Project Number: **770299**

NewsEye:

A Digital Investigator for Historical Newspapers

Research and Innovation Action
Call H2020-SC-CULT-COOP-2016-2017

D6.12: Showcase Case Studies for the user interface (b) (final)

Due date of deliverable: M36 (30 April 2020)

Actual submission date: 28 April 2020

Start date of project: 1 May 2018   **Duration:** 45 months

Partner organization name in charge of deliverable: UIBK-ICH

Project co-funded by the European Commission within Horizon 2020		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	-
RE	Restricted to a group specified by the Consortium (including the Commission Services)	-
CO	Confidential, only for members of the Consortium (including the Commission Services)	-

Revision History

Document administrative information	
Project acronym:	NewsEye
Project number:	770299
Deliverable number:	D6.12
Deliverable full title:	Showcase Case Studies for the user interface (b) (final)
Deliverable short title:	Showcase Case Studies (final)
Document identifier:	NewsEye-T64-D125-ShowcaseCaseStudies-b-Submitted-v3.0
Lead partner short name:	UIBK-ICH
Report version:	V3.0
Report preparation date:	28.04.2020
Dissemination level:	PU
Nature:	Report
Lead author:	Sarah Oberbichler (UIBK-ICH)
Co-authors:	Jani Marjanen (UH-DH), Eva Pfanzelter (UIBK), Nejma Omari (UPVM)
Internal reviewers:	Marion Ansel (BNF), Martin Gasteiner (UNIVIE)
Status:	Draft
	Final
	x Submitted

The NewsEye Consortium partner responsible for this deliverable has addressed all comments received, making changes as necessary. Changes to this document are detailed in the change log table below.

Change Log

Date	Version	Editor	Summary of changes made
02/04/2021	0.1	Sarah Oberbichler (UIBK-ICH)	First Draft
08/04/2021	0.2	Jani Marjanen (UH-DH)	Additions and light editing
12/04/2021	0.3	Sarah Oberbichler (UIBK-ICH)	Additions and light editing
14/04/2021	1.0	Eva Pfanzelter (UIBK-ICH)	Proof read and small corrections
20/04/2021	1.1	Marion Ansel (BNF)	Internal review
21/04/2021	1.2	Martin Gasteiner (UNIVIE)	Internal review
21/04/2021	1.3	Sarah Oberbichler (UIBK-ICH)	Revision based on comments from reviewers; Sent to Stefan Hechl for proof reading
24/04/2021	1.4	Stefan Hechl (UIBK-ICH)	Proofreading
26/04/2021	1.5	Sarah Oberbichler (UIBK-ICH)	Small corrections; prepared for quality management
26/04/2021	2.0	Eva Pfanzelter (UIBK-ICH)	Sent to quality manager and PI
26/04/2021	2.1	Nada Lavrac (Qualtity Manager)	External review
28/04/2021	3.0	Antoine Doucet (ULR)	Minor adjustments and submission

Executive summary

Within the project 'NewsEye: A Digital Investigator for Historical Newspapers', researchers from computer science and digital humanities have collaborated with the three national libraries of Finland, France and Austria in order to develop methods and tools for effective exploration and exploitation of digital newspaper collections. The aim is to improve the ways researchers and experts, as well as the interested general public, study European cultural heritage. To make these rich resources of cultural heritage better accessible by means of new technologies and 'big data' approaches, qualitative and quantitative methods of digital humanities have been investigated and combined.

Deliverable D6.12 is the final deliverable of Task T6.4 'Showcase case studies for the user interface'. It reports on the showcases published on the NewsEye Demonstrator, the NewsEye website or on Twitter, which were the public face of the digital humanities (DH) group's work within the project. The aim of the showcases is to present how the DH group used the tools and how it came to its results based on the project work and output. Having identified different case studies (digital history, migration, nationalism, gender, and media/journalism), the DH teams of the universities of Innsbruck, Helsinki, Paul-Valéry-Montpellier and Vienna created screencasts, NewsEye *Jupyter* notebooks, case study descriptions, blog posts and Twitter threads publicly available for different purposes and user groups. This report summarises the output.

Contents

1 Introduction

The creation of showcases, supervised by the department of contemporary history at the University of Innsbruck, was an important deliverable of the DH teams within the NewsEye project. Besides being valuable for academic use, the showcases elaborated by DH members of the universities of Innsbruck (UIBK), Helsinki (UH-DH), Paul-Valéry-Montpellier (UPVM) and Vienna (UNIVIE), as well as computer science members from the University of La Rochelle (ULR), are important in order to understand the tools and methods generated in the NewsEye project. The central aim of this report is to give an account of the published showcases.

To show how the digital humanities groups used the tools and how it came to its results is a benefit that was pursued by either showcasing the case studies on the NewsEye Demonstrator interface[1] (see Figure 1), the NewsEye website[2], or on Twitter. These different platforms were chosen in order to address different user groups. Not every platform is suitable for each type of material, which is why the Demonstrator was chosen to publish screencasts that guide users through the features and tools in the user interface. The NewsEye website was chosen to publish case study descriptions, blog posts and *Jupyter* notebooks. They are aimed at both academics who are looking for content related to their research or teaching, or are interested in the topics investigated by the DH teams (migration, gender, nationalism and media), but also at a general audience interested in the project or digital historical newspapers. The NewsEye Demonstrator screencasts were furthermore included as part of the educational material for the general public. Finally, Twitter was chosen as a the project's social media platform to reach out to people who are not explicitly looking for material on digitised newspapers but are interested in the content the project is sharing.

Different kind of communication styles were used in order to reach these different user groups. For example, case study descriptions and blog posts were kept in an academic writing style. Twitter threads, on the other hand, were aimed at a wide range of interested parties and the communication style was therefore open to a wider audience (civil society, general public, etc.). *Jupyter* notebooks and screencasts for the Demonstrator, finally, were provided in a technical but also explanatory narrative style.

2 NewsEye showcases

All the texts, screencasts, *Jupyter* notebooks and Twitter threads described in this deliverable showcase the expert knowledge of the DH team in working with the Demonstrator and further methods, and inform the aforementioned different user groups in this regard. Helping users to understand and use the tools created in the project was one of the primary motivations for the DH team to create showcases.

2.1 Screencasts on how to use the Demonstrator

Screencasts are a useful way for documenting or publicising interface features such as the NewsEye Demonstrator. The screencasts include video recordings of a computer screen with accompanying narration. The aim of the screenshots is to showcase the demonstrator exemplifying the DH group's case studies. They can be accessed via the button 'Platform help' on the Demonstrator interface,

---

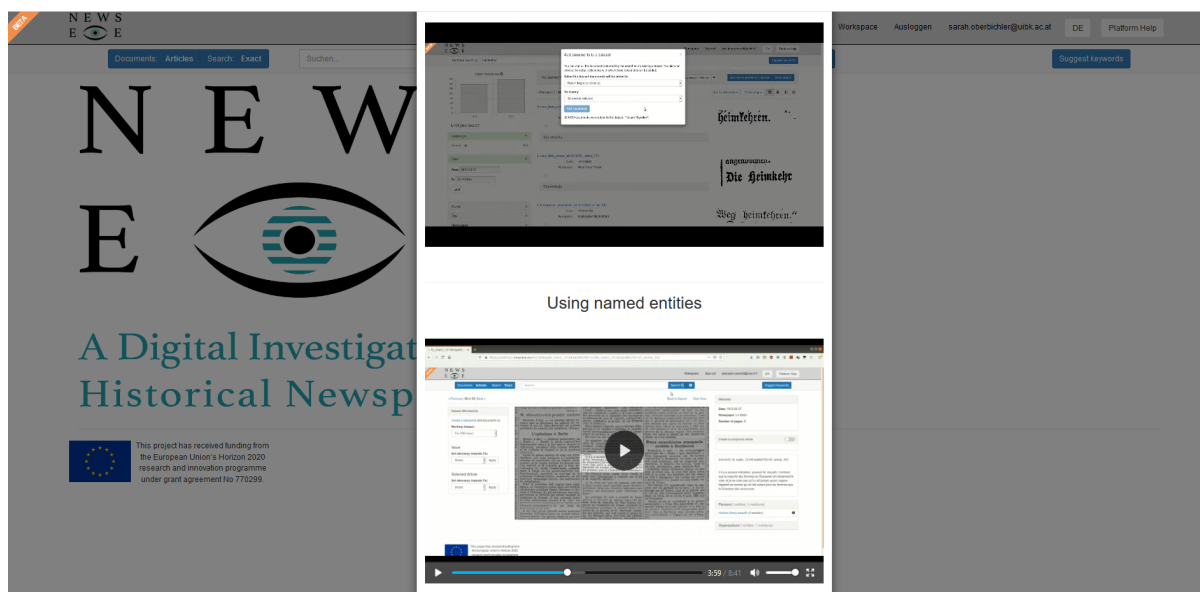[1] https://platform.newseye.eu/
[2] https://www.newseye.eu/

Figure 1: Screencasts as part of the NewsEye Demonstrator interface

but they are also available as part of the educational material for the general public[3] on the NewsEye website and on the NewsEye YouTube channel[4].

In total, three screencasts were created:

- NewsEye Demonstrator in action: creating and managing datasets [1]
  This screencast was created by Sarah Oberbichler (UIBK-ICH), Axel Jean-Caurant (ULR), and recorded by Tonica Hunter (ÖNB). It shows how datasets can be created with the demonstrator. The empirical work on specific research questions must often be based on suitable datasets (subcorpora). Especially historical, but also linguistic research questions require the limitation of the entire newspaper collection to a specific topic or a specific event. Collections are usually created with the help of keyword searches and extraction of the found newspapers, pages or articles. However, the used keywords can be ambiguous. For example, when searching for articles that report on the return of emigrants, prisoners of war etc., the German language does not always offer clear terms. The combined terms '*heimat zurückkehren*' (return home) and '*heimat rückkehr*' (returning home) in an interval of 20 words lead to many results that are related to return migration, but also to a high amount of articles that have nothing to do with this topic. The manual exclusion of irrelevant articles for the topic on return migration is extremely time-consuming and disproportionate to the gain in knowledge, especially the larger the subcorpus becomes. The screenshot shows ways how datasets can be created and organised within the Demonstrator by establishing the relevancy of the found articles.

- Using Named Entities in the NewsEye platform [2]
  This screencast was created by Nejma Omari (UPVM), Axel Jean-Caurant (URL), and recorded by Amanda Maunoury (BNF). The screencast shows how named entities (person names, locations and organisations) in the Demonstrator can be used in order to create a subcorpus for further qualitative analysis on the topic of women's suffrage.

Person entities are interesting for the gender case study because they can indicate the gender of a person if first names or titles are mentioned. In the demonstrator, these named entities appear right next to the newspaper article, which allows us to easily locate them. The work with the named entities is useful to see emerging trends. Certain places, for example, are particularly associated with the history of feminism, such as 'Hyde Park' where 500,000 women gathered to fight for their rights. Organisations can also help to find relevant articles, such as 'suffragettes', 'women's rights' and 'women's suffrage', or even 'Women's Social and Political Union' (WSPU). In addition, named entity linking can help to find relevant named organisations or places in a multilingual setting. Through the linking of recognised entities to, for example, Wikidata, this information becomes available in many different languages. This screencast guides users through a workflow using named entities within the Demonstrator.

- A Digital History Workflow – Emigration in historical newspapers and economic benefits/advertising between 1850 and 1950 [3]
  This video/screencast (see Figure 2) was created by Sarah Oberbichler (UIBK-ICH). It shows a digital history workflow using the Austrian newspapers available in the NewsEye Demonstrator: the *Arbeiter Zeitung*, the *Illustrierte Kronen Zeitung*, the *Innsbrucker Nachrichten*, and the *Neue Freie Presse* were analysed in terms of the economic benefits of emigration between 1850 and 1950. Next to the dataset functions, the tools *Generate Time Series* and *Extract Names* are explained and showcased. In addition, the export functions and the *Jupyter* notebook *Text classification for topic-specific newspaper collections* (see section '*Jupyter* notebooks') play a central role in this short video. The following research questions guide the user through the video: How can newspaper content be categorised and how can advertisements be automatically extracted from a newspaper corpus? What kind of advertisements can be found in reports on emigration? What were people selling or searching for? When and in what context did advertisements appear and how did they change over time? This video guides through a workflow using search functions, dataset functions and analysis tools in the Demonstrator, as well as tools outside the platform.
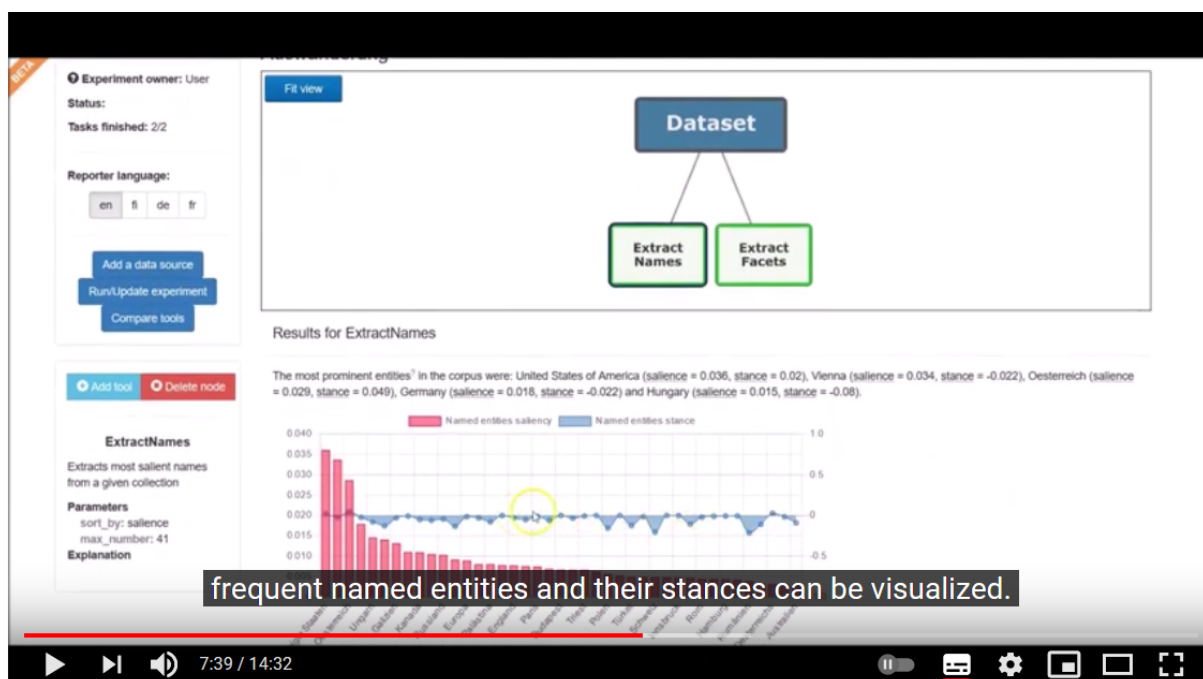

Figure 2: A screenshot of the video/screencast on 'A Digital History Workflow – Emigration in historical newspapers and economic benefits/advertising between 1850 and 1950'

2.2 NewsEye *Jupyter* Notebooks

The NewsEye *Jupyter* Notebooks demonstrate how exported collections from the NewsEye Demonstrator can be analysed using methods outside the interface. *Jupyter* Notebook is an open-source web application that enables users to create and share documents that contain live code, equations, visualisations and narrative text. The NewsEye notebooks describe digital methods for searching, selecting and analysing historical newspapers.

The aim is to showcase computational solutions connected to our case studies. The demos are mainly aimed at academics dealing with similar questions and challenges. The narrative includes code, visualisations, and text. Due to the narrative elements, the notebooks are also suitable for humanities scholars. All notebooks are available on GitHub as NewsEye GitHub repositories[5] and can be accessed via the NewsEye website as part of the open science software repositories[6], or as part of the educational material for university students and teachers[7].

```
In [29]:  import pandas as pd
          import numpy as np
          import networkx as nx
          import matplotlib.pyplot as plt

          # Build a dataframe with 4 connections
          df = pd.DataFrame({ 'from': relevance, 'to': topic})
          df

          # Build your graph
          G = nx.from_pandas_edgelist(df, 'from', 'to')


          # Plot it
          nx.draw(G, with_labels=True)
          plt.show()
```

```
C:\Users\c62255\anaconda3\lib\site-packages\networkx\drawing\nx_pylab.py:579: MatplotlibDeprecationWarning:
The iterable function was deprecated in Matplotlib 3.1 and will be removed in 3.3. Use np.iterable instead.
  if not cb.iterable(width):
```
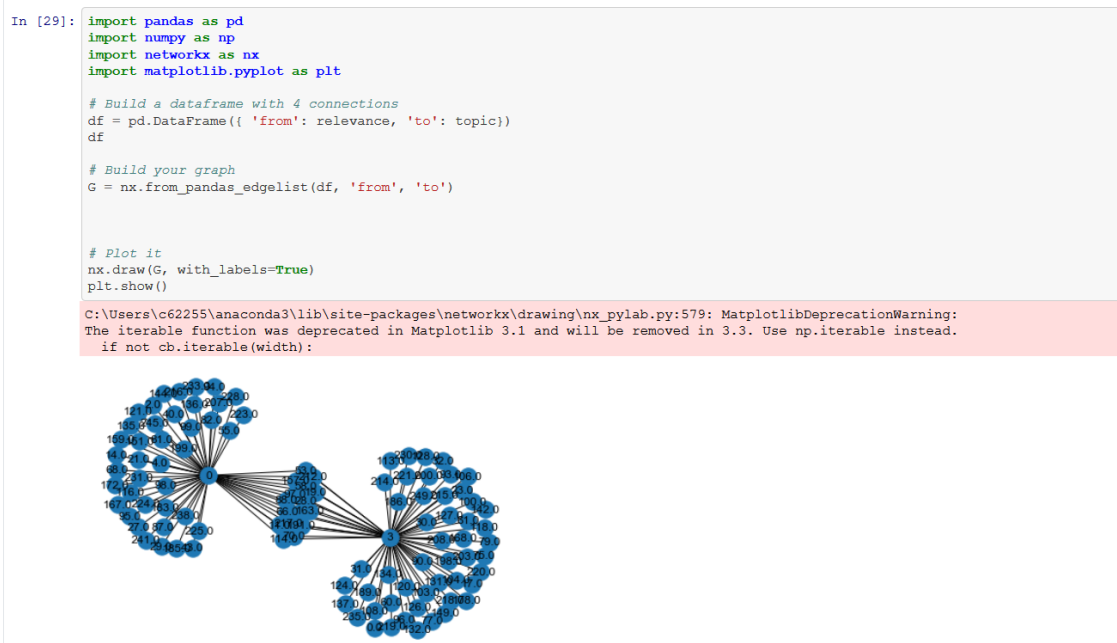

Figure 3: Section of the *Jupyter* Notebook 'Group similar newspaper articles'

Four *Jupyter* notebooks were created by DH team member Sarah Oberbichler (UIBK-ICH), and one by János Békési and Martin Gasteiner (UNIVIE). The *Jupyter* notebooks are intended to support the application of digital methods, to give ideas for additional possibilities for analysis, and to serve as a learning aid for everyone interested in working with digital methods. All notebooks analyse data extracted from the NewsEye Demonstrator:

- Text classification for topic-specific newspaper collections [4]
  This notebook, created by Sarah Oberbichler (UIBK-ICH), was used with a collection of the case study on emigration, and shows how a model can be trained to classify topic-specific collec-

---

[5] https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections
[6] https://www.newseye.eu/open-science/software-repositories/
[7] https://www.newseye.eu/resources/educational-material/syllabi-and-teaching-material-for-universities/

tions. For the training/test corpus, a collection with the keywords '*Auswander**' (emigra*), '*Aus-gewanderte*' (emigrated*), '*Emigrant**' (emigrant*), '*Emigrierte*' (emigrated*), '*Emigration*' (emig-ration), '*Kolonist** (colonist*)', and '*Ansiedler**' (settler) was created. In addition, information on pre-defined news items (news, ads, culture, etc.) were added using numbers between one and ten. For classification, Latent Dirichlet allocation (LDA) was chosen because it showed the best performance in classification (after experiments with word embeddings or a combination of LDA and word embeddings). LDA provides a way to group documents by topic, perform similarity searches and improve precision. Thanks to the Python package *sklearn*, it is relatively easy to test different classifiers for a given topic classification task. Logistic regression was chosen as binary classifier.

- Group similar newspaper articles [5]
  This notebook (Figure 3) was created by Sarah Oberbichler (UIBK-ICH) in order to support the corpus-building process on the topic of return migration. Many search terms connected with this topic are characterised by ambiguity. For example, the German words '*Heimkehr*' (returning home) or '*Rückkehr*' (returning back) lead to many articles that are relevant to the research question, but also to articles that are not relevant at all (e.g. return from a mountain hike, from work, etc.). The notebook provides explanations and code for removing those irrelevant articles from a dataset. It groups similar articles using topic modeling (gensim library) and the Jensen-Shannon distance (JSD) method, and classifies them into relevant or irrelevant ones. Manual annotations are used for evaluation and for counting the relevance probability during the classification process. While topic modeling provides a topic distribution for each article, JSD is used to measure the similarity between the topic distribution of an unseen article and the topic distribution of the training corpus.

- Discover a newspaper collection with diachronic ngram clouds [6]
  Again, this notebook was created by Sarah Oberbichler (UIBK-ICH). Ngrams are connected sequences of n items from a given text or speech sample. This means that words are not considered as individual units, but in relation to each other. For humanities scholars, ngrams can be helpful to get an overview of their collection or to identify discourse markers where a discourse is seen as a group of related texts belonging to a common theme. Ngrams are not a research result per se – which is true for any output of NLP methods – but they can help scholars find important patterns in a particular collection. This notebook shows how diachronic ngrams can be built and visualised (see Figure 4). Diachronic ngrams show the evolution of ngrams by splitting datasets into units (months or years), while the ngrams of every unit are visualised. The notebook was created with the aim of examining the perception of South Tyrol in Austrian newspapers post 1945.

- Discover media discourse in Spanish Flu newspaper coverage with diachronic bigram clouds and topic modeling [7]
  This notebook was created by Sarah Oberbichler (UIBK-ICH) with the aim of using it in workshops and for teaching purposes. It shows how the media discourse on the Spanish Flu can be discovered by diachronic bigrams and topic modeling (gensim library). The notebook was launched via *Binder*, which allows users to turn a *Jupyter* Notebook into an interactive notebook, making the code immediately reproducible by anyone without prior installations. This is especially helpful for digital humanities courses because it allows students to get started with coding and with *Jupyter* Notebooks. The notebook enables the creation of diachronic ngrams and topic visualisations based on exports from the NewsEye Demonstrator.

Figure 4: Visual output of the *Jupyter* notebook 'Discover a newspaper collection with diachronic Ngram clouds'

- Topic Modeling and Uses of the Term Telegraph in the Context of Journalism [8]
  This notebook was created by János Békési and Martin Gasteiner (UNIVIE) with the aim of investigating the usage of the term '*Telegraph*' in Austrian historical newspapers. The data used here is based on a general search for articles on the word '*Telegraph*' in the following time periods: 1864–1874, 1895–1901, 1911–1922. The search was carried out in the following German-language newspapers: The *Neue Freie Presse*, the *Innsbrucker Nachrichten*, the *Arbeiter Zeitung* and the *Illustrierte Kronen Zeitung*. The resulting data package was exported in the JSON format and processed further with regard to topic modelling. The notebooks shows how topic models can be trained and visualised, and how diachronic topic models can be created (see Figure 5).
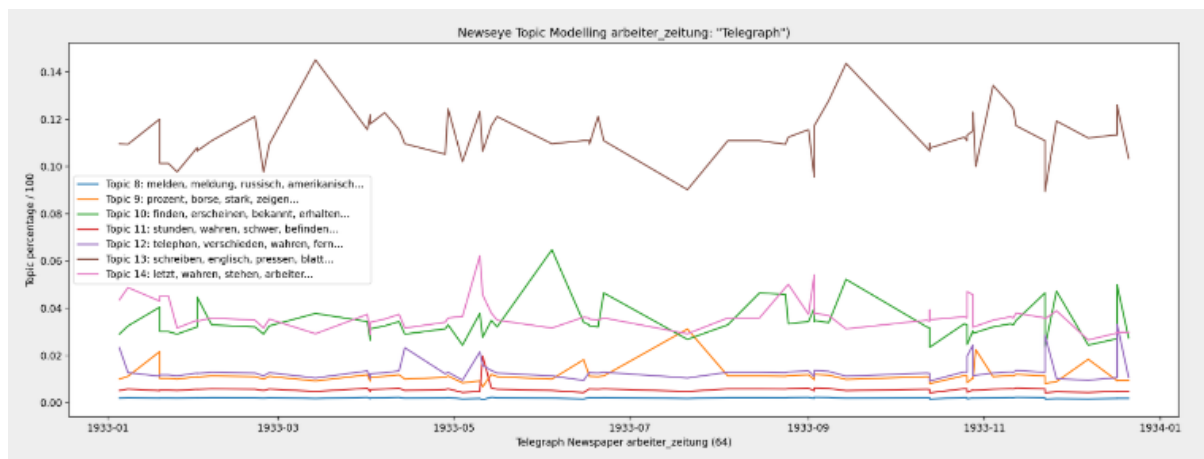
Figure 5: Visual output of the *Jupyter* Notebook 'Topic Modeling and Uses of the Term Telegpraph in the Context of Journalism'

2.3 Case study descriptions reflecting on digital methods

DH group members give insights into their research in the Case Studies[8] tab on the project website. The case study descriptions present their research and the impact of the newly developed tools and data from the NewsEye project on their research. The section consists of a short summary of the four case studies (Migration, Gender, Nationalism and Revolutions, as well as Media and Journalism) and links to more extensive reports about specific subtopics within each case study. All case studies describe one subtopic more closely, in order to show how insight is gained using digital methods in humanities' research. The descriptions and reports should mainly appeal to researchers, lay historians and other individuals who are interested in learning more about the case studies and tools from the NewsEye project.

While all information can be found on the website, the following list describes the chosen subtopics:

- Return migration between 1850 and 1950 [9]
  The topic on return migration was researched by DH team member Sarah Oberbichler (UIBK-ICH). The case study attempts to investigate return migration processes to Europe between 1850 and 1950. It does so by asking the following questions: How and in what context did Austrian daily newspapers report on returnees and how did this reporting change over time? Text mining methods were used to support the process of corpus building for further qualitative analysis.

- Women in Pants [10]
  The topic on women in pants was researched by DH team member Nejma Omari (UPVM). The case study aims to bring to light the evolution of the representation of women in the news from 1850 to 1950. The author was interested in the feminine condition and the right to wear pants: What place is given to them in the newspapers? What role do newspapers play in the struggles of women? The subtopic has a strong focus on named entities.

- 'The 'dark sides of war' (1914-1918)' [11]
  Revolutions and war were the main research interests of Barbara Klaus (UIBK-ICH). The central questions were: How are the 'dark sides of war' represented in news coverage during the war and what kind of propagandistic and patriotic purposes do these 'dark sides of war' serve in the newspapers? Frequency analyses on newspaper data support the hypotheses that many 'dark sides of war' fell prey to war-time censorship.

- Interaction between Newspapers and Readers [12]
  The interaction between newspapers and readers was investigated by DH team member Benedikt Kapferer (UIBK-ICH). His main interest was the question on how newspapers and the newsroom interacted with their readers and how the newsroom as the journalistic backstage of media production appeared on the front stage, i.e. in articles. Corpus building functions and frequency analysis supported the outcome of the analysed dataset.

2.4 Blog posts showcasing tools and project output

The project blog[9], which is managed and edited by the communications expert from the ONB, is open to entries from all project members and in all languages. The DH team has published a total of 19

---

[8] https://www.newseye.eu/case-studies/
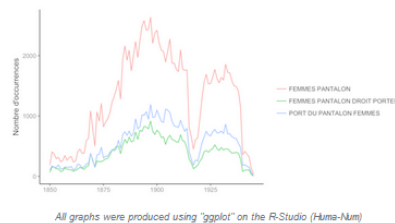[9] https://www.newseye.eu/blog/

Figure 6: Case study description 'Woman in Pants' from Nejma Omari (UPVM)

blog posts in English, French, Finnish and German (a full list of all blog posts can be found in the report on Deliverable D6.11: Use of project results for the general public). Eight of these are discussed here specifically because they show the use of tools and methods based on the DH case studies. Consequently, these blog posts use a more technical language and sometimes discuss specific tools and methods.

- Bringing together what belongs together: Thematic grouping of newspaper clippings using LDA and JSD [13]
  In this blog post by Sarah Oberbicher and Eva Pfanzelter (UIBK-ICH), the *Jupyter* Notebook on corpus building using LDA and the Jensen-Shannon Distance (JSD) method is explained and showcased.

- Newspapers as 'social' media: Crowdsourcing and user-generated content in historical perspective [14]
  In this blog post, Benedikt Kapferer (UIBK-ICH) examines different possibilities of interaction and two-way communications between readers and newspapers by finding specific newspaper content, analysing word frequencies, and close reading.

- „An unsere Leser": The interaction between newspaper and readership and the journalistic self-image [15]
  This blog post, written by Benedikt Kapferer (UIBK-ICH), deals with the question of how newspapers interacted with their readership and what this reveals about the development and the practices of journalism. The blog post also shows how phrase searches and frequency analysis

support such an analysis.

- **Als eine andere Epidemie die Welt in Atem hielt: Die Spanischen Grippe 1918/19 in der öster-reichischen Presse** [16]
 This blog post, written by Sarah Oberbichler, Stefan Hechl, and Eva Pfanzelter (UIBK-ICH), takes a closer look at the newspapers' reporting on the Spanish Flu 1918/19 and draws comparisons to the Covid-19 pandemic. The post gives insights into search processes and the importance of relative frequency analysis, and uses word clouds to detect changes in the discourse on the Spanish Flu from 1918.

- **What's the frequency, Kenneth?** [17]
  Jani Marjanen (UH-DH) explains the advantages but also difficulties of diachronic frequency analysis. For corpus linguistics, the analysis of word frequencies is much about getting to the heart of language use and hence having a balanced corpus is central. This blog post shows the importance of the interpretative component when working with frequencies (see Figure 7).

National as an example of a discourse spreading to new contexts

If we turn to the word national, the trend line shows a story of growth in frequency during the nineteenth century, but the change is more gradual and seem to be less connected to particular debates.
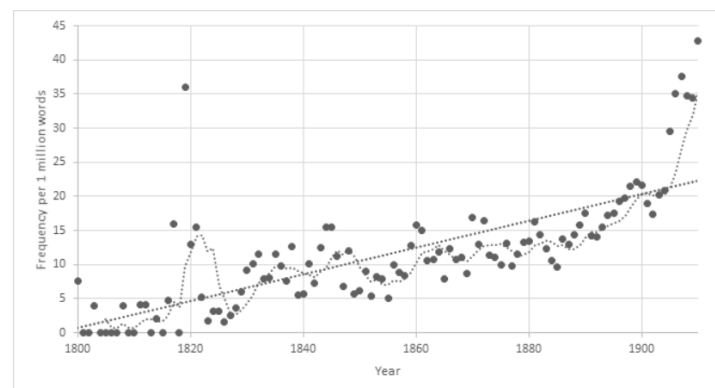
Relative frequency of the lemma nationell in Swedish-language newspapers and periodicals in Finland, 1800–1910.

Figure 7: Excerpt of the blog post 'What's the frequency, Kenneth?' by Jani Marjanen (UH-DH)

- **Identifying the 'right' widow in WW1 newspapers** [18]
  The blog post by Barbara Klaus (UIBK-ICH) takes a closer look at the coverage of war widows in the First World War. She presents results of her research and explains the central difficulties and challenges regarding the research of this topic using ANNO.

- **The fire of Wöllersdorf (1918) in Austrian Newspapers** [19]
  In this blog post, Barbara Klaus (UIBK-ICH) takes a look at the Austrian press coverage of one of the most important incidents on Austrian territory during the First World War (1914–1918). In doing so, she analyses word frequencies and investigates 'white spots' as result of censorship in historical newspapers.

- British drug advertising in the 19th century through the prism of gender [20]
  This blog post was written by Sarah Oberichler (UIBK-ICH), Khanim Garayeva, Ruben Ros, and Kalle Kusk Gjetting in the context of the Digital Humanities Hackathon (DHH) 2019 in Helsinki. It analyses British drug advertisements using word embeddings.

2.5 Twitter threads showcasing tools and project output

A thread on Twitter is a series of tweets linked to one another. Using this storytelling function, one can provide additional context (such as videos or images), an update, or an extended point by connecting multiple tweets together. That is why Twitter threads are ideal for NewsEye showcases in order to reach out to other scholars, lay historians, teachers, and other users that may be interested in the projects output.

Figure 8: Twitter thread from Stefan Hechl (UIBK-ICH) on the Spanish flu

For the DH group from the NewsEye project, the focus was on the individual case studies or currently important topics. The goal was to show how newspapers can be used to find information about a specific event or topic and what has to be considered when working with historical newspapers. The tweets include graphs, pictures and screenshots of article clippings (since visuals have a better impact). The same applies to tweets with hashtags (#). Hashtags allow users to label the content in order to quickly find other content on that same topic. Unlike previously explained showcases, our Twitter threads have a less academic style and were used to raise the interest on historical newspapers and interfaces. In total, 9 Twitter threads were created by DH members, and five of these (which are listed below) were created within the framework of this task by showing tools and project outputs (a full list of all Twitter threads can be found in Deliverable D6.11: Use of project results for the general public).

The following Twitter threads were published by DH researchers and shared by the NewsEye project account:

- [CORONAVIRUS : CE QUE NOUS APPREND LA PRESSE D'IL Y A 100 ANS] [21]
  In this French-language thread, Nejma Omari (UPVM) analysed the news coverage on Spanish Flu pandemic 1918–1919 by comparing it to the Covid-19 pandemic.

- Wie wurde vor knapp 100 Jahren über die Spanische Grippe berichtet – und wie findet man das heraus? Ein Recherche mit Beispielen [22]
  In this German-language thread (Figure 8), Stefan Hechl (UIBK-ICH) analysed newspaper coverage of the Spanish Flu pandemic 1918–1919 by using qualitative and quantitative methods.

- Todesstrafe für „Hamsterei" und eine mögliche Antwort, warum wir Klopapier horten. Eine historische Zeitreise durch die Geschichte des Hamsterns mit Hilfe digitaler österreichischer Zeitungen http://anno.onb.ac.at [23]
  In this German-language thread, Sarah Oberbichler (UIBK-ICH) analysed the news coverage on 'hamster buyers' using relative frequencies, search strategies and qualitative reading.

- Our article 'The expansion of isms, 1820–1917' is published. It links to previous work on isms/ideology, but in this project we got the opportunity to do something more computationally oriented [24]
  Jani Marjanen (UH-DH) discusses an article on isms and the use of word embeddings as a method for digital history (see Figure 9). The English-language thread discusses specific methods, as well as what kind of historical reasoning is possible based on quantitative analysis of text.
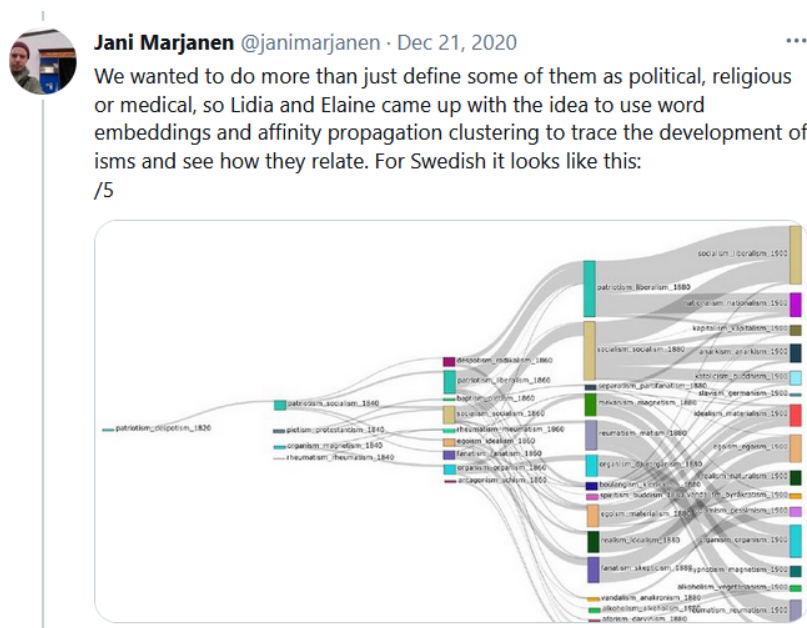

Figure 9: Twitter thread by Jani Marjanen (UH-DH) on the 'Expansion of isms, 1820-1917'

- Heute vor 75 Jahren erschien mit der „Österreichischen Zeitung" in Wien die erste Zeitung des befreiten Österreich, herausgegeben von der Roten Armee. Sie erschien bis 1955, die Redaktion war mit sowjetischen und österreichischen Journalist*innen besetzt [23]
  In this German-language thread, Stefan Hechl (UIBK-ICH) guides readers through the re-establishment of democratic newspapers in Austria after 1945 in a series of 'On this day in 1945...' tweets. He gives a general overview and uses co-occurrence and diachronic frequency analysis.

3 Dissemination

Several steps were taken in order to reach the aimed audiences (researchers as well as the general public) with the created and published showcases:

1. The content was shared on social media via Twitter and it was advertised on the homepage of the NewsEye website and included in the NewsEye Demonstrator.
2. Some of our showcases have been added as references in Wikipedia or in other publications (see public Deliverable D6.11: Use of project results for the general public).
3. Our multimedia paper in the *EuropeanaTech* issue on historical newspapers [25], for example, presents the screencast on how to create and manage datasets with the NewsEye Demonstrator.
4. Parts of almost all of the presented material were included in the educational material for school pupils, teachers, university students and interested members of the general public.
5. Some of the podcasts, such as those by Sarah Oberbichler (UIBK-ICH) on her case study on migration, by Stefan Hechl (UIBK-ICH) on the development of the Austrian 'nation', and by Benedikt Kapferer together with Stefan Hechl (both UIBK-ICH) on their case studies on nationalism and journalism as inspiration for teaching, reflect on the showcases in more detail.

With these initiatives, the DH team members tried to give an open insight into digital humanities practice throughout the project. Re-publication of blog-posts in paper journal editions, re-use of notebooks, inclusion of screencasts and so forth should be a valuable contribution when reaching out to a general public.

References

[1] *NewsEye Demonstrator in action: creating and managing datasets*. In collab. with Sarah Oberbichler, Axel Jean-Caurant and Tonica Hunter. 10th Jan. 2020. URL: https://www.youtube.com/watch?v=N85F1JXdfS4&t=189s (visited on 30/03/2021).

[2] *Using Named Entities in the NewsEye Platform*. In collab. with Nejma Omari, Axel Jean-Caurant and Amanda Maunoury. 25th Feb. 2021. URL: https://www.youtube.com/watch?v=TYsvGmCfXQg (visited on 30/03/2021).

[3] *A Digital History Workflow - Emigration and economic benefits/advertising between 1850 and 1950*. In collab. with Sarah Oberbichler. 13th Apr. 2021. URL: https://www.youtube.com/watch?v=XkQvq196KFg (visited on 02/04/2021).

[4] Sarah Oberbichler. *Text classification for topic-specific newspaper collections*. Juypter Notebook, 2020. URL: https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections/blob/master/Text_classification_of_newspaper_clippings_notebook.ipynb (visited on 02/04/2021).

[5] Sarah Oberbichler. *Group similar newspaper articles*. Jupyter Notebook, 2020. URL: https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections/blob/master/news_article_similarity_notebook.ipynb (visited on 02/04/2021).

[6] Sarah Oberbichler. *Discover a newspaper collection with diachronic Ngram clouds*. Jupyter Notebook, 2020. URL: https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections/blob/master/n-gram_clouds_notebook.ipynb (visited on 02/04/2021).

[7] Sarah Oberbichler. *Discover media discourse in Spanish Flu newspaper coverage with diachronic bi-gram clouds and topic modeling*. MyBinder, 2020. URL: https://mybinder.org/v2/gh/soberbichler/Discourse-in-Spanish-flu-coverage_Notebook/HEAD (visited on 02/04/2021).

[8]   János Békési and Martin Gasteiner. *Topic Modeling and Uses of the Term Telegraph in the Context of Journalism*. Jupyter Notebook, 2020. URL: https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections#Topic-Modeling-and-Uses-of-the-Term-Telegraph-in-the-Context-of-Journalism (visited on 02/04/2021).

[9]   Sarah Oberbichler. *Return migration between 1850 and 1950*. en. 2018. URL: https://www.newseye.eu/case-studies/case-study-1-migration/return-migration-between-1850-and-1950/ (visited on 16/03/2020).

[10]   Nejma Omari. *Women in Pants*. en. 2019. URL: https://www.newseye.eu/case-studies/case-study-2-gender/women-in-pants/ (visited on 16/03/2020).

[11]   Barbara Klaus. *The 'dark sides of war' (1914-1918)*. en. 2019. URL: https://www.newseye.eu/case-studies/case-study-3-nationalism-and-revolutions/the-dark-sides-of-war-1914-1918/ (visited on 16/03/2020).

[12]   Benedikt Kapferer. *Interaction between Newspapers and Readers*. 2020. URL: https://www.newseye.eu/case-studies/case-study-4-media-and-journalism/interaction-between-newspapers-and-readers/ (visited on 16/03/2021).

[13]   Sarah Oberbichler and Eva Pfanzelter. *Bringing together what belongs together: Thematic grouping of newspaper clippings using LDA and JSD*. NewsEye Blog. 29th Oct. 2020. URL: https://www.newseye.eu/blog/news/bringing-together-what-belongs-together/ (visited on 31/03/2021).

[14]   Benedikt Kapferer. *Newspapers as „social" media: Crowdsourcing and user-generated content in historical perspective*. NewsEye Blog. 9th Oct. 2020. URL: https://www.newseye.eu/blog/news/newspapers-as-social-media/ (visited on 31/03/2021).

[15]   Benedikt Kapferer. *„An unsere Leser": The interaction between newspaper and readership and the journalistic self-image*. NewsEye Blog. 24th Aug. 2020. URL: https://www.newseye.eu/blog/news/an-unsere-leser/ (visited on 31/03/2021).

[16]   Sarah Oberbichler, Eva Pfanzelter and Stefan Hechl. *Als eine andere Epidemie die Welt in Atem hielt: Die Spanischen Grippe 1918/19 in der österreichischen Presse*. NewsEye Blog. 6th Sept. 2020. URL: https://www.newseye.eu/blog/news/epidemie-spanische-grippe/ (visited on 31/03/2021).

[17]   Jani Marjanen. *What's the frequency, Kenneth?* NewsEye Blog. 25th June 2019. URL: https://www.newseye.eu/blog/news/what-s-the-frequency-kenneth/ (visited on 31/03/2021).

[18]   Barbara Klaus. *Identifying the 'right' widow in WW1 newspapers*. NewsEye Blog. 14th Apr. 2020. URL: https://www.newseye.eu/blog/news/identifying-the-right-widow-in-ww1-newspapers/ (visited on 31/03/2021).

[19]   Barbara Klaus. *The fire of Wöllersdorf (1918) in Austrian Newspapers*. NewsEye Blog. 29th Aug. 2019. URL: https://www.newseye.eu/blog/news/the-fire-catastrophe-in-woellersdorf-1918-in-austrian-newspapers/ (visited on 31/03/2021).

[20]   Sarah Oberbichler, Khanim Garayeva, Ruben Ros and Kalle Kusk Gjetting. *British drug advertising in the 19th century through the prism of gender*. NewsEye Blog. 6th June 2019. URL: https://www.newseye.eu/blog/news/british-drug-advertising-in-the-19th-century-through-the-prism-of-gender/ (visited on 31/03/2021).

[21]   Nejma Omari. *#CORONAVIRUS : CE QUE NOUS APPREND LA PRESSE D'IL Y A 100 ANS.* de. 2020. URL: https://mobile.twitter.com/nejmaomari/status/1238542525239091205 (visited on 16/03/2020).

[22] Stefan Hechl. *Wie wurde vor knapp 100 Jahren über die Spanische Grippe berichtet – und wie findet man das heraus? Ein Recherche mit Beispielen*. de. 2020. URL: https://mobile.twitter.com/stefanhechl/status/1238773566650359808 (visited on 16/03/2020).

[23] Sarah Oberbichler. *Todesstrafe für „Hamsterei" und eine mögliche Antwort, warum wir Klopapier horten. Eine historische Zeitreise durch die Geschichte des Hamsterns mit Hilfe digitaler öster-reichischer Zeitungen https://t.co/GEe02cKbhe https://t.co/g5q40gpcRz*. Twitter. 15th Mar. 2020. URL: https://twitter.com/SOberbichler/status/1239065898838888449 (visited on 31/03/2021).

[24] Jani Marjanen. *Our article "The expansion of isms, 1820–1917" is published. It links to previous work on isms/ideology, but in this project we got the opportunity to do something more computationally oriented.* Twitter. 21st Dec. 2020. URL: https://twitter.com/janimarjanen/status/1340924519083634689 (visited on 08/04/2021).

[25] Sarah Oberbichler, Eva Pfanzelter, Stefan Hechl and J. Marjanen. 'Doing historical research with digital newspapers - perspectives of DH scholars'. In: *EuropeanaTech Insight* 16: Newspapers (2020). URL: https://pro.europeana.eu/page/issue-11-generous-interfaces (visited on 25/11/2020).