TOWARDS DIGITAL TWINS OF TERRITORIES THROUGH SEMANTIC STORY MAPS

Valentina Bartalesi, Gianpaolo Coro, Emanuele Lenzi, Nicolò Pratelli *and* Pasquale Pagano
Institute of Information Science and Technologies "Alessandro Faedo" (ISTI) of the National Research Council of Italy (CNR), Via Moruzzi 1, 56124, Pisa, Italy
e-mail: valentina.bartalesi@isti.cnr.it

Digital maps greatly support storytelling about territories, especially when enriched with data describing cultural, societal, and ecological aspects, conveying emotional messages that describe the territory as a whole. Story maps are interactive online digital narratives that can describe a territory beyond its map by enriching the map with text, pictures, videos, and other multimedia information. This paper outlines how online story maps can fill the gap between a map and a territory in narratives to create a digital twin of different territories as inter-connected semantic stories.

Keywords: *story maps, semantic web, linked open data, digital twin of territories*

1. Introduction

A narrative is a method of articulating life experiences in a meaningful way and a conceptual basis of collective human understanding [1]. Humans use narratives, or stories, to explain and communicate abstract experiences, describe a particular point of view on some domain of interest, and share meanings among different communities. Maps have always geographically supported stories and storytelling, especially to represent the geospatial components of narratives. However, as Korzybski pointed out in his General Semantics [2], there is a perceptual gap between the territory and its map in narratives. The *perceptive* cartographic challenge for a map is when it tries to represent also the life, emotions, reality, fiction, legends, and expectations associated with the described territory. Story maps are computer science realisations of narratives based on interactive online maps enriched with text, pictures, videos, data, and other multimedia information. Story maps building platforms support diverse users to participate and possibly collaborate in the map-based story telling. Story maps can be used both to enhance the perception of the geographic space cited in narratives and also to overcome Korzybski's perceptual gap, enriching maps with media that communicate emotional messages associated with the described territory, e.g. digital audio/video material to describe the territorial complexity. Story maps satisfy these properties creating a digital twin of the territory, going beyond the standard map representation. Digitally interconnecting different story maps within a knowledge base can reveal implicit information hidden within and between the territories and allows answering previously new, complex questions.

This paper briefly outlines how online story maps can fill the gap between a map and a territory in narratives to create a digital twin of different territories as inter-connected stories. We start from a specific domain (mountain territories and value chains) and propose a general software solution for a semi-automatic transformation from text to narrative. In particular, we build a workflow to transform the textual descriptions of the value chains and territories of 23 rural European areas from 16 countries involved in the MOVING (MOuntain Valorisation through INterconnectedness and Green growth) European project [3]. MOVING aims at building co-development policy frameworks across Europe that can contribute to the resilience and sustainability of mountain areas. Community data on rural areas come in unstructured textual formats. They summarise the emotional, cultural, and societal contents to attract tourism and customers. Digitalising this information and assigning it to map locations is complex, but is crucial for extending the information reachability, effectively communicate contents, and harmonise the data to produce one overall narrative of territories.

Our semi-automatic workflow transforms the unstructured information about a territory into a story map. The workflow uses natural language processing algorithms [4] to extract terms (persons, locations, organisations, and keywords) with high importance to understand the text (named entities). Successively, it associates the extracted terms with Wikidata entries and geographic coordinates. Then, it produces and visualises as story maps a sequence of story events enriched with titles, descriptions, entities, coordinates, and multimedia data. Finally, the workflow feeds a semantic knowledge base, based on an ontology for representing narratives [5], to enable data discovery and integration. The workflow design is independent of the specific application context. Furthermore, differently from other story-map creation software [6, 7], the workflow components are open-source, and oriented to the Open Science directives of transparency and reproducibility of data, results, and processes. Moreover, different from other story-visualisation software [8, 9, 10, 11, 12], the workflow provides a semantic knowledge base that automatically inter-connects the created stories, extracts and discover knowledge from them, and comply with the Linked Open Data paradigm. This way, our workflow points towards satisfying the creation of a digital twin of the territory.

2. Methodology

2.1 Data pre-processing

We pre-processed the unstructured textual data of 23 European regions to prepare one new textual document - in the MS Excel format - for each region. For each story, we organised the new MS Excel document by describing one story event in each row. Specifically, each row reported (i) a title, (ii) a description, (iii) one representative image (optionally), (iv) hyperlinks to online multimedia material (optionally), and (v) the event *type*. The row sequence represented the events' sequence of the region's story. An event sequence reports the same concepts expressed by the paragraphs of the corresponding region description by the MOVING regional experts. The data associated with each region from the experts' descriptions and original data were mapped onto the new MS Excel documents by attaching them to the most appropriate events. The 23 newly prepared documents were sequentially passed to our workflow as input data.

2.2 Story-structure building

A story-structure building module processes the text of the input MS Excel rows (events) to enrich them with information on associated entities and locations. In particular, it uses the NLPHub [4] and the Wikidata services to extract valid entities with possible associated spatial coordinates. To keep the entity locations focused on a geographically consistent region, our module traces a bi-variate log-normal distribution on the longitude-latitude pairs. Then, it sets the boundaries to the upper and lower log-normal confidence limits over the axes, and marks all coordinates outside of these boundaries as outliers. For the entities whose coordinates are not present in Wikidata, the algorithm adds by default the coordinates of the Local Administrative Unit in which the story occurs.

2.3 Story map creation

Each MS Excel files enriched with event and coordinate information are further processed to finalise the story map representation through multimedia and hyperlink information and a structured format. First, each story event is associated with the event type and multimedia hyperlinks specified in the original input text. Second, all acronyms are expanded through a reference domain-specific dictionary to make descriptions less technical. Third, images are linked to the events if referred in the original input text; otherwise, the first image associated with the event entities on Wikidata, ordered by their position in the text, is retrieved and linked. Finally, the story's event sequence, with all associated entities, images, and links is described in the JSON format, according to the schema used by the Story Map Building and Visualising Tool (SMBVT) [13]. This JSON document is an offline realisation of the story map. The document it stored on a PostgreSQL-JSON database used by the SMBVT for fast online visualisation (Section 2.4). Finally, the script invokes a semantic triplification software that translates the JSON document into a Web Ontology Language (OWL) graph and stores it in an Apache Jena Fuseki triple store. This graph complies with

the Narrative Ontology model [5], and its scope is to populate the SMBVT knowledge base and enable semantic queries for knowledge extraction.

2.4 Story Map Building and Visualising Tool

SMBVT [14, 15] is a open-source software that represents narratives as a network of spatiotemporal events related by semantic relations (part-of, temporal, spatial, and causal relations) enriched by event components, i.e., the entities that take part in the event (e.g., persons, objects, places, concepts). For the present experiment, we used a free-to-use SMBVT instance hosted on the D4Science e-Infrastructure accessible after free registration to the platform [16]. Following the Semantic Web approach [17], SMBVT assigns to each event and event component an IRI [18]. These IRIs are mainly extracted from Wikidata [19], which SMBVT uses as an external reference knowledge base. SMBVT retrieves events, entities and all associated data from a PostgreSQL-JSON database for visualisation. It synchronises this database with an OWL-graph representation of the stories stored on an Apache Jena Fuseki server, compliant with the Narrative Ontology model. The Fuseki server provides a SPARQL endpoint to query the complete graph of collected stories. The graph automatically connects the stories through the entities shared between the events. The server allows executing SPARQL data-extraction queries on the entire story graph within or across the stories. This feature allows connecting the stories to other knowledge bases published as Linked Open Data [20] (e.g., Europeana, [21]). In particular, based on the SPARQL server, SMBVT offers an entity-search functionality through a Web interface that allows querying the entire knowledge base in a user-friendly way. This feature is crucial to explore story inter-connections, for example, to retrieve (i) the events involving a specific entity (e.g., sheep, beer, etc.), (ii) the nations sharing the same entities (e.g., products, export locations, etc.), and (iv) the most frequent entities across the stories (e.g., the most common products). SMVBT provides an online graphical interface to create and manage the stories. This interface facilitates story-event building and event sequencing and contextualisation. SMBVT visualises the produced stories as story maps placing the narrative events on an interactive map that respects an event browsing order based on the user-defined plot.

The SMBVT story map publication process returns a public link for each story map. Each publication operation overwrites the previously published application so that the public link always points to the latest story-map version.

3. Conclusions

Through semantic queries, our approach contributes to discover new knowledge from the data; for example, the territories sharing the same environmental characteristics (e.g., rivers, lakes, vineyards, chestnut trees) and issues (e.g., depopulation, emigration, climate change problems), or providing similar products (e.g., cow or sheep milk cheese). Discovering new knowledge from the data is particularly useful for mountain ecosystems to design sustainable environmental management pathways and contribute to long-term cities' ecological sustainability. The extracted knowledge therefore represent a digital twin of the territories through the interconnection of their representations as story maps. The constructed semantic knowledge base can indeed help understand and dam the vanishing of essential services in rural areas due to constant depopulation trends in rural communities. Therefore, it helps answering crucial questions about the territory socio-economic and ecosystem status. The importance of describing a territory beyond its maps is more understandable if we consider that the world's rural population will likely pass from 47% (today) to 30% by 2050 [22]. This trend poses problems due to vanishing traditional and cultural heritage aspects and human health, welfare, and cities' ecological sustainability, monitored by the United Nations' Sustainable Development Goal 11 ("sustainable cities and communities"). In this context, story maps and knowledge discovery can help citizens, investors, and governmental authorities to better understand the ecosystem and support sustainability strategies. The MOVING project has indeed adopted story maps (available at `https://www.moving-h2020.eu/story-maps/`) as effective tools for this task because the experts judged them to appropriately convey "information going beyond the map" for scientists, stakeholders, and the general public.

REFERENCES

1. Wertsch, J. V. and Roediger, H. L. Collective memory: Conceptual foundations and theoretical approaches, *Memory*, **16** (3), 318–326, (2008).

2. Korzybski, A., (1933), A non-aristotelian system and its necessity for rigour in mathematics and physics. *Science and sanity: an introduction to non-Aristotelian systems and general semantics*, Lancaster.

3. MOVING, (2020), *The MOVING European project - Mountain Valorisation through Interconnectedness and Green Growth*. Accessed 4 January 2023 `https://www.moving-h2020.eu/`.

4. Coro, G., Panichi, G., Pagano, P. and Perrone, E. Nlphub: An e-infrastructure-based text mining hub, *Concurrency and Computation: Practice and Experience*, **33** (5), e5986, (2021).

5. Meghini, C., Bartalesi, V. and Metilli, D. Representing narratives in digital libraries: The narrative ontology, *Semantic Web*, **12** (2), 241–264, (2021).

6. ArcGIS, (2022), *ArcGIS Story Map Building Tool*. `https://storymaps.arcgis.com`.

7. Timescape, (2020), *Timescape Story Map Building Tool*. `https://www.recombine.net/projects/timescape.html`.

8. Knightlab, (2020), *StoryMap JS tool for narrative building*. `https://storymap.knightlab.com`.

9. Frenvik Sveen, A., (2020), *Leaflet Storymaps template*. `https://github.com/atlefren/storymap`.

10. Map Box, (2020), *Map Box Storymaps template*. `https://www.mapbox.com/solutions/interactive-storytelling`.

11. Becker, T., Köbben, B. and Blok, C. Timemapper : visualizing moving object data using wms time and svg smil interactive animations, *Proceedings SVGOpen 2009*, SVGOpen, 7th international conference on scalable vector graphics (SVGOpen 2009), 02-04-10-2009, Mountain View, United States, (2009).

12. Odyssey, (2020), *Odyssey Story Maps template*. `http://cartodb.github.io/odyssey.js/`.

13. Bartalesi, V., Coro, G., Lenzi, E., Pratelli, N. and Pagano, P. ISTI Research Report, MOVING, D3.3, 2022, Moving d3.3 - tools for science-society-policy interfaces. using semantic story maps to describe a territory beyond its map, (2022).

14. Bartalesi, V., Coro, G., Lenzi, E., Pagano, P. and Pratelli, N. From unstructured texts to semantic story maps, *International Journal of Digital Earth*, **16** (1), 234–250, (2023).

15. Bartalesi, V., Lenzi, E. and Pratelli, N. A web tool to create and visualise semantic story maps, *Text2Story 2023 Sixth Workshop on Narrative Extraction From Texts*, CEUR, (2023).

16. Assante, M., et al. Enacting open science by d4science, *Future Generation Computer Systems*, **101**, 555–563, (2019).

17. Berners-Lee, T., Hendler, J. and Lassila, O. The semantic web, *Scientific american*, **284** (5), 34–43, (2001).

18. Dürst, M. and Suignard, M. W3C, Internationalized resource identifiers (iris), (2005).

19. Vrandecic, D. The rise of wikidata, *IEEE Intelligent Systems*, **28** (4), 90–95, (2013).

20. Bizer, C., Heath, T. and Berners-Lee, T., (2011), Linked data: The story so far. *Semantic services, interoperability and web applications: emerging concepts*, pp. 205–227, IGI global.

21. Isaac, A. and Haslhofer, B. Europeana linked open data–data. europeana. eu, *Semantic Web*, **4** (3), 291–297, (2013).

22. United Nations, (2018), *68% of the world population projected to live in urban areas by 2050, says UN*. on-line, https://www.un.org/en/desa/68-world-population-projected-live-urban-areas-2050-says-un.