Extending the Discourse Analysis Tool Suite with Whiteboards for Visual Qualitative Analysis

Tim Fischer[†], **Florian Schneider**[†], **Fynn Petersen-Frey**[†],
Anja Silvia Mollah Haque[‡], **Isabel Eiser**[‡], **Gertraud Koch**[‡], **Chris Biemann**[†]
[†] Language Technology Group, Department of Informatics, Universität Hamburg, Germany
[‡] Institute of Anthropological Studies in Culture and History, Universität Hamburg, Germany
{firstname.lastname OR florian.schneider-1}@uni-hamburg.de

Abstract

In this system demonstration paper, we describe the *Whiteboards* extension for an existing web-based platform for digital qualitative discourse analysis. *Whiteboards* comprise interactive graph-based interfaces to organize and manipulate objects, which can be qualitative research data, such as documents, images, etc., and analyses of these research data, such as annotations, tags, and code structures. The proposed extension offers a customizable view of the material and a wide range of actions that enable new ways of interacting and working with such resources. We show that the visualizations facilitate various use cases of qualitative data analysis, including reflection of the research process through sampling maps, creation of actor networks, and refining code taxonomies.

Keywords: Qualitative Data Analysis, Visualization, Discourse Analysis, Digital Humanities

1. Introduction

The Discourse Analysis Tool Suite (Schneider et al., 2023a) is an open-source working environment for digital qualitative discourse analysis in the Digital Humanities (DH). The tool is developed in close co-creation with and for researchers in the DH to make state-of-the-art machine learning technologies from Computer Vision and Natural Language Processing available to non-expert users, enabling them to manage and analyze unstructured, multi-modal data.

The overall platform is originally designed and tailored to support Grounded Theory (Strauss et al., 1996) based research such as the Sociology of Knowledge Approach to Discourse (Keller, 2011). However, many functionalities, such as the automatic pre-processing of multi-modal data (text, image, audio, and video), data exploration, as well as manual and automatic annotation, are also of great interest to other disciplines.

In this paper, we extend our existing platform with *Whiteboards* to facilitate and enrich the methodological procedures of Grounded Theory in qualitative discourse analysis with maps as known from Situational Analysis (Clarke et al., 2017). Such maps are typically utilized to organize and obtain knowledge. The proposed *Whiteboards* supports a wide range of usage scenarios, including but not limited to theoretical sampling maps, discourse actor networks, and code structure maps. The implementation is kept generic, enabling new ways of interaction with material, and paving the way for novel discourse analysis practices. Specifically, the extension offers features such as creating text, post-its, various shapes, and arrows. More importantly, it allows to link all kinds of user- and system-generated data, such as annotations, codes, documents, memos, and tags, to the boards. This intertwining of typical mind-map functionalities with objects of qualitative data analysis creates a unique user experience to view, organize, and work with the materials.

At their core, *Whiteboards* are graphs of nodes and edges where nodes represent any object (e.g., shapes, post-its, documents, or annotations) while edges encode their (custom) relationships. Thus, the proposed extension can also be seen as an interactive tool to manipulate annotated, multi-modal resources in a novel, graph-based way.

The contributions of this paper are twofold: First, we develop and describe the extension of our platform with *Whiteboards*, which allow for a customizable view of and interaction with user- and automatically-generated resources. Second, we demonstrate real-world usage scenarios of refining hierarchical code structures, creating actor networks, and reflecting with the help of sampling maps to highlight the effectiveness and usefulness of the proposed extension.

2. Related Work

Related works include software for annotation, qualitative data analysis, or the creation of mind maps.

CodeAnno (Schneider et al., 2023b), INCEpTION (Klie et al., 2018), CATMA (Gius et al., 2023), and DWTS (Koch et al., 2022; Schneider et al., 2023a) are open-source platforms for annotating text resources and offer, to a varying degree, features for qualitative data analysis. However, none of these tools offer interactive visual analysis of the material that resembles mind maps.

New/s/leak (Wiedemann et al., 2018) and Jigsaw

(Görg et al., 2014) are visual analytic systems for investigative analysts that visualize documents and their automatically extracted entities, such as people, places, and organizations. While New/s/leak offers a generated graph view of the entity network, Jigsaw offers "Tablets" to manually create such networks. *Whiteboards* cover all of Tablets' functionality and extend it to other materials like documents, tags, etc., in addition to entities.

MAXQDA is a commercial but frequently used tool for mixed-methods research in DH. It offers the *Creative Coding*[1] feature to manage code systems on a 2D canvas. While offering a creative interaction, this feature is limited to category systems. In contrast, our proposed extension enables comprehensive, custom-generated visualization and working with any material, including codes, annotations, tags, or documents.

Miro[2] is a visual workspace that started as a tool for the collaborative creation of mind maps but since then integrated tools and features to support various use cases of companies, including project planning, data-visualization, diagramming, agile processes, and much more. Miro was analyzed and evaluated as part of a manual discourse analysis in a co-creation approach (Eiser et al., 2023) to software development. It was found to be helpful and motivated the full integration of *Whiteboards*, which overcomes many disadvantages of working with third-party tools, such as cumbersome copy-and-pasting of data, losing the connection to the source material or manually applying changes in all used tools.

The resulting extension significantly differs from conventional whiteboard or mind-map tools. While providing common features such as post-its, texts, various shapes, and arrows, it further facilitates direct interaction with and manipulation of research objects, as elaborated in the following section.

3. The Whiteboards Extension

Whiteboards are an extension of our platform for interacting with business objects in a novel way on a 2D canvas. Such objects are the outcome of typical workflows within our platform and include, for example, uploaded documents, tags, codes, annotations, and memos created by users or automatically annotated named entities from the pre-processing pipeline, as provided by the platform already.

Users can create and work with multiple *Whiteboards*. In technical terms, *Whiteboards* are graphs of edges and nodes. Nodes are representations of automatically and user-created objects. There exist two conceptually different nodes: Database nodes include documents, codes, tags, memos, annotated text passages, etc. They are directly linked to the platform's database model. Whiteboard nodes include texts, post-its, and various shapes and only exist in a particular whiteboard. Similarly, there exist two different types of edges: Database edges are directly inferred from the database. For example, a database edge will be rendered automatically between document and tag nodes if the documents were assigned that tag. Further, database edges exist between annotation and code nodes if those passages were annotated with that code. While database edges cannot be customized, whiteboard edges represent customizable arrows that only exist within a particular whiteboard. As database nodes and edges are linked to the respective objects and their relations, modifications are applied in both directions: Changes to database nodes or edges in *Whiteboards* are reflected in all other parts of the application and vice versa. For example, renaming codes in *Whiteboards* applies these changes globally while connecting tags and documents with database edges assign these tags to the documents.

Its primary use case is the creation of maps and organization of knowledge. Hence, *Whiteboards* are designed to be fully interactive and allow various actions (see Figure 4): Navigation is done by panning and zooming. Database and whiteboard nodes can be added using the sidebar (a). While database edges (b) are added automatically, whiteboard edges (c) can only be created by users. Additional database edges are created by dragging&dropping from the grey border of database nodes to valid other database nodes (d). In contrast, whiteboard edges can be created by connecting two blue dots (e). Double-clicking database nodes allows the modification of the corresponding database entries, whereas whiteboard nodes and edges can be customized using the toolbar (f). It offers various options, such as changing font size, color, formatting, positioning (for whiteboard nodes) or arrowheads, thickness, and style (for whiteboard edges). The context menu offers additional actions depending on the node (g). Most importantly, the menu allows the expansion of related database nodes, adding them to the board with their respective database edges. For example, a code node allows the creation or expansion of the parent code, all the child codes, the attached memo, or all annotated passages, while a document node allows the expansion of all document tags, all annotations, the attached memo, or linked documents. This expansion feature allows an exploratory way to traverse the database, finding connections in the data while interactively unfolding a network of connected nodes.

The *Whiteboards* extension is unique by in-

[1] https://www.maxqda.com/help-mx22/06-creative-coding/idea-behind-creative-coding

[2] https://miro.com/

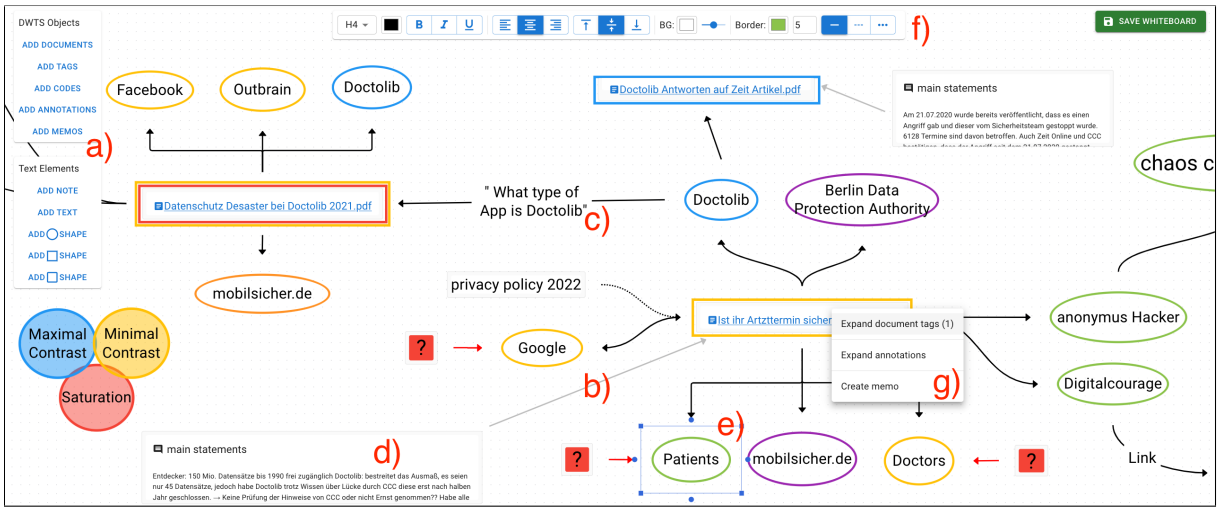

Figure 1: A sampling map created with the *Whiteboards* extension. Features are described in Section 3.

tertwining typical mind-map functionalities with database objects that result from qualitative data analysis. This way, it allows for a unique user experience. We want to highlight that every database node allows users to jump to the respective content in the application: For example, clicking a node representing a document will open the corresponding document in the Document Viewer. Similarly, clicking a node representing an annotation will open the document and jump to that exact passage. This interconnectivity is where this extension shines and is a significant motivation to fully integrate *Whiteboards* within our existing platform instead of relying on third-party tools.

From a DH perspective with a focus on discourse analysis, working with all relevant material on a board while adding custom post-its and arrows enables creative interactions. It offers a vastly different view of the material besides the existing presentations as lists, tables, or plain documents. Organizing, refining, and managing data visually is a meaningful way to understand phenomena better, identify relations, track research processes, and obtain knowledge about discourses. The use of other visualization tools, such as Miro, would be feasible in terms of technical functions but would require a change of working environment and leaving the analysis tool. In this context, the alternating translation work of the programs, which is a double effort of digital handcraft (Franken, 2020) for researchers, is no longer necessary. Further methodological advantages of using an integrated solution are described in the next section.

The *Whiteboards* extension is implemented using the libraries *React Flow*[3] and *MUI*[4].

[3]https://reactflow.dev/
[4]https://mui.com/

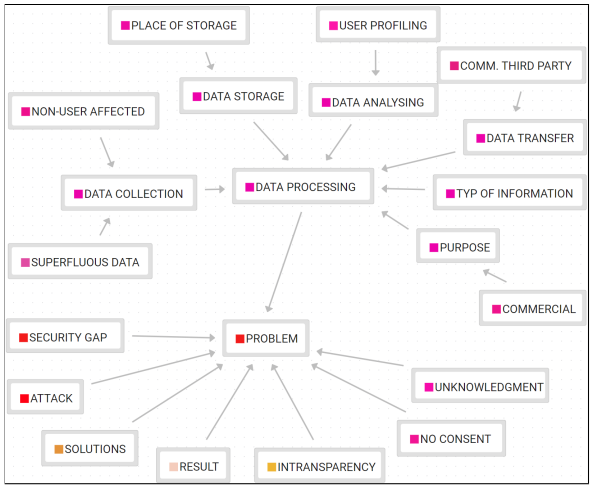

Figure 2: Result of using *Whiteboards* to refine a complex category system on data protection.

4. Exemplary Use Cases

This section discusses empirical usage scenarios of the *Whiteboards* extension. All described use cases are part of a project on a discourse analysis of the debate on E-Health in Germany. We contrast our proposed extension with Miro, which was used previously in the aforementioned project.

4.1. Code System Management

The platform already allows the creation of hierarchical code structures. It provides means to refine, merge, or delete codes, which is integral to approaches based on Grounded Theory. Coding is done on the fly while annotating a document (open coding) or in the Code Explorer, which visualizes the taxonomy as a tree view (axial and selective coding).

The proposed *Whiteboards* allow a different,

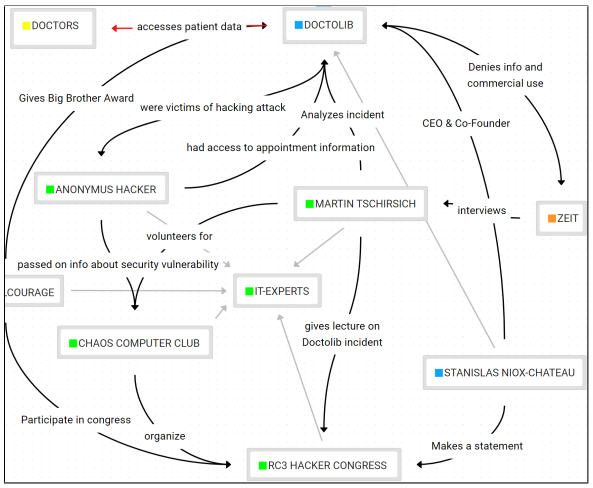

Figure 3: An excerpt of a manually created actor-network built with the *Whiteboards* extension.

graph-based view of the code taxonomy as shown in Figure 2. Consequently, the interaction also differs: The visualization of the codes can be rearranged to the user's preference by dragging & dropping nodes. Creating and changing a code hierarchy by adding or modifying parent-child relations is as easy as drawing a line to add an edge. By selecting multiple nodes followed by a quick context menu action, the corresponding codes can be merged, or a new top-level code can be introduced. This interaction makes *Whiteboards* a powerful tool for organizing and refining code structure. It allows visual and creative interpretable approaches to coding, from mere structural quantitative analysis to hermeneutic approaches to qualitative analyses.

After an extensive phase of open coding, the resulting category system of codes related to data protection was reviewed using a whiteboard specifically for this task. The category system was messy, a typical outcome of open coding. In particular, the features to merge codes and introduce new top-level codes were frequently used to refine the taxonomy. The users preferred this graph-based interaction with their codes over the list-based or tree-view-based way as offered by the Code Explorer. The users noted that they sometimes expanded annotated text passages of a code to remember what exactly this code describes and possibly rename it with a better description or identify codes that can be merged. However, they also noted that the view can become confusing if too many passages are annotated by a code. We plan to address this feedback in the next iteration. This use case demonstrated the benefit of linked material, where codes are directly linked to annotated text passages, and highlighted a key difference to Miro.

4.2. Actor Network

The platform ships with object detection and named entity recognition models that are applied during the pre-processing of texts and images. Automatically detected events, persons, and organizations are especially helpful in identifying actors involved in the discourse.

During an open coding phase and constructing the phenomenon structure, recurrent relationships among actors became evident, unveiling various actors assuming distinct roles and responsibilities. An actor-network (see Figure 3) was created with the *Whiteboards* extension to capture these relationships. Initially, the map was centered around a company and health app related to various data security problems in E-Health. Expanding on that, automatically detected entities, relevant actors, and relationships identified during close reading of the material were added to the board.

Actor networks are integral to the presentation, discussion, and reflection of results, the analysis process, and knowledge management. In this use-case, *Whiteboards* assisted in visualizing the relevant content of online articles. Furthermore, it helped in structuring the content of the discourse, an essential endeavor in discourse analysis to understand the phenomenon structure. Using *Whiteboards* instead of Miro offers the advantage of quickly inserting actors from processed research material. As codes, these can simply be added to a large number of objects, display the corresponding annotated text passages, and do not have to be laboriously inserted individually into a board.

4.3. Theoretical Sampling Map

Theoretical Sampling is an operation commonly used in Grounded Theory methodology. It involves theory-driven and selective data collection to refine and develop theories or concepts.

Following a selective coding phase, a Sampling Map (see Figure 4) was created. The main objectives are providing a comprehensive overview of the current progress and identifying missing data in the corpus to achieve theoretical saturation of the discourse.

The Sampling Map visualizes the documents selected for detailed analysis according to the Theoretical Sampling with their relevant actors. Starting from an initial document that appears to be relevant, the path through the analyzed documents can be depicted here by first analyzing very similar documents for minimal contrast for a more precise reconstruction. For an overall spectrum of the discourse, documents that are as different as possible are now used (maximum contrasting). This map helps to achieve data saturation and capture possible voids, such as identifying "silent" actors who

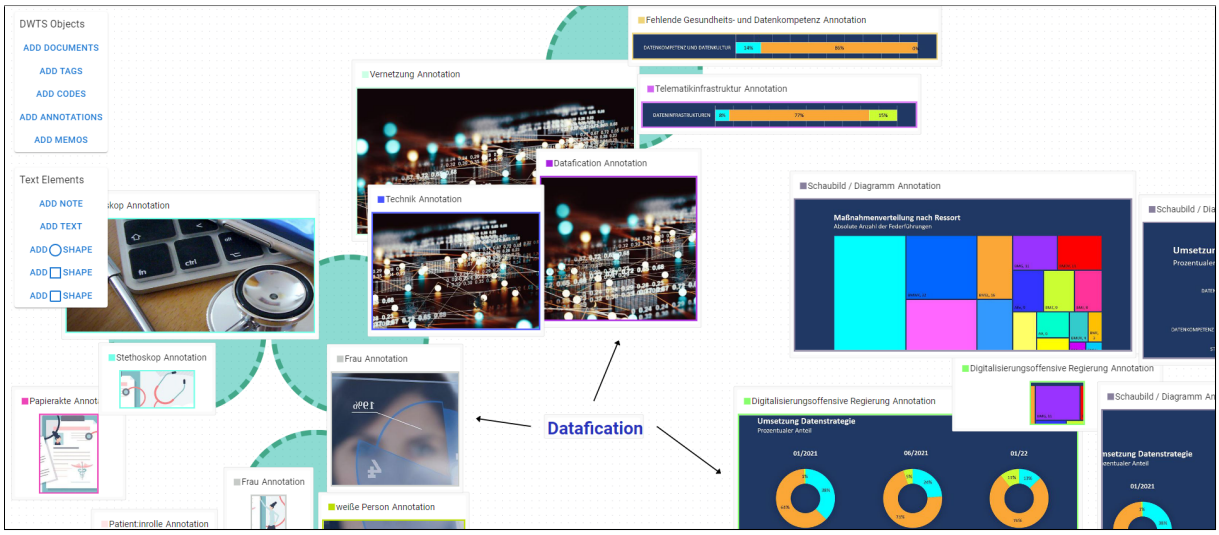

Figure 4: An excerpt of an image map to analyze visually conveyed knowledge, manually built with the *Whiteboards* extension.

are mentioned in the corpus but do not occupy a distinct speaker position.

By visualizing the material and associated actors with the *Whiteboards* extension, the users could effectively identify such gaps in their corpus. New documents were sampled explicitly in the following steps to achieve better saturation.

This conceptual use of the map highlights the need for integrated visualizations. Ensuring proximity to the data material and incorporating all analytical elements from the research process was particularly important for this scenario, making *Whiteboards* the preferred choice over Miro.

4.4. Image Map

The Discourse Analysis Tool Suite allows the processing of all common modalities: text, images, audio, and videos. The project on the debate on E-Health in Germany was conducted using mainly news articles as a data source, which are often accompanied by images.

Consequently, images were also annotated in the phase of open coding in addition text documents: They were analyzed with respect to certain attributes such as the roles of the depicted persons, the focus of the image, the type of image (e.g., photography vs. illustration), usage of the image (e.g., advertisement), etc.

Subsequently, the image map (see Figure X) was created with the *Whiteboards* extension to provide an overview of the discourse images as well as to analyze the visual dimension and the visually conveyed knowledge. In this map, the images were arranged by similarity, topic, and striking overlaps.

The image map supported the users to analyze which symbols are used for medical and technological fields and find relations to other societal discourses. Further, it facilitated the exploration of research questions related to investigating the social codes depicted in the images, comprehending the visual portrayal of health data digitization, and scrutinizing the absence of specific individuals, along with examining the portrayal of certain individuals in various roles.

5. Conclusion

In close co-creation of cultural anthropologists and computer scientists, we developed the *Whiteboards* extension, an interactive 2D canvas for viewing, organizing, and manipulating research objects.

It is fully integrated within the Discourse Analysis Tool Suite, linking nodes with their respective counterparts in the application and interleaving standard mind-map features with creative operations to manipulate database objects. This results in a unique user experience that enables an interactive and hermeneutic approach to digital discourse analysis.

We highlighted three empirical use cases, demonstrating that the proposed extension effectively supports a digital discourse analysis by reflecting with sampling maps, modeling actor networks, analyzing images, and organizing and refining complex code structures.

For future work, we aim to enhance the interaction with large datasets by implementing methods for grouping multiple objects into nodes, creating additional layers, and navigating between different layers, providing both a "distant" and "close" view of the research material. Further, we want to integrate quantitative analysis results like plots and tables as additional nodes.

6. Ethical Considerations and Limitations

The Discourse Analysis Tool Suite and the proposed extension in particular are tools intended for expert users. While the basic operations of adding, editing, moving and deleting nodes and edges are simple, working with *Whiteboards* may be challenging for new users unfamiliar with graph-based interfaces. Nevertheless, we believe that investing time in learning how to interact with the tool, despite its initial complexity, will ultimately yield significant benefits for research endeavors.

By extending our platform with *Whiteboards*, we do not introduce new ethical considerations but inherit those of our platform.

First, the platform could be misused for other work. Second, errors in the pre-processing pipeline can propagate and may lead to biased analysis results. Such errors even impact *Whiteboards* e.g., importing automatically annotated entities of the Named Entity Recognition module into a board. Third, we introduce a bias in our system design and workflow. While we implemented the platform in general and the *Whiteboards* extension in particular as flexible and generic as possible, we may still restrict users in their desired workflow with our design decisions. Finally, the platform requires users to upload their data to the system, raising concerns regarding privacy and security. We ensured that the platform is easy to deploy and can be self-hosted even by non-experts to maintain data sovereignty.

7. Bibliographical References

Adele E Clarke, Carrie Friese, and Rachel S Washburn. 2017. *Situational analysis: Grounded theory after the interpretive turn*. Sage publications.

Isabel Eiser, Tim Fischer, Florian Schneider, Gertraud Koch, Chris Biemann, and Fynn Petersen-Frey. 2023. Open Science Prinzipien und interdisziplinäre Kollaboration in D-WISE: Zwischen Hermeneutik und Digitaler Methode in der Diskursanalyse. In *9. Tagung des Verbands Digital Humanities im deutschsprachigen Raum*, page 252, Belval, Luxembourg and Trier, Germany.

Lina Franken. 2020. Kulturwissenschaftliches digitales Arbeiten: Qualitative Forschung als digitale Handarbeit? *Berliner Blätter*, 82:107–118.

Evelyn Gius, Jan Christoph Meister, Malte Meister, Marco Petris, Mareike Schumacher, and Dominik Gerstorfer. 2023. *CATMA 7 (Version 7.0)*. Zenodo.

Carsten Görg, Zhicheng Liu, and John Stasko. 2014. Reflections on the evolution of the jigsaw visual analytics system. *Information Visualization*, 13(4):336–345.

Reiner Keller. 2011. The Sociology of Knowledge Approach to Discourse (SKAD). *Human Studies*, 34:43–65.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

Gertraud Koch, Chris Biemann, Isabel Eiser, Tim Fischer, Florian Schneider, Teresa Stumpf, and Alejandra Tijerina García. 2022. D-WISE Tool Suite for the Sociology of Knowledge Approach to Discourse. In *Culture and Computing*, pages 68–83, Cham. Springer International Publishing.

Florian Schneider, Tim Fischer, Fynn Petersen-Frey, Isabel Eiser, Gertraud Koch, and Chris Biemann. 2023a. The D-WISE Tool Suite: Multi-Modal Machine-Learning-Powered Tools Supporting and Enhancing Digital Discourse Analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 328–335, Toronto, Canada. Association for Computational Linguistics.

Florian Schneider, Seid Muhie Yimam, Fynn Petersen-frey, Gerret Von Nordheim, Katharina Kleinen-von Königslöw, and Chris Biemann. 2023b. CodeAnno: Extending WebAnno with hierarchical document level annotation and automation. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 11–17, Dubrovnik, Croatia. Association for Computational Linguistics.

Anselm Strauss, Juliet Corbin, Solveigh Niewiarra, and Heiner Legewie. 1996. *Grounded Theory: Grundlagen Qualitativer Sozialforschung*. Beltz, Psychologie-Verlag-Union Weinheim.

Gregor Wiedemann, Seid Muhie Yimam, and Chris Biemann. 2018. New/s/leak 2.0 – multilingual information extraction and visualization for investigative journalism. In *Social Informatics*, pages 313–322, Cham. Springer International Publishing.