



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

## Artikel- separation

### Kontakt:

Roger Labahn

✉ roger.labahn@uni-rostock.de

### LEON XIII

«*Chien amoureux de Rome, le pape légendaire qui traita, par ses paroles, avec les nations et les temps, à son tour, l'œuvre responsable de son lieu.*»

Rome, 14 mars.

Monsieur le rédacteur en chef,  
Le Monde publie, il y a quelques jours, un télégramme de son correspondant de Rome, éclairant certains recrois de l'histoire récente des rapports entre le Vatican et le Quirinal.

Permettez-moi de venir à mon tour mettre sous les yeux de vos lecteurs tout un chapitre de cette histoire, en vous garantissant une exactitude même dans les plus minimes détails.

Avant de prendre place au Saint-Colège, Léon XIII était d'abord de Perugin, où il avait sa société toute les sympathies. Tandis que Mgr. Pucci, commandant, se chargea de Perugin, un général de grand talent et de beaucoup d'expérience, le général Cardini commandait la division militaire, dont le siège était tout à côté de l'Arco. Les deux chefs eurent occasion de se rencontrer et de s'apprécier mutuellement — ils échangeaient des idées — bref, ils devinrent bons amis.

L'événement qui donna naissance à la général Cardini est qu'il y avait eu les Vénitien d'un Pape. Le conflit entre le gouvernement italien et le Vatican demandait de lui le plus sage et les deux amis, sincèrement amis, en déclinèrent les conséquences.

Le général, dont l'éloquence s'élevait n'est pas oubliée par les vices, par les bouleaux, où il y a rien pendant quelques années, disant tout le monde par son

pour saluer le nom d'un homme que le gouvernement italien avait officiellement reconnu.

Le Pape, après, cependant, à la fois le système d'Italie avec Rome capitale, et devant sa félicitation aux troupes de son.

Il y avait encore une chose, qui n'intéressait ni la requête, ni le système, mais qui intéressait beaucoup le cardinal de Pietro — elle n'était pas bien longue — la voici.

«Le roi payera les dettes du cardinal de Pietro.»

Le général Cardini avait obtenu l'adhésion du cardinal Pucci au plan arrangé entre Victor Emmanuel et le Pape, mais, malheureusement, tout ne marcha pas à souhait.

On commença par payer les dettes du cardinal — ce qui n'était pas son petit affaire — mais le roi donna l'ordre, en suite du Parlement obligea M. Nicotri à donner sa démission de ministre de l'Intérieur.

Le personnel de police, lequel, qui avait conduit les négociations avec le Vatican, se retira avec le ministre, ne voulant pas servir son successeur, M. Crispi, dans lequel n'avait pas une grande confiance.

Le roi, qui n'était pas non plus M. Crispi, se parla bien de lui, mettre un contrat de sa politique nationale.

Il attendit le retour de M. Nicotri au pouvoir. En 1891, avait aussi de lire, et on n'avait pas trop peur.

Mais les événements se précipitèrent — Victor Emmanuel tomba malade et fut emporté presque subitement, sans pouvoir donner d'instructions à son fils — et sans informer ses ministres de ce qu'il avait préparé.

Quelques semaines après, le roi mourut

### APRES BAC-NINH

LES CONSÉQUENCES DE LA VICTOIRE

A LA POURSUITE DE L'EN-  
NEMI

Dispersion des troupes vietnamiennes — Les  
Chinois au sud — Les Français  
Nuits à l'ouest.

(par un témoin)

L'ouest, 14 mars. On mande de Hai-  
Phong, 14 mars, 7 h. 15 du matin.

«La colonne du général Négrier est en-  
trée à Bac-Ninh dans la soirée du 12.

«L'ennemi, dans la soirée de sa voir com-  
pter la retraite, s'est retiré vers la frontière, laissant le cadavre sans défense.

«Les Français, pendant leur marche en  
avant, n'ont éprouvé qu'une faible résis-  
tance.

«Leurs pertes sont insignifiantes.  
«Leur succès est complet.»

Marche en Avant.  
(par un témoin)

L'ouest, 14 mars. On télégraphie de  
Hai-Phong au Daily News, 14 mars. «Le  
moitié de la brigade du général Millot a  
attaqué et enlevé cinq forts et quatre vil-  
lages sur la route stratégique de Bac-Ninh.

«Les Chinois possèdent des fortifications  
extrêmement construites et armées sur  
les collines, mais ils n'ont opposé aucune  
résistance.

de toute. Aucune déclaration que cette affaire  
n'est qu'un demi-succès, selon le mot de  
Négrier. Tel est, ainsi l'avis de l'ensemble  
des officiers, constaté par le journal. Enfin,  
et par la Liberté, qui conclut, au gou-  
vernement «d'insister les uns de la ré-  
sistance et de la suite.»

Marche des Colonnes Millot et Négrier.  
(par un témoin)

L'ouest, 14 mars. «Le correspondant  
du Times, attaché à la colonne du général  
Millot, a télégraphié, à la date du 13. Après  
une marche difficile de quatre jours à tra-  
vers des champs de riz, évitant les routes  
directes, nord et sud du canal des Rapides,  
les troupes françaises ont passé le canal à  
un endroit au sud-est de Bac-Ninh où pré-  
cédemment il y avait déjà un poste.

«Le général Négrier a pris aujourd'hui  
cinq forts établis sur les hauteurs. Une faible  
résistance lui a été opposée. Les Fran-  
çais n'ont subi aucune perte. La santé et  
l'esprit des troupes sont excellents. Les  
Chinois ont au nombre de 2.000 hommes.  
La poursuite se continue.

Les Morts et Blessés de «Xacouang».

«Voici les noms des hommes de la com-  
pagnie de Mousquetiers, tués ou blessés, dans la  
journée du 12 mars, au Tonkin, sous les ordres  
des chefs de transmission.

Diaz, médecin, tué.  
Léon, second-maître, blessé.  
Bénédicte, matelot, blessé.

Bei der Artikelseparation (AS) werden Regionen auf einer Zeitungsseite identifiziert, die Informationen zu einem Artikel enthalten. Sie bildet die Grundlage für weitere Anwendungen im NewsEye-Workflow aus Bereichen des Natural Language Processing.

Für die AS verwenden wir einen zweistufigen Ansatz: Zunächst werden die Textzeilen aus der

Layout-Analyse für die Automatische Texterkennung zu Textblöcken zusammengefasst. Unsere Forschung konzentriert sich auf den anschließenden Schritt, das Zusammenfassen (Clustern) von Textblöcken zu Artikeln. Neben dem Einsatz traditioneller Clustering-Methoden werden vor allem Algorithmen aus dem Bereich des maschinellen Lernens eingesetzt.



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

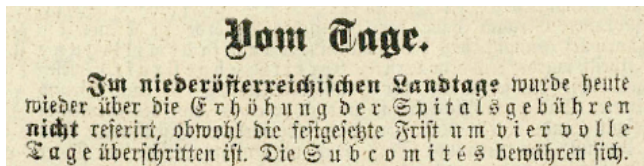


# Automatische Text- erkennung

## Kontakt:

Roger Labahn

✉ [roger.labahn@uni-rostock.de](mailto:roger.labahn@uni-rostock.de)



## Vom Tage.

Im niederösterreichischen Landtage wurde heute wieder über die Erhöhung der Spitalsgebühren nicht referirt, obwohl die festgesetzte Frist um vier volle Tage überschritten ist. Die Subcomités bewähren sich.

*Die Automatische Texterkennung liefert eine Transkription für jede Textzeile*

Die automatische Texterkennung (Automated Text Recognition, ATR) hat die Aufgabe, die in Bildern enthaltenen Textinhalte automatisch zu transkribieren. In diesem grundlegenden Schritt des Workflows werden die Textdaten extrahiert, welche in darauf folgenden Arbeitsschritten bei der Verarbeitung natürlicher Sprachelemente Anwendung finden.

In unserem Fall wenden wir Modelle auf segmentierte Textzeilenbilder an, die das Ergebnis eines Layoutanalyse-Algorithmus sind. Diese Modelle umfassen tiefe künstliche neuronale Netze, die für Sequenz-Labeling-Aufgaben entwickelt wurden. Zusätzlich enthalten die Ausgaben solcher ATR-Modelle Schätzungen für Wahrscheinlichkeiten der einzelnen Zeichen, die zur Suche in der Transkription mittels Keyword-Spotting verwendet werden können.



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu



# NewsEye Demonstrator

## Kontakt:

Antoine Doucet

✉ antoine.doucet@univ-lr.fr



Der NewsEye-Demonstrator (<https://platform.newseye.eu>) stellt die verschiedenen Ergebnisse des Projektes vor. Seine frei zugängliche Webseite ermöglicht es den Benutzern\*innen, Zeitungssammlungen in verschiedenen Sprachen zu erkunden. Diese Zeitungen, aus dem Zeitraum von 1850 bis 1950 stammend, wurden von den

drei Nationalbibliotheken des Projektes aus Wien, Paris und Helsinki bereitgestellt. Nach der Suche in den Sammlungen können die Benutzer\*innen persönliche Sammlungen anlegen, um für sie interessante Dokumente zu verwalten. Diese können anschließend in verschiedenen Formaten exportiert oder direkt im Demonstrator analysiert werden.



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu



## Semantische Textan- reicherung von OCR-Inhalten

### Kontakt:

Antoine Doucet

✉ antoine.doucet@univ-lr.fr

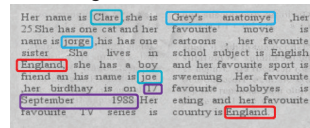
In der semantischen Textanreicherung werden Dokumente analysiert und anschließend semantische Metadaten zu deren Inhalten hinzugefügt. Bei der Vielzahl an semantischen Metadaten legen wir den Fokus auf benannte Entitäten (named entities). Genauer gesagt, wollen wir diese Entitäten in Dokumenten erfassen und zu einer Wissensdatenbank zusammenführen. Zusätzlich werden diese Entitäten mit der Haltung des Textes, in dem sie erwähnt werden, in Verbindung gebracht.

Je nach Zustand der Dokumente, bedingt durch Alterung, schlechte Lagerbedingungen und/oder die schlechte Qualität der verwendeten Druckermaterialien, kann die OCR jedoch zahlreiche Fehler aufweisen. Diese Fehler reduzieren die Leistung aller nachfolgenden Verfahren der natürlichen Sprachverarbeitung (z. B. NER, NEL und Stance Detection).

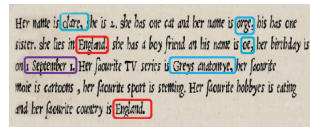
Als Lösung dieser Probleme werden in NewsEye Ansätze auf der Grundlage neuester Techniken (Neuronale Netze und Deep Learning) entworfen, die robust gegenüber OCR-Problemen

und sprachunabhängig sind. Die Ansätze von NewsEye erreichten bei der CLEF HIPE 2020 Competition zu NER- und NEL-Aufgaben in Englisch, Französisch und Deutsch von insgesamt 13 Teilnehmern in 50 von 52 Bestenlisten den ersten Platz.

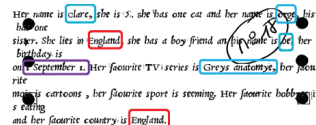
### Angemessene Qualität



### Mittlere Qualität



### Schlechte Qualität







A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

## Case Study: Gender

## Frauen (in Hosen)

**Kontakt:**

Die Case Study zu Frauen und Gender untersucht die Entwicklung der Repräsentation von Frauen in den Nachrichten zwischen 1850 und 1950. Wir interessieren uns für die Stellung der Frau in der Gesellschaft (das Frauenwahlrecht,



aber auch das Recht auf das Tragen von Hosen) sowie für jene Frauen, welche die journalistische Sphäre betraten: Journalistinnen und Schriftstellerinnen. Welchen Platz nahmen sie in den Zeitungen ein? Welche Rolle spielten die Zeitungen im Kampf um Emanzipation? Digitale Tools (z. B. Topic Modelling und Frequenzanalysen) sind in diesem Zusammenhang sehr hilfreich, um versteckte Entwicklungen und neue Trends zu entdecken.



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

# NEWS E E

## Case Study: Medien & Journalismus

Die Interaktion  
zwischen Redaktion  
und Leserschaft

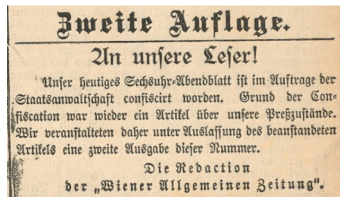
### Kontakt:

Eva Pfanzelter

✉ Eva.Pfanzelter@uibk.ac.at

## An die künftigen Abonnenten.

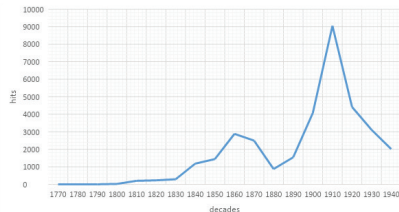
Hast du nie ein Schiff vom Stapel laufen sehen, lieber Leser? Es ist ein interessanter Anblick, und gar manche Betrachtungen und Gedanken, heiz



Wiener Allgemeine Zeitung, 13. November  
1880, Seite 1, Abendzeitung

Im „langen“ 19. Jahrhundert wurden Zeitungen zum Massenmedium schlechthin. Gleichzeitig entstand dadurch ein neuer öffentlicher Raum, in dem die Verhandlungen des modernen Nationalstaates eine wichtige Rolle spielten. Die gängige Wahrnehmung, dass es sich bei der Nachrichten- und

Meinungsverbreitung der Presse in dieser Zeit um einen einseitigen, linearen Prozess handelt, in dem die Information von der Redaktion zur Leserschaft getragen wird, ist jedoch nicht zur Gänze korrekt. Es gab unterschiedliche Möglichkeiten der Interaktion und zweiseitigen Kommunikation. In den Blick genommen wird daher in dieser Fallstudie diese Interaktion.



Absolute Häufigkeit der Phrase „An die Redaktion“  
nach Jahrzehnt



A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

# NEWS E E

## Case Study: Nationalismen

Die Sprache des  
Nationalismus  
1850-1950

### Kontakt:

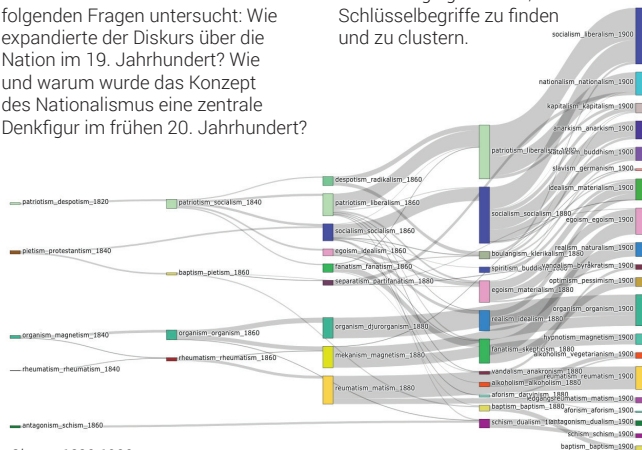
Jani Marjanen

 jani.marjanen@helsinki.fi

*Finska Språket, betraktadt såsom Nationalspråk.*

Die Case Study zu Nationalismen konzentriert sich auf die sich verändernde Sprache der Nation im späten 19. und frühen 20. Jahrhundert. Es werden die folgenden Fragen untersucht: Wie expandierte der Diskurs über die Nation im 19. Jahrhundert? Wie und warum wurde das Konzept des Nationalismus eine zentrale Denkfigur im frühen 20. Jahrhundert?

In der Fallstudie wird sowohl mit qualitativer Analyse von Quellentexten als auch mit Frequenzanalysen und Word Embeddings gearbeitet, um Schlüsselbegriffe zu finden und zu clustern.





A Digital Investigator for  
Historical Newspapers

In diesem Projekt arbeiten gemeinsam  
Geisteswissenschaftler\*innen,  
Informatiker\*innen und Bibliothekar\*innen  
an historischen Zeitungen  
aus drei Nationalbibliotheken  
(Österreich, Finnland, Frankreich)

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

## Case Study: Migration

Rückkehrmigration  
1850-1950

### Kontakt:

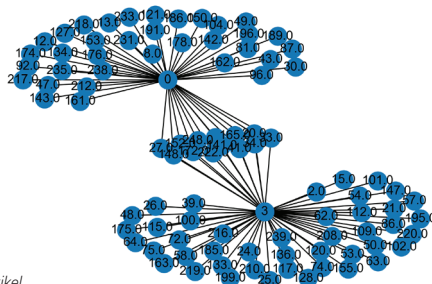
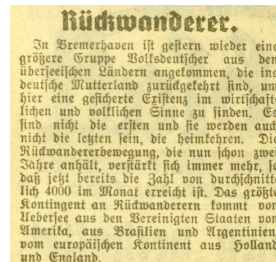
Sarah Oberbichler

✉ [sarah.oberbichler@uibk.ac.at](mailto:sarah.oberbichler@uibk.ac.at)

## Sie sind da, sie sind da!

Die Case Study zu Migration versucht, Prozesse der Remigration nach Europa zwischen 1850 und 1950 aufzuspüren, indem folgenden Fragen nachgegangen wird: Wie und in welchem Kontext berichteten österreichische Tageszeitungen über Rückkehrende und wie änderte sich diese Berichterstattung im Laufe der Zeit?

Die Fallstudie setzt digitale Methoden wie Topic Modelling und Word Embedding ein, um Subkorpora für weitere qualitative Analysen zu bilden (Diskurs- und Argumentationsanalyse).



Topic-Visualisierung für  
relevante und nicht-relevante Artikel