A Geographical Extension for NOnt Ontology (short paper)

Nicolò Pratelli[1,2]

[1]Institute of Information Science and Technologies "Alessandro Faedo", 56124 Pisa, Italy [2]Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

Abstract

Digital Libraries (DLs) are full of narratives. Besides its richness, DLs services often retrieve narrative components, but not the narratives as a whole. To formally represent this knowledge and to create and visualize narratives the Digital Humanities group of ISTI-CNR has developed the Narrative Ontology (NOnt) and the Narrative Building and Visualising Tool (NBVT). In this context, my research aims to investigate the possibility to introduce the geospatial dimension of narratives in NOnt. Moreover, my research aims to extend the functionalities of NBVT to enrich the narratives with geospatial information. As a case study, I have chosen to create narratives about mountain ecosystems and economic value chains produced within the Mountain Valorization through Interconnectedness and Green Growth (MOVING) European project (2020-2023). Currently, my research is still at an early stage and I have started to conduct a state-of-the-art study of the geospatial and spatiotemporal RDF/S representation techniques. Eventually, I will evaluate the extension of NOnt.

Keywords

Narrative, Digital Libraries, Semantic Web, Ontology, Linked Open Data, Geographical Knowledge, Story Maps

Tutor(s)

Dr. Ing. Nicola Tonellotto, Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

Dr. Valentina Bartalesi, Institute of Information Science and Technologies "Alessandro Faedo", 56124 Pisa, Italy

1. Introduction

Narratives are essential structures to document every aspect of human activity either cultural or scientific or social domain [1]. As narratives, I mean semantic networks of events that are linked to the objects of the DLs and are endowed with a set of semantic relations that connect one event to another. Currently, Digital Libraries (DLs) functionalities retrieve the digital object contained in them, but they can not retrieve semantic relations among them. Thus, DLs, do not allow to build a narrative, using their digital object. Trying to overcome some issues of narratives in DLs, the Narrative Ontology (NOnt) [2] was introduced. It has been created by the Digital Humanities group of ISTI-CNR as an extension of three standard vocabularies, i.e. the

TPDL2022: 26th International Conference on Theory and Practice of Digital Libraries, 20-23 September 2022, Padua, Italy
nicolo.pratelli@isti.cnr.it (N. Pratelli)
0000-0003-0364-922X (N. Pratelli)

CIDOC-CRM [3], FRBRoo [4], and OWL Time [4], and using the SWRL rule language [5] to express the axioms.

Indeed, a core element of narratives is the concept of the event, the spatio-temporal occurrence of a phenomenon. The concept of the event is also a key element of the Event Calculus theory (EC) [6, 7, 8], a logic language for representing actions that have a temporal duration and can overlap with each other. NOnt ontology is deeply focused on the temporal dimension of the event, peripherally considering the spatial dimension.

In this context, my research aims to enrich the formal representation of spatio-temporal dimension in narratives using Semantic Web Technologies (i.e. RDF/S, OWL, SPARQL) [9] and Linked Open Data (LOD) [10].

2. Research Problem and Motivation

First and foremost, I started my research with some questions to clarify: (i) how can be represented geospatial information in an ontological model (ii) how can be queried narratives through a geospatial dimension, and (iii) what can be the possible visualizations of geospatial information in narratives.

The main goal of my research is to design and implement a new extension of the NOnt Ontology to formally represent geographical knowledge. In particular, for geographical knowledge, I intend: (i) a point on the surface of the Earth identified by latitude and longitude; (ii) a polygon as represented through a geospatial standardised description, used by common GIS software and services; (iii) any region identified by an IRI in a standard GIS gazetteer. On the basis of this ontology extension, a new functionality of the Narrative Building and Visualising Tool (NBVT) software [11] will be implemented, which will allow the users to insert geographical knowledge in narratives. This knowledge will be visualized through the Story Maps visualization format. Story Maps are online representations of interactive maps enriched with text, pictures, videos and other multimedia content. Story maps are suitable to represent narratives with the geospatial dimension. However, the current state-of-the-art software for Story Maps has some issues and limitations, that: (i) every story is disconnected from the other; (ii) software for building Story Maps is commercial; (iii) Story Maps are not integrated with online collaborative environments.

To overcome the first issue, in my research project I will exploit the advantages provided by Semantic Web technologies. Indeed, starting from different stories, these technologies allow creating a unified knowledge graph that represents a more complete story encompassing all the other ones, following the idea that many stories can contribute creating one bigger and more complete story. To overcome the second and third issues, I will integrate free software for geospatial data visualization with a new version of NBVT, using the geographically-extended NOnt, and I will make the new NBVT interoperable with collaborative infrastructures, such as the D4Science e-Infrastructure of ISTI-CNR [12].

3. Methodological Approach

The methodological approach I will follow to model the geographical and spatial dimensions and improve the functionalities of NOnt ontology in DLs is close to the methodology that characterizes a common workflow to develop an algorithm in Computer Science [13], that is:

1. Formalisation of the problem
2. Computational analysis
3. Development of a new algorithm
4. Experimentation with a case study
5. Evaluation

This methodology phase will be adapted to my research. In particular, the first step will be the conceptualization of the domain, in which the issue is described and analyzed in its main parts. Secondly, an extension of the ontology will be developed as the specification of the conceptualization in terms of a logical theory. Based on this ontology extension, a new functionality of NBVT will be implemented that will allow the users to insert geographical knowledge in narratives. The geographical knowledge will be visualized through the Story Maps visualization format. As a case study, I will apply the new NBVT tool to the representation of the data of mountain ecosystems and economic value chains produced within the Mountain Valorization through Interconnectedness and Green Growth (MOVING) European project (2020-2023). Eventually, I will perform an evaluation of the ontology.

4. Related Works

The main aim of my research is to formally represent the spatio-temporal dimension in narratives using Semantic Web Technologies (i.e. RDF/S, OWL, SPARQL) and Linked Open Data (LOD).

Resource Description Framework (RDF) [14] is a W3C recommendation for the notation of metadata on the World Wide Web (WWW). The basic idea of RDF is a triple model *<subject>* *<predicate>* *<object>* where anything is a 'resource' identified by International Resource Identifier (IRI). Moreover, Resource Description Framework Schema (RDFS) [15] enriches and defines semantic characteristics of RDF data. RDFS allows the creation of hierarchies of classes and subclasses and properties and sub-properties. A further important standard is Web Ontology Language (OWL) [16]. OWL allows enhancing the expressivity of RDF/S languages. For example, it allows defining several property characteristics (e.g. transitivity, reflexivity) or axioms on classes (e.g. cardinality, universal and existential quantification). To retrieve and manipulate data stored in RDF/S and OWL formats SPARQL query language[17] can be used. The term Linked Data refers to a set of best practices for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF. In particular, the approach includes HTTP URIs for denoting resources, a description for every resource, and links to (and hopefully from) other datasets [18].

My research is still at an early stage, indeed, until now, I am almost concluding a state-of-the-art review about the ontologies for representing events geospatial and spatiotemporal RDF/S representation techniques.

In the Semantic Web field [19, 20], during the last few years, various formal models have been developed for representing the core concept of the event. These models include the Event Ontology [21], the Linking Open Descriptions of Events (LODE) [22], the Event-Model-F ontology [23], and the Simple Event Model (SEM) [24], among others. More general event-based formal models for representing knowledge are the CIDOC CRM [3], the Europeana Data Model [25], and the DOLCE upper level ontology [26].

On the other hand, several attempts to represent and cope with geospatial data in RDF/S have been done. One of the early works is the W3C Basic Geo Vocabulary (W3C GEO 2003) [27]. It allows the representation of points in World Geodetic System 1984 (WSG84), the most common Coordinate Reference System (CRS) for spatial data on the web. Another work is GeoRDF [28] an RDF-compatible profile for geographic information that allows representing lines and polygons too. Moreover, Open Geospatial Consortium (OGC) GeoSPARQL [29] has defined a vocabulary for representing geospatial data in RDF, and an extension to the SPARQL query language in order to process geospatial data. Another vocabulary for representing geospatial data in RDF is NeoGeo [30, 31], which is based on the GML Simple Features Profile. In this vocabulary, simple geometries are described (along with their coordinates) explicitly in RDF, and composite geometries are described as an aggregation of simple geometries. This approach allows reasoning and querying on these geometries, as well as making shared borders more explicit. Eventually, to address the problem of representing and querying incomplete geospatial information in RDF, Nikolaou and Koubarakis [32, 33] have proposed the framework RDF^i (where 'i' stands for 'incomplete'), which extends RDF with the ability to represent property values that exist but are unknown or partially known.

However, managing spatial RDF data isn't enough to deal with events. A lot of applications are related to the spatial context as well as to the temporal context. Because the geospatial objects are complex abstractions and can be a composition of other objects, simply adding a temporal dimension could lead to a higher level of complexity. So, there have been many interests in managing spatiotemporal RDF data. These models include the stRDF based on linear constraints [34], $stRDF^i$ [35], STT [36, 37], g^{st}-Store [38, 39], YAGO2 [40], and stRDFS [41].

Some research has been focused on integrating CIDOC-CRM with geospatial ontologies such as GEOSPARQL. Some of these models are CRMgeo [42] and the POI-based data model [43].

5. Impact and Expected Results

I performed the first experiment to extend NOnt with spatio-temporal knowledge. Since NOnt is an extension of CIDOC CRM, I took into account CRMgeo as reference ontology. Notice that CRMgeo is compliant with GeoSPARQL. Figure 1 shows the first attempt to model time and space using classes and properties of CRMgeo. As a notational convention, the CIDOC CRM uses the letters "E" and "P" to indicate classes and properties respectively. On the other hand, CRMgeo uses the letters "SP" and "Q" to indicate classes and properties, respectively.

Making available an ontology for representing narratives with geospatial information can improve the functionalities of the current DLs. In particular, using an ontology, DLs can provide stories about their collected digital objects instead of single items. Furthermore, by exploiting the semantic model, it is possible to create links among different narrations - e.g. linking

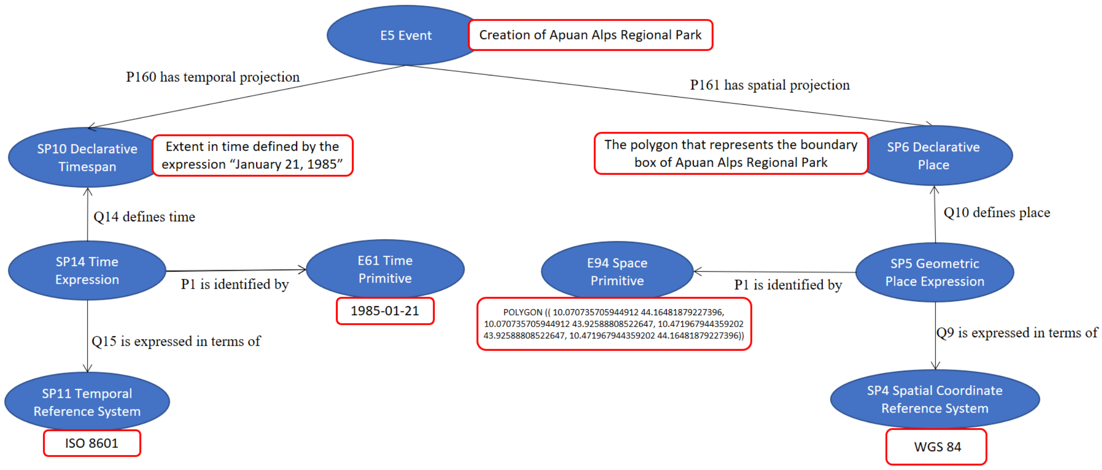

Figure 1: First attempt to model time and space using classes and properties of CRMgeo and CIDOC CRM.

narrations that refer to the same places - as well as to enrich the narrations with data included in other Knowledge Bases (KB).

From a technological point of view, using standard (e.g. GeoSPARQL) allows representing geospatial data also in other information systems, like Geographic Information Systems (GIS). Moreover, having available this data as either RDF/S or OWL graphs allows performing data analysis, and using inference engines to deduce new knowledge.

From a conceptual point of view, having a semi-automatic tool that creates narratives in form of a Story map using Semantic Web technologies allows overcoming the perceptual gap between the territory as a whole and its narrated representation [44]. Story maps can help fill the gap by linking spatiotemporal events to emotions, perceptions, and historical context. Indeed, Story maps face a perceptive cartographic challenge by enriching the maps with multimedia data to increase the overall perception of the territory beyond the map.

This approach can be used to narrate stories in different scientific fields - e.g. cultural heritage, ecosystems, sustainable resource management, product niches, tourism - and it can be beneficial for several research fields, especially for the dissemination of scientific results and for educational purposes. Moreover, a relevant contribution of a visualization tool is the power to boost the capacity to explore data, foster understanding and encourage learning, especially among different communities (such as scientists of different disciplines).

References

[1] M. Bal, Narratology: Introduction to the theory of narrative, University of Toronto Press, Toronto, 1997. ISBN: 978-0802096319.

[2] C. Meghini, V. Bartalesi, D. Metilli, Representing narratives in digital libraries: The narrative ontology, Semantic Web 12 (2021) 241–264. doi:10.3233/SW-200421, publisher: IOS Press, DOI: https://doi.org/10.3233/SW-200421.

[3] M. Doerr, C.-E. Ore, S. Stead, The cidoc conceptual reference model: a new standard for knowledge sharing, in: Tutorials, posters, panels and industrial contributions at the 26th International Conference on Conceptual Modeling, Australian Computer Society, Darlinghurst, 2007, p. 51–56.
[4] M. Doerr, C. Bekiari, P. L. Boeuf, Frbr oo , a conceptual model for performing arts, Annual Conference of CIDOC Athens (2008).
[5] Swrl: A semantic web rule language combining owl and ruleml, ???? URL: https://www.w3.org/Submission/SWRL/.
[6] R. Kowalski, M. Sergot, A logic-based calculus of events, 1989. DOI: https://doi.org/10.1007/978-3-642-83397-7_2.
[7] R. Miller, M. Shanahan, Some alternative formulations of the event calculus, Computational logic: logic programming and beyond (2002) 452–490. DOI:https://doi.org/10.1007/3-540-45632-5_17.
[8] E. Mueller, Commonsense Reasoning: An Event Calculus Based Approach, Morgan Kaufmann, Burlington, MA, 2014. ISBN: 978–0128014165.
[9] P. Hitzler, M. Krotzsch, S. Rudolph, Foundations of semantic web technologies, Chapman and Hall/CRC, 2009.
[10] C. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: Semantic services, interoperability and web applications: emerging concepts, IGI global, 2011, pp. 205–227.
[11] B. V., M. D., M. C., Constructing narratives using nbvt: a case study, in: 8th AIUCD Conference 2019, pp. 169–171, Udine, Italy, 22-25 January 2019, 2019.
[12] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, F. Sinibaldi, Enacting open science by D4Science, Future Generation Computer Systems 101 (2019) 555–563. URL: https://www.sciencedirect.com/science/article/pii/S0167739X1831464X, DOI: https://doi.org/10.1016/j.future.2019.05.063.
[13] A. Levitin, Introduction to the Design & Analysis of Algorithms, Always learning, Pearson, 2012.
[14] Rdf 1.1 semantics, ????. URL: https://www.w3.org/TR/rdf11-mt/.
[15] Rdf schema 1.1, ????. URL: https://www.w3.org/TR/rdf-schema/.
[16] Web ontology language (owl), ???? URL: https://www.w3.org/TR/owl-features/.
[17] Sparql query language for rdf, ???? URL: https://www.w3.org/TR/rdf-sparql-query/.
[18] Linked data - w3c wiki, ???? URL: https://www.w3.org/wiki/LinkedData.
[19] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific American 284, 5 (2001-05) 34–43.
[20] A. Bernstein, J. Hendler, N. Noy, A new look at the semantic web. commun, ACM 59, 9 (2016-09) 35–37.
[21] Y. Raimond, S. Abdallah, The Event Ontology, Technical Report, Queen Mary University of London, 2007. URL: http://motools.sourceforge.net/event.
[22] R. Shaw, R. Troncy, L. Hardman, Lode: Linking open descriptions of events, in: The Semantic Web, Springer, Berlin, Heidelberg, 2009, p. 153–167. DOI: https://doi.org/10.1007/978-3-642-10871-6_11.
[23] A. Scherp, T. Franz, C. Saathoff, S. Staab, F: A model of events based on the foundational ontology dolce+ dns ultralight, in: Proceedings of the fifth International Conference on Knowledge Capture, ACM, 2009, p. 137–144. DOI: https://doi.org/10. 1145/1597735.1597760.

[24] W. Van Hage, V. Malaisé, R. Segers, L. Hollink, G. Schreiber, Design and use of the simple event model (sem), Web Semantics: Science, Services and Agents on the World Wide Web 9 (2011) 128–136. DOI: https://doi.org/10.1016/j.websem.2011.03.003.
[25] M. Doerr, S. Gradmann, S. Hennicke, A. Isaac, C. Meghini, H. Sompel, The europeana data model (edm), in: World Library and Information Congress: 76th IFLA general conference and assembly, IFLA, 2010, p. 10–15.
[26] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider, Sweetening ontologies with dolce, knowledge engineering and knowledge management: Ontologies and the semantic web, 2002. DOI: https://doi.org/10.1007/3-540-45810-7_18.
[27] Basic geo (wgs84 lat/long) vocabulary, ???? URL: https://www.w3.org/2003/01/geo/.
[28] Geordf - w3c wiki, ???? URL: https://www.w3.org/wiki/GeoRDF.
[29] R. Battle, D. Kolas, Enabling the geospatial semantic web with parliament and geosparql, Semantic Web 3 (2012) 355–370. doi:10.3233/SW-2012-0065.
[30] J. M. Salas, A. Harth, Finding spatial equivalences across multiple rdf datasets (????). URL: http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource/.
[31] J. M. Salas, A. Harth, B. Norton, L. M. Vilches, A. D. León, J. Goodwin, C. Stadler, S. Anand, D. Harries, Neogeo vocabulary: Defining a shared rdf representation for geodata, 2011. URL: http://geovocab.org/doc/neogeo.html.
[32] C. Nikolaou, M. Koubarakis, Querying linked geospatial data with incomplete information, 2012. URL: https://www.madgik.di.uoa.gr/publications/querying-linked-geospatial-data-incomplete-information.
[33] C. Nikolaou, M. Koubarakis, Incomplete information in rdf, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7994 LNCS (2013) 138–152. doi:10.1007/978-3-642-39666-3_11.
[34] M. Koubarakis, K. Kyzirakos, Modeling and querying metadata in the semantic sensor web: The model strdf and the query language stsparql, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6088 LNCS (2010) 425–439. doi:10.1007/978-3-642-13486-9_29.
[35] M. Koubarakis, K. Kyzirakos, B. Nikolaou, M. Sioutis, S. Vassos, A data model and query language for an extension of rdf with time and space, 2012.
[36] M. S. Perry, A Framework to Support Spatial, Temporal and Thematic Analytics over Semantic Web Data, Ph.D. thesis, USA, 2008. AAI3324256.
[37] A. Sheth, M. Perry, Traveling the semantic web through space, time, and theme, IEEE Internet Computing 12 (2008) 81–86. doi:10.1109/MIC.2008.46.
[38] D. Wang, L. Zou, D. Zhao, g st-store: An engine for large rdf graph integrating spatiotemporal information (2014). URL: http://www.govtrack.us/data/rdf/. doi:10.5441/002/edbt.2014.66.
[39] D. Wang, L. Zou, D. Zhao, gst-store: Querying large spatiotemporal rdf graphs, Data and Information Management 1 (2017) 84–103. doi:10.1515/DIM-2017-0008.
[40] J. Hoffart, F. M. Suchanek, K. Berberich, G. Weikum, Yago2: A spatially and temporally enhanced knowledge base from wikipedia, Artificial Intelligence 194 (2013) 28–61. doi:10.1016/J.ARTINT.2012.06.001.
[41] L. Zhu, N. Li, L. Bai, Algebraic operations on spatiotemporal data based on rdf, ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 80 9 (2020) 80. URL: https:

//www.mdpi.com/2220-9964/9/2/80/htmhttps://www.mdpi.com/2220-9964/9/2/80. doi:10.3390/IJGI9020080.

[42] G. Hiebel, M. Doerr, , Øyvind Eide, Crmgeo: A spatiotemporal extension of cidoc-crm, Int J Digit Libr 18 (2017) 271–279. doi:10.1007/s00799-016-0192-4.

[43] B. Ranjgar, A. Sadeghi-Niaraki, M. Shakeri, S. M. Choi, An ontological data model for points of interest (poi) in a cultural heritage site, Heritage Science 10 (2022) 1–22. URL: https://heritagesciencejournal.springeropen.com/articles/10.1186/s40494-021-00635-9. doi:10.1186/S40494-021-00635-9/TABLES/17.

[44] A. Korzybski, Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics, International non-Aristotelian library, International Non-Aristotelian Library Publishing Company, 1933. URL: https://books.google.it/books?id=WnEVAQAAIAAJ.