NOTAE: NOT A writtEn word but graphic symbols

Eleonora Bernasconi¹, Maria Boccuzzi², Livia Briasco², Tiziana Catarci¹,
Antonella Ghignoli², Francesco Leotta¹, Massimo Mecella¹, Anna Monte³,
Nina Sietis⁴, Silvestro Veneruso¹ and Zahra Ziran²

¹*Department of Computer, Control and Management Engineering "A. Ruberti" -* Sapienza University of Rome, *Italy*

²*Department of History, Anthropology, Religions, Art, Performing Arts -* Sapienza University of Rome, *Italy*

³*Department of Humanities and Cultural Heritage -* University of Udine, *Italy*

⁴*Department of Human Arts and Philosophy -* University of Cassino and Southern Lazio, *Italy*

Abstract

Late antique and early medieval documents often include *graphic symbols*, i.e., graphic entities drawn as a visual unit whithin a written text, but communicating something other than a word of that text. The Project NOTAE aims to investigate them, in order to capture all the possible historical implications by studying their execution, models, cross influences, historical context and transmission. The project involves two approaches working in close collaboration: the historical, papyrological and palaeographical investigation and the IT research activity, which has developed the NOTAE System (fundamental tool, to fulfil the humanistic approach itself) and the NOTAE Knowledge Graph, testing also the possibility of identifying graphic symbols through software applications.

Keywords

Graphic symbols, Paleography, Digital Humanities, Image processing, Knowledge Graph

1. Introduction

The project NOTAE – *NOT A writtEn word but graphic symbols. An evidence-based reconstruction of another written world in pragmatic literacy from Late Antiquity to early medieval Europe –*, which started in July 2018, represents the first attempt to investigate the presence of graphic symbols in documentary records as a historical phenomenon from Late Antiquity to early medieval Europe: a crucial period that contributed in providing and shaping a set of graphic symbols and signs, from which later, in Carolingian age, the cultural élites of the latin West selected and reinvented the elements of their symbolic written communication [1].

Graphic symbols are meant as graphic entities, composed by graphic signs, including alphabetical ones, drawn as a visual unit within a written text, but communicating something other, or something more, than a word of that text. We say "symbol" and not sign, because there is no intrinsic prior relationship between the message-bearing graphic entity and the information conveyed by it. Even when it seems to us – men and women of the 21th century – simply and clear (as in Figure 1, left), the message is in any case to discover, because that graphic entity is an object of historical investigation.

The sources of the project are texts generated for pragmatic purposes: petitions, official and private letters, lists, receipts, authentics from relics, contracts and so on written on papyrus, wooden tablets, slates, parchment. In particular, legal documents enable to relate graphic symbols to illiterate people: the gradual introduction of signatures in the legal documentary practice meant an increasing use of graphic symbols not only by literate people writing their subscriptions in their own hands but also by illiterate contract partners or witnesses, who performed graphic symbols by their own hands in the empty space left for it in the line of their subscription written by the scribe or by a delegated third-party literate person.

In conclusion, NOTAE aims to investigate the graphic symbols in order to capture all the possible historical implications by studying their graphic execution as well as their models and cross influences, their context and transmission, with the purpose to frame also the category of illiterates in terms of gender and social status, for each significant period and region involved in the research, with the particular challenge represented by problematic evidences preserved in a problematic documentary transmission in a *longue durée*. Novel is therefore both the object and the perspective of investigation of the project [2].

Documents considered in the NOTAE project are available either in archives and libraries or through *digital reproductions* in public web repositories (e.g., https://papyri.info/). Identifying and classifying graphic symbols on such documents is not an easy task, requiring the experience and knowledge of an expert in the field. An expert who wants to identify and study graphic symbols in a specific document, inspects its digital reproduction together with associated bibliography. However, specific software tools can make this task easier by for example *(i)* make the digital reproductions easier to read, *(ii)* simplifying the insertion of reports, *(iii)* making it easier to extract information, and *(iv)* integrating the information with other data sources in order to contextualize the symbols and the containing documents.

This paper is organized as it follows. Section 2 introduces the objectives of the project and the expected tangible results. Section 3 describes the results already achieved by the project, Section 4 finally concludes the paper with challenges and final considerations.

2. Objectives and Expected Tangible Results

The NOTAE project pursues several goals within its general aim:

1. to provide an inventory as complete as possible of graphic symbols and a collection of their images, through the systematic inspection of all the documentary sources available for the period in question. For this purpose, a database will be primarily designed and implemented in order to work as a research tool of the Project;
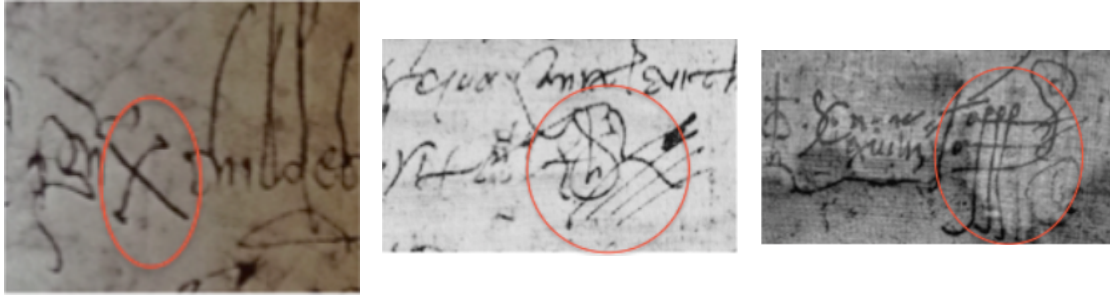
Figure 1: Examples of graphic symbols: *Left)* Lamorlaye, France (Merovingian Kingdom), 673 March 10: autograph diagonal cross (or letter χ) of Childebrando, an illiterate man. *Center)* Ravenna, Italy, 575: graphic symbol in complex structure at the end of the autograph subscription of a witness. *Right)* Hermopolis, Egypt, 561: graphic symbol in complex structure at the end of the autograph subscription of a greek notary.

2. to develop software tools to facilitate the task of identifying graphic symbols in a digital reproduction and find their positions in the document;

3. to find optimal solutions to overcome the limitations due to the original media, preservation conditions and the passage of time. Parts of the documents may be lost, and we deal with partial observations of ancient texts and symbols;

4. to conduct studies based on geographical and historical implications of the employment of such graphic symbols. For this purpose, additional software tools will be provided.

A graphic symbols' database has been designed to assist NOTAE experts. This database stores information about graphic symbols contained in the documents within the scope of the project. Documents are referenced by using identifiers that are globally recognized in the research community. Information does not only include their presence in a specific document, but also additional details such as comments about their usage. With experts continuously introducing new classifications of symbols, the database is progressively populated with graphic symbols, and this allows to increasingly refine the identification and classification process. One of the goals of this database is to be used as a reference to detect symbols in newly uploaded documents. Access to the database must be provided with two kinds of web applications: a back-office (dedicated to NOTAE experts) and a public website providing the paleography research community with access to completed and verified reports.

One of the advantage of using a database is simplifying queries and reporting tasks. In addition, data stored in a structured way allows for integration with other information sources to perform contextual analysis. As an example, documents are associated with a provenance place and an original archive. These geographical places can be projected on maps by using the information contained in historical place repositories such as Trismegistos places (https://www.trismegistos.org/geo/) or the Mapping Past Societies project of Harvard University (https://darmc.harvard.edu/). Once documents are geographically contextualized, it is possible to perform analyses about the employment of graphic symbols in specific geographic areas and time periods.

Item Type	#	Description
Graphic Symbols	3748	The main target of the project
Images of graphic symbols	3191	Number of graphic symbols with an image stored in the database
Documents	1510	Original textual unit including graphic symbols
Material Supports	1498	Physical support on which a document or part of it is preserved
People	914	Notaries and relevant people writing the documents
Bibliography	1254	Voices included to document graphic symbols

Table 1

Number of items in the NOTAE database as of 2022-03-24.

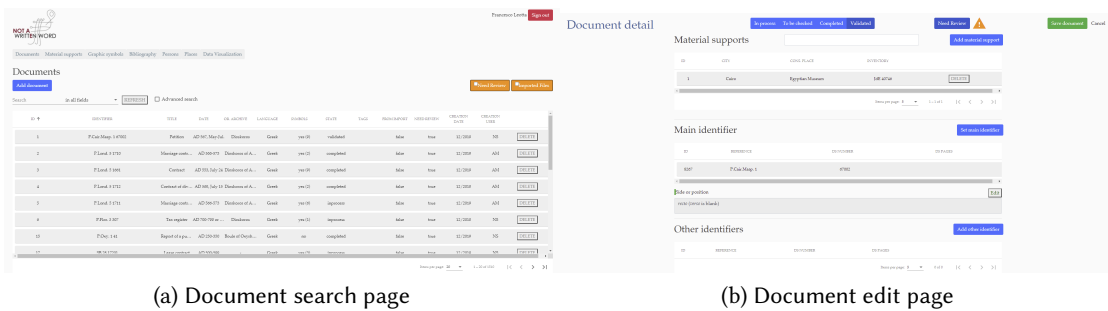(a) Document search page   (b) Document edit page

Figure 2: Screenshots from the NOTAE back-office web application.

3. Current Project Results

During these first years, several features have been implemented, satisfying most of the aims of the project.

One of the first core-tasks to be designed and built was the Graphic Symbols Database. The database structure is continuously updated and refined, following the NOTAE team's feedback. At bootstrap, this database was obviously empty. During this first period, experts have continuously populated the database with newly classifications of graphic symbols, and this allows to increasingly refine the identification and classification skills. At the moment, the database contains more than eighteen thousand identified and classified graphic symbols from documents, and it is still being updated. Table 1 shows the number of items added to the database for each category. Figure 2 shows to screenshot from the NOTAE Back-office web application.

In the following subsections we go into details of the aspects of the project more related to the computer science research community.

3.1. Symbol Identification

A tool has been designed, able to identify possible graphic symbols inside a document by matching previously identified symbols. This tool, based on template matching and the seminal
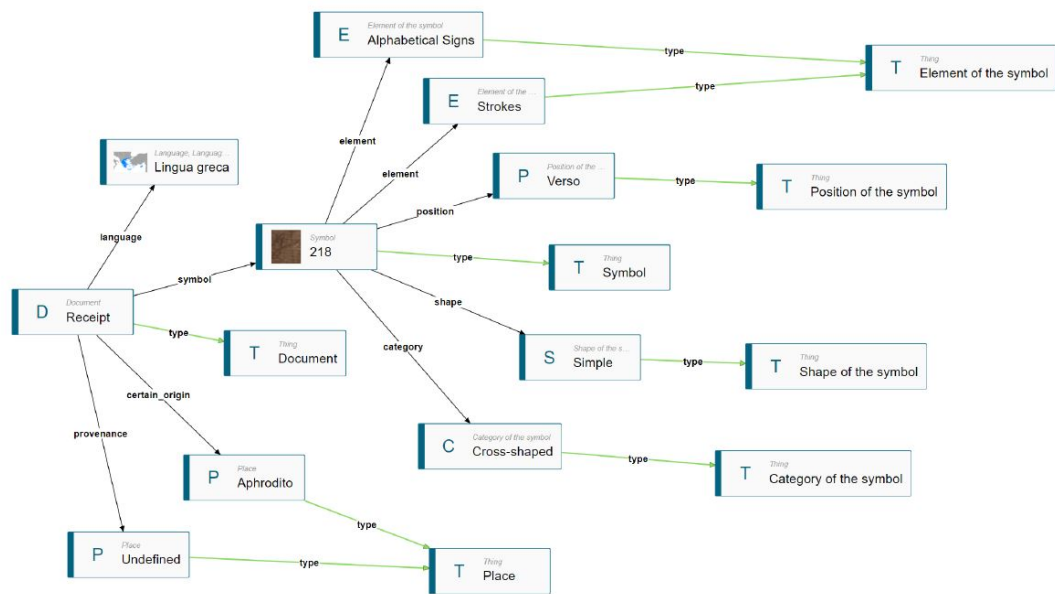
Figure 3: NOTAE KG exploration example. Starting from two terms of interest, such as the symbol with *ID 218* and *Aphrodito*, we can explore all the resources that bind them. In this case the *Receipt* entity.

OPTICS clustering algorithm [3] is intended to help the researcher who can discard wrongly identified symbols and select new ones. The tool has been published in [4]. The very same tool can be used by the final user to search symbols by drawing them using, for example, a touch screen. A second version of the tool, relying on deep neural networks instead of clustering has been published in [5].

3.2. Digital Reproduction Enhancement

Graphic symbols' sources are papyri, wooden tablets, slates and parchments. Depending on the original media, the preservation conditions and the passage of time, parts of the documents may be ruined or, in the worst case, lost. An image enhancement processing step could be necessary. In addition, a wrong assumption might make think that images are unique, but instead it is likely to have duplicates. Then, various pre-processing techniques have been implemented.

3.3. NOTAE Knowledge Graph

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.

A Knowledge Graph (KG) [6, 7] 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 Knowledge Graph [8] on top of the NOTAE Database, 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, we propose a graphical user interface which allows researchers to explore and search relations and connections between resources within the NOTAE KG. An example of such exploration is shown in Figure 3.

4. Conclusions

In this paper, we introduced the NOTAE ERC Project. The goal of the project is to study the employment of graphic symbols in documents from Late Antiquity to early medieval Europe. The project, ending in December 2023, has already achieved important results. Still, the team behind the project is working on several aspects still to tackle.

As a first point, the automatic symbol identification features are still under testing and only provided as experimental functionalities. Further tests are ongoing in order to provide the recommendation functionalities to NOTAE researchers, through the NOTAE Back-office web application, and to the full paleography research community in the form of search functionalities.

Secondarily, the integration with geographical information sources must be finalized, by also acquiring maps and allowing for advanced search functionalities. Also, the visual analytics tools allowing for the exploration of the database according to the geographic and time dimensions are currently under implementation.

Finally, the public website, which will allow the community to explore contents and items verified by NOTAE curators is currently in the design phase, with the goal of making it available to the public at the end of 2023.

Acknowledgements

References

[1] A. Ghignoli, The notae project: a research between est and west, late antiquity and early middle ages, Comparative Oriental Manuscript Studies Bullettin 5/1 (2019) 27–39. doi:https://doi.org/10.25592/uhhfdm.185.

[2] D. Internullo, Magis intellegi quam legi. segni e simboli grafici cristiani nel mediterraneo tardoantico e altomedievale, Storicamente 65 (2019-2020) 15–16. doi:https://doi.org/10.12977/stor811.

[3] M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander, Optics: Ordering points to identify the clustering structure, ACM Sigmod record 28 (1999) 49–60.

[4] M. Boccuzzi, T. Catarci, L. Deodati, A. Fantoli, A. Ghignoli, F. Leotta, M. Mecella, A. Monte, N. Sietis, Identifying, classifying and searching graphic symbols in the notae system, in: Italian Research Conference on Digital Libraries, Springer, 2020, pp. 111–122.

[5] Z. Ziran, E. Bernasconi, A. Ghignoli, F. Leotta, M. Mecella, Accurate graphic symbol detection in ancient document digital reproductions, in: International Conference on Document Analysis and Recognition, Springer, 2021, pp. 147–162.

[6] L. Ehrlinger, W. Wöß, Towards a definition of knowledge graphs, SEMANTiCS (Posters, Demos, SuCCESS) 48 (2016) 2.

[7] S. Ji, S. Pan, E. Cambria, P. Marttinen, S. Y. Philip, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems (2021).

[8] E. Bernasconi, M. Boccuzzi, T. Catarci, M. Ceriani, A. Ghignoli, F. Leotta, M. Mecella, A. Monte, N. Sietis, S. Veneruso, et al., Exploring the historical context of graphic symbols: the notae knowledge graph and its visual interface, in: IRCDL, 2021, pp. 147–154.