



HAL
open science

NewsEye: A digital investigator for historical newspapers

Antoine Doucet, Martin Gasteiner, Mark Granroth-Wilding, Max Kaiser, Minna Kaukonen, Roger Labahn, Jean-Philippe Moreux, Guenter Muehlberger, Eva Pfanzelter, Marie-Ève Thérénty, et al.

► **To cite this version:**

Antoine Doucet, Martin Gasteiner, Mark Granroth-Wilding, Max Kaiser, Minna Kaukonen, et al.. NewsEye: A digital investigator for historical newspapers. 15th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2020, Jul 2020, Ottawa, Canada. hal-03029072

HAL Id: hal-03029072

<https://hal.science/hal-03029072>

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NewsEye: A digital investigator for historical newspapers¹

Antoine Doucet¹, Martin Gasteiner², Mark Granroth-Wilding³,
Max Kaiser⁴, Minna Kaukonen³, Roger Labahn⁵,
Jean-Philippe Moreux⁶, Guenter Muehlberger⁷, Eva Pfanzer⁷,
Marie-Eve Therenty⁸, Hannu Toivonen³, Mikko Tolonen³

¹University of La Rochelle; ²University of Vienna; ³University of Helsinki; ⁴National Library of Austria; ⁵University of Rostock; ⁶National Library of France; ⁷University of Innsbruck; ⁸University Paul Valéry Montpellier

The **NewsEye H2020 project**, running from May 2018 until April 2021, is an interdisciplinary undertaking that involves 3 European national libraries, 4 humanities and social science research groups and 4 computer science research groups. **The core concept of NewsEye is a seamlessly integrated armory of tools and methods that will improve the users' capability to access, analyze and use the content in the digital libraries of historical newspapers.**

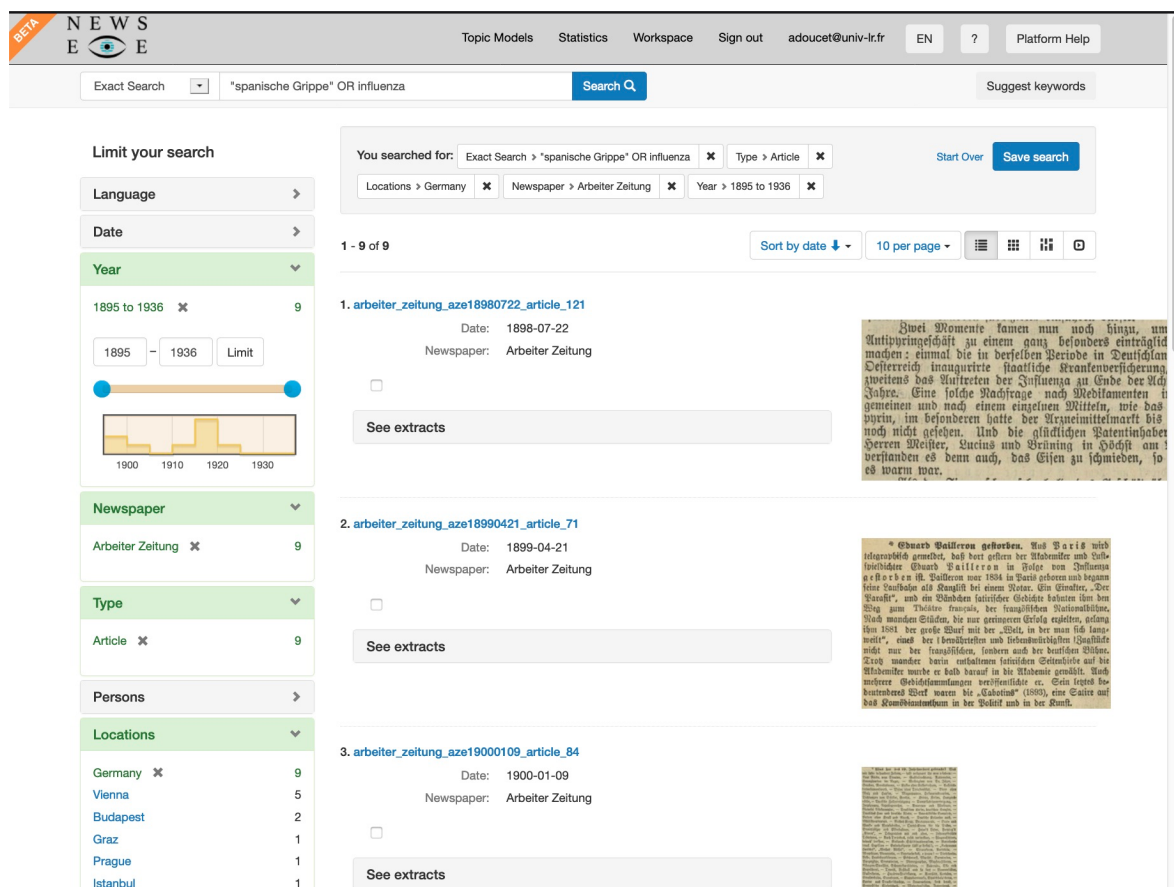


Figure 1: Beta version of the NewsEye demonstrator (June 2020)

¹ This work has been supported by the European Union Horizon 2020 research and innovation programme under grant 770299 (NewsEye).

Specifically, **in the context of historical newspaper written in German, Finnish, Swedish and French, with a focus on the period 1850-1950**, the project aims to develop a toolbox consisting of two main layers, as well as **novel research results on several topics and in several fields of digital humanities, based on documents in different languages, so as to demonstrate the potential extent of its usefulness** as a catalyst for the seamless development of novel research.

In details, the first layer of the **NewsEye toolbox focuses on tools to improve and enrich historical newspapers**, with improved text recognition and article segmentation, followed by semantic enrichment through the recognition and linking of named entities, stance detection, as well as novelty detection. A language-independent set of higher quality data results from this step, already allowing an enriched experience and access to the newspaper collections. The second layer of the toolbox provides ways to benefit from this enriched dataset, through **dynamic text analysis tools** interacting with respect to user activities: contextualized topic modeling, viewpoint and comparative analysis, etc. In addition, an **innovative personal research assistant** is able to design strategies (plans) for finding something interesting and to revise them on the fly when needed. It consists of an investigator (dynamically finding and suggesting novel ideas), a reporter (summarizing the grounds for all suggestions) and an explainer (allowing the user to understand the suggestions by herself, and to return to the original data to confirm or infirm them).

Within the project, several digital humanities case studies are led, with the aim to guide the development of adequate tools, and so as to demonstrate their potential for the development of novel research in digital humanities. **In the case studies, groups of humanities scholars carry out investigations for representative research issues, such as “gender”, “migration”, “nationalism and revolutions”, and “media”**. Since there is plenty of existing qualitative research on these topics, the project strives towards making an impact in the fields of historical research and digital humanities by combining knowledge from qualitative analyses with new findings in big data analyses provided by the new tools in this project.



By Dr. Sarah Oberbichler, University of Innsbruck

More than 50 million Europeans left their homeland voluntarily for overseas destinations between the Napoleonic Wars and the depression of the 1930s to seek a better life. In addition, there were involuntary emigrants such as prisoners of war or deported people as well as (war) refugees (esp. after the First and Second World War). No official numbers exist, but a significant amount of these voluntary and forced emigrants returned to their country of origin. Return migration, defined as the movement of emigrants back to their homelands to resettle, is therefore an important but often neglected aspect of the history of migration.

But not can emigration, be voluntary, so too can return migration; or forced (repatriation, deportation, or banishment). Every person who emigrated voluntarily and later decided to come back to their homeland had their own reasons: success, failure, homesickness, rejection of the new country, or perhaps family members asked them to return. For involuntary emigrants, some reasons may have been similar, but their return was often supported by the country of origin. However, the voluntary or forced remigration of emigrants not only leaves its traces on every person, but also in the country of return and its inhabitants. Newspapers can reveal how returnees were perceived in their countries of return and this is what I wanted to investigate. The steps for this were taken as below, starting with my research questions.

Figure 2: Short description of the case study on return migration

It is essential to understand that the NewsEye research topics and datasets are showcases, and that the seamless inclusion of additional research question is a key ambition. With this in mind, all the tools developed are language-independent, so as to be able to seamlessly integrate further datasets. In fact, **NewsEye is both open to further research cases to be studied using its tools and to the integration of additional datasets and tools through the status of associated partner.**

Further information

[1] NewsEye project Website in 4 languages: <https://www.newseye.eu/> [de] [en] [fi] [fr]

[2] NewsEye demonstrator: <https://platform.newseye.eu/>

[3] NewsEye twitter account: <https://twitter.com/NewsEyeEU>

[4] NewsEye project publications and data sets: <https://zenodo.org/communities/newseye/>

[5] NewsEye project software and source code: <https://github.com/NewsEye/>

[6] NewsEye podcasts: <https://www.univie.ac.at/newseye/>

[7] NewsEye videos: <https://www.youtube.com/channel/UCwEqOk8JRfbJeBYV-ZkPctA/playlists>