



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 [@newseyeeu](https://twitter.com/newseyeeu)

✉ roger.labahn@uni-rostock.de



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 [@newseyeeu](https://twitter.com/newseyeeu)

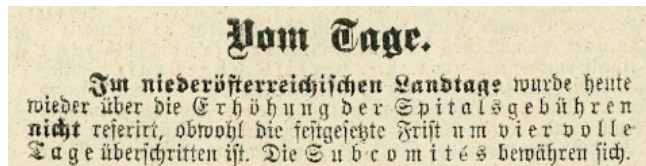


Automated Text Recognition

Contact:

Roger Labahn

✉ 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.

Automatic Text Recognition provides a transcription for each text line

Automatic Text Recognition (ATR) is the task of automatically transcribing the textual contents contained in images. This is a fundamental step in the NewsEye workflow, as it provides the textual data which is needed to further be processed in Natural Language Processing tasks.

In our case, we apply models on segmented text line images, which are the results of a Layout Analysis algorithm. These models involve deep Artificial Neural Networks designed for sequence labeling tasks. Additionally, the results of such ATR models include posterior character probabilities, which can be used to search in the transcription by means of keyword spotting.



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 @newseyeeu



NewsEye Demonstrator

Contact:

Antoine Doucet
antoine.doucet@univ-lr.fr



The NewsEye demonstrator (<https://platform.newseye.eu>) showcases the different outputs of the project. Its website, freely available, allows users to explore newspaper collections in different languages. Ranging from 1850 to 1950, these newspapers were made available by three National Libraries

of the project in Vienna, Paris and Helsinki. After searching through the collection, users can create personal collections to gather documents of interest. These collections can be exported in various formats or analysed directly inside the Demonstrator.



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 [@newseyeeu](https://twitter.com/newseyeeu)



Semantic Text Enrichment for OCR'd content

Contact:

Antoine Doucet

✉ antoine.doucet@univ-lr.fr

Semantic text enrichment consists of analyzing documents and adding semantic metadata to their contents. Among several semantic metadata, we concentrate on named entities and, more precisely, we aim to recognize these entities in documents and disambiguate them to a knowledge base. In addition, these entities will be associated with the stance of the text in which they are mentioned.

However, OCR'd documents contain numerous errors due to the state of documents, as a result of aging, poor storage conditions and/or the low quality of initial printing materials. These errors reduce the performance of all downstream natural language processing tasks (such as NER, NEL and stance detection).

To overcome these problems, NewsEye designs approaches based on the most recent techniques (neural networks and deep learning) that are robust to OCR problems and language independent. Out of 13

competitors, NewsEye approaches reached first place in 50 of 52 leaderboards on NER and NEL in English, French and German at the CLEF HIPE 2020 competition.

Reasonable quality

Her name is **Clare** she is 25. She has one cat and her name is **George**. His has one sister. She lives in **England**. She has a boy friend and his name is **Joe**. Her birthday is on **17 September 1988**. Her favourite TV series is **Grey's Anatomy**. Her favourite movie is **Cartoons**. Her favourite school subject is **English** and her favourite sport is **swimming**. Her favourite hobby is **eating** and her favourite country is **England**.

Medium quality

Her name is **Clare**, he is 25. She has one cat and her name is **George**. His has one sister. She lies in **England**. She has a boy friend and his name is **Joe**. Her birthday is on **17 September 1**. Her favourite TV series is **Greys anatomy**. Her favourite movie is **cartoons**. Her favourite sport is **swimming**. Her favourite hobby is **eating** and her favourite country is **England**.

Poor quality

Her name is **Clare**, she is 25. She has one cat and her name is **George**. His has one sister. She lies in **England**. She has a boy friend and his name is **Joe**. Her birthday is on **17 September 1**. Her favourite TV series is **Greys anatomy**. Her favourite movie is **cartoons**. Her favourite sport is **swimming**. Her favourite hobby is **eating** and her favourite country is **England**.



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 @newseyeeu

vote, the right to wear pants), as well as the female figures who infiltrate the journalistic space: female writers/journalists. What place is given to them in the newspapers? What role do newspapers play in the struggles of women? Digital tools (notably topic modeling and frequency tools) are very valuable in this case study for revealing the trends and key figures.



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 @newseyeeu

NEWS E E

Case Study: Media & Journalism

The interaction
between editorial
office and readers

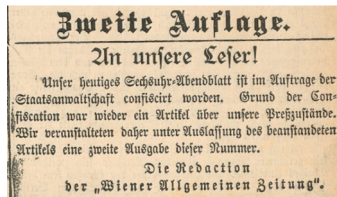
Contact:

Eva Pfanzelter

✉ Eva.Pfanzelter@uibk.ac.at

An die künftigen Abonnenten.

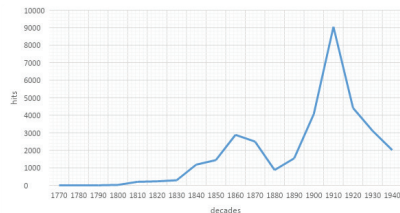
Hast du nie ein Schiff vom Stapel laufen sehen, lieber Leser? Es ist ein interessanter Anblick, und gar manche Betrachtungen und Gedanken, heiz



Wiener Allgemeine Zeitung, 13 November 1880, page 1, evening paper

In the “long” 19th century, newspapers became the mass medium par excellence. At the same time, this created a new public sphere, where the negotiations of the modern nation state played an important role. However, the common perception that the dissemination

of news and opinion by the press during this period was a one-way, linear process in which information was carried from the editorial office to the readers is not entirely correct. There were different possibilities of interaction and two-way communication. This case study therefore focuses on this interaction.



Absolute frequency of the phrase „an die Redaktion“ („to the editors“) per decade



A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 @newseyeeu

NEWS E E

Case Study on Nationalism

The language of Nationalism
1850-1950

Contact:

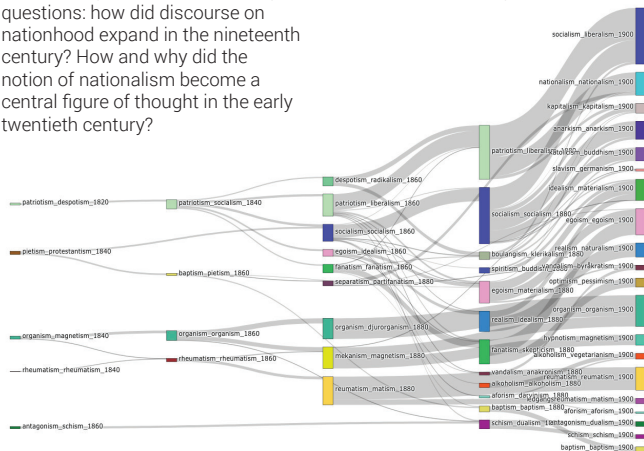
Jani Marjanen

✉ jani.marjanen@helsinki.fi

Finska Språket, betraktadt såsom Nationalspråk.

The case study on nationalism focuses on the changing language of nationhood in the late 19th and early 20th century. It asks the following questions: how did discourse on nationhood expand in the nineteenth century? How and why did the notion of nationalism become a central figure of thought in the early twentieth century?

The case combines qualitative reading of sources with frequency analysis and word embedding models to cluster key terms.





A Digital Investigator for Historical Newspapers

In this project, humanities researchers, computer scientists, and librarians work together on historical newspapers of three national libraries (Austria, Finland, France)

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

 www.newseye.eu

 @newseyeeu

Case Study on Migration

Return Migration 1850-1950

Contact:

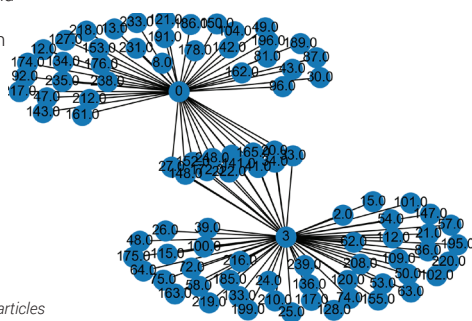
Sarah Oberbichler

✉ sarah.oberbichler@uibk.ac.at

Sie sind da, sie sind da!

The case study on migration attempts to reveal return migration processes to Europe between 1850 and 1950. It does so by asking the following questions: how and in what context were Austrian daily newspapers reporting on returnees and how did the reporting change over time?

The case uses digital methods like topic modeling and word embedding to create a subcollection for further qualitative analysis (discourse and argumentation analysis).



Topic visualization for relevant and non-relevant articles

Rückwanderer.

In Bremerhaven ist gestern wieder eine größere Gruppe Volksdeutscher aus den überseeischen Ländern angekommen, die ins deutsche Mutterland zurückgekehrt sind, um hier eine geistigere Erziehung im wirtschaftlichen und politischen Sinne zu finden. Es sind nicht die ersten und sie werden auch nicht die letzten sein, die heimkehren. Die Rückwandererbewegung, die nun schon zwei Jahre anhält, verstärkt sich immer mehr, so daß jetzt bereits die Zahl von durchschnittlich 4000 im Monat erreicht ist. Das größte Kontingent an Rückwanderern kommt von Uebersee aus den Vereinigten Staaten von Amerika, aus Brasilien und Argentinien, vom europäischen Kontinent aus Holland und England.