EuroVis Workshop on Visual Analytics (2024)
M. El-Assady and H.-J. Schulz (Editors)

Tag-Xplore: Interactive Exploration of Annotation Practices in Digital Editions

M. Blum[1], M. Sachdeva[1], Y. Stricker[2], R. Mumenthaler[2,3], and J. Bernard[1,3]

[1]University of Zürich, Zürich, Switzerland; [2]University Library of Zürich, Zürich, Switzerland; [3]Digital Society Initiative, Zürich, Switzerland

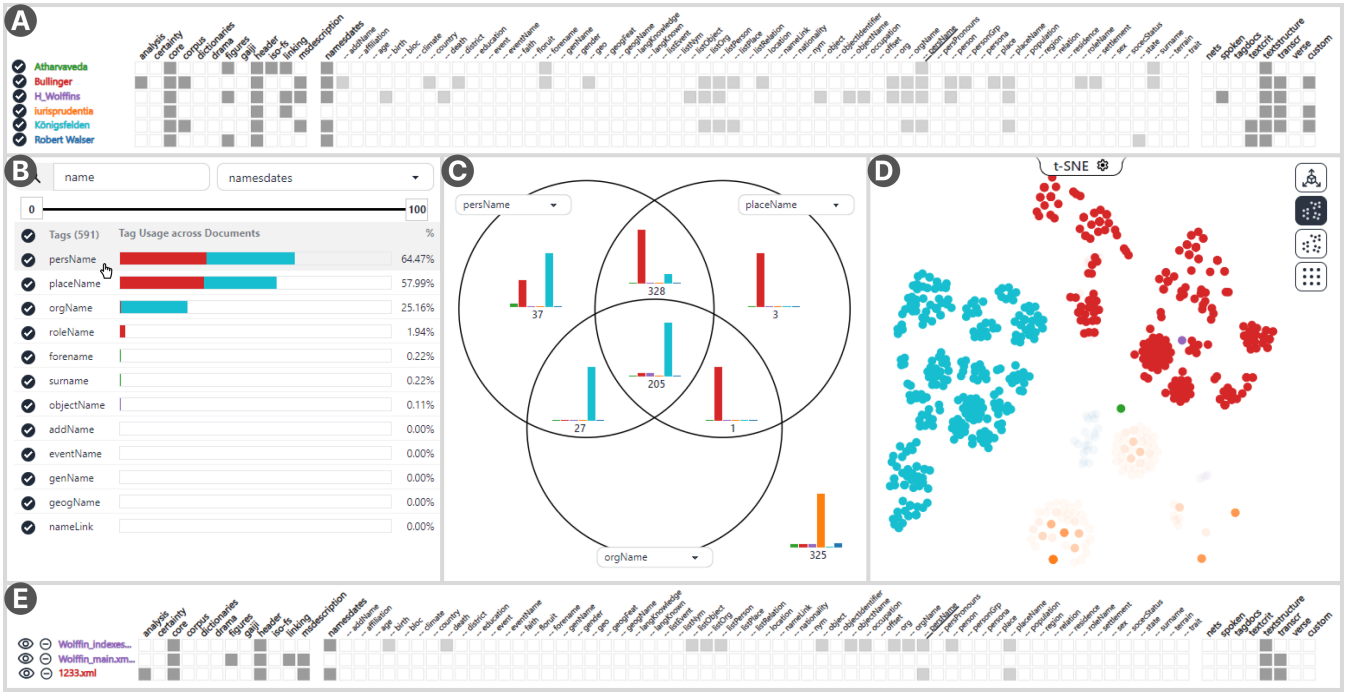

Figure 1: *Tag-Xplore enables data curators to explore annotation practices within and across digital editions. The five main coordinated views offer a systematic crosscut of attribute-document relationships from various perspectives and granularities. Specifically, views A and E allow the exploration of attributes across collections (editions), and documents respectively. Users can filter and rank attributes (B), compare attributes across documents (C), and gain an overview of documents based on attribute similarity (D). The example illustrates the investigation of annotation practices for three attributes:* persName, placeName, *and* orgName.

Abstract

Digital Editions (DE) are scholarly document collections that make research artifacts accessible to both humans and machines in a structured manner, enriched with annotations. However, the interoperability and reusability of DE can be hampered by annotation inconsistencies within DE and heterogeneous annotation practices across DE. We present Tag-Xplore, an interactive and visual exploration tool for annotation practices within and across DE. Tag-Xplore offers multiple coordinated views that provide both attribute-based and document-based access to the huge search space at multiple granularities. The approach also provides rank, filter, and comparison techniques, to further support the exploration. With Tag-Xplore, data curators can validate assumptions based on existing knowledge and generate new insights about annotation practices. We demonstrate the usefulness of Tag-Xplore with two qualitative case studies on attribute ambiguity and outlier documents.

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics; Visualization application domains;***

DOI: 10.2312/eurova.20241115

1. Introduction

In Digital Edition (DE) projects, domain experts from the Humanities curate and annotate cultural heritage artifacts in digital formats. The resulting DE open up numerous forms of usage for research in the Humanities and Social Sciences. Technical interoperability of DE, a cornerstone of the FAIR principles (findability, accessibility, interoperability, reusability) [WDA*16], is crucial to enhance research efficiency and scholarly impact. However, data curators encounter challenges in decision-making during curation due to lack of support, leading to interoperability issues [Elo20]. Firstly, curators are not always able to identify inconsistencies of annotation practices within their own DE [McD09]. Within-edition inconsistencies can occur when multiple actors are involved, or the curation phase spans long durations of time. Secondly, curators are often unaware of annotation practices of other editions [SBS*23, Fla16]. Across-edition inconsistencies lead to decreased interoperability and reusability [Kin11]. In essence, to enhance interoperability of DE, it is imperative to conduct a systematic analysis of the attributes across documents from various existing DE. However, this analysis is often infeasible for data curators in the scope of their annotation projects due to its complexity [HIL*20]. The current practice is often limited to problem-specific analysis solutions that are resource-consuming and not scalable. Data curators lack the ability to explore the complex search space defined by multiple DE consisting of hundreds of attributes and often thousands of documents per DE in a time-efficient manner.

Our approach contributes a Visual Analytics (VA) solution for the class of problems where the exploration of a large, hierarchical item space and a high-dimensional attribute space is required. We demonstrate the usefulness of our approach by applying it to the exploration of annotation practices in DE, a use case that fulfills our described data characteristics. The visual interface consists of multiple coordinated views for comparative and detailed analysis of annotation practices within DE, as well as across multiple DE. The different perspectives facilitate assumption validation and insight generation, allowing data curators to make more informed curation decisions, as we demonstrate in two short case studies. This support can enhance the interoperability and lifecycle of DE [Sab15].

2. Related Work

The Digital Humanities have been the focus of numerous VA approaches, enhancing the expert examination of research artifacts (*close reading* [JFCS17, Eve19]) and provided scholars with new abstract perspectives (*distant reading* [Mor05]). Applications of VA approaches varied, from poetry visualization [ARLC*13] and exploratory thematic analysis of digital archives [KES15] to the analysis of text transcripts [JEAGK17]. These approaches focus on the content of data collections, but do not reveal information about the metadata attributes. Metadata-focused approaches exist for the curation of photographs [AG23] and video annotations [RCLK10]. While these tools focus on metadata curation to improve interoperability, they omit the exploration of annotation practices that Tag-Xplore aims to provide. Inspiration is found in the work of Xu et al. [XEJJ14], providing data curators with multi-faceted metadata overview of large-scale collections and in the interactive sensemaking of Jigsaw [SGLS07]. Williamson [Wil15] surveys generic VA approaches for the visual representation of metadata, but does not focus on annotations of DE. A focus on exploration of annotation practices is found in *Reading Traces* [BBBD20] for digitized handmade annotations within a book collection to foster pattern discovery. However, a drawback is its close alignment with the specific case study it was developed for. In medieval studies, annotation tags are used to analyze two-variant texts [BJP*19], but their approach is limited to this specific task and cannot handle multiple documents. The *Interactive Text Mining Suite* [SD17] employs heatmaps for part-of-speech annotations and offers multifaceted corpus exploration, yet do not put emphasis on the integration of its various views into a cohesive VA tool. The *Compus* system [FD00] facilitates exploration of XML-encoded document corpora, highlighting structural patterns and discrepancies. Beyond Compus, Tag-Xplore, also enables experts to compare documents from multiple sources. To summarize, only few VA tools for the Digital Humanities support data curators in exploring annotation practices, facing metadata curation challenges of multiple DE, thousands of documents, and hundreds of attributes.

Edition Name	Documents	Total Tags	Unique Tags
Atharvaveda	2	75,301	65
Bullinger Digital	13,159	1,569,899	86
Heinrich Wölfflin	4	15,205	103
Iurisprudentia	352	534,847	31
Königsfelden	1,547	838,583	83
Robert Walser	20	94,271	70

Table 1: *The six DE are heterogeneous in document size and the amount of total and unique annotation tags (attributes) used.*

Related VA techniques that informed the design of Tag-Xplore mainly relate to the attribute/feature drill-down, ranking, and selection. Seo and Shneiderman introduced the rank-by-feature framework [SS05] to enable the exploration of ranked features and their relationships, complemented by Johansson and Johansson's use of association metrics for feature selection through ranking [JJ09]. The interactive arrangement, selection, and culling of attributes helps users to find interesting relations and cluster patterns in high-dimensional spaces [War94, Guo03, LSP*10]. Together with Ingram et al.'s Dimstiller [IMI*10], we provide feature selection for interactive dimensionality reduction, but do not focus on feature transformations. Finally, Cibulski et al. [CMP*19] study interactive feature selection in a data-specific context for multivariate time series. All these VA techniques are related, but differ in their data or machine learning focus. In particular, Tag-Xplore is special in its support for the exploration of annotation practices in DE.

3. Abstractions

We present the abstract data characteristics, before deriving requirements and tasks, and finally introducing six digital editions as a representative use case of our approach.

3.1. Data Characteristics

Tag-Xplore is designed as an exploration tool for large and complex search spaces, defined by thousands of items grouped into collections. The attribute space typically includes hundreds of different attributes, which can be grouped into semantically meaningful attribute groups. As such, both the items and attributes can be hierarchically structured, providing opportunities for drill-downs of the

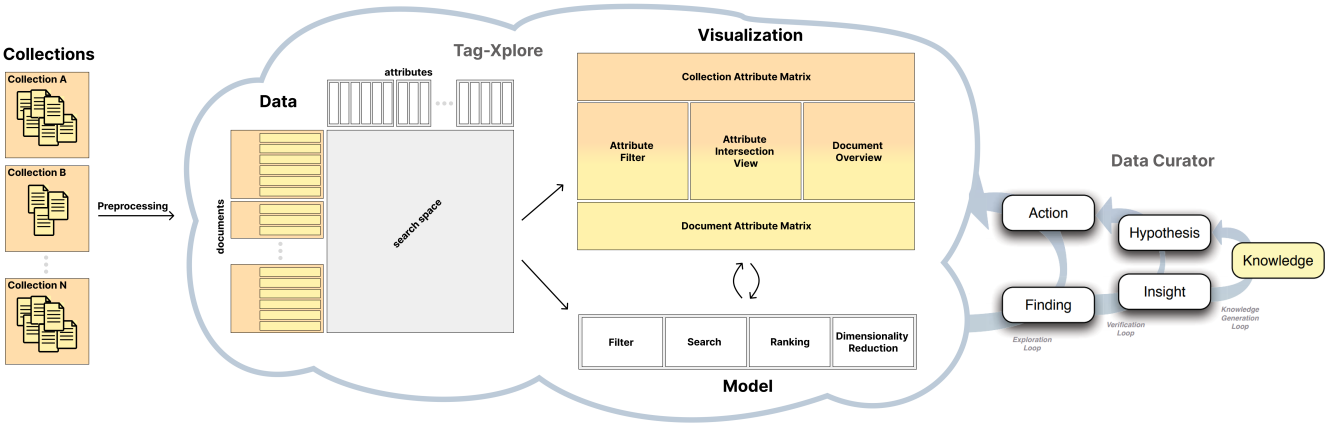

Figure 2: *Conceptualization of our approach in the context of the Knowledge Generation Model for VA [SSS*14]. Preprocessing of collections reveals the search space, spanned by hierarchically structured documents and attributes. Algorithmic models coupled with the visual interface include analysis support for overview exploration, filtering, search, ranking, and detailed analyses. The interface offers five main views and two auxiliary views, each with different perspectives on attribute-document crosscuts, differing by data granularity and overview-detail support. Through iterative exploration, data curators generate knowledge about annotation practices.*

search space. In Tag-Xplore, attributes are binary, indicating the presence or absence of an attribute for a given item.

3.2. Requirement & Task Abstraction

We use Sacha's *Knowledge Generation Model* [SSS*14] to introduce the principal analytical building blocks of Tag-Xplore. Figure 2 illustrates the instantiation of the KGM in the context of DE annotations as a representative case for the class of problems our approach is designed for. Assumptions of users can be based on existing knowledge or insights from previous exploration loops, and are usually about a specific subset of documents and/or attributes. Typically, users seek to answer attribute-based questions by investigating corresponding documents, and vice versa, document-based questions through a cross-cut with attribute interactions. This suggests perspectives on the data space that stringently take both attribute-based and the document-based characteristics into account. As usual for exploratory settings where the information need is not clearly defined at the start, one key strength that a VA solution must offer is its ability to be used iteratively in multiple cycles of exploration and verification and allow for individuals' unique analysis approaches [ITC08]. To summarize, we draw the following high-level requirements to Tag-Xplore:

$\mathbf{R}_1$: Attribute-based access to the search space
$\mathbf{R}_2$: Document-based access to the search space
$\mathbf{R}_3$: Systematic attribute-document crosscut for enhanced analysis
$\mathbf{R}_4$: Interactive exploration, iterativity, and drill-down capability

We break the identified requirements down into abstract tasks to inform the design of Tag-Xplore in a process based on two key sources: First, from influential related works on the various types and uses of DE curation, revealing remaining shortcomings. This perspective helped us to maintain a broad scope towards solving a general problem, aiming for a tool designed for broad applicability. Second, feedback from domain experts in the Digital Humanities who specialize in curating DE, one expert per studied edition, influenced our design choices. Here, we employed Munzner's Nested Model [Mun09] to achieve a thorough understanding of the domain, data, and tasks specific to curators.

Attribute-based tasks:
$\mathbf{T}_1$: Explore attributes across document collections
$\mathbf{T}_2$: Explore attributes across documents
$\mathbf{T}_3$: Rank attributes based on criteria of attribute usage
$\mathbf{T}_4$: Filter attributes for analyses on attribute usage
$\mathbf{T}_5$: Compare few attributes of interest across documents

Document-based tasks:
$\mathbf{T}_6$: Overview of documents based on attribute characteristics
$\mathbf{T}_7$: Filter documents to narrow down the search space
$\mathbf{T}_8$: Inspect document distribution across collections
$\mathbf{T}_9$: Inspect document structure in detail

3.3. Representative Case on Six Digital Editions

To design, develop, and study Tag-Xplore in a real-world setting, we use six heterogeneous DE coordinated by the Center for Digital Editions at the University of Zürich (ZDE). The documents from the DE are digital representation of research artifacts, structured and enriched with annotation tags. Naturally, the six DE form the grouping of document collections. Due to the heterogeneity of research contexts across the DE, the size and structure of the documents is highly heterogeneous (see Table 1). In the case of the Heinrich Bullinger letter edition, each document corresponds to one letter written or received by the influential reformer, totaling over 13,000 documents. All DE utilize a limited set of annotation tags, maintained by the Text Encoding Initiative guidelines (TEI) [Sch12]. Overall, the TEI standard has 586 annotation tags, grouped into so-called "modules", reflecting their semantic context. As an example, tags relevant for the annotation of screenplays are grouped in the *performance* module. In our preprocessing, we parse all documents and extract the information about their used annotation tags, i.e., if a tag appears in a specific document or not.

4. Visual Interface

Based on the data characteristics, the requirements, and tasks outlined in Section 3, we designed and developed Tag-Xplore. We first provide an overview before introducing the views in detail.

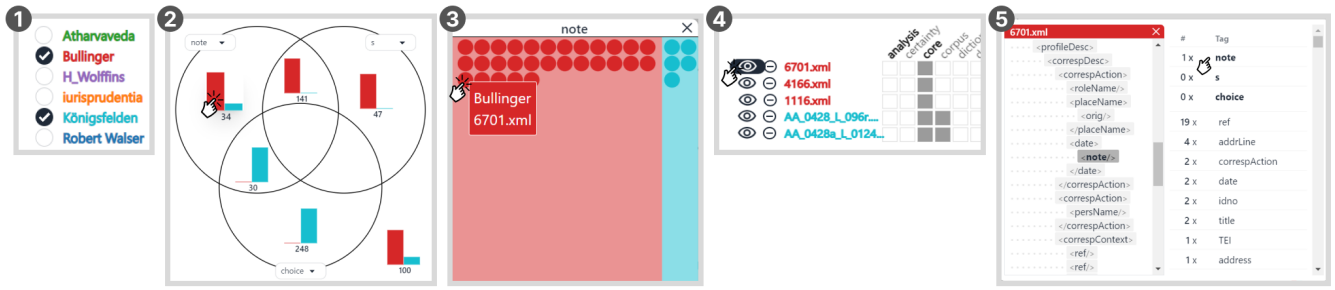

Figure 3: *Example of how different views of Tag-Xplore are used in the knowledge generation process. (1) Filtering of document search space (T_7) on collection granularity. (2) Comparison of three candidate attributes across documents (T_5). (3) Inspection of document distribution across collections (T_8), with document selection. (4) Exploration of attributes across documents (T_2). (5) Inspection of document structure (T_9). With this sequence of actions, the user gained insights about ambiguous annotation practices for annotations across two DE.*

4.1. Overview

The interface consists of five main views (see Figure 1) and two auxiliary views (see Figure 3.3 & 3.5). The different views are designed to accommodate the different perspectives and access needed for the diverse information needs. Attribute-centric access includes views for the exploration, ranking, filtering and comparison of attributes of interest ($\mathbf{R}_1$). Document-centric access includes overview, filtering and detail inspection views ($\mathbf{R}_2$). All main views of the interface provide different analysis possibilities for the document-attribute crosscut ($\mathbf{R}_3$). Differences in aggregation levels (e.g., document or collection granularity) allow for drill-down capabilities ($\mathbf{R}_4$) which is further supported by the auxiliary detail views. In addition, interactive exploration is supported by the connectivity of the views. Interactions with the visual interface (e.g., hovering over a document) or with the underlying model (e.g., filtering) are immediately reflected across all views. Figure 1 demonstrates an interaction where the user hovers over the *persName* attribute (view B), resulting in the highlighting of all documents that contain this attribute (view D). The consistent use of color to differentiate between documents from the various DE provide orientation for across-edition analysis tasks.

4.2. Views

The ***Collection Attribute Matrix*** (Figure 1A) allows the edition-level exploration of attributes ($\mathbf{T}_1$) and enables filtering of documents on an edition granularity ($\mathbf{T}_7$). In the grid heatmap, the rows represent the DE and columns represent the attributes aggregated by groups. Clicking on an attribute group will expand it horizontally and reveal the individual attributes. The coloring of the matrix cells indicates whether an edition is using a specific attribute or not. As an example, the row-wise comparison of the *birth* column in Figure 1 reveals that only one DE is annotating the birthdate of persons. Users can filter documents on a DE granularity with toggles next to the DE names ($\mathbf{T}_7$). Likewise, users can filter attributes by selecting or deselecting the individual columns ($\mathbf{T}_4$).

The ***Attribute Filter*** (Figure 1B) is the main view for ranking ($\mathbf{T}_3$) and filtering attributes ($\mathbf{T}_4$). Multi-faceted filters enable refinement of the attribute space, which can be based on attribute group, attribute name, and usage amount. By default, the attributes are ranked in descending order based on the total amount of documents using the attribute. The horizontally stacked bar charts reveal the distribution of documents across DE utilizing the attribute.

The ***Attribute Intersection View*** (Figure 1C) is tailored towards the investigation of ambiguous attribute usage. It enables the comparison of up to three attributes across documents ($\mathbf{T}_5$). Bar charts within the circle intersection areas show the number of documents from each DE that use the respective attribute combinations.

The ***Document Overview*** (Figure 1D) allows users to find clusters and outliers of documents based on attribute usage ($\mathbf{T}_6$). Each document is mapped onto a two-dimensional layout, where the proximity of documents indicates their attribute-based similarity, as a result of different dimensionality reduction techniques. The input vector for the reduction technique consists of the currently filtered attributes. Users can choose from four dimensionality reduction techniques (PCA, t-SNE, UMAP, SOM) [AHT20] and configure the respective model parameters. Figure 4 displays the four dimensionality reduction techniques applied to the same search space.

The ***Document Attribute Matrix*** (Figure 1E) allows the document-level exploration of attributes ($\mathbf{T}_2$). We make use of the grid heatmap introduced in the *Collection Attribute Matrix*, but this time the rows represent single documents. Row-wise comparisons can reveal similarities and differences across documents of interest.

The ***Document Listing*** (Figure 3.3) allows users to inspect the document distribution across DE. This enables the transition from DE aggregations to concrete document instances. The view is accessible by clicking on the bar charts in the *Attribute Filter*, an area within the *Attribute Intersection View*, or a grid cell of the self-organizing map in the *Document Overview*.

The ***Document Inspector*** (Figure 3.5) allows the inspection of the hierarchically structured document, providing detailed information about attribute annotation practices and the context in which attributes are used ($\mathbf{T}_9$). It is the lowest level of detail our visual interface offers, used to confirm or contextualize generated insights at the end of the exploration process. To provide additional information about the attributes used in the document, they are listed next to the document viewer, sorted from most to least used.

5. Case Studies

We demonstrate the usefulness of Tag-Xplore by reporting two qualitative case studies with two experts *Phillip* and *Reto*, responsible for the curation of two of the six DE described in Section 3.3. The experts were instructed to share their thoughts during exploration, leading to a think-aloud protocol aligned with the workflow.

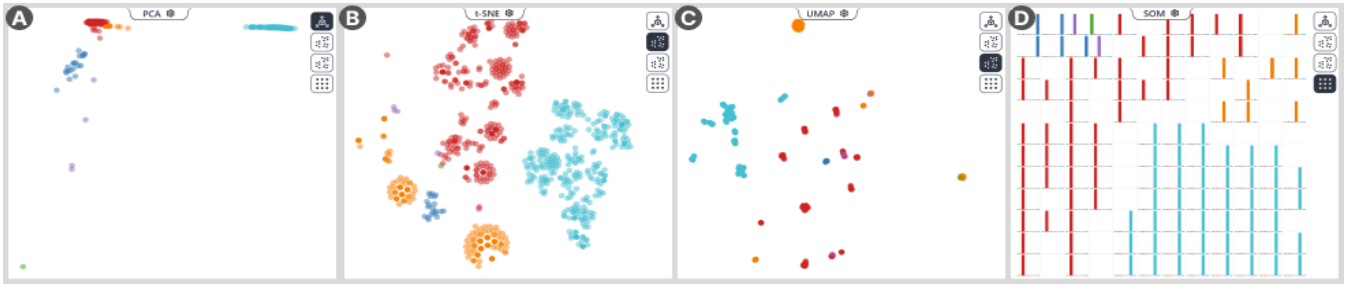

Figure 4: *Users can choose between four dimensionality reduction techniques in the* Document Overview *(from left to right: PCA, t-SNE, UMAP, SOM), and steer their model parameters. Each point represents a document, colored by its edition association. A series of patterns can be observed, revealing similarities and differences in annotation practices.*

5.1. Identifying Ambiguous Annotation Practices

Phillip, a data curator for the *Bullinger* letter edition, wants to gain insights on the progress of the letter digitalization process. He begins by filtering the search space exclusively to documents from the *Bullinger* DE ($\mathbf{T}_7$, see Step 1 of the interaction sequence for this case study depicted in Figure 3). Aware that the *s* attribute is used for the annotation of sentences in the automatic transcription process and the *note* attribute is used for footnotes, he selects both attributes in the *Attribute Intersection View* ($\mathbf{T}_5$). He finds out that many documents contain both attributes, indicating that the curation process of these documents is finished. Curious about how other DE handle their footnote annotations, he decides to add the *Königsfelden* DE to his search space ($\mathbf{T}_7$). The updated *Attribute Intersection View* reveals a shared annotation practice, as both DE employ the *note* attribute. To confirm his insight about the shared annotation practice, he selects documents from both DE via the *Document Listing* view and inspects their structure in the *Document Inspector* ($\mathbf{T}_9$). He finds out that the *Königsfelden* documents utilize the *note* attribute to provide more information about the digitalization process, rather than for annotating the text. He concludes that this is a typical instance of ambiguous attribute usage and that he would consider these alternative solutions if he would start over.

5.2. Finding Outlier Documents

Reto, who has extensive practical experience with the TEI annotation guidelines, wants to learn more about the reasons for interoperability issues across DE. He gains a first overview of the document space using the *Document Overview* ($\mathbf{T}_6$). He notices that the spatial differences between groups of documents in the PCA dimensionality reduction (see Figure 4A) hint towards cons differences across DE. He forms the assumption that the annotations from the *namesdates* group could contribute to this heterogeneity, and applies the according group filter in the *Attribute Filter* ($\mathbf{T}_4$). The updated *Document Overview* now reveals clear outlier documents. The subsequent row-wise comparison of the documents in the *Document Attribute Matrix* reveals an outlier document, which he further investigates using the *Document Inspector* ($\mathbf{T}_9$). He concludes his exploration with this newly discovered document structure type, which needs to be refined to fit the rest of the documents.

6. Discussion and Future Work

Guidance: Tag-Xplore is novel in allowing the exploration of annotation practices across multiple collections. Users can quickly gain insights from collections and attributes that were previously unknown. However, the large search space of thousands of documents and hundreds of attributes may overwhelm users. In the future, we plan to integrate guidance components to navigate this space more easily, by suggesting documents and attributes.

Knowledge Externalization: Tag-Xplore enables sense-making of the data, but the externalization of knowledge in the exploration process is not yet integrated. Leveraging VA feedback-loop principles could form an important basis for collaboration and objective discourse about interoperability, using the example of DE.

What-If: During our collaboration with data curators, we observed their strong curiosity to explore their own annotation practices and compare it with external DE. With Tag-Xplore, data curators were able to verify their assumptions, complementing the more open-ended explorations observed. In future work, a *what-if* analysis will enable data curators to adapt annotation practices, and assess the impact of refinements on interoperability.

Generalization of Use: We have designed and evaluated our approach around the use case of annotation practices of DE. However, with minor prototype adjustments, other data that fulfills the characteristics described in Section 3 can also be explored. This envisioned generalizability goes beyond document-based collections and will be the focus of our future work on Tag-Xplore.

7. Conclusion

We have presented a Visual Analytics approach for the exploration of annotation practices within and across digital editions. Tag-Xplore enables curators to efficiently navigate the search space, facilitating the filtering, exploration, ranking, and comparison of attribute usage from various perspectives of the attribute-document crosscut. This empowers users to validate their assumptions and generate new insights, thereby uncovering unexpected patterns of attribute usage within their own edition as well as in comparison with other editions. Such systematic analysis, previously challenging and resource-intensive, is now achievable with Tag-Xplore. The two case studies exemplify typical situations how data curators can generate knowledge about annotations with Tag-Xplore, enabling them to make more informed decisions during curation, which leads to enhanced interoperability of DE.

Acknowledgments We thank the Humanities and Library communities of the Digital Society Initiative for the interdisciplinary collaboration. Special thanks go to Reto Baumgartner and Phillip Ströbel for their valuable contributions to our case studies.

References

[AG23] ASKE K., GIARDINETTI M.: (mis) matching metadata: Improving accessibility in digital visual archives through the eycon project. *Computing and Cultural Heritage 16*, 4 (2023), 1–20. doi:10.1145/3594726. 2

[AHT20] AYESHA S., HANIF M. K., TALIB R.: Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion 59* (2020), 44–58. doi:10.1016/j.inffus.2020.01.005. 4

[ARLC*13] ABDUL-RAHMAN A., LEIN J., COLES K., MAGUIRE E., MEYER M., WYNNE M., JOHNSON C. R., TREFETHEN A., CHEN M.: Rule-based visual mappings–with a case study on poetry visualization. In *CGF* (2013), vol. 32, Wiley Online Library, pp. 381–390. doi:10.1111/cgf.12125. 2

[BBBD20] BLUDAU M.-J., BRÜGGEMANN V., BUSCH A., DÖRK M.: Reading traces: scalable exploration in elastic visualizations of cultural heritage data. In *CGF* (2020), vol. 39, Wiley Online Library, pp. 77–87. doi:10.1111/cgf.13964. 2

[BJP*19] BAUMANN M., JOHN M., PFLÜGER H., HERBERICHS C., VIEHHAUSER G., KNOPKI W., ERTL T.: An interactive visualization for the analysis of annotated text variance in the legendary der heiligen leben, redaktion. In *Leipzig Symposium on Visualization in Applications (LEVIA)* (2019). doi:10.31219/osf.io/wd9yz. 2

[CMP*19] CIBULSKI L., MAY T., PREIM B., BERNARD J., KOHLHAMMER J.: Visualizing time series consistency for feature selection. *J. WSCG 27*, 2 (2019). doi:10.24132/JWSCG.2019.27.2.2. 2

[Elo20] ELO K.: Big data bad metadata: A methodological note on the importance of good metadata in the age of digital history. In *Digital Histories*. Helsinki University Press, 2020, pp. 103–111. doi:10.33134/HUP-5-6. 2

[Eve19] EVE M. P.: *Close reading with computers: Textual scholarship, computational formalism, and David Mitchell's Cloud atlas*. Stanford University Press, 2019. doi:10.21627/9781503609372. 2

[FD00] FEKETE J.-D., DUFOURNAUD N.: Compus: visualization and analysis of structured documents for understanding social life in the 16th century. In *Conference on Digital Libraries* (2000), ACM, pp. 47–55. doi:10.1145/336597.336632. 2

[Fla16] FLANDERS J.: Collaboration and dissent: Challenges of collaborative standards for digital humanities. In *Collaborative research in the digital humanities*. Routledge, 2016, pp. 67–80. doi:10.4324/9781315572659. 2

[Guo03] GUO D.: Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. *Information Visualization 2*, 4 (2003), 232–246. doi:10.1057/PALGRAVE.IVS.9500053. 2

[HIL*20] HARROWER N., IMMENHAUSER B., LAUER G., MARYL M., ORLANDI T., RENTIER B., WANDL-VOGT E.: Sustainable and fair data sharing in the humanities. doi:10.7486/DRI.tq582c863. 2

[IMI*10] INGRAM S., MUNZNER T., IRVINE V., TORY M., BERGNER S., MÖLLER T.: Dimstiller: Workflows for dimensional analysis and reduction. In *IEEE Conf. Visual Analytics Science & Technology (VAST)* (2010), IEEE, pp. 3–10. doi:10.1109/VAST.2010.5652392. 2

[ITC08] ISENBERG P., TANG A., CARPENDALE S.: An exploratory study of visual information analysis. In *ACM Conf. Human Factors in Computing Systems (CHI)* (2008), pp. 1217–1226. doi:10.1145/1357054.1357245. 3

[JEAGK17] JENTNER W., EL-ASSADY M., GIPP B., KEIM D. A.: Feature alignment for the analysis of verbatim text transcripts. In *EuroVis Wkshp. Visual Analytics (EuroVA)* (2017), pp. 13–18. doi:10.2312/eurova.20171113. 2

[JFCS17] JÄNICKE S., FRANZINI G., CHEEMA M. F., SCHEUERMANN G.: Visual text analysis in digital humanities. In *Computer Graphics Forum (CGF)* (2017), vol. 36, Wiley Online Library, pp. 226–250. doi:10.1111/cgf.12873. 2

[JJ09] JOHANSSON S., JOHANSSON J.: Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Trans. Visualization & Computer Graphics (TVCG) 15*, 6 (2009), 993–1000. doi:10.1109/TVCG.2009.153. 2

[KES15] KLEIN L. F., EISENSTEIN J., SUN I.: Exploratory thematic analysis for digitized archival collections. *Digital Scholarship in the Humanities 30* (2015), i130–i141. doi:10.1093/llc/fqv052. 2

[Kin11] KING G.: Ensuring the data-rich future of the social sciences. *science 331*, 6018 (2011), 719–721. doi:10.1126/science.1197872. 2

[LSP*10] LEX A., STREIT M., PARTL C., KASHOFER K., SCHMALSTIEG D.: Comparative analysis of multidimensional, quantitative data. *IEEE Trans. Visualization & Computer Graphics (TVCG) 16*, 6 (2010), 1027–1035. doi:10.1109/TVCG.2010.138. 2

[McD09] MCDONOUGH J.: Xml, interoperability and the social construction of markup languages: the library example. *Digital Humanities Quarterly 3*, 3 (2009). URL: http://digitalhumanities.org/dhq/vol/3/3/000053/000053.html. 2

[Mor05] MORETTI F.: *Graphs, maps, trees: abstract models for a literary history*. Verso, 2005. doi:10.1093/fmls/cql045. 2

[Mun09] MUNZNER T.: A nested process model for visualization design and validation. *IEEE Trans. Visualization & Computer Graphics (TVCG) 15*, 6 (2009), 921–928. doi:10.1109/TVCG.2009.111. 3

[RCLK10] RENZEL D., CAO Y., LOTTKO M., KLAMMA R.: Collaborative video annotation for multimedia sharing between experts and amateurs. In *WISMA* (2010), Citeseer, pp. 7–14. URL: https://publications.rwth-aachen.de/record/125461. 2

[Sab15] SABHARWAL A.: *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Chandos Publishing, 2015. doi:10.1016/B978-0-08-100143-1.00002-7. 2

[SBS*23] SACHDEVA M., BLUM M., STRICKER Y., SCHRECK T., MUMENTHALER R., BERNARD J.: Ord-xplore: Bridging open research data collections through modality abstractions. doi:10.2312/evp.20231065. 2

[Sch12] SCHMIDT D.: The role of markup in the digital humanities. *Historical Social Research* (2012), 125–146. doi:10.12759/hsr.37.2012.3.125-146. 3

[SD17] SCRIVNER O., DAVIS J.: Interactive text mining suite: Data visualization for literary studies. In *CDH@ TLT* (2017), pp. 29–38. URL: https://ceur-ws.org/Vol-1786/scrivner.pdf. 2

[SGLS07] STASKO J., GORG C., LIU Z., SINGHAL K.: Jigsaw: supporting investigative analysis through interactive visualization. In *Visual Analytics Science and Technology* (2007), IEEE, pp. 131–138. doi:10.1057/palgrave.ivs.9500180. 2

[SS05] SEO J., SHNEIDERMAN B.: A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization 4*, 2 (2005), 96–113. doi:10.1057/PALGRAVE.IVS.9500091. 2

[SSS*14] SACHA D., STOFFEL A., STOFFEL F., KWON B. C., ELLIS G., KEIM D. A.: Knowledge generation model for visual analytics. *IEEE Trans. Visualization & Computer Graphics (TVCG) 20*, 12 (2014), 1604–1613. doi:10.1109/TVCG.2014.2346481. 3

[War94] WARD M. O.: Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Visualization Conference* (1994), IEEE, pp. 326–333. doi:10.1109/VISUAL.1994.346302. 2

[WDA*16] WILKINSON M. D., DUMONTIER M., AALBERSBERG I. J., APPLETON G., AXTON M., BAAK A., BLOMBERG N., BOITEN J.-W., DA SILVA SANTOS L. B., BOURNE P. E., ET AL.: The fair guiding principles for scientific data management and stewardship. *Scientific data 3*, 1 (2016), 1–9. doi:10.1038/sdata.2016.18. 2

[Wil15] WILLIAMSON E. P.: Seeing metadata: Representing digital collection metadata. doi:10.14288/sa.v1i1.186327. 2

[XEJJ14] XU W., ESTEVA M., JAIN S. D., JAIN V.: Interactive visualization for curatorial analysis of large digital collection. *Information Visualization 13*, 2 (2014), 159–183. doi:10.1177/1473871612473590. 2