of Henricus, spouses from the Diocese of Turku,
Fallschilderung (narratio)	quod ipsi olim scientes se quarto affinitatis gradu invicem fore attinentes matrimonium inter se publice per verba de presenti contraxerunt carnali copula minime subsecuta.	that they some time ago, knowing themselves to be mutually constrained by affinity of the fourth degree, publicly contracted a marriage between themselves through words of present intent, which was not followed by carnal union.
Gnadengesuch (petitio)	Cum autem, pater sancte, prefati exponentes in dicto matrimonio sic contracto absque speciali dispensatione sedis apostolice remanere non possint, supplicatur igitur sanctitati vestre pro parte dictorum coniugum quatenus ipsos ab huiusmodi excessibus absolvere cum eisque ut impedimento predicto non obstante in eorum sic contracto matrimonio libere et licite remanere possint et valeant misericorditer dispensare dignemini prolem exinde suscipiendam legitimam decernentes de gratia speciali.	Since, however, Holy Father, the aforementioned petitioners may not remain in the mentioned matrimony, which was contracted without a special dispensation from the Apostolic See, Your Holiness is therefore beseeched on behalf of the aforementioned spouses to mercifully deign to absolve them from such transgressions, and – notwithstanding the aforementioned impediment – to deign to grant dispensation to them by special grace so that they might be able and have the power to remain freely and lawfully in their marriage thus contracted, and to regard the offspring resulting from it as legitimate.
Signatur	Fiat de speciali, A[ntonius] episcopus Aprutinus, Regens.	So be it by special mandate, Antonius, bishop of Teramo, regent.
Datum	Datum Senis xij kalendas augusti anno ijº, C. de Regio.	[21.7.1460]


In welche der 17 Kategorien würden Sie den Brief einordnen?


- Paar aus Finnland
- Eheschluss trotz Verschwägerung 4. Grades
- Absolution und Ehedispens (Sondergenehmigung)
- Rechtlicher Hintergrund:
4. Laterankonzil (1215) → Eheverbot für Blutsverwandte und Verschwägte bis zum einschließlich 4. Grad


Tötungsdelikte	Raubzüge/Kriegsbeteiligung	Klosterangelegenheiten
Lockerung/ Umwandlung von Gelübden	Lockerung von Fastenvorschriften	Gottesdienst trotz Exkommunikation/ Interdikt
Körperverletzung	Besuchserlaubnis für Pilgerstätten	Beichtprivileg
Körperlicher Defekt	Absolution von Sentenzen	Kirchliche Karriere
Simonie (verbotener Ämterkauf)	Bestattung	Übriges
Sex	Sonderregeln für die Messe	


3. Wie funktioniert die automatisierte Kategorisierung?


Band	Lemma-ID	Tötungsdelikte	Körperverletzung	Raubzüge/Kri	Klosterangele	Lockerung von Fast	Gottesdien	Kirchliche Karriere
RPG I	11000093							
RPG I	11000511							x
RPG I	11000116		x					
RPG I	11000648	x						
RPG I	11000732	x						
RPG II	11100400							
RPG II	11100353							
RPG II	11100161							x
RPG II	11101027		x			x		
RPG II	11100377					x		
RPG II	11100514					x		
RPG II	11100035				x			
RPG II	11100086				x			


 rpg1.xml


 rpg2.xml


 rpg3.xml

 rpg4.xml

 rpg5.xml

 rpg6.xml

 rpg7.xml

 rpg8.xml

RPG	Nr.
1	1-775
2	1-1066
3	1-595
4	822-1845
5	931-2194
6	2019-3820
7	1478-2659
8	2093-3453

3. Wie funktioniert die automatisierte Kategorisierung?

Training und Prototyp (Konstantin Schulz)

- KI (Sprachmodell) als Basis, ähnlich wie ChatGPT:
<https://huggingface.co/ClassCat/roberta-base-latin-v2>
- Training: „gesamtes lateinisches Internet“:
<https://huggingface.co/datasets/statmt/cc100> (ca. 600 Megabytes lateinischer Rohtext)
- Spezialisierung auf Textklassifikation: Text als Input, Kategorie als Output

3. Wie funktioniert die automatisierte Kategorisierung?

Training und Prototyp (Konstantin Schulz)

- wahrscheinlichkeitsbasiert:
 - für jeden Text wird für jedes mögliche Kategorie eine Wahrscheinlichkeit ausgegeben, d.h. 17 verschiedene Werte pro Text
 - Output = Kategorie-Label mit der höchsten Wahrscheinlichkeit
- **Ergebnis:** 94,6% korrekte Zuordnungen in den Trainingsdaten

4. Wie gut sind die Ergebnisse?

https://scm.cms.hu-berlin.de/daidalos/daidalos-share/-/tree/main/tandems/NF?ref_type=heads

```
1  [
2    {
3      "lemma_id": "11200001",
4      "text": "Petrus Sphlan presb. Herbip. dioc. exponit, quod ipse olim pretermisiss min. ord. se fecit ad omnes sacros ord. promoveri et in ipsis ministravit: de abso
5      "label": "Klosterangelegenheiten"
6    },
7    {
8      "lemma_id": "11200002",
9      "text": "Johannes Ruff de Buttelspach orator cler. Constant. dioc. exponit, quod cum olim temp. guerrarum in partibus suis vigentium inter nob. et civit. inimici t
10     "label": "Gottesdienst trotz Exkommunikation/Interdikt"
11   },
12   {
13     "lemma_id": "11200003",
14     "text": "Henricus Hase laic. prof. mon. o.Cist. de Herbacht nullius dioc. propter molestias et persecutiones, que sibi per superiores et confratres suos iniuste in
15     "label": "Übriges"
16   },
17 ]
```

```
167058 "lemma_id": "11102784",
167059 "text": "Mauricius Andrevodi scol. acPetrus Beiler et Anthonius Jeimiiclerici Sedun. dioc. (coniugatus/soluta): de disp. (f.d.s. D.) Rome 26.mart. 55 sede vacante
167060 "label": "Lockerung von Fastenvorschriften"
167061 },
167062 {
167063 "lemma_id": "11102785",
167064 "text": "Johannes Weyshorn scol. de Inferiori Reuxingen Spiren. dioc. (coniugatus/coniugata): de disp. (f.d.s. D.) Rome 26. mart. 55 sede vacante PA 4 241r.",
167065 "label": "Tötungsdelikte"
167066 }
167067 ]
```

4. Wie gut sind die Ergebnisse?

Evaluation

- 9.070 Bittschriften / Zuweisungen
- 17 Kategorien
- Händische Überprüfung von 120 Zuordnungen

	ID	Vorhersage KI	Korrekte Kategorie
RPG 1 (die ersten 5)	11000001	Gottesdienst trotz Exkommunikation/Interdikt	Klosterangelegenheit
	11000002	Übriges	Kirchliche Karriere
	11000003	Gottesdienst trotz Exkommunikation/Interdikt	Übriges (def. nat.)
	11000004	Körperverletzung	Übriges (matr.)
	11000005	Gottesdienst trotz Exkommunikation/Interdikt	Kirchliche Karriere
RPG 1 (5 in der Mitte)	11000385	Körperverletzung	Körperverletzung
	11000386	Gottesdienst trotz Exkommunikation/Interdikt	Tötungsdelikte
	11000387	Körperverletzung	Körperverletzung
	11000388	Klosterangelegenheiten	Klosterangelegenheiten
	11000389	Tötungsdelikte	Raubzüge/Kriegsbeteiligung
RPG 1 (die letzten 5)	11000771	Gottesdienst trotz Exkommunikation/Interdikt	Kirchliche Karriere
	11000772	Gottesdienst trotz Exkommunikation/Interdikt	Kirchliche Karriere
	11000773	Kirchliche Karriere	Kirchliche Karriere
	11000774	Übriges	Kirchliche Karriere
	11000775	Gottesdienst trotz Exkommunikation/Interdikt	Besuchserlaubnis für Pilgerstätten

4. Wie gut sind die Ergebnisse?

RPG Band	Korrekte Zuweisungen	Falsche Zuweisungen
1	4	11
2	9	6
3	9	6
4	9	6
5	8	7
6	5	10
7	6	9
8	10	5
Insg.	60 (50%)	60 (50%)

4. Wie gut sind die Ergebnisse?

Versuch: Anpassung des Sprachmodells

- **Hypothese:** Die erste KI hat nur die ersten 30 Worte eines jeden Registereintrags beachtet. Längere Bittschriften, die ihre Thematik i.d.R. zum Ende hin preisgeben, werden falsch zugeordnet.
- **Überprüfung:** Training eines neuen Modells mit einem Kontextfenster von ca. 300 statt 30 Wörtern, Anwendung auf die 120 zuvor evaluierten Fälle
- **Ergebnis:** 48,8% Genauigkeit

4. Wie gut sind die Ergebnisse?

Versuch: Anpassung des Sprachmodells

- **Deutung:** Overfitting

„Overfitting ist, wenn das Modell einfach die Texte aus den Trainingsdaten auswendig lernt, anstatt allgemeinere Rückschlüsse zu ziehen, die sich auch auf andere Texte übertragen ließen. [...].

Overfitting erkennt man oft daran, dass die Leistung beim Training (92%) stark von der Leistung bei neuen Texten (50%) abweicht.“
(Konstantin Schulz)

4. Wie gut sind die Ergebnisse?

Und jetzt?

- Erwartungsmanagement: erster Versuch mit 50% Treffsicherheit bei 17 möglichen Kategorien nicht schlecht
- weitere Problemanalyse, z.B. durch größer angelegte Evaluation der ersten Ergebnisse
 - „Automatisierte Zuordnung funktioniert besser mit Band 8 als mit Band 1.“
 - „Automatisierte Zuordnung funktioniert besser mit Körperverletzung als mit Tötungsdelikten.“
- anderes KI-Basismodell nutzen mit besserer Dokumentation
 - aktuelles KI-Basismodell ist quasi anonym veröffentlicht, d.h. ohne Dokumentation des Entstehungsprozesses
 - Problem: Dadurch ist die Bewertung von Vor- und Nachteilen des Modells stark erschwert.

4. Wie gut sind die Ergebnisse?

Und jetzt?

- Trainingsmaterial anpassen
 - Worttrennung im Digitalisat oft mangelhaft
 - Auflösung von Abkürzungen im Text → Domänenwissen notwendig
 - Kategorien zusammenfassen, z.B.:
 - „Gewaltdelikte“: Tötungsdelikte, Körperverletzung, Raubzüge/Kriegsbeteiligung
 - „Sondererlaubnis“: Lockerung von Fastenvorschriften, Besuchserlaubnis von Pilgerstätten, Lockerung/Umwandlung von Gelübden, Beichtprivilegien, Sonderregel Messe

- Repertorium Poenitentiariae Germanicum IV (1996). Verzeichnis der in den Supplikenregistern der Pönitentie Pius' II. vorkommenden Personen, Kirchen und Orte des Deutschen Reiches (1458-1464), hrsg. vom Deutschen Historischen Institut in Rom; Text bearb. von Ludwig Schmugge mit Patrick Hersperger u. Béatrice Wiggerhauser. Indices bearb. von Hildegard Schneider-Schmugge und Ludwig Schmugge, Max Niemeyer Verlag: Tübingen.
- Salonen, Kirsi; Schmugge, Ludwig (2009). A Sip from the Well of Grace: Medieval Texts from the Apostolic Penitentiary. Studies in Medieval and Early Modern Canon Law, Vol.7. The Catholic University of America Press.
- Schmugge, Ludwig; Hersperger, Patrick; Wiggerhauser, Béatrice (1996). Die Supplikenregister der päpstlichen Pönitentie aus der Zeit Pius' II. (1458-1464). Max Niemeyer Verlag: Tübingen.

Informationen zum Daidalos-Projekt	Dokumentation der vorgestellten Textklassifikation
https://daidalos-projekt.de/	https://scm.cms.hu-berlin.de/daidalos/daidalos-share/-/tree/main/tandems/NF?ref_type=heads
