Algorithmic Edition

Sebastian Enns, Maximilian Michel, Andreas Kuczera;

Introduction

Digital editions are a cornerstone of the Digital Humanities, offering enriched resources that combine content with context (Sahle 2017). These editions serve as essential tools for research, incorporating structured annotations that support both traditional inquiry and computational analysis (Pagel et al. 2020). However, their reliance on visual platforms prioritizes accessibility over the algorithmic possibilities of digital texts (Buzzetti/Rehbein 2008). To address this, scholars propose a “text-as-network” paradigm, emphasizing the interconnectivity of data points to unlock their computational potential (Witt 2018: 219).

Annotations play a crucial role in this context, as they provide added layers of meaning to specific parts of a text, such as historical context, linguistic features, or thematic connections (Reiter 2020: 126). These annotations not only support human interpretation but are also key to preparing texts for computational analysis, transforming them into machine-readable datasets that facilitate deeper exploration (Pagel et al. 2020). Therefore, annotations are not merely passive metadata but active participants in scholarly discourse, connecting texts with insights derived from close, data-driven analysis (Pagel et al. 2020; Reiter 2020: 135). This function becomes even more pronounced when text fragments and their annotated context could be extracted, reassembled, and grouped around specific topics or themes.

In this paper, we explore how digital editions can serve as a foundation for what we call **algorithmic editions**. Central to this transition is the ability to extract, combine, and contextualize annotated text fragments, enabling their use for human and computational analysis. By leveraging technologies and concepts such as Labeled-Property Graphs (LPG), Applied Text As Graph (ATAG) and Explicitly Notated Citations (ENC), we demonstrate how these approaches make it possible to utilize text fragments, bridging the gap between traditional digital and algorithmic editions.

Towards an Algorithmic Edition

While the vision of algorithmic editions is promising, significant challenges arise when examining the current state of digital editions and annotation practices. These challenges stem from technical and yet unresolved limitations, which must be addressed to further explain the potential of an algorithmic edition.

Anchoring Problems

A foundational issue is the fragility of anchoring third-party annotations to textual fragments. While scholars conceptualize annotations as precise references to specific parts of a text, the reality of implementation is often tied to the technical framework of web browsers. As Koolen and Boot (2020) point out, browsers only interpret annotations as parts of a webpage's HTML structure, creating a disconnect between scholarly intent and technical

execution. This reliance on HTML structure leads to instability, where changes in layout or relocation of web content render anchors invalid, undermining their reliability for long-term use. Moreover, one of the greatest challenges in this context lies in the flexible granularity when referencing these fragments, which cannot be adequately identified due to the lack of chapter, paragraph, or line references (Bernhart/Hahn 2014: 227; Ralle 2016: 155).

Lack of Persistent Linking

Annotations and text fragments in current digital editions are typically tied to the layout or URL of a single webpage. Furthermore, general URLs are often unreliable due to frequent changes or the unavailability of linked resources (Bleier 2021: 4; Stronks/Boot 2007: 169). As a result, permalinks and persistent identifiers, such as DOIs, have become increasingly important (Bleier 2021: 4). However, permalinks typically point to static content and fail to account for parameterized or dynamic views of texts in digital editions (Stäcker 2020: 9). Additionally, their reliability largely depends on the institution responsible for their maintenance (Bleier 2021: 4). This lack of persistent linking prevents their reuse across multiple editions or representations of the same text (Koolen/Boot 2020). As a result, fragments are siloed within specific platforms, limiting their scalability and their ability to support broader, cross-edition computational studies.

Early solutions like Canonical Text Services (CTS) URNs aimed to provide persistent, structured references for textual passages, particularly in classical corpora (Tiepmar/Heyer 2017). However, their strictly hierarchical model limits their applicability in networked, overlapping annotation contexts. Similarly, robust hyperlinks as proposed by Phelps and Wilensky (2000) addressed URL instability through lexical signatures, but remain tied to surface-level document retrieval and lack structural or semantic depth. These models highlight the foundational need for persistent and content-aware referencing.

Applied Text as Graph (ATAG)

Labeled-property graphs (LPG) allow for the representation of complex relationships and multidimensional aspects. They consist of nodes, edges, and properties, providing a flexible framework for text elements and their connections (Kuczera 2022: 103). The ATAG approach expands this by segmenting text into manageable units within a graph structure. Each character is transformed into a node with a unique identifier, connected sequentially by edges that define their relationships and order (Kuczera 2024).

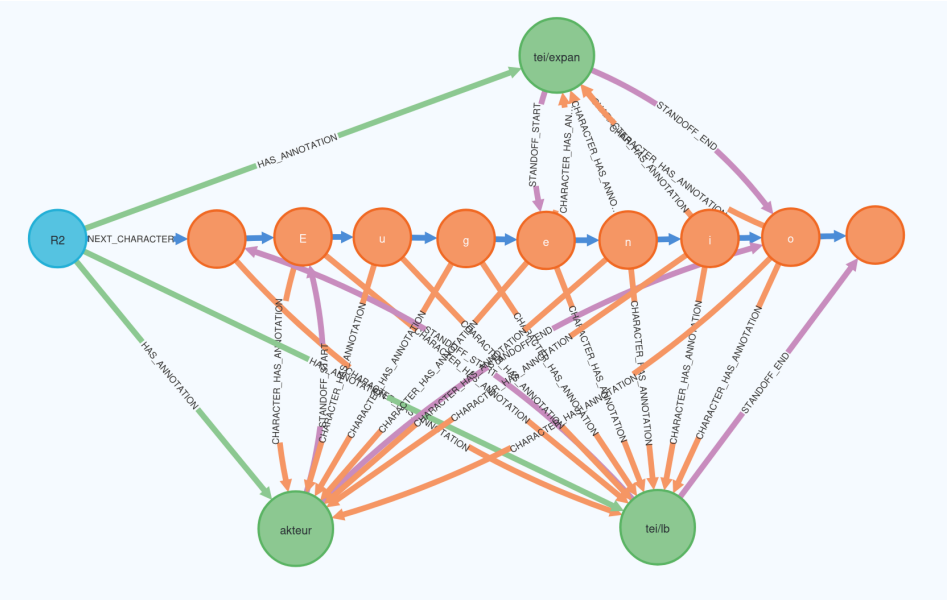

Figure 1: A LPG visualizing the text "Eugenio" with letters as orange nodes connected by edges to green annotation nodes, which are linked to the central text node R2 (Kuczera 2024: 2).

Figure 1 shows a graph fragment for the text "Eugenio" with associated annotations. Orange nodes represent individual characters linked sequentially, while green nodes indicate annotations tied to the text chain, illustrating their relationships and scope. A central blue node integrates all elements, reflecting the character sequence and text properties. Each node is uniquely referenced by a UUID, enabling multiple points of connection in a networked structure.

Explicitly Notated Citations (ENC)

In the digital edition *The Socinian Correspondence*, we introduced dynamic placement of text markers and the creation of citation links, as illustrated in Figure 2. These permanent links enable access to both the broader context and the specifically cited text passage. When the link is used, the relevant page in the edition opens and automatically navigates to the marked section. This method is also applicable to annotations and other objects within the edition, leveraging their persistent identifiers.

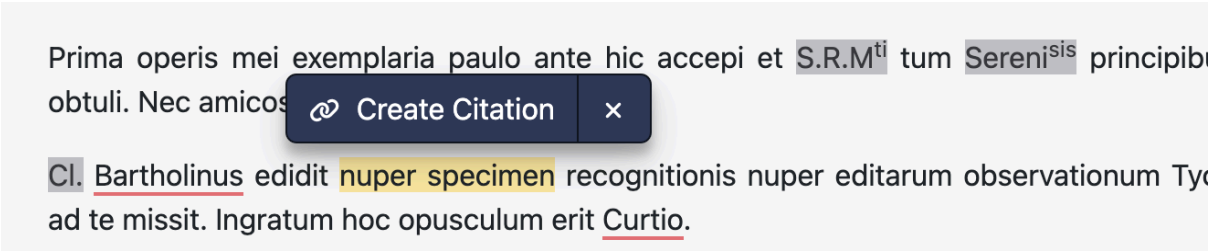

Figure 2: Screenshot of the "Citation" feature in the digital edition *The Socinian Correspondence*, demonstrating the creation of permanent references to specific text passages.

However, a key issue arises as these citation links depend on the start and end indices of the marked text. Changes to the text can lead to broken or inconsistent references, compromising the reliability of the links.

The **Explicitly Notated Citations (ENC)** method is an advancement in citation practices for digital editions (Enns/Kuczera 2025). ENC replaces traditional start- and end-index parameters with unique character identifiers from the ATAG chain. Each citation uses the UUIDs of the first and last characters in the marked section, along with a hash parameter that encodes the entire ATAG chain for integrity verification. If the UUIDs or hash do not match, it indicates a change in the referenced text, and users are notified accordingly. Figure 3 illustrates the structure of an ENC-link compared to the previously mentioned citation link.

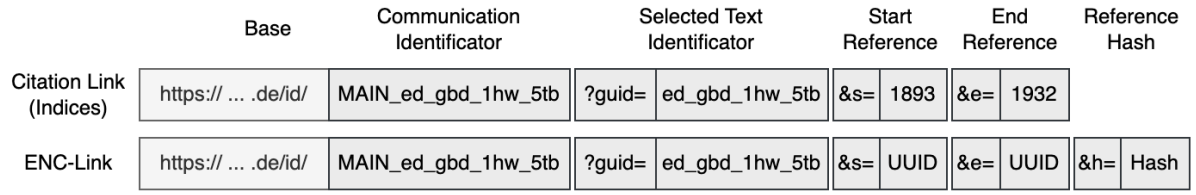

Figure 3: Overview of the structure of a citation link using indices compared to an ENC-link, which utilizes unique identifiers instead of indices and includes a hash for verification.

ENC resolves key issues in digital citation by creating stable, dynamic, and granular links. These links point not only to the specific ATAG chain but also to its broader textual context. To ensure persistence, the referenced text and its annotations can be exported in JSON format and versioned using Git, with commit hashes added to the citation link. This allows previous text states to be retrieved and referenced, ensuring that citations remain valid even after changes to the source text. By supporting ENC-links as endpoints for cited texts and annotations in machine-readable formats like JSON, ENC unlocks new possibilities for algorithmic editions. This capability enables automated processing, deeper analysis, and integration with computational tools, advancing the scope of digital scholarly work.

Algorithmic Edition

An algorithmic edition combines ATAG with ENC-links and enables machine-readability and computational engagement with traditional human interaction. Unlike conventional digital editions, which rely heavily on visual interfaces like websites, algorithmic editions also focus on structuring texts and annotations for algorithmic analysis and integration. They prioritize precise access to textual content, annotations, and contextual metadata. Every element – texts, annotations, and their relationships – is uniquely addressable, enabling seamless retrieval and integration into computational workflows.

Using methods such as ENC-links, algorithmic editions allow direct access to specific text segments. These links return metadata like semantic annotations, layout details, and linguistic information (e.g., part-of-speech tags), transforming the edition into a comprehensive data repository for both human and algorithmic use. This granularity supports modern scholarly practices, including detailed citation, intertextual analysis, and exploration of complex textual networks. Such a network-oriented approach views texts as

interconnected nodes within broader systems, emphasizing relationships not only between textual elements but also across datasets (Witt 2018).

Notably, an algorithmic edition does not require a visual interface like a website; it can exist as a purely machine-readable structure, enabling integration into diverse digital ecosystems. This flexibility enhances its usability, fostering interdisciplinary collaboration and innovation. While visual presentation remains vital for many scholars, the underlying structure is optimized for machine interaction, ensuring texts and metadata are equally accessible to human readers and algorithms.

In essence, an algorithmic edition is not merely a digital reproduction of a text but a dynamic system designed for comprehensive exploration. Through its structured approach, it facilitates new insights into both textual content and the cultural practices encoded within (Krämer 2023: 7). It redefines digital editions, integrating texts, data, and technology to advance scholarly practice.

Conclusion

Algorithmic editions redefine digital scholarship by integrating ATAG and ENC techniques to create stable, granular access to texts and annotations. ATAG structures texts as networks of interlinked nodes, while ENC ensures reliable citation through unique identifiers and hash-based verification. Together, these approaches facilitate seamless human and machine interaction, enabling interdisciplinary research and data-driven insights. By treating texts as dynamic networks, algorithmic editions enhance accessibility, computational analysis, and scholarly exploration, setting a foundation for the future of digital humanities.

Bibliography

Bernhart, Toni / Hahn, Carolin (2014): "Datenmodellierung in digitalen Briefeditionen und ihre interpretatorische Leistung. Ontologien, Textgenetik und Visualisierungsstrategien. Workshop im Jacob-und-Wilhelm-Grimm-Zentrum der Humboldt- Universität zu Berlin, 15./16. Mai 2014" Editio, vol. 28, no. 1, 2014, pp. 225-229. DOI: https://doi.org/10.1515/editio-2014-015.

Bleier, Roman (2021): "How to cite this digital edition?" DHQ: Digital Humanities Quarterly 15 (3). Available at: https://digitalhumanities.org/dhq/vol/15/3/000561/000561.html (Accessed: April 23, 2025).

Buzzetti, Dino / Rehbein, Malte (2008): **“**Towards a model for dynamic text editions”. In: Lisa Lena Opas-Hänninen, Mikko Jokelainen, Ilkka Juuso und Tapio Seppänen (Hg.): Digital Humanities 2008. Oulu: University of Oulu, S. 78–81.

Enns, Sebastian / Kuczera, Andreas (2025): “Explicitly Notated Citations. Von digitalen zu algorithmischen Editionen”. DHd 2025 Under Construction, Bielefeld, Deutschland. *Zenodo*. DOI: https://doi.org/10.5281/zenodo.15112014.

Koolen, Marijn / Boot, Peter (2020): "Facilitating Reusable Third-Party Annotations in Digital Editions". Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization, edited by Julia Nantke and Frederik Schlupkothen, Berlin, Boston: De Gruyter, 2020, pp. 177-200. DOI: https://doi.org/10.1515/9783110689112-009.

Krämer, Sybille (2023): “Should We Really ‘Hermeneutise’ the Digital Humanities? A Plea for the Epistemic Productivity of a ‘Cultural Technique of Flattening’ in the Humanities.” Journal of Cultural Analytics, vol. 7, no. 4, Jan. 2023, DOI: https://doi.org/10.22148/001c.55592.

Kuczera, Andreas (2022): “TEI Beyond XML – Digital Scholarly Editions as Provenance Knowledge Graphs.” In: Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera and Joris van Zundert (eds.): Graph Technologies in the Humanities - Proceedings 2020, published at http://ceur-ws.org/Vol-3110.

Kuczera, Andreas (2024): “Applied Text as Graph (ATAG)”. DHd 2024 Quo Vadis (DHd2024), Zenodo. DOI: https://doi.org/10.5281/zenodo.10698323

Pagel, Janis / Reiter, Nils / Rösiger, Ina / Schulz, Sarah (2020): Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt, edited by Nils Reiter, Axel Pichler and Jonas Kuhn, Berlin, Boston: De Gruyter, 2020, pp. 125-142. DOI: https://doi.org/10.1515/9783110693973-006.

Phelps, Thomas A. / Wilensky, Robert (2000): “Robust Hyperlinks Cost Just Five Words Each.” *UCB/CSD-00-1091*, Technical Report. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2000/5442.html (Accessed: April 23, 2025).

Ralle, Inga Hanna (2016): "Maschinenlesbar – menschenlesbar: Über die grundlegende Ausrichtung der Edition" Editio, vol. 30, no. 1, 2016, pp. 144-156. DOI: https://doi.org/10.1515/editio-2016-0009.

Reiter, Nils / Pichler, Axel / Kuhn, Jonas (2020): Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt, Berlin, Boston: De Gruyter, 2020. DOI: https://doi.org/10.1515/9783110693973.

Sahle, Patrick (2017): "Digitale Edition." In: Digital Humanities: Eine Einführung, edited by Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 234–249. Stuttgart: J.B. Metzler. DOI: https://doi.org/10.1007/978-3-476-05446-3_17.

Stäcker, Thomas (2020): ">A Digital Edition is Not Visible<--Some Thoughts on the Nature and Persistence of Digital Editions." *ZfdG - Zeitschrift für digitale Geisteswissenschaften*, Nr. 5. Forschungsverbund Marbach Weimar Wolfenbüttel. DOI: https://doi.org/10.26083/tuprints-00019469.

Stronks, Els / Boot, Peter (2007): "Learned Love. Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)." Data Archiving and Networked Services (DANS).

Tiepmar, Jochen / Heyer, Gerhard (2017): "An Overview of Canonical Text Services," Linguistics and Literature Studies, Vol. 5, No. 2, pp. 132 - 148, 2017. DOI: https://doi.org/10.13189/lls.2017.050209.

Witt, Jeffrey C. (2018): “Digital Scholarly Editions and API Consuming Applications“. *Digital Scholarly Editions as Interfaces* 12: 219–47. ISBN: 9783748109259.