Semantic Web 14 (2023) 155–158
DOI 10.3233/SW-223187
IOS Press

Editorial of the Special issue on Cultural heritage and semantic web

Mehwish Alam [a,b,*], Victor de Boer [c], Enrico Daga [d], Marieke van Erp [e], Eero Hyvönen [f] and Albert Meroño-Peñuela [g]

[a] *FIZ-Karlsruhe, Leibniz Institute for Information Infrastructure, Germany*
[b] *Karlsruhe Institute of Technology, Germany*
E-mail: mehwish.alam@kit.edu
[c] *Department of Computer Science, Vrije Universiteit Amsterdam, the Netherlands*
E-mail: v.de.boer@vu.nl
[d] *Knowledge Media Institute, STEM Faculty, The Open University, United Kingdom*
E-mail: enrico.daga@open.ac.uk
[e] *DHLab, KNAW Humanities Cluster, the Netherlands*
E-mail: marieke.van.erp@dh.huc.knaw.nl
[f] *Semantic Computing Research Group (SeCo), Aalto University and University of Helsinki (HELDIG Centre), Finland*
E-mail: eero.hyvonen@aalto.fi
[g] *Department of Informatics, King's College London, United Kingdom*
E-mail: albert.merono@kcl.ac.uk

1. Preface

The increasing accessibility and affordability of services and computational resources to collect, enhance, analyze and republish data continues to impact and increase the structuring of many fields of scholarship. With more Cultural Heritage (CH) and humanities datasets available than ever, Digital Humanities (DH) brings together an exciting blend of researchers and practitioners from different disciplines, such as social sciences, arts, humanities, cultural heritage, library science, and computer science. This highly interdisciplinary community brings promising visions of our digital pasts and futures, but also humanistic and cultural concerns about their use. Data management, data ethics, data curation, data cleaning, data provenance, data integration, and semantics are prevalent in DH from idiosyncratic and varied perspectives. Semantic technologies have shown, in various venues through the last decade, a key role and deep penetration in cultural heritage and DH workflows through unique methods, adoption stories, and contributions to a harmonic ecosystem for Semantic data-intensive technologies. On the one hand, semantic technologies have been effective at addressing challenges and research questions from humanities scholars, such as working with data of limited scale, vague and yet valuable evidence, and the need for richer contexts. On the other hand, humanities scholars are continuously given a new technological landscape to reflect on and develop new thoughts, closing a virtuous circle.

This circle has unlocked knowledge that would have remained unknown otherwise, such as extensive human-driven analysis of semantically rich cultural datasets; large knowledge modeling efforts that have crystalized into ontologies, controlled vocabularies, and conceptual models such as CIDOC-CRM, the Europeana Data Model, and FRBRoo; and numerous 5-star Linked Data datasets that populate the Linked Open Data cloud today as a

*Corresponding author. E-mail: mehwish.alam@kit.edu.

category in its own right. With the recent advances and quickly transforming the landscape in Knowledge Graphs, property graphs, deep learning, automated knowledge base construction, language models, computer vision, and multimodality, in our opinion that the paths of semantics and CH/DH, and to a broader extent those of Cultural AI, have only started crossing.

This special issue focuses on contributions along the lines of how semantic web technologies are fulfilling the needs of cultural heritage practitioners and organizations. Notably, it has welcomed submissions on open research problems related to cultural heritage preservation, valorization, engagement, and ethics for which Semantic Web technologies could provide a viable approach. Moreover, it has welcomed research efforts in collaboration with archivists, historians, curators, philologists, cultural critics, musicologists, and other humanists that generally deal with information that is subjective, vague, fragmentary, uncertain, and contradictory; and yet still provides valuable evidence that poses a challenge to knowledge bases and even to Artificial Intelligence research as a whole.

2. Topics of interest

The issue solicited contributions specifically around the topics of:

- Cultural Heritage Ontologies and Vocabularies;
- Semantic technologies for Digital Humanities;
- Linked Data, Knowledge Graphs, and Digital Humanities;
- Semantic and Exploratory Search for Humanities and Cultural Heritage;
- Geo-semantics for Humanities and Digital Heritage;
- Knowledge Driven NLP for Digital Humanities;
- Supporting Humanities scholars accessing Semantic data;
- Effects of knowledge/ontology engineering in CH/DH scholarship;
- Social Semantics for Humanities data;
- Semantics for Audio-Visual Data;
- Semantics for Ethics, norms, and regulatory compliance in the Heritage Sector;
- Semantic Digital Rights and Privacy in Cultural Heritage;
- Historical Entity reconciliation on the Semantic Web;
- Semantics and Hermeneutics;
- Semantic Social Networks in Heritage data;
- Semantics in Digital Libraries;
- Machine Learning for Knowledge Graphs in Digital Humanities;
- Visual Intelligence and Semantics in Digital Heritage;
- Multimodal Processing and Cultural Heritage;
- Semantics for Intangible Heritage.

3. Content

The special issue attracted 20 submissions covering our key areas of research, i.e., semantic web technologies and cultural heritage. Out of these, 12 papers were accepted after two rounds of reviews, indicating an acceptance rate of 60%. Each paper was reviewed by at least 3 expert reviewers. In the following, we will provide a broad overview of all the accepted papers.

- The first paper entitled “Generation of Training Data for Named Entity Recognition of Artworks” by Nitisha Jain, Alejandro Sierra, Jan Ehmueller, and Ralf Krestel creates training datasets for named entity recognition (NER) in the context of the cultural heritage domain. NER plays an essential role in many natural language processing systems. The authors present a framework with a heuristic-based approach to creating high-quality training data by leveraging existing cultural heritage resources from knowledge bases such as Wikidata.

- In the second paper "Move Cultural Heritage Knowledge Graphs in Everyone's Pocket" by Maria Angela Pellegrino, Vittorio Scarano, and Carmine Spagnuolo, the authors present a domain-agnostic Knowledge Graph exploitation approach based on virtual assistants as they naturally enable question-answering features where users formulate questions in natural language directly by their smartphones. The authors discuss the design and implementation of the proposed approach within an automatic community-shared software framework (a.k.a. generator) of virtual assistant extensions and its evaluation on a standard benchmark of question-answering systems. Finally, according to a taxonomy of the Cultural Heritage field, the authors present a use case for each category to show the applicability of the proposed approach in the Cultural Heritage domain.
- The third paper with the title "Transdisciplinary approach to archaeological investigations in a Semantic Web perspective" by Vincenzo Lombardo, Tugce Karatas, Monica Gulmini, Laura Guidorzi, and Debora Angelici presents a modular computational ontology for the interlinked representation of all the facts related to the archaeological and archaeometric analyses and interpretations, also connected to the recording catalogues. The computational ontology is compliant with CIDOC-CRM reference models CRMarchaeo and CRMsci. It introduces several novel classes and properties to merge the two worlds in a joint representation. The ontology is used in "Beyond Archaeology", a methodological project for establishing a transdisciplinary approach to archaeology and archaeometry, interlinked through a semantic model of processes and objects.
- The fourth paper on "Question Answering with Deep Neural Networks for Semi-Structured Heterogeneous Genealogical Knowledge Graphs" by Omri Suissa, Maayan Zhitomirsky-Geffet, and Avshalom Elmalech proposes an end-to-end approach for question answering using genealogical family trees by representing genealogical data as knowledge graphs, converting them to texts, combining them with unstructured texts, and training a transformer-based question answering model. A comparison between the fine-tuned model (Uncle-BERT) trained on the auto-generated genealogical dataset and state-of-the-art question-answering models was performed to evaluate the need for a dedicated approach.
- The fifth paper entitled "RelTopic: A Graph-Based Semantic Relatedness Measure in Topic Ontologies and Its Applicability for Topic Labeling of Old Press Articles" was written by Mirna El Ghosh, Nicolas Delestre, Jean-Philippe Kotowicz, Cecilia Zanni-Merk, and Habib Abdulrab. The paper describes an approach called RelTopic, which considers the semantic properties of entities in ontologies. Thus, correlations of nodes and weights of nodes and edges are assessed. The pertinence of RelTopic is evaluated for the topic labeling of old press articles. For this purpose, a topic ontology representing the articles, named Topic-OPA, is derived from open knowledge graphs by applying a SPARQL-based automatic approach. A use-case is presented in the context of the old French newspaper Le Matin. The generated topics are evaluated using a dual evaluation approach with the help of human annotators.
- The sixth paper with the title "Linking Women Editors of Periodicals to the Wikidata Knowledge Graph" was written by Katherine Thornton, Kenneth Seals-Nutt, Marianne Van Renmoortel, Julie M. Birkholz, and Pieterjan De Potter. The paper details how a machine, through the power of the Semantic Web, can compile scattered and diverse materials and information to construct stories. Through the example of the WeChangEd research project on women editors of periodicals in Europe from 1710–1920, the authors detail how to move from an isolated archive to a structured data model linked to Wikidata. This is leveraged by a Stories Services API that generates multimedia stories related to people, organizations, and periodicals. As more humanists, social scientists and other researchers choose to contribute their data to Wikidata, further projects can benefit from this approach.
- The seventh paper entitled "A Shape Expression approach for assessing the quality of Linked Open Data in Libraries" by Gustavo Candela, Pilar Escobar, María Dolores Sáez, and Manuel Marco-Such proposes a methodology to create Shape Expressions definitions to validate LOD datasets published by libraries. The methodology was then applied to four use cases based on datasets published by relevant institutions. It intends to encourage institutions to use ShEx to validate LOD datasets and promote the reuse of LOD, made openly available by libraries.
- The eighth paper "Typed properties and negative typed properties: dealing with type observations and negative statements in the CIDOC CRM" by Athanasios Velios, Carlo Meghini, Martin Doerr, and Stephen Stead discusses techniques for expressing such observations within the context of the CIDOC CRM in both OWL and RDFS are explored. OWL cardinality restrictions are considered and new special properties deriving from the

CIDOC CRM are proposed, namely 'typed properties' and 'negative typed properties' which allow stating the types of multiple individuals and the absence of individuals. The nature of these properties is then explored in relation to their correspondence to longer property paths, their hierarchical arrangement, and relevance to thesauri.

- The ninth paper is a survey article on "Semantic models and services for conservation and restoration of cultural heritage: a comprehensive survey" by Efthymia Moraitou, Yannis Christodoulou, and George Caridakis, discusses semantic models relevant to the CnR knowledge domain (as the name indicates). The scope, development methodology, and coverage of CnR aspects are described and discussed. Furthermore, the evaluation, deployment, and current exploitation of each model are examined, with a focus on the types and variety of services provided to support the CnR professional.
- The tenth paper entitled "Analyzing Biography Collections Historiographically as Linked Data: Case National Biography of Finland" by Minna Tamper, Petri Leskinen, Eero Hyvönen, Risto Valjus, and Kirsi Keravuori, focuses on analyzing biographical collections available as Linked Open Data. The National Biography data, extracted from a textual repository of biographies, was available as part of the Linked Open Data service and semantic portal "BiographySampo – Finnish Biographies on the Semantic Web" in Finland. The paper presents various results related to, e.g., how specific prosopographical groups, such as women or professional groups, are represented and portrayed in the collection, using, e.g., statistics and network analyses of the biographees and the texts. The analyses are based on using the semantic portal interface and Google Colab and Python scripting on top of the SPARQL endpoint. The presented approach can also be applied to similar biography collections in other countries.
- The tool & system report "Casual Learn: A Linked Data-Based Mobile Application for Learning about Local Cultural Heritage" was written by Adolfo Ruiz-Calleja, Pablo García-Zarza, Guillermo Vega-Gorgojo, Miguel L. Bote-Lorenzo, Eduardo Gómez-Sánchez, Juan I. Asensio-Pérez, Sergio Serrano-Iglesias, and Alejandra Martínez-Monés. The paper presents *Casual Learn*, an application that proposes ubiquitous learning tasks about Cultural Heritage. Casual Learn leverages a dataset of 10,000 contextualized learning tasks that were semi-automatically generated out of open data from the Web. Casual Learn offers these tasks to learners according to their physical location. For example, it may suggest describing the characteristics of the Gothic style when passing by a Gothic Cathedral. Additionally, Casual Learn has an interactive mode where learners can geo-search available tasks.
- The application report "Linked Open Images: Visual Similarity for the Semantic Web" by Lukas Klic presents ArtVision, a Semantic Web application that integrates computer vision APIs with the ResearchSpace platform, allowing for the matching of similar artworks and photographs across cultural heritage image collections. Using the images and artwork data of Pharos collections, this paper outlines the methodologies used to integrate visual similarity data from a number of computer vision APIs, allowing users to discover similar artworks and generate canonical URIs for each artwork.

This issue presents a broad view of current state-of-the-art research on the topic of Cultural Heritage and Semantic Web technology. We would like to thank the authors for submitting their research work on diverse and interesting topics. We would like to express our gratitude to the many reviewers for their commitment to ensuring a rigorous and open review process and thereby making this special issue a success.