Digital Humanities and Semantic Web: The New Frontiers of Transdisciplinary Knowledge

Agostino Sorbara
Pegaso International

The Digital Humanities is a field of study, research, and teaching, that arises from the union of humanities and digital disciplines. It includes research, analysis and dissemination of knowledge through computer media. Through this research we will analyze the way in which computers process the Digital Humanities datasets so that they can be exchanged on the internet. In this paper, we will illustrate the strengths of their representation in terms of quantitative data through the semantic web. Finally, guidelines will be provided to create models to provide added value to the knowledge society.

Keywords: digital humanities, semantic web, RDF, linked data, SPARQL, SKOS, OWL

INTRODUCTION

The Digital Humanities is a field of study, research, and teaching, that arises from the union of humanities and digital disciplines. It includes research, analysis and dissemination of knowledge through computer media. In addition to having a solid humanistic background, the researchers of Digital Humanities must be able handle cultural content with the appropriate information technology tools.

Digital Humanities is a diverse and still emerging field that encompasses the practice of humanities research through information technology, and the exploration of how the humanities may evolve through integration with technology, media, and computational methods.

Through this research we will analyze the way in which computers process the Digital Humanities datasets so that they can be exchanged on the internet. In this paper, we will illustrate the strengths of their representation in terms of quantitative data through the semantic web.

The technologies semantic web used are RDF, Linked data, SPARQL, SKOS and OWL, which offer excellent opportunities to represent Digital Humanities in quantitative (digital) data.

From the point of view of scientific research, the semantic web assemble heterogeneous data, such as texts, images, sounds, videos, etc., aggregating them thus favoring their cataloging. Shifting focus, from data publication, to data analysis, leads scientific research to face new challenges, such as the knowledge extraction, the aggregated visualization of data, machine learning and the knowledge discovery. Finally, guidelines will be provided to create models to provide added value to the knowledge society.

RDF

RDF (Resource Description Framework) is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.

LINKED DATA

The collection of Semantic Web technologies provides an environment where application can query that data, draw inferences using vocabularies, etc.

However, to make the Web of Data a reality, it is important to have the huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but relationships among data should be made available, too, to create a Web of Data. This collection of interrelated datasets on the Web can also be referred to as Linked Data.

Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web. Almost all applications listed, say collection of Semantic Web Case Studies and Use Cases are essentially based on the accessibility of, and integration of Linked Data at various level of complexities.

SPARQL

SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language) is, a semantic query language for databases - able to retrieve and manipulate data stored in RDF format. SPARQL allows for a query to consist of triple patterns, conjunctions, disjunctions, and optional patterns.

SKOS

SKOS (Simple Knowledge Organization System) is a common data model for sharing and linking knowledge organization systems via the Web.

Many knowledge organization systems, such as thesauri, taxonomies, classification schemes and subject heading systems, share a similar structure, and are used in similar applications. SKOS captures much of this similarity and makes it explicit, to enable data and technology sharing across diverse applications.

The SKOS data model provides a standard, low-cost migration path for porting existing knowledge organization systems to the Semantic Web. SKOS also provides a lightweight, intuitive language for developing and sharing new knowledge organization systems. It may be used on its own, or in combination with formal knowledge representation languages such as the OWL.

OWL

OWL (Ontology Web Language) is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit. OWL documents, known as ontologies, can be published in the World Wide Web and may refer to or be referred from other OWL ontologies. OWL is part of the World Wide Web Consortium Semantic Web technology stack, which includes RDF, RDFS, SPARQL, etc.

BIG DATA AND SMART DATA

The conversion of Big Data into Smart Data introduced a methodological change in the Digital Humanities. When inserting Big Data and Smart Data in the context of the Digital Humanities, one of the main challenges to be overcome is that of using them in historical contexts. In processes that transform unstructured data into structured and semi-structured, the Smart Data strategy pushes data service providers to target the machine understandable, processable and operable (instead of the machine simply readable). This is to ensure that accurate data can be obtained in highly efficient interconnection, citation and data transfer processes.

DATA MODELLING

Data modeling is one of the main tasks that Digital Humanities researchers have to perform. Digital Humanities researchers are engaged in creating databases, digital editions, geographic information systems, research collections, digital libraries, and heterogeneous data collection.

Data modeling refers to the activity of designing a model of a subset of the real (or imaginary) world to meet a specific set of user requirements using one or more meta-models.

In general, a data model consists of a conceptual part that defines the semantic part of the data, the relevant entities, their characteristics and their relationships, and a logical part that expresses a subset of the conceptual model.

For the modeling of data in the Digital Humanities two characteristics are of particular importance:

- objects of data modeling activities are artifacts and many of their activities are intentionally created;
- the objects of humanistic research have a long history which must be adequately managed.

GUIDELINES TO CREATE MODELS TO PROVIDE ADDED VALUE TO THE KNOWLEDGE SOCIETY

The guidelines for the theoretical reference model consist of values and methods.

Values:

- authors must be capable of self-criticism;
- authors must encourage experimentation;
- there must be broad collaboration between the different authors.

Methods:

- switch from the big data to smart data;
- aggregate data;
- data design;
- thick mapping;
- dynamic interactive archives;
- production of the distributed knowledge;
- virtual learning environments;
- reusable content.

CONCLUSION

From the point of view of scientific research, the semantic web assemble heterogeneous data, such as texts, images, sounds, videos, etc., aggregating them thus favoring their cataloging.

Shifting focus, from data publication, to data analysis, leads scientific research to face new challenges, such as the knowledge extraction, the aggregated visualization of data, machine learning and the knowledge discovery.

Today, advanced technologies such as Big Data and Smart Data allow Digital Humanities researchers to manage large volumes of heterogeneous data among themselves.

Having guidelines for modeling facilitates the work of researchers.

REFERENCES

Ackoff, R.L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Beynon, W.M., Russ, S., & McCarty, W. (2006). Human computing: modelling with meaning. *Literary and Linguistic Computing*, *21*, 141–157.

Borgman, C. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Ciula A., & Eide, Ø. (2014). Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory. *1st International Conference on Digital Access to Textual Cultural Heritage (DATeCH)*, pp. 35–41. Madrid, Spain.

Ciula A., & Eide, Ø. (2017). Modelling in digital humanities: Sign in context. *Digital Scholarship in the Humanities*, *32*(1).

Ciula, A., & Marras, C. (2016). Circling around texts and language: towards ''pragmatic modelling''. *Digital Humanities Quarterly*, *10*(3).

Eide, Ø. (2015). Ontologies, data modeling. *Journal of the Text Encoding Initiative*, *8*.

Flanders, J., & Jannidis, F. (2015). *Knowledge Organization and Data Modeling*. Humanities.

Gardiner, E., & Musto, R.G. (2015). *The Digital Humanities: A Primer for Students and Scholars*. Cambridge University Press.

Hyvönen, E. (2012). *Publishing and using cultural heritage linked data on the semantic web*. Morgan & Claypool.

Hyvönen, E. (n.d.). *Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Knowledge Discovery*. IOS Press.

Iafrate, F. (2015). *From big data to smart data*. Wiley.

Jannidis, F., & Flanders, J. (2012). *Knowledge Organization and Data Modeling in the Humanities*. Workshop Brown University.

Jannidis, F., & Flanders, J. (2013). A Concept of Data Modeling for the Humanities. *Digital Humanities*, pp. 237–239.

Josh, H. (2015*). A Guide to Digital Humanities: Values and Methods*. Northwestern University Library.

Kaplan, F. (2015). A map for big data research in digital humanities. *Frontiers in Digital Humanities*, *2*, 1-7.

Kralemann, B., & Lattmann, C. (2013). Models as icons: modeling models in the semiotic framework of Peirce's theory of signs. *Synthese*, *190*(3), 397–420.

Mastandrea, P. (2018). *L'orizzonte delle Digital Humanities*. Umanistica Digitale.

Meroño-Peñuela, A. (n.d.). Digital Humanities on the Semantic Web: Accessing Historical and Musical Linked Data. *Journal of Catalan Intellectual History*.

Schöch, C. (2013). Big? smart? clean? messy? Data in the humanities. *Journal of Digital Humanities*, *2*(3), 2–13.

Sheth, A. (2014). *Transforming big data into smart data*. 30th IEEE International Conference on Data Engineering (ICDE).

Sprugnoli, R., Pardelli, G., Boschetti, F., & Del Gratta, R. (2019). *Un'Analisi Multidimensionale della Ricerca Italiana nel Campo delle Digital Humanities e della Linguistica Computazionale*. Umanistica Digitale.
Tomasi F. (2018). Modelling in the Digital Humanities. *Historical Social Research*, *31*, 170-179.
Zeng, M.L. (2017). *Smart Data for Digital Humanities*. School of Library & Information Science.