On a digital journey into yesterday's future: Zeit.shift – preserving Tyrol's cultural text heritage

Johanna Walcher[c]**, Andrea Abel**[a]**, Johannes Andresen**[b]**, Paolo Brasolin**[a]**, Isabella Dissertori**[b]**, Eva Eberwein**[c]**, Greta Franzini**[a]**, Silvia Gstrein**[c]**, Maritta Horwath**[c]**, Christian Koessler**[c]**, Barbara Laner**[c]**, Verena Lyding**[a]**, Karin Pircher**[b]**, Egon W. Stemle**[a]

[a]*Eurac Research,*
Drususallee 1, 39100 Bozen, Italy

[b]*Landesbibliothek Dr. Friedrich Teßmann,*
A.-Diaz-Straße 8, 39100 Bozen, Italy

[c]*Universitäts- und Landesbibliothek Tirol,*
Innrain 50, 6020 Innsbruck, Austria

E-mail: johanna.walcher@uibk.ac.at, andrea.abel@eurac.edu, johannes.andresen@tessmann.it, paolo.brasolin@eurac.edu, isabella.dissertori@tessmann.it, eva.eberwein@uibk.ac.at, greta.franzini@eurac.edu, silvia.gstrein@uibk.ac.at, maritta.horwath@uibk.ac.at, christian.koessler@uibk.ac.at, barbara.laner@uibk.ac.at, verena.lyding@eurac.edu, karin.pircher@tessmann.it, egon.stemle@eurac.edu

Abstract

This paper outlines the participatory approach in Zeit.shift and provides insight into strategies and challenges of the exploratory approach to public involvement in this project.

*Speaker: Johanna Walcher

1.   Introduction

Zeit.shift is a cross-border *Interreg*-project in the field of Digital Humanities of the *Landesbibliothek Dr. Friedrich Teßmann*, the *Universitäts- und Landesbibliothek Tirol* and *Eurac Research* (October 2020–June 2023). The project pursues the goal of indexing historical newspapers from North, South and East Tyrol with the help of citizen scientists. The aim is the establishment of a long-term, cross-border institutional cooperation for the preservation, development and communication of the cultural text heritage of Tyrol and South Tyrol. Initially, the written cultural assets were scattered throughout all parts of the region and were not accessible in their entirety in one institution. Also, the documents were only partially digitized and their accessibility as well as awareness among the population were insufficient. Through a cross-border digitization and dissemination strategy the text collections will be virtually brought together for the first time in a single web platform and they will be processed in a user-oriented manner with the help of computer-linguistic approaches. German-language historical daily newspapers are used as case example, which as sources are also threatened by decay.

One focus of the project is the active involvement of interested citizens. They are encouraged to actively participate and invited to contribute to text indexing through online annotation of the text material. This offers them the opportunity to learn more about their own cultural heritage and discover historical texts as sources. A common web platform, whose launch is imminent, is the basis for collaboration even beyond the project duration. Thus, the focus of the project is twofold: on the one hand, digitize historical Tyrolean newspapers (1850–1950), improve their searchability and make them accessible to systematic research and, on the other hand, to test the potential of a participatory approach to indexing and crowd-correction of automated transcriptions – being in line with the motto of the Austrian Citizen Science Conference 2022 "Citizen Science – Warum (eigentlich nicht)?" ("Citizen Science – why not?").

2.   Preserve – Develop – Communicate

The goals in Zeit.shift will be achieved through the three milestones: "Preserve – Develop – Communicate". In the context of preservation, approximately 500,000 pages of German-language local and regional newspapers were digitized, the full-text was recognized by OCR (Optical Character Recognition using ABBYY FineReader Engine v. 11) and methods of computational linguistics (Named Entity Recognition, topic labelling) are used to enhance the text content and its searchability. In addition, citizen scientists are invited to content-based tagging and geolocation of newspaper snippets. More specifically, advertisements were selected as a use case and semi-automatically extracted from the newspapers. After manually sorting out non-relevant information, the clippings were published for tagging on the external platform *Historypin* (https://www.historypin.org/en/zeit-shift/).

The second milestone aims at the development and implementation of a citizen science initiative for the valorization and mediation of regional historical text documents. Within this framework, workshops have been organized for various stakeholders such as libraries, archives, museums and chroniclers. These institutions act as multipliers to

reach as many people as possible. The cooperation of the volunteers does not pursue data generation as the main goal; the main importance is within milestone three to communicate cultural text heritage, introduce people to historical text sources and offer a discovery trip through historical newspapers.

The central project aim of Zeit.shift is to make the digitized newspaper and associated metadata publicly accessible via a web platform. The content generated by the volunteers will be added to the platform and helps to improve the searchability of texts. Thereby, the data and effort are given back to the public and the resulting benefits can be shared by all. Until launch, the resources can be accessed via the online portals of the *Universitäts- und Landesbibliothek Tirol* (https://diglib.uibk.ac.at/obvuibz) and of the *Landesbibliothek Dr. Friedrich Teßmann* (https://digital.tessmann.it/).

3.   Participatory approach in Zeit.shift

Via the platform *Historypin* (hosted externally), citizens can participate in the text indexing by describing and geolocating advertisements from historical daily newspapers. The collection is made up of newspaper advertisements from the newspaper holdings digitized in the project. The indexing of this special type of newspaper text provides a glimpse into the economic situation of Tyrol in past times. Newspaper advertisements contain valuable information about everyday life and reflect the interests of the citizens of that time. Moreover, volunteers have the opportunity to compare the mentioned places of the past with today's. They actively participate in the text indexing and contribute to the better usability and searchability of the historical text material. Likewise, participatory activities in general are intended to stimulate and promote interest in the cultural textual heritage of the region. In order to cover as many interest groups as possible, advertisements from newspapers of different cities were chosen. The primary target group for this activity is adults interested in history, such as library users and chroniclers. Outputs for the project include names for improving Named Entity Recognition and Named Entity Search functionality. So far 1668 unique semantic tags and 440 comments, including geotags, have been added by volunteers.
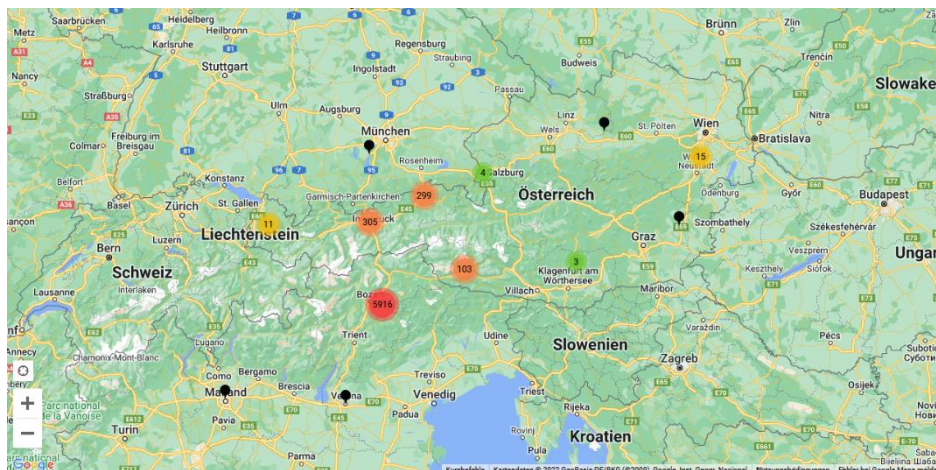

Figure 1: Geolocalization of newspaper ads (Kartendaten © 2022 GeoBasis-DE/BKG (©2009), Google; via https://www.historypin.org/en/zeit-shift/)

The second application and connected activity (used separately from the annotation-task via Historypin) involves an online game, called *Ötzit!,* developed by *Eurac Research*. In this game, Ötzi is rescued from dangerous animals by correctly typing words in Fraktur-script. The game is primarily aimed at German-speaking students (approximate age: eleven to fourteen), but is generally open to anyone interested. The aim of the game is to increase awareness of the digitized newspaper collections and to practice reading historical documents in Fraktur script. The transcriptions generated by the players will be analyzed and used to correct the OCR and to improve the searchability of the digitized collections. To date 2268 games have been played with 1662 unique devices used and thereby about 6700 words have been transcribed.
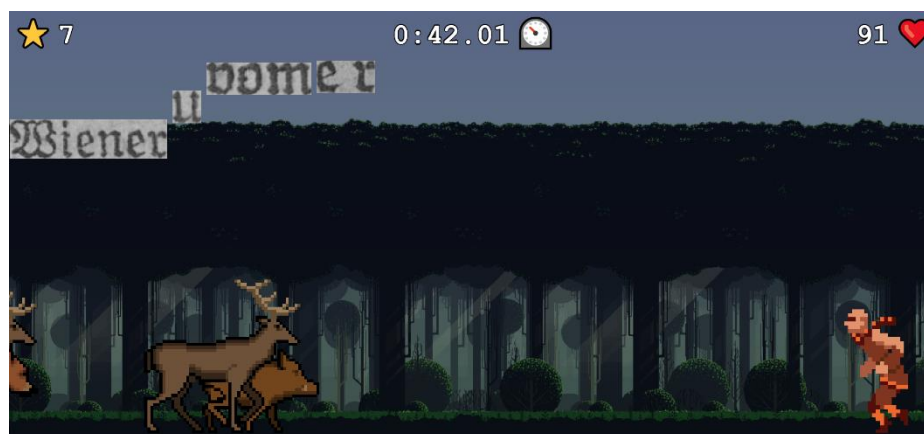

Figure 2: Ötzit!-game (CC BY 4.0 Paolo Brasolin c/o IAL @ Eurac Research)

Next to its many enrichments, the participatory approach can also be challenging: Zeit.shift attempts to go beyond mere OCR correction as a participatory activity. Here, for example, data quality poses a challenge because OCR quality varies widely. To allow for a low-threshold activity, there were no strict rules for the tagging of the newspaper advertisements, resulting in a variance of tags that need to be manually normalized by the Zeit.shift-staff, which requires an additional effort. The Covid-19 pandemic made direct interaction with volunteers more difficult and in addition, the lack of a dedicated community management position in the project, which would have allowed for more intensive supervision of the participatory approach in the project, became noticeable.

The aim of the participatory approach in Zeit.shift is first and foremost to create awareness of cultural text heritage. By providing opportunities to discover interesting information in the newspaper texts about everyday life in Tyrol 100 years ago and about any other relevant topic of interest, the project conveys the value of historical text sources. This puts the focus of the project in the context of "Cultural Heritage", an area that has enjoyed increasing popularity in recent years [3].

4. Conclusion

Research benefits from (meta-)data enrichment and the possibility of systematic searchability of the historical press. In total, about 13 terabytes of digitized data have been generated and made accessible through OCR text recognition. Both citizen science tasks primarily aim at engagement and yet are designed to test their effectiveness for research-oriented tasks, especially for topic classification and geolocation (*Historypin*) as well as for OCR post-correction (*Ötzit!*). The contribution to the public is provided by free access to a large historical text collection and easy contact with the source material via citizen science applications.

Together with the citizen scientists, hidden archival treasures are brought to light and local knowledge is improved. The participation of citizen scientists allows to work on a larger number of texts and at the same time communicating their content to them. In the process, the volunteers can gain expert knowledge and skills as well as new insights. In this way, they make an invaluable contribution to the awareness of research as an integral part of a functioning society, which is of pressing importance, especially in response to the skepticism about science [2].

Acknowledgements

References

[1] Franzini, G., Stemle, E.W., Lyding, V., Abel, A., Andresen, J., Pircher, K., Gstrein, S., Laner, B., Walcher, J., Horwath, M., Kössler, C. (2022): Citizen Humanities in Tyrol: A case study on historical newspapers, in: Ciracì, F., Miglietta, G., Gatto, C. (eds.) *Proceedings of the Eleventh AIUCD Annual Conference (AIUCD 2022)*. Lecce, 1–3 June. ISBN: 978-88-9425-356-6. DOI: http://doi.org/10.6092/unibo/amsacta/6848

[2] Göbel, C, Henke, J., Mauermeister, S. (2020): Kultur und Gesellschaft gemeinsam erforschen Überblick und Handlungsoptionen zu Citizen Science in den Geistes- und Sozialwissenschaften. Institut für Hochschulforschung (HoF) an der Martin-Luther-Universität: Halle-Wittenberg 2020. ISBN: 978-3-937573-79-3

[3] Hetland, P., Pierroux, P., Esborg, L. (eds.) (2020): A History of Participation in Museums and Archives. London: Routledge. DOI: https://doi.org/10.4324/9780429197536

PoS(ACSC2022)019