



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ

Ηλεκτρονική Υπηρεσία Δημοσίευσης των Συζητήσεων του Κοινοβουλίου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΕΡΕΤΑΚΟΥ ΓΕΩΡΓΙΟΥ

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Μάριος Κόνιαρης
Ε.ΔΙ.Π. Ε.Μ.Π.

Αθήνα, Ιούλιος 2025



Ηλεκτρονική Υπηρεσία Δημοσίευσης των Συζητήσεων του Κοινοβουλίου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΕΡΕΤΑΚΟΥ ΓΕΩΡΓΙΟΥ

Επιβλέπων: Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων: Μάριος Κόνιαρης
Ε.ΔΙ.Π. Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10η Ιουλίου 2025.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

.....
Γ. Ματσόπουλος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2025



Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.

Γεώργιος Σερετάκος, 2025.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....

Γεώργιος Σερετάκος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

10η Ιουλίου 2025

Περίληψη

Η δημοσιοποίηση και μελέτη του κοινοβουλευτικού λόγου αποτελεί θεμέλιο της δημοκρατικής διακυβέρνησης, της πολιτικής ανάλυσης και της επιστημονικής έρευνας στις κοινωνικές επιστήμες. Το υλικό, ωστόσο, που διατίθεται από το κοινοβούλιο στο κοινό παρουσιάζει σημαντικά προβλήματα προσβασιμότητας, διαλειτουργικότητας και δυνατότητας επεξεργασίας. Η απουσία ενιαίας, ανοιχτής και επεξεργάσιμης μορφής καθιστά δύσκολη την αναζήτηση, ανάλυση και αξιοποίηση των κοινοβουλευτικών συζητήσεων από ερευνητές, δημοσιογράφους, και πολίτες.

Η παρούσα διπλωματική εργασία προτείνει και υλοποιεί ολοκληρωμένο σύστημα για την αναπαράσταση, οργάνωση και ανάκτηση του κοινοβουλευτικού λόγου αξιοποιώντας ανοικτά πρότυπα και τεχνολογίες του Σηματολογικού Ιστού. Στόχος είναι η δομημένη αποτύπωση του περιεχομένου, η ενίσχυσή του με μεταδεδομένα και η υποστήριξη στοχευμένης αναζήτησης. Μέσω της προσέγγισης αυτής, οι κοινοβουλευτικές συζητήσεις καθίστανται περισσότερο προσβάσιμες και αξιοποιήσιμες στο πλαίσιο της επιστημονικής έρευνας. Το σύστημα ενισχύει τη μελέτη της θεσμικής λειτουργίας, της πολιτικής ρητορικής και των κοινωνικών εξελίξεων, γεφυρώνοντας την τεχνολογική και την κοινωνικοπολιτική διάσταση της πληροφορίας.

Η εργασία εντάσσεται στο διεπιστημονικό πεδίο μεταξύ Πληροφορικής, Πολιτικής Επιστήμης και Ψηφιακής Ανθρωπιστικής Έρευνας, αναδεικνύοντας τις δυνατότητες των τεχνολογιών πληροφορίας στην ενίσχυση της διαφάνειας, της πρόσβασης στη γνώση και της δημοκρατικής λογοδοσίας. Παράλληλα, προτείνει ένα λειτουργικό πρότυπο για την ψηφιακή επεξεργασία των κοινοβουλευτικών δεδομένων, το οποίο μπορεί να αξιοποιηθεί ως βάση για μελλοντικές εφαρμογές μελετών λόγου, θεσμικής ανάλυσης ή ανοικτών δεδομένων.

Λέξεις Κλειδιά

Ελληνικό Κοινοβούλιο, Ανοιχτά Διασυνδεδεμένα Δεδομένα, Ανάκτηση Πληροφορίας

Abstract

The publication and study of parliamentary discourse are fundamental to democratic governance, political analysis, and scholarly research in the social sciences. However, the material made available by the parliament to the public poses significant challenges in terms of accessibility, interoperability, and structured processing. The absence of a unified, open, and editable format hinders the ability of researchers, journalists, and citizens to search, analyze, and make use of parliamentary discussions.

This diploma thesis proposes and implements an integrated system for the representation, organization, and retrieval of parliamentary discourse, leveraging open standards and Semantic Web technologies. The goal is to achieve structured content representation, metadata enrichment, and support for targeted search. This approach renders parliamentary discussions more accessible and amenable to scientific study. The system facilitates the study of institutional functions, political rhetoric, and social developments, bridging the technological and sociopolitical dimensions of information.

This thesis lies at the intersection of Computer Science, Political Science, and Digital Humanities, highlighting the potential of information technologies to advance transparency, knowledge accessibility, and democratic accountability. In parallel, it introduces a practical framework for the digital processing of parliamentary data, which can serve as a foundation for future applications in discourse analysis, institutional research, and open data initiatives.

Keywords

Hellenic Parliament, Open Linked Data, Information Retrieval

στην οικογένειά μου

Ευχαριστίες

Ξεκινώντας, θα ήθελα να εκφράζω τη βαθύτατη ευγνωμοσύνη μου στον κύριο Μάριο Κόνιαρη. Οι εβδομαδιαίες μας συναντήσεις προσέφεραν πολύτιμη καθοδήγηση. Οι γνώσεις, η επιμονή του, αλλά πρωτίστως το ενδιαφέρον του για εμένα με βοήθησαν να βελτιώσω τις ιδέες μου και να ολοκληρώσω την διπλωματική μου εργασία. Η ευγνωμοσύνη μου δεν γίνεται να μην επεκταθεί και στον κύριο Παναγιώτη Τσανάκα, που με εμπιστεύτηκε να φέρω εις πέρας αυτό το απαιτητικό θέμα που διαπραγματεύεται σε αυτήν την εργασία.

Στη συνέχεια, θα ήθελα να ευχαριστήσω την μητέρα μου Ρεγγίνα και τον παππού μου Τίμο, οι οποίοι στάθηκαν δίπλα μου, δίνοντας μου ατελείωτη υποστήριξη, αγάπη και κατανόηση κατά τη διάρκεια όλων των ακαδημαϊκών μου χρόνων.

Τέλος, θέλω να πω ένα μεγάλο ευχαριστώ και στους φίλους μου, των οποίων η παρέα συνέβαλε στη δημιουργία μιας ευχάριστης ατμόσφαιρας που με βοήθησε να ξεπεράσω ακόμη και τις πιο δύσκολες και πιεστικές ημέρες.

Αθήνα, Ιούλιος 2025

Γεώργιος Σερετάκος

Περιεχόμενα

Περίληψη	4
Abstract	5
Ευχαριστίες	7
1 Εισαγωγή	13
1.1 Περιγραφή Προβλήματος	13
1.2 Αντικείμενο Διπλωματικής	14
1.3 Οργάνωση Κειμένου	14
2 Θεωρητικό Υπόβαθρο και Σχετικές Εργασίες	16
2.1 Θεωρητικό υπόβαθρο	16
2.1.1 Σημασιολογικός Ιστός (Semantic Web)	16
2.1.2 Ανοιχτά Διασυνδεδεμένα Δεδομένα (Linked Open Data)	18
2.1.3 Τεχνολογίες Δομημένης Αναπαράστασης Δεδομένων	19
2.1.4 Ανοιχτά Κυβερνητικά Δεδομένα	21
2.1.5 Τεχνικές Ανάκτησης Πληροφορίας	22
2.1.6 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP)	24
2.2 Σχετικές Εργασίες και Παραδείγματα	25
2.2.1 Δομημένη Αναπαράσταση Κοινοβουλευτικών Κειμένων	26
2.2.2 Μεθοδολογικές Προσεγγίσεις και Τεχνολογικές Εφαρμογές στην Ανάλυση του Κοινοβουλευτικού Λόγου	27
2.2.3 Επιπτώσεις για τη Μελλοντική Έρευνα	28
3 Σχεδιασμός Συστήματος	29
3.1 Λειτουργικές Απαιτήσεις	29
3.1.1 Εισαγωγή, Εμπλουτισμός και Διαχείριση Δεδομένων	29
3.1.2 Ανάλυση Περιεχομένου με Φυσική Γλώσσα (NLP)	29
3.1.3 Αναζήτηση Πλήρους Κειμένου	30
3.1.4 Αναζήτηση Πρακτικών	30
3.1.5 Αναζήτηση Ομιλητών	31
3.1.6 Προβολή Στατιστικών Στοιχείων	31
3.2 Αρχιτεκτονική Δομή	31

4 Μοντέλο Δεδομένων Συστήματος - Βάση Δεδομένων	33
4.1 Εννοιολογική Επισκόπηση των Οντοτήτων	33
4.2 Σχήμα της Βάσης Δεδομένων	34
4.2.1 Μοντέλο Οντοτήτων Συσχετίσεων	34
4.2.2 Αναλυτική Περιγραφή Οντοτήτων	34
4.2.3 Συσχετίσεις μεταξύ των Οντοτήτων	43
5 Λειτουργίες και Περιγραφή Τελικού Συστήματος	45
5.1 Παρουσίαση Υποσυστημάτων	45
5.1.1 Υποσύστημα Εισαγωγής και Εμπλουτισμού Δεδομένων	46
5.1.2 Υποσύστημα Επεξεργασίας Ομιλιών με Φυσική Γλώσσα (NLP)	48
5.1.3 Υποσύστημα Αναζήτησης Πλήρους Κειμένου με Elasticsearch	49
5.1.4 Υποσύστημα Διεπαφής Χρήστη και Παρουσίασης Περιεχομένου	50
5.2 Διασύνδεση και Ανάκτηση Δεδομένων μέσω API	51
5.2.1 /api/search-debates	51
5.2.2 /api/search-in-debate	53
5.2.3 /api/debates/metadata/[id]	53
5.2.4 /api/search-speakers	54
5.2.5 /api/search-speaker-speeches	55
5.2.6 /api/speakers/metadata/[id]	57
5.2.7 Δημόσια GET Endpoints του Συστήματος Διαχείρισης Περιεχομένου	58
6 Αξιολόγηση του Συστήματος	59
6.1 Μεθοδολογία Ελέγχου	59
6.2 Ερωτηματολόγιο Αξιολόγησης	60
6.2.1 Ενότητα 1: Pre-Test Questionnaire	61
6.2.2 Ενότητα 2: Task Scenarios	61
6.2.3 Ενότητα 3: System Usability Scale (SUS)	63
6.2.4 Ενότητα 4: Technology Acceptance Model (TAM)	63
6.2.5 Ενότητα 5: Post-Test Interview Guide	64
6.3 Συμπεράσματα από την Αξιολόγηση του Συστήματος	64
7 Επίλογος	67
7.1 Σύνοψη και Γενικά Συμπεράσματα	67
7.2 Μελλοντικές Επεκτάσεις	68
Βιβλιογραφία	73

Κατάλογος Σχημάτων

3.1	Η αρχιτεκτονική του συστήματος, με έμφαση στις ροές δεδομένων μεταξύ των επιπέδων και την αλληλεπίδραση με εξωτερικές υπηρεσίες.	32
4.1	Αναπαράσταση οντοτήτων και συσχετίσεων του μοντέλου δεδομένων του συστήματος	34
5.1	Διαδικασία επεξεργασίας, εμπλουτισμού, ευρετηρίασης και παρουσίασης των κοινοβουλευτικών δεδομένων.	45
5.2	Στάδια μετατροπής, εμπλουτισμού και αποθήκευσης των κοινοβουλευτικών πληροφοριών	46
5.3	Διάγραμμα ροής του υποσυστήματος ευρετηρίασης και αναζήτησης	50
5.4	Ενδεικτική απάντηση του endpoint /api/search-debates	52
5.5	Ενδεικτική απάντηση του endpoint /api/search-in-debate	53
5.6	Ενδεικτική απάντηση του endpoint /api/debates/metadata/[id]	54
5.7	Ενδεικτική απάντηση του endpoint /api/search-speakers	55
5.8	Ενδεικτική απάντηση του endpoint /api/search-speaker-speeches	56
5.9	Ενδεικτική απάντηση του endpoint /api/speakers/metadata/[id]	57

Κατάλογος Εικόνων

2.1	Η εξέλιξη του διαδικτύου προς τον Σημασιολογικό Ιστό και τον Web 5.0, με έμφαση στη σημασιολογία των πληροφοριών και των κοινωνικών διασυνδέσεων.	17
2.2	Απεικόνιση του οικοσυστήματος Ανοικτών Συνδεδεμένων Δεδομένων με κόμβους ως σύνολα δεδομένων και συνδέσμους ως σημασιολογικές σχέσεις. . . .	18
2.3	Η εξέλιξη του Νέφους Συνδεδεμένων Ανοικτών Δεδομένων (2007–2017), με έμφαση στην αύξηση του αριθμού και της διασύνδεσης συνόλων δεδομένων .	19
2.4	Παράδειγμα δομής εγγράφου σε XML για την αναπαράσταση στοιχείων ενός πολιτικού προσώπου	20
2.5	Παράδειγμα δομής εγγράφου σε JSON για την αναπαράσταση στοιχείων ενός πολιτικού προσώπου	21
5.1	Απόσπασμα XML εγγράφου της συνεδρίασης της 1ης Μαρτίου 1994 με πρόλογο και επιμέρους ομιλίες (preface, speech)	47
5.2	Αναδομημένη μορφή εισαγωγικού μέρους και ομιλιών της συζήτησης 1ης Μαρτίου 1994 σε JSON	47

Κατάλογος Πινάκων

4.1	Πεδία της οντότητας Debate	36
4.2	Πεδία της οντότητας Speech	39
4.3	Πεδία της οντότητας Speaker	41
4.4	Λίστα θεματικών ενοτήτων στις οποίες κατηγοριοποιούνται οι κοινοβουλευτικές συζητήσεις του συστήματος	42
4.5	Πεδία της οντότητας Topic	42
4.6	Πεδία της οντότητας Political Party	43
4.7	Συσχετίσεις μεταξύ των οντοτήτων του συστήματος	44
5.1	Συνοπτική παρουσίαση των βασικών API endpoints της εφαρμογής.	51
5.2	Παράμετροι αιτήματος στο endpoint /api/search-debates	52
5.3	Παράμετροι αιτήματος στο endpoint /api/search-in-debate	53
5.4	Παράμετροι αιτήματος στο endpoint /api/debates/metadata/[id]	54
5.5	Παράμετροι αιτήματος στο endpoint /api/search-speakers	55
5.6	Παράμετροι αιτήματος στο endpoint /api/search-speaker-speeches	56
5.7	Παράμετροι αιτήματος στο endpoint /api/speakers/metadata/[id]	57
5.8	Ενδεικτικές παράμετροι URL για δημόσιες κλήσεις στο API του συστήματος διαχείρισης περιεχομένου.	58
6.1	Ερωτηματολόγιο Πριν τη Δοκιμή: Ερωτήσεις και Επιλογές Απαντήσεων	61
6.2	Σενάρια Χρήσης: Περιγραφή και Κριτήρια Επιτυχίας	62
6.3	Ερωτήσεις Κλίμακας Χρηστικότητας Συστήματος (SUS)	63
6.4	Ερωτήσεις Μοντέλου Αποδοχής Τεχνολογίας (TAM) ανά Διάσταση	63
6.5	Οδηγός Συνέντευξης Μετά τη Δοκιμή: Ανοιχτές Ερωτήσεις ανά Κατηγορία . .	64

Εισαγωγή

1.1 Περιγραφή Προβλήματος

Η δημοσιοποίηση και μελέτη του κοινοβουλευτικού λόγου αποτελεί θεμέλιο της δημοκρατικής διακυβέρνησης, της πολιτικής ανάλυσης και της επιστημονικής έρευνας στις κοινωνικές επιστήμες. Ωστόσο, το διατιθέμενο, από το κοινοβούλιο στο κοινό, υλικό παρουσιάζει σημαντικά προβλήματα προσβασιμότητας, διαλειτουργικότητας και δυνατότητας επεξεργασίας. Η απουσία ενιαίας, ανοιχτής και επεξεργάσιμης μορφής καθιστά δύσκολη την αναζήτηση, ανάλυση και αξιοποίηση των κοινοβουλευτικών συζητήσεων από ερευνητές, δημοσιογράφους και πολίτες· μια τάση που έχει ήδη αναγνωριστεί και διεθνώς, με αρκετά κοινοβούλια να αναζητούν σύγχρονες ψηφιακές λύσεις για τη μετάβασή τους στο διαδίκτυο [1].

Η ανάγκη για διαφανή, τεκμηριωμένη και σημασιολογικά εμπλουτισμένη αναπαράσταση του κοινοβουλευτικού λόγου έχει αναδειχθεί σε πληθώρα πρόσφατων μελετών. Η πρωτοβουλία Parla-CLARIN [2], για παράδειγμα, προτείνει ένα σύνολο προτύπων για την ενιαία κωδικοποίηση κοινοβουλευτικών σωμάτων λόγου, ακριβώς για την αποφυγή ανομοιομορφίας και την ενίσχυση της αναζητησιμότητας και της επαναχρησιμοποίησης του περιεχομένου. Παρόμοια ζητήματα έχουν καταγραφεί και στο πλαίσιο άλλων εθνικών κοινοβουλίων. Στην περίπτωση της Αυστραλίας [3], η επεξεργασία και θεματική αξιοποίηση των πρακτικών παρέμεινε δύσκολη, παρότι είχαν ψηφιοποιηθεί επί δύο δεκαετίες. Αντίστοιχα, υπογραμμίζεται η ανάγκη ενσωμάτωσης σημασιολογικών μοντέλων και ανοικτών δεδομένων για να καταστεί ο κοινοβουλευτικός λόγος ουσιαστικά προσβάσιμος [4].

Όσον αφορά την Ελλάδα, η κατάσταση είναι ακόμη πιο περιοριστική. Το υλικό είναι συχνά διαθέσιμο αποκλειστικά σε αρχεία μορφής PDF ή DOCX, χωρίς ενσωματωμένα μεταδεδομένα, μηχανισμό αναζήτησης ή δυνατότητα θεματικής ομαδοποίησης. Η έλλειψη ενιαίου τρόπου αναπαράστασης του περιεχομένου και η απουσία ανοικτής υποδομής δυσχεραίνουν κάθε προσπάθεια συστηματικής μελέτης του κοινοβουλευτικού λόγου, είτε από την επιστημονική κοινότητα είτε από την κοινωνία των πολιτών. Το πρόβλημα αυτό αναδεικνύεται και σε διεθνές επίπεδο: σύμφωνα με πρόσφατα ευρήματα [5], μόλις το 24% των εθνικών κοινοβουλίων παρέχει πλήρως δομημένα δεδομένα των πρακτικών τους, ενώ λιγότερα από τα μισά προσφέρουν δυνατότητες σημασιολογικής αναζήτησης.

1.2 Αντικείμενο Διπλωματικής

Η παρούσα διπλωματική εργασία επιδιώκει να ανταποκριθεί στις προκλήσεις αυτές, προτείνοντας ένα ενιαίο, επεκτάσιμο και σημασιολογικά εμπλουτισμένο πληροφοριακό σύστημα για την οργάνωση, αναπαράσταση και αξιοποίηση των κοινοβουλευτικών συζητήσεων. Το αντικείμενο της εργασίας αφορά στη δημιουργία μιας ολοκληρωμένης ηλεκτρονικής υπηρεσίας που καθιστά τις κοινοβουλευτικές συζητήσεις εύκολα προσβάσιμες, αναζητήσιμες και αξιοποιήσιμες από ερευνητές, πολίτες και θεσμικούς φορείς. Η υπηρεσία αυτή επιτυγχάνεται μέσω του μετασχηματισμού του κοινοβουλευτικού λόγου σε δομημένη πληροφορία, εμπλουτισμένη με σημασιολογικά μεταδεδομένα. Παράλληλα, παρέχεται δημόσιο API μέσω του οποίου επιτρέπεται η προγραμματιστική πρόσβαση στα δεδομένα, με δυνατότητες φιλτραρίσματος και αναζήτησης περιεχομένου.

Κεντρικός στόχος της προσέγγισης είναι η μετατροπή του περιεχομένου των συνεδριάσεων από στατική μορφή σε σημασιολογικά οργανωμένη πληροφορία, ενισχυμένη με στοιχεία όπως η ταυτότητα του ομιλητή, το θεσμικό και χρονικό πλαίσιο και η θεματική ταξινόμηση. Μέσω αυτής της υποδομής, προσφέρονται βελτιωμένες δυνατότητες αναζήτησης, πλοήγησης και διάθεσης των δεδομένων, συμβάλλοντας ουσιαστικά στην ενίσχυση της διαφάνειας και της συμμετοχής στον δημόσιο διάλογο. Η αξία τέτοιων συστημάτων έχει ήδη επιβεβαιωθεί από διεθνείς προσεγγίσεις που αξιοποιούν τεχνολογίες του Semantic Web για τον εμπλουτισμό και τη διασύνδεση πρακτικών κοινοβουλίων, διευκολύνοντας την αναζήτηση, ανάλυση και κατανόηση του πολιτικού λόγου [6]. Παράλληλα, έχουν προταθεί ευέλικτες υποδομές για την ανάλυση και τη διαλειτουργικότητα δεδομένων κοινοβουλευτικών συνεδριάσεων, βασισμένες σε ανοικτά πρότυπα και μοντέλα δεδομένων, με στόχο την ενίσχυση της διαφάνειας και της προσβασιμότητας των κοινοβουλευτικών αρχείων [7].

Η παρούσα εργασία τοποθετείται σε αυτό το πλαίσιο και φιλοδοξεί να αποτελέσει ένα λειτουργικό πρότυπο για την ελληνική περίπτωση, καλύπτοντας το τεχνολογικό, επιστημονικό και θεσμικό έλλειμμα στη διαχείριση του κοινοβουλευτικού λόγου και την παροχή σχετικών υπηρεσιών στο ευρύ κοινό.

1.3 Οργάνωση Κειμένου

Η παρούσα διπλωματική εργασία οργανώνεται σε επτά κεφάλαια. Στο Κεφάλαιο 1 παρουσιάζεται το πρόβλημα που αποτελεί το σημείο αφετηρίας της εργασίας. Στο Κεφάλαιο 2 αναπτύσσεται το θεωρητικό πλαίσιο της εργασίας, με έμφαση σε έννοιες που αφορούν τον Semantic Web, τη σημασιολογική διασύνδεση δεδομένων και την επεξεργασία φυσικής γλώσσας. Στο Κεφάλαιο 3 περιγράφονται οι απαιτήσεις του συστήματος, τόσο ως προς τις λειτουργίες που καλείται να υποστηρίξει όσο και ως προς την αρχιτεκτονική του. Στο Κεφάλαιο 4 παρουσιάζεται ο σχεδιασμός του συστήματος και η δομή της βάσης δεδομένων. Στο Κεφάλαιο 5 περιγράφεται η υλοποίηση των επιμέρους υποσυστημάτων και η διασύνδεση των δεδομένων μέσω API. Στο Κεφάλαιο 6 παρουσιάζεται η μεθοδολογία αξιολόγησης του συστήματος, καθώς και τα βασικά συμπεράσματα που προέκυψαν από τη δοκιμή και αξιολόγησή του από τελικούς χρήστες. Η εργασία ολοκληρώνεται με το Κεφάλαιο 7, στο οποίο διατυπώνονται τα συνολικά συμπεράσματα της ερευνητικής μας προσπάθειας και προτείνονται ενδεικτικές

κατευθύνσεις για μελλοντική έρευνα και εξέλιξη του συστήματος.

Οι υπολογιστικοί πόροι που χρησιμοποιήθηκαν στο πλαίσιο της παρούσας εργασίας παραχωρήθηκαν με την υποστήριξη του Εθνικού Δικτύου Υποδομών Τεχνολογίας και Έρευνας – ΕΔΥΤΕ (GRNET).

Θεωρητικό Υπόβαθρο και Σχετικές Εργασίες

Το παρόν κεφάλαιο παρουσιάζει το θεωρητικό υπόβαθρο και παραδείγματα σχετικών εργασιών που υποστηρίζουν την ανάπτυξη ενός συστήματος επεξεργασίας και ανάλυσης κοινοβουλευτικών δεδομένων. Ειδικότερα, αναλύονται έννοιες και προσεγγίσεις που σχετίζονται με τη σημασιολογική αναπαράσταση, τη διασύνδεση και την αξιοποίηση νομοθετικού και πολιτικού περιεχομένου, καθώς και οι βασικές αρχές που διέπουν τη δομημένη παρουσίαση δημόσιων εγγράφων. Παράλληλα, παρουσιάζονται τεχνικές ανάκτησης πληροφορίας και ανάλυσης φυσικής γλώσσας, οι οποίες επιτρέπουν την κατανόηση, την αναζήτηση και τη σύνθεση πληροφοριών από μεγάλες συλλογές κοινοβουλευτικού λόγου. Στο δεύτερο μέρος του κεφαλαίου, παρατίθενται και σχολιάζονται ενδεικτικές ερευνητικές εργασίες και εφαρμογές που αξιοποιούν συναφείς τεχνολογίες, ώστε να αναδειχθούν πρακτικά παραδείγματα και να υποστηριχθεί το πλαίσιο υλοποίησης της παρούσας εργασίας.

2.1 Θεωρητικό υπόβαθρο

2.1.1 Σημασιολογικός Ιστός (Semantic Web)

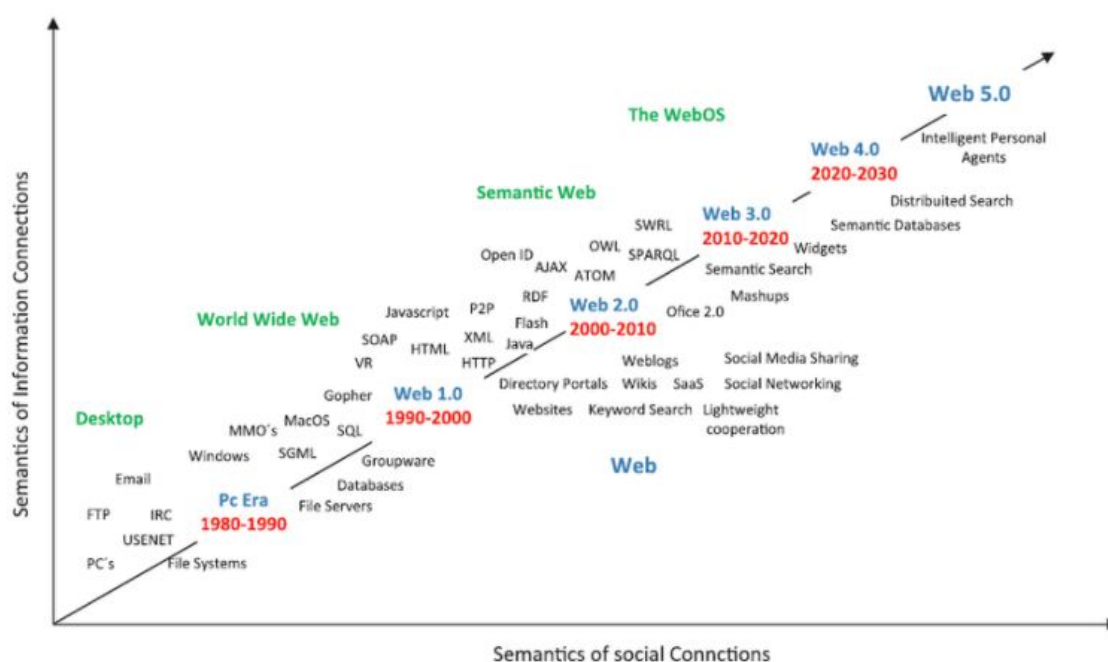
Η ανάπτυξη του διαδικτύου έχει φέρει επανάσταση στον τρόπο με τον οποίο επικοινωνούμε, έχουμε πρόσβαση σε πληροφορίες και διεξάγουμε επιχειρηματικές δραστηριότητες. Από τα πρώτα βήματά του, ως ένα απλό δίκτυο διασυνδεδεμένων εγγράφων, το διαδίκτυο έχει υποστεί αξιοσημείωτες αλλαγές με την πάροδο των ετών. Μια χαρακτηριστική εξέλιξη αποτελεί η μετάβαση στον Σημασιολογικό Ιστό [8], που επέτρεψε την τυποποιημένη αναπαράσταση γνώσης, την ολοκλήρωση ετερογενών δεδομένων και την υποστήριξη ευφών αυτοματοποιημένων διεργασιών.

Στα πρώτα στάδια του ιστού, οι πληροφορίες παρουσιάζονταν κυρίως σε αδόμητες μορφές, γεγονός που καθιστούσε δύσκολη έως και αδύνατη την αποτελεσματική επεξεργασία και ερμηνεία του περιεχομένου από τις μηχανές. Ωστόσο, καθώς ο ιστός μεγάλωνε σε μέγεθος και πολυπλοκότητα, προέκυψε η ανάγκη για έναν πιο έξυπνο και αποτελεσματικό τρόπο οργάνωσης και κατανόησης του τεράστιου όγκου των διαθέσιμων πληροφοριών. Έτσι γεννιέται η έννοια του σημασιολογικού ιστού, η οποία αποσκοπούσε στο να προσδώσει στο περιεχόμενο του ιστού σαφές νόημα και να επιτρέψει στις μηχανές να κατανοήσουν τα δεδομένα.

Η μετάβαση στον σημασιολογικό ιστό φέρνει πολλές υποσχέσεις για διάφορους τομείς. Ενδεικτικά, στον τομέα της ανάκτησης πληροφοριών και των μηχανών αναζήτησης, ο σημασιολογικός ιστός μπορεί να βελτιώσει την ακρίβεια και τη συνάφεια των αποτελεσμάτων

αναζήτησης, κατανοώντας το νόημα και το πλαίσιο πίσω από τα ερωτήματα των χρηστών. Σε έναν άλλο τομέα, όπως αυτού του ηλεκτρονικού εμπορίου, ο σημασιολογικός ιστός επιτρέπει πιο έξυπνες συστάσεις προϊόντων και εξατομικευμένες εμπειρίες αγορών με βάση τη βαθύτερη κατανόηση των προτιμήσεων και των αναγκών των πελατών. Τέλος, στην υγειονομική περίθαλψη, ο σημασιολογικός ιστός διευκολύνει τη διαλειτουργικότητα και την διασύνδεση των ιατρικών δεδομένων, οδηγώντας σε βελτιωμένη περίθαλψη των ασθενών, ερευνητική συνεργασία και λήψη αποφάσεων βάσει δεδομένων.

Επιπλέον, ο σημασιολογικός ιστός έχει θετικές επιδράσεις στις εφαρμογές τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML). Παρέχοντας ένα τυποποιημένο πλαίσιο για την αναπαράσταση δεδομένων και την ολοκλήρωση της γνώσης, ο σημασιολογικός ιστός μπορεί να βελτιώσει την εκπαίδευση και την απόδοση των μοντέλων τεχνητής νοημοσύνης, επιτρέποντάς τους να κάνουν πιο τεκμηριωμένες προβλέψεις και συστάσεις. Ανοίγει επίσης το δρόμο για την ανάπτυξη ευφύων πρακτόρων και chatbots που μπορούν να κατανοούν και να απαντούν σε ερωτήματα χρηστών με πιο φυσικό και συνειδητό τρόπο.



Εικόνα 2.1: Η εξέλιξη του διαδικτύου προς τον Σημασιολογικό Ιστό και τον Web 5.0, με έμφαση στη σημασιολογία των πληροφοριών και των κοινωνικών διασυνδέσεων.

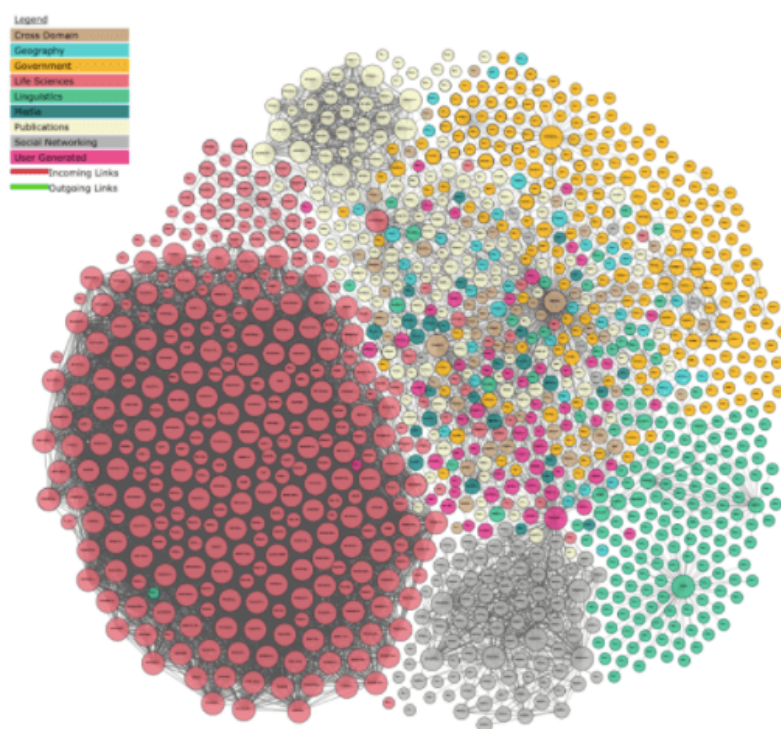
Κατά συνέπεια, όπως περιγράφεται στην Εικόνα 3.1, η μετάβαση προς τον Σημασιολογικό Ιστό ή τον Ιστό 3.0 είναι ήδη σε εξέλιξη. Η προβλεπόμενη μετάβαση στο Web 4.0, (ή Ευφυής Ιστός), είναι ακόμα σε μεταβατικό στάδιο, όπου αναμένεται να αξιοποιηθεί η τεχνητή νοημοσύνη και η μηχανική μάθηση για να παρέχεται στον χρήστη προσαρμοσμένες υπηρεσίες και περιεχόμενο.

2.1.2 Ανοιχτά Διασυνδεδεμένα Δεδομένα (Linked Open Data)

Η ιδέα των Ανοιχτών Συνδεδεμένων Δεδομένων [9] αποτελεί θεμελιώδη συνιστώσα του Σημασιολογικού Ιστού, προάγοντας τη διαλειτουργικότητα και την επαναχρησιμοποίηση δεδομένων μέσω της τυποποιημένης και ανοικτής δημοσίευσης και σύνδεσης δομημένης πληροφορίας. Για τη μηχανικά αναγνώσιμη αναπαράσταση της πληροφορίας, χρησιμοποιείται το πρότυπο RDF (Resource Description Framework) [10], το οποίο υιοθετεί την τριπλή δομή υποκειμένου-κατηγορουμένου-αντικειμένου και διευκολύνει την ενοποίηση ετερογενών πηγών δεδομένων.

Η αναζήτηση αυτών των δεδομένων πραγματοποιείται μέσω της γλώσσας SPARQL (SPARQL Protocol and RDF Query Language) [11], η οποία επιτρέπει τη διατύπωση πολύπλοκων ερωτημάτων και την επιστροφή δομημένων αποτελεσμάτων από αποθηκευμένες RDF οντότητες. Η χρήση παγκοσμίως μοναδικών URIs διασφαλίζει την ακριβή ταυτοποίηση και διασύνδεση εννοιών και πόρων σε διαφορετικά σύνολα δεδομένων.

Το γράφημα LOD Cloud στην Εικόνα 2.2 αποτυπώνει το παγκόσμιο οικοσύστημα των συνδεδεμένων συνόλων δεδομένων, όπου κάθε κόμβος αντιπροσωπεύει ένα σύνολο και κάθε σύνδεσμος δηλώνει την υφιστάμενη σημασιολογική σχέση μεταξύ τους. Το μοντέλο αυτό αναδεικνύει τις δυνατότητες πλοήγησης, συνδυαστικής ανάλυσης και εμπλουτισμού πληροφορίας που προσφέρει η διασύνδεση των δεδομένων.

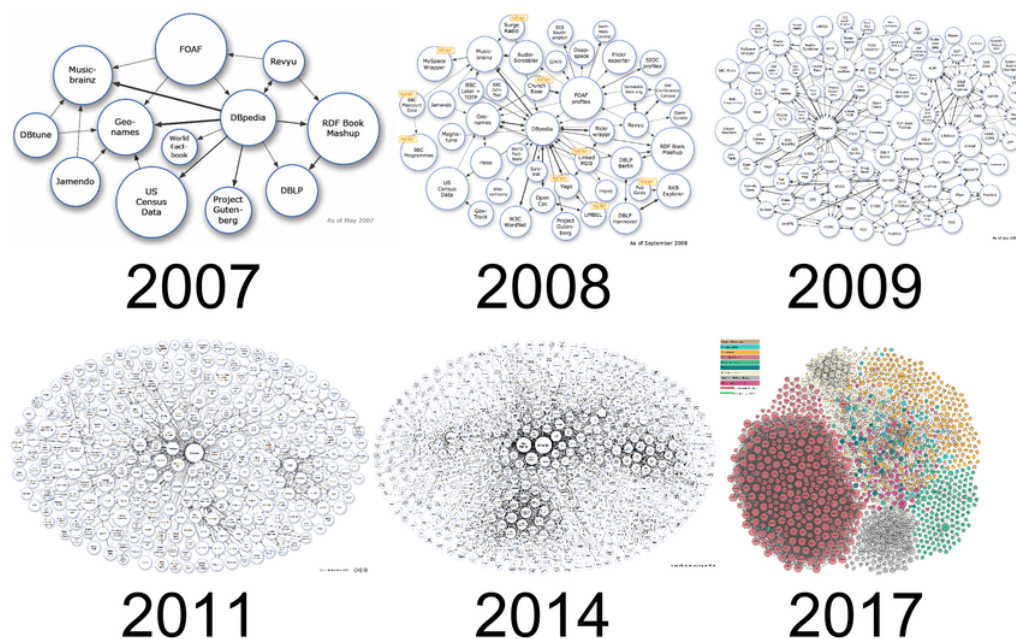


Εικόνα 2.2: Απεικόνιση του οικοσυστήματος Ανοιχτών Συνδεδεμένων Δεδομένων με κόμβους ως σύνολα δεδομένων και συνδέσμους ως σημασιολογικές σχέσεις.

Το Wikidata¹, ως μία από τις σημαντικότερες βάσεις γνώσης του διαδικτύου, παρέχει δομημένα δεδομένα μέσω RDF και υποστηρίζει ερωτήματα SPARQL για την εξαγωγή πληροφοριών με βάση σημασιολογικές σχέσεις. Στην παρούσα εργασία, τα δεδομένα που ανα-

κτώνται από το Wikidata χρησιμοποιούνται είτε για αποθήκευση στο σύστημα διαχείρισης περιεχομένου, είτε για δυναμική ανάκτηση κατά τη φόρτωση της σελίδας κάθε ομιλητή.

Η Εικόνα 2.3 απεικονίζει τη ραγδαία ανάπτυξη των Συνδεδεμένων Ανοικτών Δεδομένων (LOD) από το 2007 έως σήμερα, καταγράφοντας την αύξηση των συνόλων δεδομένων που δημοσιεύονται σύμφωνα με τις αρχές των LOD. Η εξέλιξη αυτή αντικατοπτρίζει την ευρεία υιοθέτηση των διασυνδεδεμένων δεδομένων από φορείς δημοσίου, επιχειρήσεις και οργανισμούς, διευκολύνοντας την ενσωμάτωση, ανακάλυψη και αξιοποίηση πληροφορίας στο διαδίκτυο.



Εικόνα 2.3: Η εξέλιξη του Νέφους Συνδεδεμένων Ανοικτών Δεδομένων (2007-2017), με έμφαση στην αύξηση του αριθμού και της διασύνδεσης συνόλων δεδομένων

2.1.3 Τεχνολογίες Δομημένης Αναπαράστασης Δεδομένων

Στον τομέα του Σημασιολογικού Ιστού και των Συνδεδεμένων Δεδομένων περιλαμβάνονται διάφορες τεχνολογίες για την αποτελεσματική αναπαράσταση, διασύνδεση και διαλειτουργικότητα των δεδομένων. Οι τεχνολογίες αυτές προσφέρουν το πλαίσιο για την οργάνωση, τη σύνδεση και τη συλλογή γνώσης από διάφορα σύνολα δεδομένων. Μια από τις πρώτες και ευρέως χρησιμοποιούμενες γλώσσες για την κωδικοποίηση και ανταλλαγή δομημένων δεδομένων είναι η XML (eXtensible Markup Language)², η οποία επιτρέπει την αναπαράσταση δεδομένων σε ιεραρχική μορφή με τη χρήση ετικετών.

Παρόλο που η XML παραμένει ιδιαίτερα χρήσιμη σε πολλούς τομείς, η ανάγκη για μια ελαφρύτερη και πιο φιλική στη χρήση μορφή οδήγησε στη διάδοση της JSON (JavaScript

¹ Πηγή: https://www.wikidata.org/wiki/Wikidata:Main_Page

Πηγή Εικόνας 3.2: <https://lod-cloud.net/>

² XML: <https://en.wikipedia.org/wiki/XML>

³ JSON: <https://www.json.org/>

Πηγή Εικόνας 3.3: https://www.researchgate.net/figure/Growth-of-Linked-Open-Data-since-2007-1-The-amount-of-data-sets-published-as-LOD-have_fig2_331748480

Object Notation)³. Η JSON είναι μια απλή και ευανάγνωστη μορφή αναπαράστασης δεδομένων, που βασίζεται σε ζεύγη κλειδιών-τιμών και φωλιασμένες δομές, επιτρέποντας την εύκολη διαχείριση και ανταλλαγή δεδομένων σε σύγχρονες διαδικτυακές εφαρμογές.

Στο πλαίσιο της παρούσας υπηρεσίας, τα κοινοβουλευτικά δεδομένα μετασχηματίζονται από μορφή XML σε JSON, ώστε να αναπαρίστανται αποδοτικά οι ομιλίες και τα συναφή μεταδεδομένα. Η πληροφορία αποθηκεύεται σε σύστημα διαχείρισης περιεχομένου και ευρετηριάζεται με τρόπο που επιτρέπει την αποδοτική αναζήτηση και ανάκτηση.

Extensible Markup Language (XML)

Η XML είναι μια ευέλικτη γλώσσα σήμανσης που χρησιμοποιείται ευρέως για τη δόμηση και ανταλλαγή δεδομένων στο διαδίκτυο. Αναπτύχθηκε από την Κοινοπραξία του Παγκόσμιου Ιστού (W3C)⁴ το 1996 ως μια προσαρμόσιμη λύση που επιτρέπει τη δημιουργία εξειδικευμένων γλωσσών για ποικίλες εφαρμογές και πλατφόρμες.

Η σύνταξη της βασίζεται σε ιεραρχικά διατεταγμένες ετικέτες εντός αγκυλών (< >), καθιστώντας την κατανοητή τόσο από ανθρώπους όσο και από μηχανές. Κάθε έγγραφο XML έχει μία ριζική ετικέτα που περικλείει όλες τις υπόλοιπες, οι οποίες με τη σειρά τους μπορεί να είναι φωλιασμένες για την αναπαράσταση πολύπλοκων δομών.

Η κύρια καινοτομία της XML έγκειται στην απλή και αυστηρά ορισμένη σύνταξή της. Τα στοιχεία περιέχονται σε αγκύλες (<) / (>) και διατάσσονται ιεραρχικά χρησιμοποιώντας μια μέθοδο βασισμένη σε ετικέτες. Τόσο οι άνθρωποι όσο και οι μηχανές μπορούν να ερμηνεύσουν τα δεδομένα, δεδομένου ότι αυτές οι ετικέτες καθορίζουν τη δομή και το νόημά τους.

```

1. <person>
2.   <name> Apostolos Kaklamanis </name>
3.   <age> 88 </age>
4.   <address>
5.     <street> Main Street </street>
6.     <city> Athens </city>
7.     <country> Greece </country>
8.   </address>
9. </person>

```

Εικόνα 2.4: Παράδειγμα δομής εγγράφου σε XML για την αναπαράσταση στοιχείων ενός πολιτικού προσώπου

Στο παραπάνω παράδειγμα, οι ετικέτες XML αναπαριστούν δομημένα χαρακτηριστικά ενός ατόμου, με τη ριζική ετικέτα "person" να περιλαμβάνει φωλιασμένες πληροφορίες όπως όνομα, ηλικία και διεύθυνση. Η χρήση αυτής της ιεραρχικής δομής καθιστά την πληροφορία ευανάγνωστη και εύκολα επεξεργάσιμη.

Στο πλαίσιο των συνδεδεμένων δεδομένων, η XML επιτρέπει τη μετατροπή μη δομημένων κειμένων σε μορφές αναγνώσιμες από μηχανήματα, διευκολύνοντας την περαιτέρω επεξερ-

⁴ XML W3C: <https://www.w3.org/XML/>

γασία, διασύνδεση και αναζήτηση μέσω τεχνολογιών του Σημαιολογικού Ιστού. Παράλληλα, χάρη στη γενική και καθολική σύνταξή της, αποτελεί εργαλείο ενοποίησης ετερογενών πηγών δεδομένων σε ενιαία αναπαράσταση.

JavaScript Object Notation (JSON)

Η JavaScript Object Notation (JSON) είναι μια ευρέως χρησιμοποιούμενη μορφή αναπαράστασης δεδομένων, που αναπτύχθηκε στις αρχές της δεκαετίας του 2000 ως ελαφριά εναλλακτική της XML. Χρησιμοποιεί απλή σύνταξη βασισμένη σε ζεύγη κλειδιού-τιμής (key-value pairs), όπου κάθε τιμή μπορεί να είναι απλός τύπος (αριθμός, λογική τιμή, συμβολοσειρά) ή σύνθετη δομή όπως πίνακας [] ή αντικείμενο {}.

Η ευανάγνωστη σύνταξη και η υποστήριξή της από δημοφιλείς γλώσσες προγραμματισμού (π.χ. JavaScript, Python) καθιστούν τη JSON κατάλληλη για αποθήκευση και μεταφορά δεδομένων σε σύγχρονες διαδικτυακές εφαρμογές.

```

1. {
2.   "name": "Apostolos Kaklamanis",
3.   "age": 88,
4.   "address": {
5.     "street": "Main Street",
6.     "city": "Athens",
7.     "country": "Greece"
8.   }
9. }
```

Εικόνα 2.5: Παράδειγμα δομής εγγράφου σε JSON για την αναπαράσταση στοιχείων ενός πολιτικού προσώπου

Η JSON, όπως και η XML, διαδραματίζει σημαντικό ρόλο στην ενοποίηση δεδομένων από ετερογενείς πηγές και την ανταλλαγή πληροφοριών μεταξύ διαφορετικών συστημάτων. Ιδιαίτερα στον Σημαιολογικό Ιστό και στα Συνδεδεμένα Δεδομένα (Linked Data), η JSON συνδυάζεται συχνά με το JSON-LD (JSON for Linked Data), το οποίο επεκτείνει τη βασική μορφή JSON ώστε να υποστηρίζει σημασιολογικές συνδέσεις μεταξύ των δεδομένων.

Η ευκολία χρήσης, η αποδοτικότητα και η διαλειτουργικότητα της JSON την έχουν καταστήσει τη δεδομένη επιλογή για την πλειονότητα των εφαρμογών Ιστού και των API, ενώ παραμένει εξίσου χρήσιμη σε εφαρμογές κινητών συσκευών, IoT και μικροϋπηρεσίες.

2.1.4 Ανοιχτά Κυβερνητικά Δεδομένα

Τα διασυνδεδεμένα ανοιχτά κυβερνητικά δεδομένα (LOGD) [12], είναι μια «ιδέα» που προωθεί την δημοσιοποίηση, τη διαλειτουργικότητα και την ενσωμάτωση των κυβερνητικών δεδομένων στο διαδίκτυο. Εφαρμόζει τις αρχές των Συνδεδεμένων Ανοικτών Δεδομένων στο πεδίο της κυβέρνησης, αποσκοπώντας να μεγιστοποιήσει το πλέγμα των ανοικτών δεδομένων και να προωθήσει την καινοτομία, τη συνεργασία και τη διαφάνεια.

Τα LOGD διατίθενται ως RDF τριπλέτες με χρήση μοναδικών αναγνωριστικών (URIs), διευκολύνοντας τη σύνδεση και συνδυαστική ανάλυση με άλλα σύνολα δεδομένων. Η προσέγγιση αυτή ενισχύει την πρόσβαση των πολιτών και των φορέων σε πληροφορία, προάγοντας τη λογοδοσία και τη συμμετοχή στις δημοκρατικές διαδικασίες.

Επιπλέον, τα ανοικτά κυβερνητικά δεδομένα υποστηρίζουν τη συνεργασία μεταξύ διαφορετικών τομέων, καθώς και τη δημιουργία νέας γνώσης μέσω της διασύνδεσης ετερογενών πηγών. Κυβερνήσεις διεθνώς δημοσιεύουν ποικίλα σύνολα δεδομένων (π.χ. πληθυσμιακά, οικονομικά, νομοθετικά), τα οποία διατίθενται μέσω εξειδικευμένων πυλών, API και SPARQL endpoints.

Ένα βασικό πεδίο εφαρμογής των LOGD είναι η τυποποιημένη αναπαράσταση νομικών και κοινοβουλευτικών εγγράφων, με το LegalDocML να αποτελεί το κυρίαρχο πρότυπο. Η μορφή αυτή χρησιμοποιήθηκε και στην παρούσα εργασία ως αρχικό format εισόδου, πριν τα δεδομένα μετατραπούν σε JSON για περαιτέρω επεξεργασία και αξιοποίηση.

LegalDocML

Το LegalDocML [13], γνωστό παλαιότερα ως Akoma Ntoso, αποτελεί διεθνές πρότυπο βασισμένο σε XML, το οποίο αναπτύσσεται από τον οργανισμό Oasis Open για την αναπαράσταση και διαχείριση κοινοβουλευτικών, νομοθετικών και δικαστικών εγγράφων. Υποστηρίζει τη διαλειτουργικότητα, την προσβασιμότητα και τη σημασιολογικά εμπλουτισμένη αναπαράσταση των εγγράφων, εναρμονιζόμενο με τις αρχές των Συνδεδεμένων Δεδομένων.

Το πρότυπο αυτό ορίζει πλήθος σημασιολογικά δομημένων στοιχείων όπως επικεφαλίδες, άρθρα, παραγράφους και παραπομπές, καθιστώντας δυνατή την ενιαία και ομοιόμορφη αναπαράσταση νομικών εγγράφων. Η χρήση του διευκολύνει την επεξεργασία, την ανάλυση και την επαναχρησιμοποίησή τους από πληροφοριακά συστήματα.

Η ένταξή του στο πλαίσιο των Ανοικτών Συνδεδεμένων Δεδομένων ενισχύει τη δυνατότητα διασύνδεσης με συναφή δεδομένα και πηγές — όπως μεταδεδομένα, δικαστικές αποφάσεις και νομοθετικές βάσεις — δημιουργώντας ολοκληρωμένα γραφήματα γνώσης. Έτσι, επιτρέπεται η προηγμένη αναζήτηση, η σημασιολογική επεξεργασία και η αυτοματοποιημένη αξιοποίηση νομοθετικού περιεχομένου.

Συνεπώς, κυβερνήσεις, δικαστήρια και άλλοι οργανισμοί σε όλο τον κόσμο έχουν υιοθετήσει ευρέως το πρότυπο αυτό για την παρουσίαση νομοθετικών κειμένων. Η χρήση του διασφαλίζει ότι οι κυβερνητικές πληροφορίες είναι τυποποιημένες και διαλειτουργικές, προωθώντας τη διεθνή συνεργασία, τη συγκριτική νομοθετική έρευνα και τη δημιουργία κυβερνητικών οντολογιών και βάσεων γνώσης.

2.1.5 Τεχνικές Ανάκτησης Πληροφορίας

Η Ανάκτηση Πληροφορίας (Information Retrieval - IR) [14] αποτελεί τον επιστημονικό κλάδο που ασχολείται με την αναπαράσταση, αποθήκευση, οργάνωση και ανάκτηση πληροφορίας από μεγάλες συλλογές μη-δομημένων ή ημι-δομημένων δεδομένων, όπως κείμενα, πολυμεσικά αρχεία ή ακόμη και σύνθετα έγγραφα, όπως τα κοινοβουλευτικά πρακτικά που εξετάζει η παρούσα διπλωματική εργασία.

Ο κεντρικός στόχος ενός συστήματος ανάκτησης πληροφορίας είναι να εντοπίσει και να επιστρέψει εκείνα τα έγγραφα από μία συλλογή, τα οποία ικανοποιούν καλύτερα τις πληροφοριακές ανάγκες του χρήστη, όπως αυτές εκφράζονται μέσω ενός ερωτήματος. Το σύστημα επιχειρεί να υπολογίσει τη συνάφεια (relevance) κάθε εγγράφου σε σχέση με το ερώτημα, χρησιμοποιώντας διάφορα μαθηματικά μοντέλα και μετρικές.

Για την υλοποίηση αυτής της λειτουργικότητας έχουν προταθεί διάφορα υπολογιστικά μοντέλα, τα οποία επιτρέπουν την αναπαράσταση εγγράφων και ερωτημάτων, καθώς και τον υπολογισμό της μεταξύ τους συνάφειας. Μεταξύ των πιο διαδεδομένων είναι το Boolean Model, το Probabilistic Model και το Vector Space Model (VSM). Το VSM, ειδικότερα, επιτρέπει την απεικόνιση εγγράφων και ερωτημάτων ως διανύσματα σε έναν πολυδιάστατο χώρο, διευκολύνοντας τον υπολογισμό της μεταξύ τους ομοιότητας με μετρικές όπως το συνημίτονο της γωνίας. Τέτοιες προσεγγίσεις είναι ιδιαίτερα χρήσιμες όταν εξετάζονται μεγάλα κείμενα, όπως κοινοβουλευτικά πρακτικά, στα οποία απαιτείται αποδοτική αναζήτηση και κατάταξη αποτελεσμάτων βάσει σημασιολογικής συνάφειας.

Vector Space Model

Η παρούσα εργασία αξιοποιεί το Μοντέλο Διανυσματικού Χώρου (Vector Space Model – VSM) [15] για την αναπαράσταση και ευρετηρίαση των κοινοβουλευτικών συζητήσεων. Στο VSM, κάθε συζήτηση εκφράζεται ως διάνυσμα όρων, όπου κάθε διάσταση αντιστοιχεί σε λέξη και η τιμή της αποτυπώνει τη σημασία της στο κείμενο.

Η εκτίμηση της σημασίας γίνεται με τη μέτρηση tf-idf (Term Frequency – Inverse Document Frequency), η οποία ενισχύει όρους που εμφανίζονται συχνά σε συγκεκριμένα κείμενα αλλά όχι στο σύνολο του σώματος. Με αυτόν τον τρόπο, λέξεις-κλειδιά με ιδιαίτερη θεματική αξία (π.χ., ‘προϋπολογισμός’, ‘νομοσχέδιο’, ‘περιβάλλον’) αποκτούν υψηλότερη βαρύτητα σε σχέση με κοινές λέξεις όπως ‘και’, ‘ο’, ‘της’.

Η συνημιτονοειδής ομοιότητα (cosine similarity) χρησιμοποιείται για τη σύγκριση ερωτημάτων με συζητήσεις, υπολογίζοντας τη γωνία μεταξύ των αντίστοιχων διανυσμάτων: όσο μικρότερη η γωνία, τόσο μεγαλύτερη η συνάφεια.

Η αποδοτική αναζήτηση επιτυγχάνεται μέσω ενός αντίστροφου ευρετηρίου (inverted index), που καταγράφει για κάθε όρο τα σχετικά έγγραφα, μειώνοντας το κόστος αναζήτησης και βελτιώνοντας την κλιμάκωση του συστήματος.

Για την υποστήριξη αυτής της μεθοδολογίας σε πραγματικές συνθήκες, απαιτείται μια ισχυρή και επεκτάσιμη υποδομή αναζήτησης που να επιτρέπει γρήγορη ευρετηρίαση, αποτελεσματική διαχείριση μεγάλου όγκου εγγράφων και σημασιολογική αναζήτηση. Στην παρούσα εργασία, αυτό επιτυγχάνεται μέσω της αξιοποίησης του Elasticsearch⁵, το οποίο καλύπτει λειτουργικά και απόδοτικά τις παραπάνω ανάγκες.

Elasticsearch

Η παρούσα εργασία υλοποιεί την υποδομή ευρετηρίασης και αναζήτησης μέσω του Elasticsearch⁶, το οποίο επιτρέπει την αποδοτική οργάνωση, αποθήκευση και αναζήτηση εγγράφων βάσει όρων. Η πλατφόρμα βασίζεται σε αρχές του Μοντέλου Διανυσματικού Χώρου

και εφαρμόζει τεχνικές επεξεργασίας φυσικής γλώσσας για να βελτιώσει τη σημασιολογική συνάφεια των αποτελεσμάτων.

Η διαδικασία ευρετηρίασης προϋποθέτει την προκαταρκτική επεξεργασία του κειμένου, κατά την οποία αφαιρούνται λέξεις χωρίς πληροφοριακή αξία (stopwords) και οι υπόλοιπες κανονικοποιούνται μέσω τεχνικών όπως η στέλεχωση (stemming) ή η λεμματοποίηση (lemmatization). Έτσι, λέξεις με παρόμοια σημασία μπορούν να αναγνωριστούν ως σχετικές, ανεξαρτήτως γραμματικής μορφής.

Η ανάκτηση των εγγράφων στηρίζεται στην έννοια της σημασιολογικής συνάφειας, με τα αποτελέσματα να κατατάσσονται βάσει της ομοιότητας μεταξύ των ερωτημάτων και των κειμένων, όπως αυτή προκύπτει από τη διανυσματική τους αναπαράσταση. Το σύστημα επιτρέπει επίσης σύνθετα ερωτήματα που συνδυάζουν περιεχόμενο και μεταδεδομένα (όπως ο ομιλητής ή η ημερομηνία συζήτησης), εξυπηρετώντας πιο στοχευμένες πληροφοριακές ανάγκες.

Η αρχιτεκτονική αυτή επιτρέπει την επεκτάσιμη διαχείριση μεγάλου όγκου κοινοβουλευτικών δεδομένων, παρέχοντας ταχεία και αξιόπιστη αναζήτηση σε σχεδόν πραγματικό χρόνο. Αυτό επιτυγχάνεται μέσω της χρήσης ανεστραμμένων ευρετηρίων (inverted indexes), τα οποία μειώνουν δραστικά το πλήθος των συγκρίσεων κατά την αναζήτηση, ενισχύοντας την αποδοτικότητα του συστήματος.

2.1.6 Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP)

Η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) [16] αποτελεί έναν ταχέως αναπτυσσόμενο κλάδο της τεχνητής νοημοσύνης, ο οποίος επικεντρώνεται στη μελέτη και ανάπτυξη αλγορίθμων και τεχνικών που επιτρέπουν στους υπολογιστές να κατανοούν, να επεξεργάζονται και να παράγουν κείμενο και λόγο σε φυσική γλώσσα.

Στο πλαίσιο της ανάλυσης κοινοβουλευτικών πρακτικών, οι τεχνικές NLP διαδραματίζουν κρίσιμο ρόλο, καθώς παρέχουν τα εργαλεία για την εξαγωγή χρήσιμης πληροφορίας από εκτενή κείμενα, την αυτόματη κατηγοριοποίηση ομιλιών, την ανάλυση συναισθήματος και τη σημασιολογική χαρτογράφηση των θεματικών ενοτήτων που καλύπτονται στις συζητήσεις.

Ανάλυση Συναισθήματος

Η ανάλυση συναισθήματος (Sentiment Analysis) [17] είναι η διαδικασία προσδιορισμού της συναισθηματικής φόρτισης (polarity) ενός κειμένου, δηλαδή αν εκφράζει θετικό, αρνητικό ή ουδέτερο συναίσθημα. Ειδικά στις κοινοβουλευτικές συζητήσεις, η ανάλυση συναισθήματος προσφέρει πολύτιμες πληροφορίες σχετικά με τον τόνο και τη στάση των ομιλητών απέναντι σε συγκεκριμένα θέματα, νομοσχέδια ή πολιτικές αποφάσεις.

Για την εφαρμογή ανάλυσης συναισθήματος, μπορούν να αξιοποιηθούν τόσο λεξικογραφικές προσεγγίσεις (lexicon-based) που βασίζονται σε προ-υπάρχουσες λίστες θετικά και αρνητικά φορτισμένων λέξεων, όσο και μέθοδοι μηχανικής μάθησης (machine learning-

⁵ Elasticsearch Guide: <https://medium.com/@abhijeetkumar7565/elasticsearch-simplified-guide-to-quick-data-retrieval-e6608750de04>

⁶ Elasticsearch Documentation: <https://www.elastic.co/guide/en/elasticsearch/reference/current/index.html>

based), όπου ταξινομητές εκπαιδεύονται σε ετικετοποιημένα δεδομένα κοινοβουλευτικών συζητήσεων.

Σημασιολογική Ανάλυση και Εξαγωγή Θεμάτων (Topic Modelling)

Η σημασιολογική ανάλυση και η εξαγωγή θεμάτων (Topic Modelling) [18] αποτελούν κρίσιμες τεχνικές για την ανίχνευση και κατανόηση των βασικών θεματικών εννοιών που αναπτύσσονται στις κοινοβουλευτικές συζητήσεις. Μέσω αυτών, τα πολυδιάστατα κείμενα των πρακτικών αποκτούν εννοιολογική δομή, επιτρέποντας την ομαδοποίηση ομιλιών και τη χαρτογράφηση των βασικών θεμάτων που απασχολούν το κοινοβούλιο.

Ένα από τα πιο διαδεδομένα μοντέλα εξαγωγής θεμάτων είναι το Latent Dirichlet Allocation (LDA), το οποίο υποθέτει ότι κάθε έγγραφο αποτελείται από ένα μείγμα θεμάτων, και κάθε θέμα αποτελείται από ένα σύνολο όρων με συγκεκριμένες πιθανότητες εμφάνισης. Η εφαρμογή του LDA ή άλλων σύγχρονων αλγορίθμων (π.χ., BERTopic) επιτρέπει τη χαρτογράφηση των κοινοβουλευτικών πρακτικών, προσφέροντας μια συνοπτική και κατανοητή εικόνα των θεματικών τάσεων.

Σύνοψη Κειμένου (Text Summarization)

Η σύνοψη κειμένου (Text Summarization) [19] στοχεύει στη δημιουργία συντομότερων εκδοχών των κοινοβουλευτικών πρακτικών, διατηρώντας ταυτόχρονα τις πιο σημαντικές πληροφορίες και έννοιες. Ειδικά για τους πολίτες ή τους ερευνητές, οι οποίοι χρειάζονται γρήγορη επισκόπηση μεγάλων συζητήσεων, η τεχνική αυτή παρέχει μια αποτελεσματική λύση. Διακρίνεται σε:

- **Extractive Summarization:** επιλογή των πιο σημαντικών προτάσεων απευθείας από το κείμενο.
- **Abstractive Summarization:** δημιουργία νέων προτάσεων που συνοψίζουν το περιεχόμενο.

Οι σύγχρονες τεχνικές σύνοψης αξιοποιούν μοντέλα βαθιάς μάθησης (Deep Learning), όπως τα Transformer-based models (π.χ., BART, T5), τα οποία έχουν επιδείξει εξαιρετικές επιδόσεις στη σύνοψη μεγάλων πολιτικών κειμένων.

2.2 Σχετικές Εργασίες και Παραδείγματα

Η μελέτη του κοινοβουλευτικού λόγου γνωρίζει αυξανόμενο ερευνητικό ενδιαφέρον, χάρη στη διαθεσιμότητα μεγάλων ποσοτήτων δεδομένων και την πρόοδο της υπολογιστικής γλωσσολογίας. Πέρα από την απλή ανάγνωση ή κατηγοριοποίηση, αξιοποιούνται σύγχρονες τεχνικές που επιτρέπουν την εξαγωγή θεματικής, συναισθηματικής, επιχειρηματολογικής και πολυτροπικής πληροφορίας. Στην ενότητα αυτή παρουσιάζονται ενδεικτικές ερευνητικές προσπάθειες και εργασίες που αξιοποιούν τεχνητή νοημοσύνη, τεχνολογίες σημασιολογικού ιστού και εργαλεία οπτικοποίησης για την ανάλυση του κοινοβουλευτικού περιεχομένου.

2.2.1 Δομημένη Αναπαράσταση Κοινοβουλευτικών Κειμένων

Στην ενότητα αυτή παρουσιάζονται διεθνείς και ελληνικές ερευνητικές πρωτοβουλίες και έργα που επικεντρώνονται στην τυποποιημένη και σημασιολογικά εμπλουτισμένη αναπαράσταση κοινοβουλευτικών δεδομένων, προτείνοντας ενοποιημένα σχήματα επισήμανσης και πρότυπα κατάλληλα για πολιτική και γλωσσική ανάλυση.

Διεθνείς Πρωτοβουλίες

Σε διεθνές επίπεδο, το έργο ParlaMint II: Advancing Comparable Parliamentary Corpora Across Europe [20], στοχεύει στη συγκρότηση συγκρίσιμων, πολυγλωσσικών κοινοβουλευτικών σωμάτων κειμένων για περισσότερες από είκοσι ευρωπαϊκές χώρες. Η προσέγγιση αυτή βασίζεται στην υιοθέτηση ενοποιημένων σχημάτων μεταδεδομένων και τυποποιημένης σήμανσης βάσει του προτύπου TEI (Text Encoding Initiative), επιτυγχάνοντας υψηλό βαθμό διαλειτουργικότητας. Το σύνολο δεδομένων συνοδεύεται από αναλυτική τεκμηρίωση, εργαλεία μετατροπής και παραδείγματα χρήσης, διευκολύνοντας την αξιοποίησή του από διαφορετικές ερευνητικές κοινότητες.

Επιπλέον, η μετατροπή των πρακτικών του Ευρωπαϊκού Κοινοβουλίου [21] δημιούργησε ένα δομημένο και σημασιολογικά εμπλουτισμένο σώμα, το οποίο επιτρέπει σύνθετα ερωτήματα και αναλύσεις μέσω αναπαραστάσεων RDF και καθιερωμένων οντολογιών για τη μοντελοποίηση κοινοβουλευτικών οντοτήτων και σχέσεων. Ακόμη, έρευνα για το Κοινοβούλιο του Ηνωμένου Βασιλείου [22] τεκμηρίωσε προσπάθειες ανάπτυξης σωμάτων λόγου κοινοβουλευτικών ομιλιών με επισημάνσεις συναισθήματος, δείχνοντας πώς τα δομημένα δεδομένα επιτρέπουν προηγμένες υπολογιστικές αναλύσεις, ενώ ανέδειξε προκλήσεις ως προς τη σημασιολογική συνέπεια σε μεγάλες συλλογές δεδομένων.

Στην περίπτωση της Ολλανδίας, το σώμα DutchParl [23] υπογράμμισε τη σημασία της τυποποιημένης δομής εγγράφων, πρωτοστατώντας στα οφέλη της γραμματικής ανάλυσης για νομοθετικά έγγραφα. Παράλληλα, επιμέρους κοινοβούλια, όπως εκείνο της Ισπανίας [24] ή αυτό του Καναδά [25], έχουν αναπτύξει εξειδικευμένες ψηφιακές βιβλιοθήκες που υποστηρίζουν σύνθετα ερωτήματα πάνω σε οντολογικά εμπλουτισμένες συλλογές, αναδεικνύοντας τις δυνατότητες δομημένης αποθήκευσης και ανάκτησης.

Ελληνικές Πρωτοβουλίες

Σε εθνικό επίπεδο, η διπλωματική εργασία “Οι Συζητήσεις του Ελληνικού Κοινοβουλίου ως Ανοιχτά Διασυνδεδεμένα Δεδομένα” [26] προτείνει μια μεθοδολογία μετατροπής των πρακτικών της Βουλής των Ελλήνων σε LOD. Τα πρακτικά, τα οποία διατίθεντο έως πρόσφατα αποκλειστικά σε μορφές όπως PDF ή DOC, μετατρέπονται αρχικά σε αρχεία XML και στη συνέχεια σε σημασιολογικά σχήματα όπως RDFS, LegalDocML και TEI ParlaMint. Η προσέγγιση αυτή επιτρέπει την οντολογική αναπαράσταση εννοιών, σχέσεων και μεταδεδομένων, ενισχύοντας τη δυνατότητα σημασιολογικής αναζήτησης και την εξαγωγή ουσιαστικής πληροφορίας, προσαρμοσμένης στις ιδιαιτερότητες των ελληνικών κοινοβουλευτικών συζητήσεων.

Παλαιότερη ελληνική έρευνα [27] δημιούργησε ένα εκτενές σύνολο δεδομένων επικεντρωμένο στη γλωσσολογική και πολιτική ανάλυση, αποτελώντας σημαντικό βήμα προς την

προσβασιμότητα των κοινοβουλευτικών δεδομένων. Ωστόσο, το έργο αυτό στερούνταν διεθνών προτύπων ή τεχνικών σημασιολογικού εμπλουτισμού και, όπως συνέβη και με επόμενες εργασίες, τα δεδομένα δεν δημοσιεύθηκαν.

Μια άλλη σημαντική έρευνα [28] επεξεργάστηκε και ανέλυσε περισσότερες από ένα εκατομμύριο ομιλίες της Βουλής των Ελλήνων (1989–2015), μετατρέποντάς τις σε δομημένη, μηχανικά αναγνώσιμη μορφή, με δυνατότητα για δυναμική θεματική μοντελοποίηση. Η πλατφόρμα επέτρεψε πλοήγηση κατά θέματα και χρονικές περιόδους, προσφέροντας πληροφορίες για συνεισφορές βουλευτών, πολιτικών κομμάτων και περιφερειών. Ωστόσο, όπως και σε προηγούμενες εργασίες, ούτε το μοντέλο ούτε τα δεδομένα δημοσιεύθηκαν.

Τέλος, η επίσημη ιστοσελίδα της Βουλής των Ελλήνων, παρέχει μόνο βασική πρόσβαση στα κοινοβουλευτικά πρακτικά. Όπως επισημαίνεται και στη μελέτη [29], στερείται προηγμένου δομικού σημασιολογικού εμπλουτισμού και τυποποιημένης σήμανσης. Οι ελλείψεις αυτές αποτέλεσαν κίνητρο για την ανάπτυξη μιας πιο ολοκληρωμένης προσέγγισης στη διαχείριση των ελληνικών κοινοβουλευτικών δεδομένων, επεκτείνοντας προηγούμενες προσπάθειες στη δομημένη αναπαράσταση και ανάλυση.

2.2.2 Μεθοδολογικές Προσεγγίσεις και Τεχνολογικές Εφαρμογές στην Ανάλυση του Κοινοβουλευτικού Λόγου

Στην ενότητα αυτή παρουσιάζονται σύγχρονες εφαρμογές και μεθοδολογίες που αξιοποιούν τεχνικές ανάλυσης για την εξαγωγή θεματικής, συναισθηματικής, επιχειρηματολογικής και πολυτροπικής πληροφορίας από τον κοινοβουλευτικό λόγο.

Θεματική Μοντελοποίηση και Πολιτική Ατζέντα Η πολιτική ατζέντα του Ευρωπαϊκού Κοινοβουλίου έχει μελετηθεί σε βάθος δεκαπέντε ετών μέσω Dynamic Topic Modeling πάνω σε περισσότερες από 150.000 παρεμβάσεις ευρωβουλευτών [30]. Η τεχνική αυτή επιτρέπει την ανίχνευση θεματικών τάσεων στο χρόνο, συνδυάζοντας ποσοτική και ποιοτική αποτύπωση της εξέλιξης του πολιτικού λόγου, αναδεικνύοντας τη δυναμική φύση της κοινοβουλευτικής δραστηριότητας.

Συναισθηματική Ανάλυση και Στάση Ομιλητών Ιδιαίτερη έμφαση δίνεται στη μελέτη της συναισθηματικής διάστασης και της στάσης των ομιλητών μέσω βιβλιογραφικών επισκοπήσεων που συγκεντρώνουν περισσότερες από 70 σχετικές μελέτες [31]. Οι προσεγγίσεις αυτές εξετάζουν τις προκλήσεις της συναισθηματικής ανάλυσης σε πολιτικό λόγο, όπου συχνά κυριαρχούν ειρωνεία, θεσμική γλώσσα και έμμεσες αναφορές, υπογραμμίζοντας την ανάγκη για εξειδικευμένα εργαλεία.

Επιχειρηματολογία και Ρητορική Δομή Η εξόρυξη επιχειρημάτων (argument mining) στο κοινοβουλευτικό πλαίσιο έχει εφαρμοστεί σε πολυτροπικά σύνολα δεδομένων που συνδυάζουν κείμενα πρακτικών και ηχητικό υλικό [32], αλλά και σε πραγματικά - ακατέργαστα - πρακτικά με στόχο την εξαγωγή απόψεων και επιχειρημάτων [33]. Οι προσεγγίσεις αυτές αναδεικνύουν τις γλωσσικές προκλήσεις και την ανάγκη για ακριβείς δομές επισήμανσης.

Ανάλυση Κοινωνικών Δικτύων Η ανάλυση κοινωνικών σχέσεων με μοντέλα κοινωνικών δικτύων βοηθά να κατανοήσουμε πώς αλληλεπιδρούν οι ευρωβουλευτές [34]. Χρησιμοποιώντας τη σειρά των παρεμβάσεων και τη θεματική τους συνάφεια, η μέθοδος αυτή συμβάλλει στην κατανόηση της πολιτικής συμπεριφοράς και της επικοινωνίας τους.

Γραφήματα Γνώσης και Σημασιολογική Αναπαράσταση Η αναπαράσταση των δεδομένων με γραφήματα γνώσης (knowledge graphs) είναι ένα σημαντικό εργαλείο για την ανάλυσή τους. Σε αυτό το πλαίσιο, εφαρμόστηκαν τεχνικές για να εντοπιστούν οντότητες και σχέσεις σε πάνω από 900.000 ομιλίες του φινλανδικού κοινοβουλίου [35], κάνοντας πιο εύκολη τη μελέτη πολιτισμικών και γλωσσικών χαρακτηριστικών.

Πολυτροπική Ανάλυση Η πολυτροπική ανάλυση (multimodal analysis) συνδυάζει πληροφορίες από κείμενο, ήχο, εικόνα και μεταδεδομένα για μια πιο ολοκληρωμένη κατανόηση του πολιτικού λόγου [36, 37]. Στην Ελλάδα, ο κ. Γκούμας [28] έδειξε ότι, όταν συνδυαστεί η θεματική μοντελοποίηση με την πολυτροπική ανάλυση, μπορούμε να καταγράψουμε και να κατανοήσουμε την εξέλιξη της πολιτικής ατζέντας με μεγαλύτερη ακρίβεια στο πέρασμα του χρόνου.

Οπτικοποίηση και Προσβασιμότητα Τέλος, η οπτικοποίηση των κοινοβουλευτικών δεδομένων βοηθά ώστε η πολύπλοκη πληροφορία να γίνεται κατανοητή και από μη ειδικούς χρήστες. Έχουν δημιουργηθεί διαδραστικά συστήματα που συνδυάζουν ανάλυση συναισθήματος και θεμάτων [38], προσφέροντας εύχρηστα εργαλεία για εξερεύνηση των δεδομένων.

2.2.3 Επιπτώσεις για τη Μελλοντική Έρευνα

Οι τάσεις αυτές δείχνουν ότι τα μελλοντικά συστήματα επεξεργασίας κοινοβουλευτικών δεδομένων θα ενσωματώνουν σημασιολογική ανάλυση ενισχυμένη από LLMs, μεθόδους μηχανικής μάθησης για διαμοιρασμό γνώσης με διατήρηση της ιδιωτικότητας, πολυτροπική ενσωμάτωση δεδομένων και αυξημένη τυποποίηση για διευκόλυνση δια-κοινοβουλευτικών και πολυγλωσσικών αναλύσεων.

Η παρούσα εργασία ευθυγραμμίζεται με αυτές τις τάσεις, συνδυάζοντας:

- πρακτική ενσωμάτωση LLMs για ανάλυση συναισθήματος, ταξινόμηση θεμάτων και σύνοψη,
- συνδυασμό αυτοματισμού και ανθρώπινης εποπτείας για την αποφυγή προκαταλήψεων,
- αξιοποίηση του προτύπου LegalDocML για διαλειτουργικότητα,
- επεκτάσιμη αρχιτεκτονική για την ενσωμάτωση πολυτροπικών προσεγγίσεων.

Οι σχεδιαστικές αυτές επιλογές τοποθετούν την εργασία στο σημείο συνάντησης εφαρμοσμένης τεχνολογίας και ερευνητικών καινοτομιών, καλύπτοντας άμεσες ανάγκες ενώ δημιουργούν τις βάσεις για μελλοντικές δυνατότητες.

Σχεδιασμός Συστήματος

Στο παρόν κεφάλαιο περιγράφονται αναλυτικά οι απαιτήσεις του πληροφοριακού συστήματος που αναπτύχθηκε με στόχο την αναζήτηση, παρουσίαση και ανάλυση των πρακτικών της Βουλής των Ελλήνων. Η ανάλυση εστιάζει τόσο στις βασικές λειτουργικότητες της εφαρμογής όσο και στις τεχνικές και λειτουργικές απαιτήσεις που καλύπτουν τις ανάγκες των τελικών χρηστών.

3.1 Λειτουργικές Απαιτήσεις

Η παρούσα ενότητα περιγράφει αναλυτικά τις λειτουργικές απαιτήσεις του συστήματος και τις βασικές λειτουργικότητες που υλοποιούνται για την κάλυψη των αναγκών των τελικών χρηστών. Οι απαιτήσεις αυτές προκύπτουν από την ανάγκη για συστηματική εισαγωγή, επεξεργασία, αναζήτηση και παρουσίαση των πρακτικών της Ελληνικής Βουλής, καθώς και για τον εμπλουτισμό τους με μεταδεδομένα και ανάλυση περιεχομένου. Κάθε λειτουργικότητα που περιγράφεται στη συνέχεια αντιστοιχεί σε επιμέρους στόχους του συστήματος και υλοποιείται μέσω κατάλληλων τεχνικών για τη διαχείριση περιεχομένου, την αναζήτηση, την ανάλυση φυσικής γλώσσας και την άντληση πληροφορίας από εξωτερικές πηγές.

3.1.1 Εισαγωγή, Εμπλουτισμός και Διαχείριση Δεδομένων

Το σύστημα επιτρέπει την αυτοματοποιημένη εισαγωγή κοινοβουλευτικών πρακτικών από αρχεία σε μορφή LegalDocML XML, τα οποία μετατρέπονται σε δομημένη μορφή κατάλληλη για αποθήκευση και περαιτέρω επεξεργασία. Κατά την εισαγωγή, το περιεχόμενο εμπλουτίζεται με επιπλέον πληροφορίες, που αντλούνται δυναμικά από εξωτερικές πηγές. Όλα τα δεδομένα αποθηκεύονται σε σύστημα που επιτρέπει τη διαχείρισή τους από διαχειριστές μέσω φιλικού γραφικού περιβάλλοντος.

3.1.2 Ανάλυση Περιεχομένου με Φυσική Γλώσσα (NLP)

Το σύστημα ενσωματώνει τεχνικές επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP) για την αυτοματοποιημένη ανάλυση του περιεχομένου των κοινοβουλευτικών πρακτικών. Μετά την εισαγωγή των πρακτικών στο σύστημα, εκτελούνται διαδικασίες για την παραγωγή μιας συνοπτικής περίληψης και την ανάλυση του συναισθηματικού ύφους των επιμέρους ομιλιών.

Στη συνέχεια, οι θεματικές ενότητες κάθε πρακτικού προσδιορίζονται με βάση την παραχθείσα περίληψη, ώστε να αποτυπώνονται τα βασικά ζητήματα που αναδείχθηκαν στη συζήτηση. Τα αποτελέσματα της ανάλυσης αξιοποιούνται επίσης για την ενίσχυση της λειτουργίας αναζήτησης, επιτρέποντας την εφαρμογή φίλτρων βάσει θεματικής ταξινόμησης, συναισθηματικών χαρακτηριστικών και ρητορικών προθέσεων.

3.1.3 Αναζήτηση Πλήρους Κειμένου

Το περιεχόμενο των κοινοβουλευτικών πρακτικών ευρετηριάζεται σε κατάλληλη μορφή για να υποστηρίξει αποτελεσματική αναζήτηση πλήρους κειμένου. Ο χρήστης μπορεί να εντοπίσει άμεσα συγκεκριμένους όρους ή φράσεις μέσα στις ομιλίες, με τα σχετικά αποσπάσματα να επισημαίνονται στη ροή του κειμένου για εύκολη ανάγνωση. Η δυνατότητα αυτή συμπληρώνει τις προηγμένες τεχνικές ανάλυσης περιεχομένου και διευκολύνει την ακριβή και εστιασμένη ανάκτηση πληροφορίας.

3.1.4 Αναζήτηση Πρακτικών

Η αναζήτηση πρακτικών αποτελεί έναν από τους δύο βασικούς πυλώνες του συστήματος. Ο χρήστης ξεκινά από τη σελίδα αναζήτησης πρακτικών, όπου μπορεί να εφαρμόσει φίλτρα όπως λέξεις-κλειδιά, θεματικές ενότητες, ημερομηνίες, συμμετέχοντες ομιλητές και άλλα κριτήρια. Το σύστημα επιστρέφει μία λίστα αποτελεσμάτων, με κάθε αποτέλεσμα να αντιστοιχεί σε ένα συγκεκριμένο πρακτικό συνεδρίασης.

Για κάθε πρακτικό που εμφανίζεται στα αποτελέσματα, ο χρήστης έχει τη δυνατότητα να επιλέξει αν θέλει να εξετάσει τα Μεταδεδομένα Συνεδρίασης ή το Πλήρες Περιεχόμενο Συνεδρίασης. Οι δύο αυτές όψεις δεν αποτελούν ξεχωριστές αναζητήσεις, αλλά εναλλακτικές προβολές των δεδομένων που σχετίζονται με το επιλεγμένο πρακτικό.

Μεταδεδομένα Συνεδρίασης

Στην προβολή μεταδεδομένων παρουσιάζονται συνοπτικά οι βασικές πληροφορίες για τη συνεδρίαση, όπως η ημερομηνία, η κοινοβουλευτική περίοδος και σύνοδος, η περίληψη της συζήτησης, οι συμμετέχοντες ομιλητές και οι θεματικές ενότητες που καλύφθηκαν. Αυτή η προβολή διευκολύνει τον χρήστη να αποκτήσει μία γρήγορη εικόνα της συνεδρίασης, χωρίς να χρειαστεί να διαβάσει το πλήρες περιεχόμενό της αναλυτικά.

Πλήρες Περιεχόμενο Συνεδρίασης

Στην προβολή πλήρους περιεχομένου εμφανίζεται ολόκληρο το κείμενο των ομιλιών της συνεδρίασης. Ο χρήστης μπορεί να εφαρμόσει πρόσθετα φίλτρα κατά την ανάγνωση, όπως φιλτράρισμα για ομιλίες με υψηλή συναισθηματική ένταση ή από συγκεκριμένους ομιλητές. Τα αποτελέσματα αυτά επισημαίνονται στη ροή του κειμένου, ώστε να είναι ευδιάκριτα κατά την ανάγνωση και να διευκολύνουν την εστιασμένη μελέτη των παρεμβάσεων.

Έτσι, η αναζήτηση πρακτικών επιτρέπει την εύκολη εύρεση συνεδριάσεων και την προβολή τους με τρόπο που προσαρμόζεται στις ανάγκες του χρήστη.

3.1.5 Αναζήτηση Ομιλητών

Η αναζήτηση ομιλητών αποτελεί τον δεύτερο βασικό πυλώνα του συστήματος. Ο χρήστης ξεκινά από τη σελίδα αναζήτησης ομιλητών, όπου μπορεί να εφαρμόσει φίλτρα όπως ονοματεπώνυμο, πολιτικό κόμμα, ηλικία, θεματικές ενότητες και συναισθηματικά χαρακτηριστικά του λόγου τους. Το σύστημα επιστρέφει μία λίστα αποτελεσμάτων, με κάθε αποτέλεσμα να αντιστοιχεί σε έναν συγκεκριμένο ομιλητή.

Για κάθε ομιλητή που εμφανίζεται στα αποτελέσματα, ο χρήστης έχει τη δυνατότητα να επιλέξει αν θέλει να εξετάσει τα Προσωπικά Στοιχεία του Ομιλητή ή τις Κοινοβουλευτικές Συμμετοχές του. Οι δύο αυτές όψεις δεν αποτελούν ξεχωριστές αναζητήσεις, αλλά εναλλακτικές προβολές των δεδομένων που σχετίζονται με τον επιλεγμένο ομιλητή.

Προσωπικά Στοιχεία Ομιλητή

Στην προβολή προσωπικών στοιχείων παρουσιάζονται πληροφορίες για τον ομιλητή, όπως βιογραφικά δεδομένα, πολιτική ιδιότητα, φωτογραφία, εκπαιδευτικό υπόβαθρο και σύνδεσμοι σε εξωτερικές πηγές. Αυτή η προβολή παρέχει μία συνοπτική εικόνα της ταυτότητας και της σταδιοδρομίας του ομιλητή.

Κοινοβουλευτικές Συμμετοχές Ομιλητή

Στην προβολή κοινοβουλευτικών συμμετοχών εμφανίζονται όλες οι συνεδριάσεις και οι ομιλίες στις οποίες έχει συμμετάσχει ο ομιλητής. Ο χρήστης μπορεί να εφαρμόσει φίλτρα για να περιορίσει τα αποτελέσματα σε συγκεκριμένα χρονικά διαστήματα, θεματικές ενότητες ή συναισθηματικά χαρακτηριστικά. Επίσης, υποστηρίζεται αναζήτηση πλήρους κειμένου στις ομιλίες του, με άμεση επισήμανση των σχετικών αποσπασμάτων.

Έτσι, η αναζήτηση ομιλητών διευκολύνει την εύρεση πληροφοριών για κάθε πολιτικό πρόσωπο και την προβολή τους με τρόπο που καλύπτει τις ανάγκες του χρήστη.

3.1.6 Προβολή Στατιστικών Στοιχείων

Το σύστημα παρέχει βασικά στατιστικά στοιχεία, όπως ο συνολικός αριθμός συμμετοχών ανά ομιλητή και ο αριθμός διαφορετικών ομιλητών ανά συνεδρίαση, ώστε να διευκολύνεται η επισκόπηση των δεδομένων. Οι πληροφορίες αυτές εμφανίζονται συγκεντρωτικά και βοηθούν τον χρήστη να κατανοήσει καλύτερα τη συμμετοχή και τη δραστηριότητα των πολιτικών προσώπων.

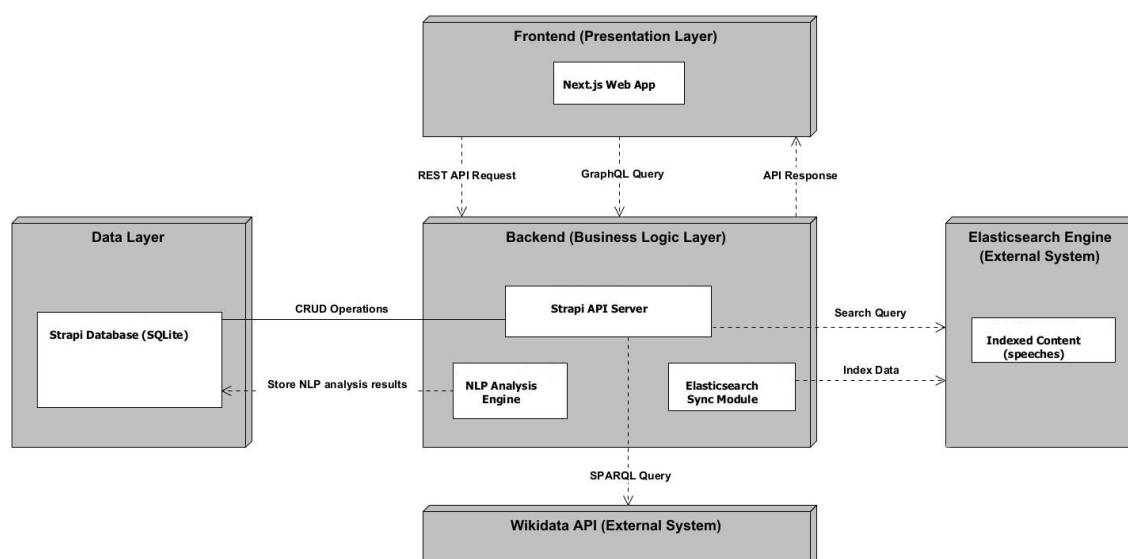
3.2 Αρχιτεκτονική Δομή

Οι λειτουργικές απαιτήσεις που παρουσιάστηκαν στην προηγούμενη ενότητα υλοποιούνται μέσα από μια αρχιτεκτονική που οργανώνει τις βασικές ζώνες λειτουργικότητας και τις ροές δεδομένων με σαφή και ξεκάθαρο τρόπο. Η αρχιτεκτονική του συστήματος ακολουθεί πολυεπίπεδη προσέγγιση (multi-tier architecture), διαχωρίζοντας τις αρμοδιότητες κάθε επιπέδου. Κάθε επίπεδο είναι υπεύθυνο για συγκεκριμένα καθήκοντα, διευκολύνοντας τη

συντήρηση, την επεκτασιμότητα και την επαναχρησιμοποίηση του κώδικα. Συγκεκριμένα, διακρίνονται τα εξής κύρια επίπεδα:

- **Επίπεδο Παρουσίασης (Frontend):** Παρέχει τη διεπαφή χρήστη μέσω διαδικτυακής εφαρμογής, επιτρέποντας την πλοήγηση, αναζήτηση και προβολή των κοινοβουλευτικών δεδομένων.
- **Επίπεδο Επιχειρησιακής Λογικής (Backend):** Υπεύθυνο για την επεξεργασία των αιτημάτων των χρηστών, τη σύνθεση των απαντήσεων, και τη διασύνδεση με τα υποσυστήματα δεδομένων και εξωτερικές πηγές πληροφόρησης.
- **Επίπεδο Δεδομένων (Data Layer):** Περιλαμβάνει τη βάση δεδομένων, όπου αποθηκεύονται τα δεδομένα του συστήματος.
- **Εξωτερικά Υποσυστήματα Αναζήτησης και Εμπλουτισμού:** Η αναζήτηση πραγματοποιείται μέσω ανεξάρτητου κόμβου ευρετηρίασης, ενώ η εμπλουτισμένη πληροφόρηση αντλείται δυναμικά από εξωτερικές πηγές μέσω διεπαφών προγραμματισμού (API).

Η συνολική αρχιτεκτονική του συστήματος απεικονίζεται στο Σχήμα 3.1, το οποίο παρουσιάζει τη δομική διάκριση των τριών βασικών επιπέδων (παρουσίασης, επιχειρησιακής λογικής και δεδομένων), καθώς και τις ροές δεδομένων και τις διασυνδέσεις με εξωτερικά υποσυστήματα, όπως η μηχανή αναζήτησης και η πηγή εμπλουτισμένων μεταδεδομένων.



Σχήμα 3.1: Η αρχιτεκτονική του συστήματος, με έμφαση στις ροές δεδομένων μεταξύ των επιπέδων και την αλληλεπίδραση με εξωτερικές υπηρεσίες.

Κεφάλαιο 4

Μοντέλο Δεδομένων Συστήματος - Βάση Δεδομένων

Το παρόν κεφάλαιο επικεντρώνεται στην αναλυτική περιγραφή του μοντέλου δεδομένων του συστήματος και της βάσης δεδομένων που υλοποιήθηκε, η οποία υποστηρίζει τη λειτουργικότητα της εφαρμογής, όπως αυτή περιγράφηκε στο προηγούμενο κεφάλαιο. Παρουσιάζονται οι βασικές οντότητες που αποτυπώνουν τα κοινοβουλευτικά πρακτικά, τα πεδία που τις συγκροτούν, καθώς και οι μεταξύ τους σχέσεις.

4.1 Εννοιολογική Επισκόπηση των Οντοτήτων

Παρακάτω παρουσιάζονται οι οντότητες που αντιπροσωπεύουν κεντρικές έννοιες του συστήματος και συμβάλλουν στην οργάνωση και ανάλυση των κοινοβουλευτικών δεδομένων.

- **Debate:** Αναπαριστά μια κοινοβουλευτική συζήτηση, η οποία περιλαμβάνει τις ομιλίες που πραγματοποιήθηκαν, τους συμμετέχοντες ομιλητές και τις θεματικές ενότητες που συζητήθηκαν, παρέχοντας παράλληλα βασικά μεταδεδομένα για το πλαίσιο διεξαγωγής και το περιεχόμενό της.
- **Speech:** Αναπαριστά μια κοινοβουλευτική ομιλία, η οποία αποδίδεται σε συγκεκριμένο ομιλητή, εντάσσεται σε μια συγκεκριμένη κοινοβουλευτική συζήτηση και αναλύεται ως προς το συναισθηματικό της φορτίο και τη ρητορική πρόθεση του ομιλητή.
- **Political Party:** Αναπαριστά ένα πολιτικό κόμμα, το οποίο συνδέεται με τους ομιλητές που έχουν υπάρξει μέλη του σε οποιαδήποτε χρονική περίοδο. Επιπλέον, περιλαμβάνει πληροφορίες όπως η επίσημη ονομασία και το έμβλημα του.
- **Speaker:** Αναπαριστά έναν ομιλητή, ο οποίος μπορεί να συμμετέχει σε πολλαπλές κοινοβουλευτικές συζητήσεις, είτε λαμβάνοντας τον λόγο είτε παρευρισκόμενος στη διαδικασία. Επιπλέον, συνδέεται με αναγνωριστικά δεδομένα, βιογραφικές πληροφορίες, καθώς και άλλα χαρακτηριστικά που αποτυπώνουν την πολιτική και επαγγελματική του ταυτότητα.
- **Topic:** Αναπαριστά τις θεματικές κατηγορίες στις οποίες ταξινομούνται οι κοινοβουλευτικές συζητήσεις και ομιλίες. Κάθε θεματική ενότητα αντανακλά ένα συγκεκριμένο

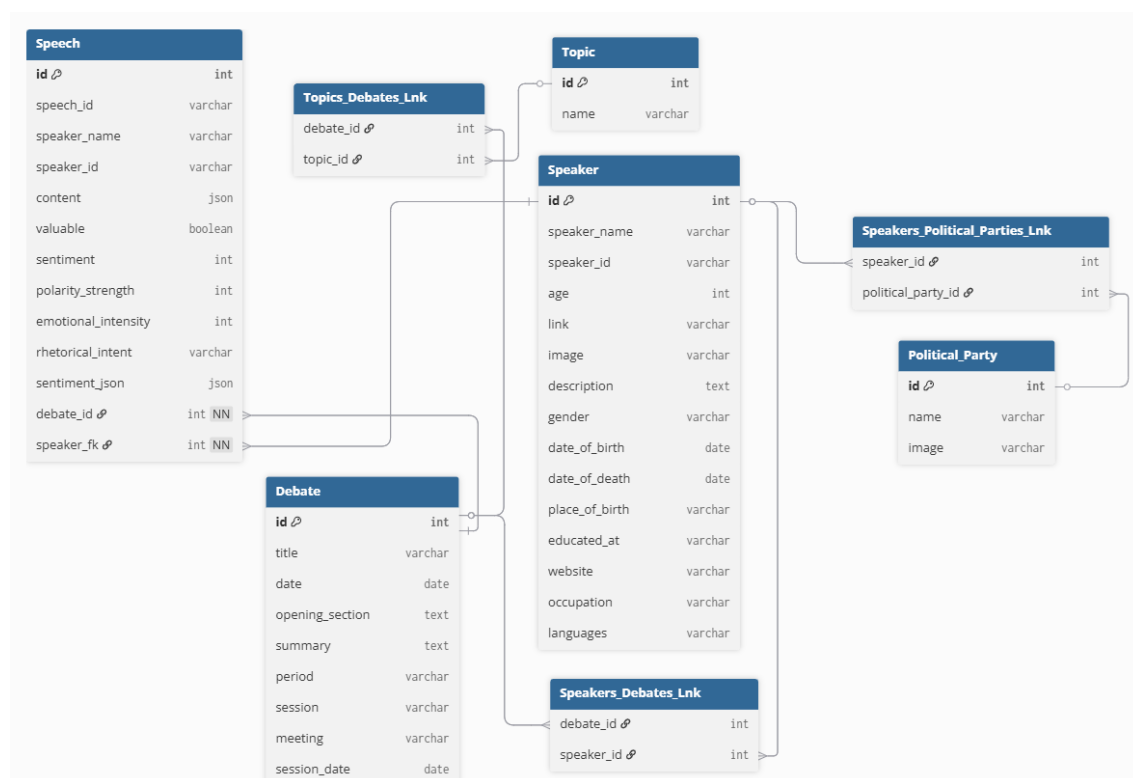
ζήτημα ή πεδίο πολιτικού ενδιαφέροντος, διευκολύνοντας την ανάλυση και την κατηγοριοποίηση του περιεχομένου των συζητήσεων.

4.2 Σχήμα της Βάσης Δεδομένων

Το σχήμα της βάσης δεδομένων αποτελείται από τις κύριες οντότητες του συστήματος, οι οποίες περιγράφουν τις βασικές συνιστώσες των κοινοβουλευτικών πρακτικών. Στις επόμενες ενότητες παρουσιάζονται τα πεδία και οι σχέσεις κάθε οντότητας.

4.2.1 Μοντέλο Οντοτήτων Συσχετίσεων

Στην παρούσα ενότητα παρουσιάζεται το διάγραμμα οντοτήτων-συσχετίσεων (Entity - Relationship Diagram) το οποίο απεικονίζει τη συνολική δομή των βασικών οντοτήτων και των μεταξύ τους σχέσεων, όπως αυτές έχουν διαμορφωθεί στο πλαίσιο του συστήματος.



Σχήμα 4.1: Αναπαράσταση οντοτήτων και συσχετίσεων του μοντέλου δεδομένων του συστήματος

4.2.2 Αναλυτική Περιγραφή Οντοτήτων

Σχήμα της οντότητας Debate

Η οντότητα Debate αναπαριστά μία κοινοβουλευτική συζήτηση, η οποία περιλαμβάνει τις βασικές πληροφορίες της συνεδρίασης, τις ομιλίες που πραγματοποιήθηκαν, καθώς και πρόσθετα μεταδεδομένα σχετικά με τη δομή και το περιεχόμενό της. Η οντότητα αυτή συνδέεται με άλλες οντότητες, όπως οι Speech, Speaker και Topic, προκειμένου να αναπαραστήσει

με ακρίβεια τη συνολική διαδικασία των κοινοβουλευτικών συνεδριάσεων.

Πεδία της Οντότητας Debate

Παρακάτω παρατίθενται και επεξηγούνται τα πεδία της οντότητας Debate.

- **id:** Το μοναδικό αναγνωριστικό της συζήτησης στη βάση δεδομένων (πρωτεύον κλειδί).
- **title:** Ο τίτλος της συζήτησης.

Ενδεικτική αναπαράσταση: ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 1997-02-03

- **date:** Η ημερομηνία διεξαγωγής της συζήτησης.

Ενδεικτική αναπαράσταση: 03/01/1999

- **opening_section:** Το εισαγωγικό τμήμα της κοινοβουλευτικής συζήτησης, το οποίο περιλαμβάνει την έναρξη της συνεδρίασης, τις πρώτες διατυπώσεις του προεδρεύοντος, καθώς και τυχόν διαδικαστικές αναφορές ή οργανωτικά ζητήματα που τίθενται πριν την κύρια συζήτηση.

Ενδεικτική αναπαράσταση: Αθήνα, σήμερα την 1η Μαρτίου 1999, ημέρα Δευτέρα και ώρα 18.10' συνήλθε στην Αίθουσα των συνεδριάσεων του Βουλευτηρίου η Βουλή σε ολομέλεια, για να συνεδριάσει υπό την προεδρία του Γ' Αντιπροέδρου αυτής κ. ΛΟΥΚΑ ΑΠΟΣΤΟΛΙΔΗ.

- **summary:** Περιεκτική και δομημένη ανασκόπηση των κυριότερων θεμάτων, βασικών επιχειρημάτων και σημαντικότερων συμπερασμάτων της κοινοβουλευτικής συζήτησης, με έκταση περίπου 150-200 λέξεων.

Ενδεικτική αναπαράσταση: Η κοινοβουλευτική συζήτηση επικεντρώθηκε σε κρίσιμα ζητήματα του αγροτικού τομέα και της απασχόλησης. Στην αγροτική παραγωγή, αναδείχθηκε το πρόβλημα της μειωμένης ελαιοπαραγωγής στην Κρήτη λόγω ακραίων καιρικών φαινομένων, με τον Υπουργό να προτείνει στήριξη μέσω ειδικών προγραμμάτων παρά την έλλειψη ασφαλιστικής κάλυψης από τον ΕΛΓΑ. Προτάθηκαν τροποποιήσεις στον κανονισμό ασφάλισης και ενίσχυσης των ΠΣΕΑ . . .

- **period:** Ο αριθμός της κοινοβουλευτικής περιόδου στην οποία εντάσσεται η συζήτηση. Η κοινοβουλευτική περίοδος αντιστοιχεί στη διάρκεια μιας νομοθετικής θητείας και αποτελεί το ανώτατο χρονικό πλαίσιο που περιλαμβάνει πολλαπλές συνόδους.

Ενδεικτική αναπαράσταση: Θ' ΠΕΡΙΟΔΟΣ (ΠΡΟΕΔΡΕΥΟΜΕΝΗΣ ΔΗΜΟΚΡΑΤΙΑΣ)

- **session:** Ο αριθμός της κοινοβουλευτικής συνόδου, η οποία αποτελεί επιμέρους χρονική διαίρεση της κοινοβουλευτικής περιόδου. Κάθε σύνοδος αντιστοιχεί συνήθως σε ένα κοινοβουλευτικό έτος και περιλαμβάνει τις συνεδριάσεις που πραγματοποιούνται εντός του διαστήματος αυτού.

Ενδεικτική αναπαράσταση: ΣΥΝΟΔΟΣ Γ'

- **meeting**: Ο αριθμός της συγκεκριμένης συνεδρίασης, η οποία αποτελεί μία μεμονωμένη σύγκληση του κοινοβουλίου στο πλαίσιο μιας συνόδου. Κατά τη διάρκεια της συνεδρίασης διεξάγονται συζητήσεις, ψηφοφορίες και νομοθετικό έργο.

Ενδεικτική αναπαράσταση: ΣΥΝΕΔΡΙΑΣΗ ΠΖ'

- **session_date**: Η ημερομηνία διεξαγωγής της συνεδρίασης στην οποία πραγματοποιήθηκε μέρος της συζήτησης. Σε περιπτώσεις όπου η συζήτηση εκτείνεται σε περισσότερες από μία ημέρες, η τιμή του πεδίου αυτού αντιστοιχεί στην ημερομηνία κάθε επιμέρους συνεδρίασης, σε αντίθεση με την ημερομηνία έναρξης της συζήτησης που αποθηκεύεται στο πεδίο date. Έτσι, το session_date αποτυπώνει με ακρίβεια τη χρονική κατανομή της συζήτησης σε διαφορετικές συνεδριάσεις.

Ενδεικτική αναπαράσταση: Δευτέρα 1 Μαρτίου 1999

Όνομα Πεδίου	Τύπος	Περιγραφή
id	Integer	Μοναδικό αναγνωριστικό της συζήτησης στη βάση δεδομένων.
title	String	Τίτλος της συζήτησης.
date	Date	Ημερομηνία διεξαγωγής της συζήτησης.
opening_section	String	Εισαγωγικό τμήμα της συνεδρίασης, με χαιρετισμούς, διαδικαστικά και οργανωτικά ζητήματα.
summary	Text	Δομημένη ανασκόπηση των θεμάτων, επιχειρημάτων και συμπερασμάτων της συζήτησης.
period	String	Κοινοβουλευτική περίοδος στην οποία εντάσσεται η συζήτηση.
session	String	Κοινοβουλευτική σύνοδος στο πλαίσιο της περιόδου.
meeting	String	Αριθμός της συγκεκριμένης συνεδρίασης εντός της συνόδου.
session_date	String	Ημερομηνία κάθε επιμέρους συνεδρίασης.

Πίνακας 4.1: Πεδία της οντότητας *Debate*

Σχήμα της οντότητας *Speech*

Η οντότητα *Speech* αναπαριστά μία ομιλία που πραγματοποιείται στο πλαίσιο μιας κοινοβουλευτικής συζήτησης. Περιλαμβάνει βασικές πληροφορίες σχετικά με το περιεχόμενο της ομιλίας, τον ομιλητή που την εκφώνησε, καθώς και πρόσθετα μεταδεδομένα που αφορούν την ανάλυση του συναισθήματος και τη ρητορική πρόθεση του ομιλητή. Η οντότητα αυτή συνδέεται με άλλες οντότητες, όπως οι *Debate*, *Speaker*, *Topic* και *Political Party*, προκειμένου να αποτυπώσει με ακρίβεια τη συμμετοχή των βουλευτών και την εξέλιξη της συζήτησης.

Πεδία της Οντότητας *Speech*

Παρακάτω παρατίθενται και επεξηγούνται τα πεδία της οντότητας *Speech*.

- **id**: Το μοναδικό αναγνωριστικό της συζήτησης στη βάση δεδομένων (πρωτεύον κλειδί).
- **speech_id**: Εναλλακτικός μοναδικός αναγνωριστικός κωδικός της ομιλίας, όπως αυτός αποδίδεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης.

Ενδεικτική αναπαράσταση: debate_2002-02-01_1_speech_103

- **speaker_name**: Το πλήρες ονοματεπώνυμο του ομιλητή, όπως αυτό καταγράφεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης..

Ενδεικτική αναπαράσταση: ΑΛΕΞΑΝΔΡΟΣ ΜΠΑΛΤΑΣ

- **speaker_id**: Εναλλακτικός μοναδικός αναγνωριστικός κωδικός του ομιλητή, όπως αυτός αποδίδεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης..

Ενδεικτική αναπαράσταση: alexandros_mpaltas

- **content**: Το πλήρες κείμενο της ομιλίας, όπως αυτή εκφωνήθηκε στο πλαίσιο της συζήτησης. Αποθηκεύεται σε δομή JSON, η οποία οργανώνει το περιεχόμενο σε λίστα παραγράφων, διατηρώντας τη μορφοποίηση και τη δομή του λόγου.

Ενδεικτική αναπαράσταση:

```

1.  {
2.    [
3.      "Εννοώ ευρωεκλογές, κύριε συνάδελφε. Είναι και αυτές εκλογές.",
4.      "Περιμέναμε, λοιπόν, και από σας μια θέση, μια τοποθέτηση..."
5.    ]
6.  }
7.  }
```

- **valuable**: Δείκτης (Boolean) που δηλώνει αν η ομιλία θεωρείται ουσιαστική για τη συζήτηση. Ομιλίες με λιγότερους από 150 χαρακτήρες θεωρούνται ότι δεν προσθέτουν χρήσιμο περιεχόμενο στη ροή της συζήτησης.

- **sentiment**: Η συνολική εκτίμηση του συναισθηματικού τόνου της ομιλίας ως θετικού, αρνητικού ή ουδέτερου.

Επιτρεπόμενες τιμές: -1 (Αρνητικό), 0 (Ουδέτερο), 1 (Θετικό).

- **polarity_strength**: Η ένταση της θετικής ή αρνητικής στάσης, δηλαδή πόσο καθαρά εκφράζεται η θέση υπέρ ή κατά μέσα στην ομιλία.

Επιτρεπόμενες τιμές: -1 (Πολύ αρνητικό), 0 (Ουδέτερο), 1 (Πολύ θετικό).

- **emotional_intensity**: Η ένταση των συναισθημάτων που εκφράζονται, ανεξάρτητα από το αν είναι θετικά ή αρνητικά, δηλαδή πόσο φορτισμένη συναισθηματικά είναι η ομιλία.

Επιτρεπόμενες τιμές: 1 (Χαμηλή), 2 (Μέτρια), 3 (Υψηλή).

- **rheterical_intent**: Η κυρίαρχη ρητορική πρόθεση του ομιλητή κατά την ομιλία.

Επιρεπόμενες τιμές:

- Inform/Explain (Ενημέρωση/Επεξήγηση)
- Persuade/Advocate (Πειθώ/Υποστήριξη)
- Motivate/Inspire (Κίνητρο/Έμπνευση)
- Criticize/Denounce (Κριτική/Καταγγελία)
- Commemorate/Honor (Ανάμνηση/Τιμή)
- Other (Άλλο)

- **sentiment_json**: Πλήρης αναπαράσταση της συναισθηματικής ανάλυσης της ομιλίας σε μορφή JSON, η οποία περιλαμβάνει τις τιμές και τις επεξηγήσεις για κάθε κατηγορία ανάλυσης (sentiment, polarity_strength, rheterical_intent, emotional_intensity).

Ενδεικτική αναπαράσταση:

```

1.  {
2.    "speech_analysis": {
3.      "sentiment": {
4.        "value": 1,
5.        "explanation": "Θετικός τόνος για την ανάγκη βελτίωσης
και επίλυσης του προβλήματος της δαχοκτονίας."
6.      },
7.      "polarity_strength": {
8.        "value": 1,
9.        "explanation": "Ισχυρή θετική τάση λόγω προτάσεων για
βελτίωση και έμφασης στην επίλυση ενός χρόνιου ζητήματος."
10.     },
11.     "rheterical_intent": {
12.       "value": "Criticize/Denounce",
13.       "explanation": "Ο ομιλητής ασκεί κριτική στο υπάρχον
σύστημα και προτείνει αλλαγές, με στόχο την οριστική επίλυση του
προβλήματος."
14.     },
15.     "emotional_intensity": {
16.       "value": 2,
17.       "explanation": "Μέτρια συναισθηματική ένταση με έμ-
φαση στην επείγουσα ανάγκη επίλυσης ενός σημαντικού αγροτικού
ζητήματος."
18.     }
19.   }
20. }
```

Όνομα Πεδίου	Τύπος	Περιγραφή
id	Integer	Μοναδικό αναγνωριστικό της ομιλίας στη βάση δεδομένων (πρωτεύον κλειδί).
speech_id	String	Εναλλακτικός μοναδικός αναγνωριστικός κωδικός της ομιλίας από το αρχείο LegalDocML XML.
speaker_name	String	Ονοματεπώνυμο του ομιλητή, όπως καταγράφεται στο αρχείο LegalDocML XML.
speaker_id	String	Εναλλακτικός μοναδικός αναγνωριστικός κωδικός του ομιλητή από το αρχείο LegalDocML XML.
content	JSON	Το πλήρες κείμενο της ομιλίας σε μορφή JSON, οργανωμένο σε λίστα παραγράφων.
valuable	Boolean	Δείκτης αν η ομιλία προσφέρει χρήσιμο περιεχόμενο στη ροή της συζήτησης.
sentiment	Integer	Συνολική εκτίμηση του συναισθηματικού τόνου της ομιλίας.
polarity_strength	Integer	Η ένταση της θετικής ή αρνητικής στάσης, δηλαδή πόσο καθαρά εκφράζεται η θέση υπέρ ή κατά.
emotional_intensity	Integer	Η ένταση των συναισθημάτων που εκφράζονται, ανεξαρτήτως του προσήμου τους.
rhetorical_intent	String	Κύρια ρητορική πρόθεση του ομιλητή, π.χ. Inform/Explain κλπ.
sentiment_json	JSON	Πλήρης αναπαράσταση της συναισθηματικής ανάλυσης με τιμές και επεξηγήσεις για κάθε κατηγορία.

Πίνακας 4.2: Πεδία της οντότητας *Speech*

Σχήμα της οντότητας *Speaker*

Η οντότητα *Speaker* αναπαριστά τα πρόσωπα που συμμετέχουν στις κοινοβουλευτικές συζητήσεις ως ομιλητές. Κάθε ομιλητής μπορεί να ανήκει σε πολιτικό κόμμα και να έχει συμμετάσχει σε πολλαπλές ομιλίες κατά τη διάρκεια διαφορετικών ή της ίδιας κοινοβουλευτικής συζήτησης. Η οντότητα αυτή συνδέεται με τις *Speech*, *Debate* και *Political Party*, επιτρέποντας την παρακολούθηση της κοινοβουλευτικής δραστηριότητας των ομιλητών και τη συσχέτισή τους με τις πολιτικές θέσεις που εκφράζουν.

Πεδία της Οντότητας *Speaker*

Παρακάτω παρατίθενται και επεξηγούνται τα πεδία της οντότητας *Speaker*.

- **id:** Το μοναδικό αναγνωριστικό του ομιλητή στη βάση δεδομένων (πρωτεύον κλειδί).
- **speaker_name:** Το όνομα του ομιλητή, όπως αποδίδεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης.

Ενδεικτική αναπαράσταση: ANNA ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ

- **speaker_id**: Μοναδικός αναγνωριστικός κωδικός του ομιλητή, όπως αποδίδεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης.

Ενδεικτική αναπαράσταση: anna_diamantopoulou

- **link**: Σύνδεσμος προς τη σελίδα του ομιλητή στο Wikidata, όπως αποδίδεται στο αρχείο LegalDocML XML της συγκεκριμένης κοινοβουλευτικής συζήτησης.

Ενδεικτική αναπαράσταση: <https://www.wikidata.org/wiki/Q242305>

- **image**: Η εικόνα του ομιλητή, η οποία αποθηκεύεται ως πεδίο τύπου Media και μπορεί να προέρχεται είτε από το Wikidata είτε από άλλη εξωτερική πηγή.

- **description**: Περιγραφή του ομιλητή, συνήθως συνοπτική αναφορά στους βασικούς ρόλους ή επιτεύγματά του, όπως καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: Πρώην υπουργός και Επίτροπος της Ε.Ε., υποψήφια Πρόεδρος «ΠΑΣΟΚ-Κίνημα Αλλαγής»

- **gender**: Το φύλο του ομιλητή, όπως αυτό καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: γυναίκα

- **date_of_birth**: Ημερομηνία γέννησης του ομιλητή, όπως αυτή καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: Τρίτη 1 Ιανουαρίου 1924

- **date_of_death**: Ημερομηνία θανάτου του ομιλητή (αν υπάρχει), όπως αυτή καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: Δευτέρα 8 Αυγούστου 2011

- **place_of_birth**: Ο τόπος γέννησης του ομιλητή, όπως αυτός καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: Αθήνα

- **educated_at**: Εκπαιδευτικά ιδρύματα στα οποία φοίτησε ο ομιλητής, όπως αυτά καταγράφονται στο Wikidata. Αποθηκεύεται ως τιμές χωρισμένες με κόμμα.

Ενδεικτική αναπαράσταση: Aristotle University of Thessaloniki, Panteion University

- **website**: Επίσημος ιστότοπος του ομιλητή, όπως αυτός καταγράφεται στο Wikidata.

Ενδεικτική αναπαράσταση: <http://www.diamantopoulou.gr/>

- **occupation**: Το επάγγελμα ή οι κύριες ιδιότητες του ομιλητή, όπως αυτές καταγράφονται στο Wikidata. Αποθηκεύεται ως τιμές χωρισμένες με κόμμα.

Ενδεικτική αναπαράσταση: μηχανικός, πολιτικός

- **languages**: Οι γλώσσες που ομιλεί ο ομιλητής, όπως αυτές καταγράφονται στο Wikidata. Αποθηκεύεται ως τιμές χωρισμένες με κόμμα.

Ενδεικτική αναπαράσταση: νέα ελληνική γλώσσα, Αγγλικά, Ιταλικά

Όνομα Πεδίου	Τύπος	Περιγραφή
id	Integer	Μοναδικό αναγνωριστικό του ομιλητή στη βάση δεδομένων.
speaker_name	String	Το όνομα του ομιλητή από το αρχείο Legal-DocML XML.
speaker_id	String	Μοναδικός κωδικός του ομιλητή από το LegalDocML XML.
link	String	Σύνδεσμος προς τη σελίδα του ομιλητή στο Wikidata.
image	Media	Εικόνα του ομιλητή.
description	Text	Σύντομη περιγραφή του ομιλητή, βασισμένη στο Wikidata.
gender	String	Φύλο του ομιλητή.
date_of_birth	String	Ημερομηνία γέννησης του ομιλητή.
date_of_death	String	Ημερομηνία θανάτου του ομιλητή (αν υπάρχει).
place_of_birth	String	Τόπος γέννησης του ομιλητή.
educated_at	Text	Εκπαιδευτικά ιδρύματα που φοίτησε ο ομιλητής.
website	String	Επίσημος ιστότοπος του ομιλητή.
occupation	Text	Επαγγέλματα ή ιδιότητες του ομιλητή.
languages	Text	Γλώσσες που ομιλεί ο ομιλητής.

Πίνακας 4.3: Πεδία της οντότητας *Speaker*

Σχήμα της οντότητας *Topic*

Η οντότητα *Topic* αναπαριστά τις θεματικές ενότητες στις οποίες εντάσσονται οι κοινοβουλευτικές συζητήσεις. Κάθε θεματική ενότητα αντικατοπτρίζει ένα συγκεκριμένο πεδίο ενδιαφέροντος ή νομοθετικής δραστηριότητας, διευκολύνοντας την ταξινόμηση και την ανάλυση των συζητήσεων. Η οντότητα αυτή συνδέεται με την οντότητα *Debate*, επιτρέποντας την ομαδοποίηση των σχετικών συζητήσεων με βάση κοινές θεματικές αναφορές.

Οι θεματικές ενότητες καθορίστηκαν μετά από μελέτη υπάρχουσών κατηγοριοποιήσεων της Βουλής των Ελλήνων και αντίστοιχων εφαρμογών ανοιχτής διακυβέρνησης. Αρχικά, χρησιμοποιήθηκαν οι κατηγορίες της πλατφόρμας *Vouliwatch*¹, οι οποίες στη συνέχεια προσαρμόστηκαν και εμπλουτίστηκαν με βάση το σύστημα ταξινόμησης *EuroVoc*² της Ευρωπαϊκής Ένωσης, ώστε να διασφαλιστεί συνέπεια και πληρότητα. Η τελική λίστα παρουσιάζεται στον ακόλουθο πίνακα.

¹ Vouliwatch: <https://vouliwatch.gr/>

² Eurovoc: https://showvoc.op.europa.eu/#/datasets/OP_EuroVoc/data

ΑΑ	Θεματική Ενότητα
1	Οικονομική Πολιτική και Δημοσιονομικά
2	Ανάπτυξη, Ανταγωνιστικότητα και Επενδύσεις
3	Εργασιακές Σχέσεις και Κοινωνική Ασφάλιση
4	Παιδεία, Θρησκευματα, Έρευνα και Καινοτομία
5	Υγεία και Κοινωνική Πρόνοια
6	Εξωτερική Πολιτική και Ευρωπαϊκές Υποθέσεις
7	Εθνική Άμυνα
8	Δικαιοσύνη και Ανθρώπινα Δικαιώματα
9	Δημόσια Τάξη και Προστασία του Πολίτη
10	Περιβάλλον, Ενέργεια και Κλιματική Αλλαγή
11	Υποδομές, Μεταφορές και Δίκτυα
12	Αγροτική Ανάπτυξη και Τρόφιμα
13	Τουρισμός
14	Πολιτισμός και Αθλητισμός
15	Ψηφιακή Διακυβέρνηση και Επικοινωνίες
16	Στεγαστική Πολιτική και Αστικός Σχεδιασμός
17	Δημόσια Διοίκηση και Αποκέντρωση
18	Μετανάστευση και Άσυλο
19	Ναυτιλία και Νησιωτική Πολιτική
20	Θεσμικά Ζητήματα και Κοινοβουλευτικές Διαδικασίες
21	Άλλο / Γενικά Θέματα

Πίνακας 4.4: Λίστα θεματικών ενότητων στις οποίες κατηγοριοποιούνται οι κοινοβουλευτικές συζητήσεις του συστήματος

Πεδία της Οντότητας Topic

Παρακάτω παρατίθενται και επεξηγούνται τα πεδία της οντότητας Topic.

- **id**: Το μοναδικό αναγνωριστικό της θεματικής ενότητας στη βάση δεδομένων (πρωτεύον κλειδί).
- **name**: Η ονομασία της θεματικής κατηγορίας, η οποία περιγράφει το αντικείμενο της συζήτησης.

Ενδεικτική αναπαράσταση: Εξωτερική Πολιτική και Ευρωπαϊκές Υποθέσεις

Όνομα Πεδίου	Τύπος	Περιγραφή
id	Integer	Μοναδικό αναγνωριστικό της θεματικής ενότητας στη βάση δεδομένων.
name	String	Η ονομασία της θεματικής κατηγορίας που περιγράφει το αντικείμενο της συζήτησης.

Πίνακας 4.5: Πεδία της οντότητας Topic

Σχήμα της οντότητας Political Party

Η οντότητα Political Party αναπαριστά τα πολιτικά κόμματα που συμμετέχουν στις κοινοβουλευτικές διαδικασίες. Κάθε ομιλητής (οντότητα Speaker) είναι συνδεδεμένος με ένα

ή περισσότερα πολιτικά κόμματα των οποίων έχει υπάρξει μέλος σε οποιαδήποτε χρονική περίοδο, και τα οποία πιθανώς επηρεάζουν το περιεχόμενο και τη ρητορική της ομιλίας του.

Πεδία της Οντότητας **Political Party**

Παρακάτω παρατίθενται και επεξηγούνται τα πεδία της οντότητας **Political Party**.

- **id**: Το μοναδικό αναγνωριστικό του πολιτικού κόμματος στη βάση δεδομένων (πρωτεύον κλειδί).
- **name**: Η επίσημη ονομασία του πολιτικού κόμματος, όπως αυτή είναι καταχωρισμένη στην αντίστοιχη εγγραφή του στο Wikidata, σύμφωνα με τα επίσημα δεδομένα της παράταξης.
- **image**: Η επίσημη οπτική αναπαράσταση του πολιτικού κόμματος, όπως το λογότυπο του ή άλλο σχετικό οπτικό στοιχείο. Η εικόνα αντλείται από το Wikidata είτε από άλλη εξωτερική πηγή και αποθηκεύεται ως πεδίο τύπου Media.

Όνομα Πεδίου	Τύπος	Περιγραφή
id	Integer	Μοναδικό αναγνωριστικό του πολιτικού κόμματος στη βάση δεδομένων.
name	String	Επίσημη ονομασία του πολιτικού κόμματος από το Wikidata.
image	Media	Οπτική αναπαράσταση του κόμματος (π.χ. λογότυπο).

Πίνακας 4.6: Πεδία της οντότητας **Political Party**

4.2.3 Συσχετίσεις μεταξύ των Οντοτήτων

Οι οντότητες του συστήματος συνδέονται μεταξύ τους μέσω σχέσεων που επιτρέπουν την οργάνωση και ανάλυση των κοινοβουλευτικών δεδομένων. Παρακάτω αναλύονται οι κύριες συσχετίσεις μεταξύ των οντοτήτων:

Debate – Speech

Κάθε κοινοβουλευτική συζήτηση μπορεί να περιλαμβάνει πολλαπλές ομιλίες, ενώ κάθε ομιλία ανήκει σε μία μοναδική συζήτηση. Ο περιορισμός αυτός αποτυπώνεται έμμεσα μέσω του μοναδικού περιορισμού στο πεδίο `speech_id`, το οποίο διασφαλίζει ότι κάθε ομιλία είναι μοναδική και αντιστοιχεί αποκλειστικά σε μία συγκεκριμένη συζήτηση.

Debate – Speaker

Οι ομιλητές που συμμετέχουν σε μια συζήτηση μπορεί είτε να έχουν εκφωνήσει ομιλία είτε απλώς να παρίστανται. Ένας ομιλητής μπορεί να συμμετέχει σε πολλές συζητήσεις και κάθε συζήτηση μπορεί να έχει πολλούς συμμετέχοντες.

Debate – Topic

Κάθε συζήτηση κατηγοριοποιείται κατά μέγιστο σε 2 από τις 21 προκαθορισμένες θεματικές

ενότητες, αποτυπώνοντας τα αντικείμενά της. Αντίστοιχα, μια θεματική ενότητα μπορεί να χαρακτηρίζει περισσότερες από μία κοινοβουλευτικές συζητήσεις.

Speaker – Speech

Κάθε ομιλία αποδίδεται σε έναν μοναδικό ομιλητή, ο οποίος μπορεί να έχει πραγματοποιήσει πολλαπλές ομιλίες στο πλαίσιο διαφορετικών ή της ίδιας συζήτησης.

Speaker – Political Party

Ένας ομιλητής μπορεί να έχει υπάρξει μέλος ενός ή περισσότερων πολιτικών κομμάτων κατά τη διάρκεια της κοινοβουλευτικής του πορείας, ενώ κάθε πολιτικό κόμμα συνδέεται με τους ομιλητές που το έχουν εκπροσωπήσει.

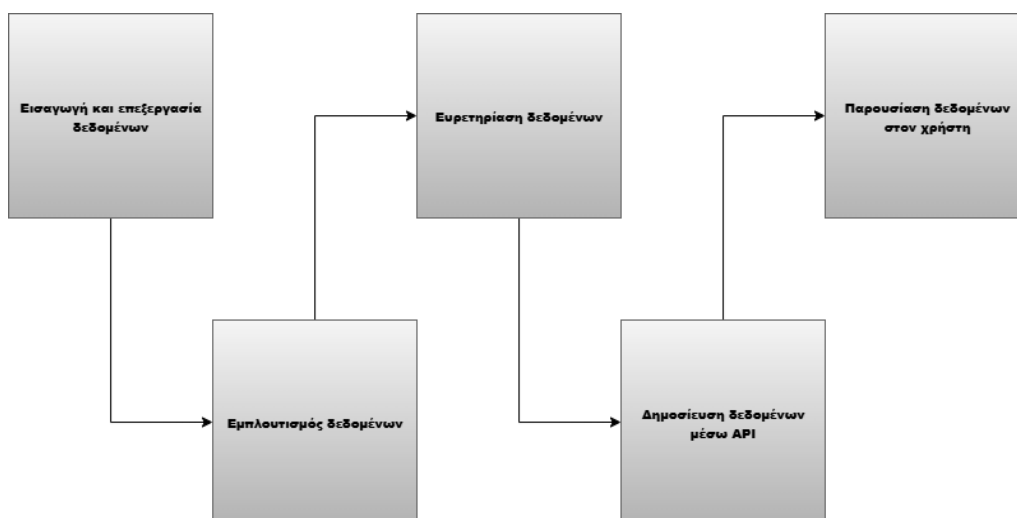
Συσχέτιση	Τύπος Σχέσης	Περιγραφή
Debate - Speech	One-to-Many	Κάθε συζήτηση περιλαμβάνει πολλές ομιλίες, ενώ κάθε ομιλία ανήκει σε μία συζήτηση.
Debate - Speaker	Many-to-Many	Πολλοί ομιλητές μπορεί να συμμετέχουν σε μία συζήτηση, και κάθε ομιλητής μπορεί να συμμετέχει σε πολλές συζητήσεις.
Debate - Topic	Many-to-Many	Κάθε συζήτηση συνδέεται με πολλές θεματικές ενότητες και κάθε θεματική ενότητα με πολλές συζητήσεις.
Speaker - Speech	One-to-Many	Κάθε ομιλία αποδίδεται σε έναν ομιλητή, ο οποίος μπορεί να έχει εκφωνήσει πολλές ομιλίες.
Speaker - Political Party	Many-to-Many	Ένας ομιλητής μπορεί να έχει συνδεθεί με πολλά πολιτικά κόμματα και κάθε κόμμα με πολλούς ομιλητές.

Πίνακας 4.7: Συσχετίσεις μεταξύ των οντοτήτων του συστήματος

Λειτουργίες και Περιγραφή Τελικού Συστήματος

Το παρόν κεφάλαιο παρουσιάζει τα αποτελέσματα της επεξεργασίας και αξιοποίησης των κοινοβουλευτικών δεδομένων μέσα από το σύστημα που αναπτύχθηκε. Τα αρχεία μετατράπηκαν σε εύχρηστες, δομημένες εγγραφές που μπορούν να αναζητηθούν και να προβληθούν μέσω μιας φιλικής διαδικτυακής διεπαφής. Ο εμπλουτισμός των δεδομένων με επιπλέον πληροφορίες, οι λειτουργίες αναζήτησης και πλοήγησης, καθώς και οι δυνατότητες που παρέχονται στους χρήστες, αποτυπώνουν την ολοκληρωμένη εμπειρία που προσφέρει το σύστημα για την εξερεύνηση του κοινοβουλευτικού λόγου.

Η συνολική ροή επεξεργασίας, εμπλουτισμού, ευρετηρίασης και παρουσίασης των δεδομένων απεικονίζεται στο Σχήμα 5.1, το οποίο συνοψίζει τα βασικά στάδια από την εισαγωγή των δεδομένων μέχρι την παρουσίασή τους στον τελικό χρήστη.



Σχήμα 5.1: Διαδικασία επεξεργασίας, εμπλουτισμού, ευρετηρίασης και παρουσίασης των κοινοβουλευτικών δεδομένων.

5.1 Παρουσίαση Υποσυστημάτων

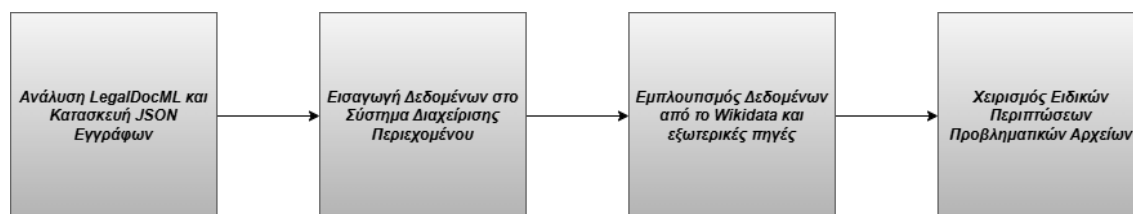
Η λειτουργία της εφαρμογής βασίζεται σε διακριτά στάδια επεξεργασίας δεδομένων, τα οποία καλύπτουν τη ροή από την εισαγωγή και οργάνωση των κοινοβουλευτικών αρχείων έως

τον εμπλουτισμό, την αποθήκευση και την ευρετηρίασή τους για αναζήτηση. Στην παρούσα ενότητα παρουσιάζονται αναλυτικά τα επιμέρους υποσυστήματα που υποστηρίζουν τη μετατροπή, την ενσωμάτωση επιπλέον πληροφορίας και την προετοιμασία του περιεχομένου για ανάκτηση μέσω της τελικής διεπαφής.

5.1.1 Υποσύστημα Εισαγωγής και Εμπλουτισμού Δεδομένων

Το πρώτο στάδιο αφορά την αρχική επεξεργασία των κοινοβουλευτικών αρχείων, με στόχο την αναδόμηση, αποθήκευση και ενίσχυση των δεδομένων. Το υποσύστημα εισαγωγής και εμπλουτισμού υλοποιεί αυτή τη διαδικασία μέσα από διακριτά βήματα, τα οποία απεικονίζονται συνοπτικά στο παρακάτω διάγραμμα.

Η διαδικασία εισαγωγής δεδομένων είναι πλήρως αυτοματοποιημένη, καλύπτοντας όλα τα στάδια από τη μετατροπή των αρχείων LegalDocML έως και την αποθήκευσή τους στο σύστημα διαχείρισης περιεχομένου. Ο συνολικός μηχανισμός έχει σχεδιαστεί ώστε να επιτρέπει την ταυτόχρονη επεξεργασία μεγάλου όγκου αρχείων, χωρίς την ανάγκη χειροκίνητης παρέμβασης. Με την ολοκλήρωση της εισαγωγής, τα δεδομένα προωθούνται για ανάλυση φυσικής γλώσσας και ευρετηρίαση ώστε να υποστηριχθεί η αναζήτηση πλήρους κειμένου, όπως περιγράφεται στα επόμενα υποσυστήματα.



Σχήμα 5.2: Στάδια μετατροπής, εμπλουτισμού και αποθήκευσης των κοινοβουλευτικών πληροφοριών

Ανάλυση και Εξαγωγή Δεδομένων από LegalDocML

Η διαδικασία ξεκινά με την ανάλυση των αρχείων συζητήσεων της Βουλής σε μορφή LegalDocML (XML), από τα οποία εξάγονται τα απαραίτητα δεδομένα σε JSON μορφή, συμβατή με τα εσωτερικά σχήματα της εφαρμογής. Τα αρχεία ακολουθούν το πρότυπο Akoma Ntoso, το οποίο χρησιμοποιείται διεθνώς για την αναπαράσταση κοινοβουλευτικών εγγράφων.

Η επεξεργασία εντοπίζει και αναδομεί βασικά στοιχεία, όπως τα μεταδεδομένα, τον πρόλογο και τις επιμέρους ομιλίες κάθε συνεδρίασης. Στο παρακάτω απόσπασμα παρουσιάζεται ενδεικτικά η αρχική μορφή του XML εγγράφου, όπως προκύπτει από το αρχείο συζήτησης της 1ης Μαρτίου 1994. Το συγκεκριμένο απόσπασμα περιλαμβάνει μόνο την ενότητα του προλόγου και των ομιλιών και δεν αντιστοιχεί στο σύνολο του εγγράφου:

```

53 <preface>
54 <container name="parliament_details">
55 <rp>Ρ Α Κ Τ Ι Κ Α Β Ο Υ Λ Η Σ</rp>
56 <rp>Η' ΠΕΡΙΟΔΟΣ (ΠΡΟΕΔΡΕΥΟΜΕΝΗΣ ΔΗΜΟΚΡΑΤΙΑΣ)</rp>
57 <rp>Σ Υ Ν Ο Δ Ο Σ Α' </rp>
58 <rp>ΣΥΝΕΑΡΙΑΣΗ ΟΒ' </rp>
59 <rp>Τρίτη 1 Μαρτίου 1994</rp>
60 </container>
61 </preface>
62 <debateBody>
63 <debateSection name="opening">
64 <rp>Αθήνα, σήμερα την 1η Μαρτίου 1994, ημέρα Τρίτη και ώρα 18.10' , συνήλθε στην Αίθουσα συνεδριάσεων του Βουλευτηρίου η Βουλή σε ολομέλεια για να συνεδριάσει
65 </debateSection>
66 <debateSection name="main_debate_section">
67 > <speech by="apostolos kaklamanis" eid="debate_1994-03-01_1_speech_1">...
312 </speech>
313 <speech by="apostolos kaklamanis" eid="debate_1994-03-01_1_speech_2">
314 <from>Απόστολος Κακλαμάνης</from>
315 <rp>Κύριοι συνάδελφοι, ο Πρόεδρος της Κυβερνήσεως έστειλε στον Πρόεδρο της Βουλής των Ελλήνων το κατωτέρω έγγραφο:</rp>
316 <rp>ΘΕΜΑ: Κυβερνητική μεταβολή.</rp>
317 <rp>Κύριε Πρόεδρε,</rp>
318 <rp>Έχουμε την τιμή να παρακαλέσουμε, όπως ανακοινώσετε στο Σώμα ότι:</rp>
319 <rp>α) Με το αριθ. 31 προεδρικό διάταγμα, που δημοσιεύθηκε στο ΦΕΚ αριθ. 23/Α/25.2.94, έγιναν αποδεκτές οι παρατηρήσεις που υπέβαλαν οι κύριοι:</rp>
320 <rp>Γεωργιος Γεννηματάς του Θεοδώρα από τη θέση του Υπουργού Οικονομικών και Αλέξανδρος Παπαδόπουλος του Κωνσταντίνου από τη θέση του Υφυπουργού Οικονομικών</rp>
321 <rp>β) Με το αριθ. 32 προεδρικό διάταγμα, που δημοσιεύθηκε στο ΦΕΚ αριθ. 23/Α/25.2.94, διορίστηκαν οι κύριοι:</rp>
322 <rp>Αλέξανδρος Παπαδόπουλος του Κωνσταντίνου, ως Υπουργός Οικονομικών και</rp>
323 <rp>Νικόλαος Κυριαζίδης του Παναγιώτου, ως Υφυπουργός Οικονομικών.</rp>
324 <rp>Ο Προϋπουργός</rp>
325 <rp>Ανδρέας Γ. Παπανδρέου</rp>
326 </speech>
327 <speech by="apostolos kaklamanis" eid="debate_1994-03-01_1_speech_3">
328 <from>Απόστολος Κακλαμάνης</from>
329 <rp>Κύριοι Βουλευτές, ευσερχόμαστε στην ημερήσια διάταξη</rp>
330 <rp>ΤΗΣ ΝΟΜΟΘΕΤΙΚΗΣ ΕΡΓΑΣΙΑΣ</rp>
331 <rp>Έχω την τιμή να ανακοινώσω προς το Σώμα ότι ο Πρόεδρος της Προανακριτικής Επιτροπής "Για την παρακολούθηση των τηλεφωνικών συνδιαλέξεων των πολιτικών αν
332 <rp>Εγκρίνει η Βουλή</rp>
333 </speech>
334 <speech by="bouleutis/es" eid="debate_1994-03-01_1_speech_4">
335 <from>Βουλευτής/ες</from>
336 <rp>Μάλιστα, μάλιστα.</rp>
337 </speech>

```

Εικόνα 5.1: Απόσπασμα XML εγγράφου της συνεδρίασης της 1ης Μαρτίου 1994 με πρόλογο και επιμέρους ομιλίες (preface, speech)

Στη συνέχεια, το περιεχόμενο μετατρέπεται σε κατάλληλα δομημένο JSON, με διατήρηση της ιεραρχικής δομής και εμπλουτισμό με επιπλέον μεταδεδομένα. Παρακάτω παρουσιάζεται η αντίστοιχη απόδοση του προλόγου και των ομιλιών:

```

333 "debateBody": {
334 {
335 "debateSection": {
336 {
337 "s": {
338 "name": "opening"
339 },
340 "p": {
341 "text": "Αθήνα, σήμερα την 1η Μαρτίου 1994, ημέρα Τρίτη και ώρα 18.10' , συνήλθε στην Αίθουσα συνεδριάσεων του Βουλευτηρίου η Βουλή σε Ολομέλεια για να συνεδριάσει υπό την Προεδρία του Προέδρου κ. ΑΠΟΣΤΟΛΟΥ ΚΑΚΛΑΜΑΝΗ.",
342 },
343 },
344 {
345 "s": {
346 "name": "main_debate_section"
347 },
348 "speech": [
349 {
350 "s": {
351 "by": "apostolos kaklamanis",
352 "eid": "debate_1994-03-01_1_speech_1"
353 },
354 "from": {
355 "text": "Απόστολος Κακλαμάνης"
356 },
357 "p": {
358 "text": "Κύριοι συνάδελφοι, αρχίζει η συνεδρίαση.",
359 "text": "Παρακαλείται η Γραμματεία και Ραδιόλη Σήμα, Βουλευτής Παναγιώτης, να διαβάσει τις αναφορές προς το Σώμα.",
360 "text": "Ανακοινώνονται από τη Γραμματεία τα ακόλουθα:",
361 "text": "Α' ΚΑΤΑΒΕΒΗ ΑΝΑΦΟΡΑ",
362 "1. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία το Πρόγραμμα Αξιοανακρίσεων του Πανελληνίου Συλλόγου Ανηλίκων ζητεί την ένταξη των νέων ανηλίκων στην παραγωγική διαδικασία.",
363 "2. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία ο Πόλεμος του Αφγανιστάν Παναγιώτης θέτει διαμνηστικό για την καλύτερη ίδρυση του τμήματος κοινωνικής διατήρησης στην Κομοτηνή.",
364 "3. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία οι Αγοραστές Συνεταιρισμού Μαυροκορφή και Κορύμβου Έβρου ζητούν να σωθεί η ποσότητα της καλλιεργητικής καπνού στις περιοχές τους.",
365 "4. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Καστανών Έβρου ζητεί την επανακελευστική του Τελωνείου Καστανών.",
366 "5. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Ορμενίου Έβρου ζητεί να της παραχωρηθούν εκτάσεις από τον ΕΟΦ προκειμένου να ανεγερθούν ο' αυτές οι εγκαταστάσεις του Πελοποννήσου.",
367 "6. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία το Εργαστηριακό Κέντρο Διαιτητικής και Περιφέρειας Νομού Έβρου διαμνηστικό για την καλύτερη των νέων θέσεων εργασίας του ΟΑΕΒ 15.",
368 "7. Ο Βουλευτής Έβρου κ. ΑΠΟΣΤΟΛΟΣ ΦΩΤΙΑΔΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Ορμενίου Έβρου ζητεί την αποδέσμευση ορυκτοειδών της ΕΤΕ στην περιοχή τους.",
369 "8. Ο Βουλευτής Ηρακλείου κ. ΕΡΜΑΝΩΝΑ ΦΡΑΓΚΙΑΔΟΥΚΗ κατέθεσε αναφορά με την οποία η Κοινότητα Πετών Νομού Ηρακλείου ζητεί να πληρωθεί η θέση της μάστις από Αγοραστικό Ταμείο Πετών.",
370 "9. Ο Βουλευτής Ηρακλείου κ. ΕΡΜΑΝΩΝΑ ΦΡΑΓΚΙΑΔΟΥΚΗ κατέθεσε αναφορά με την οποία η Κοινότητα Τερέλίου Νομού Ηρακλείου ζητεί ίδρυση λυκείου στο Τερέλι.",
371 "10. Ο Βουλευτής Ηρακλείου κ. ΕΡΜΑΝΩΝΑ ΦΡΑΓΚΙΑΔΟΥΚΗ κατέθεσε αναφορά με την οποία η Κοινότητα Κάτω Άστρων Νομού Ηρακλείου ζητεί χρηματοδότηση της κοινότητας για τη διαμόρφωση του υγειονομικού σταθμού.",
372 "11. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία ο Πρόεδρος της Κοινότητας Αγίου Κων/νου Αιτωλ/νίας ζητεί την οικονομική ενίσχυση της φιλομαρικής αρχήσας της κοινότητάς του.",
373 "12. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Γαβρολίνας Αιτωλ/νίας ζητεί την κατασκευή γηπέδου υπόκειται ο' αυτή.",
374 "13. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Τερέλίου Νομού Ηρακλείου ζητεί τη μίσση μέσων για την παύση της ανεργίας στην περιοχή του.",
375 "14. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Αρμυρίου Τριφυλίας Αιτωλ/νίας ζητεί έκτακτη οικονομική ενίσχυση για την κάλυψη λειτουργικών αναγκών της.",
376 "15. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία εκπαιδευτικοί ζητούν να μη διαγράφονται από την επετηρίδα πριν το 55ο έτος της ηλικίας τους.",
377 "16. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία ο Αγοραστικός Σύλλογος Βόνιτσας διαμνηστικό για την μετονομασία του αερολιμένα του Ακτίου.",
378 "17. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Φυτειών Αιτωλ/νίας ζητεί την πρόληψη γρουατών.",
379 "18. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Κιφέλης Αγίντας ζητεί τον καθορισμό νέων όρων δόμησης.",
380 "19. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία ο Ποδοσφαιρικός Αθλητικός Σύλλογος Βόνιτσας Αιτωλ/νίας ζητεί χρηματοδότηση για την κάλυψη αναγκών της ποδοσφαιρικής ομάδας Βόνιτσας.",
381 "20. Ο Βουλευτής Αιτωλ/νίας κ. ΧΡΗΣΤΟΣ ΚΟΚΚΙΝΟΒΑΣΙΛΗΣ κατέθεσε αναφορά με την οποία η Κοινότητα Φυτειών Αιτωλ/νίας ζητεί το δικαίωμα της εθνικής οδού εντός του συν/σμού Κουβούρα.",
382 "21. Ο Βουλευτής Ηλείας κ. ΒΑΣΙΛΕΙΟΣ ΚΟΡΚΟΡΑΝΟΣ κατέθεσε αναφορά με την οποία η Κοινότητα Αντίστασης Ηλείας ζητεί χρηματοδότηση για την κάλυψη κοινωνικών αναγκών.",

```

Εικόνα 5.2: Αναδομημένη μορφή εισαγωγικού μέρους και ομιλιών της συζήτησης 1ης Μαρτίου 1994 σε JSON

Η διαδικασία εξαγωγής περιλαμβάνει επίσης ελέγχους εγκυρότητας και καταγραφή περιπτώσεων ασυνεπειών (όπως διπλοεγγραφές ή μη ευθυγραμμισμένες αναφορές σε ομιλητές). Οι εγγραφές αυτές καταγράφονται σε αρχεία καταγραφής χωρίς να διακόπτεται η ροή εισαγωγής.

Το τελικό JSON που παράγεται είναι βελτιστοποιημένο για εισαγωγή και συσχέτιση στο υπόλοιπο σύστημα, διατηρώντας όλα τα κρίσιμα στοιχεία κάθε κοινοβουλευτικής συνεδρίασης με συνέπεια και πληρότητα.

Εμπλουτισμός Δεδομένων από Εξωτερικές Πηγές

Για την ενίσχυση του περιεχομένου κάθε εγγραφής ομιλητή, το σύστημα αντλεί αυτόματα πρόσθετες πληροφορίες από το Wikidata, ώστε να παρουσιαστεί ένα πιο πλήρες και κατατοπιστικό προφίλ στο περιβάλλον χρήστη. Πέρα από τα δεδομένα που προέρχονται από τα αρχεία συζητήσεων, συλλέγονται επιπλέον στοιχεία, όπως ημερομηνία και τόπος γέννησης, πολιτικός φορέας, εκπαιδευτικό υπόβαθρο, επάγγελμα, γλώσσες και σύνδεσμος προς την προσωπική του ιστοσελίδα.

Η σύνδεση με το Wikidata βασίζεται στο μοναδικό `entityId` κάθε ομιλητή, το οποίο εντοπίζεται κατά την ανάλυση των αρχείων LegalDocML. Για κάθε ομιλητή εκτελείται ένα δυναμικό ερώτημα (SPARQL query) προς την επίσημη υπηρεσία Wikidata SPARQL endpoint, το οποίο συλλέγει τα απαραίτητα μεταδεδομένα, με έλεγχο εγκυρότητας μέσω του Wikidata Query Service¹.

Η ενσωμάτωση των δεδομένων πραγματοποιείται κατά την καταχώριση του ομιλητή, με τα σταθερά χαρακτηριστικά του (όπως ημερομηνία γέννησης, πολιτικός φορέας, εκπαιδευτικά ιδρύματα) να αποθηκεύονται στη βάση δεδομένων, ενώ δυναμικά στοιχεία, όπως οι θέσεις που έχει διατελέσει, να ανακτώνται σε πραγματικό χρόνο κατά την προβολή του προφίλ στο frontend. Με αυτόν τον τρόπο διασφαλίζεται ότι οι πληροφορίες που προβάλλονται είναι πλήρεις, έγκυρες και εμπλουτισμένες από έγκυρες εξωτερικές πηγές.

Εισαγωγή Δεδομένων στο Σύστημα Διαχείρισης Περιεχομένου

Τέλος, τα δεδομένα εισάγονται στο σύστημα διαχείρισης περιεχομένου, το οποίο λειτουργεί και ως κύρια βάση δεδομένων της εφαρμογής.

Κατά την εισαγωγή, οι βασικές οντότητες —συζητήσεις, ομιλητές, πολιτικά κόμματα και ομιλίες— καταχωρίζονται οργανωμένα στο σύστημα διαχείρισης περιεχομένου και συνδέονται μεταξύ τους, εξασφαλίζοντας την πληρότητα και τη συνοχή των δεδομένων. Κάθε ομιλία συσχετίζεται ρητά με τον αντίστοιχο ομιλητή και τη συνεδρίαση στην οποία ανήκει, ενώ καταγράφονται επιπλέον πληροφορίες, όπως οι πολιτικοί φορείς και οι εικόνες των ομιλητών, ώστε να σχηματιστεί μια πλήρης και αξιόπιστη βάση δεδομένων, έτοιμη για χρήση και ανάκτηση.

Κάθε ενέργεια υλοποιείται ως αυτόνομο βήμα, ώστε να εξασφαλίζεται η ανεξαρτησία και επεκτασιμότητα της διαδικασίας. Ιδιαίτερη έμφαση δίνεται στην ακρίβεια των σχέσεων μεταξύ των εγγραφών και στην καταγραφή μοναδικών αναγνωριστικών, ώστε να διατηρείται η συνέπεια του περιεχομένου.

5.1.2 Υποσύστημα Επεξεργασίας Ομιλιών με Φυσική Γλώσσα (NLP)

Το υποσύστημα NLP εμπλουτίζει τις ομιλίες και τις συζητήσεις με χρήσιμα νοηματικά χαρακτηριστικά, προσφέροντας στους χρήστες μια πιο βαθιά κατανόηση του κοινοβουλευτι-

¹ Πηγή: <https://query.wikidata.org>

κού λόγου. Κάθε κοινοβουλευτική συζήτηση εμπλουτίζεται με σύντομη περίληψη, θεματική ταξινόμηση σε έως δύο κατηγορίες από τις 21 προκαθορισμένες (Πίνακας 4.4), καθώς και ανάλυση συναισθηματικού τόνου, ρητορικής πρόθεσης, πολικότητας και συναισθηματικής έντασης.

Η ανάλυση πραγματοποιείται επιλεκτικά σε εκείνες τις ομιλίες της κάθε συζήτησης, οι οποίες χαρακτηρίζονται ως “valuable”, δηλαδή περιέχουν τουλάχιστον 150 χαρακτήρες, ώστε να αποφεύγεται η επεξεργασία τυπικών ή διαδικαστικών παρεμβάσεων (π.χ. «Ευχαριστώ κύριε Πρόεδρε») που δεν προσφέρουν ουσιαστικό περιεχόμενο στη συζήτηση. Η θεματική κατηγοριοποίηση βασίζεται στο κείμενο της περίληψης κάθε ομιλίας, η οποία αποτυπώνει συνοπτικά τα κύρια ζητήματα που θίγονται.

Οι πληροφορίες αυτές καθιστούν πιο εύκολη την πλοήγηση, την αναζήτηση και την ανάλυση των δεδομένων, καθώς επιτρέπουν στους χρήστες να εντοπίζουν γρήγορα τις πιο σχετικές, έντονες ή ενδιαφέρουσες παρεμβάσεις. Ο χρήστης μπορεί πλέον να πραγματοποιεί συνδυαστικές αναζητήσεις, εντοπίζοντας πού ειπώθηκαν συγκεκριμένες φράσεις ή συζητήθηκαν συγκεκριμένα θέματα, αλλά και να περιορίζει περαιτέρω τα αποτελέσματα με βάση τη ρητορική πρόθεση ή τον συναισθηματικό τόνο της ομιλίας. Για παράδειγμα, μπορεί να αναζητήσει παρεμβάσεις που συνδυάζουν πρόθεση “Πειθώ/Υποστήριξη” με πολύ θετικό συναισθηματικό φορτίο, ώστε να εστιάσει στις πιο ενθουσιώδεις και υποστηρικτικές δηλώσεις.

Με αυτόν τον τρόπο, διευκολύνεται η μελέτη των προθέσεων των ομιλητών, η παρακολούθηση των απόψεων που εξελίσσονται με την πάροδο των ετών και η εντοπισμός υποστηριζόμενων θέσεων, χωρίς αυτές να «χάνονται» μέσα στον τεράστιο όγκο πληροφοριών. Η επεξεργασία πραγματοποιείται αυτόματα μέσω ενός εξειδικευμένου γλωσσικού μοντέλου για την ελληνική γλώσσα (Krikti²), το οποίο ενσωματώθηκε στο σύστημα μέσω FastAPI υπηρεσίας.

5.1.3 Υποσύστημα Αναζήτησης Πλήρους Κειμένου με Elasticsearch

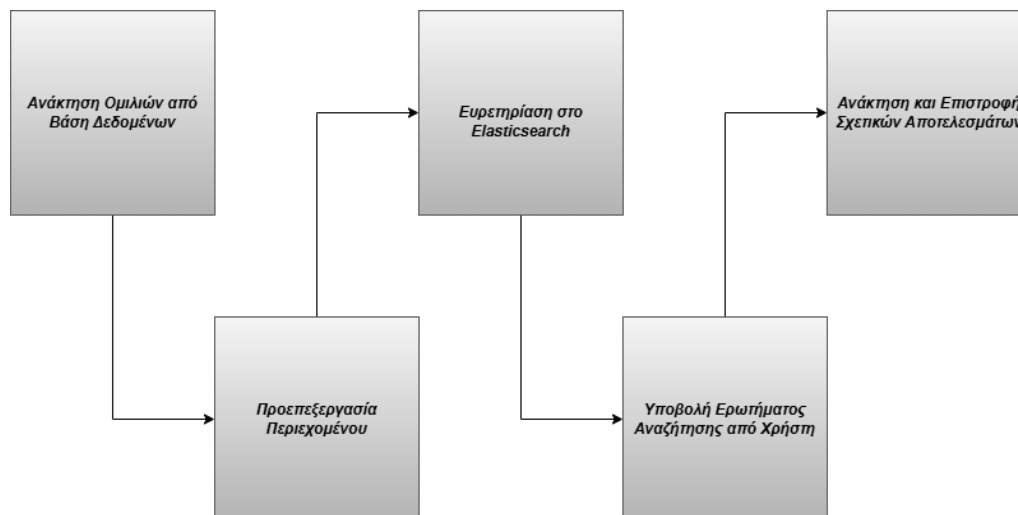
Το υποσύστημα αναζήτησης πλήρους κειμένου προσφέρει στους χρήστες τη δυνατότητα να εντοπίζουν γρήγορα και στοχευμένα κοινοβουλευτικές ομιλίες, αξιοποιώντας τρία διαφορετικά επίπεδα αναζήτησης που καλύπτουν τόσο το περιεχόμενο όσο και τη νοηματική ανάλυση των δεδομένων. Συγκεκριμένα, το σύστημα επιτρέπει:

- **Αναζήτηση πλήρους κειμένου:** Ο χρήστης μπορεί να εισάγει οποιαδήποτε λέξη ή φράση και να αναζητήσει σε όλο το σώμα των ομιλιών, εντοπίζοντας ακριβώς πού και από ποιόν ομιλητή χρησιμοποιήθηκε.
- **Αναζήτηση συναισθηματικού φορτίου:** Οι ομιλίες μπορούν να εντοπιστούν ανάλογα με τον συναισθηματικό τους τόνο (π.χ. πολύ θετικός, ουδέτερος, αρνητικός), αλλά και με βάση το επίπεδο της συναισθηματικής έντασης, διευκολύνοντας τη μελέτη της στάσης και της συναισθηματικής εμπλοκής των ομιλητών σε διάφορα θέματα.
- **Αναζήτηση ρητορικής πρόθεσης:** Το σύστημα επιτρέπει την ανεύρεση ομιλιών σύμφωνα με τη ρητορική τους πρόθεση, όπως «Πειθώ/Υποστήριξη», «Κριτική/Καταγγελία», «Ενημέρωση/Εξήγηση», αποκαλύπτοντας τη στάση και τη στρατηγική του ομιλητή.

²Πηγή: <https://huggingface.co/papers/2505.13772>

Στα αποτελέσματα της αναζήτησης, οι σχετικές φράσεις τονίζονται μέσα στα κείμενα, ώστε ο χρήστης να εντοπίζει άμεσα τα σημεία ενδιαφέροντος.

Το συνολικό pipeline της λειτουργίας συνοψίζεται στο Σχήμα 5.3, το οποίο απεικονίζει τη ροή των δεδομένων από την ανάκτηση και προεπεξεργασία των ομιλιών έως και την υποβολή του ερωτήματος από τον χρήστη και την εμφάνιση των σχετικών αποτελεσμάτων.



Σχήμα 5.3: Διάγραμμα ροής του υποσυστήματος ευρετηρίασης και αναζήτησης

5.1.4 Υποσύστημα Διεπαφής Χρήστη και Παρουσίασης Περιεχομένου

Η διεπαφή χρήστη αποτελεί το τελικό στάδιο του συστήματος και είναι υπεύθυνη για την παρουσίαση των κοινοβουλευτικών δεδομένων με τρόπο κατανοητό και εύχρηστο. Μέσα από μία σύγχρονη και φιλική διεπαφή, οι χρήστες μπορούν να περιηγηθούν στα πρακτικά της Βουλής, να αναζητήσουν συζητήσεις και ομιλίες με ποικίλα κριτήρια και να εξερευνήσουν τις παρεμβάσεις των ομιλητών σε βάθος.

Η διεπαφή οργανώνεται σε δύο κύριες ενότητες: **Αναζήτηση Ομιλητών** και **Αναζήτηση Πρακτικών**. Στην πρώτη, προβάλλονται τα προσωπικά στοιχεία και οι κοινοβουλευτικές συμμετοχές κάθε πολιτικού προσώπου, με δυνατότητα φιλτραρίσματος και εστιασμένης αναζήτησης στις ομιλίες του. Στη δεύτερη, εμφανίζονται τα μεταδεδομένα και το πλήρες περιεχόμενο κάθε συνεδρίασης, με δυνατότητα πλοήγησης και εφαρμογής φίλτρων κατά την ανάγνωση (όπως υψηλή συναισθηματική ένταση ή συγκεκριμένος ομιλητής), ώστε οι σχετικές ομιλίες να επισημαίνονται μέσα στο κείμενο και να εντοπίζονται εύκολα.

Η διεπαφή υποστηρίζει συνδυαστικές αναζητήσεις με φίλτρα όπως λέξεις-κλειδιά, θεματικές ενότητες, συναισθηματικός τόνος, ρητορική πρόθεση και χρονικά διαστήματα, ενώ τα αποτελέσματα παρουσιάζονται με τονισμένα τα πιο σχετικά αποσπάσματα για άμεσο εντοπισμό. Οι υποκατηγορίες των δύο βασικών ενότητων συμπληρώνουν τα αποτελέσματα, επιτρέποντας στον χρήστη να μελετήσει αναλυτικά τις παρεμβάσεις και να εντοπίζει ευκολότερα τις τάσεις, προθέσεις και επιχειρήματα που καταγράφονται στα πρακτικά.

5.2 Διασύνδεση και Ανάκτηση Δεδομένων μέσω API

Η ανάκτηση και προβολή των δεδομένων του συστήματος γίνεται μέσω ενός συνόλου API endpoints, που προσφέρουν πρόσβαση σε μεταδεδομένα και πλήρες κειμενικό περιεχόμενο. Τα endpoints αυτά χρησιμοποιούνται τόσο από το frontend για την παρουσίαση των δεδομένων, όσο και από εξωτερικά συστήματα.

Για τα API requests που απαιτούν αυθεντικοποίηση, χρησιμοποιείται **Bearer Token** στο πεδίο Authorization των HTTP headers, μαζί με Content-Type: application/json.

Τα API endpoints διακρίνονται σε δύο κατηγορίες: προσαρμοσμένα endpoints στο Next.js backend για σύνθετα ή επεξεργασμένα αποτελέσματα, και προκαθορισμένα GET endpoints του συστήματος διαχείρισης περιεχομένου για άμεση πρόσβαση στις βασικές οντότητες.

Στον Πίνακα 5.1 παρουσιάζεται συνοπτικά το σύνολο των βασικών διαθέσιμων API endpoints, με σύντομη περιγραφή της λειτουργικότητάς τους.

Endpoint	Σύντομη περιγραφή
/api/search-debates	Αναζητά συζητήσεις της Βουλής βάσει φίλτρων (π.χ. λέξη-κλειδί, θεματική, χρονική περίοδος, συναίσθημα, ρητορική πρόθεση). Επιστρέφει λίστα συνεδριάσεων και χαρακτηριστική ομιλία ανά συνεδρίαση.
/api/search-in-debate	Αναζητά συγκεκριμένες ομιλίες μέσα σε μια συνεδρίαση, βάσει φίλτρων (π.χ. ομιλητής, συναίσθημα, ένταση, ρητορική πρόθεση, φράση-κλειδί).
/api/debates/metadata/[id]	Επιστρέφει τα μεταδεδομένα μιας συγκεκριμένης συνεδρίασης της Βουλής, όπως περίοδος, σύννοδος, ημερομηνία, θεματικές και συμμετέχοντες.
/api/search-speakers	Αναζητά ομιλητές με βάση φίλτρα (π.χ. φύλο, ηλικία, κόμμα) και επιστρέφει τον πιο αντιπροσωπευτικό λόγο τους για φίλτρα αναζήτησης ομιλιών (π.χ. συναίσθημα).
/api/search-speaker-speeches	Ανακτά συνεδριάσεις στις οποίες συμμετείχε ένας ομιλητής, μαζί με την πιο χαρακτηριστική ομιλία του για κάθε συνεδρίαση.
/api/speakers/metadata/[id]	Παρέχει τα βιογραφικά στοιχεία ενός ομιλητή (όνομα, ηλικία, εκπαίδευση, γλώσσες, κόμμα, φωτογραφία).
Δημόσια GET Endpoints	Προκαθορισμένα GET endpoints του συστήματος για βασικές οντότητες (speakers, debates, speeches, political parties, topics), χωρίς αυθεντικοποίηση.

Πίνακας 5.1: Συνοπτική παρουσίαση των βασικών API endpoints της εφαρμογής.

Στις υποενότητες που ακολουθούν, κάθε endpoint περιγράφεται αναλυτικά, με τις παραμέτρους που υποστηρίζει, ενδεικτικές απαντήσεις και τη λειτουργία του.

5.2.1 /api/search-debates

Το /api/search-debates επιτρέπει την εύρεση κοινοβουλευτικών συζητήσεων που ικανοποιούν συγκεκριμένα κριτήρια αναζήτησης, όπως θεματική ενότητα, χρονικό διάστημα, συναίσθημα, ρητορική πρόθεση και φράση-κλειδί. Επιστρέφει λίστα συζητήσεων μαζί με

μία αντιπροσωπευτική ομιλία από κάθε συζήτηση, εφόσον εντοπίζεται η φράση-κλειδί ή η το συναισθηματικό φορτίο.

Παράμετρος	Τύπος	Περιγραφή
keyPhrase	String	Φράση-κλειδί που αναζητείται στο περιεχόμενο των ομιλιών.
rhetoricalIntent	String	Ρητορική πρόθεση (π.χ. Criticize/Denounce).
sentiment	String	Επίπεδο συναισθήματος (π.χ. very_negative).
highIntensity	Boolean	Επιστρέφονται μόνο ομιλίες υψηλής έντασης (true/false).
startDate	Date	Ημερομηνία έναρξης εύρους αναζήτησης.
endDate	Date	Ημερομηνία λήξης εύρους αναζήτησης.
session	String	Συγκεκριμένη κοινοβουλευτική σύνοδος.
topics	Array	Λίστα θεματικών ενοτήτων.
speakers	Array	Λίστα συμμετεχόντων ομιλητών.
sortBy	String	Κριτήριο ταξινόμησης (newest ή oldest).
page	Integer	Αριθμός σελίδας αποτελεσμάτων.
pageSize	Integer	Πλήθος αποτελεσμάτων ανά σελίδα.

Πίνακας 5.2: Παράμετροι αιτήματος στο endpoint /api/search-debates

Η απάντηση είναι μία λίστα από συζητήσεις που ικανοποιούν τα φίλτρα, συνοδευόμενες από την «καλύτερη» ομιλία που περιέχει τη φράση-κλειδί (όπου υπάρχει). Περιλαμβάνει επίσης συνολικό αριθμό αποτελεσμάτων και σελίδων.

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.4.

```

1  {
2    "debates": [
3      {
4        "id": 9553,
5        "documentId": "3cbbd69a-12e8-487e-aa8e-e28862cc1b97",
6        "date": "2023-11-06",
7        "title": "ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 2023-11-06",
8        "session": "Σύνοδος Α΄ Σύνοδος (03/07/2023 - 06/10/2024)",
9        "topics": ["Υγεία και Κοινωνική Πρόνοια"],
10       "top_speech": {
11         "speaker_name": "ΑΔΑΜΑΝΤΙΑ ΜΑΝΩΛΑΚΟΥ",
12         "content": "Οι συνέπειες είναι άμεσες... ",
13         "sentiment_label": "very_negative",
14         "rhetorical_intent": "Criticize/Denounce"
15       }
16     },
17     ...
18   ],
19   "totalPages": 1,
20   "totalDebates": 15
21 }

```

Σχήμα 5.4: Ενδεικτική απάντηση του endpoint /api/search-debates

5.2.2 /api/search-in-debate

Το `/api/search-in-debate` επιτρέπει την αναζήτηση ομιλιών εντός μιας συγκεκριμένης συζήτησης (debate) με βάση κριτήρια όπως συναισθημα, ρητορική πρόθεση, φράση-κλειδί, υψηλή ένταση και ομιλητές. Επιστρέφει τις ομιλίες της συζήτησης που πληρούν τα επιλεγμένα φίλτρα.

Παράμετρος	Τύπος	Περιγραφή
debateId	String	Το μοναδικό αναγνωριστικό της συζήτησης.
keyPhrase	String	Φράση-κλειδί που αναζητείται στο περιεχόμενο των ομιλιών.
sentiment	String	Επίπεδο συναισθήματος (π.χ. neutral, positive).
rhetoicalIntent	String	Ρητορική πρόθεση (π.χ. Inform/Explain).
highIntensity	Boolean	Επιστρέφονται μόνο ομιλίες υψηλής έντασης (true/false).
speakers	Array	Λίστα ονομάτων ομιλητών.

Πίνακας 5.3: Παράμετροι αιτήματος στο endpoint `/api/search-in-debate`

Η απάντηση είναι μία λίστα από ομιλίες της συζήτησης που ικανοποιούν τα φίλτρα, με βασικά στοιχεία όπως `speaker_name`, `content`, `speech_number`, `sentiment`, `rhetoical_intent` και `high_intensity`.

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.5.

```

1  {
2    "speeches": [
3      {
4        "id": "debate_2023-11-01_1_speech_1",
5        "speaker_name": "ΓΕΩΡΓΙΟΣ ΓΕΩΡΓΑΝΤΑΣ",
6        "speaker_id": "f5e264b6-f7d1-444b-a17a-47bbca7e52df",
7        "content": "Κυρίες και κύριοι συνάδελφοι, καλημέρα. Αρχίζει η συνεδρίαση...",
8        "speech_number": 1,
9        "sentiment_label": "neutral",
10       "rhetoical_intent": "Inform/Explain",
11       "high_intensity": false
12     },
13     {
14       "id": "debate_2023-11-01_1_speech_201",
15       "speaker_name": "ΑΘΑΝΑΣΙΟΣ ΜΠΟΥΡΑΣ",
16       "speaker_id": "91f4311f-b9a8-4bb8-ba4e-b6c491b2a595",
17       "content": "Δεν μπορείτε να κάνετε παρέμβαση σε αυτό το στάδιο. Θα απαντήσετε...",
18       "speech_number": 201,
19       "sentiment_label": "neutral",
20       "rhetoical_intent": "Inform/Explain",
21       "high_intensity": false
22     }
23   ]
24 }
25 
```

Σχήμα 5.5: Ενδεικτική απάντηση του endpoint `/api/search-in-debate`

5.2.3 /api/debates/metadata/[id]

Το `/api/debates/metadata/[id]` επιστρέφει τα πλήρη μεταδεδομένα μίας συγκεκριμένης συζήτησης στη Βουλή. Περιλαμβάνει τίτλο, ημερομηνία, περίοδο, σύνοδο, περίληψη, θεμα-

τικές ενότητες και λίστα των ομιλητών που συμμετείχαν, μαζί με βασικά στοιχεία και (όπου υπάρχουν) εικόνες τους.

Παράμετρος	Τύπος	Περιγραφή
id	String	Το documentId της συζήτησης, το οποίο χρησιμοποιείται για την ανάκτηση των μεταδεδομένων.

Πίνακας 5.4: Παράμετροι αιτήματος στο endpoint `/api/debates/metadata/[id]`

Η απάντηση είναι ένα αντικείμενο με βασικά μεταδεδομένα της συζήτησης, μία περίληψη, τις θεματικές ενότητες και τη λίστα των ομιλητών με τα documentId, ονόματα και συνδέσμους εικόνων τους (εάν υπάρχουν).

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.6.

```

1 {
2   "debate": {
3     "documentId": "c94db788-5126-48bc-9ed0-f63d5c8afe37b",
4     "title": "ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 1992-06-10",
5     "opening_section": "Αθήνα, σήμερα στις 10 Ιουνίου 1992, ημέρα Τετάρτη και ώρα 18,46' συνήλθε στην Αίθουσα συνεδριάσεων του Κοινοβουλίου . . .",
6     "summary": "Το έγγραφο αναλύει εκτενώς τα τρέχοντα κοινοβουλευτικά ζητήματα και τις κυβερνητικές απαντήσεις σε πολλαπλά επίπεδα . . .",
7     "period": "Ζ' ΠΕΡΙΟΔΟΣ",
8     "session_date": "Τετάρτη 10 Ιουνίου 1992",
9     "session": "Σύννοδος Β' (07/10/1991 - 19/06/1992)",
10    "meeting": "ΣΥΝΕΔΡΙΑΣΗ ΡΜΘ",
11    "topics": [
12      { "name": "Δικαιοσύνη και Ανθρώπινα Δικαιώματα" },
13      { "name": "Οικονομική Πολιτική και Δημοσιονομικά" }
14    ],
15    "speakers": [
16      {
17        "documentId": "f60b460c-1141-484c-9a6b-91ad4e504b87",
18        "speaker_id": "basilis_kedikoglou",
19        "speaker_name": "ΒΑΣΙΛΗΣ ΚΕΔΙΚΟΓΛΟΥ",
20        "image": {
21          "url": "/uploads/kedikoglou.jpg"
22        }
23      },
24      {
25        "documentId": "b4cca003-eedc-452d-817f-c69d6a732e57",
26        "speaker_id": "ioannis_skoularikis",
27        "speaker_name": "ΙΩΑΝΝΗΣ ΣΚΟΥΛΑΡΙΚΗΣ",
28        "image": {
29          "url": "/uploads/skoularikis.jpg"
30        }
31      },
32      {
33        "documentId": "56a0a19b-7d3c-4c61-b2a1-5278f7c8da27",
34        "speaker_id": "dimitrios_androulakis",
35        "speaker_name": "ΔΗΜΗΤΡΙΟΣ ΑΝΔΡΟΥΛΑΚΗΣ",
36        "image": null
37      }
38    ]
39  }
40 }
41
```

Σχήμα 5.6: Ενδεικτική απάντηση του endpoint `/api/debates/metadata/[id]`

5.2.4 /api/search-speakers

Το `/api/search-speakers` επιτρέπει την αναζήτηση ομιλητών που πληρούν συγκεκριμένα κριτήρια, όπως ηλικία, φύλο, πολιτικό κόμμα, φράση-κλειδί και άλλα χαρακτηριστικά. Επιστρέφει λίστα ομιλητών, καθένας από τους οποίους συνοδεύεται από την πιο αντιπροσωπευτική του ομιλία και τη συζήτηση στην οποία ανήκει.

Παράμετρος	Τύπος	Περιγραφή
keyPhrase	String	Φράση-κλειδί που αναζητείται στο περιεχόμενο των ομιλιών.
sentiment	String	Επίπεδο συναισθήματος (π.χ. very_negative).
rhetoricalIntent	String	Ρητορική πρόθεση (π.χ. Criticize/Denounce).
highIntensity	Boolean	Επιστρέφονται μόνο ομιλίες υψηλής έντασης (true/false).
speakerName	String	Όνομα συγκεκριμένου ομιλητή.
parties	Array	Λίστα πολιτικών κομμάτων.
gender	String	Φύλο ομιλητή (άνδρας/γυναίκα).
ageRange	Object	Εύρος ηλικίας με πεδία min και max.
page	Integer	Αριθμός σελίδας αποτελεσμάτων.

Πίνακας 5.5: Παράμετροι αιτήματος στο endpoint /api/search-speakers

Η απάντηση είναι μία λίστα από ομιλητές που ικανοποιούν τα φίλτρα, συνοδευόμενοι από μία κορυφαία ομιλία τους και τα βασικά στοιχεία της συζήτησης στην οποία ανήκει. Επιστρέφεται επίσης συνολικός αριθμός αποτελεσμάτων και σελίδων.

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.7.

```

1 {
2   "result": [
3     {
4       "documentId": "c5f3ae3b-943d-4a26-acaa-92bd00ac89d0",
5       "speaker_name": "ΑΘΑΝΑΣΙΟΣ ΔΗΜΗΤΡΑΚΟΠΟΥΛΟΣ",
6       "age": 86,
7       "gender": "άνδρας",
8       "imageUrl": "/uploads/thanasis_dimitrakopoulos_eurokinissi_7411309df8.jpg",
9       "top_speech": {
10        "id": "debate_1992-06-10_1_speech_247",
11        "score": 1.52,
12        "speaker_name": "ΑΘΑΝΑΣΙΟΣ ΔΗΜΗΤΡΑΚΟΠΟΥΛΟΣ",
13        "content": "Το δημόσιο χρέος να τραβάει την ανηφόρα... η <span style='font-weight: bold; color: #f9d342;'>ανεργία</span> να καλπάζει...",
14        "sentiment_label": "",
15        "rhetorical_intent": "",
16        "high_intensity": false
17      },
18      "debate": {
19        "documentId": "c94db788-5126-48bc-9ed0-f63d5c4fe37b",
20        "date": "1992-06-10",
21        "title": "ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 1992-06-10",
22        "session": "Σύνοδος Β' (07/10/1991 - 19/06/1992)",
23        "period": "Ζ' ΠΕΡΙΟΔΟΣ"
24      }
25    },
26    ...
27  ],
28  "totalPages": 2,
29  "totalSpeakers": 82
30 }

```

Σχήμα 5.7: Ενδεικτική απάντηση του endpoint /api/search-speakers

5.2.5 /api/search-speaker-speeches

Το /api/search-speaker-speeches επιστρέφει τις συζητήσεις στις οποίες συμμετείχε ένας συγκεκριμένος ομιλητής, μαζί με την πιο αντιπροσωπευτική ομιλία του σε κάθε συζήτηση. Υποστηρίζει φίλτρα όπως φράση-κλειδί, θέματα, χρονικό διάστημα, συναίσθημα, ρητορική πρόθεση και ένταση.

Παράμετρος	Τύπος	Περιγραφή
speakerId	String	Το μοναδικό αναγνωριστικό του ομιλητή.
keyPhrase	String	Φράση-κλειδί που αναζητείται στο περιεχόμενο των ομιλιών.
sentiment	String	Επίπεδο συναισθήματος.
rhetoricalIntent	String	Ρητορική πρόθεση.
highIntensity	Boolean	Επιστρέφονται μόνο ομιλίες υψηλής έντασης (true/false).
startDate	String	Ημερομηνία έναρξης χρονικού φίλτρου (YYYY-MM-DD).
endDate	String	Ημερομηνία λήξης χρονικού φίλτρου (YYYY-MM-DD).
topics	Array	Λίστα θεμάτων.
session	String	Συγκεκριμένη σύνοδος.
sortBy	String	Ταξινόμηση (newest/oldest).
page	Integer	Αριθμός σελίδας αποτελεσμάτων.

Πίνακας 5.6: Παράμετροι αιτήματος στο endpoint `/api/search-speaker-speeches`

Η απάντηση είναι μία λίστα από συζητήσεις όπου συμμετείχε ο ομιλητής, κάθε μία συνοδευόμενη από την πιο χαρακτηριστική ομιλία του στη συγκεκριμένη συζήτηση, μαζί με βασικά στοιχεία για τη συζήτηση (τίτλος, ημερομηνία, θέματα).

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.8.

```

1  {
2    "debates": [
3      {
4        "id": 12449,
5        "documentId": "34b01bce-b13e-4e79-92bf-0b74f716e2e1",
6        "date": "2014-02-06",
7        "title": "ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 2014-02-06",
8        "session": "Σύνοδος Β' Σύνοδος (07/10/2013 - 04/06/2014)",
9        "period": "ΙΕ' ΠΕΡΙΟΔΟΣ",
10       "topics": ["Υγεία και Κοινωνική Πρόνοια"],
11       "top_speech": {
12         "id": "debate_2014-02-06_1_speech_123",
13         "score": 8.41,
14         "speaker_name": "ΙΩΑΝΝΗΣ ΚΕΦΑΛΟΓΙΑΝΝΗΣ",
15         "content": "Εάν λάβουμε υπ' όψιν την καλπάζουσα <span style=\\"font-weight: bold; color: #f9d342; \">ανεργία</span>...",
16         "sentiment_label": "",
17         "rhetorical_intent": "",
18         "high_intensity": false
19       }
20     },
21     {
22       "id": 12416,
23       "documentId": "1094634e-68be-48a5-aea8-dc72ddd496bf",
24       "date": "2013-12-06",
25       "title": "ΠΡΑΚΤΙΚΑ ΒΟΥΛΗΣ 2013-12-06",
26       "session": "Σύνοδος Β' Σύνοδος (07/10/2013 - 04/06/2014)",
27       "period": "ΙΕ' ΠΕΡΙΟΔΟΣ",
28       "topics": ["Υγεία και Κοινωνική Πρόνοια", "Οικονομική Πολιτική και Δημοσιονομικά"],
29       "top_speech": {
30         "id": "debate_2013-12-06_1_speech_39",
31         "score": 8.68,
32         "speaker_name": "ΙΩΑΝΝΗΣ ΚΕΦΑΛΟΓΙΑΝΝΗΣ",
33         "content": "Δεν ακούστηκαν προτάσεις για την αξιοποίηση των περίπου 6 δισεκατομμυρίων ευρώ...",
34         "sentiment_label": "",
35         "rhetorical_intent": "",
36         "high_intensity": false
37       }
38     }
39   ],
40   "totalPages": 1,
41   "totalDebates": 6,
42   "speaker_name": "ΙΩΑΝΝΗΣ ΚΕΦΑΛΟΓΙΑΝΝΗΣ"
43 }
44

```

Σχήμα 5.8: Ενδεικτική απάντηση του endpoint `/api/search-speaker-speeches`

5.2.6 /api/speakers/metadata/[id]

Το `/api/speakers/metadata/[id]` επιστρέφει τα βασικά μεταδεδομένα ενός ομιλητή με βάση το μοναδικό `documentId` του. Περιλαμβάνει πληροφορίες όπως όνομα, ημερομηνία γέννησης, ηλικία, τόπο γέννησης, εκπαίδευση, πολιτικό κόμμα, γλώσσα, επάγγελμα και σύνδεσμο σε καταχώρηση Wikidata.

Παράμετρος	Τύπος	Περιγραφή
id	String	Το <code>documentId</code> του ομιλητή για τον οποίο ζητούνται τα μεταδεδομένα.

Πίνακας 5.7: Παράμετροι αιτήματος στο endpoint `/api/speakers/metadata/[id]`

Η απάντηση περιλαμβάνει τις βασικές πληροφορίες του ομιλητή, τη φωτογραφία, τη λίστα πολιτικών κομμάτων με τα οποία έχει συνδεθεί, καθώς και πρόσθετα στοιχεία όπως εκπαίδευση και επάγγελμα.

Ενδεικτική απάντηση (JSON) παρουσιάζεται στο Σχήμα 5.9.

```

1  {
2    "speaker": {
3      "documentId": "22e41a8c-15d3-407b-9743-c8d92f95b717",
4      "speaker_name": "ΑΓΓΕΛΟΣ ΤΟΛΚΑΣ",
5      "age": 47,
6      "date_of_birth": "1978-08-06",
7      "date_of_death": null,
8      "description": "Έλληνας πολιτικός",
9      "educated_at": "Aristotle University of Thessaloniki",
10     "languages": "νέα ελληνική γλώσσα",
11     "link": "https://www.wikidata.org/wiki/Q7338911",
12     "occupation": "πολιτικός",
13     "place_of_birth": "Νάουσα Ημαθίας",
14     "website": null,
15     "image": {
16       "url": "/uploads/tolkas_aggelos0989_b6ee1e6047.jpg",
17       "formats": {
18         "thumbnail": {
19           "name": "thumbnail_tolkas_aggelos0989.jpg",
20           "hash": "thumbnail_tolkas_aggelos0989_b6ee1e6047",
21           "ext": ".jpg",
22           "mime": "image/jpeg",
23           "width": 156,
24           "height": 156,
25           "size": 5.12,
26           "url": "/uploads/thumbnail_tolkas_aggelos0989_b6ee1e6047.jpg"
27         }
28       }
29     },
30     "political_parties": [
31       {
32         "name": "Πανελλήνιο Σοσιαλιστικό Κίνημα",
33         "image": {
34           "formats": null,
35           "url": "/uploads/image_663609b1b4.svg"
36         }
37       }
38     ]
39   }
40 }
```

Σχήμα 5.9: Ενδεικτική απάντηση του endpoint `/api/speakers/metadata/[id]`

5.2.7 Δημόσια GET Endpoints του Συστήματος Διαχείρισης Περιεχομένου

Το σύστημα διαχείρισης περιεχομένου της εφαρμογής παρέχει δημόσια πρόσβαση μέσω GET αιτημάτων σε όλες τις βασικές οντότητες του συστήματος: Speakers, Debates, Political Parties, Topics και Speeches. Η πρόσβαση αυτή επιτρέπει την ανάκτηση των σχετικών δεδομένων χωρίς αυθεντικοποίηση και αξιοποιείται τόσο στο frontend όσο και σε βοηθητικά σενάρια.

Για την εξειδίκευση των αποτελεσμάτων μπορούν να χρησιμοποιηθούν διάφορες παραμετροποιήσεις στο URL, όπως αυτές που φαίνονται στον ακόλουθο πίνακα:

Παράμετρος URL	Περιγραφή
?populate=*	Φόρτωση όλων των σχετικών οντοτήτων (relations) με την κύρια εγγραφή.
?filters[field][\$eq]=value	Φιλτράρισμα αποτελεσμάτων με βάση τιμές συγκεκριμένων πεδίων.
?pagination[limit]=-1	Ανάκτηση όλων των εγγραφών χωρίς σελιδοποίηση.
?fields[0]=fieldname	Ανάκτηση μόνο επιλεγμένων πεδίων της εγγραφής.

Πίνακας 5.8: Ενδεικτικές παράμετροι URL για δημόσιες κλήσεις στο API του συστήματος διαχείρισης περιεχομένου.

Αξίζει να σημειωθεί ότι, για λόγους ανεξαρτησίας από το εσωτερικό id που αποδίδεται αυτόματα από τη βάση δεδομένων, όλες οι αναζητήσεις και συνδέσεις αυτές βασίζονται στο πεδίο **documentId**, το οποίο ορίζεται ρητά και διατηρείται σταθερό μεταξύ διαφορετικών συστημάτων.

Για περισσότερες λεπτομέρειες σχετικά με τη σύνταξη και τις δυνατότητες του API, μπορεί να γίνει αναφορά στην επίσημη τεκμηρίωση του συστήματος διαχείρισης περιεχομένου³ που χρησιμοποιήθηκε.

³ Strapi CMS Documentation: <https://docs.strapi.io/dev-docs/api/rest>

Αξιολόγηση του Συστήματος

Στην παρούσα ενότητα περιγράφεται η μεθοδολογία που ακολουθήθηκε για την αξιολόγηση του συστήματος από τους χρήστες.

6.1 Μεθοδολογία Ελέγχου

Η αξιολόγηση της Ηλεκτρονικής Υπηρεσίας Δημοσίευσης των Συζητήσεων του Κοινοβουλίου πραγματοποιήθηκε με στόχο να αποτιμηθεί τόσο η χρηστικότητα όσο και η αποδοχή της τεχνολογίας από διαφορετικές κατηγορίες χρηστών. Για την επίτευξη μιας πολυδιάστατης αποτίμησης υιοθετήθηκε μια μεικτή προσέγγιση, η οποία περιελάμβανε ποσοτικές και ποιοτικές τεχνικές συλλογής δεδομένων.

Στο επίκεντρο της αξιολόγησης βρέθηκαν δύο καθιερωμένα μοντέλα:

- Το System Usability Scale (SUS) [39], ένα συνοπτικό αλλά αξιόπιστο εργαλείο μέτρησης της αντιλαμβανόμενης χρηστικότητας.
- Το Technology Acceptance Model (TAM) [40], ένα θεωρητικό μοντέλο πρόβλεψης της αποδοχής τεχνολογιών από τους χρήστες.

Το SUS αποτελείται από 10 δηλώσεις και χρησιμοποιεί κλίμακα Likert¹ πέντε σημείων, όπου οι χρήστες εκφράζουν τον βαθμό συμφωνίας ή διαφωνίας τους. Ο συνολικός δείκτης (SUS score) κυμαίνεται από 0 έως 100 και δίνει μια συνολική εικόνα της χρηστικότητας του συστήματος, χωρίς όμως να εστιάζει σε συγκεκριμένα προβλήματα ή λειτουργίες. Χρησιμοποιείται ευρέως λόγω της ταχύτητας συμπλήρωσής του και της ευκολίας ερμηνείας του, ενώ έχει αποδειχθεί αξιόπιστο ακόμα και σε δείγματα μικρού μεγέθους.

Το TAM, από την άλλη, είναι ένα θεωρητικό εργαλείο βασισμένο στην Theory of Reasoned Action, που εστιάζει στις αντιλήψεις των χρηστών σχετικά με την τεχνολογία και πώς αυτές επηρεάζουν την πρόθεση χρήσης. Το μοντέλο αναλύει τέσσερις βασικές διαστάσεις:

- **Perceived Usefulness (PU):** Η πεποίθηση ότι η χρήση του συστήματος βελτιώνει την απόδοση ή αποτελεσματικότητα.
- **Perceived Ease of Use (PEOU):** Η πεποίθηση ότι το σύστημα είναι εύκολο στη χρήση.
- **Attitude Toward Use (ATU):** Η συνολική στάση απέναντι στη χρήση του συστήματος.

¹ Likert: https://en.wikipedia.org/wiki/Likert_scale

- **Behavioral Intention to Use (BI):** Η πρόθεση του χρήστη να χρησιμοποιήσει το σύστημα στο μέλλον.

Το ερωτηματολόγιο TAM που χρησιμοποιήθηκε περιλάμβανε 15 δηλώσεις με κλίμακα 5 σημείων, ώστε να καλυφθούν οι παραπάνω διαστάσεις καθώς και δείκτες πραγματικής χρήσης (Actual System Use), σε περίπτωση επαναλαμβανόμενων χρηστών.

Η διαδικασία αξιολόγησης οργανώθηκε σε πέντε διακριτές φάσεις, καθεμία από τις οποίες αντιστοιχεί σε μία από τις ενότητες του ερωτηματολογίου :

- **Pre-Test Questionnaire:** Καταγραφή βασικών δημογραφικών στοιχείων, επαγγελματικού υποβάθρου, εμπειρίας με κοινοβουλευτικά δεδομένα, επιπέδου τεχνολογικής επάρκειας και αρχικών προσδοκιών από την πλατφόρμα.
- **Task Scenarios:** Εκτέλεση τριών ρεαλιστικών σεναρίων χρήσης που αξιολογούν την ευχρηστία και αποτελεσματικότητα βασικών λειτουργιών του συστήματος, όπως η αναζήτηση, πλοήγηση και αξιοποίηση μεταδεδομένων.
- **System Usability Scale (SUS):** Συμπλήρωση του πρότυπου ερωτηματολογίου SUS, που αποτελείται από 10 δηλώσεις και μετρά τη συνολική εμπειρία χρηστικότητα του συστήματος σε κλίμακα 1-5.
- **Technology Acceptance Model (TAM):** Αξιολόγηση της αποδοχής της τεχνολογίας μέσω δηλώσεων που καλύπτουν τέσσερις βασικές διαστάσεις: Perceived Usefulness, Perceived Ease of Use, Attitude Toward Using, και Behavioral Intention to Use.
- **Post-Test Interview Guide:** Συλλογή ποιοτικής ανατροφοδότησης μέσω ανοιχτού τύπου ερωτήσεων που διερευνούν τις εντυπώσεις, τις δυσκολίες, τις προτάσεις βελτίωσης και τη γενική στάση του χρήστη απέναντι στο σύστημα.

Η συνδυασμένη χρήση των εργαλείων TAM και SUS παρείχε μια πλήρη εικόνα για την εμπειρία και πρόθεση χρήσης του συστήματος, συνδυάζοντας υποκειμενικές απόψεις με αντικειμενικά δεδομένα απόδοσης.

6.2 Ερωτηματολόγιο Αξιολόγησης

Το ερωτηματολόγιο αξιολόγησης σχεδιάστηκε ως ενιαία ηλεκτρονική φόρμα, με σκοπό τη συλλογή ποσοτικών και ποιοτικών δεδομένων σχετικά με την εμπειρία χρήσης της πλατφόρμας. Η δομή του βασίστηκε στα καθιερωμένα μοντέλα System Usability Scale (SUS) και Technology Acceptance Model (TAM), και εμπλουτίστηκε με σενάρια χρήσης, ερωτήσεις δημογραφικών χαρακτηριστικών και ανοιχτές ερωτήσεις για ανατροφοδότηση.

Η φόρμα είναι προσβάσιμη στον ακόλουθο σύνδεσμο: [Evaluation Form](#).

Η παρουσίαση που ακολουθεί αποτυπώνει αναλυτικά το περιεχόμενο κάθε ενότητας του ερωτηματολογίου, με τις ερωτήσεις και τις διαθέσιμες επιλογές απάντησης, όπως εμφανίζονται στην ηλεκτρονική φόρμα. Κάθε ενότητα συνοδεύεται από σύντομη περιγραφή του σκοπού της, ενώ για λόγους αναγνωσιμότητας οι ερωτήσεις έχουν οργανωθεί σε πίνακες κατά θεματική κατηγορία.

6.2.1 Ενότητα 1: Pre-Test Questionnaire

Η πρώτη ενότητα του ερωτηματολογίου αποσκοπεί στη συλλογή βασικών πληροφοριών για το προφίλ των συμμετεχόντων πριν την αλληλεπίδρασή τους με το σύστημα. Συγκεκριμένα, περιλαμβάνει ερωτήσεις δημογραφικού χαρακτήρα (ηλικία, φύλο, επίπεδο εκπαίδευσης, επαγγελματικό υπόβαθρο), ερωτήσεις για την εμπειρία τους με κοινοβουλευτικά δεδομένα, την εξοικείωσή τους με την τεχνολογία, καθώς και τις προσδοκίες τους από μια σχετική πλατφόρμα. Η ενότητα αυτή μας επιτρέπει να κατηγοριοποιήσουμε τους χρήστες και να ερμηνεύσουμε τα αποτελέσματα της αξιολόγησης με μεγαλύτερη ακρίβεια.

Demographic Information	
What is your age group?	18-24, 25-34, 35-44, 45-54, 55-64, 65+
What is your gender?	Male, Female, Prefer not to say
What is your education level?	High school, Bachelor's degree, Master's degree, Doctorate, Other
What is your professional background?	Open-ended
Experience with Parliamentary Records	
How often do you access parliamentary records?	Never, Rarely, Occasionally, Frequently, Very frequently
What sources do you currently use to access parliamentary information?	Official parliament website, News media, Academic databases, Social media, Other
For what purposes do you typically seek parliamentary information?	Research, Professional work, Educational purposes, Personal interest, Civic engagement, Other
Technical Proficiency	
How would you rate your general comfort level with technology?	Very uncomfortable, Somewhat uncomfortable, Neutral, Somewhat comfortable, Very comfortable
How often do you use search engines?	Never, Rarely, Monthly, Weekly, Daily, Multiple times daily
How often do you use digital archives or databases?	Never, Rarely, Monthly, Weekly, Daily
Expectations	
What features would you expect in a parliamentary records portal?	Open-ended
What challenges do you currently face when accessing parliamentary information?	Open-ended
What would make a parliamentary records portal valuable to you?	Open-ended

Πίνακας 6.1: Ερωτηματολόγιο Πριν τη Δοκιμή: Ερωτήσεις και Επιλογές Απαντήσεων

6.2.2 Ενότητα 2: Task Scenarios

Η δεύτερη ενότητα του ερωτηματολογίου περιλαμβάνει τρία σενάρια χρήσης που εξετάζουν βασικές λειτουργίες της πλατφόρμας, όπως η αναζήτηση συνεδριάσεων και ομιλιών, η πρόσβαση σε μεταδεδομένα και πλήρες περιεχόμενο, η χρήση φίλτρων, και η αναζήτηση

εντός συζητήσεων. Ο στόχος αυτής της ενότητας είναι η αξιολόγηση της χρηστικότητας και της αποτελεσματικότητας του περιβάλλοντος χρήσης κατά την εκτέλεση ρεαλιστικών εργασιών. Κάθε σενάριο συνοδεύεται από συγκεκριμένα κριτήρια επιτυχίας, βάσει των οποίων εκτιμάται η επιτυχής ολοκλήρωσή του.

Task 1: Search Debates and Explore Proceedings Scenario: "You heard that an conversation about 'Δικαιοσύνη και Ανθρώπινα Δικαιώματα' happened between two main political figures and want to read more about it. Therefore: • Research parliamentary debates where both speakers 'ΚΥΡΙΑΚΟΣ ΜΗΤΣΟΤΑΚΗΣ' and 'ΑΛΕΞΗΣ ΤΣΙΠΡΑΣ' participated. • Focus on the years 2015-2017 (01/01/2015 – 31/12/2017). • Find debates discussing 'Δικαιοσύνη και Ανθρώπινα Δικαιώματα'. • Find and review the oldest of these debates (should be 2015-04-06). • In that debate, filter speaker 'ΚΥΡΙΑΚΟΣ ΜΗΤΣΟΤΑΚΗΣ' to highlight all his speeches. Success Criteria: • User Successfully navigates to search-debates page • User applies participating speakers, date range and topics filter successfully • User sorts debates by date • User navigates to the full proceedings of the first result debate (2015-04-06) • User applies speaker filter and locates the first highlighted speech
Task 2: Debate Enrichment Search Scenario: "In February 2021 there were intense discussions on education. You care about all the supporting opinions. Therefore: • Find speeches mentioning the key-phrase 'εκπαιδευτικό σύστημα'. • With rhetorical intent 'Πειθώ / Υποστήριξη' and very positive sentiment 'Πολύ θετικό'. • Filter between 01/02/2021 – 28/02/2021. • Click the first debate of that period (should be 2021-02-09) and read the summary. Success Criteria: • User navigates to search-debates page • Filters by key-phrase • Filters by rhetorical intent and sentiment • Filters by date range • Navigates to metadata of the debate and reads the summary (2021-02-09)
Task 3: Dive into Political Figures Scenario: "You want to search parliamentary speakers based on filters. Specifically: • Find all 'male' speakers belonging to political party 'Νέα Δημοκρατία' and under 60 years old. • Identify which of them mentioned 'άνεργία' in their speeches. • Locate speaker 'ΙΩΑΝΝΗΣ ΚΕΦΑΛΟΓΙΑΝΝΗΣ' and review educational background. • Find all debates this speaker participated in and count in how many he explicitly mentioned 'άνεργία'. Success Criteria: • User navigates to search-speakers page • Filters by political party, age range, and gender • Filters by key-phrase • Navigates to speaker bio and locates educational background • Navigates to speaker debates/speeches • Applies key-phrase filter and locates total occurrences

Πίνακας 6.2: Σενάρια Χρήσης: Περιγραφή και Κριτήρια Επιτυχίας

6.2.3 Ενότητα 3: System Usability Scale (SUS)

Η τρίτη ενότητα του ερωτηματολογίου περιλαμβάνει την πρότυπη κλίμακα System Usability Scale (SUS), η οποία χρησιμοποιείται για τη μέτρηση της συνολικής ευχρηστίας της πλατφόρμας, όπως αντιλαμβάνονται οι χρήστες.

Question
1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Πίνακας 6.3: Ερωτήσεις Κλίμακας Χρησικότητας Συστήματος (SUS)

6.2.4 Ενότητα 4: Technology Acceptance Model (TAM)

Η τέταρτη ενότητα του ερωτηματολογίου βασίζεται στο θεωρητικό πλαίσιο Technology Acceptance Model (TAM) και εξετάζει τους παράγοντες που διαμορφώνουν την πρόθεση των χρηστών να χρησιμοποιήσουν την πλατφόρμα.

Perceived Usefulness (PU)
Using the portal would enhance my effectiveness in accessing relevant parliamentary records.
The portal helps me complete research tasks more easily and efficiently.
I would find the portal useful in my academic, professional, or personal activities related to parliamentary information.
Using the portal would improve the overall quality of my work involving parliamentary information.
Perceived Ease of Use (PEOU)
Learning to operate the portal would be easy for me.
I would find it easy to get the portal to do what I want it to do.
The portal's interface would be intuitive and easy to navigate.
I would find the search and filtering features easy to use.
It would be easy for me to become skillful at using the portal.
Attitude Toward Using (ATU)
Using the portal is a good idea.
Using the portal would be pleasant.
Using the portal would make working with parliamentary records more interesting.
Behavioral Intention to Use (BI)
I intend to use the Greek Parliament Debates Portal when it becomes available.
I plan to use the portal regularly.
I would recommend the portal to others who need parliamentary information.

Πίνακας 6.4: Ερωτήσεις Μοντέλου Αποδοχής Τεχνολογίας (TAM) ανά Διάσταση

6.2.5 Ενότητα 5: Post-Test Interview Guide

Η πέμπτη και τελευταία ενότητα του ερωτηματολογίου περιλαμβάνει ανοικτού τύπου ερωτήσεις που στοχεύουν στη συλλογή ποιοτικής ανατροφοδότησης από τους χρήστες, μετά την ολοκλήρωση της αλληλεπίδρασής τους με την πλατφόρμα. Οι ερωτήσεις έχουν οργανωθεί σε θεματικές κατηγορίες που καλύπτουν τις γενικές εντυπώσεις του χρήστη, την εμπειρία του με τα σενάρια χρήσης, την αποτίμηση συγκεκριμένων λειτουργιών, καθώς και τη συνολική του τοποθέτηση συγκριτικά με άλλες μεθόδους πρόσβασης στην κοινοβουλευτική πληροφορία.

General Impressions
What are your overall impressions of the Greek Parliament Session Records Portal?
What aspects of the portal did you find most useful?
What aspects did you find most challenging or confusing?
Task Reflection
Which tasks did you find easiest or most difficult to complete and why?
Feature Assessment
How would you rate the effectiveness and intuitiveness of the search and filtering system?
How would you rate the presentation of parliamentary records?
What additional features would you like to see in the portal?
Comparison and Context
How does this portal compare to other methods you've used to access parliamentary information?
Would this portal meet your needs for parliamentary information? Why or why not?
Final Thoughts
What are your suggestions for improving the portal?

Πίνακας 6.5: Οδηγός Συνέντευξης Μετά τη Δοκιμή: Ανοιχτές Ερωτήσεις ανά Κατηγορία

6.3 Συμπεράσματα από την Αξιολόγηση του Συστήματος

Στην αξιολόγηση συμμετείχαν συνολικά 12 άτομα, τα οποία συμπλήρωσαν το ερωτηματολόγιο και ολοκλήρωσαν τα σενάρια χρήσης. Η ανάλυση που ακολουθεί βασίζεται στις απαντήσεις τους και παρουσιάζει τα βασικά ευρήματα ανά ενότητα.

Pre-Test Questionnaire

Η ανάλυση του Pre-Test Questionnaire ανέδειξε χρήσιμες πληροφορίες για το προφίλ και τις προσδοκίες των χρηστών. Στην πλειοψηφία τους, οι συμμετέχοντες είναι εξοικειωμένοι με την τεχνολογία και διαθέτουν υψηλή τεχνική επάρκεια, ωστόσο έχουν περιορισμένη εμπειρία με κοινοβουλευτικά δεδομένα και βασίζονται κυρίως σε μέσα ενημέρωσης και κοινωνικά δίκτυα για την ενημέρωσή τους.

Οι κυριότερες προσδοκίες των ερωτηθέντων για τις δυνατότητες του συστήματος αφορούν:

- Ισχυρά και εύχρηστα εργαλεία αναζήτησης και φιλτραρίσματος.
- Ευανάγνωστα και αξιόπιστα πρακτικά, με δυνατότητες σχολιασμού και κοινοποίησης.
- Σαφή, γρήγορη και πλήρη πληροφόρηση, με διασύνδεση συζητήσεων, νομοθεσίας και περιλήψεις σε απλή γλώσσα.

Συνολικά, οι χρήστες επιθυμούν μια πλατφόρμα που να υπερνικά τα προβλήματα των υφιστάμενων λύσεων, όπως η σύγχυση και οι καθυστερήσεις, προσφέροντας ακρίβεια, ταχύτητα και ευχρηστία στην παρουσίαση της κοινοβουλευτικής πληροφορίας.

Task Scenarios

Στην ενότητα των σεναρίων χρήσης, οι συμμετέχοντες κλήθηκαν να ολοκληρώσουν τρεις ρεαλιστικές εργασίες, αξιολογώντας βασικές λειτουργίες της πλατφόρμας όπως η αναζήτηση, η πλοήγηση και η χρήση μεταδεδομένων.

Οι χρήστες επέδειξαν αποτελεσματική κατανόηση και χρήση του συστήματος, ολοκληρώνοντας με επιτυχία τα περισσότερα από τα καθορισμένα κριτήρια επιτυχίας. Η μόνη αξιοσημείωτη δυσκολία εντοπίστηκε στην κατανόηση και χρήση των πιο σύνθετων φίλτρων, ειδικά εκείνων που σχετίζονται με τα χαρακτηριστικά NLP, όπως η ρητορική πρόθεση του ομιλητή.

Αποτελέσματα SUS

Η κλίμακα SUS κατέγραψε θετικά αποτελέσματα, υποδεικνύοντας υψηλή συνολική χρηστικότητα. Ο υπολογισμένος δείκτης SUS ανήλθε σε περίπου 85/100, τιμή που σύμφωνα με τη βιβλιογραφία αντιστοιχεί σε εξαιρετικό επίπεδο χρηστικότητας και υποδηλώνει υψηλή ικανοποίηση των χρηστών από τη διεπαφή και τη λειτουργικότητα του συστήματος.

Το σύστημα δεν κρίθηκε περιττά πολύπλοκο ή δυσνόητο, καθώς η συντριπτική πλειονότητα των συμμετεχόντων διαφώνησε με σχετικές αρνητικές δηλώσεις. Οι χρήστες ένιωσαν σίγουροι κατά τη χρήση, ενώ η ανάγκη για τεχνική υποστήριξη κρίθηκε ελάχιστη.

Η ενσωμάτωση των λειτουργιών αξιολογήθηκε θετικά, χωρίς να παρατηρηθούν σημαντικές ασυνέπειες. Συνολικά, τα αποτελέσματα της SUS υποδηλώνουν ότι το σύστημα είναι φιλικό προς τον χρήστη, εύκολα κατανοητό ακόμη και από αρχάριους, και πληροί υψηλά πρότυπα χρηστικότητας.

Αποτελέσματα TAM

Perceived Usefulness (PU) Οι απαντήσεις δείχνουν ότι οι χρήστες θεωρούν το σύστημα ιδιαίτερα χρήσιμο, καθώς διευκολύνει την πρόσβαση και την έρευνα σε κοινοβουλευτικά δεδομένα, βελτιώνοντας τόσο την ποιότητα όσο και την αποτελεσματικότητα της εργασίας τους. Αναφέρουν ότι μπορεί να καλύψει ανάγκες σε ακαδημαϊκό, επαγγελματικό και προσωπικό επίπεδο.

Perceived Ease of Use (PEOU) Η πλειονότητα των χρηστών βρήκε το σύστημα εύκολο στη χρήση, με κατανοητή και φιλική διεπαφή. Οι βασικές λειτουργίες, όπως η αναζήτηση και τα φίλτρα, κρίθηκαν απλές, ενώ εκφράστηκε η πεποίθηση ότι απαιτείται ελάχιστος χρόνος εξοικείωσης.

Attitude Toward Using (ATU) Οι χρήστες έχουν θετική στάση απέναντι στη χρήση του συστήματος, θεωρώντας ότι αποτελεί μια καλή και ευχάριστη ιδέα που κάνει την ενασχόληση με τα κοινοβουλευτικά δεδομένα πιο ενδιαφέρουσα.

Behavioral Intention to Use (BI) Τέλος, οι συμμετέχοντες δήλωσαν πρόθυμοι να χρησιμοποιήσουν το σύστημα όταν γίνει διαθέσιμο, ενώ αρκετοί εξέφρασαν την πρόθεση να το χρησιμοποιούν τακτικά και να το προτείνουν και σε άλλους.

Post-Test Interview

Οι τελικές ερωτήσεις μετά τη δοκιμή ανέδειξαν πολύ θετική εικόνα της πλατφόρμας. Οι χρήστες εκτίμησαν την καθαρή διεπαφή, την ευκολία χρήσης, την ταχύτητα και τη δυνατότητα φιλτραρίσματος. Βρήκαν ιδιαίτερα χρήσιμα τα φίλτρα αναζήτησης και την παρουσίαση των δεδομένων, ενώ δυσκολεύτηκαν μόνο σε πιο σύνθετα φίλτρα, όπως το φίλτρο συναισθηματικής φόρτισης.

Η αναζήτηση θεωρήθηκε ακριβής και γρήγορη, ενώ η παρουσίαση των δεδομένων ήταν σαφής και εύκολη στην κατανόηση. Ως προτάσεις βελτίωσης αναφέρθηκαν η προσθήκη περισσότερων φωτογραφιών πολιτικών, η ενσωμάτωση ενός AI assistant για βοήθεια κατά την πλοήγηση και την αναζήτηση, ή μιας λειτουργίας text-to-speech για ακουστική ανάγνωση των πρακτικών, καθώς και η καλύτερη τοποθέτηση των φίλτρων στις σελίδες αναζήτησης. Συνολικά, οι χρήστες θεώρησαν την πλατφόρμα καλύτερη από άλλες μεθόδους που έχουν χρησιμοποιήσει και δήλωσαν ότι κάλυψε πλήρως τις ανάγκες τους.

Τελικά Συμπεράσματα

Η αξιολόγηση της πλατφόρμας ανέδειξε μια ιδιαίτερα θετική εμπειρία χρήσης και αποδοχή του συστήματος από τους ερωτηθέντες/αξιολογητές. Παρά την περιορισμένη προηγούμενη ενασχόληση των συμμετεχόντων με κοινοβουλευτικά δεδομένα, χρησιμοποίησαν με ευκολία τις βασικές λειτουργίες, ολοκλήρωσαν επιτυχώς τα σενάρια και αξιοποίησαν αποτελεσματικά τα διαθέσιμα φίλτρα και μεταδεδομένα.

Η πλατφόρμα αποτιμήθηκε ως εύχρηστη, αποτελεσματική και κατάλληλα σχεδιασμένη για να καλύπτει ανάγκες αναζήτησης και έρευνας, με τους χρήστες να εκφράζουν θετική στάση και διάθεση να τη χρησιμοποιήσουν και να τη συστήσουν.

Εντοπίστηκαν κυρίως δυσκολίες στη χρήση των πιο σύνθετων φίλτρων, ενώ οι προτάσεις βελτίωσης αφορούσαν συμπληρωματικές δυνατότητες, όπως περισσότερες φωτογραφίες, ενσωμάτωση AI assistant και καλύτερη οργάνωση των φίλτρων.

Συνολικά, η πλατφόρμα ανταποκρίνεται στις αρχικές προσδοκίες των χρηστών, υπερβαίνει τις αδυναμίες των υφιστάμενων λύσεων και προσφέρει ένα αξιόπιστο και ελκυστικό εργαλείο για την αναζήτηση και επεξεργασία κοινοβουλευτικών δεδομένων, με περιθώρια βελτίωσης στις λεπτομέρειες της εμπειρίας χρήσης.

Επίλογος

Στην τελευταία αυτή ενότητα συνοψίζουμε παρουσιάζοντας συμπεράσματα σχετικά με το σύνολο της εργασίας καθώς και κάποια από τα βασικά προβλήματα που αντιμετωπίσαμε. Τέλος προτείνουμε πιθανές μελλοντικές επεκτάσεις.

7.1 Σύνοψη και Γενικά Συμπεράσματα

Τα κοινοβουλευτικά πρακτικά που καθίστανται διαθέσιμα στο κοινό παρέχονται με τρόπο που δεν ευνοούν την αποτελεσματική αναζήτηση και σημασιολογική ανάλυση. Αυτό οφείλεται στην απουσία ενός ενιαίου και οργανωμένου πλαισίου που θα βοηθούσε στη συστηματική κατανόηση και αξιοποίηση των δεδομένων. Έτσι, οι δυνατότητες για ανάλυση περιεχομένου και αξιοποίηση σύγχρονων τεχνικών αναζήτησης των κοινοβουλευτικών πρακτικών παραμένουν σε μεγάλο βαθμό ανεκμετάλλετες. Για να υποστηριχθούν προηγμένες αναζητήσεις και ουσιαστική ανάλυση, απαιτούνται τόσο μια δομημένη αναπαράσταση των κοινοβουλευτικών δεδομένων όσο και ένα σύστημα που να την αξιοποιεί.

Ο κύριος στόχος της εργασίας είναι η δημιουργία μιας ολοκληρωμένης ψηφιακής υποδομής για την επεξεργασία, σημασιολογικό εμπλουτισμό και διάθεση των κοινοβουλευτικών συζητήσεων της Ελληνικής Βουλής, με τρόπο που να υποστηρίζει προηγμένες δυνατότητες αναζήτησης και ανάλυσης. Μέσω της μετατροπής των δεδομένων σε ενιαία μορφή, του εμπλουτισμού με εξωτερικές πηγές γνώσης και της ευρετηρίασης με εργαλεία πλήρους κειμένου, παρέχεται ένα εργαλείο που ενισχύει τη διαφάνεια, την ερευνητική αξιοποίηση και την πρόσβαση των πολιτών στο κοινοβουλευτικό έργο. Επιπλέον, έχει αναπτυχθεί δημόσιο API που επιτρέπει προγραμματιστική πρόσβαση στα δεδομένα και υποστηρίζει φιλτραρισμένη ανάκτηση μέσω ειδικών endpoints. Έτσι, η εργασία συμβάλλει ουσιαστικά στη σύνδεση τεχνολογίας και δημοκρατίας, προωθώντας την αξιοποίηση ανοικτών δεδομένων.

Μια σημαντική πρόκληση αφορούσε τον προσδιορισμό των χρήσιμων δεδομένων από τα αρχεία XML. Το σύνολο των διαθέσιμων πληροφοριών συχνά περιλάμβανε πεδία με περιορισμένη ή ασαφή χρησιμότητα. Έγινε επομένως αναγκαίος ένας ενδιάμεσος σχεδιασμός επιλογής, όπου κρίθηκε ποια δεδομένα άξιζε να διατηρηθούν και να ενσωματωθούν στο τελικό μοντέλο, ανάλογα με τη δυνατότητα αξιοποίησής τους σε λειτουργίες αναζήτησης, οπτικοποίησης ή μελλοντικής ανάλυσης.

Καθ' όλη τη διάρκεια της υλοποίησης, προέκυψαν σημαντικές προγραμματιστικές και εννοιολογικές προκλήσεις που ανέδειξαν την πολυπλοκότητα της εργασίας. Ένα από τα

πλέον καθοριστικά προβλήματα αφορούσε την ασυνέπεια και ετερογένεια στα αρχικά δεδομένα. Πολλές οντότητες που θα έπρεπε να ταυτίζονται, όπως συγκεκριμένοι ομιλητές ή συνεδριάσεις, εμφανίζονταν πολλαπλά με μικρές διαφοροποιήσεις στα ονόματα ή στα αναγνωριστικά τους. Για παράδειγμα, το ίδιο πρόσωπο εντοπιζόταν με διαφορετικά `speaker_id`, όπως `panagiotis_kritikos` και `panagiotis_n._kritikos`, ενώ η ίδια σύνοδος αναφερόταν είτε ως “ΣΥΝΟΔΟΣ Α” είτε με παραμορφωμένες ή μη τυποποιημένες μορφές, όπως “Σ Υ Ν Ο Δ Ο Σ Α”.

Επιπλέον, ο εμπλουτισμός των δεδομένων με εξωτερικές πηγές, όπως το Wikidata, αποκάλυψε περιπτώσεις λανθασμένων αντιστοιχίσεων — ομιλητές είχαν συσχετιστεί με ανακριβή Wikidata URLs, οδηγώντας σε εσφαλμένες ή παραπλανητικές πληροφορίες. Η αντιμετώπιση αυτών των θεμάτων απαίτησε την ανάπτυξη μηχανισμών κανονικοποίησης και επικύρωσης, τόσο στο επίπεδο των εσωτερικών αναγνωριστικών όσο και στη διασύνδεση με εξωτερικές πηγές, ώστε να διασφαλιστεί η ορθότητα και η συνέπεια του συνόλου των δεδομένων.

Τέλος, ιδιαίτερη έμφαση δόθηκε στον σχεδιασμό του συστήματος αναζήτησης. Η πρόκληση εδώ ήταν διπλή: αφενός η διατήρηση της λειτουργικότητας πολλαπλών φίλτρων, και αφετέρου η διατήρηση μιας καθαρής και κατανοητής διεπαφής για τον τελικό χρήστη. Για να αποφευχθεί η σύγχυση σε συνδυαστικές αναζητήσεις, τα φίλτρα οργανώθηκαν σε κατηγορίες – ομιλητών και συνεδριάσεων – ώστε να παραμένουν σαφή τα πλαίσια εφαρμογής τους. Η υλοποίηση συνδυαστικών ερωτημάτων αποδείχθηκε απαιτητική, καθώς απαιτούσε δύο παράλληλες αναζητήσεις: μία στη βάση δεδομένων για τη φιλτραρισμένη επιλογή μεταδεδομένων και μία στο σύστημα αναζήτησης πλήρους κειμένου για εντοπισμό σχετικών περιεχομένων. Τα αποτελέσματα έπρεπε να συνδυαστούν με εύρεση τομής, ώστε να προκύπτουν έγκυρα και συνεπή σύνολα δεδομένων προς παρουσίαση.

7.2 Μελλοντικές Επεκτάσεις

Έχοντας ως σημείο εκκίνησης την παρούσα εργασία και τα συμπεράσματα που προέκυψαν από αυτήν, μπορούν να προταθούν διάφορες μελλοντικές επεκτάσεις για τη βελτίωση και την ενίσχυση των αποτελεσμάτων της.

Μια πρώτη προοπτική αφορά την ενσωμάτωση εργαλείων οπτικοποίησης δεδομένων. Η ανάπτυξη διαδραστικών γραφημάτων, χρονικών αξόνων και διαγραμμάτων συναφών εννοιών θα μπορούσε να ενισχύσει τη χρησιμότητα της εφαρμογής και να προσφέρει στους χρήστες πιο διαισθητικούς τρόπους κατανόησης του κοινοβουλευτικού λόγου. Μέσω τέτοιων αναπαραστάσεων θα είναι δυνατή η ανάδειξη θεματικών τάσεων, η απεικόνιση της συμμετοχής των ομιλητών σε βάθος χρόνου, καθώς και η ανάλυση της δραστηριότητας με βάση πολιτικές περιόδους, κόμματα ή θεματικές ενότητες.

Μια άλλη κατεύθυνση επέκτασης σχετίζεται με τη διασύνδεση του συστήματος με άλλες πηγές δημοσίου ενδιαφέροντος, όπως νομοθετικά κείμενα, πρακτικά επιτροπών ή δεδομένα του Gov.gr¹, ενισχύοντας έτσι το οικοσύστημα των διασυνδεδεμένων δεδομένων ελληνικής διακυβέρνησης. Η διαλειτουργικότητα μεταξύ ετερογενών πηγών μπορεί να υποστηρίξει τη διεξαγωγή πιο σύνθετων και διατομεακών ερωτημάτων. Επιπλέον, η αυτόματη συσχέτιση των κοινοβουλευτικών ομιλιών με εξωτερικές πηγές επικαιρότητας, όπως ειδησεογραφικά άρθρα ή επίσημα δελτία τύπου, μπορεί να προσδώσει απαραίτητα συμπραζόμενα και να ενισχύσει

τη σύνδεση του περιεχομένου με το εκάστοτε κοινωνικοπολιτικό πλαίσιο.

Παράλληλα, η υλοποίηση ενός πολυγλωσσικού περιβάλλοντος, με αξιοποίηση τεχνολογιών αυτόματης μετάφρασης, θα μπορούσε να επιτρέψει την πρόσβαση στο περιεχόμενο σε μη ελληνόφωνους χρήστες, ερευνητές ή διεθνείς φορείς. Η δυνατότητα πολυγλωσσικής αναζήτησης και πλοήγησης προάγει τη διαφάνεια και καθιστά την εφαρμογή συμβατή με τις αρχές της ανοικτής και εξωστρεφούς διακυβέρνησης.

Μια ακόμη ουσιαστική δυνατότητα εξέλιξης αφορά την ενσωμάτωση μηχανισμών σχολιασμού και αξιολόγησης του περιεχομένου από την κοινότητα των χρηστών. Η δυνατότητα προσθήκης σχολίων, επισημάνσεων ή ερωτήσεων σε συγκεκριμένες ομιλίες θα μπορούσε να μετατρέψει την εφαρμογή από στατικό εργαλείο αναζήτησης σε διαδραστικό πεδίο πολιτικού διαλόγου και συλλογικής ανάλυσης. Ταυτόχρονα, η αξιολόγηση περιεχομένου βάσει χρησιμότητας ή σαφήνειας μπορεί να αναδείξει κομβικές παρεμβάσεις και να ενισχύσει τη διάχυση τεκμηριωμένης πληροφορίας.

Τέλος, ιδιαίτερο ενδιαφέρον παρουσιάζει η δυνατότητα αξιοποίησης τεχνικών μηχανικής μάθησης για τη βελτιστοποίηση της εμπειρίας αναζήτησης. Μοντέλα κατάταξης (learning to rank), τα οποία μαθαίνουν από δεδομένα χρήστη ποια αποτελέσματα θεωρούνται πιο χρήσιμα για κάθε ερώτημα, ή ερωταποκρίσεων (question answering) μπορούν να ενισχύσουν τη σχετικότητα των αποτελεσμάτων, καθιστώντας την αναζήτηση περισσότερο στοχευμένη και αποδοτική. Περαιτέρω, η ενσωμάτωση ενός έξυπνου ψηφιακού βοηθού συνομιλίας (chatbot) θα μπορούσε να επιτρέψει την υποβολή ερωτημάτων σε φυσική γλώσσα και την επιστροφή απαντήσεων βασισμένων στο σύνολο των κοινοβουλευτικών δεδομένων. Η προσέγγιση αυτή ενισχύει τον διαλογικό χαρακτήρα της πλατφόρμας και τη μετατρέπει σε εργαλείο ανακάλυψης γνώσης και ερευνητικής διερεύνησης.

Αυτές οι πιθανές επεκτάσεις, οι οποίες βασίζονται στα θεμέλια που έθεσε η παρούσα διπλωματική εργασία, ανοίγουν τον δρόμο για καινοτόμες λύσεις που μπορούν να βελτιώσουν περαιτέρω την προσβασιμότητα, τη χρηστικότητα και τον κοινωνικό και ερευνητικό αντίκτυπο των κοινοβουλευτικών δεδομένων στη νέα ψηφιακή εποχή.

¹ Gov.gr: <https://howto.gov.gr/course/view.php?id=7>

Βιβλιογραφία

- [1] Inter-Parliamentary Union. *How World's Parliaments Are Moving Online*. <https://www.ipu.org/digital-stories/how-worlds-parliaments-are-moving-online>, 2024.
- [2] Tomaž Erjavec, Darja Fišer και Kaja Zupan. *The Parla-CLARIN Recommendations for Encoding Corpora of Parliamentary Proceedings*. *Journal of the Text Encoding Initiative*, 14, 2022. <https://journals.openedition.org/jtei/4133>.
- [3] Tomás Martins, João Almeida, Henrique Costa και Jorge Almeida. *Digitization of the Australian Parliamentary Debates, 1998-2022*. *arXiv preprint*, 2023. <https://doi.org/10.48550/arXiv.2304.04561>.
- [4] Henna Poikkimäki, Petri Leskinen, Minna Tamper και Eero Hyvönen. *Analyses of Networks of Politicians Based on Linked Data: Case ParliamentSampo - Parliament of Finland on the Semantic Web*, σελίδα 585-592. Springer International Publishing, 2022. http://dx.doi.org/10.1007/978-3-031-15743-1_53.
- [5] Inter Parliamentary Union. *World e-Parliament Report 2024*. Τεχνική Αναφορά με αριθμό 978-92-9142-903-5, Inter-Parliamentary Union, Geneva, Switzerland, 2024. [parliament-report](https://www.ipu.org/parliament-report).
- [6] Damir Juric, Laura Hollink και Geert Jan Houben. *Bringing Parliamentary Debates to the Semantic Web. Proceedings of the DeRiVE Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web at ISWC*, τόμος 902, σελίδες 51-60. CEUR Workshop Proceedings, 2012. [juric2012](https://www.ceur-ws.org/viewdoc/getDocument?id=12345).
- [7] Polykarpos Meladianos, Giannis Nikolentzos, François Rousseau και Michalis Vazirgiannis. *An Open Framework for Analyzing Public Parliaments Data*. *Journal of Big Data*, 10(1):1-20, 2023. <https://doi.org/10.1186/s40537-023-00831-3>.
- [8] N. Shadbolt, T. Berners-Lee και W. Hall. *The Semantic Web Revisited*. *IEEE Intelligent Systems*, 21(3):96-101, 2006. <https://doi.org/10.1109/MIS.2006.62>.
- [9] Stacy Allison-Cassin και Tim Sherratt. *Introduction to Linked Data for Beginners*, 2020. <https://doi.org/10.46430/phen0068>.
- [10] Guus Schreiber και Yves Raimond. *RDF 1.1 Primer*, 2014. W3C Working Group Note, 24 June 2014, <https://www.w3.org/TR/rdf11-primer/>.
- [11] Steve Harris και Andy Seaborne. *SPARQL 1.1 Query Language*, 2013. W3C Recommendation, 21 March 2013, <https://www.w3.org/TR/sparql11-query/>.

- [12] Nigel Shadbolt, Kieron O'Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, Wendy Hall και m.c. schraefel. *Linked Open Government Data: Lessons from Data.gov.uk*. *IEEE Intelligent Systems*, 27(3):16–24, 2012. <https://doi.org/10.1109/MIS.2012.23>.
- [13] Grant Vergottini. *The Pocket Guide to Akoma Ntoso (Revision 1)*, 2018. OASIS LegalDocumentML (LegalDocML) TC, Accessed: 2025-03-31, https://groups.oasis-open.org/higherlogic/ws/public/document?document_id=62993.
- [14] Christopher D. Manning, Prabhakar Raghavan και Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, UK, 2009. <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>.
- [15] S. K. M. Wong και Vijay V. Raghavan. *Vector space model of information retrieval: a reevaluation*. *Proceedings of the 7th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '84, σελίδα 167–185, Swindon, GBR, 1984. BCS Learning & Development Ltd. <https://dl.acm.org/doi/10.5555/636805.636816>.
- [16] Daniel Jurafsky και James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rdη έκδοση, 2025. Online manuscript released January 12, 2025, <https://web.stanford.edu/~jurafsky/slp3/>.
- [17] K. L. Tan, C. P. Lee και K. M. Lim. *A Survey of Sentiment Analysis: Approaches, Datasets, and Future Directions*. *Applied Sciences*, 13(7):4550, 2023. <https://doi.org/10.3390/app13074550>.
- [18] Aly Abdelrazek, Yomna Eid, Eman Gawish, Walaa Medhat και Ahmed Hassan. *Topic modeling algorithms and applications: A survey*. *Information Systems*, 112:102131, 2023. <https://doi.org/10.1016/j.is.2022.102131>.
- [19] Supriyono, Aji Prasetya Wibawa, Suyono και Fachrul Kurniawan. *A survey of text summarization: Techniques, evaluation and challenges*. *Natural Language Processing Journal*, 7:100070, 2024. <https://doi.org/10.1016/j.nlp.2024.100070>.
- [20] Tomaž Erjavec, Darja Fišer, Andrej Pančur και άλλοι. *ParlaMint II: advancing comparable parliamentary corpora across Europe*. *Language Resources and Evaluation*, 2024. <https://doi.org/10.1007/s10579-024-09798-w>.
- [21] A. Van Aggelen, L. Hollink, M. Kemman, M. Kleppe και H. Beunders. *The debates of the European Parliament as Linked Open Data*. *Semantic Web*, 8(2):271–281, 2017. <https://doi.org/10.3233/sw-160227>.
- [22] Gavin Abercrombie και Riza Theresa Batista-Navarro. *A Sentiment-labelled Corpus of Hansard Parliamentary Debate Speeches*. *ParlaCLARIN*, 2018. http://lrec-conf.org/workshops/lrec2018/W2/pdf/book_of_proceedings.pdf.

- [23] M. Marx και A. Schuth. *DutchParl: A corpus of parliamentary documents in Dutch*. *Language Resources and Evaluation*, 44(4):357–370, 2010. https://www.researchgate.net/publication/254913976-DutchParl_A_corpus_of_parliamentary_documents_in_Dutch.
- [24] Manuel Gómez Juan M. Fernández-Luna, Juan F. Huete και Carlos J. Martín-Dancausa. *Development of the XML Digital Library from the Parliament of Andalucía for Intelligent Structured Retrieval*. *Foundations of Intelligent Systems*. An, S. Matwin, Z.W. Ras και D. Slezak, επιμελητές, τόμος 4994 στο *Lecture Notes in Artificial Intelligence*, σελίδες 417–423. Springer-Verlag, Berlin, Heidelberg, 2008. https://doi.org/10.1007/978-3-540-68123-6_45.
- [25] Kaspar Beelen, Timothy Thijm, Christopher Cochrane, Kees Halvemaan, Graeme Hirst, Michael Kimmins, Sander Lijbrink, Maarten Marx, Nona Naderi, Ludovic Rheault, Roman Polyanovsky και Tanya Whyte. *Digitization of the Canadian Parliamentary Debates*. *Canadian Journal of Political Science*, 50:1–16, 2017. <https://doi.org/10.1017/S0008423916001165>.
- [26] Ιωάννης Παπανικολάου. *Οι Συζητήσεις του Ελληνικού Κοινοβουλίου ως Ανοιχτά Διασυνδεδεμένα Δεδομένα*. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, 2024. <http://artemis.cslab.ece.ntua.gr:8080/jspui/handle/123456789/19258>.
- [27] Konstantina Dritsa, Katerina Thoma, John Pavlopoulos και Panos Louridas. *A Greek Parliament Proceedings Dataset for Computational Linguistics and Political Analysis*, 2022. Preprint, <https://doi.org/10.48550/arxiv.2210.12883>.
- [28] D. Gkoumas, M. Pontiki, K. Papanikolaou και H. Papageorgiou. *Exploring the Political Agenda of the Greek Parliament Plenary Sessions*. *Proceedings of the Language Resources and Evaluation Conference*, σελίδες 1723–1728, 2018. <http://lrec-conf.org/workshops/lrec2018/W2/pdf/8.W2.pdf>.
- [29] E. Berkovitz, Y. Mazuz και M. Fire. *Open Framework for Analyzing Public Parliaments Data*. *Proceedings of the International Conference on Digital Government Research*, σελίδες 289–298, 2022. <https://doi.org/10.1186/s40537-023-00831-3>.
- [30] Derek Greene και James P. Cross. *Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach*. *Political Analysis*, 25(1):77–94, 2017. <https://doi.org/10.1017/pan.2016.7>.
- [31] Gavin Abercrombie και Riza Batista-Navarro. *Sentiment and Position-Taking Analysis of Parliamentary Debates: A Systematic Literature Review*. *Journal of Computational Social Science*, 3(1):245–270, 2020. <https://doi.org/10.1007/s42001-019-00060-w>.
- [32] Rafael Mestre, Razvan Milicin, Stuart E. Middleton, Matt Ryan, Jiatong Zhu και Timothy J. Norman. *M-Arg: Multimodal Argument Mining Dataset for Political Debates with Audio and Transcripts*. *Proceedings of the 8th Workshop on Argument Mining*. Khalid Al-Khatib, Yufang Hou και Manfred Stede, επιμελητές, σελίδες 78–88,

- Punta Cana, Dominican Republic, 2021. Association for Computational Linguistics. <https://aclanthology.org/2021.argmining-1.8/>.
- [33] Steven Grijzenhout, Valentin Jijkoun και Maarten Marx. *Opinion Mining in Dutch Hansards*. Τεχνική Αναφορά με αριθμό, ISLA, Informatics Institute, University of Amsterdam. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a31415a5295e795ed30bec05f22d24502f40d967>.
- [34] Stefanie Walter, Lucy Kinski και Zsófia Boda. *Who talks to whom? Using social network models to understand debate networks in the European Parliament*. *European Union Politics*, 24(2):410–423, 2023. <https://doi.org/10.1177/14651165221137994>.
- [35] Minna Tamper, Rafael Leal, Laura Sinikallio, Petri Leskinen, Jouni Tuominen και Eero Hyvönen. *Extracting Knowledge from Parliamentary Debates for Studying Political Culture and Language*. *Proceedings of the International Workshop on Knowledge Graph Generation from Text (Text2KG 2022)*, τόμος 3184 στο *CEUR Workshop Proceedings*. CEUR-WS.org, 2022. http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_5.pdf.
- [36] L. Kaffee και others. *Wikidata subsetting: Approaches, tools, and evaluation*. *Journal of Semantic Web*, 15(2):233–249, 2024. <https://journals.sagepub.com/doi/10.3233/SW-233491>.
- [37] M. Höppner και T. Scheffler. *Predictive Analytics in Parliamentary Proceedings: Machine Learning Approaches for Policy Analysis*. *PMC Bioinformatics*, 10(319836), 2023.
- [38] Laura South, Michail Schwab, Nick Beauchamp, Lu Wang, John Wihbey και Michelle A. Borkin. *DebateVis: Visualizing Political Debates for Non-Expert Users*. *2020 IEEE Visualization Conference (VIS)*, σελίδες 241–245, 2020. <https://doi.org/10.1109/VIS47514.2020.00055>.
- [39] James R. Lewis. *The System Usability Scale: Past, Present, and Future*. *International Journal of Human-Computer Interaction*, 34(7):577–590, 2018. <https://doi.org/10.1080/10447318.2018.1455307>.
- [40] Nikola Marangunić και Andrina Granić. *Technology acceptance model: a literature review from 1986 to 2013*. *Universal Access in the Information Society*, 14(1):81–95, 2015. <https://doi.org/10.1007/s10209-014-0348-1>.