

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
«ΣΥΣΤΗΜΑ ΜΕΤΑΓΡΑΜΜΑΤΙΣΜΟΥ ΕΓΓΡΑΦΩΝ  
ΒΙΒΛΙΟΘΗΚΗΣ»

Μεταγραμματος ελληνικών  
χαρακτήρων σε λατινικούς



Κείμενο

Εισάγετε το κείμενο για μεταγραμματος

Αρχαία Ελληνικά

ALA-LC Romanization

ISO843:1997

Καθαρισμός

Υποβολή

Του φοιτητή  
Ηλία Τσιτσιλέγκα  
Αρ. Μητρώου: it185301

Επιβλέπων  
Όνοματεπώνυμο: Χαράλαμπος Μπράτσας  
Βαθμίδα: Επίκουρος Καθηγητής

Ημερομηνία 20/09/2024

Τίτλος Δ.Ε. Σύστημα Μεταγραμματισμού Εγγραφών Βιβλιοθήκης

Κωδικός Δ.Ε. 23314

Όνοματεπώνυμο φοιτητή: Ηλίας Τσιτσιλέγκας

Όνοματεπώνυμο εισηγητή Χαράλαμπος Μπράτσας

Ημερομηνία ανάληψης Δ.Ε. 24-11-2023

Ημερομηνία περάτωσης Δ.Ε. ...

*Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.*

*Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Ηλία Τσιτσιλέγκα που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.*

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

*«Αφιέρωση»*

*Στους «δικούς μου» ανθρώπους*



## Πρόλογος

Οι προπτυχιακές σπουδές ανατροφοδότησαν, γενικά, την επιθυμία μου να καλλιεργώ τον τρόπο σκέψης μου, ώστε να δημιουργώ αυτοματοποιημένα εργαλεία για την επίλυση προβλημάτων, προς διευκόλυνση όσων εργάζονται με ηλεκτρονικά μέσα.

Πιο συγκεκριμένα, στο πλαίσιο συνεργασίας της Εθνικής Βιβλιοθήκης της Ελλάδος (ΕΒΕ) με το Τμήμα Πληροφορικής και Ηλεκτρονικών Μηχανικών του ΔΠΙΑΕ και το Open Knowledge Foundation Greece, προκειμένου να αποκτήσει η πρώτη ένα εργαλείο απόδοσης των ονομάτων Ελλήνων συγγραφέων και νομικών προσώπων με λατινικούς χαρακτήρες, ο όρος “μεταγραμματισμός” κίνησε ιδιαίτερα την περιέργειά μου και με κατεύθυνε προς το συγκεκριμένο έργο: τη δημιουργία, δηλαδή, ενός προγράμματος μεταγραμματισμού των χαρακτήρων της ελληνικής γλώσσας, αρχαίας και νέας, σε λατινικούς χαρακτήρες, που θα διευκολύνει τον διαμοιρασμό μεταδεδομένων της ΕΒΕ με άλλες βιβλιοθήκες, σε εθνικό και σε παγκόσμιο επίπεδο, αλλά και τη συνεργασία των ξένων βιβλιοθηκών μεταξύ τους.

Η ελληνική γλώσσα είναι από τις λίγες που δεν γράφονται με λατινικούς χαρακτήρες. Ταυτόχρονα, είναι η γλώσσα στην οποία έχει γραφεί ένα πολύ σπουδαίο κεφάλαιο του ανθρώπινου πολιτισμού, από την αρχαιότητα μέχρι και σήμερα, στο οποίο ανατρέχει η παγκόσμια επιστημονική κοινότητα. Αυτά τα δύο χαρακτηριστικά της ελληνικής γλώσσας πείθουν για την αναγκαιότητα να κατασκευαστεί ένα εργαλείο μεταγραμματισμού, που θα διευκολύνει την πρόσβαση στην τεράστια ελληνική βιβλιογραφία και θα αξιοποιείται και στο μέλλον.

## Περίληψη

Ο στόχος της παρούσας εργασίας ήταν η δημιουργία ενός εργαλείου μεταγραμματισμού των χαρακτήρων της αρχαίας ελληνικής και της νεοελληνικής γλώσσας, προκειμένου να εφαρμόζεται από την Εθνική Βιβλιοθήκη της Ελλάδος και όχι μόνο, στη διαδικασία διαμοιρασμού μεταδεδομένων μεταξύ των βιβλιοθηκών σε τοπικό και διεθνές επίπεδο.

Προετοιμάζοντας την κατασκευή του προγράμματος μεταγραμματισμού, έγινε βιβλιογραφική ανασκόπηση, από την οποία προέκυψε η κατάλληλη επιλογή των ανοιχτών τεχνολογιών ως υποδομών και της μεθόδου μεταγραμματισμού, έναντι άλλων μη πρόσφορων για εφαρμογή μεγάλης κλίμακας και πολλαπλών επιλογών. Συγκεκριμένα, στη βάση των προτύπων μεταγραμματισμού ISO843 και American Romanization Tables για τα αρχαία και για τα νέα ελληνικά και με τις HTML, CSS και JavaScript τεχνολογίες για την κατασκευή ιστοσελίδας και γραφικού περιβάλλοντος, δημιουργήθηκε ένα εργαλείο μεταγραμματισμού με τη μέθοδο των πινάκων αντιστοίχισης (mapping method).

Η πιλοτική εφαρμογή του από καταλογογράφους της Εθνικής Βιβλιοθήκης Ελλάδος σε ένα σύνολο δεδομένων αναφοράς τεκμηρίωσε τόσο την ευχρηστία του εργαλείου (ευκολία, άμεση πρόσβαση και ταχεία ανταπόκριση), όσο και την ακρίβεια των αποτελεσμάτων του σε πολύ υψηλό ποσοστό επιτυχίας, ακόμα και σε ειδικές περιπτώσεις που γεννώνται από τις ιδιαιτερότητες της αρχαίας ελληνικής γλώσσας (πολυτονικό σύστημα) και εκείνες της νεοελληνικής ως μετεξέλιξής της (μονοτονικό σύστημα, λέξεις που δασύνονταν στην αρχαία ελληνική). Συγκριτικά με την αντίστοιχη εκτέλεση του μεταγραμματισμού στο ίδιο σύνολο δεδομένων αναφοράς από το ChatGPT, το εν λόγω εργαλείο αποδεικνύεται πιο αξιόπιστο. Με το βοηθητικό πρόγραμμα ελέγχου, μάλιστα, που προτείνεται στην παρούσα εργασία, μπορεί να βελτιωθεί περαιτέρω, αξιοποιώντας έτσι ακόμα ένα σπουδαίο πλεονέκτημα της mapping method, αυτό της επεκτασιμότητας, προκειμένου να αξιοποιείται ως εργαλείο και στο μέλλον.

# «TRANSLITERATION SYSTEM FOR LIBRARY DOCUMENTS»

«Ilias Tsitsilegkas»

## **Abstract**

The aim of this project was to develop a tool for the transliteration of characters from Ancient and Modern Greek, intended for use by the National Library of Greece and beyond, in the process of metadata sharing between libraries both locally and internationally.

In preparation for developing the transliteration tool, a literature review was conducted, which led to the appropriate selection of open technologies as infrastructure and the choice of the transliteration method, as opposed to others that are unsuitable for large-scale applications and multiple options. Specifically, based on the transliteration standards ISO 843 and American Romanization Tables for both Ancient and Modern Greek, and using HTML, CSS, and JavaScript technologies for the creation of the website and graphical interface, a transliteration tool was developed using the mapping method.

A pilot implementation of the tool by cataloguers from the National Library of Greece on a dataset demonstrated both its usability (ease of use, immediate access, and quick response) and the accuracy of its results, achieving a high success rate. This success was evident even in special cases arising from the complexities of Ancient Greek (the polytonic system) and Modern Greek as its evolution (the monotonic system, words that had the insignia “daseîa” in Ancient Greek). Compared to performing the same transliteration on the dataset using ChatGPT, this tool demonstrated higher reliability. Moreover, with the control utility proposed in this project, the tool can be further improved, capitalizing on another significant advantage of the mapping method, namely scalability, to ensure it can be utilized as a future tool.

## Ευχαριστίες

Για την εκπόνηση της διπλωματικής μου εργασίας, θα ήθελα να ευχαριστήσω πρώτ' απ' όλους, τον καθηγητή μου, κο Χαράλαμπο Μπράτσα, Επίκουρο Καθηγητή του Τμήματος Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΠΙΑΕ, ο οποίος μου ενέπνευσε το ενδιαφέρον για το συγκεκριμένο ερευνητικό θέμα.

Επίσης, οφείλω να αποδώσω τη μέγιστη εκτίμησή μου στο προσωπικό της Εθνικής Βιβλιοθήκης της Ελλάδας, που συμμετείχε στην δοκιμαστική εφαρμογή του προγράμματος μεταγραμματισμού και όχι μόνο. Ιδιαίτερα, θέλω να ευχαριστήσω θερμά την κα Σοφία Ζαπουνίδου, η οποία συστηματικά με στήριξε στην συγκέντρωση του υλικού που αξιοποιήθηκε για την κατασκευή του εργαλείου μεταγραμματισμού, με αμείωτο ενδιαφέρον για τη βελτίωση των mappings που συνθέτουν το προτεινόμενο πρόγραμμα.

Τέλος, ευχαριστώ από καρδιάς τους φίλους μου, για την ηθική τους συμπαράσταση, περισσότερο απ' όλους όμως, την οικογένειά μου, η οποία με στήριξε συναισθηματικά σε αυτήν την προσπάθεια.

# Περιεχόμενα

Πρόλογος.....	v
Περίληψη.....	vi
Abstract .....	vii
Ευχαριστίες .....	viii
Περιεχόμενα .....	ix
Κατάλογος Πινάκων.....	xi
Κατάλογος εικόνων .....	xi
Συνομογραφίες.....	xiii
Κεφάλαιο 1ο: Μεταγραμματισμός και βιβλιοθήκες.....	1
1.1 Εισαγωγή.....	1
1.2 Τι είναι “μεταγραμματισμός”;.....	1
1.3 Πρότυπα μεταγραμματισμού.....	1
1.4 Γιατί χρησιμοποιείται από τις βιβλιοθήκες.....	2
1.5 Ερευνητικό ερώτημα εργασίας.....	2
1.6 Επίλογος.....	2
Κεφάλαιο 2ο: Βιβλιογραφική ανασκόπηση .....	3
2.1 Εισαγωγή.....	3
2.2 Εθνική Βιβλιοθήκη της Ελλάδος (ΕΒΕ) .....	3
2.2.1 Εθνικό Βιβλιογραφικό και Καταλογογραφικό Κέντρο .....	3
2.2.2 Πρότυπα μεταγραμματισμού στην ΕΒΕ.....	4
2.3 Αναφορά σε διαθέσιμες μεθόδους.....	4
2.3.1 Πλεονεκτήματα και Μειονεκτήματα.....	5
2.3.3 Μέθοδοι που έχουν χρησιμοποιηθεί σε transliteration.....	10
2.4 Επίλογος.....	10
Κεφάλαιο 3ο: Μέθοδοι.....	12
3.1 Εισαγωγή.....	12
3.2 Γενικές αποφάσεις/προδιαγραφές .....	12
3.3 ISO 843 .....	13
3.3.1 Σύντομη περιγραφή του προτύπου.....	13
3.3.2 Βασικοί κανόνες.....	13
Υλοποίηση κανόνων στο εργαλείο μεταγραμματισμού .....	15
3.4 Πρότυπο ALA-LC Romanization table for Greek .....	19
3.4.1 Σύντομη περιγραφή του προτύπου.....	19

3.4.2	Βασικοί κανόνες.....	19
3.4.3	Υλοποίηση κανόνων στο εργαλείο μεταγραμματισμού .....	21
3.5	Σχεδιασμός αξιολόγησης.....	34
3.5.1	Δημιουργία συνόλου δεδομένων αναφοράς (gold dataset) .....	34
3.5.2	Προετοιμασία prompts για ChatGPT .....	34
3.6	Γραφικό περιβάλλον.....	36
3.6.1	Αρχές δημιουργίας GUI που εφαρμόστηκαν .....	36
3.6.2	Δοκιμή από το προσωπικό της ΕΒΕ για σχόλια σχετικά με την ευχρηστία.....	37
3.7	Επίλογος.....	38
Κεφάλαιο 4ο: Αποτελέσματα .....		39
4.1	Εισαγωγή.....	39
4.2	Σύντομη περιγραφή της εφαρμογής με εικόνες.....	39
4.3	Στατιστικά αποτελέσματα αξιολόγησης.....	43
4.3.1	Αξιολόγηση εργαλείου Transliterator .....	43
4.3.2	Αξιολόγηση ChatGPT .....	43
4.4	Βοηθητικό πρόγραμμα ελέγχου.....	44
4.5	Επίλογος.....	48
Κεφάλαιο 5ο: Συμπεράσματα ή/και προτάσεις βελτίωσης .....		49
5.1	Εισαγωγή.....	49
5.2	Ανοικτές τεχνολογίες που επιλέχθηκαν .....	49
5.3	Για μέθοδο που επιλέχθηκε - τα πλεονεκτήματα, τα μειονεκτήματα της μεθόδου όπως εφαρμόστηκε στο εργαλείο, πιθανοί περιορισμοί, συσχετισμός με άλλες σύγχρονες μεθόδους .....	49
5.4	Προκλήσεις για κάθε πρότυπο. ....	50
5.5	Σχολιασμός αποτελεσμάτων αξιολόγησης. Γιατί πήγε το εργαλείο καλύτερα από το Chat Gpt; 50	
Επίλογος.....		52
ΒΙΒΛΙΟΓΡΑΦΙΑ.....		53
Παράρτημα Α: Κατάλογος δασυνόμενων λέξεων και ονομάτων .....		55

## Κατάλογος Πινάκων

Πίνακας 3.1: Πίνακας μεταγραμματος προτύπου ISO 843 .....	15
Πίνακας 3.2: Μετατροπή των μη αλφαβητικών χαρακτήρων.....	17
Πίνακας 3.3: Πίνακας αντιστοίχισης “ALA-LC Romanization tables - Greek” .....	20
Πίνακας 4.1: Πίνακας στατιστικών επιτυχίας προγράμματος μεταγραμματος .....	43
Πίνακας 4.2: Πίνακας στατιστικών επιτυχίας ChatGPT .....	44

## Κατάλογος εικόνων

Εικόνα 2.1: Μεταγραμματος της Αλφαβήτας με την χρήση if ελέγχου.....	8
Εικόνα 2.2: Μεταγραμματος της Αλφαβήτας με την χρήση Switch ελέγχου.....	9
Εικόνα 3.1: Αντιστοίχιση Νέων Ελληνικών ISO 843 .....	16
Εικόνα 3.2: Αντιστοίχιση Αρχαίων Ελληνικών ISO 843 .....	16
Εικόνα 3.3: Σημεία στίξης και ειδικοί χαρακτήρες.....	17
Εικόνα 3.4: Τονισμοί με οξεία .....	18
Εικόνα 3.5: Τονισμοί με περισπωμένη.....	18
Εικόνα 3.6: Παράδειγμα με διαλυτικά .....	18
Εικόνα 3.7: Αντιστοιχίες με δύο χαρακτήρες.....	18
Εικόνα 3.8: Έλεγχος αντιστοιχιών με δύο χαρακτήρες.....	19
Εικόνα 3.9: Αντιστοιχίες Νέων Ελληνικών ALA-LC Romanization tables .....	21
Εικόνα 3.10: Αντιστοιχίες Αρχαίων Ελληνικών ALA-LC Romanization tables.....	22
Εικόνα 3.11: Τονισμοί, οξείες και βαρείες.....	22
Εικόνα 3.12: Τονισμοί - Περισπωμένη .....	22
Εικόνα 3.13: Αντιστοιχίες με απόστροφο .....	23
Εικόνα 3.14: Αντιστοιχίες ειδικών χαρακτήρων και σημείων στίξης.....	23
Εικόνα 3.15: Τονισμοί με δασεία.....	23
Εικόνα 3.16: Έλεγχος μονού χαρακτήρα .....	24
Εικόνα 3.17: Μεταγραμματος Β χαρακτήρα στα αρχαία ελληνικά.....	24
Εικόνα 3.18: Μεταγραμματος Β στα Νέα Ελληνικά .....	24
Εικόνα 3.19: Δασυνόμενες λέξεις 3 χαρακτήρων .....	25
Εικόνα 3.20: Δασυνόμενες λέξεις με 4 χαρακτήρες.....	26
Εικόνα 3.21: Δασυνόμενες λέξεις με 5 χαρακτήρες.....	26
Εικόνα 3.22: Έλεγχος δασυνόμενων λέξεων για τα Νέα Ελληνικά.....	27
Εικόνα 3.23: Αντιστοιχίες διφθόγγων, στην αρχή της λέξης.....	28
Εικόνα 3.24: Αλγόριθμος ελέγχου των starters.....	28
Εικόνα 3.25: Αντιστοιχίες άρθρων.....	29
Εικόνα 3.26: Βοηθητικός αλγόριθμος αντιστοιχίσεων .....	31
Εικόνα 3.27: Συνέχεια βοηθητικού αλγόριθμου αντιστοιχίσεων .....	32
Εικόνα 3.28: Ολοκλήρωση βοηθητικού αλγόριθμου αντιστοιχίσεων.....	33
Εικόνα 3.29: Γραφικό περιβάλλον στα ελληνικά.....	36
Εικόνα 3.30: Γραφικό περιβάλλον στα αγγλικά.....	37
Εικόνα 4.1: Παράδειγμα 1.....	39
Εικόνα 4.2: Παράδειγμα 2.....	40

Εικόνα 4.3: Παράδειγμα 1, δασυνόμενης λέξης στα Νέα Ελληνικά.....	41
Εικόνα 4.4: Παράδειγμα 1, δασυνόμενης λέξης στα Αρχαία Ελληνικά .....	41
Εικόνα 4.5: Παράδειγμα 2, δασυνόμενης λέξης στα Νέα Ελληνικά.....	42
Εικόνα 4.6: Παράδειγμα 2, δασυνόμενης λέξης στα Αρχαία Ελληνικά .....	42
Εικόνα 4.7: Περιοχή εισαγωγής mappings στο βοηθητικό πρόγραμμα.....	44
Εικόνα 4.8: Παραδείγματα για έλεγχο .....	45
Εικόνα 4.9: Κώδικας ελέγχων .....	46
Εικόνα 4.10: Παράδειγμα αποτελεσμάτων ελέγχου 1.....	47
Εικόνα 4.11: Παράδειγμα αποτελεσμάτων ελέγχου 2.....	47
Εικόνα 4.12: Μετρητές σωστού / λάθους.....	47
Εικόνα 4.13: Παράδειγμα αποτελεσμάτων μετρητών ελέγχου .....	48

## Συντομογραφίες

ALA	American Library Association
BGN	Board on Geographic Names
CSS	Cascading Style Sheets
DIN	Normenausschuss Information und Dokumentation
GMM	Gaussian Mixture Model
HHI	Herfindahl-Hirschman Index
HTML	HyperText Markup Language
ISO	International Standards Organization
ISSN	International Standard Serial Number
LC	Library of Congress
OCR	Optical Character Recognition
PCGN	Permanent Committee on Geographical Names for British Official Use
VIAF	Virtual International Authority File
WFSTs	Weighted Finite-State Transducers
MTSR	Metadata and Semantics Research
ΔΠΙΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
ΕΒΕ	Εθνική Βιβλιοθήκη της Ελλάδας
ΕΛΟΤ	Ελληνικός Οργανισμός Τυποποίησης



## Κεφάλαιο 1ο: Μεταγραμματισμός και βιβλιοθήκες

### 1.1 Εισαγωγή

Το κύριο αντικείμενο της εργασίας είναι η δημιουργία ενός εργαλείου μεταγραμματισμού. Σε αυτό το κεφάλαιο επιχειρούνται η αποσαφήνιση της έννοιας του μεταγραμματισμού, η παρουσίαση των προτύπων διαδικασίας του μεταγραμματισμού, καθώς και η τεκμηρίωση της χρησιμότητάς του για τις βιβλιοθήκες.

### 1.2 Τι είναι “μεταγραμματισμός”;

Μεταγραμματισμός είναι η διαδικασία κατά την οποία οι χαρακτήρες ενός συστήματος γραφής μετατρέπονται σε χαρακτήρες ενός άλλου συστήματος [1]. Ο όρος αποδίδεται στα αγγλικά ως “transliteration”. Οι κανόνες μεταγραμματισμού συνήθως αποτυπώνονται σε πίνακες αντιστοίχισης, όπου παρουσιάζεται η μετατροπή ανά χαρακτήρα. Σε περιπτώσεις όπου ένας χαρακτήρας δεν μπορεί να αποδοθεί στο αλφαβητικό σύστημα μετατροπής, είτε γιατί δεν υπάρχει ανάλογος χαρακτήρας είτε επειδή το σύστημα μετατροπής έχει λιγότερους ή περισσότερους χαρακτήρες από το αρχικό αλφαβητικό σύστημα, χρησιμοποιούνται ειδικοί χαρακτήρες ή διακριτικά σημεία.

Ο μεταγραμματισμός στηρίζεται αποκλειστικά στη μετατροπή χαρακτήρων και δε λαμβάνει υπόψη τη φωνολογία των λέξεων. Το γεγονός αυτό, αφενός, καθιστά εφικτή την αντίστροφη διαδικασία μετατροπής μιας μεταγεγραμμένης λέξης στο αρχικό αλφάβητο (επαναγραμματισμός) [1], αφετέρου, επιτρέπει τον μεταγραμματισμό από όλους, ακόμα και από άτομα που δε γνωρίζουν κανένα από τα δύο συστήματα γραφής.

Ο μεταγραμματισμός δεν πρέπει να συγχέεται με τη διαδικασία της μεταγραφής. Και σε αυτή τη διαδικασία γίνεται μετατροπή χαρακτήρων από ένα αλφαβητικό σύστημα σε ένα άλλο. Η διαφορά έγκειται στο ότι η διαδικασία της μεταγραφής λαμβάνει υπόψη τη φωνολογία των λέξεων και δεν είναι απολύτως αντιστρέψιμη [1]. Ο όρος αποδίδεται στα αγγλικά ως “transcription”.

Ένας επιπλέον όρος που συναντάται στη βιβλιογραφία είναι ο όρος “ρομανοποίηση” (romanization). Ο όρος αυτός αφορά στη μετατροπή ενός μη λατινικού συστήματος γραφής σε λατινικούς χαρακτήρες. Για τη ρομανοποίηση μπορεί να χρησιμοποιηθεί είτε η μέθοδος του μεταγραμματισμού είτε η μέθοδος της μεταγραφής [1].

### 1.3 Πρότυπα μεταγραμματισμού

Υπάρχουν πολλά πρότυπα μεταγραμματισμού για κάθε σύστημα γραφής, τα οποία αναπτύσσονται είτε σε διεθνές είτε σε εθνικό είτε σε τοπικό επίπεδο. Σε διεθνές επίπεδο αναπτύσσονται από τον Παγκόσμιο Οργανισμό Τυποποίησης (International Standards Organization), σε εθνικό επίπεδο από τα εθνικά γραφεία τυποποίησης ή άλλες εθνικές υπηρεσίες, σε τοπικό επίπεδο από οποιοδήποτε φορέα που έχει σχετικό ενδιαφέρον.

Όσον αφορά το ελληνικό σύστημα γραφής και τη μετατροπή του σε λατινικούς χαρακτήρες, υπάρχει το διεθνές πρότυπο ISO 843 [2], το οποίο προβλέπει κανόνες και για τη διαδικασία μεταγραμματισμού και για τη διαδικασία της μεταγραφής. Στην Ελλάδα, αυτό το πρότυπο έχει μεταφραστεί από τον Ελληνικό Οργανισμό Τυποποίησης (ΕΛΟΤ), λαμβάνοντας την αρίθμηση ΕΛΟΤ 743 [1]. Το διεθνές πρότυπο ISO έχει υιοθετηθεί από τις γεωγραφικές/χαρτογραφικές υπηρεσίες Ηνωμένων Πολιτειών Αμερικής και Βρετανίας ως “BGN/PCGN ρομανοποίηση” [3]. Άλλες χώρες έχουν αναπτύξει τα δικά τους πρότυπα

μεταγραμματισμού για τα ελληνικά, όπως για παράδειγμα η Γερμανία, η οποία έχει αναπτύξει το πρότυπο DIN 31634:2011 [4]. Ένα πολύ γνωστό τοπικό πρότυπο είναι ο πίνακας ρομανοποίησης για τα ελληνικά που ανέπτυξαν η Αμερικάνικη Ένωση Βιβλιοθηκών (American Library Association - ALA) σε συνεργασία με τη Βιβλιοθήκη του Κογκρέσου (Library of Congress - LC) [5].

### 1.4 Γιατί χρησιμοποιείται από τις βιβλιοθήκες

Πάγια πρακτική στις βιβλιοθήκες είναι οι διαμοιρασμός μεταδεδομένων (υπό τη μορφή εγγραφών) σχετικά με εκδόσεις ή οντότητες που σχετίζονται με αυτές, π.χ., συγγραφείς, εκδότες, τόποι, κ.λπ. Στο πλαίσιο διαμοιρασμού εγγραφών σε διεθνές επίπεδο, οι βιβλιοθήκες υιοθέτησαν τη μέθοδο του μεταγραμματισμού, η οποία δεν απαιτεί από τους χρήστες να γνωρίζουν τα διαφορετικά συστήματα γραφής και, επιπλέον, επιτρέπει τον επαναγραμματισμό.

Διεθνείς φορείς, όπως είναι το Διεθνές Κέντρο ISSN (ISSN International Centre), το οποίο αποδίδει πρότυπους αριθμούς σε περιοδικές εκδόσεις (International Standard Serial Number - ISSN), υπαγορεύουν τον μεταγραμματισμό των τίτλων περιοδικών εκδόσεων, ονομάτων εκδοτών και τόπων έκδοσης σύμφωνα με το διεθνές πρότυπο ISO 843 [6]. Βιβλιοθήκες που συμμετέχουν στο Πρόγραμμα Συνεργατικής Καταλογογράφησης της Βιβλιοθήκης του Κογκρέσου ακολουθούν τους κανόνες ρομανοποίησης που ανέπτυξαν οι ALA-LC [5, 107].

Ανάλογα με τις συνεργασίες που έχει (ή που θέλει να αποκτήσει), μία βιβλιοθήκη επιλέγει ποια πρότυπα μεταγραμματισμού θα εφαρμόσει.

### 1.5 Ερευνητικό ερώτημα εργασίας

Το ερευνητικό ερώτημα που προκύπτει είναι αν μπορεί να δημιουργηθεί ένα αυτόματο σύστημα μεταγραμματισμού, το οποίο να αναγνωρίζει ορθά όλους τους χαρακτήρες της αρχαίας και της νέας ελληνικής και να τους μεταγραμματίζει γρήγορα, αξιόπιστα και με ακρίβεια, υλοποιώντας τους κανόνες του κάθε προτύπου.

### 1.6 Επίλογος

Μεταγραμματισμός είναι η διαδικασία μετατροπής χαρακτήρων από ένα σύστημα γραφής σε ένα άλλο. Διακρίνεται από τη συγγενή έννοια της μεταγραφής σε ό,τι αφορά τον τύπο της μετατροπής και τη δυνατότητα αντιστροφής της διαδικασίας: η μετατροπή γίνεται σύμφωνα με τους κανόνες κάποιου προτύπου, δεν είναι φωνολογική και είναι απόλυτα αντιστρέψιμη. Γι' αυτούς τους λόγους, οι βιβλιοθήκες προτιμούν τον μεταγραμματισμό για την ανταλλαγή βιβλιογραφικών δεδομένων με άλλες βιβλιοθήκες που χρησιμοποιούν διαφορετικές γλώσσες και συστήματα γραφής, των οποίων η γνώση δεν απαιτείται από τους καταλογογράφους για την εφαρμογή του εργαλείου. Υπάρχουν αρκετά πρότυπα σε διεθνές, εθνικό ή τοπικό επίπεδο. Για την ελληνική γλώσσα και τους ελληνικούς χαρακτήρες, τα πιο διαδεδομένα συστήματα μεταγραμματισμού είναι το ISO 843 [2] και ο Πίνακας ρομανοποίησης των ALA-LC [5].

## Κεφάλαιο 2ο: Βιβλιογραφική ανασκόπηση

### 2.1 Εισαγωγή

Στο παρόν κεφάλαιο, αναφέρονται ιστορικά στοιχεία της Εθνικής Βιβλιοθήκης της Ελλάδας, προβάλλεται ο σκοπός της, περιγράφεται η λειτουργία της ως βιβλιογραφικού και καταλογογραφικού κέντρου και παρουσιάζονται οι μέθοδοι μεταγραμματισμού που έχει υιοθετήσει.

Περιγράφεται η μέθοδος αντιστοίχισης (mapping method), η οποία υιοθετείται για τη δημιουργία του εργαλείου μεταγραμματισμού στην παρούσα εργασία. Τεκμηριώνεται η επιλογή της και παρατίθενται τα πλεονεκτήματα και τα μειονεκτήματά της.

Επιπλέον, επιχειρείται ένα θεωρητικό συμπλήρωμα γενικά γύρω από τις μεθόδους του μεταγραμματισμού, παραδοσιακές και σύγχρονες, είτε αφορούν τον μεταγραμματισμό της ελληνικής γλώσσας είτε άλλων γλωσσών, μη ευρωπαϊκών.

### 2.2 Εθνική Βιβλιοθήκη της Ελλάδος (ΕΒΕ)

Η Εθνική Βιβλιοθήκη της Ελλάδος (ΕΒΕ) ιδρύθηκε με διάταγμα του Κυβερνήτη Ιωάννη Καποδίστρια τον Μάιο του 1832. Η επωνυμία της, σύμφωνα με το διάταγμα, ήταν “Δημοσία Βιβλιοθήκη”. Διευθυντής ορίστηκε ο Γεώργιος Γεννάδιος. Αρχικά, λειτούργησε στο Ορφανοτροφείο της Αίγινας μαζί με άλλα εκπαιδευτικά και πνευματικά ιδρύματα. Το 1834 μεταφέρθηκε στην Αθήνα, η οποία ορίστηκε ως η νέα πρωτεύουσα του νεοσυσταθέντος ελληνικού κράτους. Το 1866 συνενώθηκε με τη βιβλιοθήκη του Οθώνειου Πανεπιστημίου και, για πρώτη φορά, λαμβάνει την ονομασία “Εθνική Βιβλιοθήκη της Ελλάδος”. Το 1888 τέθηκε ο θεμέλιος λίθος του νεοκλασικού κτιρίου με δωρεά από τους αδερφούς Βαλλιάνο. Η μεταφορά της Εθνικής Βιβλιοθήκης στο εμβληματικό Βαλλιάνειο Μέγαρο έγινε το 1903. Η έδρα της Εθνικής Βιβλιοθήκης της Ελλάδος μεταφέρθηκε το 2017 από το Βαλλιάνειο Μέγαρο στο Κέντρο Πολιτισμού Ίδρυμα Σταύρος Νιάρχος [8].

Σύμφωνα με την ισχύουσα νομοθεσία [9], η ΕΒΕ σήμερα αποτελεί Νομικό Πρόσωπο Δημοσίου Δικαίου εποπτευόμενο από το Υπουργείο Παιδείας. Ο σκοπός της είναι [9]:

“ο εντοπισμός, η συγκέντρωση, η οργάνωση, η περιγραφή και η διαφύλαξη στο διηνεκές τεκμηρίων του ελληνικού πολιτισμού και της πρόσληψής του διαχρονικά, καθώς και σημαντικών αντιπροσωπευτικών τεκμηρίων της πνευματικής παραγωγής του ανθρώπου. Η ΕΒΕ διασφαλίζει την ισότιμη πρόσβαση στα τεκμήρια αυτά με γνώμονα την ελευθερία της γνώσης, της πληροφόρησης και της έρευνας.”

#### 2.2.1 Εθνικό Βιβλιογραφικό και Καταλογογραφικό Κέντρο

Στο πλαίσιο υλοποίησης του σκοπού της, η ΕΒΕ οφείλει να λειτουργεί ως Εθνικό Βιβλιογραφικό και Καταλογογραφικό Κέντρο της χώρας. Ειδικότερα [9]:

“(α) αναπτύσσει και συμμετέχει στη διαμόρφωση εθνικών βιβλιογραφικών προτύπων και κατευθυντήριων οδηγιών για κάθε θέμα που αφορά τη λειτουργία και τις υπηρεσίες των βιβλιοθηκών της χώρας, (ββ) συντάσσει και δημοσιεύει την τρέχουσα και την παλαιότερη Εθνική Βιβλιογραφία, (γγ) συντάσσει και διαθέτει στις άλλες βιβλιοθήκες πρότυπες εγγραφές βιβλιογραφικής περιγραφής για τα ελληνικά δημοσιεύματα, (δδ) παρέχει διαρκή πληροφόρηση και υποστήριξη στις ελληνικές βιβλιοθήκες σε θέματα που αφορούν τη βιβλιογραφική περιγραφή, την τεκμηρίωση, τη διατήρηση, την ψηφιοποίηση και τη

μικροφωτογράφιση των συλλογών τους, (εε) παρέχει έγκυρη πληροφόρηση και υλικό για την ελληνική εκδοτική παραγωγή που αφορά τον ελληνικό πολιτισμό και τη χώρα, στο πλαίσιο της συνεργασίας της με άλλες εθνικές βιβλιοθήκες και διεθνείς οργανισμούς, καθώς και στο πλαίσιο της συμμετοχής της σε προγράμματα διαδανεισμού.”

Ο ρόλος της ΕΒΕ ως Εθνικού Βιβλιογραφικού Κέντρου συνοψίζεται στο να καταγράφει την ετήσια πνευματική παραγωγή, ενώ ο ρόλος της ως Εθνικού Καταλογογραφικού Κέντρου επικεντρώνεται στη δημιουργία εγγραφών, τις οποίες ελεύθερα διαμοιράζει η ΕΒΕ σε άλλες ελληνικές βιβλιοθήκες. Επιπλέον και προκειμένου να προβάλλει τον ελληνικό πολιτισμό, η ΕΒΕ διαμοιράζει τα δεδομένα της και σε βιβλιοθήκες του εξωτερικού [9].

### 2.2.2 Πρότυπα μεταγραμματισμού στην ΕΒΕ

Προκειμένου να διευκολυνθεί ο διαμοιρασμός εγγραφών με βιβλιοθήκες της Ελλάδας ή του εξωτερικού, η ΕΒΕ φροντίζει, ώστε οι εγγραφές που δημιουργεί να είναι σύμφωνες με διεθνή πρότυπα του βιβλιοθηκονομικού χώρου. Επιπλέον, λόγω της χρήσης του ελληνικού αλφαβήτου, το οποίο δε χρησιμοποιεί λατινικούς χαρακτήρες, η ΕΒΕ προχωρά σε ρομανοποίηση ονομάτων φυσικών προσώπων και συλλογικών οργάνων, ώστε η πλέον βασική πληροφορία, το όνομα, να μπορεί να αναγνωστεί από συστήματα ή ανθρώπους που δε γνωρίζουν το ελληνικό αλφάβητο.

Η μετατροπή των ελληνικών χαρακτήρων σε λατινικούς χαρακτήρες (ρομανοποίηση) γίνεται με τη μέθοδο του μεταγραμματισμού. Η μέθοδος του μεταγραμματισμού, όπως έχει προαναφερθεί, στηρίζεται στη μετατροπή των χαρακτήρων από ένα σύστημα γραφής σε ένα άλλο, αδιαφορώντας για την προφορά των λέξεων. Η μέθοδος αυτή έχει επιλεγεί από τις βιβλιοθήκες παγκοσμίως. Το πρότυπο που χρησιμοποιεί η ΕΒΕ είναι ο Πίνακας Ρομανοποίησης για τα ελληνικά που ανέπτυξαν η Αμερικάνικη Ένωση Βιβλιοθηκών (American Library Association - ALA) σε συνεργασία με τη Βιβλιοθήκη του Κογκρέσου (Library of Congress - LC) [5].

Ο λόγος που αρχικά επιλέχθηκε από την ΕΒΕ ο Πίνακας ρομανοποίησης ήταν η συστηματική αποστολή εγγραφών στο Διεθνές Αρχείο Καθιερωμένων Όρων VIF (Virtual International Authority File). Στο αρχείο αυτό, το κύριο σύστημα ρομανοποίησης είναι ο Πίνακας των ALA-LC, καθώς τα δεδομένα που συγκεντρώνονται στο VIF προέρχονται, ως επί το πλείστον, από βιβλιοθήκες των ΗΠΑ, μέλη του Προγράμματος Συνεργατικής Καταλογογράφησης που συντονίζει η Βιβλιοθήκη του Κογκρέσου.

Οι συνεργασίες της ΕΒΕ έχουν αυξηθεί και, στο πλαίσιο διαμοιρασμού εγγραφών με εθνικές βιβλιοθήκες, έχει αποφασίσει την εφαρμογή και του διεθνούς προτύπου ISO 843 (έλαβε την αριθμηση ΕΛΟΤ 743 στα ελληνικά) [1-2].

## 2.3 Αναφορά σε διαθέσιμες μεθόδους

### Mapping Method (Μέθοδος Αντιστοίχισης)

Η μέθοδος αντιστοίχισης, γνωστή και ως Mapping στη JavaScript, αναφέρεται στη διαδικασία σύνδεσης κλειδιών (keys) με τιμές (values), δημιουργώντας μία δομή δεδομένων που επιτρέπει την αποτελεσματική αποθήκευση και ανάκτηση πληροφοριών. Αυτή η έννοια υλοποιείται κυρίως μέσω των αντικειμένων (objects) και, πιο πρόσφατα, μέσω της κλάσης Map.

Σύμφωνα με τον Flanagan, τα αντικείμενα στη JavaScript λειτουργούν ως συσχετιστικοί πίνακες, επιτρέποντας την αντιστοίχιση μεταξύ ονομάτων ιδιοτήτων (κλειδιών) και τιμών. Η σύνταξη γι' αυτήν την αντιστοίχιση συχνά παρουσιάζεται ως `key:value` μέσα σε ένα αντικείμενο. [10]

### Η χρησιμότητα της Mapping Method στην Εφαρμογή Μεταγραμματισμού

Η μέθοδος αντιστοίχισης (Mapping) στη JavaScript παίζει καίριο ρόλο στην εφαρμογή τεχνικών μεταγραμματισμού. Σε μια πρόσφατη μελέτη, οι Chen et al. (2022) διαπίστωσαν ότι η χρήση δομών αντιστοίχισης στον μεταγραμματισμό μπορεί να βελτιώσει την απόδοση των μετασχηματισμών κώδικα κατά 30% σε σύνθετα έργα ανάπτυξης λογισμικού [11].

Ο Rodriguez (2021), στη διδακτορική του διατριβή, υποστηρίζει ότι οι τεχνικές αντιστοίχισης είναι θεμελιώδεις για την ανάπτυξη ευέλικτων συστημάτων μεταγραμματισμού. Συγκεκριμένα, αναφέρει: "Η χρήση των Map objects στη JavaScript επιτρέπει την κατασκευή πολύπλοκων σχημάτων μετασχηματισμού, με τρόπο που είναι τόσο αποδοτικός, όσο και εύκολα συντηρήσιμος" [12].

### Η σημασία της Mapping Method

Η χρήση της μεθόδου αντιστοίχισης (Mapping) είναι ζωτικής σημασίας στην εφαρμογή του μεταγραμματισμού, καθώς μεγιστοποιεί την ακρίβεια, ως εκ τούτου, και την αξιοπιστία στη μετατροπή χαρακτήρων και στοιχείων. Χωρίς την εφαρμογή της, τα ποσοστά αναντιστοιχίας, ασυνέπειας και ανακρίβειας στα αποτελέσματα του μεταγραμματισμού θα ήταν υψηλά. Η προσεχτική σχεδίαση και εφαρμογή της μεθόδου αντιστοίχισης μπορεί να καλύψει πλήρως τις ανάγκες και τις προσδοκίες των χρηστών και της ΕΒΕ.

#### 2.3.1 Πλεονεκτήματα και Μειονεκτήματα

Τα πλεονεκτήματα της μεθόδου αντιστοίχισης είναι τα εξής:

##### 1. Σαφής Καθορισμός Αντιστοιχίσεων:

Κάθε ελληνικός χαρακτήρας ή συνδυασμός χαρακτήρων αντιστοιχίζεται ρητά με έναν ή περισσότερους λατινικούς χαρακτήρες. Για παράδειγμα, ο ελληνικός χαρακτήρας "Α" αντιστοιχεί στον λατινικό "A", ενώ ο συνδυασμός "Θ" αντιστοιχεί στο "Th".

Για να επιτευχθεί αυτό, χρησιμοποιείται ένας mapping table (πίνακας αντιστοιχίσεων), που περιλαμβάνει όλους τους πιθανούς συνδυασμούς και τις αντίστοιχες μεταγραφές τους.

##### 2. Επεξεργασία Κειμένου:

Όταν ο χρήστης εισάγει ένα ελληνικό κείμενο ή πρόταση στην εφαρμογή, η μέθοδος αντιστοίχισης αναλαμβάνει την επεξεργασία της, διαβάζοντας κάθε χαρακτήρα ή συνδυασμό χαρακτήρων και αντικαθιστώντας τους με τους αντίστοιχους λατινικούς χαρακτήρες, σύμφωνα με τους κανόνες που καθορίζονται στο mapping table.

##### 3. Διαχείριση Ειδικών Περιπτώσεων:

Σε ορισμένες περιπτώσεις, οι κανόνες μεταγραμματισμού ενδέχεται να απαιτούν πιο σύνθετη λογική, όπως τη διαχείριση δίψηφων ή τρίψηφων γραμμάτων (π.χ. "Γκ" να γίνει "Gk") ή την προσαρμογή ανάλογα με τη θέση του χαρακτήρα στη λέξη.

Για να αντιμετωπιστούν αυτές οι ειδικές περιπτώσεις, η μέθοδος αντιστοίχισης ενσωματώνει αλγοριθμική λογική, η οποία αναγνωρίζει τη θέση του χαρακτήρα ή του συνδυασμού χαρακτήρων και εφαρμόζει τον κατάλληλο κανόνα αντιστοίχισης.

#### 4. Ευκολία Συντήρησης και Προσαρμογής:

Η μέθοδος αντιστοίχισης σχεδιάζεται με τέτοιο τρόπο, ώστε να είναι εύκολα παραμετροποιήσιμη. Αυτό σημαίνει ότι μελλοντικές αλλαγές ή προσθήκες στους κανόνες μεταγραμματισμού μπορούν να ενσωματωθούν εύκολα στο σύστημα, χωρίς μεγάλες τροποποιήσεις στον κώδικα.

Ως μειονεκτήματα της μεθόδου αντιστοίχισης μπορούν να θεωρηθούν τα παρακάτω:

##### 1. Ασυνέπεια Μεταξύ Προτύπων:

Τα διάφορα πρότυπα μεταγραμματισμού (όπως το ISO 843 και τα ALA-LC) μπορεί να παράγουν διαφορετικά αποτελέσματα για την ίδια είσοδο, οδηγώντας σε ασυνέπειες. Αυτό μπορεί να είναι προβληματικό, όταν απαιτείται ομοιομορφία, όπως σε βιβλιογραφικές βάσεις δεδομένων ή επίσημα έγγραφα.

##### 2. Διαχείριση Ειδικών Περιπτώσεων:

Η μέθοδος mapping μπορεί να δυσκολεύεται να διαχειριστεί ειδικές γλωσσικές περιπτώσεις, όπως οι δίφθογγοι, οι συνδυασμοί γραμμάτων ή τα διακριτικά. Αυτά τα στοιχεία μπορούν να περιπλέξουν τη διαδικασία mapping και να οδηγήσουν σε λάθη, αν ο προγραμματιστής δεν τα προβλέψει και δεν τα διαχειριστεί σωστά.

##### 3. Προκλήσεις Λόγω Συμφραζομένων:

Οι μέθοδοι mapping μπορεί να μην λαμβάνουν υπόψη το συμφραζόμενο στο οποίο χρησιμοποιείται ένας χαρακτήρας ή συνδυασμός χαρακτήρων. Για παράδειγμα, ο ίδιος ελληνικός χαρακτήρας μπορεί να μεταγράφεται διαφορετικά ανάλογα με τη θέση του σε μια λέξη ή πρόταση. Μια αυστηρή μέθοδος mapping μπορεί να παραβλέψει αυτές τις λεπτομέρειες.

### 2.3.2 Η Προτίμηση της Μεθόδου Αντιστοίχισης έναντι των Δομών if και switch

Η χρήση της μεθόδου αντιστοίχισης (mapping) στη διαδικασία ανάπτυξης του λογισμικού προσφέρει σημαντικά πλεονεκτήματα έναντι των παραδοσιακών δομών ελέγχου, όπως οι if και switch. Παρότι οι τελευταίες διατηρούν τη χρησιμότητά τους σε απλές περιπτώσεις, η μέθοδος αντιστοίχισης αποδεικνύεται ανώτερη, λόγω αποτελεσματικότητας, όταν απαιτείται μεγαλύτερη ευελιξία και επεκτασιμότητα. Οι λόγοι είναι οι εξής:

**Απλοποίηση της δομής του κώδικα:** Σε σύνθετες εφαρμογές, όπως ο μεταγραμματισμός, η χρήση δομών if και switch μπορεί να συντηρηθεί δύσκολα, εξαιτίας της γραμμικής ανάπτυξης και πολυπλοκότητας. Η μέθοδος αντιστοίχισης προσφέρει μια πιο ευέλικτη και λιτή προσέγγιση, επιτρέποντας τη σύνδεση κλειδιών με τις αντίστοιχες λειτουργίες, χωρίς την ανάγκη πολλαπλών συνθηκών.

**Επεκτασιμότητα:** Σε αντίθεση με τις δομές if και switch, όπου κάθε νέα επιλογή απαιτεί την προσθήκη νέου κώδικα, η μέθοδος αντιστοίχισης επιτρέπει την προσθήκη νέων λειτουργιών, μέσω απλής εισαγωγής νέων κλειδιών και τιμών στο αντικείμενο. Αυτό μειώνει την πολυπλοκότητα του κώδικα και διευκολύνει την ανάπτυξή του.

**Βελτίωση της συντηρησιμότητας:** Οι δομές if και switch ενδέχεται να προκαλέσουν δυσκολίες στη συντήρηση του κώδικα, ειδικά όταν αυξάνεται η πολυπλοκότητα των επιλογών. Η μέθοδος αντιστοίχισης επιτρέπει πιο ευέλικτη τροποποίηση, χωρίς να απαιτείται εκ νέου αναδιάρθρωση μεγάλων τμημάτων του προγράμματος.

**Μείωση του κινδύνου σφαλμάτων:** Ο απλός τρόπος με τον οποίο μπορούν να εισαχθούν νέες επιλογές μέσω της μεθόδου αντιστοίχισης μειώνει τις πιθανότητες σφαλμάτων, σε σύγκριση με τη χρήση δομών if και switch, όπου το αυξημένο μέγεθος του κώδικα αυξάνει την πιθανότητα λαθών.

**Ταχύτερη απόδοση σε εφαρμογές μεγάλης κλίμακας:** Τόσο σε εφαρμογές μεγάλης κλίμακας, όσο και σε εφαρμογές πολλαπλών επιλογών, η μέθοδος αντιστοίχισης (mapping) υπερτερεί έναντι των παραδοσιακών δομών συνθήκης if και switch. Αυτές οι δομές εκτελούν διαδοχικούς ελέγχους μέχρι να εντοπιστεί η κατάλληλη συνθήκη. Αυτή η διαδικασία, ακριβώς επειδή καθιστά πολύπλοκο τον έλεγχο, επιβραδύνει την απόδοση. Και το μειονέκτημα αυτό μεγεθύνεται, όταν η εφαρμογή απαιτεί πολλαπλές επιλογές ή όταν πρόκειται για εφαρμογή μεγάλης κλίμακας. Αντίθετα, η μέθοδος αντιστοίχισης, χρησιμοποιώντας hash tables ή Map objects, επιτρέπει πολύ γρήγορη πρόσβαση στη σωστή λειτουργία ή τιμή, εξασφαλίζοντας έτσι υψηλή απόδοση ακόμα και σε απαιτητικές εφαρμογές.[13]

Στις παρακάτω φωτογραφίες 2.1, 2.2 διαπιστώνεται η δυσκολία ανάγνωσης και περαιτέρω επέκτασης και συντήρησης του προγράμματος.

```

function convertGreekToLatinUsingIf(char) {
  if (char === 'Α') return 'A';
  else if (char === 'Β') return 'V';
  else if (char === 'Γ') return 'G';
  else if (char === 'Δ') return 'D';
  else if (char === 'Ε') return 'E';
  else if (char === 'Ζ') return 'Z';
  else if (char === 'Η') return 'I';
  else if (char === 'Θ') return 'Th';
  else if (char === 'Ι') return 'I';
  else if (char === 'Κ') return 'K';
  else if (char === 'Λ') return 'L';
  else if (char === 'Μ') return 'M';
  else if (char === 'Ν') return 'N';
  else if (char === 'Ξ') return 'X';
  else if (char === 'Ο') return 'O';
  else if (char === 'Π') return 'P';
  else if (char === 'Ρ') return 'R';
  else if (char === 'Σ') return 'S';
  else if (char === 'Τ') return 'T';
  else if (char === 'Υ') return 'Y';
  else if (char === 'Φ') return 'F';
  else if (char === 'Χ') return 'Ch';
  else if (char === 'Ψ') return 'Ps';
  else if (char === 'Ω') return 'O';
  else if (char === 'α') return 'a';
  else if (char === 'β') return 'v';
  else if (char === 'γ') return 'g';
  else if (char === 'δ') return 'd';
  else if (char === 'ε') return 'e';
  else if (char === 'ζ') return 'z';
  else if (char === 'η') return 'i';
  else if (char === 'θ') return 'th';
  else if (char === 'ι') return 'i';
  else if (char === 'κ') return 'k';
  else if (char === 'λ') return 'l';
  else if (char === 'μ') return 'm';
  else if (char === 'ν') return 'n';
  else if (char === 'ξ') return 'x';
  else if (char === 'ο') return 'o';
  else if (char === 'π') return 'p';
  else if (char === 'ρ') return 'r';
  else if (char === 'σ' || char === 'ς') return 's';
  else if (char === 'τ') return 't';
  else if (char === 'υ') return 'y';
  else if (char === 'φ') return 'f';
  else if (char === 'χ') return 'ch';
  else if (char === 'ψ') return 'ps';
  else if (char === 'ω') return 'o';
  else if (char === ' ') return ' ';
  else return '*';
}

console.log("Γειά σου Κόσμε".split('').map(convertGreekToLatinUsingIf).join(''));

```

Εικόνα 2.1: Μεταγραμματισμός της Αλφαβήτας με την χρήση if ελέγχου.

```

function convertGreekToLatinUsingSwitch(char) {
  switch (char) {
    case 'Α': return 'A';
    case 'Β': return 'B';
    case 'Γ': return 'G';
    case 'Δ': return 'D';
    case 'Ε': return 'E';
    case 'Ζ': return 'Z';
    case 'Η': return 'I';
    case 'Θ': return 'Th';
    case 'Ι': return 'I';
    case 'Κ': return 'K';
    case 'Λ': return 'L';
    case 'Μ': return 'M';
    case 'Ν': return 'N';
    case 'Ξ': return 'X';
    case 'Ο': return 'O';
    case 'Π': return 'P';
    case 'Ρ': return 'R';
    case 'Σ': return 'S';
    case 'Τ': return 'T';
    case 'Υ': return 'Y';
    case 'Φ': return 'F';
    case 'Χ': return 'Ch';
    case 'Ψ': return 'Ps';
    case 'Ω': return 'O';
    case 'α': return 'a';
    case 'β': return 'v';
    case 'γ': return 'g';
    case 'δ': return 'd';
    case 'ε': return 'e';
    case 'ζ': return 'z';
    case 'η': return 'i';
    case 'θ': return 'th';
    case 'ι': return 'i';
    case 'κ': return 'k';
    case 'λ': return 'l';
    case 'μ': return 'm';
    case 'ν': return 'n';
    case 'ξ': return 'x';
    case 'ο': return 'o';
    case 'π': return 'p';
    case 'ρ': return 'r';
    case 'σ': case 'ς': return 's';
    case 'τ': return 't';
    case 'υ': return 'y';
    case 'φ': return 'f';
    case 'χ': return 'ch';
    case 'ψ': return 'ps';
    case 'ω': return 'o';
    case ' ': return ' ';
    default: return '*';
  }
}

console.log("Γειά σου Κόσμε".split(' ').map(convertGreekToLatinUsingSwitch).join(' '));

```

Εικόνα 2.2: Μεταγραμματισμός της Αλφαβήτας με την χρήση Switch ελέγχου.

Και η πολυπλοκότητα αυξάνεται με την προσθήκη νέων χαρακτήρων, καθώς αυξάνονται και οι συνθήκες.

### 2.3.3 Μέθοδοι που έχουν χρησιμοποιηθεί σε transliteration

#### 1. Μέθοδος Αντιστοίχισης vs Στατιστική Προσέγγιση

Η μέθοδος αντιστοίχισης, η οποία χρησιμοποιείται στη συγκεκριμένη εφαρμογή μεταγραμματισμού, βασίζεται στη δημιουργία προκαθορισμένων αντιστοιχιών μεταξύ χαρακτήρων των αλφαβήτων. Στο πλαίσιο αυτό, ο μεταγραμματισμός πραγματοποιείται μέσω σταθερών και προβλέψιμων κανόνων, που εξασφαλίζουν την ορθότητα των αποτελεσμάτων. Η προσέγγιση που περιγράφεται στη διατριβή "**Automatic Validation of Terminology Translation Consistency**" επικεντρώνεται στη χρήση στατιστικών μοντέλων, για την επικύρωση της συνέπειας στις μεταφράσεις, μέσα από τη συνεκτίμηση του εύρους πιθανοτήτων και της συχνότητας των αποτελεσμάτων της μεταγραφής. Συγκεκριμένα, η στατιστική αυτή προσέγγιση αξιοποιεί το **Gaussian Mixture Model** (GMM) και δείκτες όπως ο **Herfindahl-Hirschman Index** (HHI), για να μετρήσει τη συνέπεια μεταγραφής όρων σε ένα πολύγλωσσο σώμα κειμένων.

Η μέθοδος αντιστοίχισης, αν και αποτελεσματική σε καλά δομημένα συστήματα, έχει περιορισμένες δυνατότητες σε περιπτώσεις όπου τα δεδομένα είναι σύνθετα ή ασαφή. Από την άλλη, το στατιστικό μοντέλο αποβλέπει στον έλεγχο της ποιότητας, επομένως και εγκυρότητας, της μεταγραφής. Επιπλέον, προσφέρει ευελιξία και δυνατότητα βελτίωσης μέσω μηχανικής μάθησης, διαχειριζόμενο μεγαλύτερα σύνολα δεδομένων και αντιμετωπίζοντας πιο περίπλοκες μετατροπές.

#### 2. Μέθοδος Αντιστοίχισης vs Αντίστροφος Μεταγραμματισμός

Η μέθοδος αντιστοίχισης ακολουθεί μια ευθεία διαδικασία μετατροπής χαρακτήρων από το ελληνικό στο λατινικό αλφάβητο, μέσω προκαθορισμένων κανόνων. Η μέθοδος αντίστροφου μεταγραμματισμού, που αναλύεται στη διατριβή "**Machine Transliteration: A Backward Approach**", χρησιμοποιεί μια γενετική προσέγγιση, για να ανασυνθέσει λέξεις από την αρχική γλώσσα σε μια άλλη, μέσω φωνητικών και πιθανολογικών μοντέλων. Αυτή η προσέγγιση είναι πιο περίπλοκη και ενσωματώνει αλγορίθμους όπως οι **Weighted Finite-State Transducers** (WFSTs) και τεχνικές πιθανολογίας, οι οποίες καθιστούν δυνατή την αντιμετώπιση ασαφών ή ανακριβών δεδομένων, όπως οι θορυβώδεις αναγνώσεις από OCR.

## 2.4 Επίλογος

Η Εθνική Βιβλιοθήκη της Ελλάδας, ως εθνικός φορέας συλλογής, ταξινόμησης και διαφύλαξης της πνευματικής κληρονομιάς και παραγωγής του ελληνικού πολιτισμού, συνεργάζεται και αλληλεπιδρά με άλλες βιβλιοθήκες του εσωτερικού και του εξωτερικού. Ειδικότερα, ως Εθνικό Βιβλιογραφικό και Καταλογογραφικό Κέντρο, για τον διαμοιρασμό βιβλιογραφικών δεδομένων έχει υιοθετήσει τα πρότυπα μεταγραμματισμού ALA-LC, λόγω της συνεργασίας της με το Διεθνές Αρχείο Καθιερωμένων Όρων (VIAF), και ISO843, λόγω της συνεργασίας της με άλλες εθνικές βιβλιοθήκες. Στην εφαρμογή των προτύπων μεταγραμματισμού, τα μέγιστα και βέλτιστα αποτελέσματα μπορεί να επιτύχει η ΕΒΕ

με τον σχεδιασμό μεθόδων αντιστοίχισης (mapping methods), με τρόπο που να αξιοποιούνται τα πλεονεκτήματά τους και να ελαχιστοποιούνται τα μειονεκτήματα.

Η μέθοδος mapping προτιμήθηκε λόγω του λιτού περιεχομένου της, ταυτόχρονα ευανάγνωστου, που επιταχύνει ως τέτοια την επεξεργασία των μεταδεδομένων, έναντι άλλων παραδοσιακών μεθόδων (if, switch), οι οποίες ενδείκνυνται για μειωμένο όγκο δεδομένων και μικρής εμβέλειας πίνακες και λειτουργούν πιο αργά, ενώ δεν επιτρέπουν εύκολη επεκτασιμότητα και συντήρηση. Τέλος, παρουσιάστηκαν και σύγχρονες μέθοδοι μεταγραμματισμού, που δημιουργήθηκαν για τη μεταγραφή μη ευρωπαϊκών γλωσσών (στατιστική προσέγγιση, αντίστροφος μεταγραμματισμός), κατάλληλες για πιο απαιτητικές και σύνθετες εφαρμογές και με διαφορετική στόχευση.

## Κεφάλαιο 3ο: Μέθοδοι

### 3.1 Εισαγωγή

Στο κεφάλαιο αυτό, παρουσιάζεται αιτιολογημένα η επιλογή των τεχνολογιών που προτείνονται στην εφαρμογή του εργαλείου του μεταγραμματισμού για τις ανάγκες της ΕΒΕ. Υιοθετούνται η γλώσσα σήμανσης για ιστοσελίδες HTML, η γλώσσα υπολογιστή CSS και η γλώσσα προγραμματισμού JavaScript. Περιγράφονται τα δύο βασικά πρότυπα στα οποία στηρίζεται ο σχεδιασμός του εργαλείου, το διεθνές πρότυπο ISO 843 και το τοπικό πρότυπο ALA-LC. Αναπτύσσονται οι παρεμβάσεις που προτείνονται στην παρούσα εργασία για τη βέλτιστη απόδοση του προγράμματος μεταγραμματισμού. Σχεδιάζονται, τέλος, η πιλοτική εφαρμογή του από καταλογγράφους της ΕΒΕ με οδηγό τρία κατευθυντήρια ερωτήματα, καθώς και η συνακόλουθη αξιολόγηση ως προς τη χρήση και τα αποτελέσματα της εφαρμογής.

### 3.2 Γενικές αποφάσεις/προδιαγραφές

Για την ανάπτυξη της εφαρμογής μεταγραμματισμού που θα χρησιμοποιηθεί από την Εθνική Βιβλιοθήκη της Ελλάδος (ΕΒΕ), η επιλογή των τεχνολογιών ήταν πολύ σημαντική. Οι τεχνολογίες που επιλέχθηκαν είναι οι **HTML**, **CSS** και **JavaScript**. Αυτές είναι οι πιο διαδεδομένες παγκοσμίως στο χώρο της ανάπτυξης ιστοσελίδων και παρέχουν τα απαραίτητα εργαλεία που απαιτούνται για την ανάπτυξη διαδραστικών και φιλικών προς τον χρήστη εφαρμογών.

Αρχικά, η **HTML** (HyperText Markup Language) αποτελεί τη βάση πάνω στην οποία αναπτύχθηκε όλη η ιστοσελίδα. Οι λόγοι για τους οποίους προτάσσεται στην παρούσα εργασία είναι οι εξής:

- **Δημιουργία Δομής Σελίδας:** η HTML, ως η πιο δημοφιλής για την ευχρηστία της γλώσσα σήμανσης υπερκειμένου (αποδοτικότητα και ικανοποίηση από τον χρήστη), χρησιμοποιείται για τη δημιουργία της βασικής δομής της ιστοσελίδας που φιλοξενεί το εργαλείο μεταγραμματισμού. Αυτό περιλαμβάνει τα πεδία κειμένου, τα κουμπιά δράσης, τα εικονίδια και άλλες πολλές διαδραστικές λειτουργίες [14].
- **Συμβατότητα με Περιηγητές:** είναι η πιο βασική και ευρέως υποστηριζόμενη γλώσσα για την κατασκευή ιστοσελίδων. Με αυτόν τον τρόπο, εξασφαλίζεται ότι η εφαρμογή θα είναι προσβάσιμη μέσω οποιουδήποτε προγράμματος περιήγησης χρησιμοποιούν οι χρήστες [14].

Η **CSS** (Cascading Style Sheets) χρησιμοποιήθηκε για την εμφάνιση και τη διάταξη των στοιχείων της εφαρμογής. Σημαντικοί λόγοι για την χρήση της CSS τεχνολογίας:

- **Σχεδιασμός Διεπαφής Χρήστη:** Η CSS χρησιμοποιείται για το σχεδιασμό των οπτικών στοιχείων της εφαρμογής, προσφέροντας μια ευχάριστη και εύχρηστη εμπειρία προς τον χρήστη. Σχεδιάστηκαν ενότητες με ευανάγνωστο κείμενο, καθαρή διάταξη και προσαρμόσιμα στοιχεία, που ανταποκρίνονται στις ανάγκες των χρηστών [15].
- **Ανταποκρίσιμος Σχεδιασμός:** επιτρέπει τη δημιουργία μιας δεκτικής και ανταποκρίσιμης (responsive) ιστοσελίδας, η οποία έχει την δυνατότητα να προσαρμόζεται αυτόματα με την συσκευή και το μέγεθος της οθόνης (π.χ., επιτραπέζιοι

υπολογιστές, laptops, tablets, smartphones). Αυτό είναι σημαντικό χαρακτηριστικό, καθώς οι χρήστες μπορούν να χρησιμοποιήσουν αρκετές διαφορετικές συσκευές, για να έχουν πρόσβαση στο εργαλείο [15].

Η **JavaScript** είναι η γλώσσα προγραμματισμού που χρησιμοποιήθηκε για τη δημιουργία των διαδραστικών στοιχείων. Στην εφαρμογή αυτή, η JavaScript επιλέχθηκε για τα παρακάτω πλεονεκτήματά της:

- **Διαδραστικότητα:** Ενισχύει τη διαδραστικότητα της εφαρμογής, επιτρέποντας άμεση ανταπόκριση στις εισαγωγές του χρήστη και την αποδοτική επεξεργασία δεδομένων [10].
- **Αποτελεσματική Διαχείριση Δεδομένων:** Διευκολύνει την εκτέλεση αλγορίθμων απευθείας στον φυλλομετρητή, μειώνοντας την ανάγκη για εξωτερικές κλήσεις [10].
- **Ενσωμάτωση με Άλλες Τεχνολογίες:** Επιτρέπει τη συνεχή ενσωμάτωσή της με τις τεχνολογίες HTML και CSS, διασφαλίζοντας πλήρη παραμετροποίηση της εφαρμογής [10].

Μία σημαντική, επίσης, απόφαση που λήφθηκε, για την απλοποίηση, την διαχώριση και την εύκολη επεκτασιμότητα του προγράμματος μεταγραμματισμού, είναι να εγγραφούν οι αντιστοιχίσεις ελληνικών και λατινικών χαρακτήρων σε 4 διαφορετικά mappings, δύο για κάθε πρότυπο, για τα νέα ελληνικά και τα αρχαία ελληνικά.

### 3.3 ISO 843

#### 3.3.1 Σύντομη περιγραφή του προτύπου

Το διεθνές πρότυπο ISO 843 πρωτοδημοσιεύθηκε το 1968 με τον τίτλο “International system for the transliteration of Greek characters into Latin characters” από τον Διεθνή Οργανισμό Τυποποίησης (International Organization for Standardization - ISO) [16]. Όπως φαίνεται από τον τίτλο, η πρώτη έκδοση του προτύπου επικεντρωνόταν μόνο στη διαδικασία του μεταγραμματισμού. Το πρότυπο αναθεωρήθηκε το 1997, συμπεριλαμβάνοντας και τη μέθοδο της μεταγραφής. Σύμφωνα με τις διαδικασίες του ISO, όλα τα πρότυπα αναθεωρούνται ανά τακτά διαστήματα. Η τελευταία αναθεώρηση του προτύπου επιβεβαίωσε πως δε χρειάζεται κάποια αλλαγή και πως το πρότυπο συνεχίζει να ισχύει με τη μορφή που έλαβε στην έκδοση του 1997 [17]. Το πρότυπο μεταφράστηκε στα ελληνικά από τον Ελληνικό Οργανισμό Τυποποίησης, λαμβάνοντας την αρίθμηση ΕΛΟΤ 743 [1].

#### 3.3.2 Βασικοί κανόνες

Το πρότυπο περιλαμβάνει 2 πίνακες μετατροπής χαρακτήρων. Ο ένας για μεταγραμματισμό και ο άλλος για μεταγραφή. Για τους σκοπούς της παρούσας εργασίας χρησιμοποιήθηκε ο πίνακας μεταγραμματισμού. Επιπλέον, το συγκεκριμένο πρότυπο διαθέτει πίνακα μετατροπής για μη αλφαβητικούς χαρακτήρες (π.χ., τόνος, ερωτηματικό, ψιλή, δασεία κ.λπ.), παρέχοντας τη δυνατότητα επιλογής του μονοτονικού ή του πολυτονικού συστήματος. Η μετατροπή των μη αλφαβητικών χαρακτήρων γίνεται με την προσθήκη ειδικών σημείων πάνω στους αλφαβητικούς χαρακτήρες. Για παράδειγμα, το πεζό γράμμα “α” μεταγραμματίζεται ως “a”, ενώ το τονούμενο πεζό γράμμα “ά” μεταγραμματίζεται ως “á”.

Το πρότυπο προβλέπει δύο εξαιρέσεις. Η μία αφορά τη μεταγραφή των διφθόγγων και η δεύτερη τον χαρακτήρα “ς” (σίγμα τελικό).

Όσον αφορά τους συνδυασμούς δύο ή περισσότερων χαρακτήρων, το πρότυπο αναφέρει πως ο μεταγραμματισμός γίνεται για τον κάθε χαρακτήρα ξεχωριστά. Ο κανόνας αυτός δεν ισχύει μόνο για τους διφθόγγους “ΑΥ, αυ, ΕΥ, ευ, ΟΥ, ου”, οι οποίοι μεταγραμματίζονται ως “ΑU, au, EU, eu, OU, ou”. Δηλαδή, ο χαρακτήρας “υ”, κατ’εξαιρέση, μεταγραμματίζεται ως “u” και όχι ως “y”, όπως αναφέρει ο κύριος πίνακας μεταγραμματισμού.

Η δεύτερη εξαίρεση αφορά στο τελικό σίγμα (ς), που μεταγραμματίζεται σε λατινικό s, το ίδιο όπως και ο ελληνικός χαρακτήρας σ (σίγμα - sigma). Το πρότυπο αναφέρει, στο κείμενο της εξαίρεσης, οδηγίες για τον επαναγραμματισμό, δηλαδή την αντίστροφη μετατροπή από λατινικούς χαρακτήρες σε ελληνικούς, ορίζοντας: όταν εντοπίζεται “s” στο τέλος μιας λέξης, να αντικαθίσταται με τελικό σίγμα (ς), αντί για σίγμα (σ). Στους σκοπούς δημιουργίας της εφαρμογής δεν περιλαμβάνεται ο επαναγραμματισμός. Γι’ αυτό το λόγο, το τελευταίο σημείο της εξαίρεσης δε λήφθηκε υπόψη κατά την υλοποίηση του σχετικού αλγορίθμου μεταγραμματισμού.

Ο πίνακας αντιστοίχισης του προτύπου ISO 843 (ή ΕΛΟΤ 743) είναι ο εξής:

Πίνακας 3.1: Πίνακας μεταγραμματισμού προτύπου ISO 843

	Ελληνικοί χαρακτήρες	Λατινικοί χαρακτήρες
1	Α, α	A, a
2	Β, β	V, v
3	Γ, γ	G, g
4	Δ, δ	D, d
5	Ε, ε	E, e
6	Ζ, ζ	Z, z
7	Η, η	$\bar{I}, \bar{i}$ ή $I^-, i^-$
8	Θ, θ	TH, th
9	Ι, ι	I, i
10	Κ, κ	K, k
11	Λ, λ	L, l
12	Μ, μ	M, m
13	Ν, ν	N, n
14	Ξ, ξ	X, x
15	Ο, ο	O, o
16	Π, π	P, p
17	Ρ, ρ	R, r
18	Σ, σ, ς	S, s, s (2)
19	Τ, τ	T, t
20	Υ, υ	Y, y
21	Φ, φ	F, f
22	Χ, χ	CH, ch
23	Ψ, ψ	PS, ps
24	Ω, ω	$\bar{O}, \bar{o}$ ή $O^-, o^-$

### Υλοποίηση κανόνων στο εργαλείο μεταγραμματισμού

Το αρχικό βήμα ήταν η δημιουργία κώδικα για την αναγνώριση χαρακτήρων της αρχαίας και της νέας ελληνικής. Για την ανίχνευση των χαρακτήρων της αρχαίας ελληνικής με τη χρήση πνευμάτων και διαφόρων τονισμών, χρησιμοποιήθηκαν διάφοροι ιστότοποι, όπως η Wikipedia [18], το Polytonic Project [19] και το Lexilogos [20]. Έπειτα, όλοι αυτοί οι χαρακτήρες αντιστοιχίστηκαν σε λατινικούς χαρακτήρες. Στην παρακάτω εικόνα 3.1, παρουσιάζεται η αναγνώριση χαρακτήρων της νέας ελληνικής και η αντιστοιχισή τους σε λατινικούς χαρακτήρες ή σε ειδικούς χαρακτήρες που ορίζει το πρότυπο, π.χ. το γράμμα Ήτα “Η” αντιστοιχίζεται στον ειδικό χαρακτήρα “ $\bar{I}$ ”. Ο πίνακας αντιστοίχισης της νέας ελληνικής περιλαμβάνει και κεφαλαίους και πεζούς χαρακτήρες.

```
//ISO 843 Modern

//Single-letters
//Uppercase
'Α': 'A', 'Β': 'V', 'Γ': 'G', 'Δ': 'D', 'Ε': 'E', 'Ζ': 'Z', 'Η': 'I', 'Θ': 'Th',
'Ι': 'I', 'Κ': 'K', 'Λ': 'L', 'Μ': 'M', 'Ν': 'N', 'Ξ': 'X', 'Ο': 'O', 'Π': 'P',
'Ρ': 'R', 'Σ': 'S', 'Τ': 'T', 'Υ': 'Y', 'Φ': 'F', 'Χ': 'Ch', 'Ψ': 'Ps', 'Ω': 'O',
//Lowercase
'α': 'a', 'β': 'v', 'γ': 'g', 'δ': 'd', 'ε': 'e', 'ζ': 'z', 'η': 'i', 'θ': 'th',
'ι': 'i', 'κ': 'k', 'λ': 'l', 'μ': 'm', 'ν': 'n', 'ξ': 'x', 'ο': 'o', 'π': 'p',
'ρ': 'r', 'σ': 's', 'ς': 's', 'τ': 't', 'υ': 'y', 'φ': 'f', 'χ': 'ch', 'ψ': 'ps', 'ω': 'o',
```

Εικόνα 3.1: Αντιστοίχιση Νέων Ελληνικών ISO 843

Στην εικόνα 3.2, παρουσιάζεται η αναγνώριση χαρακτήρων της αρχαίας ελληνικής και η αντιστοίχισή τους σε λατινικούς χαρακτήρες ή σε ειδικούς χαρακτήρες που ορίζει το πρότυπο. Οι χαρακτήρες της αρχαίας ελληνικής μπορεί να λαμβάνουν πολλά διακριτικά ταυτόχρονα, π.χ. ένα πνεύμα μαζί με τόνο. Για κάθε συνδυασμό απαιτήθηκε η προσθήκη επιπλέον χαρακτήρων, τονούμενοι ή μη τονούμενοι χαρακτήρες, χαρακτήρες με ή χωρίς πνεύματα, χαρακτήρες με συνδυασμό πνευμάτων και τονισμού (π.χ. ένα φωνήεν με ψιλή και τόνο). Ο πίνακας αντιστοίχισης της αρχαίας ελληνικής περιλαμβάνει και κεφαλαίους και πεζούς χαρακτήρες.

```
//ISO 843 Ancient

//Single-letters
//Uppercase
'Α': 'A', 'Β': 'V', 'Γ': 'G', 'Δ': 'D', 'Ε': 'E', 'Ζ': 'Z', 'Η': 'I', 'Θ': 'Th',
'Ι': 'I', 'Κ': 'K', 'Λ': 'L', 'Μ': 'M', 'Ν': 'N', 'Ξ': 'X', 'Ο': 'O', 'Π': 'P',
'Ρ': 'R', 'Σ': 'S', 'Τ': 'T', 'Υ': 'Y', 'Φ': 'F', 'Χ': 'Ch', 'Ψ': 'Ps', 'Ω': 'O',
//Lowercase
'α': 'a', 'β': 'v', 'γ': 'g', 'δ': 'd', 'ε': 'e', 'ζ': 'z', 'η': 'i', 'θ': 'th',
'ι': 'i', 'κ': 'k', 'λ': 'l', 'μ': 'm', 'ν': 'n', 'ξ': 'x', 'ο': 'o', 'π': 'p',
'ρ': 'r', 'σ': 's', 'ς': 's', 'τ': 't', 'υ': 'y', 'φ': 'f', 'χ': 'ch', 'ψ': 'ps', 'ω': 'o',
```

Εικόνα 3.2: Αντιστοίχιση Αρχαίων Ελληνικών ISO 843

Το πρότυπο ISO 843 επιτρέπει τον μεταγραμματισμό όλων των μη αλφαβητικών χαρακτήρων. Όποιος φορέας υλοποιεί το πρότυπο μπορεί να επιλέξει αν θα εφαρμόσει τον μεταγραμματισμό μόνο για το μονοτονικό ή και για το πολυτονικό σύστημα. Η ΕΒΕ αποφάσισε την υλοποίηση του προτύπου με το μονοτονικό σύστημα. Έτσι, αναπτύχθηκε κώδικας μόνο για τους χαρακτήρες της πρώτης στήλης στον πίνακα 3.2.



```
//Oxia ( ` )
//Uppercase
'A': 'Á', 'E': 'É', 'H': 'Ή', 'I': 'Ϊ', 'O': 'Ó', 'Y': 'Ύ', 'Ω': 'Ώ',
```

Εικόνα 3.4: Τονισμοί με οξεία

```
//Perispomeni ( ~ )
//Lowercase
'| 'ã': 'á', 'ñ': 'í', 'ĩ': 'í', 'ũ': 'ý', 'õ': 'ó',
```

Εικόνα 3.5: Τονισμοί με περισπωμένη

```
'ĩ': 'í',
```

Εικόνα 3.6: Παράδειγμα με διαλυτικά

Το πρότυπο ISO 843, αν και μεταγραμματίζει τον χαρακτήρα “Υ, υ” ως “Y, y”, στις περιπτώσεις των διφθόγγων “ΑΥ, αυ, ΕΥ, ευ, ΟΥ, ου” ορίζει πως ο χαρακτήρας “Υ, υ” πρέπει να μεταγραμματιστεί ως “U, u”. Η εικόνα 3.7, παρουσιάζει την αντιστοίχιση των διφθόγγων “ΑΥ, αυ, ΕΥ, ευ, ΟΥ, ου” και η εικόνα 3.8, παρουσιάζει τον αλγόριθμο που δημιουργήθηκε για τον χειρισμό της συγκεκριμένης περίπτωσης. Ο κώδικας πρώτα ανιχνεύει τις αντιστοιχίσεις που περιλαμβάνουν δύο χαρακτήρες και μετά τις αντιστοιχίσεις του ενός μόνο χαρακτήρα.

```
//2 letters
//Shmeiosh 1
//Uppercase
'Aυ': 'Au', 'Ευ': 'Eu', 'Ου': 'Ou',
'Aύ': 'Aú', 'Εύ': 'Eú', 'Ού': 'Oú',
//Lowercase
'αυ': 'au', 'ευ': 'eu', 'ου': 'ou',
'αύ': 'aú', 'εύ': 'eú', 'ού': 'oú',
```

Εικόνα 3.7: Αντιστοιχίες με δύο χαρακτήρες

```

//Handle 2-letter combinations
const twoLetterCombo = text[i] + text[i+1];
if(transliterationMap.hasOwnProperty(twoLetterCombo)) {
    result += transliterationMap[twoLetterCombo];
    i += 2; // Move to the next pair of characters
    isWordStart = false;
    continue;
}

```

Εικόνα 3.8: Έλεγχος αντιστοιχιών με δύο χαρακτήρες

Το πρότυπο, επιπλέον, ορίζει ως εξαίρεση τον μεταγραμματισμό του τελικού σίγμα “ς” με τον λατινικό χαρακτήρα “s”. Για την περίπτωση αυτή, δε χρειάστηκε επιπλέον κώδικας, πέρα από την προσθήκη της αντιστοίχισης στους μονούς χαρακτήρες.

### 3.4 Πρότυπο ALA-LC Romanization table for Greek

Οι πίνακες ρομανοποίησης, δηλαδή μεταγραμματισμού χαρακτήρων μη λατινικών αλφαβήτων σε λατινικούς χαρακτήρες, αναπτύσσονται από κοινού από τη Βιβλιοθήκη του Κογκρέσου (Library of Congress - LC) και την Αμερικάνικη Βιβλιοθηκονομική Ένωση (American Library Association - ALA). Οι πίνακες δημιουργήθηκαν για τη διευκόλυνση της καταλογογράφησης ξενόγλωσσου υλικού και την υποστήριξη εργασιών, όπως διακίνηση υλικού, προσκτήσεις, παραλαβή περιοδικών εκδόσεων, ταξιθέτηση, εξυπηρέτηση χρηστών και, κυρίως, για την προβολή δεδομένων σε καταλόγους βιβλιοθηκών [21]. Οι πίνακες ρομανοποίησης έχουν φτάσει τους 77 και αφορούν σε γλώσσες και διαλέκτους από όλο τον κόσμο.

#### 3.4.1 Σύντομη περιγραφή του προτύπου

Ο ισχύων πίνακας ρομανοποίησης για τα Ελληνικά δημοσιεύθηκε το 1997 και ανανεώθηκε το 2010 [5]. Λαμβάνει υπόψη του αν το υπό μεταγραμματισμό κείμενο είναι στα αρχαία ή στα νέα ελληνικά, καθώς και την προέλευση των νεοελληνικών λέξεων.

#### 3.4.2 Βασικοί κανόνες

Σε γενικές γραμμές, το πρότυπο ορίζει τους ίδιους κανόνες μεταγραμματισμού των χαρακτήρων του ελληνικού αλφαβήτου, με μόνη διαφορά τον μεταγραμματισμό του χαρακτήρα “Β,β”, ο οποίος, ανάλογα αν είναι στα αρχαία ή στα νέα ελληνικά, μεταγραμματίζεται διαφορετικά. Επιπλέον, κανόνες υπάρχουν για τον μεταγραμματισμό των διφθόγγων “γκ, μπ, ντ”, ανάλογα με τη θέση που έχουν σε μία λέξη (αρχή, μέση, τέλος), και για τους διφθόγγους “αυ, ευ, ηυ, ου, υι, και ων”, όπου το γράμμα “υ” μεταγραμματίζεται κατ’ εξαίρεση ως “u”.

Η πιο σημαντική εξαίρεση αφορά στον μεταγραμματισμό χαρακτήρων των νεοελληνικών λέξεων, καθώς, σύμφωνα με το πρότυπο, πρέπει να ληφθεί υπόψη η προέλευση της λέξης, δηλαδή, αν η νέα ελληνική λέξη προέρχεται από αρχαία ελληνική λέξη, η οποία στην αλεξανδρινή γραφή της έπαιρνε δασεία (´). Σε αυτή την περίπτωση, η δασεία πρέπει να αποδοθεί

### Κεφάλαιο 3

ως υπάρχουσα στη νεοελληνική λέξη. Ο χαρακτήρας που χρησιμοποιείται για τον μεταγραμματισμό της δασείας είναι ο λατινικός χαρακτήρας “H, h”. Για παράδειγμα, η λέξη “ιστορία” μεταγραμματίζεται ως “historia”, επειδή στα αρχαία ελληνικά η λέξη λάμβανε δασεία (ιστορία). Η δασεία στα αρχαία ελληνικά χρησιμοποιούνταν σε φωνήεντα ή διφθόγγους στην αρχή μίας λέξης, όπως και σε κάποιες λέξεις που ξεκινούσαν με το γράμμα (P, ρ). Στο πρότυπο ορίζεται πως ο μεταγραμματισμός του δασυόμενου “P” ή “ρ” θα γίνεται μόνον όταν η δασεία εμφανίζεται στο κείμενο. Αν εμφανίζεται, μόνον τότε μεταγραμματίζεται ο χαρακτήρας ως Rh/rh.

Ο πίνακας αντιστοίχισης για το “ALA-LC Romanization tables - Greek”, είναι ο εξής:

Πίνακας 3.3: Πίνακας αντιστοίχισης “ALA-LC Romanization tables - Greek”

<i>Vernacular</i>	<i>Romanization</i>	<i>Vernacular</i>	<i>Romanization</i>
A	A	α	a
B	B (Ancient and Medieval) V (Modern)	β	b (Ancient and Medieval) v (Modern)
Γ	G	γ	g
		γγ	ng
Γκ	Gk	γκ	gk (initially and finally) nk (medially)
		γξ	nx
		γχ	nch
Δ	D	δ	d
E	E	ε	e
Z	Z	ζ	z
H	Ē	η	ē
Θ	Th	θ	th
I	I	ι	i
K	K	κ	k
Λ	L	λ	l
M	M	μ	m
Μπ	B	μπ	b (initially) mp (medially and finally)
N	N	ν	n
Nτ	Ḑ	ντ	ḑ (initially) nt (medially and finally)
Ξ	X	ξ	x

Ο	O	ο	o
Π	P	π	p
Ρ	R	ρ	r
Ρ	Rh (see notes)	ῥ	rh (see notes)
Σ, C	S	σ, c, ς (final)	s
Τ	T	τ	t
Υ	Y	υ	y (u in diphthongs αυ, ευ, ηυ, ου, υι, and ωυ)
Φ	Ph	φ	ph
Χ	Ch	χ	ch
Ψ	Ps	ψ	ps
Ω	Ō	ω	ō

### 3.4.3 Υλοποίηση κανόνων στο εργαλείο μεταγραμματισμού

Το αρχικό βήμα ήταν η δημιουργία κώδικα για την αναγνώριση χαρακτήρων της αρχαίας και της νέας ελληνικής. Για την ανίχνευση των χαρακτήρων της αρχαίας ελληνικής με τη χρήση πνευμάτων και διαφόρων τονισμών χρησιμοποιήθηκαν διάφοροι ιστότοποι, όπως η Wikipedia [18], το Polytonic Project [19] και το Lexilogos [20]. Έπειτα, όλοι αυτοί οι χαρακτήρες αντιστοιχίστηκαν σε λατινικούς χαρακτήρες. Στην εικόνα 3.9, παρουσιάζεται η αναγνώριση χαρακτήρων της νέας ελληνικής και η αντιστοίχισή τους σε λατινικούς χαρακτήρες ή σε ειδικούς χαρακτήρες που ορίζει το πρότυπο, π.χ. το γράμμα ωμέγα “Ω” αντιστοιχίζεται στον ειδικό χαρακτήρα “Ō”. Ο πίνακας αντιστοίχισης της νέας ελληνικής περιλαμβάνει και κεφαλαίους και πεζούς χαρακτήρες.

```
//Romanization Modern
//Single-letters
//Uppercase
'Α': 'A', 'Β': 'V', 'Γ': 'G', 'Δ': 'D', 'Ε': 'E', 'Ζ': 'Z', 'Η': 'Ē', 'Θ': 'Th',
'Ι': 'I', 'Κ': 'K', 'Λ': 'L', 'Μ': 'M', 'Ν': 'N', 'Ξ': 'X', 'Ο': 'O', 'Π': 'P',
'Ρ': 'R', 'Σ': 'S', 'Τ': 'T', 'Υ': 'Y', 'Φ': 'Ph', 'Χ': 'Ch', 'Ψ': 'Ps', 'Ω': 'Ō',
//Lowercase
'α': 'a', 'β': 'v', 'γ': 'g', 'δ': 'd', 'ε': 'e', 'ζ': 'z', 'η': 'ē', 'θ': 'th',
'ι': 'i', 'κ': 'k', 'λ': 'l', 'μ': 'm', 'ν': 'n', 'ξ': 'x', 'ο': 'o', 'π': 'p',
'ρ': 'r', 'σ': 's', 'ς': 's', 'τ': 't', 'υ': 'y', 'φ': 'ph', 'χ': 'ch', 'ψ': 'ps', 'ω': 'ō',
```

Εικόνα 3.9: Αντιστοιχίσεις Νέων Ελληνικών ALA-LC Romanization tables

Στις εικόνες 3.10, 3.11, 3.12, παρουσιάζονται οι αντιστοιχίσεις χαρακτήρων της αρχαίας ελληνικής με τους λατινικούς ή ειδικούς χαρακτήρες που ορίζει το πρότυπο. Επειδή στα αρχαία ελληνικά χρησιμοποιούνται τα πνεύματα και οι τόνοι, προστέθηκαν σχόλια για την ευαναγνωσία του κώδικα. Ο πίνακας αντιστοίχισης της αρχαίας ελληνικής περιλαμβάνει και κεφαλαίους και πεζούς χαρακτήρες.

### Κεφάλαιο 3

```
//Romanization Ancient

//Single-letters
//Uppercase
'Α': 'A', 'Β': 'B', 'Γ': 'G', 'Δ': 'D', 'Ε': 'E', 'Ζ': 'Z', 'Η': 'Ē', 'Θ': 'Th',
'Ι': 'I', 'Κ': 'K', 'Λ': 'L', 'Μ': 'M', 'Ν': 'N', 'Ξ': 'X', 'Ο': 'O', 'Π': 'P',
'Ρ': 'R', 'Σ': 'S', 'Τ': 'T', 'Υ': 'Y', 'Φ': 'Ph', 'Χ': 'Ch', 'Ψ': 'Ps', 'Ω': 'Ō', 'Ϝ': 'W', 'Ϙ': 'K',
//Lowercase
'α': 'a', 'β': 'b', 'γ': 'g', 'δ': 'd', 'ε': 'e', 'ζ': 'z', 'η': 'ē', 'θ': 'th',
'ι': 'i', 'κ': 'k', 'λ': 'l', 'μ': 'm', 'ν': 'n', 'ξ': 'x', 'ο': 'o', 'π': 'p',
'ρ': 'r', 'σ': 's', 'ς': 's', 'τ': 't', 'υ': 'y', 'φ': 'ph', 'χ': 'ch', 'ψ': 'ps', 'ω': 'ō', 'ϝ': 'w', 'ϙ': 'k',
```

Εικόνα 3.10: Αντιστοιχίσεις Αρχαίων Ελληνικών ALA-LC Romanization tables

```
//Tonismoι
//Oxia ( ´ )
//Uppercase
'A': 'A', 'E': 'E', 'H': 'Ē', 'I': 'I', 'O': 'O', 'Y': 'Y', 'Ō': 'Ō',
'A': 'A', 'E': 'E', 'H': 'Ē', 'I': 'I', 'O': 'O', 'Y': 'Y', 'Ō': 'Ō', //+Apostrofos
//Lowercase
'á': 'a', 'é': 'e', 'ĥ': 'ē', 'í': 'i', 'ó': 'o', 'ú': 'y', 'ώ': 'ō',
'έ': 'e', 'ι': 'i', 'ο': 'o', 'ύ': 'y', // diff ascii code
'ǎ': 'a', 'ě': 'e', 'ĥ': 'ē', 'ĩ': 'i', 'ö': 'o', 'Û': 'y', 'ώ': 'ō', 'ǒ': 'ō', //+Apostrofos
//Baria ( ` )
//Uppercase
'A': 'A', 'E': 'E', 'H': 'Ē', 'I': 'I', 'O': 'O', 'Y': 'Y', 'Ō': 'Ō',
//Lowercase
'â': 'a', 'è': 'e', 'ĥ': 'ē', 'ì': 'i', 'ò': 'o', 'ù': 'y', 'ώ': 'ō',
'â': 'a', 'è': 'e', 'ĥ': 'ē', 'ò': 'o', //Diff ascii code
'ǎ': 'a', 'ĥ': 'ē', 'ĩ': 'i', //+Apostrofos
```

Εικόνα 3.11: Τονισμοί, οξείες και βαρείες

```
//Perispomeni ( ~ )
//Uppercase
'Ā': 'A', 'Ĥ': 'Ē', 'Ī': 'I', 'Ō̄': 'Ō', //+Apostrofos
'Ā̇': 'A', 'Ĥ̇': 'Ē', 'Ō̇': 'Ō', //+Apostrofos +iota
'Ā̃': 'A', 'Ĥ̃': 'Ē', 'Ī̃': 'I', 'Ū̃': 'Y', 'Ō̃': 'Ō', //+Daseia
'Ā̄': 'A', 'Ĥ̄': 'Ē', 'Ō̄': 'Ō', //+Daseia +
//Lowercase
'ā': 'a', 'ĥ̃': 'ē', 'ĩ̃': 'i', 'ũ̃': 'y', 'ō̃': 'ō',
'ā̇': 'a', 'ĥ̇': 'ē', 'ĩ̇': 'i', 'ũ̇': 'y', 'ō̇': 'ō',
'ā̃': 'a', 'ĥ̃': 'ē', 'ĩ̃': 'i', 'ũ̃': 'y', 'ō̃': 'ō', //+Apostrofos
'ā̄': 'a', 'ĥ̄': 'ē', 'ō̄': 'ō', //+Apostrofos+
'ā̄': 'a', 'ĥ̄': 'ē', 'ĩ̄': 'i', 'ũ̄': 'y', 'ō̄': 'ō', //+Daseia
'ā̄̇': 'a', 'ĥ̄̇': 'ē', 'ō̄̇': 'ō', //+Daseia +iota
```

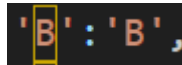
Εικόνα 3.12: Τονισμοί - Περισπωμένη



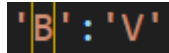
```
//Single letter combinations
result += transliterationMap[text[i]] || '*';
i++;
isWordStart=false;
```

Εικόνα 3.16: Έλεγχος μονού χαρακτήρα

Το πρότυπο ορίζει πως ο χαρακτήρας “B,β” μεταγραμματίζεται διαφορετικά για τα αρχαία και τα νέα ελληνικά. Στα νέα ελληνικά μεταγραμματίζεται ως “V,v”, ενώ στα αρχαία ελληνικά ως “B,b”. Έτσι, στους πίνακες αντιστοίχισης χαρακτήρων της αρχαίας και της νέας ελληνικής χρησιμοποιήθηκαν διαφορετικές αντιστοιχίσεις για το γράμμα “B” (βλ εικόνες ...).



Εικόνα 3.17: Μεταγραμματισμός B χαρακτήρα στα αρχαία ελληνικά



Εικόνα 3.18: Μεταγραμματισμός B στα Νέα Ελληνικά

Η μεγαλύτερη πρόκληση που θέτει το ALA-LC Romanization Table είναι ο σεβασμός στην προέλευση των λέξεων. Δηλαδή, αν μία λέξη της νέας ελληνικής προέρχεται από μία δασυνόμενη λέξη της αρχαίας ελληνικής, τότε η νεοελληνική λέξη μεταγραμματίζεται σαν να υπάρχει η δασεία. Προκειμένου να αναγνωρίζονται οι νεοελληνικές λέξεις που προέρχονται από δασυνόμενες λέξεις, έπρεπε να γίνει μία σχετική έρευνα. Ανιχνεύθηκαν σε λεξικά και ιστότοπους της αρχαίας ελληνικής 293 δασυνόμενες λέξεις και ονόματα (βλ. Παράρτημα Α). Έπειτα, έγινε μία έρευνα στο λεξικό Μπαμπινιώτη, για να ανιχνευθούν νεοελληνικές λέξεις και ονόματα που προέρχονται από τις δασυνόμενες λέξεις. Με βάση αυτή την έρευνα, καταγράφηκαν 402 θέματα λέξεων (ρίζες) εις διπλούν, και με αρχικό γράμμα κεφαλαίο και με αρχικό γράμμα πεζό (σύνολο 804 θέματα-ρίζες). Επειδή ο μεταγραμματισμός στην ΕΒΕ χρησιμοποιείται κυρίως για τα κύρια ονόματα, διεξήχθη επιπλέον έρευνα σε Κατάλογο Ελληνικών Επιθέτων [22], με στόχο να επιβεβαιωθεί η ορθότητα των καταγεγραμμένων θεμάτων-ριζών και να μη γίνει λανθασμένος μεταγραμματισμός σε επίθετα που χρησιμοποιούν τον ίδιο συνδυασμό γραμμάτων, αλλά δεν προέρχονται από αρχαία ελληνική δασυνόμενη λέξη, π.χ. είναι αραβικής, σλαβικής ή τουρκικής προέλευσης. Πρέπει να διευκρινιστεί ότι τα θέματα-ρίζες λέξεων που ανιχνεύθηκαν αποτελούν συνδυασμούς 3-5 γραμμάτων και λαμβάνουν υπόψη και τον τονισμό. Για παράδειγμα, έχουν προστεθεί τα θέματα-ρίζες “αγια” και “Αγια”, αλλά και “αγία” και “Αγία”.

Η παρακάτω εικόνα 3.19, παρουσιάζει τον πίνακα αντιστοίχισης θεμάτων-ριζών με 3 χαρακτήρες.

```
//Special Occasions

//3 letters
//Uppercase
'Αι': 'Ha', 'Αψύ': 'Ha',
'Επτ': 'He', 'Εξτ': 'He', 'Εξη': 'He', 'Ερπ': 'He', 'Εωλ': 'He',
'Ηρα': 'Hē', 'Ηρω': 'Hē', 'Ηρω': 'Hē', 'Ηρω': 'Hē', 'Ηβη': 'Hē', 'Ηδου': 'Hē', 'Ηδού': 'Hē', 'Ηλο': 'Hē', 'Ηλω': 'Hē', 'Ημι': 'Hē', 'Ημι': '',
'Ηντ': 'Hē', 'Ηττ': 'Hē', 'Ηττ': 'Hē',
'Ιππ': 'Hī',
'Οκ': 'Ho', 'Ορκ': 'Ho', 'Ορκ': 'Ho', 'Ολκ': 'Ho', 'Ολκ': 'Ho', 'Ολο': 'Ho', 'Ολό': 'Ho', 'Ολο': 'Ho', 'Ομο': 'Ho', 'Ομό': 'Ho', 'Οπλ': 'Ho',
'Οπλ': 'Ho', 'Οση': 'Ho', 'Οσα': 'Ho', 'Οσο': 'Ho', 'Οτσι': 'Ho',
'Ορο': 'Hō', 'Ορα': 'Hō', 'Ορα': 'Hō',
//Lowercase
'αι': 'ha', 'αψύ': 'ha',
'επτ': 'he', 'εξτ': 'he', 'εξη': 'he', 'ερπ': 'he', 'εωλ': 'he',
'ηρα': 'hē', 'ηρω': 'hē', 'ηρω': 'hē', 'ηρω': 'hē', 'ηβη': 'hē', 'ηδου': 'hē', 'ηδού': 'hē', 'ηλο': 'hē', 'ηλω': 'hē', 'ημι': 'hē', 'ημι': 'hē', 'ημι': 'hē',
'ηντ': 'hē', 'ηττ': 'hē', 'ηττ': 'hē',
'ιππ': 'hī',
'ορκ': 'ho', 'ορκ': 'ho', 'οκ': 'ho', 'ολκ': 'ho', 'ολκ': 'ho', 'ολο': 'ho', 'ολό': 'ho', 'ολο': 'ho', 'ομο': 'ho', 'ομό': 'ho', 'οπλ': 'ho', 'οπλ': 'ho',
'οση': 'ho', 'οσα': 'ho', 'οσο': 'ho', 'οτσι': 'ho',
'ορο': 'hō', 'ορα': 'hō', 'ορα': 'hō',
```

Εικόνα 3.19: Δασυνόμενες λέξεις 3 χαρακτήρων

Η παρακάτω εικόνα 3.20, παρουσιάζει τον πίνακα αντιστοίχισης θεμάτων-ριζών με 4 χαρακτήρες. Αξίζει να σημειωθεί, με αφορμή αυτήν την εικόνα, ότι σε κάποιες περιπτώσεις όπου η έρευνα αποκάλυψε ότι ένα θέμα-ρίζα δεν προέρχεται από δασυνόμενη λέξη, η αντιστοίχιση έγινε χωρίς τη χρήση του χαρακτήρα “H, h”. Για παράδειγμα, προστέθηκε κανόνας για τον μεταγραμματισμό της λέξης “όρκα” σε “orka”, ώστε να μη γίνεται λανθασμένος μεταγραμματισμός, καθώς στον κώδικα ορίστηκε, επίσης, ότι όσες λέξεις ξεκινούν από τον συνδυασμό “όρκ” μεταγραμματίζονται ως “horK”, λόγω προέλευσης από την αρχαιοελληνική δασυνόμενη λέξη “όρκος”.

Ένα επιπλέον παράδειγμα είναι οι περιπτώσεις 'Αγιου': 'A' και 'Αγιο': 'Ha', όπου εδώ επιδιώκεται οι λέξεις που ξεκινούν με “Αγιο” να παίρνουν τον χαρακτήρα “H” στην αρχή της λέξης και να ακολουθούν τον κανόνα, αλλά δεν θέλουμε τις λέξεις που ξεκινούν με “Αγίου” να ακολουθήσουν τον ίδιο κανόνα και τις εξαιρούμε. Γι’ αυτό και έχει προστεθεί η περίπτωση χωρίς το “H”. Και ο λόγος που προστέθηκαν και οι δύο περιπτώσεις, παρόλο που και οι δύο ξεκινάνε με τους χαρακτήρες “Αγιο”, είναι επειδή, έτσι όπως έχει γραφτεί ο κώδικας, πρώτα θα γίνει αναζήτηση στους συνδυασμούς με τους περισσότερους χαρακτήρες και, εφόσον, δεν βρει κάποια αντιστοιχία, θα προσανατολίσει την αναζήτησή του σε συνδυασμούς λιγότερων χαρακτήρων. Στο παράδειγμά μας, επομένως, ο κώδικας ξεκινάει να διαβάζει τους 5 χαρακτήρες, όπου και ανήκει ο συνδυασμός 'Αγιου': 'A', και έτσι παραλείπεται τελείως ο συνδυασμός 'Αγιο': 'Ha' των 4 χαρακτήρων.



```

//Handle 5-letter combinations
const fiveLetterCombo = text[i] + text[i+1] + text[i+2]+ text[i+3]+ text[i+4];
if (transliterationMapStarters.hasOwnProperty(fiveLetterCombo)){
    result += transliterationMapStarters[fiveLetterCombo];
    i += 1; // Move to the next pair of characters
    isWordStart = false;
    continue;
}
//Handle 4-letter combinations
const fourLetterCombo = text[i] + text[i+1] + text[i+2]+ text[i+3];
if (transliterationMapStarters.hasOwnProperty(fourLetterCombo)){
    result += transliterationMapStarters[fourLetterCombo];
    i += 1; // Move to the next pair of characters
    isWordStart = false;
    continue;
}
//Handle 3-letter combinations
const threeLetterCombo = text[i] + text[i+1] + text[i+2];
if (transliterationMapStarters.hasOwnProperty(threeLetterCombo)){
    result += transliterationMapStarters[threeLetterCombo];
    i += 1; // Move to the next pair of characters
    isWordStart = false;
    continue;
}

```

Εικόνα 3.22: Έλεγχος δασυνόμενων λέξεων για τα Νέα Ελληνικά

Οι δασυνόμενες λέξεις έχουν προστεθεί στο “transliterationMapStarters”, το οποίο είναι ένα ειδικό mapping με όλους τους συνδυασμούς έναρξης μίας λέξης, που θέλουμε να ελέγχουμε. Και επειδή ο κανόνας μας αναφέρεται σε τέτοιους συνδυασμούς, γι’ αυτό και προστέθηκαν σε αυτόν τον πίνακα. Επίσης, στον σχετικό αλγόριθμο παρατηρούμε ότι, παρόλο που διαβάζονται 3 με 5 χαρακτήρες και η αντικατάσταση γίνεται με 2 χαρακτήρες, ενώ ο δείκτης *i* αυξάνεται μόνο κατά 1 μονάδα. Αυτό συμβαίνει για την μεγαλύτερη ευκολία και του χρήστη, καθώς και για την αύξηση της ταχύτητας του κώδικα, καθώς έχουμε πολλά keys να δείχνουν την ίδια θέση στην μνήμη.

Μία επιπλέον εξαίρεση στον βασικό πίνακα μεταγραμματισμού είναι ότι μερικοί χαρακτήρες ή ομάδα χαρακτήρων, ανάλογα με το πού βρίσκονται μέσα στην λέξη -αρχή, μέση ή τέλος- αντικαθίστανται από διαφορετικούς χαρακτήρες. Για παράδειγμα, ο συνδυασμός των χαρακτήρων “Μπ” στην αρχή της λέξης θα πρέπει να αντικατασταθεί από τον χαρακτήρα “B”. Ενώ στην περίπτωση όπου ο συνδυασμός αυτός βρίσκεται στην μέση της λέξης ή στο τέλος, θα αντικατασταθεί με τους χαρακτήρες “mp”, που είναι και ο βασικός μεταγραμματισμός. Η εικόνα 3.23 παρουσιάζει τις διαφορετικές αντιστοιχίσεις των διφθόγγων “μπ, ντ, γκ”, ενώ η εικόνα 3.24 παρουσιάζει τον αλγόριθμο χειρισμού αυτής της περίπτωσης.

```

//Starters
//Single Letter
'Υ': 'Hy', 'υ': 'hy',

//2-letters Starters
//Uppercase
'Μπ': 'B', 'Ντ': 'D',

//Lowercase
'ντ': 'd', 'μπ': 'b',
//2 letter output
'Γκ': 'Gk', 'γκ': 'gk',

```

Εικόνα 3.23: Αντιστοιχίσεις διφθόγων, στην αρχή της λέξης

```

let result = '';
let i = 0;
let isWordStart = true;

while (i < text.length) {

  if (text[i] === ' '){
    result += text[i];
    isWordStart = true;
    i++;
    continue;
  }
  if(isWordStart){
    const twoLetterCombo = text[i] + text[i+1];
    //Starters (2 Letters)
    if(transliterationMapStarters.hasOwnProperty(twoLetterCombo)) {
      result += transliterationMapStarters[twoLetterCombo];
      i += 2; // Move to the next pair of characters
      isWordStart = false;
      continue;
    }else if(transSpecialStarters.hasOwnProperty(twoLetterCombo)){
      result += transSpecialStarters[twoLetterCombo];
      i += 1; // Move to the next pair of characters
      isWordStart = false;
      continue;
    }else{
      //Handle 5-letter combinations
      const fiveLetterCombo = text[i] + text[i+1] + text[i+2]+ text[i+3]+ text[i+4];
      if (transliterationMapStarters.hasOwnProperty(fiveLetterCombo)){
        result += transliterationMapStarters[fiveLetterCombo];
        i += 1; // Move to the next pair of characters
        isWordStart = false;
        continue;
      }
    }
  }
}

```

Εικόνα 3.24: Αλγόριθμος ελέγχου των starters

Στην ουσία έχει δημιουργηθεί μία νέα κατηγορία, ένα νέο mapping, που ονομάζεται `transSpecialStarters`, κυρίως για την ειδική περίπτωση του μεταγραμματισμού άρθρων στη νέα ελληνική. Ο λόγος που τα άρθρα έχουν αποτελέσει ειδική κατηγορία στη νέα ελληνική είναι για να αποφευχθεί η ταύτιση του άρθρου “ο, κενό” με την κατάληξη λέξης “ο, κενό”. Προστέθηκε, λοιπόν, μία Boolean μεταβλητή, ώστε να μπορεί να ελέγχεται πότε μία λέξη ξεκινάει και πότε όχι. Η μεταβλητή αυτή είναι γνωστή και ως “flag”. Όταν εκτελείται ο αλγόριθμος, στην αρχή της πρότασης η μεταβλητή έχει την τιμή “true” (`flag=true`), ομοίως όταν ξεκινάει μία λέξη. Ο αλγόριθμος, λοιπόν, ελέγχει κάθε φορά το flag, για να ξέρει με ποιο mapping θα ξεκινήσει να ελέγχει για τυχόν αντιστοιχίες.

Ξεκινώντας ο αλγόριθμος, η μεταβλητή (flag) έχει την τιμή “true”, γιατί αρχίζει η πρόταση.

Όταν εντοπίσει μία αντιστοιχία στον πίνακα, αυξάνει τον δείκτη “i” και θέτει νέα τιμή στο `flag=false`, όσο ο δείκτης βρίσκεται εντός λέξης. Έτσι, ο αλγόριθμος ανακατευθύνεται στο ανάλογο mapping, όσο ισχύει η μεταβλητή `flag=false`. Η μεταβλητή ανακτά την τιμή true (`flag=true`), όταν ο αλγόριθμος διαβάζει τον χαρακτήρα “κενό” από μόνο του “.”. Ο κενός χαρακτήρας δηλώνει την λήξη μιας λέξης και την προετοιμασία έναρξης της επόμενης. Εδώ να τονιστεί ότι είναι σημαντική η σωστή σύνταξη των λέξεων. Ένα γνωστό λάθος που παρατηρείται είναι, για παράδειγμα, η έλλειψη κενού χαρακτήρα μετά τον χαρακτήρα κόμμα “,”, η οποία θα μπορούσε να επηρεάσει την σωστή λειτουργία του κώδικα.

```
const transSpecialStarters = {
  'ο' : 'Ho', 'Η' : 'Hē', 'ο' : 'ho', 'η' : 'hē'
}
const transliterationMapStarters = {
  //Starters
  //Single Letter
  'Υ' : 'Hy', 'υ' : 'hy',
  //2-letters Starters
  //Uppercase
  'Μπ' : 'B', 'Ντ' : 'D',
  //Lowercase
  'ντ' : 'd', 'μπ' : 'b',
  //2 letter output
  'Γκ' : 'Gk', 'γκ' : 'gk',
}
```

Εικόνα 3.25: Αντιστοιχίες άρθρων

Στην αρχή, ο αλγόριθμος ξεκινάει με την συνθήκη (`text === ""`), αναζητώντας τον χαρακτήρα που, βάσει της υπόδειξης του δείκτη “i”, ισούται με το κενό. Σε περίπτωση που η συνθήκη αυτή αληθεύει, παραλείπει τις εντολές που ακολουθούν και προχωράει στην επόμενη θέση / χαρακτήρα. Στην συνέχεια, γίνεται ο έλεγχος της μεταβλητής/flag “isWordStart”, προκειμένου

να διερευνηθούν τα ανάλογα ειδικά Mappings. Στην περίπτωση που η μεταβλητή είναι “True”, τότε σημαίνει ότι έχουμε την έναρξη μίας λέξης και οι έλεγχοι που γίνονται είναι οι εξής: αρχικά, αναζητά μέσω του mapping “transliterationMapStarters” την αντιστοίχιση 2 χαρακτήρων στην αρχή λέξης. Εάν αυτή δεν βρεθεί, ο αλγόριθμος συνεχίζει με την αναζήτηση άρθρων από το Mapping “transSpecialStarters”. Εάν δεν προκύψει και πάλι αντιστοίχιση, θα ελέγξει εάν υπάρχει συνδυασμός 2 χαρακτήρων που αντιστοιχίζεται με κάποιες δασυνόμενες λέξεις. Εφόσον δεν εντοπίσει και πάλι κάποια αντιστοιχία, συνεχίζει, πάλι μέσω του “transliterationMapStarters”, με τον έλεγχο συνδυασμών 5 χαρακτήρων, μετά 4 χαρακτήρων, ακολούθως 3 χαρακτήρων, για να καλυφθούν όλες οι περιπτώσεις των δασυνόμενων λέξεων στη νέα ελληνική. Τέλος, ελέγχει συνδυασμούς, αρχικά, 3 χαρακτήρων και, στη συνέχεια, 2 χαρακτήρων, μέσω του βασικού, πλέον, πίνακα “transliterationMap” (default mapping). Εφόσον δεν ανιχνευθεί πάλι καμία αντιστοίχιση, προχωρά στον έλεγχο αντιστοίχισης 1 χαρακτήρα, ο οποίος, εάν δεν εντοπιστεί, αντικαθίσταται με έναν αστερίσκο, ο οποίος δηλώνει την έλλειψη αντιστοίχισης στα mappings. Στην περίπτωση που ο δείκτης flag έχει την τιμή false, οι μοναδικοί έλεγχοι που θα πραγματοποιηθούν είναι από το “transliterationMap” για τους συνδυασμούς χαρακτήρων με πλήθος από 3 έως 1.

```

while (i < text.length) {
  if (text[i]=== ' '){
    result += text[i];
    isWordStart = true;
    i++;
    continue;
  }
  if(isWordStart){
    const twoLetterCombo = text[i] + text[i+1];
    //Starters (2 Letters)
    if(transliterationMapStarters.hasOwnProperty(twoLetterCombo)) {
      result += transliterationMapStarters[twoLetterCombo];
      i += 2; // Move to the next pair of characters
      isWordStart = false;
      continue;
    }else if(transSpecialStarters.hasOwnProperty(twoLetterCombo)){
      result += transSpecialStarters[twoLetterCombo];
      i += 1; // Move to the next pair of characters
      isWordStart = false;
      continue;
    }else{
      //Handle 5-letter combinations
      const fiveLetterCombo = text[i] + text[i+1] + text[i+2]+ text[i+3]+ text[i+4];
      if (transliterationMapStarters.hasOwnProperty(fiveLetterCombo)){
        result += transliterationMapStarters[fiveLetterCombo];
        i += 1; // Move to the next pair of characters
        isWordStart = false;
        continue;
      }
      //Handle 4-letter combinations
      const fourLetterCombo = text[i] + text[i+1] + text[i+2]+ text[i+3];
      if (transliterationMapStarters.hasOwnProperty(fourLetterCombo)){
        result += transliterationMapStarters[fourLetterCombo];
        i += 1; // Move to the next pair of characters
        isWordStart = false;
      }
    }
  }
}

```

Εικόνα 3.26: Βοηθητικός αλγόριθμος αντιστοιχίσεων

```

//Handle 3-letter combinations
const threeLetterCombo = text[i] + text[i+1] + text[i+2];
if (transliterationMapStarters.hasOwnProperty(threeLetterCombo)){
  result += transliterationMapStarters[threeLetterCombo];
  i += 1; // Move to the next pair of characters
  isWordStart = false;
  continue;
}
//Handle 3-letter combinations
const threeLetterCombos = text[i] + text[i+1] + text[i+2];
if (transliterationMap.hasOwnProperty(threeLetterCombos)){
  result += transliterationMap[threeLetterCombos];
  i += 1; // Move to the next pair of characters
  isWordStart = false;
  continue;
}
//Handle 2-letter combinations
const twoLetterCombo = text[i] + text[i+1];
if(transliterationMap.hasOwnProperty(twoLetterCombo)) {
  result += transliterationMap[twoLetterCombo];
  i += 2; // Move to the next pair of characters
  isWordStart = false;
  continue;
}
const oneLetterCombo = text[i];
if(transliterationMapStarters.hasOwnProperty(oneLetterCombo)){
  result += transliterationMapStarters[oneLetterCombo];
  i += 1; // Move to the next pair of characters
  isWordStart = false;
  continue;
}
//Single letter combinations
result += transliterationMap[text[i]] || '*';
i++;
isWordStart=false;}
}else{

```

Εικόνα 3.27: Συνέχεια βοηθητικού αλγόριθμου αντιστοιχίσεων

```
    }else{
      //Handle 3-letter combinations
      const threeLetterCombo = text[i] + text[i+1] + text[i+2];
      if (transliterationMap.hasOwnProperty(threeLetterCombo)){
        result += transliterationMap[threeLetterCombo];
        i += 2; // Move to the next pair of characters
        isWordStart = false;
        continue;
      }
      //Handle 2-letter combinations
      const twoLetterCombo = text[i] + text[i+1];
      if(transliterationMap.hasOwnProperty(twoLetterCombo)) {
        result += transliterationMap[twoLetterCombo];
        i += 2; // Move to the next pair of characters
        isWordStart = false;
        continue;
      }
      //Single letter combinations
      result += transliterationMap[text[i]] || '*';
      i++;
      isWordStart=false;
    }
  }
  return result;
}
```

Εικόνα 3.28: Ολοκλήρωση βοηθητικού αλγόριθμου αντιστοιχίσεων

### 3.5 Σχεδιασμός αξιολόγησης

Προκειμένου να αξιολογηθεί το εργαλείο, δημιουργήθηκε από το προσωπικό της ΕΒΕ ένα μικρό σύνολο δεδομένων αναφοράς (gold dataset). Παρακάτω παρουσιάζεται ο τρόπος που δημιουργήθηκαν τα δεδομένα αναφοράς.

#### 3.5.1 Δημιουργία συνόλου δεδομένων αναφοράς (gold dataset)

Πριν αρχίσουν να συλλέγονται δεδομένα, αποφασίστηκε ότι το σύνολο που θα δημιουργηθεί πρέπει να περιλαμβάνει περιπτώσεις που καλύπτουν όλους τους κανόνες και τις εξαιρέσεις των προτύπων. Η ΕΒΕ χρησιμοποιεί τον μεταγραμματισμό κυρίως για τη μετατροπή ονομάτων. Επιλέχθηκαν, λοιπόν, ονόματα της αρχαίας και της νέας ελληνικής, με χρήση πολυτονικού και μονοτονικού συστήματος, όπως επίσης και φράσεις με λέξεις οι οποίες περιλαμβάνουν διφθόγγους σε διάφορα σημεία (αρχή, μέση, τέλος λέξης) ή προέρχονται από δασυνόμενες λέξεις της αρχαίας ελληνικής.

Τα δεδομένα αναφοράς έχουν αναρτηθεί σε excel αρχείο με 385 γραμμές. Από τα 385 ονόματα και φράσεις, 18 είναι στα αρχαία ελληνικά και τα υπόλοιπα 367 στα νέα ελληνικά. Βιβλιοθηκονόμοι του Τμήματος Καταλόγων της ΕΒΕ μεταγραμμάτισαν το όνομα ή τη φράση της κάθε γραμμής σύμφωνα με τους κανόνες των δύο προτύπων. Το excel αρχείο έχει αναρτηθεί στο αποθετήριο της Ευρωπαϊκής Επιτροπής Zenodo και είναι διαθέσιμο στη διεύθυνση:

<https://zenodo.org/doi/10.5281/zenodo.13318808>

#### 3.5.2 Προετοιμασία prompts για ChatGPT

Η αξιολόγηση του εργαλείου που δημιουργήθηκε στο πλαίσιο της παρούσας εργασίας περιλαμβάνει τη σύγκριση των δεδομένων αναφοράς όχι μόνο με τα αποτελέσματα του εργαλείου, αλλά και με τα αποτελέσματα που παράγει το ChatGPT.

Για την παραγωγή αποτελεσμάτων από το εργαλείο τεχνητής νοημοσύνης, συντάχθηκαν 3 κατευθυντήριες ερωτήσεις για την εξαγωγή αποτελεσμάτων σύμφωνα με:

1. το πρότυπο ISO 843
2. το πρότυπο ALA-LC Romanization table για τα αρχαία ελληνικά
3. το πρότυπο ALA-LC Romanization table για τα νέα ελληνικά.

Σε κάθε κατευθυντήρια ερώτηση ενσωματώθηκαν οι οδηγίες και οι εξαιρέσεις του προτύπου και, μετά, δινόταν η λίστα των 385 ονομάτων και φράσεων που περιλαμβάνονται στα δεδομένα αναφοράς. Ειδικά στο 3ο ερώτημα, προστέθηκε στην εκπαίδευση και ο κατάλογος δασυνόμενων λέξεων, που παρατίθεται στο Παράρτημα Α.

#### Κατευθυντήριο ερώτημα (prompt) για το πρότυπο ISO 843

1. *Να μεταγράψεις τα ακόλουθα ελληνικά ονόματα και φράσεις σύμφωνα με το πρότυπο ISO 843 που μπορείς να εντοπίσεις στην παρακάτω διεύθυνση: [https://en.wikipedia.org/wiki/ISO\\_843](https://en.wikipedia.org/wiki/ISO_843). Δεν υπάρχουν διαφορετικοί κανόνες για τα αρχαία και τα νέα ελληνικά.*

2. Ο μεταγραμματισμός πρέπει να ακολουθεί το μονοτονικό σύστημα. Αυτό σημαίνει ότι πρέπει να μεταγράφονται και οι τόνοι. Λοιποί ειδικοί χαρακτήρες και διακριτικά, π.χ. ψιλή, δασεία, πρέπει να αγνοούνται.
3. Τήρησε τους ειδικούς κανόνες για τη μεταγραφή των διφθόγγων αυ, ευ και ου.
4. Χρησιμοποίησε μόνο τους ειδικούς χαρακτήρες που προβλέπονται από το πρότυπο.

**Κατευθυντήριο ερώτημα (prompt) για το πρότυπο ALA-LC Romanization table για τα αρχαία ελληνικά**

1. Να μεταγράψεις τα παρακάτω αρχαία ελληνικά ονόματα ακολουθώντας τον Πίνακα Ρομανοποίησης των ALA-LC για τα ελληνικά, που μπορείς να βρεις στην παρακάτω διεύθυνση: <https://www.loc.gov/catdir/cpsd/romanization/greek.pdf>.
2. Να εξετάσεις τους διαφορετικούς κανόνες μεταξύ της μεταγραφής των αρχαίων και των νέων ελληνικών και να εφαρμόσεις μόνον αυτούς που αφορούν τα αρχαία ελληνικά.
3. Να τηρείς τους ειδικούς κανόνες για τη μεταγραφή των διφθόγγων και τη θέση τους στις λέξεις.
4. Να σεβαστείς την ύπαρξη δασείας στην αρχή μιας λέξης. Να την μεταγράψεις ως «H» ή «h», ανάλογα με το αν το γράμμα που έχει δασεία είναι κεφαλαίο γράμμα ή όχι.
5. Ο πίνακας ρομανοποίησης δεν παρέχει κανόνες για τα σημεία στίξης. Μην τα μεταγράψεις, απλώς να τα αντιγράψεις ως έχουν. Τα μόνα σημεία στίξης που πρέπει να μετατραπούν είναι τα εξής: ελληνικό ερωτηματικό “;” (μεταγράφεται ως “?”), ελληνική άνω τελεία “.” (μεταγράφεται ως “;”), αριστερό ελληνικό εισαγωγικό “«” (μεταγράφεται ως “«”), δεξιό ελληνικό εισαγωγικό “»” (μεταγράφεται ως “»”).

**Κατευθυντήριο ερώτημα (prompt) για το πρότυπο ALA-LC Romanization table για τα νέα ελληνικά**

1. Να μεταγράψεις τα παρακάτω νεοελληνικά ονόματα ακολουθώντας τον Πίνακα Ρομανοποίησης των ALA-LC για τα ελληνικά που μπορείς να βρεις στην παρακάτω διεύθυνση: <https://www.loc.gov/catdir/cpsd/romanization/greek.pdf>.
2. Να εξετάσεις τους διαφορετικούς κανόνες μεταξύ της μεταγραφής των αρχαίων και των νέων ελληνικών και να εφαρμόσεις μόνον αυτούς που αφορούν τη νέα ελληνική γλώσσα.
3. Να τηρείς τους ειδικούς κανόνες για τη μεταγραφή των διφθόγγων και τη θέση τους στις λέξεις.
4. Ο πίνακας ρομανοποίησης δεν παρέχει κανόνες για τα σημεία στίξης. Μην τα μεταγράψεις, απλώς να τα αντιγράψεις ως έχουν. Τα μόνα σημεία στίξης που πρέπει να μετατραπούν είναι τα εξής: ελληνικό ερωτηματικό “;” (μεταγράφεται ως “?”), ελληνική άνω τελεία “-” (μεταγράφεται ως “;”), αριστερό ελληνικό εισαγωγικό “«” (μεταγράφεται ως “«”), δεξιό ελληνικό εισαγωγικό “»” (μεταγράφεται ως “»”).
5. Οι κανόνες ρομανοποίησης σέβονται την προέλευση των νεοελληνικών λέξεων: αν μια νεοελληνική λέξη προέρχεται από μια αρχαία ελληνική λέξη που παίρνει δασεία, τότε η νεοελληνική λέξη πρέπει να μεταγράφεται σαν να υπάρχει ακόμα η δασεία. Για παράδειγμα,

το νεοελληνικό όνομα *Ελένη* μεταγράφεται ως *Helepē*, επειδή προέρχεται από το αρχαιοελληνικό όνομα *Έλενη* που είχε δασεία στο πρώτο γράμμα *Ε*. Η δασεία μεταγράφεται ως «*Η*» ή «*h*», ανάλογα με το αν το αρχικό γράμμα της λέξης είναι κεφαλαίο ή όχι.

6. Να εξετάσεις τον ακόλουθο κατάλογο αρχαίων ελληνικών λέξεων που παίρνουν δασεία [εδώ ακολούθησε ο κατάλογος λέξεων του παραρτήματος Α].

Τα αποτελέσματα της αξιολόγησης παρουσιάζονται στο 4ο κεφάλαιο “Αποτελέσματα”.

### 3.6 Γραφικό περιβάλλον

#### 3.6.1 Αρχές δημιουργίας GUI που εφαρμόστηκαν

Για τη δημιουργία του γραφικού περιβάλλοντος, λήφθηκε υπόψη κυρίως το σενάριο χρήσης της Εθνικής Βιβλιοθήκης της Ελλάδος. Ο στόχος ήταν να σχεδιαστεί ένα απλό γραφικό περιβάλλον, που να μη χρειάζεται ειδική επεξήγηση για τη χρήση του και, επιπλέον, να είναι αρκετά εύχρηστο, ώστε να αποφεύγονται πολλές κινήσεις αντιγραφής και επικόλλησης. Για την εύκολη ενσωμάτωση του εργαλείου σε διαφορετικές εφαρμογές, τέθηκε ένας επιπλέον στόχος: αυτός της εφαρμογής ανοιχτών τεχνολογιών. Συγκεκριμένα, αξιοποιήθηκαν οι τεχνολογίες HTML, CSS και JavaScript.

Οι καταλογογράφοι, κατά κανόνα, πληκτρολογούν σύντομα κείμενα (κυρίως ονόματα φυσικών προσώπων) και τα επικολλούν μέσα στο σύστημα καταλογογράφησης σε συγκεκριμένα πεδία. Για την αποφυγή πολλαπλών κινήσεων αντιγραφής και επικόλλησης, ο μεταγραμματισμός γίνεται ταυτόχρονα και για τα δύο πρότυπα και η αντιγραφή γίνεται με τη χρήση ενός κουμπιού στα δεξιά.

Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς

Κείμενο

Εισάγετε το κείμενο για μεταγραμματισμό

Αρχαία ελληνικά

ALA-LC romanization

ISO843:1997

Καθαρισμός Υποβολή

Disclaimer: Για τον μεταγραμματισμό ονομάτων και φράσεων στα νέα ελληνικά σύμφωνα με τους κανόνες του ALA-LC Romanization table, επιβεβαιώστε την προέλευση των λέξεων (αν δηλαδή προέρχονται από δασυνόμενες λέξεις της αρχαίας ελληνικής) μέσα από το online λεξικό αρχαίας ελληνικής των Liddell-Scott στον δικτυακό τόπο του Κέντρου Ελληνικής Γλώσσας.

Created by: Ilias Tsitsilegkas

Εικόνα 3.29: Γραφικό περιβάλλον στα ελληνικά

Αναλυτικά, στο πρώτο πεδίο ο χρήστης μπορεί να εισάγει το κείμενο που θέλει να μεταγραμματίσει. Και στην συνέχεια, εφόσον πατήσει το κουμπί «Υποβολή», η ιστοσελίδα

επιστρέφει τα αποτελέσματα και από τα δύο πρότυπα. Στο πάνω πεδίο εμφανίζει το αποτέλεσμα με βάση το πρότυπο «ALA-LC Romanization» και στο κάτω πεδίο, με βάση το πρότυπο «ISO 843». Ο χρήστης μπορεί έπειτα να πάρει το αποτέλεσμα που τον ενδιαφέρει, είτε χρησιμοποιώντας το «Ctrl + C», αφού επιλέξει το κείμενο που τον ενδιαφέρει, είτε πατώντας το εικονίδιο της αντιγραφής που βρίσκεται δίπλα από τα δύο αποτελέσματα.

Η προεπιλεγμένη ρύθμιση της ιστοσελίδας είναι στα Νέα Ελληνικά και στην ελληνική γλώσσα. Εάν ο χρήστης ενδιαφέρεται για τον μεταγραμματισμό ενός κειμένου που είναι στα Αρχαία Ελληνικά, μπορεί να τσεκάρει το μικρό κουτάκι που έχει την λέξη «Αρχαία ελληνικά» από δίπλα.

Επιπλέον, υπάρχει η δυνατότητα ο χρήστης να αλλάζει την γλώσσα της σελίδας στα Αγγλικά πατώντας πάνω στην Αγγλική σημαία.

Εικόνα 3.30: Γραφικό περιβάλλον στα αγγλικά

Επιπρόσθετα, υπάρχει ένα κουμπί «Καθαρισμός», το οποίο, όταν πατηθεί, διαγράφει το περιεχόμενο όλων των πεδίων. Τέλος, διατίθεται ένα disclaimer στο κάτω μέρος της ιστοσελίδας, η οποία παρέχει και έναν υπερσύνδεσμο προς το έγκριτο λεξικό Lidell-Scott της αρχαίας ελληνικής, ώστε ο χρήστης να έχει την δυνατότητα να επιβεβαιώσει την προέλευση της λέξης.

### 3.6.2 Δοκιμή από το προσωπικό της ΕΒΕ για σχόλια σχετικά με την ευχρηστία

Προκειμένου να ληφθεί ανατροφοδότηση από το προσωπικό της ΕΒΕ σχετικά με το εργαλείο Transliterator, το τμήμα Καταλόγων ανέθεσε σε δύο υπαλλήλους να εργαστούν πιλοτικά με το εργαλείο στο πλαίσιο της καθημερινής εργασίας. Η πιλοτική δοκιμή έγινε το πρώτο δεκαπενθήμερο του Αυγούστου 2024 και στέφθηκε με επιτυχία, καθώς οι καταλογογράφοι της ΕΒΕ έκριναν πως το εργαλείο, εκτός από ακριβές στα αποτελέσματα που δίνει, είναι και εύχρηστο. Βρήκαν πολύ βοηθητική την ύπαρξη του κουμπιού “αντιγραφή”, προκειμένου εύκολα να αντιγράφουν και να επικολλούν τον μεταγραμματισμό στις εγγραφές που επεξεργάζονται. Μία

πρόταση που έκαναν και έγινε αποδεκτή ήταν η προσθήκη ενός ενημερωτικού κειμένου, τύπου “disclaimer”, στο κάτω μέρος της οθόνης, από το οποίο εύκολα μπορούν να παραπέμπονται στο online λεξικό Lidell-Scott και να ελέγχουν κάποια αρχαιοελληνική λέξη σε αυτό.

### 3.7 Επίλογος

Για τον σχεδιασμό του εργαλείου μεταγραμματισμού προς εξυπηρέτηση των αναγκών της EBE, επιλέχθηκαν οι τεχνολογίες HTML, CSS και JavaScript, ως οι πλέον κατάλληλες, καθώς συνθέτουν ένα περιβάλλον εργασίας εύκολα προσβάσιμο από οποιονδήποτε χρήστη, διαδραστικό, φιλικό, ευανάγνωστο ως προς το περιεχόμενο, σαφές ως προς τη δομή, προσαρμόσιμο σε κάθε συσκευή και αποδοτικό ως προς τα επιθυμητά αποτελέσματα. Αξιοποιήθηκε το διεθνές πρότυπο ISO 843, στην αναθεωρημένη έκδοση του 1997 και στην ελληνική του μετάφραση (ΕΛΟΤ 743), μόνον ως προς τον πίνακα μεταγραμματισμού. Αξιοποιήθηκε, επιπλέον, ο πίνακας μετατροπής μη αλφαβητικών χαρακτήρων, με το πλεονέκτημα της επιλογής μονοτονικού και πολυτονικού συστήματος. Στους βασικούς κανόνες του εν λόγω προτύπου έγιναν παρεμβάσεις βελτίωσης σε ό,τι αφορά τις εξαιρέσεις που δεν καλύπτονται από το ΕΛΟΤ 743 και Romanization table του προτύπου ALA-LC (αναθεωρημένη έκδοση του 2010). Σε ό,τι αφορά το δεύτερο αυτό πρότυπο, η σύνταξη πινάκων στο προτεινόμενο εργαλείο μεταγραμματισμού άντλησε το πλεονέκτημα να ερευνάται η προέλευση των νεοελληνικών λέξεων από αντίστοιχες αρχαίες, ώστε να αποδίδεται με ακρίβεια η δασεία. Αυτό ενσωματώθηκε στον γραφικό περιβάλλον του εργαλείου που προτείνεται εδώ, ώστε να αποδίδεται με ακρίβεια ο τονισμός των λέξεων, συγκεκριμένα η δασεία. Επιπλέον, έγιναν παρεμβάσεις σε ό,τι αφορά τους συνδυασμούς χαρακτήρων, ανάλογα και με τη θέση τους στην λέξη. Μετά την ολοκλήρωση σχεδιασμού του εργαλείου, εφαρμόστηκε αυτό πιλοτικά από το προσωπικό της EBE σε μία βάση δεδομένων αναφοράς που το ίδιο το προσωπικό συνέταξε. Στις πρώτες εντυπώσεις των χρηστών επισημάνθηκαν η ευκολία στη χρήση και η αποτελεσματικότητα του εργαλείου. Επίσης, υιοθετήθηκε πρότασή τους για περαιτέρω διευκόλυνσή τους, σε ό,τι αφορά την αναζήτηση της προέλευσης νεοελληνικής λέξης από την αρχαία γλώσσα.

## Κεφάλαιο 4ο: Αποτελέσματα

### 4.1 Εισαγωγή

Σε αυτήν την ενότητα παρουσιάζεται η εφαρμογή του εργαλείου μεταγραμματισμού με εικόνες και παραδείγματα. Καταγράφονται τα αποτελέσματα της αξιολόγησης, όπως αυτά προέκυψαν κατά την πιλοτική εφαρμογή του από το προσωπικό της ΕΒΕ σε στατιστικούς πίνακες, σε παράλληλη σύγκριση με τα εξαγόμενα από την εκτέλεση της ίδιας διαδικασίας από το ChatGPT. Τέλος, προτείνεται ένα βοηθητικό πρόγραμμα ελέγχου για τη μεγιστοποίηση των ποσοστών επιτυχίας του υπό αξιολόγηση εργαλείου.



### 4.2 Σύντομη περιγραφή της εφαρμογής με εικόνες

Η εφαρμογή που αναπτύχθηκε μπορεί εύκολα να χρησιμοποιηθεί από Έλληνες και ξένους χρήστες και, επιπλέον, να ενσωματωθεί σε οποιοδήποτε δικτυακό τόπο. Τα παρακάτω δείγματα οθόνης παρουσιάζουν τη χρήση του εργαλείου με πραγματικά δεδομένα στις εικόνες 4.1 και 4.2

The screenshot shows a web application titled "Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς" (Transcription of Greek characters to Latin). It features two flags, the Greek and the British, indicating language support. The interface includes a text input field for the Greek text "Αλεξάνδρου, Νικολέττα". Below this, there is a checkbox for "Αρχαία Ελληνικά" (Ancient Greek) which is unchecked. There are two output fields: one for "ALA-LC Romanization" showing "Alexandrou, Nikoletta" and another for "ISO843:1997" showing "Alexándrou, Nikolétta". Each output field has a copy icon to its right. At the bottom, there are two blue buttons: "Καθαρισμός" (Clean) and "Υποβολή" (Submit).

Εικόνα 4.1: Παράδειγμα 1


**Transliteration of Greek characters  
into Latin**


Text

Ancient Greek

ALA-LC Romanization





ISO843:1997



Εικόνα 4.2: Παράδειγμα 2

Ακολουθούν δείγματα οθόνης με μερικές δασυνόμενες λέξεις στα νέα ελληνικά, για να παρατηρηθούν οι εξαιρέσεις που αναφέρθηκαν νωρίτερα. Στην εικόνα 4.3 αναγνωρίζεται ότι το επίθετο “Ηρακλείδης” προέρχεται από το αρχαίο όνομα “Ηρακλῆς”, που έπαιρνε δασεία στα αρχαία ελληνικά.


### Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς


Κείμενο

Αρχαία Ελληνικά

ALA-LC Romanization

ISO843:1997



 

Καθαρισμός
Υποβολή

Εικόνα 4.3: Παράδειγμα 1, δασυνόμενης λέξης στα Νέα Ελληνικά

Χρησιμοποιώντας ως παράδειγμα το ίδιο επίθετο, αυτή τη φορά καταγράφηκε με πνεύματα της αρχαίας ελληνικής. Και πάλι, όπως φαίνεται στην εικόνα 4.4, το εργαλείο μεταγραμματισμού αναγνωρίζει ορθά τους αρχαίους ελληνικούς χαρακτήρες (εν προκειμένω το δασυνόμενο Ή) και μεταγραμματίζει με τη χρήση του λατινικού χαρακτήρα “H”.


### Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς


Κείμενο

Αρχαία Ελληνικά

ALA-LC Romanization

ISO843:1997



 

Καθαρισμός
Υποβολή

Εικόνα 4.4: Παράδειγμα 1, δασυνόμενης λέξης στα Αρχαία Ελληνικά

Οι παρακάτω εικόνες 4.5 και 4.6, παρουσιάζουν ένα ακόμα παράδειγμα με το μικρό όνομα Ερμιόνη, το οποίο προέρχεται από το αρχαίο όνομα “Ερμιόνη” που έπαιρνε δασεία στα αρχαία ελληνικά.


### Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς


Κείμενο

Αρχαία Ελληνικά

ALA-LC Romanization



 

ISO843:1997

Εικόνα 4.5: Παράδειγμα 2, δασυνόμενης λέξης στα Νέα Ελληνικά


### Μεταγραμματισμός ελληνικών χαρακτήρων σε λατινικούς


Κείμενο

Αρχαία Ελληνικά

ALA-LC Romanization

ISO843:1997

Εικόνα 4.6: Παράδειγμα 2, δασυνόμενης λέξης στα Αρχαία Ελληνικά

### 4.3 Στατιστικά αποτελέσματα αξιολόγησης

Η αξιολόγηση του εργαλείου, συγκρίνοντας τα αποτελέσματά του με το σύνολο δεδομένων αναφοράς (gold dataset), πραγματοποιήθηκε στις 13 Αυγούστου 2024. Στο διάστημα 13-14 Αυγούστου έγινε η υποβολή των κατευθυντήριων ερωτημάτων στο ChatGPT και η αξιολόγηση των αποτελεσμάτων σε σχέση με τα δεδομένα αναφοράς. Να σημειωθεί πως η σύγκριση μεταξύ των δεδομένων αναφοράς με το αποτέλεσμα του κάθε εργαλείου έγινε με τη χρήση του MS Excel και συγκεκριμένα της συνάρτησης EXACT.

Παρακάτω παρουσιάζονται τα αποτελέσματα των αξιολογήσεων και των δύο εργαλείων, τα οποία έχουν αναρτηθεί μαζί με τα δεδομένα αναφοράς στο αποθετήριο της Ευρωπαϊκής Επιτροπής Zenodo και είναι διαθέσιμα στη διεύθυνση: <https://zenodo.org/doi/10.5281/zenodo.13318808>

#### 4.3.1 Αξιολόγηση εργαλείου Transliterator

Το εργαλείο Transliterator είχε 100% επιτυχία στον μεταγραμματισμό ονομάτων και φράσεων σύμφωνα με το πρότυπο ISO 843.

Στην αξιολόγηση της επιτυχίας του εργαλείου όσον αφορά το πρότυπο ALA-LC Romanization Table, οι έλεγχοι καταγράφονται ξεχωριστά, καθώς αυτό προβλέπει διαφορετικούς κανόνες για αρχαία και νέα ελληνικά.

Η επιτυχία του εργαλείου ήταν απόλυτη για τα αρχαία ελληνικά: και οι 18 περιπτώσεις μεταγράφηκαν σωστά. Στα νέα ελληνικά, επίσης το ποσοστό επιτυχίας ήταν υψηλό. Εννέα περιπτώσεις δεν μεταγράφηκαν σωστά, επειδή το εργαλείο απέτυχε να αναγνωρίσει είτε ότι οι νεοελληνικές λέξεις προέρχονται από δασυνόμενες είτε ότι πρόκειται για τα άρθρα “ο”, “η”. Στη δεύτερη περίπτωση απέτυχε, καθώς πριν από τα άρθρα υπήρχε παρένθεση ή εισαγωγικά, π.χ. (ο , «η.

Ο Πίνακας 4.1 παρουσιάζει τα αποτελέσματα της διαδικασίας αξιολόγησης.

Πίνακας 4.1: Πίνακας στατιστικών επιτυχίας προγράμματος μεταγραμματισμού

Πρότυπο	Transliterator	Ποσοστό
ISO 843	385/385	100%
ALA-LC Romanization Table - Αρχαία ελληνικά	18/18	100%
ALA-LC Romanization Table - Νέα ελληνικά	358/367	95,37%

#### 4.3.2 Αξιολόγηση ChatGPT

Το ChatGPT επέδειξε χαμηλά ποσοστά επιτυχίας (μόλις 59,74%) για το πρότυπο ISO, παρόλο που ως πρότυπο έχει πιο απλούς κανόνες και λίγες εξαιρέσεις. Τα κύρια λάθη αφορούσαν στον μεταγραμματισμό των τονούμενων χαρακτήρων και των διφθόγγων “αυ, ευ, ου”.

Το υψηλότερο ποσοστό επιτυχίας, 83,33%, εντοπίζεται στον μεταγραμματισμό αρχαίων ελληνικών ονομάτων σύμφωνα με το πρότυπο ALA-LC Romanization Table. Οι αστοχίες στα αρχαία ελληνικά σχετίζονται με τον ορθό μεταγραμματισμό του γράμματος “η”, του “υ” και φωνηέντων με περισπωμένη, π.χ. Μηνᾶς.

Στα νέα ελληνικά το ποσοστό επιτυχίας έπεσε στο 58,31%, κυρίως λόγω της αποτυχίας του εργαλείου να αναγνωρίσει ποιες νεοελληνικές λέξεις προέρχονται από δασυνόμενες της αρχαίας ελληνικής. Άλλα λάθη αφορούσαν στον μεταγραμματισμό των χαρακτήρων “β, η, φ, χ” και των διφθόγγων “αυ, ευ, ου, ντ, γκ, μπ”.

Ο Πίνακας 4.2 παρουσιάζει τα αποτελέσματα της διαδικασίας αξιολόγησης.

Πίνακας 4.2: Πίνακας στατιστικών επιτυχίας ChatGPT

Πρότυπο	ChatGPT	Ποσοστό
ISO 843	230/385	59,74%
ALA-LC Romanization Table - Αρχαία ελληνικά	15/18	83,33%
ALA-LC Romanization Table - Νέα ελληνικά	214/367	58,31%

#### 4.4 Βοηθητικό πρόγραμμα ελέγχου

Μετά την ολοκλήρωση της αποτίμησης των αποτελεσμάτων, αποφασίστηκε η δημιουργία ενός βοηθητικού προγράμματος ελέγχου, που θα αξιοποιεί τα δεδομένα αναφοράς και τα εκάστοτε αποτελέσματα του μεταγραμματισμού, ώστε να μπορεί να τα βελτιώνει. Το πρόγραμμα μεταγραμματισμού μπορεί να βελτιωθεί στο μέλλον με νέες προσθήκες αντιστοιχίσεων και κώδικα. Το βοηθητικό πρόγραμμα ελέγχου θα αξιοποιεί τα δεδομένα αναφοράς, ελέγχοντας την ορθότητα των νέων αντιστοιχίσεων και με γνώμονα πάντα τη βελτίωση του συνολικού ποσοστού πετυχημένων μεταγραμματισμών.

Ο τρόπος που λειτουργεί το βοηθητικό πρόγραμμα είναι ο εξής:

Αρχικά ο χρήστης μπορεί να ανανεώσει τα functions / Mapping στον παρακάτω κώδικα:

```
//=====
//PUT FUNCIONS HERE
{
  function modernGreekToLatinTransliterationIso843(text) { ...
  }

  function ancientGreekToLatinTransliterationIso843(text) { ...
  }

  function modernGreekToLatinTransliterationRomanization(text) { ...
  }

  function ancientGreekToLatinTransliterationRomanization(text) { ...
  }
}
//=====
```

Εικόνα 4.7: Περιοχή εισαγωγής mappings στο βοηθητικό πρόγραμμα

Και στην συνέχεια, σύμφωνα με τα δεδομένα αναφοράς, έχει δημιουργηθεί μία μεγάλη βάση δεδομένων αντιστοίχισης, όπου στο πεδίο “Input” τοποθετείται η λέξη που μας ενδιαφέρει, στο πεδίο

“expected\_ALA\_LC” το αναμενόμενο σωστό αποτέλεσμα για το πρότυπο “Romanization Tables” και στο πεδίο “expected\_ISO843” το αναμενόμενο σωστό αποτέλεσμα για το πρότυπο “Iso 843”.

Το βοηθητικό πρόγραμμα περιέχει τουλάχιστον 393 παραδείγματα και υπάρχει το περιθώριο για προσθήκη επιπλέον παραδειγμάτων.

```
// Example test cases with expected outputs
const testCases = [
  { input: 'Αἴσωπος', expected_ALA_LC: 'Aisōpos', expected_ISO843: 'Aisōpos' },
  { input: 'Ἀριστοτέλης', expected_ALA_LC: 'Aristotelēs', expected_ISO843: 'Aristotélīs' },
  { input: 'Ἀριστοφάνης', expected_ALA_LC: 'Aristophanēs', expected_ISO843: 'Aristofánīs' },
  { input: 'Βερενίκη', expected_ALA_LC: 'Berenikē', expected_ISO843: 'Verenikī' },
  { input: 'Βυζάντιος, Ἀλέξανδρος', expected_ALA_LC: 'Byzantios, Alexandros', expected_ISO843: 'Vyzántios, Aléxandros' },
  { input: 'Ἑλένη', expected_ALA_LC: 'Helenē', expected_ISO843: 'Eléni' },
  { input: 'Ἑλλάνικος', expected_ALA_LC: 'Hellanikos', expected_ISO843: 'Ellánikos' },
  { input: 'Εὐδήμος ὁ Ῥόδιος', expected_ALA_LC: 'Eudēmos ho Rhodios', expected_ISO843: 'Eúdīmos o Ródios' },
  { input: 'Ζηνοβία', expected_ALA_LC: 'Zēnobia', expected_ISO843: 'Zinová' },
  { input: 'Ἡρακλείδης', expected_ALA_LC: 'Hērakleidēs', expected_ISO843: 'Írakleídīs' },
  { input: 'Ἡράκλειτος', expected_ALA_LC: 'Hērakleitos', expected_ISO843: 'Írakleitos' },
  { input: 'Ἡσύχιος', expected_ALA_LC: 'Hēsychiōs', expected_ISO843: 'Ísychiōs' },
  { input: 'Ἰάμβλιχος', expected_ALA_LC: 'Iamblichos', expected_ISO843: 'Iámvlīchos' },
  { input: 'Στοβαῖος, Ἰωάννης', expected_ALA_LC: 'Stobaios, Iōannēs', expected_ISO843: 'Stovaiōs, Iōánnīs' },
  { input: 'Στράβων', expected_ALA_LC: 'Strabōn', expected_ISO843: 'Strávōn' },
  { input: 'Ὀραπόλλων', expected_ALA_LC: 'Hōrapollōn', expected_ISO843: 'Ōrapóllōn' },
  { input: 'Ἀγαθοκλῆς', expected_ALA_LC: 'Agathoklēs', expected_ISO843: 'Agathoklīs' },
  { input: 'Μηνᾶς', expected_ALA_LC: 'Mēnas', expected_ISO843: 'Mīnás' },
  { input: 'Ἀγραφιῶτη, Ἀναστασία', expected_ALA_LC: 'Agraphiōtē, Anastasia', expected_ISO843: 'Agrafiōtī, Anastasī' },
  { input: 'Ἀλεξάνδρου, Νικολέττα', expected_ALA_LC: 'Alexandrou, Nikoletta', expected_ISO843: 'Alexándrou, Nikolé' },
  { input: 'Ἀλεξίου, Πηνελόπη', expected_ALA_LC: 'Alexiou, Pēnelopē', expected_ISO843: 'Alexíou, Pīnelōpī' },

```

Εικόνα 4.8: Παραδείγματα για έλεγχο

Όλα αυτά τα παραδείγματα ελέγχονται από τον παρακάτω αλγόριθμο, ο οποίος είναι υπεύθυνος για την εμφάνιση των αποτελεσμάτων.

```

// Function to run the tests
function runTests() {
  const isAncientGreek = false; // Adjust as needed for your test cases

  testCases.forEach((test, index) => {
    const { input, expected_ALA_LC, expected_ISO843 } = test;

    let alaLcResult, iso843Result;

    if (isAncientGreek) {
      alaLcResult = ancientGreekToLatinTransliterationRomanization(input);
      iso843Result = ancientGreekToLatinTransliterationIso843(input);
    } else {
      alaLcResult = modernGreekToLatinTransliterationRomanization(input);
      iso843Result = modernGreekToLatinTransliterationIso843(input);
    }

    // Validate results
    const alaLcPassed = alaLcResult === expected_ALA_LC;
    const iso843Passed = iso843Result === expected_ISO843;

    // Update counters
    if (alaLcPassed && iso843Passed) {
      passedCount++;
      console.log(`Test case ${index + 1}: OK`);
    } else {
      failedCount++;
      console.log(`Test case ${index + 1}: NOT OK`);
      console.log(`  Input: ${input}`);
      if (!alaLcPassed) {
        console.log(`  ALA-LC Expected: ${expected_ALA_LC}`);
        console.log(`  ALA-LC Actual: ${alaLcResult}`);
      }
      if (!iso843Passed) {
        console.log(`  ISO 843 Expected: ${expected_ISO843}`);
        console.log(`  ISO 843 Actual: ${iso843Result}`);
      }
    }
    console.log('---');
  });

  // Print summary
  console.log(`Total Passed: ${passedCount}`);
  console.log(`Total Failed: ${failedCount}`);
}

```

Εικόνα 4.9: Κώδικας ελέγχων

Τα αποτελέσματα που επιστρέφονται έχουν την εξής δομή:

```

----
Test case 248: OK
----
Test case 249: NOT OK
  Input: Δέλτα, Πηνελόπη Σ.,
  ALA-LC Expected: Dέλta, Pīnelórī S.,
  ALA-LC Actual: Delta, Pēnelopē S.,
----

```

Εικόνα 4.10: Παράδειγμα αποτελεσμάτων ελέγχου 1

Όταν το αποτέλεσμα είναι σωστό, επιστρέφει το μήνυμα “OK” ως ένδειξη ότι όλα είναι σωστά.

Σε περίπτωση λάθους, επιστρέφει το μήνυμα “NOT OK”, ακολουθούμενο από το input, και μαζί με το όνομα του προτύπου, ακολουθούμενο από το αποτέλεσμα που αναμενόταν να εμφανιστεί, και, τέλος, μαζί με το αποτέλεσμα που τελικά προέκυψε από τις αντιστοιχίσεις, οι οποίες πραγματοποιήθηκαν με την χρήση των Mappings. Σε περίπτωση που και τα δύο πρότυπα εντοπίσουν λάθος, εμφανίζονται και τα δύο.

```

Test case 137: NOT OK
  Input: Εμείς οι ... εξωγήϊνοι;
  ALA-LC Expected: Emeis oi exōgēinoi?
  ALA-LC Actual: Emeis hoi * exōgēinoi?
  ISO 843 Expected: Emeís oi ... exōgí ĩnoi?
  ISO 843 Actual: Emeís oi * exōgí ĩnoi?

```

Εικόνα 4.11: Παράδειγμα αποτελεσμάτων ελέγχου 2

Επιπλέον, στον παραπάνω κώδικα, παρατηρείται ότι προστέθηκαν 2 counters.

```

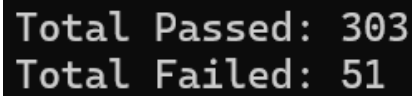
// Initialize counters
let passedCount = 0;
let failedCount = 0;

```

Εικόνα 4.12: Μετρητές σωστού / λάθους

Αυτοί προστέθηκαν, για να είναι πιο γρήγορος ο έλεγχος της ορθότητας των νέων στοιχείων που έχουν προστεθεί στα Mappings. Δηλαδή, κάθε φορά που τρέχει το script, εμφανίζεται το πλήθος των σωστών και των λανθασμένων αποτελεσμάτων. Με αυτόν τον τρόπο, ο χρήστης μπορεί με μία γρήγορη ματιά να ελέγξει την επίπτωση των αλλαγών που πραγματοποιήθηκαν στα Mappings. Όταν ο μετρητής των

λανθασμένων αυξηθεί, σημαίνει ότι οι αλλαγές που έγιναν επηρεάζουν αρνητικά τα αποτελέσματα. Αυτό σημαίνει ότι θα πρέπει να μελετηθούν περισσότερο οι προσθήκες που πραγματοποιήθηκαν, προκειμένου να μην επηρεαστούν τα υπόλοιπα αποτελέσματα.



Total Passed:	303
Total Failed:	51

Εικόνα 4.13: Παράδειγμα αποτελεσμάτων μετρητών ελέγχου

### 4.5 Επίλογος

Μετά την περιγραφή της προτεινόμενης εφαρμογής με εικόνες και παραδείγματα, παρατέθηκαν τα στατιστικά αποτελέσματα ως προς την επιτυχία της, όπως αυτά προέκυψαν από τη σύγκριση των δεδομένων αναφοράς με τα αποτελέσματα του κάθε εργαλείου/πίνακα βάσει της συνάρτησης EXACT. Διαπιστώθηκε η απόλυτη επιτυχία του προγράμματος μεταγραμματισμού σε ό,τι αφορά τους πίνακες αντιστοίχισης που βασίστηκαν στο πρότυπο ISO 843 και στο πρότυπο ALA-LC Romanization table για τα αρχαία ελληνικά, ενώ υψηλή ήταν η αποδοτικότητα του πίνακα αντιστοίχισης που αφορούσε τη ρομανοποίηση για τα νέα ελληνικά (ALA-LC Romanization table για τα νέα ελληνικά). Αντίστοιχα, η αξιολόγηση των αποτελεσμάτων από την επεξεργασία του ChatGPT εμφάνισε σαφώς χαμηλότερα ποσοστά επιτυχίας.

Για τη διόρθωση και βελτιστοποίηση του προτεινόμενου εργαλείου μεταγραμματισμού, δημιουργήθηκε ως πρόταση ένα βοηθητικό πρόγραμμα ελέγχου, με δυνατότητα περαιτέρω ανάπτυξης, το οποίο παρουσιάστηκε αναλυτικά.

## Κεφάλαιο 5ο: Συμπεράσματα ή/και προτάσεις βελτίωσης

### 5.1 Εισαγωγή

Σε αυτήν την ενότητα, παρουσιάζονται συνοπτικά τα εργαλεία που χρησιμοποιήθηκαν για την δημιουργία του προγράμματος μεταγραμματισμού: οι τεχνολογίες και η μέθοδος. Ειδικά για τη μέθοδο που επιλέχθηκε, παρουσιάζονται επιγραμματικά τα πλεονεκτήματα και τα μειονεκτήματά της και η προοπτική να διευρυνθεί η εφαρμογή της σε μεγαλύτερα κείμενα. Τέλος, γίνεται μία συγκριτική αποτίμηση των αποτελεσμάτων αξιολόγησης του εργαλείου μεταγραμματισμού και του λογισμικού ChatGPT, με αντίστοιχη ερμηνευτική προσέγγιση.

### 5.2 Ανοικτές τεχνολογίες που επιλέχθηκαν

Ως οι πλέον κατάλληλες για την κατασκευή του εργαλείου μεταγραμματισμού ανοικτές τεχνολογίες επιλέχθηκαν:

- η τεχνολογία HTML για την δημιουργία της δομής,
- η τεχνολογία CSS, για την εμφάνιση της ιστοσελίδας,
- και η τεχνολογία JavaScript, για τη λειτουργικότητά της.

Αυτές είναι οι πιο δημοφιλείς παγκοσμίως.

### 5.3 Για μέθοδο που επιλέχθηκε - τα πλεονεκτήματα, τα μειονεκτήματα της μεθόδου όπως εφαρμόστηκε στο εργαλείο, πιθανοί περιορισμοί, συσχετισμός με άλλες σύγχρονες μεθόδους

Επιλέχθηκε η μέθοδος αντιστοίχισης (mapping method), η οποία υπερτερεί έναντι άλλων, για τους εξής λόγους:

- Καθορίζει τις αντιστοιχίσεις σε μεγάλο εύρος και με ακρίβεια.
- Διαχειρίζεται καλύτερα την ανάγνωση του κειμένου.
- Έχει τη δυνατότητα ανάγνωσης και ανίχνευσης σύνθετων συνδυασμών.
- Χαρακτηρίζεται από ευκολία στη συντήρηση, επέκταση και προσαρμοστικότητα, χαρακτηριστικά που την καθιστούν αξιοποιήσιμη και στο μέλλον.

Ωστόσο, υπάρχουν και κάποια μειονεκτήματα:

- Αδυνατεί να ανταποκριθεί σε διαφορετικά πρότυπα μεταγραμματισμού, όταν αυτά “επιστρέφουν” αποτελέσματα ανόμοια στην έξοδο, ενώ η είσοδος είναι κοινή.
- Δεν μπορεί να διαχειριστεί τις ειδικές περιπτώσεις εύστοχα, χωρίς τη βοήθεια επιπλέον κώδικα.
- Αδυνατεί να αναγνωρίσει στη λέξη τη θέση ενός χαρακτήρα.

Η μέθοδος αντιστοίχισης διέπεται από σταθερότητα και προβλεψιμότητα στην εφαρμογή της, εξασφαλίζει απλότητα, συνέπεια και ταχύτητα στη μετατροπή, γεγονός που την καθιστά αποτελεσματική για καθορισμένα και σταθερά σύνολα δεδομένων. Η μέθοδος αντιστοίχισης διακρίνεται για την **απλότητα και την αποτελεσματικότητά της** σε συστήματα με σταθερούς κανόνες, όπου απαιτείται ταχύτητα και συνέπεια.

Το στατιστικό μοντέλο μεταγραμματισμού χαρακτηρίζεται από ευελιξία και προσαρμοστικότητα, είναι πιο περίπλοκη προσέγγιση και ενδείκνυται για περιπτώσεις όπου απαιτείται διαχείριση ασαφών και θορυβωδών δεδομένων. Είναι κυρίως ένα εργαλείο ελέγχου της ποιότητας και εγκυρότητας των αποτελεσμάτων της μεταγραφής.

Η αντίστροφη προσέγγιση προσφέρει μεγαλύτερη **προσαρμοστικότητα σε σύνθετα δεδομένα**, καθώς επιτρέπει την ανάκτηση πολλαπλών πιθανών αποτελεσμάτων και την κατάταξή τους με βάση την πιθανότητα ακρίβειας. Αξίζει, βέβαια, να σημειωθεί ότι η αντίστροφη προσέγγιση σχεδιάστηκε πρωτίστως για να μεταγράφει, και όχι μόνο για να μεταγραμματίζει, και δεν περιορίζεται σε ονόματα και τίτλους, αλλά εφαρμόζεται σε μεγαλύτερα κείμενα.

### 5.4 Προκλήσεις για κάθε πρότυπο.

Στο πρότυπο ALA-LC romanization table για τη νεοελληνική γλώσσα και, συγκεκριμένα, για τις δασυνόμενες λέξεις, αξιοποιήθηκε μία λίστα 804 θεμάτων λέξεων με 3-5 χαρακτήρες. Για τον μεταγραμματισμό μεγαλύτερων κειμένων, θα χρειαστεί η προσθήκη περισσότερων θεμάτων-ριζών, που να χρησιμοποιούν μεγαλύτερο αριθμό χαρακτήρων. Μια σχετική έρευνα των υπαλλήλων της ΕΒΕ εντόπισε επιπλέον 24 θέματα-ρίζες με συνδυασμό 6-7 χαρακτήρων. Για τον χειρισμό αυτών των περιπτώσεων, χρειάζεται η ανάπτυξη επιπλέον κώδικα.

### 5.5 Σχολιασμός αποτελεσμάτων αξιολόγησης. Γιατί πήγε το εργαλείο καλύτερα από το Chat Gpt;

Το εργαλείο Transliterator, το οποίο σχεδιάστηκε στην παρούσα εργασία, είχε πολύ μεγαλύτερα ποσοστά επιτυχίας και στα δύο πρότυπα, συγκρινόμενο με το ChatGPT. Το πρότυπο ISO 843 έχει απλούς κανόνες και ελάχιστες εξαιρέσεις. Το Transliterator εφάρμοσε κανόνες και εξαιρέσεις με απόλυτη επιτυχία, σε αντίθεση με το ChatGPT. Η μόνη υπόθεση που μπορούμε να κάνουμε γι' αυτό είναι ότι το ChatGPT ως εργαλείο έχει εκπαιδευθεί με δεδομένα από τις ΗΠΑ, όπου για τον μεταγραμματισμό ελληνικών χαρακτήρων χρησιμοποιείται σχεδόν αποκλειστικά το πρότυπο των ALA-LC.

Όσον αφορά στο πρότυπο ALA-LC Romanization Table, το εργαλείο Transliterator είχε, επίσης, μεγάλο ποσοστό επιτυχίας (100% για τα αρχαία ελληνικά και 95,37% για τα νέα ελληνικά). Προφανώς, η προσθήκη θεμάτων-ριζών για την ανίχνευση νεοελληνικών λέξεων που προέρχονται από αρχαίες δασυνόμενες λέξεις συνέβαλε στην ακρίβεια των αποτελεσμάτων. Η αποτυχία ανίχνευσης άρθρων που ακολουθούν ειδικούς χαρακτήρες, όπως παρενθέσεις και εισαγωγικά, μπορεί να λυθεί με τροποποίηση του κώδικα στο μέλλον.

Το ChatGPT επέδειξε υψηλό ποσοστό για τα αρχαία ελληνικά (83,33%). Αυτό που προβληματίσε ήταν πως δεν ήταν συνεπές στα λάθη που έκανε. Δηλαδή, μεταγραμμάτιζε άλλοτε ορθά και άλλοτε λανθασμένα τους ίδιους χαρακτήρες. Όσον αφορά στα νέα ελληνικά, το ποσοστό επιτυχίας έπεσε στο 58,31%. Στο κατευθυντήριο ερώτημα είχε συμπεριληφθεί κατάλογος δασυνόμενων λέξεων της αρχαίας ελληνικής, ωστόσο, το εργαλείο δε φαίνεται να αξιοποίησε επαρκώς αυτή τη λίστα. Προφανώς, απαιτείται περισσότερη εκπαίδευση του ChatGPT με σχετικά δεδομένα. Δε γνωρίζουμε αν και σε ποια έκταση μια τέτοια εκπαίδευση είναι εφικτή, καθώς σχετικές πληροφορίες υπάρχουν σε βιβλία και

λεξικά για τα οποία υπάρχουν πνευματικά δικαιώματα και δεν είναι ακόμα γνωστό σε ποιον βαθμό μπορεί το ChatGPT να αξιοποιεί υλικό προστατευόμενο από τη νομοθεσία περί πνευματικής ιδιοκτησίας.

## Επίλογος

Η δημιουργία ενός εργαλείου μεταγραμματισμού για την ελληνική γλώσσα κρίθηκε χρήσιμη και αναγκαία για την διευκόλυνση και την αξιοπιστία στο διαμοιρασμό μεταδεδομένων μεταξύ των βιβλιοθηκών, τοπικών και εθνικών, πρωτίστως για την Εθνική Βιβλιοθήκη της Ελλάδας. Η ελληνική γλώσσα είναι μία γλώσσα ζωντανή για τουλάχιστον 2.500 χρόνια. Καθώς εξελίχθηκε από την αρχαία έως τη νεοελληνική μορφή της, υπέστη πολλές αλλαγές στην ορθογραφία της. Αυτή η ιδιαιτερότητά της αποτέλεσε μία πολύ απαιτητική συνθήκη στη διαδικασία κατασκευής ενός αποτελεσματικού προγράμματος για το μεταγραμματισμό της σε λατινικούς χαρακτήρες.

Ωστόσο, η προσπάθεια αυτή υπήρξε πολύ ενδιαφέρουσα και γόνιμη. Μέσα από τη βιβλιογραφική έρευνα των προτύπων, των απαιτούμενων τεχνολογιών και των μεθόδων για τον μεταγραμματισμό των ελληνικών χαρακτήρων σε λατινικούς (Romanization), προέκυψε το σύστημα μεταγραμματισμού εγγραφών βιβλιοθήκης που προτείνεται στην παρούσα εργασία.

Οι δυσκολίες για τον μεταγραμματισμό μιας γλώσσας που δεν γράφεται με λατινικούς χαρακτήρες και οι αντίστοιχες ειδικές περιπτώσεις αντιμετωπίστηκαν με την σύνταξη εξειδικευμένων αλγορίθμων, που να λειτουργούν αποτελεσματικά με ταχύτητα, συνέπεια και ακρίβεια. Επίσης, επιδιώχθηκε ο σχεδιασμός ενός γραφικού περιβάλλοντος φιλικού προς τον χρήστη, συγκεκριμένα, εύχρηστου, ευανάγνωστου, με σαφή δομή και αμεσότητα στη διάδραση.

Η δοκιμαστική λειτουργία του προγράμματος στην Εθνική Βιβλιοθήκη της Ελλάδας απέδωσε πολύ υψηλά ποσοστά επιτυχίας. Στην περίπτωση, μάλιστα, του μεταγραμματισμού των χαρακτήρων της αρχαίας ελληνικής γλώσσας, η επιτυχία ήταν απόλυτη.

Για την αντιμετώπιση των όποιων αστοχιών προέκυψαν ή θα προκύπτουν στο μέλλον από τη χρήση του εργαλείου, δημιουργήσαμε και ένα βοηθητικό πρόγραμμα ελέγχου, με επιπλέον αλγόριθμους και κώδικες. Αυτό το συμπληρωματικό εργαλείο, ουσιαστικά, επιταχύνει τον έλεγχο της ορθότητας των αποτελεσμάτων και διευκολύνει περισσότερο την επεκτασιμότητα και την ευκολία στη συντήρηση του προγράμματος μεταγραμματισμού.

Αξίζει να σημειωθεί ότι το θέμα της παρούσας διπλωματικής εργασίας παρουσιάζεται στο άρθρο "Building a transliteration tool to enhance the exchange of metadata about Greek authors" [23] των C. Bratsas, L. Ioannidis, I. Tsitsilegkas, M. Gerolimos, E. Koufakou, S. Zarpounidou, το οποίο έχει γίνει αποδεκτό από το MTSR και θα συμπεριληφθεί ως εισήγηση στο 18ο International Conference on Metadata and Semantics Research, που θα πραγματοποιηθεί στις 20-22 Νοεμβρίου 2024.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] ΕΛΟΤ, *ΕΛΟΤ 743 Πληροφορίες και τεκμηρίωση - Μετατροπή των Ελληνικών χαρακτήρων με χαρακτήρες Λατινικούς*. Αθήνα: Ελληνικός Οργανισμός Τυποποίησης, 2001.
- [2] ISO, *ISO 843 Information and documentation - Conversion of Greek characters into Latin characters*. Genève: International Standards Organization, 1999.
- [3] “BGN/PCGN romanization,” Wikipedia. Ημερομηνία πρόσβασης: 19 Αυγούστου 2024. Διαθέσιμο σε: [https://en.wikipedia.org/wiki/BGN/PCGN\\_romanization](https://en.wikipedia.org/wiki/BGN/PCGN_romanization)
- [4] DIN, *DIN 31364 Information und Dokumentation - Umschrift des griechischen Alphabets*. Berlin: Deutsches Institut für Normung, 2011.
- [5] American Library Association and Library of Congress, *Romanization table for Greek*. Washington, DC, ALA-LC, 2010. Ημερομηνία πρόσβασης: 19 Αυγούστου 2024. Διαθέσιμο σε: <https://www.loc.gov/catdir/cpsd/romanization/greek.pdf>
- [6] ISSN, *ISSN Manual: 24 Alternate Graphic Representation*. Paris: ISSN, 2015. Ημερομηνία πρόσβασης: 19 Αυγούστου 2024. Διαθέσιμο σε: [https://www.issn.org/wp-content/uploads/2021/09/ISSNManual\\_ENG2015\\_23-01-2015.pdf](https://www.issn.org/wp-content/uploads/2021/09/ISSNManual_ENG2015_23-01-2015.pdf)
- [7] PCC, *PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets*. Washington, DC, Program for Cooperative Cataloging, 2017. Ημερομηνία πρόσβασης: 19 Αυγούστου 2024. Διαθέσιμο σε: <https://loc.gov/aba/pcc/bibco/documents/PCCNonLatinGuidelines.pdf>
- [8] “Εθνική Βιβλιοθήκη της Ελλάδος,” Βικιπαίδεια. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: [https://el.wikipedia.org/wiki/Εθνική\\_Βιβλιοθήκη\\_της\\_Ελλάδος](https://el.wikipedia.org/wiki/Εθνική_Βιβλιοθήκη_της_Ελλάδος)
- [9] Ελλάδα, “Ν. 4452, Ρύθμιση θεμάτων του Κρατικού Πιστοποιητικού Γλωσσομάθειας, της Εθνικής Βιβλιοθήκης της Ελλάδας και άλλες διατάξεις,” *Εφημερίδα της Κυβερνήσεως*, Αρ.17, 15 Φεβρουαρίου 2017. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: <https://www.nlg.gr/wp-content/uploads/2018/05/%CE%9D.4452-17.pdf>
- [10] D. Flanagan, *JavaScript: The Definitive Guide* (7th ed.). O'Reilly Media, 2020.
- [11] L. Chen, H. Wang, and Y. Zhang, "Performance Improvements in Code Transformation Using Mapping Structures," *J. Softw. Eng. Appl.*, vol. 15, no. 3, pp. 123-135, 2022.
- [12] A. Rodriguez, "Advanced Metaprogramming Techniques in Modern JavaScript," Ph.D. dissertation, Dept. Comp. Sci., Massachusetts Inst. of Tech., Cambridge, MA, 2021.
- [13] J. Rodriguez, "Mapping Techniques in Code Transformation Systems," Ph.D. dissertation, Univ. of California, Berkeley, CA, USA, 2021.
- [14] J. Duckett, *HTML and CSS: Design and Build Websites*. Wiley, 2011.
- [15] A. Marcotte, *Responsive Web Design*. A Book Apart, 2011.

[16] ISO, *ISO/R 843 International system for the transliteration of Greek characters into Latin characters*. Genève: International Standards Organization, 1968.

[17] ISO, *ISO 843:1997, Information and documentation — Conversion of Greek characters into Latin characters*. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: <https://www.iso.org/standard/5215.html>

[18] “Πολυτονική γραφή της νέας ελληνικής γλώσσας,” Wikipedia. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: [https://el.wikipedia.org/wiki/Πολυτονική\\_γραφή\\_της\\_ελληνικής\\_γλώσσας](https://el.wikipedia.org/wiki/Πολυτονική_γραφή_της_ελληνικής_γλώσσας)

[19] The Polytonic Project, *Μετατροπή μονοτονικοῦ κειμένου σε πολυτονικό*. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: <https://thepolytonicproject.gr/tonizo/>

[20] Lexilogos, *Ancient Greek Keyboard*. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: [https://www.lexilogos.com/keyboard/greek\\_ancient.html](https://www.lexilogos.com/keyboard/greek_ancient.html)

[21] Library of Congress, “*Romanization Landscape*”. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: [https://www.loc.gov/catdir/cpsd/romlandscape\\_Oct2011.html](https://www.loc.gov/catdir/cpsd/romlandscape_Oct2011.html)

[22] “Κατηγορία:Ανδρικά επώνυμα (νέα ελληνικά),” Wiktionary. Ημερομηνία πρόσβασης: 22 Αυγούστου 2024. Διαθέσιμο σε: [https://el.wiktionary.org/w/index.php?title=Κατηγορία:Ανδρικά\\_επώνυμα\\_\(νέα\\_ελληνικά\)](https://el.wiktionary.org/w/index.php?title=Κατηγορία:Ανδρικά_επώνυμα_(νέα_ελληνικά))

[23] C. Bratsas, L. Ioannidis, I. Tsitsilegkas, M. Gerolimos, E. Koufakou, S. Zapounidou, "Building a transliteration tool to enhance the exchange of metadata about Greek authors," MTSR 2024, Lect. Notes Comput. Sci. [In Press]

## Παράρτημα Α: Κατάλογος δασυνόμενων λέξεων και ονομάτων

### Α

#### Κύρια ονόματα

Άγνη, Άγνοδίκη, Άδης, Αἴμος, Αἴμων, Ἀλιάκμων, Ἀλίαρτος, Ἀλικαρνασσός, Ἀλόνησος, Ἀμαδρυάς

#### Λοιπές λέξεις

άβρός, ἄγιος, ἀγνίζω, ἀγνός, ἀδρός, ἀθρόος (στην αττική διάλεκτο), αἷμα, αἵρεση, αἵρεση, αἰρετός, αἰρέω-ῶ, αἰ ἄλαι (= ἡ αλυκή), ἄλας, ἄλατα, ἄλατι, ἀλιεία, ἀλιεύω, ἀλίπαστος, ἀλίπεδο, ἄλις (= αρκετά), ἀλίσκομαι, ἄλλομαι (= πηδῶ), ἄλμα, ἄλμη, ἀλμυρός, ἄλς (ὑφαλος), ἀλτήρ, ἀλτήρες, ἄλτης, ἀλυκή, ἀλυσίδα, ἄλυσις, ἄλώνι, ἀλωνίζω, ἄλως, ἄλωση, ἄμα, ἄμαξα, ἀμαρτάνω, ἀμαρτία, ἄμιλλα, ἀνύτω (αλλά και ἀνύω), ἀπαλός, ἄπαξ, ἀπλός, ἀπλοῦς, ἀπλώνω, ἄπτομαι, ἀπτός, ἄρμα - ἄρματα (αναφέρεται στο ιππήλατο ὄχημα - διαφορετικό ἀπὸ τὰ ἄρματα λαμβάνει ψιλή γιατί προέρχεται ἀπὸ τὴ λατινική λέξη arma = ὄπλα), ἄρμη, ἀρμόζω, ἀρμονία, ἀρμός, ἄρπα, ἀρπάζω, ἄρπαξ, ἀφή, ἀψίδα, ἀψικόρος, ἀψιμαχία, ἀψίς, ἀψύς

---

### Ε

#### Κύρια ονόματα

Ἐβραῖος, Ἐβρος, Ἐκάβη, Ἐκάτη, Ἐκτορας, Ἐκτωρ, Ἐλένη, Ἐλικών, Ἐλικώνας, Ἐλλάς, Ἐλλη, Ἐλλην, Ἐρμῆς, Ἐρμιόνη

#### Λοιπές λέξεις

ἑαυτός, ἑβδομάς, ἑδρα, ἑδῶλιον, εἴλωτας, εἰμαρμένη, εἰρκτή (κάθειρξη), εἰρμός, ἕκαστος, ἑκατό, ἐκών, ἐλιγμός, ἑλικά(ς), ἐλίσσομαι, ἑλκηθρο, ἑλκος, ἐλκύω, ἑλλανοδίκης, ἑλληνικός, ἑλξη, ἐλονοσία, ἑλος, ἑνα, ἑνεκα ἢ ἑνεκεν, ἐνικός, ἐνώνω, ἑνωσι, ἑξη, ἐξῆς, ἑξι, ἐορτή, ἐπομένως, ἐπτά, ἑρμα, ἑρμαιο, ἑρμαφρόδιτος, ἑρμηνεία, ἑρμηνεύω, ἑρμητικός, ἐρπετό, ἑρπω, ἑση (ἔφεση), ἑσμός, ἑσπέρα, ἑσπερινός, ἑστία, ἐστιατόριο, ἐταιρεία, ἐταῖρος, ἑτοιμος, εὕρεση, εὐρετήριο, εὕρημα, εὐρίσκω, ἑψω (= βράζω), ἑωλος (= παλιός, ὄχι πρόσφατος), ἑως (= αυγή και ὄχι ἡ πρόθεση)

---

### Η

#### Κύρια ονόματα

Ἡβη, Ἡλιαία, Ἡλιος, Ἡρα, Ἡρακλείδης, Ἡράκλειον, Ἡράκλειος, Ἡράκλειτος, Ἡρακλῆς, Ἡρόδοτος, Ἡρόστρατος, Ἡρώ, Ἡρώδης, Ἡρωδιανός, Ἡρων, Ἡρώνδας, Ἡσαΐας, Ἡσίοδος, Ἡσύχιος, Ἡφαιστος

#### Λοιπές λέξεις

ἡγεμόνας, ἡγεσία, ἡγήτορας, ἡγοῦμαι, ἡγούμενος, ἡδονή, ἡδύς, ἡκιστα, ἡκω, ἡλικία, ἡλιξ (= συνομήλικος, σύντροφος), ἡλος, ἡμέρα, ἡμερος, ἡμι- (αχώριστο μόριο π.χ. ἡμισφαίριο), ἡμισυ,

ἥμισυς, ἡ ἡνία και τὰ ἡνία (= χαλινός), ἡνίοχος, ἥπαρ, ἥρωας, ἥρως, ἡσυχία, ἡσυχος, ἦπτα, ἡφαίστειον

---

## I

### Κύρια ονόματα

Ἰερουσαλήμ, Ἰερώνυμος, Ἰππίας, Ἰπποκλείδης, Ἰπποκράτης, Ἰππόλυτος, Ἴππος

### Λοιπές λέξεις

ἴδρυμα, ἰδρύω, ἰδρώνω, ἰδρώς, ἰδρώτας, ἰερά, ἰέρακας, ἰέραξ, ἱερεύς, ἱερή, ἱερός, ἱερων, ἴζημα, ἴημι, ἱκανός, ἱκεσία, ἱκετεύω, ἱκέτης, ἱκνός, ἰλαρά, ἰλαρός, ἴλεως, ἰμάντας, ἰμάτιο, ἰμάτιον, ἴμερος (=πόθος), ἴνα, ἱππικό, ἵπταμαι, ἴσταμαι, ἰστίο, ἱστορία, ἰστός, ἴστωρ (γεν. -ορος = ἔμπειρος, γνώστης)

---

## O

### Κύρια ονόματα

Ὅμηρος

### Λοιπές λέξεις

ὀδηγός, ὀδηγῶ, ὀδός, ὀλκή, ὀλκός, ὄλμος, ὀλόκληρος, ὄλος, ὀμάδα, ὀμαλός, ὀμας, ὀμιλία, ὀμιλος, ὀμιλῶ, ὀμίχλη, ὀμο- (πρώτο συνθετικό σε λέξεις ὅπως ὀμοεθνής, ὀμόλογος, ὀμότεχνος κ.λπ.), ὀμοιος, ὀμοῦ, ὄμως, ὀπλή, ὄπλο, ὄπλον, ὄραμα, ὄραση, ὄρασις, ὄρατός, ὀρίζω, ὄριο, ὄριον, ὄρκος, ὄρμαθός, ὄρμη, ὄρμῶνη, ὄρμος, ὄρμῶ, ο ὄρος (ἀλλὰ το ὄρος=βουνό), ὄσιος, ὄσος, ὄστις, ὄτε

---

## Υ

### Κύρια ονόματα

Ὑδρα, Ὑμητός, Ὑπαπαντή, Ὑπάτη, Ὑπατία, Ὑπερείδης, Ὑπερίων, Ὑρκανός, Ὑψηλάντης

### Λοιπές λέξεις

Ὅλες οι λέξεις ἀπὸ ὑ ἐκτὸς ἀπὸ τῆ λέξη ὑψιλον (ῥ ψιλόν)

---

## Ω

### Κύρια ονόματα

Ὦρα

### Λοιπές λέξεις

ῥῶδε, ῥῶρα, ῥῶραϊος, ῥῶριμος, ῶς, ῶσαύτως, ῶστε