IRCDL 2021
Italian Research Conference on Digital Libraries

Proceedings of the 17th Italian Research Conference on Digital Libraries

Padua, Italy (virtual event due to the Covid-19 pandemic), February 18-19, 2021

Edited by

Dennis Dosso *1
Stefano Ferilli *2
Paolo Manghi *3
Antonella Poggi *4
Giuseppe Serra *5
Gianmaria Silvello *1

*1 University of Padua, Department of Information Engineering, Padua, Italy
*2 University of Bari, Department of informatics, Bari, Italy
*3 International National Research Council (CNR), Pisa, Italy
*4 Sapienza University of Rome, Faculty of Arts and Humanities, Rome, Italy
*5 University of Udine, Department of Mathematics and Computer Science, Udine, Italy

Table of Contents

Full Papers

Short Papers/Extended Abstracts

Exploring the Historical Context of Graphic Symbols: the NOTAE Knowledge Graph and its Visual Interface ⋆

Eleonora Bernasconi[1], Maria Boccuzzi[1], Tiziana Catarci[1], Miguel Ceriani[2], Antonella Ghignoli[1], Francesco Leotta[1], Massimo Mecella[1], Anna Monte[1], Nina Sietis[1], Silvestro Veneruso[1], and Zahra Ziran[1]

[1] Sapienza Università di Roma, ITA
{name.surname}@uniroma1.it
[2] Università degli Studi di Bari Aldo Moro, ITA
miguel.ceriani@uniba.it

Abstract. Graphic symbols i.e. graphic entities drawn as a visual unit in a written text and representing something other or something more than a word of that text are the research object of the NOTAE project, which investigates them in the documentary practice of the late Roman State and Post-Roman Kingdoms (400-800 AD). While research results from the project are stored by filling forms resulting from the analysis of ancient documents, we argue that the availability of a navigable knowledge graph can ease the work of researchers at finding non trivial implications in data. In this paper, we propose a first version of the NOTAE Knowledge Graph, and we outline future works and possible synergies.

Keywords: Knowledge Graph · Metadata · Semantic Web · Linked Open Data · Visual search interface.

1 Introduction

With the gradual introduction of subscription and the increasing use of papyrus the presence of graphic symbols became widespread in late Roman legal practice as it already was for some time in other expressions of the so-called "pragmatic literacy" (defined in [10] as the "literacy of one who has to read or write in the course of transacting any kind of business"); it continued in Post-Roman

---
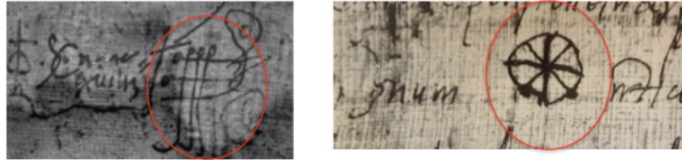
Fig. 1. Examples of graphic symbols. *Left)* graphic symbol in complex structure at the end of the subscription written by a Greek notary (Hermoupolis, Egypt, 561 AD). *Right)* symbol (Greek cross and diagonal cross crossing each other within a circle) drawn by an illiterate seller in his own hand on the sale contract (Ravenna, Italy, 572 AD).

kingdoms as part of the same historical process of reception of the late antique documentary practice. A graphic symbol is meant as a graphic device composed by graphic signs (included alphabetical ones) drawn as a visual unit in a written text and representing something other or something more than a word of that text. The message it carries on is to be discovered, not only by definition, because there is no intrinsic prior relationship between the message-bearing graphic entity and the information it conveys, but also in historical terms: a huge quantity of graphic symbols of different forms (in appearance strange or common, unique or recurring, original or archetypal) was drawn by a huge quantity of persons, both literates of any degree (professional scribes, bureaucrats, economic lites) and illiterates, in different typologies of documents, with different functions and expectations, in different places, different social, cultural and economic contexts. Examples of graphic symbols are shown in Fig. 1.

For illiterates, for example, both in the Greek-Latin graphic and linguistic koin of the late Roman state and in the Post-Roman kingdoms as long as Latin functioned as language of vertical communication, performing graphic symbols in their own hand certainly meant a way of taking an active part in the writing process: in other words, this phenomenon can be seen as the "other side" of the written world and considered as a matter of literacy in a wider sense. Not only for this reason the project NOTAE represents the first attempt to investigate this material from a novel perspective, as indicated in [6].

With the term NOTAE system, we refer to the information system that let human experts involved in NOTAE to insert, update and query the research insights gathered during the project activities. In particular, by manual inspection of ancient documentary sources, researchers maintain a database of documents, physical supports (papyrus, wooden tablet, slate, parchment) containing texts and graphic symbols, if any, together with a wide range of related information. Part of this activity could be improved in future by the employment of automatic machine learning and image processing methods [2].

One of the foreseen outcomes for the project, is to discover geographical and historical implications of the employment of graphic symbols, and this requires to provide researchers with advanced query and visualization functionalities. In this paper, in particular, we introduce the first version of the NOTAE Knowledge

Graph (KG). A KG [5] is a knowledge base that uses a graph-structured data model or topology to integrate data. KGs are often used to store interlinked descriptions of entities objects, events, situations or abstract concepts with a defined semantics.

By building the NOTAE KG, we aim at *(i)* introducing a common vocabulary for researchers in the area, *(ii)* sharing a common understanding of how concepts are related, *(iii)* enabling the reuse of domain knowledge, and *(iv)* making domain assumptions explicit. In addition, to build a proper KG for the NOTAE project, we also propose a graphical user interface which allows researchers *(i)* to explore the NOTAE KG, *(ii)* to search and explore relations and connections between resources, *(iii)* to make historical-geographical implications, and *(iv)* to discover new facts about the research field.

In this paper, we envision a tool to be employed in the broad field of digital humanities, conveying information not only from the NOTAE project but also from the vast set of related repositories and services available on the Internet. The challenge here is to enhance all of these resources, turning them into a semantically enriched ecosystem to ease information accessibility and knowledge discovery. At the best of our knowledge, even though some of the available services allow to download data as Linked Open Data (LOD), none of them provides an integrated view of the research field nor allow to draw historical implications and explore data in a simple and intuitive manner. The availability of a tool such as the one proposed in this paper would foster collaboration among the researchers in the area, and could attract curious [9], casual, users by easing the diffusion of niche topics like those regarding ancient documentary texts on which NOTAE project focuses.

The rest of the work is organized as follows. Section 2 reviews the major information systems and tools in the area of ancient written sources and related historical data. Section 3 describes the first version of the NOTAE KG and its visual interface. Finally, Section 4 concludes the paper also highlighting future works.

2   Related work

The world of digital humanities offers several different digital services and Web applications providing information about ancient documents, historical facts and geographical entities, many of which are based and/or can be downloaded as semantically enriched datasets following the LOD principles. The ones that can potentially be of interest for the historical context considered by the NOTAE project are listed below.

Pleiades[3] [7] contains data about geographical places (e.g., cities) relevant for the study of ancient literature and history. Pleiades allows the download of their data in JSON, CSV, and RDF/TTL format from a Github repository.

Papyri.info[4] is a research engine integrating several different ancient document

---

[3] see `https://pleiades.stoa.org/`
[4] see `http://www.papyri.info/`

databases. It models relationship between documents from different sources using RDF triples.

MANTIS is the semantically enriched database of the American Numismatic Society[5], a research institute devoted to study coins from all periods and cultures. In MANTIS, each record can be exported as JSON-LD, Turtle RDF/XML, and many others.

Open Context[6] collects, among other resources, archaeological reports. The data can be exported in tabular (CSV) or Geo-JSON form.

Trismegistos [4] is a metadata platform[7] for the study of texts from the Ancient World. It contains data about ancient documents, people, and places. Trismegistos does not currently provide a data export feature, but it does share its data with Papyri.info in the form of periodic database dumps.

EDH[8] is a search tool for Latin epigraphic data. It provides a data dump including RDF (inscriptions including prosopography) and GeoJSON formats.

All of them have embraced the LOD philosophy as a mean to connect their data to the broader digital cultural heritage infrastructure. Unfortunately, all of these efforts are jeopardized by the lack of a common vocabulary and the lack of integration, which are among the NOTAE KG goals.

In addition to these services, a wide set of websites is available, containing digital reproductions of ancient original documents. Many of these websites are provided by university departments, libraries and archives, and do not provide standard ways to access data. An approach similar to ours, with a narrower target, has been proposed in [3] for studying people (producing prosopography) who lived in Ptolemaic Egypt, based on the Trismegistos data. Finally, Pelagios [8] is an international consortium that uses the LOD approach and the Pleiades gazetteer to join up a variety of online resources that refer to places in the ancient world. One of the potential synergies of the NOTAE KG could be the participation to the Pelagios consortium.

3   The NOTAE Knowledge Graph

The NOTAE KG is built on top of the NOTAE Database, which consists of data manually inserted by domain experts as a result of their research activity in the context of the project. As the NOTAE Database is a relational database, an ETL process is executed daily to synchronize it with the KG.

The NOTAE KG makes use of multiple vocabularies, providing a set of classes and properties to describe the given domain. Vocabularies, expressed using the RDFS and OWL standards, make data integration easier by reducing diversity in describing things. The data model is defined incorporates our own vocabulary, described below, as well as the standard vocabularies listed below.

---

[5] see `http://numismatics.org/search`

[6] see `https://opencontext.org/`

[7] see `https://www.trismegistos.org/`

[8] Epigraphic Database Heidelberg - see `https://edh-www.adw.uni-heidelberg.de/`

Table 1. Classes

name	description
FROM NOTAE NAMESPACE	
Document	A document is a written text generated for pragmatic purposes. Main focus of our research is the documentary records (such as contracts, lists, petitions, official and private letters, et cetera) dating from the 4th to the 8th centuries and written in Greek or Latin, on hard materials (such as slate fragments and wooden tablets) or on soft ones (papyrus, parchment).
Symbol	Graphic symbols are meant as graphic entities composed of graphic signs (included alphabetical ones) drawn as a visual unit within a written text but communicating something other or something more than a word of that text.
Category	Category defines the ideal geometric shape that a symbol designs or which can be inscribed in, for example, square, circle, and cross.
Shape	Shape is a way of defining a symbol based on its apparent structure. It can be simple, in the case of a single and easily recognizable symbol, or complex, in symbols with interconnected shapes or groups of symbols.
Element	An element is what a symbol is made of. It can be alphabetical, if the symbol is made of letters, tachygraphic, in case of stenography, and simple strokes for geometrical purposes.
Position	Position indicates the place in the doc. where a symbol is drawn: it specifies both the side and the exact line(s) of the doc. on which the symbol appears.
FROM OTHER NAMESPACES	
schema:Place	Indicates entities that have a somewhat fixed, physical extension.
dbo:Language	Represents the language of an object.

Schema.org[9] was born in 2011 by a collaboration of Google, Microsoft, Yahoo!, and Yandex to mark up website content with metadata about the website itself. Here we use the schema.org class "Place.

DBpedia is one of the principal Linked Open Datasets[10], automatically extracted from multiple language versions of Wikipedia pages. We use the class "Language" from the DBpedia ontology.

FOAF[11] provides terms for describing people and organizations, documents associated with them, and social connections between people. We used properties from FOAF to connect images of symbols to the Symbol class.

geo[12] provides terms for specifying geographical coordinates. We used geo to define latitude and longitude of the instances of the class "Place".

An RDF schema is composed of properties and classes. Tables 1 and 2 list respectively classes and properties employed to define NOTAE KG.

The data contained in the knowledge graph are stored in a Blazegraph triple-storeand made available through a SPARQL endpoint[13]. For what concerns the visualization and the exploration of the NOTAE KG[14] (see Fig. 2), we adapted the tool developed in the context of the ARCA project [1].

---

[9] see http://schema.org/

[10] see http://dbpedia.org/ontology/ and http://dbpedia.org/resource/

[11] Friend Of A Friend - see http://xmlns.com/foaf/0.1/

[12] see http://www.w3.org/2003/01/geo/

[13] see http://notae-system.diag.uniroma1.it:9999/blazegraph/#query

[14] see http://notae-system.diag.uniroma1.it:8888. For reviewing purposes please use the following credentials, i.e., login: IRCDL2021@notae.it and password: ircdl2021.

Table 2. Properties

| name	description	domain	range	cardinality
FROM NOTAE NAMESPACES				
symbols	Connects a doc. to its symbols	Document	Symbol	multiple
provenance	Represents the provenance, i.e., the finding place of a document	Document	schema:Place	functional
shape	Represents the shape of a symbol	Symbol	Shape	functional
position	Represents the position of a symbol with respect to the document	Symbol	Position	functional
category	Is the category of a symbol	Symbol	Category	functional
element	Represents the graphical elements contained in a symbol	Symbol	Element	max(3)
certain_origin	Represents the certain origin of a document, i.e., its creation place	Document	schema:Place	functional
uncertain_origin	Is the uncertain origin of a doc.	Document	schema:Place	functional
undefined_origin	Is the undefined origin of a doc.	Document	schema:Place	functional
FROM OTHER NAMESPACES				
rdfs:label	Represents the human-readable name of a given resource	rdfs:Resource	rdfs:Literal	functional
dc:date	A point or period of time associated with the estimation of the document production	rdfs:Resource	rdfs:Literal	max(2)
dc:language	A language of the document	rdfs:Resource	dbo:Language	functional
dc:description	An account of the document	rdfs:Resource	rdfs:Resource	functional
foaf:depiction	Represents the image of a symbol	owl:Thing	dbo:Image	functional
geo:lat	The latitude of a SpatialThing	geo:SpatialThing	xmlns:basednear	functional
geo:lon	The longitude of a SpatialThing	geo:SpatialThing	xmlns:basednear	functional
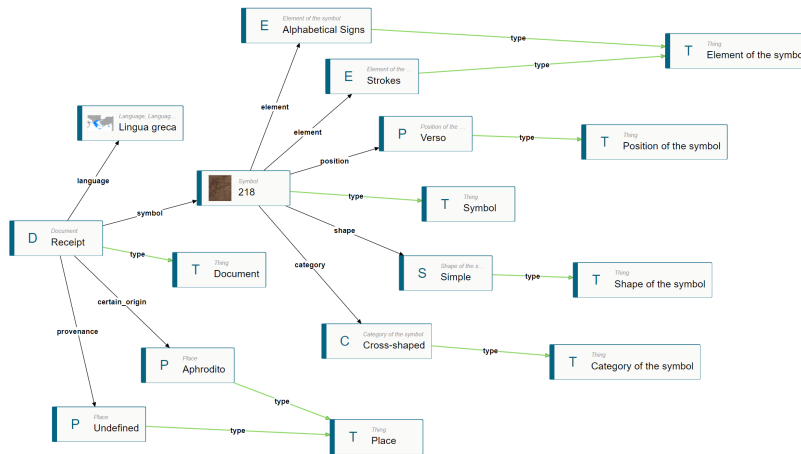
Fig. 2. NOTAE KG exploration example.

Let's now see an example of how we can deduce historical implications from the data in the knowledge graph. Starting from two terms of interest (see Fig. 2), such as the symbol with ID 218 and Aphrodito, we can find all the entities that bind the two. In this case, the "Receipt" document. The same method can be

applied to any class or instance of the graph. From this example, emerge how easy it is to find all documents that contain a symbol or to visually trace (by looking at the links that appear on the whiteboard) to the fact that a searched symbol is in a searched document.

4   Conclusions and Future work

In this paper, we presented the NOTAE Knowledge Graph (KG) as a tool to aid the researchers involved in the project and curious users to identify historical and geographical implications. The KG can be explored through a handy graphical Web application, which is already available for testing, and through a SPARQL endpoint, which allows the integration into external services.

Nonetheless, what has been presented in this paper, only represents the first version of a tool that is intended to fully cover the outcomes of the NOTAE project, by also integrating additional resources as those introduced in Section 2.

The final aim of the platform is to represent a reference in ancient so-called "pragmatic literacy", covering the main target of the project, i.e., the analysis and categorization of graphic symbols and related historical and sociological aspects.

References

1. Bernasconi, E., Ceriani, M., Mecella, M., Catarci, T., Capanna, M.C., Fazio, C.D., Marcucci, R., Pender, E., Petriccione, F.M.: ARCA. semantic exploration of a bookstore. In: Tortora, G., Vitiello, G., Winckler, M. (eds.) AVI '20: International Conference on Advanced Visual Interfaces, Island of Ischia, Italy, September 28 - October 2, 2020. pp. 78:1–78:3. ACM (2020)
2. Boccuzzi, M., Catarci, T., Deodati, L., Fantoli, A., Ghignoli, A., Leotta, F., Mecella, M., Monte, A., Sietis, N.: Identifying, classifying and searching graphic symbols in the notae system. In: Italian Research Conference on Digital Libraries. pp. 111–122. Springer (2020)
3. Broux, Y., Depauw, M.: Developing onomastic gazetteers and prosopographies for the ancient world through named entity recognition and graph visualization: Some examples from trismegistos people. In: SocInfo Workshops (2014)
4. Depauw, M., Gheldof, T.: Trismegistos: An interdisciplinary platform for ancient world texts and related information. pp. 40–52 (07 2014)
5. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. SEMANTiCS (Posters, Demos, SuCCESS) **48**, 1–4 (2016)
6. Ghignoli, A.: The NOTAE Project: a Research between Est and West, Late Antiquity and Early Middle Ages. Comparative Oriental Manuscript Studies Bulletin **5/1**, 27–41 (2019)
7. Gleyzes, M., Perret, L., Kubik, P.: Pleiades system architecture and main performances. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences pp. 537–542 (2012)
8. Isaksen, L., Simon, R., Barker, E.T.E., de Soto Can?amares, P.: Pelagios and the emerging graph of ancient world data. In: WebSci '14: Proceedings of the 2014 ACM conference on Web science. pp. 197–201. ACM (June 2014)

9. Mecella, M., Leotta, F., Marrella, A., Palucci, F., Seri, C., Catarci, T.: Encouraging persons to visit cultural sites through mini-games. EAI Endorsed Trans. Serious Games **4**(14), e3 (2018)
10. Parkes, M.B.: Scribes, scripts and readers: studies in the communication, presentation and dissemination of medieval texts. Hambledon Press London and Rio Grande, OH (1991)