



Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Μουσειολογίας

From OCR to Quantitative Analysis:

A Multi-Method Framework for AI-Driven Archive Enrichment and Accessibility

Ευάγγελος Βάρθης, evarthis@ionio.gr, Πέτρος Κостаγιόλας, pkostagiolas@ionio.gr, Μάριος Πούλος, mpoulos@ionio.gr - Εργαστήριο Τεχνολογίας της Πληροφορίας

Εισαγωγή – Περίληψη

Τα ψηφιακά αρχεία έχουν τεράστια πολιτιστική αξία, ωστόσο η χρησιμότητά τους συχνά περιορίζεται από έλλειψη σημασιολογικών μεταδεδομένων, γλωσσικά εμπόδια και δυσκολία μεταγραφών.

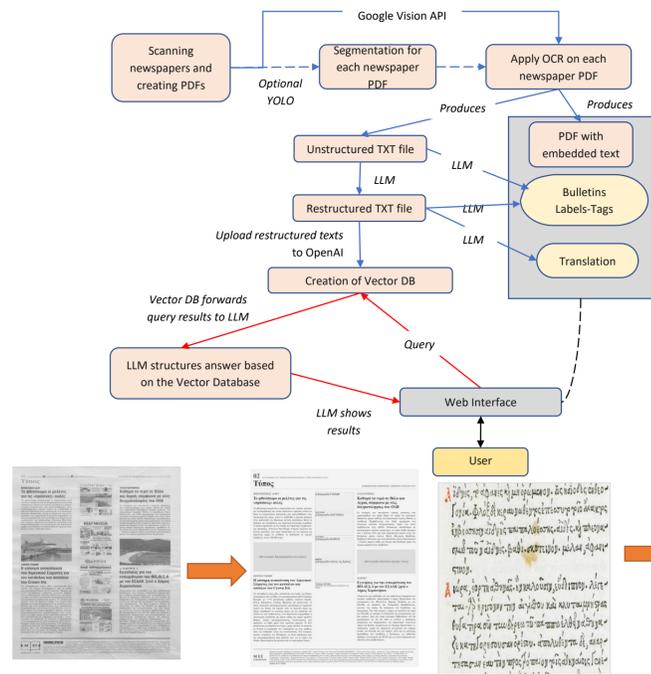
Προτείνουμε ένα ολοκληρωμένο πλαίσιο που ενσωματώνει τρεις διακριτές μεθοδολογικές προσεγγίσεις με στόχο την υπέρβαση των παραπάνω περιορισμών και τη βελτίωση της προσβασιμότητας, της λειτουργικής αξιοποίησης και της επιστημονικής κατανόησης των ιστορικών κειμένων.

Ποσοτική ανάλυση μεγάλων κειμένων: Ανάλυση μεγάλων αδόμητων σωμάτων όπως π.χ. η *Patrologia Graeca*, με απώτερο σκοπό την χαρτογράφηση βιβλικών μοτίβων, εύρεση θεματικών συγγενειών μεταξύ των συγγραφέων καθώς και τα κύρια θέματα που τους απασχολούσαν μέσα στους αιώνες με εφαρμογή **topic modelling** και **στατιστικές μεθόδους** πάνω στα βιβλικά εδάφια.

Μεταγραφή και αναζήτηση μέσω Τεχνητής Νοημοσύνης: Ροή εργασίας πάνω σε σαρωμένα αρχεία εφημερίδων και χειρόγραφα, με εφαρμογή προχωρημένα πολυγλωσσικά OCR και M-LLMs για να επιτρέψει την αυτόματη μεταγραφή, σύνοψη, μετάφραση και σημασιολογική αναζήτηση.

Μεταγραφή μέσω Crowd-sourcing: Οικονομικά αποδοτική και απλή πλατφόρμα που συνδυάζει τη διόρθωση και την μεταγραφή δύσκολων εγγράφων (Πάπυροι κτλ.) με προσθήκη μεταδεδομένων που δημιουργούνται από τον χρήστη για τη δημιουργία πινάκων περιεχομένων με δυνατότητα πλοήγησης.

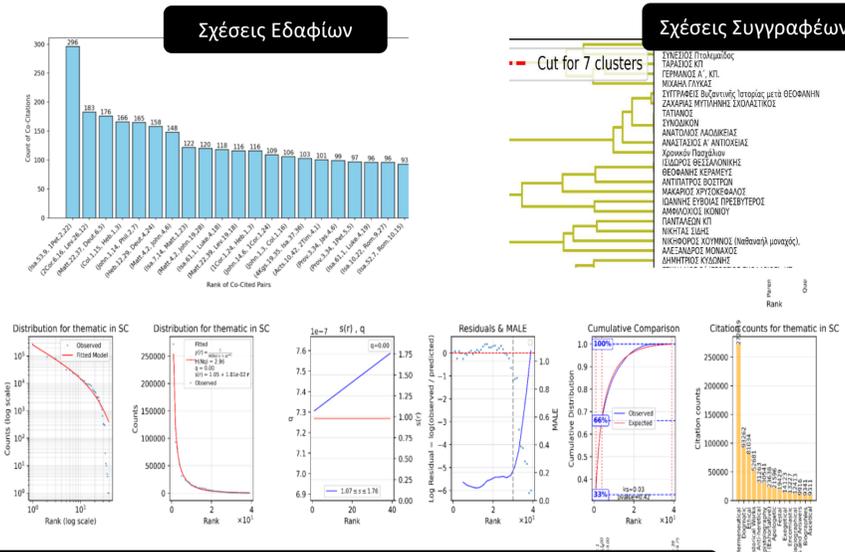
OCR και M-LLMs workflow για Μεταγραφή Τυπωμένων και Χειρόγραφων Εγγράφων



Αιδώς, τὸ ἀφανὲς καὶ μὴ ὁράμενον. ὡς Ἡσίοδος ἐν Ἑσο- γόνα. Φιλὸς δὲ κεραυνωθέντος ἐπέσσυτο τοῦ ἀνακτοῦ ἐμβήσκη αἰδῆς παταλοῦσης, αἰδῆς, καὶ πλεονα- σμῶ τοῦ ν, αἰδῆς, βαθῦ, σκοτεινόν. μέλαν, ἀφανι- στικόν. Ἀτώρα, ἐορτὴ Ἀθηνῶν. ἦν κολοῦσιν, εὐδαιμονόν. λέγε- ται γὰρ Ἡρόδοτον τὴν Αἰγίσθου καὶ Κλυταμνήστρας θυγατέρα σὺν Τυνδάρεω τῷ πάπῳ ἐλθεῖν Ἀθῆνα- ζε κατηγορούσασαν Ὀρέστου, ἀπολυθέντα δὲ, ἀναρ- τήσασαν ἐαυτὴν προσβάλαιον τοῖς Ἀθηναίοις γενέ- σθαι. κατὰ χρῆσμον δὲ ἐπὶ αὐτῆ συντελεσθαι τὴν ἐορτήν.

Varthis, E., & Poulos, M. (2026). A Streamlined AI-Powered Workflow for Enriching Library Newspaper Archives, accepted and pending for publication in *Journal of Web Librarianship* (Taylor & Francis)

Ποσοτική Ανάλυση Μεγάλων Δεδομένων



Στατιστικά συμπεράσματα για τις κατανομές των εδαφίων

Figure 1. Decreasing number of verse-references across different thematics and their interpolation using the MZM model for the OT, NT, and SC. Panels show: (a) log-log plots, (b) raw (linear-scale) plots, (c) the rank-dependent function $s(r)$ and q , (d) residuals, mean absolute log error (MALE), (e) comparison of cumulative plots, KS statistic and (f) top 15 ranks.

Crowdsourcing Platform για Μεταγραφή και Δημιουργία TOCs σε δύσκολα έγγραφα

Αναζήτηση & Πίνακας υπερσυνδέσμων μέσα στο έγγραφο

Δυνατότητα μεταγραφής μέσω OCR, LLM, Φωνής

Varthis, E., & Poulos, M. (2025). Assigning thematic headings to sections of scanned documents for semantic enrichment in libraries: an alternative-complementary approach to the complexity of AI techniques. *Library Hi Tech News*.



research-day
ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

<https://conferences.ionio.gr/research-day>



Ημέρα Έρευνας 2026

Τρίτη 17 Μαρτίου 2026 - Ιόνιο Πανεπιστήμιο