

Alim (Archivio della Latinità Italiana del Medioevo)

A Digital Library with text analysis tools (*Lexicon*)

Elisabetta Bartoli

Università di Siena

elisabetta.bartoli@unisi.it

Abstract

This article surveys the development of *ALIM* (Archivio della Latinità Italiana del Medioevo) from its origins in the 1990s to the present. Established under the auspices of the Unione Accademica Nazionale to collect Italian-area Latin texts for the prospective compilation of a dictionary of European Medieval Latin, *ALIM* has received sustained institutional support, including PRIN funding. The paper highlights the project's defining features, including the integration of documentary and literary texts, curated Collections, and a sophisticated search syntax, as well as subsequent enhancements such as *ALIM Plus*, which enables users to include or exclude texts beyond the original chronological (7th–14th centuries) and geographical (Italian area) scope. As a digital library, *ALIM* provides open-access texts in multiple formats, including TEI-encoded XML, while also supporting advanced linguistic analysis. In particular, the *Lexicon* tool, designed by F. Stella and developed by L. Tassarolo, facilitates investigations of lexical overlap and exclusivity across corpora. Closely associated with the *ALIM* project is *EVT* (dev. R. Rosselli Del Turco), a tool that has become widely adopted for the creation of digital critical editions.

Keywords: Digital Library; XML-TEI; Medieval Latin; Linguistic Analysis

Il contributo ripercorre la storia di ALIM, acronimo di Archivio della Latinità Italiana del Medioevo, dalle sue origini negli anni '90 del secolo scorso ad oggi. Promosso dall'Unione Accademica Nazionale per reperire materiali di area italiana analizzabili per il progetto di un dizionario della latinità medievale europea, ALIM è un progetto che ha ricevuto vari finanziamenti come PRIN. Il saggio ne illustra sinteticamente le peculiarità (tra cui la presenza di testi documentari, testi letterari e di collezioni), la sintassi di ricerca, le implementazioni sopraggiunte nel corso degli anni come il recente ALIM Plus, che permette di includere o escludere dalla ricerca testi fuori dai parametri cronologici e geografici previsti nel progetto originale (VII-XIV sec. e testi latini di area italiana). ALIM è una Digital Library che mette a disposizione dell'utente testi in open access, scarabilli in vari formati – incluso l'XML con markup TEI - ma consente anche ricerche linguistiche approfondite: grazie a Lexicon, strumento progettato da F. Stella e realizzato da L. Tassarolo, permette indagini di sovrapposibilità o esclusività del lessico anche tra due corpora. Legato al progetto ALIM è stato

sviluppato EVT (svil. R. Rosselli Del Turco) uno strumento ormai largamente usato per le edizioni critiche digitali.

Parole Chiave: Biblioteca digitale; XML-TEI; Latino medievale; Analisi linguistica

1. History

*ALIM*¹ is a searchable Digital Library that brings together texts written in Latin and composed in Italy between the 8th and the 15th centuries. The *ALIM* project dates back to the 1990s and is still ongoing. It was initially developed to complement the linguistic information that, for earlier centuries, can be found in the *Grande Dizionario della latinità medievale italiana*, promoted by the Unione Accademica Nazionale and edited by Francesco Araldi and Pasquale Smiraglia.²

Linked to the study of linguistic materials promoted by *ALIM*, over the years several updates to the *Lexicon latinitatis italicae Medii Aevi*, edited by Antonio De Prisco, have been published. These appeared in print in the *Bulletin du Cange* and are available online among Brepols' Dictionaries.

From its beginnings, *ALIM* has therefore had a strong linguistic focus. As a digital resource, one of its primary aims is to produce analysable materials for the project promoted by the Union Académique Internationale of Brussels: a dictionary of European medieval Latin intended to integrate and complement the *Glossarium ad Scriptores Mediae et Infimae Latinitatis* devised by Charles du Fresne, Seigneur du Cange (Amiens 1610 – Paris 1688), whose first edition dates back to 1678. The importance of the *ALIM* project has also been recognized at the national level: *ALIM* has repeatedly received funding as a PRIN project—in 1996 (PI: Giovanni Polara), 1997, 1999, 2001, 2003, 2009 (PI: Gian Carlo Alessio), 2005 (PI: Antonio De Prisco), 2012 (PI: Francesco Stella), and 2017 (PI: Edoardo D'Angelo). At the conclusion of the 2017 PRIN, the conference *ALIM and Beyond* was held in Venice in January 2020.³ In the *Publications* section of the *ALIM* website, a selected bibliography of research outputs related to the project and produced by the research team is available.⁴

One of the first presentations of the *ALIM* project is due to Gian Carlo Alessio, who illustrated its mechanisms at the 8th Meeting on Humanities Computing, held in 1999 at the University of Verona [2]. Since that now distant date, numerous changes have taken place, and the members of the project's scientific team have progressively developed a series of reflections, some of which have led to substantial innovations ([14],[15]). These aimed both at greater flexibility in the analytical tool and at the necessary alignment with best practices that were being codified in the

¹ <https://alim.unisi.it>.

² This work was reprinted as *Latinitatis Italicae Medii Aevi lexicon* in [3].

³ <https://alim.unisi.it/alim-and-beyond-workshop-internazionale-venezia-ca-foscari-27-28-gennaio-2020/>

⁴ <https://alim.unisi.it/pubblicazioni/>

field of computing. This latter aspect was central to the transition from *ALIM1*⁵ to *ALIM2*⁶, which took place in 2016. The proprietary platform (IBM Lotus Domino) was dismissed and an open-source system with code was adopted.⁷ The website (Figure 1) has a section dedicated to public, freely accessible documentation.⁸



Fig. 1 - *Alim* Homepage

From that moment on, texts that in the first version of *ALIM* had been marked up in HTML have been subjected to declarative TEI markup, with the possibility of freely downloading – alongside TXT, HTML, and PDF files – the XML format of every textual resource present in the Digital Library; in the previous version of *ALIM*, it was possible to download a Word file. The choice of TEI markup, widely discussed by the research team, has made *ALIM* capable of interacting with major institutions and has facilitated the exchange of materials, as occurred with *CLARIN*⁹ and the *Corpus Corporum*,¹⁰ in compliance with interoperability practices that are now an essential element for the sustainability of digital resources.

Among the enhancements to the search engine that have made this tool capable of increasingly sophisticated and precise analyses, mention should be made of the integration of *Lexicon*,¹¹ and the development of the EVT program (edited by Roberto Rosselli Del Turco). EVT was initially

⁵ Alim1 is still visible here <http://www.alim.dfl.univr.it/>.

⁶ This is exemplarily illustrated in [12].

⁷ <https://it.alim.unisi.it/architettura-informatica-della-piattaforma-alim2/>

⁸ <https://alim.unisi.it/documentazione/>.

⁹ <https://www.clarin-it.it/>. See [7].

¹⁰ <https://mlat.uzh.ch/browser?path=/>

¹¹ See below, 4. Lexicon.

applied to the edition of Latin texts from the Germanic area (the *Edictum Rothari*) and has now reached version beta 2.1,¹² becoming one of the most widely used tools for digital editions.¹³

Alongside its attention to the technological domain, *ALIM* has co-financed the production of numerous *editiones principes*, later published in the ENTMI series of SISMEL (*Edizione nazionale dei testi Mediolatini d'Italia*). The Latin text, without critical apparatus, has been made available in the Digital Library and represents a scientific enhancement of great importance.¹⁴

ALIM is a resource appreciated both nationally and internationally, to the extent that it is cited among Digital Libraries on numerous websites, including those of the American Academy in Rome,¹⁵ the University of Wisconsin–Madison,¹⁶ the *Biblissima* Project,¹⁷ and NUME (Research Group on Medieval Latin). A selection of such citations can be found in the Publications section of the website.¹⁸

2. The texts of the *DL*, the choice of source texts, and search operations

Born out of the need for lexical indexing for the creation of lemmas, *ALIM* is a Digital Library characterized by a search engine and by the dual presence of documentary sources and literary sources – a distinctive feature that makes this tool unique in the current landscape of digital libraries. The criteria for entering texts into the DL have always been inspired by the greatest possible fidelity to the source edition from which the text is taken: for this reason, in addition to divisions into chapters and paragraphs, paragraph breaks, line breaks, page changes, and every other detail relating to the layout of the original edition are preserved. This is because *ALIM* was also conceived as a tool that, besides allowing certain search operations, enabled users to download and have at their disposal a faithful copy of a text's critical edition.

ALIM, like any digital library, must comply with the copyright regulations of the original editions, regulated at the European level and made increasingly stringent over the years.¹⁹ For this reason, the texts included in the DL come from editions that are more or less recent, if not produced within the project itself. This is a highly relevant factor on the textual and linguistic

¹² <https://visualizationtechnology.wordpress.com/2026/01/20/evt-2-beta-2-1-available-for-download/>.

¹³ For example, see the critical edition of *De rebus Siculis Carmen* by Peter of Eboli, produced for *ALIM* by Fulvio Delle Donne [9].

¹⁴ Among these, for example [10],[8],[4],[5],[1].

¹⁵ <https://www.aarome.org/it/node/701>.

¹⁶ <https://search.library.wisc.edu/database/UWI18143>.

¹⁷ <https://baobab.biblissima.fr/fr/ressource-909>.

¹⁸ <https://alim.unisi.it/pubblicazioni/>.

¹⁹ Directive (EU) 2019/790 on Copyright in the Digital Single Market. The topic has already been discussed in Ferrarini [11]: 11-12.

level and must be taken into account when using the texts and commenting on the results obtained through searches: the source editions are clearly indicated to allow users to carry out quick and essential checks.

The more than six thousand documents in the documentary section reproduce texts derived from major diplomatic collections, including Cava dei Tirreni, Lucca, Trani, Arezzo, Bari, Naples, and others. Users can select or search for a document according to the collection that preserves it, the date, or the place where it was where it was produced.

The texts in the literary section – fewer than one thousand – can be selected by author, title, typology (verse or prose), period, or literary genre. This last classification is not without problems, due to the different medieval sensibility in the perception of literary genres compared with the modern era and to certain overlaps; nevertheless, it is functional for research purposes and has therefore been retained.

3. Search Operations: *ALIM* as a Precision Tool

Once it has been determined whether to query the entire textual section ("documentary" or "literary") or only a subset of texts selected according to the criteria outlined above, a range of searches may be performed, either for individual words or for pairs of terms in proximity. Wildcard characters and logical operators (AND / OR) are supported. Beyond its importance for lexicography and historical linguistics, *ALIM* constitutes an invaluable tool for the compilation of critical apparatuses and the preparation of critical editions, particularly in the case of certain literary genres for which it offers a highly specialized corpus. This corpus includes several first editions produced within the framework of the project itself. Consequently, *ALIM* is widely used by philologists, even though its use is not always explicitly acknowledged in scholarly publications.

The classification of texts by literary genre enables targeted searches, which are often essential for lexical analysis/verification or for the selection of textual variants. By way of example, while preparing an edition of the *Rationes* by Bernard, a twelfth-century master of epistolography, it became necessary to verify a phrase printed by the previous editor, Ludwig Rockinger [14], as *petitio hortoria*, a reading that is semantically implausible [7]. The manuscript, however, reads *hortatoria*, a relatively rare term. A search within *ALIM*'s rhetorical sources immediately provides a highly relevant parallel in a technical rhetorical treatise: the phrase *petitio hortatoria* occurs in Bono of Lucca's *Cedrus Libani* [16]. Numerous further examples could be adduced. In rhetorical and ars dictaminis contexts, for instance, the term *appositio* carries a specific technical meaning: restricting the search to the rhetorical section yields four relevant occurrences, whereas a non-selective search produces seventeen occurrences in *ALIM*, not all of which are significant. In an even larger digital library, such as *Corpus Corporum*, the same query would return 145 items requiring individual examination.

4. Lexicon

In 2016, *ALIM* incorporated a tool of considerable importance for lexical research, both in terms of its analytical capabilities and its ease of use: *Lexicon*.²⁰ Developed by Luigi Tassarolo on a project by Francesco Stella, it continues the tradition of the *Analisi lessicale* software by Carboni-Poli for Windows 98. It can be accessed directly from *ALIM*'s homepage, where it allows operations to be performed on the entire digital library, encompassing both literary and documentary sources. *Lexicon* is capable of generating indexes, frequency distributions, collocational data, and concordances for individual texts (Fig. 2); most importantly, it enables systematic comparisons between two texts or between two *corpora*, assessing the degree of overlap or exclusivity of linguistic forms within the materials analysed (Fig. 3).

In addition to the documentation provided on the tool's homepage, a detailed guide has been published by Alessandra Di Meglio.²¹ The integration of *Lexicon* has significantly enhanced and further refined the possibilities for linguistic analysis within *ALIM*.

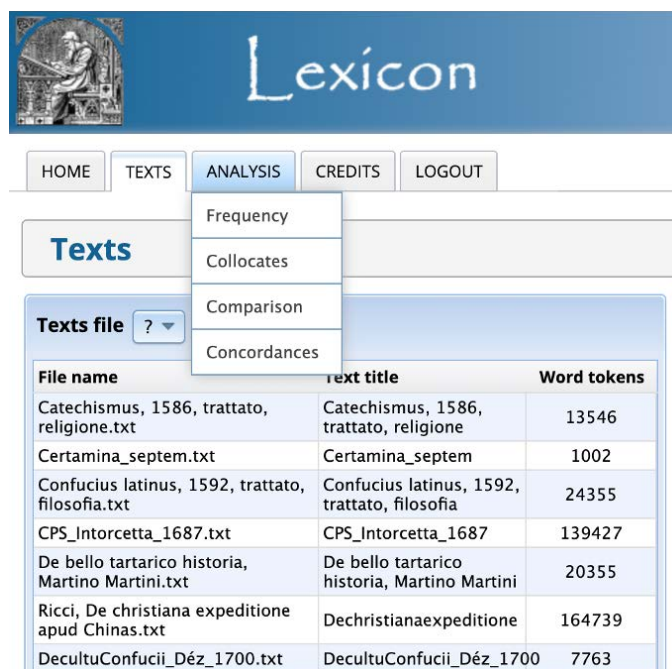


Fig. 2 - Lexicon's set of analysis tools

²⁰ <https://www.lexicon.unisi.it/public/>.

²¹ https://dhow.hypotheses.org/files/2019/02/Tutorial_lexicon_dimeglio.pdf.

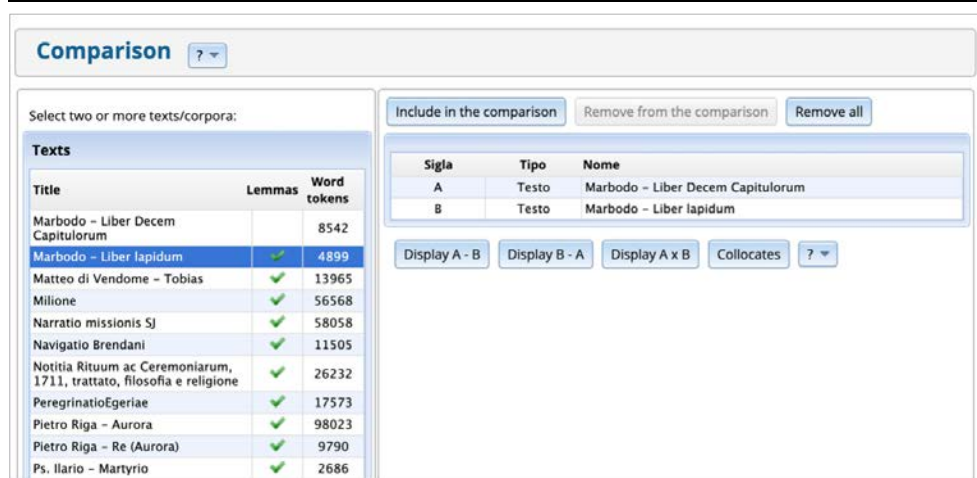


Fig. 3 - Lexicon's comparison between texts

5. The collections

The collections, visible on the homepage (Fig. 4), represent a section of the *ALIM* website where the two areas – the documentary and the literary – enter into dialogue. They were created to go beyond the literary genres catalogued in the DL and aim to provide examples of thematic and interdisciplinary pathways that are useful not only from a linguistic perspective but also from a historical-literary one.



Fig. 4 - Collections displayed on the Homepage

They were designed and set up for the first time in 2016, when the transition from *ALIM1* to *ALIM2* – also thanks to the WordPress model – allowed for greater flexibility in the autonomous management of these materials, without the need to request costly external intervention each time.

New Collections are created whenever there is a desire to highlight the addition of new materials to *Alim's DL*. The Collections are also conceived as a useful teaching tool and are intended for an audience made up not only of specialists who use this resource for research, but also of school

teachers and students, since open-access resources are part of the collective heritage and should be disseminated as widely as possible.

6. Present and future: synergy with other funded projects

A long-standing resource such as *ALIM* has, over the years, necessarily broadened its horizons; some innovations initially conceived to meet the needs of a funded project can become a defining strength in their own right. This has been the case with the *editiones principes*, an innovation introduced as early as 2012 and subsequently expanded over time. The opportunity to equip the digital library (with unpublished and scientifically vetted texts) constitutes an absolute and increasingly significant value – *ALIM* has also served as a testing ground for digital critical editions with highly sophisticated markup, such as the one produced by Paolo Monella for *Orso Beneventano* [12].²²

The section devoted to unpublished texts is set to grow further thanks to the synergy with several currently funded projects, listed below, in which *ALIM* is involved. For this reason, it will be necessary to further refine the guidelines for XML encoding, taking into account this textual typology—which lacks a pre-existing source edition—as well as a set of shared criteria to be followed for the markup of *editiones principes* to be included in the DL.²³

With a view to expanding the DL and promoting the sustainability of the resource, even when it is not independently funded, *ALIM* is involved in numerous projects as a platform that will host digitized texts or first editions produced by research groups. Only those projects that are currently ongoing or have recently concluded are mentioned below, listed in alphabetical order.

DICTAMINA, a Franco-Italian-German project promoted by the École française de Rome, RWTH Aachen University, the University of Siena, and Ca' Foscari University of Venice; it envisages the production of first editions or transcriptions of collections of Medieval Latin letters (2022–2026).²⁴

ELA – Eurasian Latin Archive: University of Siena, PI Francesco Stella. The project involves data extraction and the identification of lexical thresholds in the Latin language (Latin texts relating to Asia from the 13th to the 17th centuries).²⁵

²² A simplified markup version can be found on the *ALIM* website (<https://alim.unisi.it/dl/resource/9928>).

²³ https://www.dropbox.com/scl/fo1ag6q5vxd97fo1ukwk1gz/teiHeader_template.xml?rlkey=x9ij9da9n2pf8fbqj6umzlix&e=1&dl=0.

²⁴ <https://dictamina.hypotheses.org/> Sono stati immessi in *Alim* Pietro da Loro, Alectorius (ed. M. Vescovo); *Dictamina* ms. Parigi BNF 8661 (ed. M. Vescovo); *Dictamina* ms. Roma, Bibl. Angelica 514 (ed. D. Internullo). In preparation: Alberto di Asti, Flores dictandi (ed. V. G. Nitti).

²⁵ <https://ela.unisi.it/>.

HISTORIALE MARE and *IMPERI SITUS*, University of Suor Orsola Benincasa and University of Basilicata, PIs respectively Edoardo D'Angelo and Fulvio Delle Donne.²⁶ These are two funded PRIN projects that involve the study of numerous historiographical texts, analyzed in terms of sources, theory, training, oratory and rhetoric, authorship, and literary development between the Middle Ages and Humanism–Renaissance (2025–2028). The *editiones principes* and some digitizations will be incorporated into *ALIM*.²⁷

MEWIL– Medieval Women In Letters and *EPISTOLAE*, F-Cur UNISI, PI Elisabetta Bartoli, concluded in June 2025,²⁸ made it possible to create the *Epistolae mulierum* Collection, comprising new materials and others drawn from the *Epistolae* website (<https://epistolae.ctl.columbia.edu/>), created and directed by Joan Ferrante at Columbia University and now transferred, by agreement with her, to the UNISI server.²⁹ In the coming months, the Latin texts of *Epistolae*, already available for consultation, will be converted into XML and uploaded to *ALIM* to enable querying.

UMAS – Università Medievali di Arezzo e Siena (Region of Tuscany grants), PI E. Bartoli. As a continuation of Francesco Stella's project on the *Studium* of Arezzo,³⁰ it envisages for *ALIM* the digitization of medieval documents relating to the two universities and the first edition of several epistolary collections.

7. *ALIM plus*

Some of the projects mentioned above required greater flexibility regarding the texts eligible for inclusion in the DL, which in its original form provided only for texts from the 8th to the 15th centuries, written in Latin and produced in Italy. While it was desirable for the DL to be enriched with texts from other geographical areas and outside that limited chronological range, it was equally important to safeguard the original project without altering its nature. For this reason, *ALIM Plus* was conceived and developed: an extension that allows the user, through a simple flag, to include or exclude from the search those texts that do not comply with the original parameters of the project (Fig. 5). Users can choose whether to restrict the search to the *ALIM* Archive (Italy, 8th–15th centuries) or to include the other texts as well.

²⁶ <https://imperisitus.unibas.it/>.

²⁷ The project includes the following editions: *Commentum ad Vitam aut facta Alphonsi I regis Antonii Panormitae di E.S. Piccolomini* (by A. Iacono); *Vita Altobelli Claravallensis Tudertini di Fabrizio Delfino da Acquasparta* and *Vita Raginaldi principis* written by Peter of Blois (by E. D'Angelo); *Historia Manfredi regis* written by Niccolò di Iamsilla (by T. De Angelis), *Denominatio regionalum Imperii* written by Goffredo da Viterbo (by E. Pacia).

²⁸ <https://medioeva.unisi.it/mewil-medieval-women-in-letters/>.

²⁹ <https://epistolae.unisi.it/>

³⁰ <https://centroideugsu.unisi.it/lo-studium-arerino-del-200/>.



Fig. 5 - Alim plus

References

- [1] Alessio Gian Carlo. ed. 2023. "*Dictamina*" di Bonfiglio d'Arezzo. SISMEL.
- [2] Alessio, Gian Carlo. 2003. "Il progetto ALIM (Archivio della Latinità Italiana del Medioevo)." In *Biblioteche elettroniche. Letture in Internet: una risorsa per la ricerca e per la didattica*, edited by Francesco Santi, 73-81. SISMEL.
- [3] Arnaldi, Francesco, and Pasquale Smiraglia. ed. 2001. *Grande Dizionario della latinità medievale italiana. Editio altera aucta addendis quae confecerunt L. Celentano, A. De Prisco, A.V. Nazzaro, I. Polara, P. Smiraglia, M. Turriani*. SISMEL.
- [4] Bartoli, Elisabetta. ed. 2014. *Maestro Guido, "Trattati e raccolte epistolari"*. SISMEL.
- [5] Bartoli, Elisabetta. ed. 2019. *Maestro Bernardo, "Introductiones prosaici dictaminis"*. SISMEL.

- [6] Bartoli, Elisabetta. ed. 2023. *Maestro Bernardo, "Rationes dictandi"*. SISMEL.
- [7] Boschetti, Federico, Roberto Del Gratta, Monica Monachini, Marina Buzzoni, Paolo Monella, Roberto Rosselli Del Turco. 2021. "«Tea for two»: the Archive of the Italian Latinity of the Middle Ages meets the CLARIN infrastructure. In *Selected Papers from the CLARIN Annual Conference 2020*, edited by Costanza Navarretta and Maria Eskevich, 37-46. <https://doi.org/10.3384/ecp1805>. Linköping ElectronicConference Proceedings.
- [8] D'Angelo, Edoardo. ed. 2014. *De rebus circa regni Siciliae curiam gestis, Epistola ad Petrum de desolatione Siciliae*. SISMEL.
- [9] Delle Donne, Fulvio, Antonella Brecia and Martina Pavoni. ed. 2000. *Petrus de Ebulo, De rebus Siculis Carmen*. Digital edition with [EVT – Edition Visualization Technology](#). BUP – Basilicata University Press.
- [10] Delle Donne, Fulvio. ed. 2007. *Una silloge epistolare della seconda metà del XIII secolo. I «Dictamina» provenienti dall'Italia meridionale del ms. Paris, Bibl. Nat. Lat. 8567*. SISMEL.
- [11] Ferrarini, Edoardo. 2017. "Alim ieri e oggi", *Umanistica Digitale* 1: 7-17. <https://doi.org/10.6092/issn.2532-8816/7193>.
- [12] Monella, Paolo. 2017. *Ursus from Benevento, "De nomine", from the "Adbreviatio artis grammaticae", codex Casanatensis 1086, ff. 1r-11r, digital edition*. In *ALIM – Archivio della Latinità Italiana del Medioevo*, Palermo. Edition <https://urlsand.esvalabs.com/u=http%3A%2F%2Fwww.unipa.it%2Fpaolo.monella%2Fursus%2F&e=ae0ec65b&h=cb82650c&f=y>.
- [13] Rockinger, Ludwig. 1863-4. "Briefsteller und Formelbücher des 11. Bis 14. Jahrhunderts". *Quellen und Erörterungen zur bayerischen un deutschen Geschichte* 9: 21.
- [14] Stella, Francesco. 2013. "Encoding Issues in Philological Digital Editions." *Horizons: Seoul Journal of Humanities* 4: 239-254.
- [15] Stella, Francesco. 2015. "Il problema della codifica nelle edizioni filologiche digitali." In *Nel segno del testo. Edizioni, materiali e studi per Oronzo Pecere*, a cura di Lucio Del Corso, Franco De Vivo e Antonio Stramaglia, 347-360. Gonnelli.
- [16] Vecchi, Giovanni. ed. 1963. *Magistri Boni Lucensis "Cedrus Libani"*. Società tipografica Modenese.