Digital Patrology: A Theory and Critical Database Approach in Digital Libraries for converting historical scripts into texts and sources of modern perception

Nikolaos MYRIDIS, Ass.Professor, Aristotle University of Thessaloniki, Greece

Keywords: *Digital Libraries – Scripts collections – Patrology.*

CHNT Reference: Myridis N. (2021). 'Digital Patrology: A Theory and Critical DataBase Approach in Digital Libraries for converting historical scripts into texts and sources of modern perception, in CHNT – ICOMOS Editorial board. *Proceedings of the 26th International Conference on Cultural Heritage and New Technologies.* Heidelberg: Propylaeum.

1. Introduction

In a previous work a transformation of ancient scripts through AI was presented (Myridis and Sarakatsianou, 2020). Following that work, in the same context, a second perspective is introduced herein.

The principles of data mining cope with a huge amount of information pieces –individual or functionally related each other. In this context the sources of information diversify according to three major conditions. Ad hoc, the source may produce *variable*, *static* or *quasi-static* information. The category of variable information includes applications where the information regards domains with a strongly dynamic environment. The latter is the case of contemporary scientific knowledge. In the static information frame the thematic area of interest is related to arbitrary kind of historical scripts or findings. It could be referred for instance the ancient or medieval literature, the premature scientific knowledge or the archaeological findings. Finally, the quasi-static mode of information regards assemblies carrying unknown, vital data. These assemblies result from new historical, scientific, research sources which append their content to the already existing information realm. In fact, these vital data are mainly semantic data (i.e. content data) that by their nature enclose deep meanings that cannot be immediately understood, without the use of knowledge-based system. Semantic technology can uncover latent correlations and produce new scientific knowledge. Of course, provided that the content of a text, on the one hand, has been enriched with additional meaning, and on the other hand, a first semantic classification has been made based on a visible and common semantic content (for example, grouping based on specific or general subjects, historical events, controlled vocabularies, or keywords/free vocabularies). These rich semantic data create a new data-web that can be maximally exploited by the dynamic use of semantic web technologies.

This paper launches a new perspective for data mining in fields of distributed information with quasi-static form, which can be expanded to variable sources. More precisely this perspective optimally analyzes phenomena related to data management, information extraction and archiving in literature and relative scripts. Initially, a description of the information of interest is constructed, using the realm of relevant scripts. The substrate for this descriptive process is based on set theory (Cenzer et. al.,

2020) and databases elements (Connolly and Begg, 2015). Therefore, the analysis of the thematic content results. The fundamental key in this process is the definition of cardinal characteristics (. Thus, crucial rules and terms are introduced resulting to information extraction with optimal *mining order*. In this presentation a first application of the aforementioned analysis is additionally given. This application concerns as Data of Interest (DOI) a corpus of Byzantine Scripts Collection (BySC) mainly written in Greek language, namely the so called 'Patrology'.

2. Data of Interest (DOI): Patrology

The term *Patrology* is originated by the Greek word 'Πατρολογία' which refers to the whole of scripts of Christian Church's Saint Fathers. In this work the scripts of Orthodox Patrology are considered. The chronological borders/windows of interest vary according to each script's collection or application etc. Most known edited collections of Patrology are: (1) Patrologia Graeca & Patrologia Latins of Migne, (2) Sources Chretiennes of CNRS, France. An excellent reference work for Greek Orthodox Patrology is given in Christou (2004). The fundamental aspects concerning Patrology for the herein proposed project are: (1) Author (name), (2) Title of the script (3) Subject.

3. Theory and Critical Database Approach

The overall work is preceded by the construction of a new theory for database approach. This theory is also preceded by a set theory foundation. The latter aspects are considered through the lens of data mining. Fundamental spaces are defined for the construction of the overall DB which hosts the DOI of BySC. These spaces are related to the above-mentioned aspects of Patrology (§2) and they are (Fig.1(a)):

1. **Π-space.** A space related to the entire work (scripts) of an individual author. This space does not contain the corresponding texts but indexing to the semantic information. It may also contain information concerning conventional bibliographic description (i.e., descriptive, administrative, technical, etc.) of the texts.
2. **Θ-space.** A space defining the thematic content of a specific unity. This space is related to the semantics of the texts' contents and their semantic classification into wider or narrower semantic clusters.
3. **π-space.** This space includes the per se total work of an individual author. This space contains the texts of the works.

The above-mentioned spaces refer to the following basic information triad: *bibliographic metadata–semantic metadata–content*. There are also three more sub-spaces, namely *ε, τ* and *δ*. The construction of the BySC DataBase (Digital Patrology) is based on the aforementioned generating spaces. The proposed scheme is shown in Fig.1(b).

4. The Byzantine Scripts Collection (BySC) application

The database constructed in this work has been structured on the theory of mathematical spaces. According to normalization level, the perceived database is 5NF (5th Normal Form). The cardinal relations, resulting from the generating spaces, describe the interaction between *Π* and *Θ* spaces. These are m:m relations. The relations of *π*, *δ* and *τ* subspaces are 1:m, 1:m and 1:1 respectively.

Following the DB design a specific application is presented in this section concerning the BySC of Patrology's scripts. A huge part of this collection could be characterized as theological scripts. In this context scripts from the early centuries of Christianity (1st & 2nd centuries AD) until the first decades of modern Greek nation (~1830) are considered. The application is implemented in the MSAccess environment. This fact tends to be a critical choice for diverse applications in the same context. The tables corresponding to the generating spaces are depicted as a star diagram in Fig.2. Finally, a snapshot of the constructed BySC Database is given in Fig.3. The users of this DB can thus read the scripts of Patrology through the implementation of the BySC DB without having the heavy -and maybe desperate- work of conventionally read the overall(!) scripts' collection of Patrology under the conventional rules of an ancient era. Moreover, this work of Patrology's study gains even more advantages such as: (1) Modern computational speed, (2) Search using modern terms, etc.

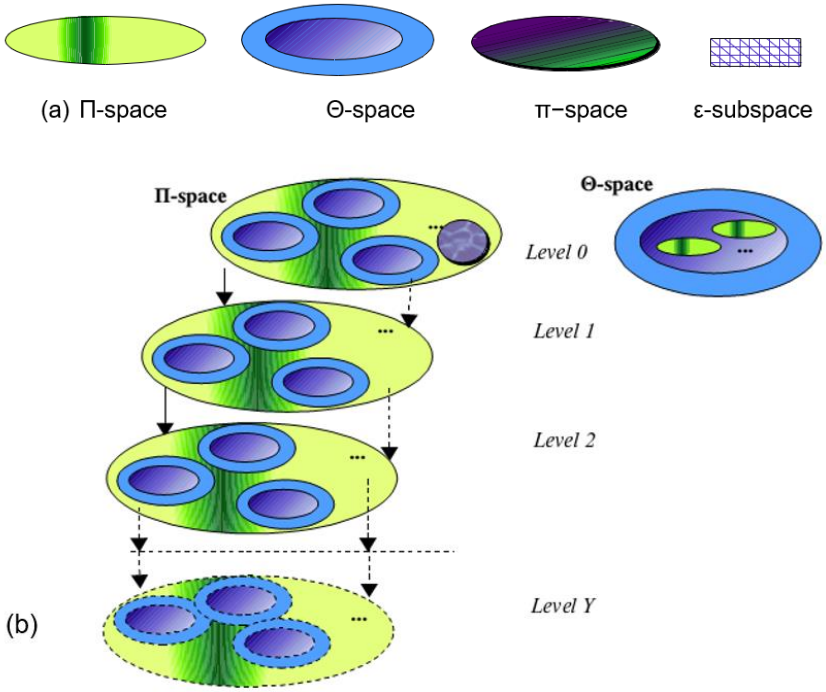

Figure 1. (a) The generating spaces of BySC DataBase (b) The proposed scheme of the BySC Database

5. Conclusion

In this paper, a database indenting to store and exploit information of script's collection, with focus on Patrology (a Byzantine scripts collections), is designed and constructed thus forming a Digital Patrology. The cardinal design is deduced from bibliographic-based information and semantic information. The relevant mathematical spaces and subspaces (that is, discrete spaces for description, semantics, and textual data) are defined by the criterion of mining order maximization. The herein presented implementation and the according methodology can critically benefit fields relative to literature, as linguistics, sociology, theology, history etc. It can also become the main source of input for a machine learning application that will export new scientific knowledge.

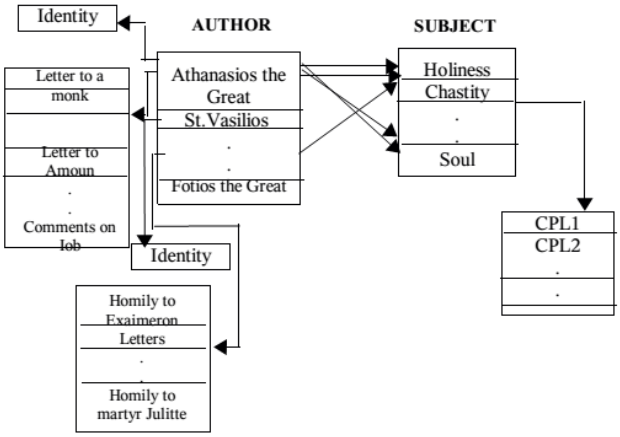

Figure 2. Tables of the generating spaces as a star diagram.

Figure 3. The final BySC DB in MS Access

Author Contributions

Conceptualization: Myridis Nikolaos
Data curation: Myridis Nikolaos
Formal Analysis: Myridis Nikolaos
Funding acquisition: Myridis Nikolaos
Investigation: Myridis Nikolaos
Methodology: Myridis Nikolaos
Project Administration: Myridis Nikolaos
Resources: Myridis Nikolaos
Software: Myridis Nikolaos
Supervision: Myridis Nikolaos
Validation: Myridis Nikolaos
Visualization: Myridis Nikolaos
Writing – original draft: Myridis Nikolaos
Writing – review & editing: Myridis Nikolaos

References

Cenzer, D., Larson, J., Porter, C. and Zapletal, J., 2020. *Set Theory and Foundations of Mathematics: An Introduction to Mathematical Logic: Volume I: Set Theory*

Christou, P. (2004). *Greek Patrology, vols.A-E,* , 3rd ed, Kyromanos Pbls. (in Greek)

Connolly, T.M. and Begg, C.E. (2015). *Database systems: a practical approach to design, implementation, and management*. Pearson.

Myridis, N., and Sarakatsianou, D. (2020). 'Climax: A historical script´s AI transformation for human behavioural estimation'. W. Börner and CHNT Organization Committee. *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, Vienna.*