Communication Papers of the 20th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 63–70
DOI: 10.15439/2025F2208
ISSN 2300-5963 ACSIS, Vol. 45

Ontological support for integration computer tools in digital humanities research

Iwona Grabska-Gradzińska
0000-0002-5799-5438
Institute of Applied Computer Science,
Faculty of Physics, Astronomy
and Applied Computer Science,
Jagiellonian University
ul. Łojasiewicza 11, 30-348 Kraków, Poland
Email: iwona.grabska@uj.edu.pl

Barbara Strug, Grażyna Ślusarczyk
0000-0002-2204-507X
0000-0003-1032-1644
Institute of Applied Computer Science,
Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University
ul. Łojasiewicza 11, 30-348 Kraków, Poland
Email: {barbara.strug,grazyna.slusarczyk}@uj.edu.pl

***Abstract*—As knowledge embedded in literary texts is often expressed in informal and implicit ways, in recent years the field of digital humanities has witnessed the emergence of many tools designed to support text analysis and creation of digital editions. However, these tools operate on heterogeneous, differently structured data coming from various sources. The integration of existing tools, which would allow for creating an effective system dedicated to international cooperation in the field of literary research, can be supported by the ontological approach. Therefore in this paper the concept of the ontology dedicated for reasoning about literary text properties, is presented. The formal representation of knowledge embedded in text together with its syntactic and semantic schema facilitates heterogeneous data integration and helps to bridge the semantic gap across various editorial projects.**

I. Introduction

LITERARY works, especially historical texts, whether preserved in print or manuscript form, are cultural artifacts, bearing witness to the development of literary expression, editorial practice, and printing craftsmanship. They posses varying degrees of aesthetic and stylistic merit and serve as a reflection of their epochs, encapsulating the intellectual climate and the needs of its contemporary readership[1]. Moreover, such texts are repositories of a wide range of data that are relevant not only to literary studies but also to historical and archival research [2].

In recent years, the field of digital humanities has witnessed the consolidation of data and metadata description standards, alongside the emergence of a wide array of tools designed to support the creation of digital editions[3]. Among these, the Text Encoding Initiative (TEI) [4], an XML-based markup language, has become the de facto standard for encoding source texts in scholarly editions. TEI enables the detailed representation of textual structure, editorial interventions, and semantic annotation, making it particularly suitable for the complex requirements of philological and historical scholarship [5].

Complementing TEI, the Dublin Core metadata standard serves as a widely adopted framework for the description of bibliographic and cataloging metadata. Its simplicity and interoperability have made it a preferred choice in diverse digital contexts, ranging from library systems and digital repositories to content management platforms. Importantly, Dublin Core plays a crucial role in computational ontologies and the semantic web: its metadata elements are formally defined as RDF properties, allowing for their integration into knowledge graphs, linked data infrastructures, and ontology-driven information systems. As such, Dublin Core facilitates machine-readable, semantically enriched descriptions of resources that can be processed, queried, and related across heterogeneous data environments.

However, the above mentioned tools operate on structured or unstructured data coming from different sources and organized in different ways. The integration of existing tools would allow for creating an effective system dedicated to international cooperation in the field of literary research[6]. Such integration can be supported by the ontological approach, as ontologies are well known to be an effective way of tackling the problem of interoperability among data [7].

In this paper the concept of the ontology for digital text edition, which facilitates reasoning about literary text properties, is presented. The formal representation of knowledge embedded in text together with its syntactic and semantic schema facilitates heterogeneous data integration and helps to bridge the semantic gap across various editorial projects. Moreover it gives the possibility of hiding the technical aspects of defining SPARQL queries behind an ontology-based user interface.

II. Computer tools in text analysis

As knowledge conveyed in historical and literary texts is often expressed in informal and implicit ways it requires readers and/or editors to engage in processes of inference, association, allusion recognition, and narrative reconstruction. There are many tools available to support such endeavours and facilitate text analysis and specialized research.

A wide range of tools has been developed to leverage the aforementioned standards, supporting various stages of the digital editorial workflow. These tools facilitate the preparation of critical editions, the indexing and semantic search of metadata, the alignment of descriptive metadata with encoded

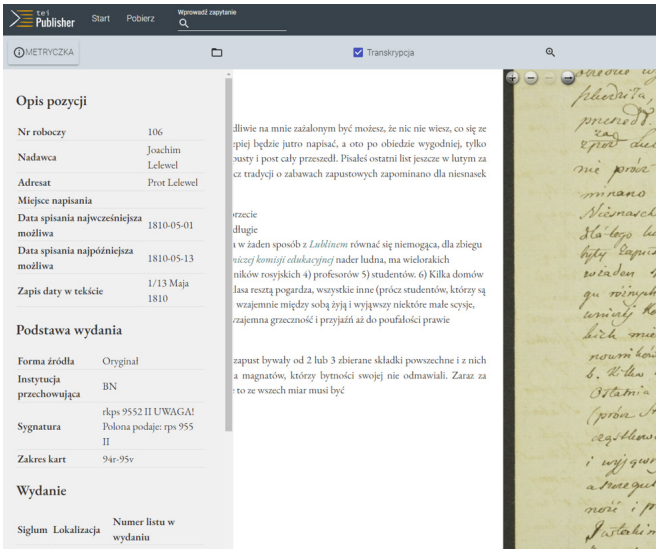


Fig. 1. Digital edition featuring metadata (left), diplomatic transcription (center), and facsimile (right), rendered using TEI Publisher.

Fig. 2. The process of converting a `.docx` document into TEI XML, enabling structured semantic encoding in accordance with the Text Encoding Initiative guidelines.

textual content, as well as inferencing and knowledge extraction based on the structured representations of texts.

For instance, TEI-aware editing environments such as oXygen XML Editor allow for sophisticated markup and validation workflows, while platforms like TEI Publisher enable the dynamic presentation of TEI-encoded editions [8]. In Fig. 1 a digital edition of a document featuring its metadata, diplomatic transcription, and facsimile in TEI Publisher is shown.

A. Annotation of data-rich texts

Documents or corpora that include elements suitable for annotation, or clusters of related documents sharing interconnected data, are called data-rich texts. Each document contains data that can be interpreted, encoded, and annotated. However, the knowledge embedded in literary texts is not readily accessible as it is typically unstructured and implicit, making its extraction and interpretation a complex process. A text containing numerous named entities like people, places, organizations, events, or objects, can be seen as data-rich, provided these entities are identified and annotated. Only after such annotation text can be treated as data in the sense of being searchable, processable, and inferable.

Fig. 2 shows a tool for converting nonstructured DOCX file into TEI-tagged format. Being a part of TEI Publisher functionality, it enables the editor to tag selected text fragments using various colours for different types of concepts.

Fig. 3 illustrates functionality of TEI Publisher that supports semi-automated semantic annotation of texts. This feature is particularly useful when working with documents already converted to XML but lacking comprehensive markup for named entities such as persons, places, organizations, and similar categories. The tool enables editors to manually identify the first occurrence of a particular entity in the text. Based on this input, the system suggests subsequent occurrences, which can then be reviewed, accepted, or rejected by the editor. This interactive process facilitates efficient and consistent entity tagging across large corpora, significantly reducing the manual workload while preserving scholarly control over annotation accuracy.

When multiple versions of a text are available (e.g. in the case of several distinct editions the comparison of which is mediated through the critical apparatus), the automation of metadata referencing become essential for optimizing editorial workflows. Therefore tools that enable the systematic identification of similarities and differences between textual variants are needed.

One such tool is LERA [9] (Locate, Explore, Retrace and Apprehend complex text variants), developed at the Institute of Computer Science, Martin Luther University Halle-Wittenberg, presented in Fig. 4. While LERA is not intended for direct use by end-users or general readers, it provides editors with robust functionalities for the analysis and preparation of critical apparatuses. The tool facilitates the alignment of variant passages, detection of editorial interventions, and generation of visual comparisons, thereby significantly enhancing the precision and efficiency of textual collation. The outputs produced by LERA can then be integrated into the scholarly edition and presented to readers as part of a structured, editorially curated critical apparatus.

A different situation arises when the textual versions in question are not variants of the same language text, but rather distinct manifestations of the same work—such as an original text and its translation. In such cases, the juxtaposition of text segments does not aim to uncover the stages of textual development or authorial revision, but rather to enable comparative analysis of aligned passages across linguistic boundaries. This form of alignment and visualization is especially common in digital editions of translated works, where paired segments

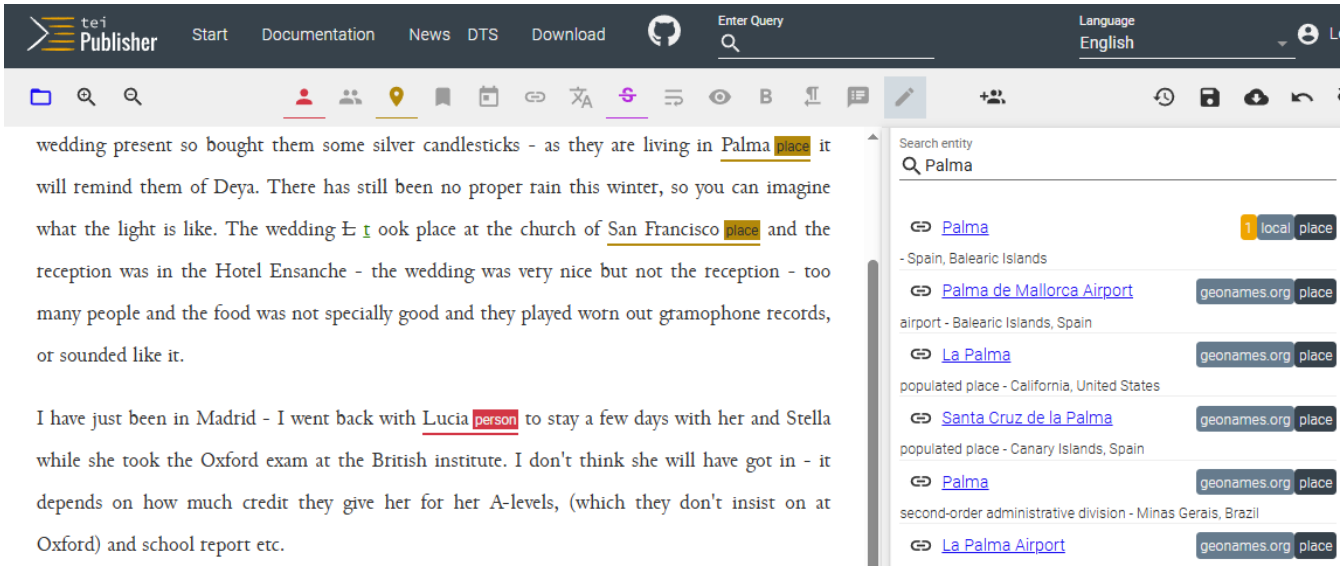


Fig. 3. *Semi-automated entity annotation in TEI Publisher.* The interface displays a TEI-encoded text undergoing semantic enrichment. The editor selects the first occurrence of an entity (e.g., a person, place, or organization), prompting the system to suggest subsequent instances for validation. On the right, external knowledge sources—such as Wikidata or GeoNames—provide contextual metadata and authoritative identifiers for the highlighted entity, aiding disambiguation and promoting interoperability with linked data frameworks.

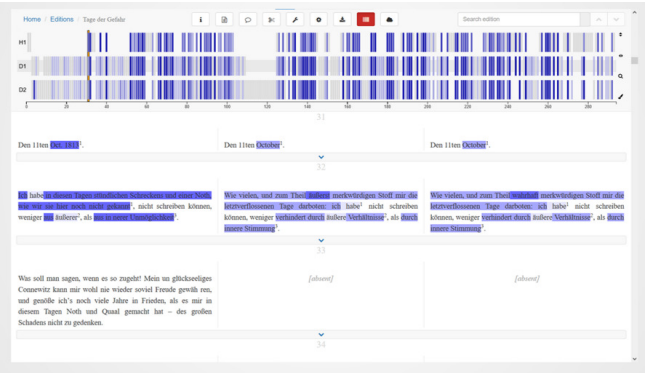


Fig. 4. LERA digital tool for inspecting similarities and differences between multiple versions of a text.

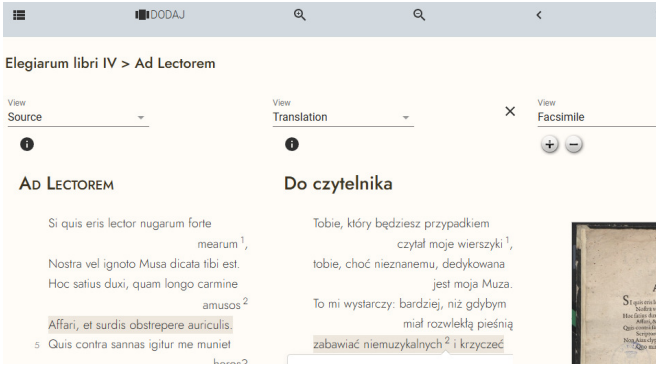


Fig. 5. TEI Publisher edition with interactive alignment of original and translated verses, along with dynamic display of footnotes.

are displayed in parallel to facilitate critical comparison, translation analysis, and reader comprehension.

Fig. 5 illustrates an example of such an edition, showcasing side-by-side rendering of an original text and its translation, structured to highlight the correspondence between semantically or syntactically related segments. This approach is particularly valuable in philological, literary, and translation studies, as it supports inquiries into translation strategies, interpretive choices, and textual equivalence.

B. Integration with external data services and metadata aggregation systems

The analysis and contextualization of texts are further enhanced by the availability of publicly accessible knowledge bases and data services that provide structured information across a wide range of domains. These include both state-maintained resources such as bibliographic and authority databases (e.g., the Integrated Authority File in Germany [10] or the National Library of Poland's catalogs [11]) and collaboratively curated general-purpose knowledge graphs, most notably Wikidata [12]. Additionally, there are domain-specific platforms that offer structured datasets focused on particular types of information or scholarly needs, such as GeoNames [13] for geographical entities, or Trismegistos [14], which aggregates metadata related to the ancient world.

By linking digital texts to these external repositories through unique identifiers and semantic relationships, researchers can enrich their editions with authoritative context, facilitate interoperability across systems, and enable advanced forms of cross-referencing, data integration, and computational analysis. These connections are foundational to the construction of a semantic ecosystem in the digital humanities, where texts

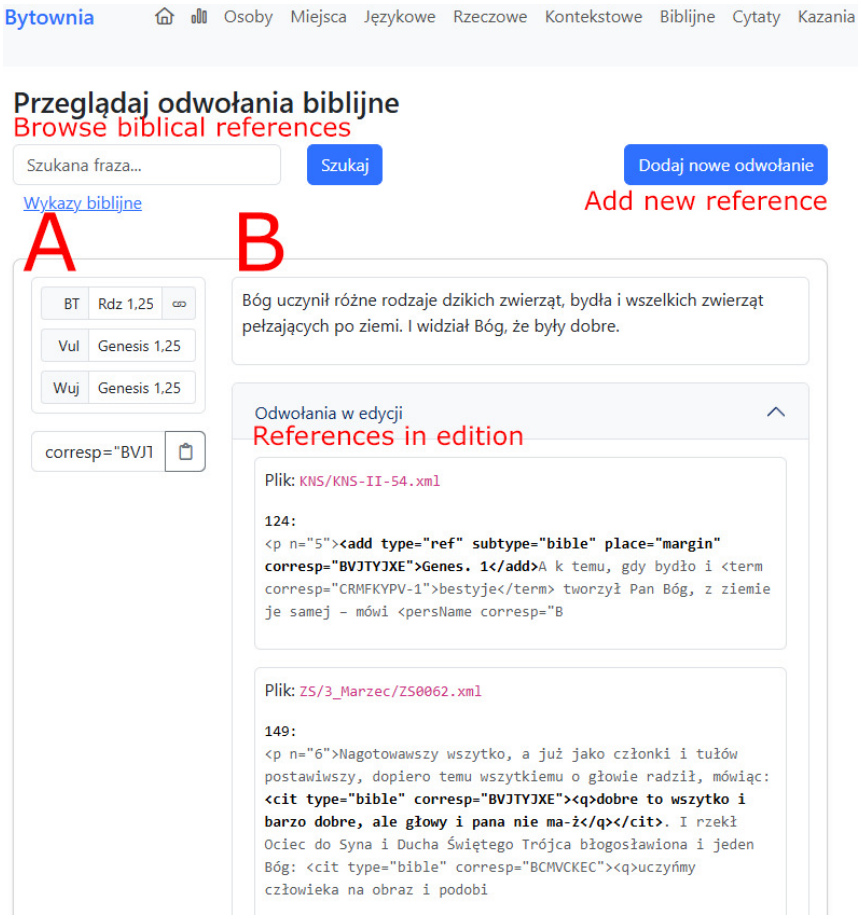


Fig. 6. Bytownia – the tool, which allows for tracing searched phrases in the text, is presented. On the left-hand side the searched biblical references are shown (A), while on the right-hand side the occurrences of these references within the analyzed text are marked (B).

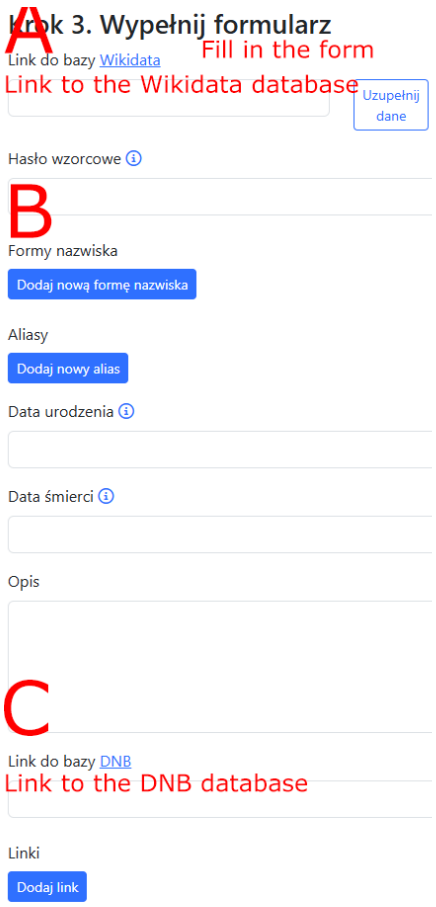


Fig. 7. Mechanism for synchronizing external knowledge services data (A, B)— enriching them with edition-specific aliases and notes (C).

are not isolated artifacts but interconnected components of a broader web of knowledge.

In Fig. 3 an additional aid for the editor in the form of the integration of external data sources (displayed on the right-hand side of the interface) can be seen. It provides encyclopedic or authority-based information about the selected entity. These linked data services enrich the annotation process by offering context from established knowledge bases, thereby supporting disambiguation and encouraging the use of persistent identifiers (e.g. Wikidata, GeoNames). This functionality not only supports the creation of richly encoded digital editions, but also fosters semantic interoperability with broader data ecosystems.

In the context of more specialized references and scholarly-oriented editions, it can be particularly beneficial to develop a service that aggregates data from publicly available external knowledge bases while incorporating detailed contextual information added during the editorial process.

Such a mechanism has been implemented in the emerging digital edition of the works of Piotr Skarga at the Jagiellonian University. The nature of Skarga's biblical references, which include direct quotations, marginal annotations, and allusions to biblical figures, necessitated the creation of a custom service designed to streamline the insertion and management of biblical citations. This tool enables editors to enrich the source text with canonical references (sigla) drawn from multiple biblical translations simultaneously. Moreover, it supports dynamic tracking of the function and rhetorical significance of each referenced passage within Skarga's homiletic practice.

The interface presented in Fig. 6 allows editors to view and manage a comprehensive list of already linked biblical references, facilitating an analytical perspective on intertextuality and the exegetical structure of Skarga's sermons. This approach exemplifies how domain-specific annotation services, integrated with external authoritative sources, can enhance both editorial precision and scholarly interpretation in digital critical editions.

This service not only enables the enrichment of textual objects by drawing data from selected external APIs (see Fig. 7), but it also exposes its own API, allowing the aggregated metadata to be reused across multiple editorial projects.

More importantly, by providing structured access to its inter-

nal data, the service facilitates the integration of the metadata repository with a computational ontology. This interoperability enables the implementation of reasoning tools that can operate over the entire body of accumulated knowledge. Through such semantic integration, the system supports advanced inferential functions, such as detecting implicit relationships, classifying textual phenomena, or identifying patterns in citation practices, thereby significantly enhancing the interpretive potential of the edition.

In this way, the service functions as both a dynamic editorial aid and a knowledge infrastructure, supporting not only the manual annotation process but also machine-assisted scholarly inquiry through semantic enrichment, linked data principles, and ontology-driven reasoning.

C. The heterogeneity of solutions adopted across individual editions

Print editors historically used typographic and layout conventions to signal textual structures and distinctions between content types. Contemporary digital editions take this further by employing markup systems to define the structure and semantics of texts in an unambiguous way. As a result, digital editions are not only readable artifacts but also structured repositories of knowledge organized according to standards that facilitate computational analysis and interoperability.

Knowledge encoding in textual corpora demands interpretive judgment, contextual awareness, and a nuanced understanding of textual semantics, making it inherently complex and variable. The conventions for formatting metadata and embedding semantic information within texts are themselves products of editorial practice. Over the years, as digital editions have evolved, these practices have developed organically through the cumulative experience of individual editors, resulting in a rich but highly heterogeneous ecosystem of encoding approaches. The diversity is particularly evident when comparing editions of varying genres and document types including poetry, prose, archival materials, historical records, and derivative or auxiliary texts.

Fortunately, despite this conceptual and methodological variety, the underlying data format in most digital editions is standardized through the TEI standard, based on XML, ensures a hierarchical and semantically rich data structure. This allows for a precise modeling of internal relationships within a text, preserving its logical organization while enabling efficient search, processing, and analytical operations using digital tools. The high granularity of TEI markup enhances editorial transparency and supports the comparison of variant textual versions across editions.

However, the implementation of TEI is far from monolithic. The manner in which individual elements and tags are defined and applied varies significantly between projects. These differences often stem not only from innovative or experimental approaches to the encoding process, but also from the intrinsic diversity of the source materials being edited. The needs of a critical edition of 17th-century poetry, for instance, differ considerably from those of a diplomatic transcription of archival correspondence or a scholarly apparatus for historical charters.

As a result, while the use of TEI fosters a degree of interoperability and standardization, the full standardization of semantic structures across digital editions remains a challenging and, to some extent, aspirational goal. It requires ongoing efforts in community coordination, the development of best practices, and the possible alignment with formal ontologies and controlled vocabularies that can help bridge semantic gaps across editorial projects.

As many knowledge relationships embedded in a corpus of source texts remain opaque to the reader, it is valuable to support text analysis with tools grounded in formal representations of knowledge. Viewing a digital edition as a knowledge base accessible via APIs allows external systems to query, extract, and infer information from the encoded content. Formal knowledge representation frameworks enable the imposition of an additional conceptual layer over the edition, allowing textual content to be interpreted in terms of concepts and their interrelations. This, in turn, facilitates the application of computational ontologies and graph-based knowledge systems.

III. Ontology-based data organization

Ontologies can be used to unambiguously describe elements such as textual variant types, individual roles, and historical-cultural contexts. When integrated with rule-based systems or inference engines, they enable the generation of new knowledge from existing data. Ultimately, this supports the creation of shared conceptual models that facilitate the comparison and integration of diverse digital editions.

Ontology-based data organization can be described as a methodological approach for structuring and managing information using ontology, that is formal specification of concepts and their interrelationships within a specified domain. This approach offers a semantically rich alternative to traditional data management systems by enabling machine-readable representations of knowledge and facilitating better data integration, retrieval, and analysis [15]. In contrast to taxonomies or thesauri, ontologies not only define hierarchical relationships but also allow for defining complex relations, adding specific constraints, and rules that reflect domain knowledge requirements.

In recent decades, the implementation of ontology-based systems has gained acceptance in various disciplines, including biomedical, chemical or geographic information systems, and more recently, the digital humanities. The ontological approach allows for supporting the creation of metadata standards and fostering cross-disciplinary data sharing. Important examples include the CIDOC Conceptual Reference Model (CIDOC-CRM), which provides a semantic framework for cultural heritage information [16], [17], and the Europeana Data Model (EDM), which integrates diverse museum and library collections across Europe.

In digital humanities, ontology-based data organization plays a critical role as there is a need for modeling complex

cultural and historical data. These domains often involve heterogeneous sources such as manuscripts, artifacts, maps, or even recorded oral documents, which often require specialized and contextualized interpretation. Ontologies can facilitate the semantic annotation and linking of these sources, enabling potential users to query datasets using concept-based rather than keyword-based search.

Moreover, ontology-driven tools support the development of digital editions, scientific annotation platforms, and visualizations that provide researchers with help in carrying out humanities research. By enabling options such as disambiguation, provenance tracking, or reasoning from cultural data, ontologies enhance the transparency and reproducibility of research work [18]. The reusability and sustainability of use is further supported by the fact that ontological models usually follow FAIR data principles (Findable, Accessible, Interoperable, Reusable) [19], [20]. The principles focus on so called machine-actionability, which means the capacity of computational systems to find, access, interoperate, and reuse data, and do it with none or minimal user intervention. As the users more and more depend on computational support to deal with data, especially large amount of data with complex interrelations, such an approach can support data accessibility for non-technical users.

Despite its advantages, mentioned above, ontology-based data organization poses some challenges, particularly in the digital humanities context. They include the need for domain-specific modeling expertise, and the problems of understanding between domain experts and data specialists. Another issue come from the possible conflict between formal logic, required by data modeling, and interpretive flexibility common in humanities research. In addition there is a problem of accepting common shared vocabularies among scholars with diverse theoretical backgrounds. Nevertheless, ongoing collaborative efforts and community-driven ontology development, such as the TEI [20] and the OntoMedia [21] ontology, continue to bridge these gaps. Thus ontology-based data organization presents an approach to managing and analyzing complex data in the digital humanities that can open new possibilities in data sharing and using. By adding rich semantic structures to raw data, such an approach can empower researchers to find new insights by following linked data and advance collaborative research by opening access to previously inaccessible sources.

IV. Ontology for digital text edition

Integration of data and metadata related to literary texts coming form different, often big and heterogeneous sources, like library or museum data bases and digital archived, requires harmonisation into a unified view. Such integration is facilitated by developing a domain-specific ontology.

Many methodologies for building ontologies have been proposed [22], [23]. Modern agile methodologies like eXtreme Design (XD) [24], Modular Ontology modeling (MOM) [25], Simplified Agile Methodology for Ontology Development (SAMOD) [26], are typically based on ontology requirements

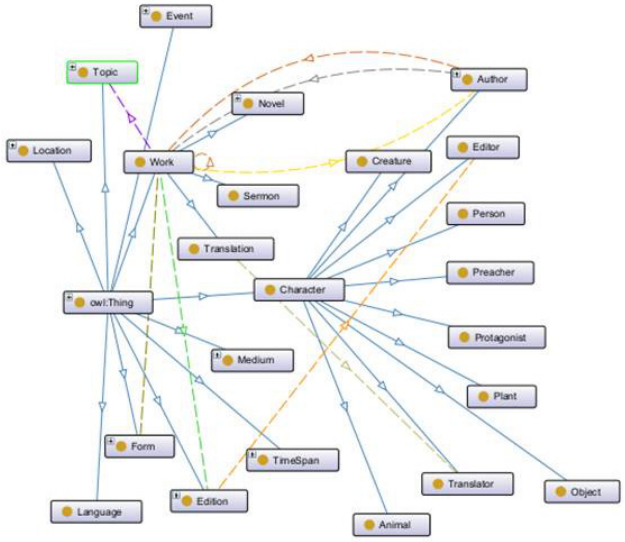


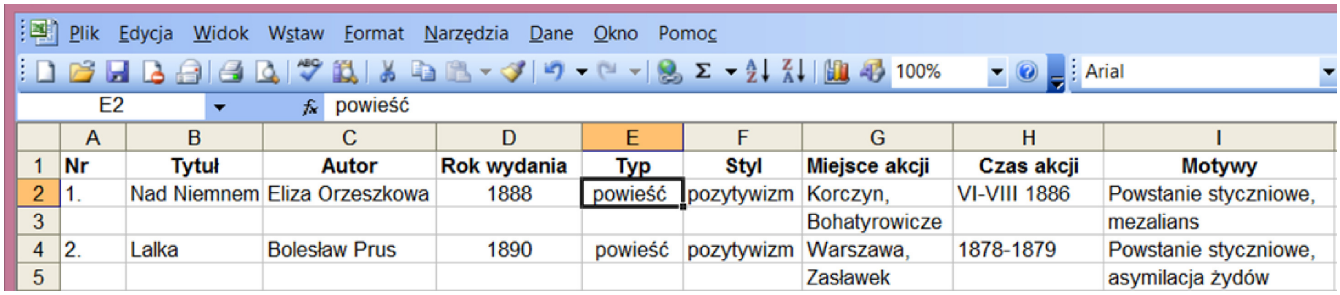
Fig. 8. A fragment of TEON ontology

or ontology design patterns. A common way to express ontology requirements, which specify the intended task of the resulting ontology, is through Competency Questions (CQ) [27] illustrating the typical information needs that one would require the ontology to respond to. These questions can be then formalized as queries over some test set of data Ain order to test the created ontology [28].

Our prototype ontology has been built upon a set of competency questions developed by philologist experts. An example of a competency question is "What literary motifs are discussed in the positivist novel?". These questions refer to data related to literary texts and constrain the scope of knowledge to be represented in an ontology. On this basis, classes, properties and relations of the Text Encoding Ontology (TEON) were defined. Such questions can be then used to test if the ontology contains enough information to answer them.

The most important classes available in the current version of the ontology are *Work*, *Character*, *Style, Topic, Location, TimeSpan, Edition, Form, Medium, Institution, Event, Language* and *Gender*. Class *Work* contains data related to a literary text, and has subclasses specifying work genre as *Novel, Letter, Essay, Poem, Sermon, Diary* etc. Class *Character* contains subclasses *Author, Editor, Person, Translator, Protagonist, Preacher Animal, Plant, Creature* and *Object*, which represent possible variants of characters related to works. For example the class *Author* provide access to information about the person who is the author of at least one work, while the class *Creature* describes a fantastic creature appearing in a literary work. In Fig. 8 a snapshot of the fragment of the ontology is depicted.

We assume that the input data concerning literary text are stored in the form of the csv files. A fragment of the file storing data about literature of positivism, converted to .xls format for clarity of presentation, is shown in Fig. 9. In this figure several



Plik Edycja Widok Wstaw Format Narzędzia Dane Okno Pomoc

E2 | powieść

	A	B	C	D	E	F	G	H	I
1	**Nr**	**Tytuł**	**Autor**	**Rok wydania**	**Typ**	**Styl**	**Miejsce akcji**	**Czas akcji**	**Motywy**
2	1.	Nad Niemnem	Eliza Orzeszkowa	1888	powieść	pozytywizm	Korczyn,	VI-VIII 1886	Powstanie styczniowe,
3							Bohatyrowicze		mezalians
4	2.	Lalka	Bolesław Prus	1890	powieść	pozytywizm	Warszawa,	1878-1879	Powstanie styczniowe,
5							Zasławek		asymilacja żydów

Fig. 9. A fragment of the csv test file containing information on works in the positivism style

entries related to work title, author, year of issue, type, style, place of action, time of action and motives can be seen.

These data are transformed into RDF triples using R2RML (Relational to Resource Description Framework Mapping Language) language [29], [30]. This language provides capabilities to connect the structure of the csv data to the ontology vocabulary. The obtained triples become instances which populate the ontology and form a data graph also referred to as a RDF triplestore. This store is then used as a data sources for SPARQL queries that support implementation of the competency questions. In Fig. 10 a data graph with individual instances coming from the entries shown in Fig. 9.

The considered triplestore supports SPARQL queries related to data concerning literary texts. In Fig. 11 a query related to authors who refer to the January Uprising (pol. Powstanie Styczniowe) in their works is presented. The result obtained using this query contains two author names: Eliza Orzeszkowa and Bolesław Prus.

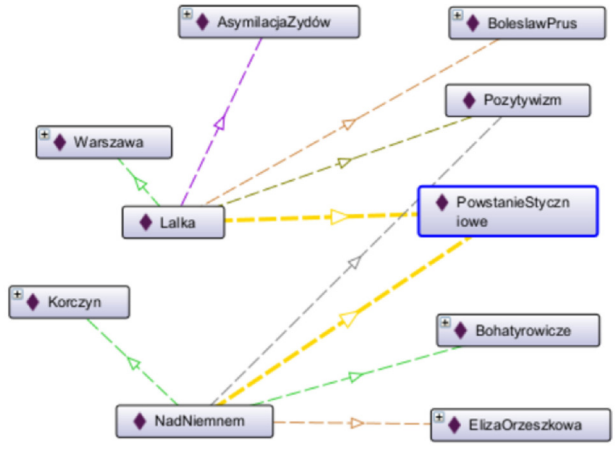


Fig. 10. A data graph with selected instances

```
PREFIX uj:
<http://www.semanticweb.org/al2020/ontologies/2025/2/ontoDigitalHum#>

SELECT ?autor WHERE {
  ?dzielo uj:refersToEvent uj:PowstanieStyczniowe .
  ?dzielo uj:hasAuthor ?autor .
}
```

Fig. 11. An example of a sparql query

V. Conclusion

This paper presents a current version of the ontology that will provide a semantic schema for the computer system making different resources accessible for researchers in different areas of humanities research.

The described methodology is a part of an ongoing research within the framework of the Jagiellonian University Flagship Programmes "Digital Humanities Lab" and "European Heritage in the Jagiellonian Library: Digital Authoring of the Berlin Collections". It aims at providing a shared ontology-based tool supporting the creation of a knowledge graph storing information from existing datasets in different, and usually incompatible and non-interoperable data formats, scattered among different places. It will also offer a unified way of querying available literary resources and their metadata.

References

[1] J. Gruchała, "Editing: a Knowledge and a Skill" (in polish). *Wielogłos*, no. 3 (13), 2012, pp. 157–164. DOI: 10.4467/2084395XWI.12.012.0867. ISSN 1897-1962.

[2] J. Gruchała, "The Virtual Publisher and the User of a Digital Edition" (in polish). In: Elżbieta Wichrowska (ed.), *The European Literary Canon*, pp. 282–288. ISBN 978-83-235-0834-2, 2012.

[3] E.F. Cavanaugh and J.E. Stertzer, "Building Accessibility: Platforms and Methods for the Development of Digital Editions and Projects", in *Digital Editing and Publishing in the Twenty-First Century*, eds. J. O'Sullivan, M. Pidd, S. Whittle, B. Wessels, M. Kurzmeier, and Ó. Murphy, Scottish Universities Press, 2025. DOI: 10.62637/sup.GHST9020.5. ISBN 978-1-917341-04-2.

[4] TEI Consortium. "TEI P5: Guidelines for Electronic Text Encoding and Interchange". Version 4.9.0, last updated 24 January 2025. https://www.tei-c.org/Guidelines/P5/

[5] E. Spadini and J.L. Palenzuela, "Re-using Data from Editions", in *Digital Editing and Publishing in the Twenty-First Century*, eds. J. O'Sullivan, M. Pidd, S. Whittle, B. Wessels, M. Kurzmeier, and Ó. Murphy, Scottish Universities Press, 2025. DOI: 10.62637/sup.GHST9020.8. ISBN 978-1-917341-04-2.

[6] G. Franzini, M. Kestemont, G. Rotari, M.Jander, J. Ochab, E. Franzini, J. Byszuk, and J. Rybicki. "Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm". *Frontiers in Digital Humanities*, vol. 5, 2018. DOI: 10.3389/fdigh.2018.00004.

[7] R. Kishore, R. Sharman and R. Ramesh (2004). "Computational Ontologies and Information Systems I: Foundations", Communications of the Association for Information Systems. 14. 158-183. 10.17705/1CAIS.01408.

[8] TEI Publisher. "An environment for publishing TEI documents". Version 9.0.9, maintained by the eXist Solutions team, 2025. https://teipublisher.com

[9] M. Pöckelmann, A. Medek, J. Ritter, and P. Molitor. "LERA - An interactive platform for synoptical representations of multiple text witnesses" In: *Digital Scholarship in the Humanities (DSH)*. Oxford University Press 2022. DOI: 10.1093/llc/fqac021

[10] German National Library. *Gemeinsame Normdatei (GND)*. Available at: https://www.dnb.de/EN/gnd [Accessed: 20.05.2025.

[11] Biblioteka Narodowa. *National Library of Poland's catalogues*. Available at: https://katalogi.bn.org.pl/ [Accessed: 20.05.2025].

[12] Wikimedia Foundation. *Wikidata: A free knowledge base that anyone can edit*. Available at: https://www.wikidata.org/ [Accessed: 20.05.2025].

[13] GeoNames. *GeoNames geographical database*. Available at: https://www.geonames.org/ [Accessed: 20.05.2025].

[14] Trismegistos. *Trismegistos: An interdisciplinary portal of the ancient world*. Available at: https://www.trismegistos.org/ [Accessed: 20.05.2025].

[15] T.R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, 5, pp 199—220, 1993. Doi: 10.1006/knac.1993.100

[16] M. Doerr, "The CIDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability of Metadata," *AI Magazine*, 24, pp 75–92, 2003. DOI: 10.1609/aimag.v24i3.1720

[17] CIDOC-CIDOC "Conceptual Reference Mode", https://cidoc-crm.org/

[18] S. Schreibman, R. Siemens, and J. Unsworth, (Eds.), *A New Companion to Digital Humanities*, John Wiley and Sons; 2016. DOI:10.1002/9781118680605

[19] M. Wilkinson, M. Dumontier, I. Aalbersberg, et al., "The FAIR Guiding Principles for scientific data management and stewardship", *Sci Data* 3, 160018, 2016. DOI: 10.1038/sdata.2016.18

[20] J. Tello, M. Göbel, U. Veentjer, S. Funk, N. Rißler-Pipka, and K. Du, . "FAIR Derived Data in TEI and Its Publication in the TextGrid Repository," *Journal of the Text Encoding Initiative* 18, 2024 DOI: 10.4000/13llz.

[21] F. Lawrence, M. Tuffield, M. Jewell, A. Prugel-Bennett, D. Millard, M. Nixon, M.C. Schraefel, and N. Shadbolt, "OntoMedia - Creating an Ontology for Marking Up the Contents of Heterogeneous Media", Ontology Patterns for the Semantic Web ISWC-05 Workshop, Galway, Ireland, 2005.

[22] N. Noy and D. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," *Knowledge Systems Laboratory,* 32, 2001.

[23] H.S. Pinto, C. Tempich and S. Staab, "Ontology Engineering and Evolution in a Distributed World Using DILIGENT," *In: Staab, S., Studer, R. (eds) Handbook on Ontologies. International Handbooks on Information Systems,* Springer, Berlin, Heidelberg, pp. 153–176, 2009. DOI:10.1007/978-3-540-92673-3_7

[24] E. Blomqvist, K. Hammar and V. Presutti, "Engineering ontologies with patterns— the eXtreme design methodology," In: Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, and Valentina Presutti, Eds., *Ontology Engineering with Ontology Design Patterns,* vol. 25 of Studies on the Semantic Web. IOS Press, 2016. DOI: 10.3233/978-1-61499-676-7-23

[25] P. Hitzler and A. Krisnadhi "A Tutorial on Modular Ontology Modeling with Ontology Design Patterns: The Cooking Recipes Ontology," CoRR, 2018. DOI:10.48550/arXiv.1808.08433

[26] S. Peroni, "A Simplified Agile Methodology for Ontology Development," *In: Dragoni, M., Poveda-Villalón, M., Jimenez-Ruiz, E. (eds) OWL: Experiences and Directions – Reasoner Evaluation. OWLED ORE 2016*, Lecture Notes in Computer Science, vol. 10161, 2017. Doi:10.1007/978-3-319-54627-8

[27] M. Grüninger and M.S. Fox, "The role of competency questions in enterprise engineering," *In Asbjorn Rolstadas, Ed., Benchmarking—Theory and Practice,* Springer, pp. 22–31, 1995. DOI:10.1007/978-0-387-34847-6_3

[28] C.M. Keet and A. Ławrynowicz, "Test-driven development of ontologies," In Harald Sack, Eva Blomqvist, Mathieu d'Aquin, Chiara Ghidini, Simone Paolo Ponzetto, and Christoph Lange, Eds., *The Semantic Web. Latest Advances and New Domains—13th International Conference,* ESWC, Heraklion, Crete, LNCS, vol. 9678, pp. 642–657, 2016, DOI: 10.48550/arXiv.1512.06211

[29] J. F. Sequeda, "On the Semantics of R2RML and its Relationship with the Direct Mapping," *In ISWC (Posters and Demos)*, pp. 193–196, 2013.,

[30] M. Rodriguez-Muro and M. Rezk, "Efficient SPARQL-to-SQL with R2RML mappings," *Journal of Web Semantics*, 33, pp. 141–169, 2015, DOI: 10.1016/j.websem.2015.03.001.