NewsEye: A digital investigator for historical newspapers[1]

Antoine Doucet[1], Martin Gasteiner[2], Mark Granroth-Wilding[3],
Max Kaiser[4], Minna Kaukonen[3], Roger Labahn[5],
Jean-Philippe Moreux[6], Guenter Muehlberger[7], Eva Pfanzelter[7],
Marie-Eve Therenty[8], Hannu Toivonen[3], Mikko Tolonen[3]

[1]*University of La Rochelle;* [2]*University of Vienna;* [3]*University of Helsinki;* [4]*National Library of Austria;* [5]*University of Rostock;* [6]*National Library of France;* [7]*University of Innsbruck;* [8]*University 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.**
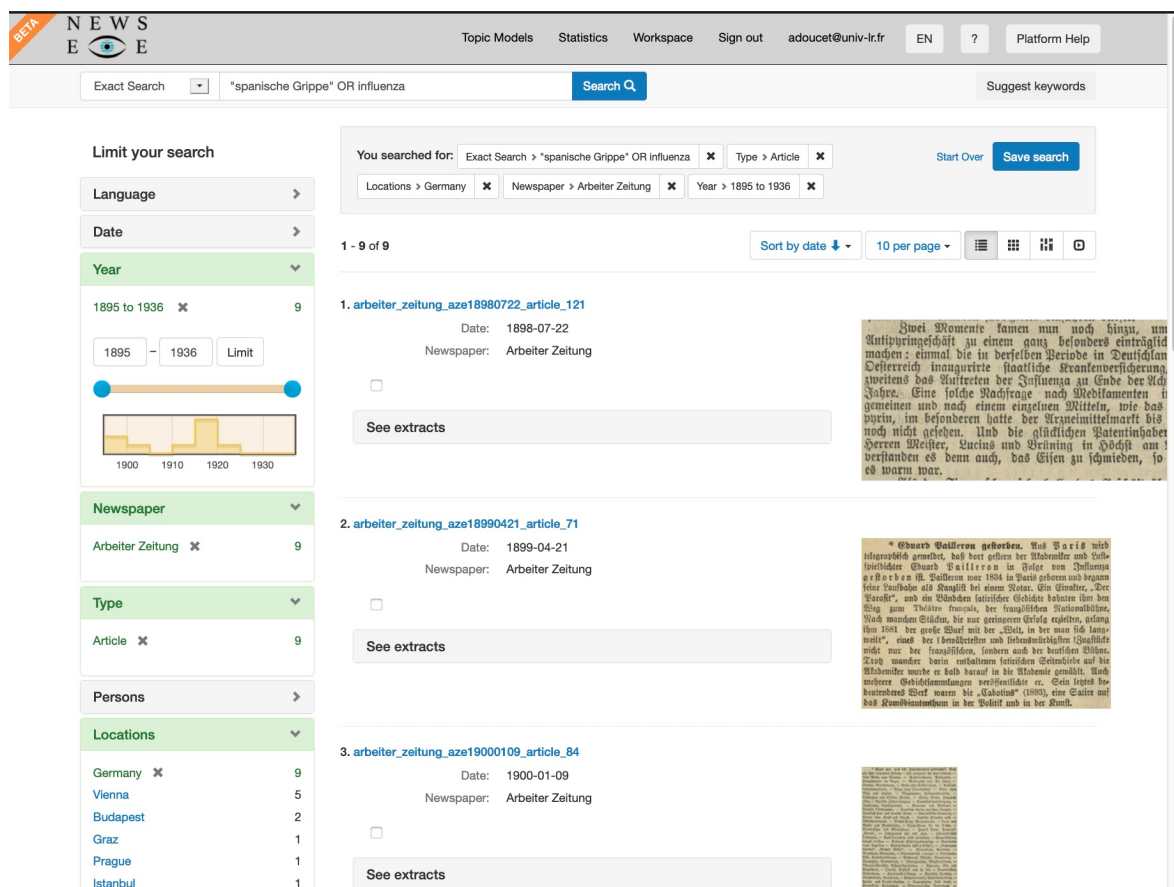

Figure 1: Beta version of the NewsEye demonstrator (June 2020)

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.

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