Ατομική Διπλωματική Εργασία

Αξιολόγηση πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του Γενικού Κανονισμού Προστασίας Δεδομένων

Γιώργος Ζαμπά

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2020

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αξιολόγηση πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του Γενικού Κανονισμού Προστασίας Δεδομένων.

Γιώργος Ζαμπά

Επιβλέπων Καθηγητής

Γιώργος Παπαδόπουλος

Συν-επιβλέπων

Χρίστος Μεττούρης

Ευαγγελία Βανέζη

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2020

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή του Τμήματος Πληροφορικής κ. Γιώργο Παπαδόπουλο που με επέλεξε και ανέλαβε την επίβλεψη της παρούσας εργασίας.
Ιδιαίτερες ευχαριστίες για τη σημαντική βοήθειά τους καθ' όλη την διάρκεια αυτής της μελέτης, θα ήθελα να εκφράσω στον κ. Χρίστο Μεττούρη και την κυρία Ευαγγελία Βανέζη. Η βοήθεια τους ήταν καθοριστική στην ολοκλήρωση του έργου.

Περίληψη

Ο γενικός κανονισμός για την προστασία δεδομένων (General Data Protection Regulation) της Ευρωπαϊκής Ένωσης τέθηκε σε ισχύ στις 25 Μαΐου 2018. Κάθε οργανισμός που επεξεργάζεται προσωπικά δεδομένα πολιτών της ΕΕ πρέπει να συμμορφώνεται με τον προαναφερθέντα κανονισμό. Συνεπώς, πρέπει να υπάρχει η απαραίτητη γνώση για να μπορεί ένας οργανισμός να δημιουργήσει μια πλατφόρμα όπου να τηρεί τις προϋποθέσεις του κανονισμού και να εγγυάται διατήρηση αυτής της δομής στο μέλλον.

Το θέμα της παρούσας διπλωματικής εργασίας είναι η αξιολόγηση διαδικτυακών πλατφορμών και της πολιτικής επεξεργασίας των προσωπικών δεδομένων (privacy policy) που αυτές περιέχουν με βάση τον κανονισμό προστασίας των δεδομένων, GDPR.

Για το πρώτο μέρος της διπλωματικής εργασίας έγινε μια έρευνα πάνω στο GDPR για τους κανονισμούς και τα δικαιώματα του χρήστη των δεδομένων. Η συγκεκριμένη έρευνα μας βοήθησε να καταλήξουμε στους κύριους όρους που πρέπει να περιέχουν οι πολιτικές απορρήτου διαδικτυακών πλατφορμών για να εξασφαλίσουν τον εναρμονισμό με τον κανονισμό GDPR και να πληροφορούν σωστά και πλήρως τους χρήστες για τα δικαιώματα τους και τη χρήση των δεδομένων τους. Αφού εντοπίστηκε ότι υπάρχει κενό, αρχικά δημιουργήσαμε μια λίστα όρων για πολιτικές απορρήτου βάσει του GDPR. Στη συνέχεια μελετήθηκε ένας αριθμός από πλατφόρμες και βρέθηκαν διάφοροι όροι πολιτικών βασιζόμενοι στο GDPR. Παρατηρήθηκαν ομοιότητες και διαφορές κατά την πλοήγηση μας στις διαδικτυακές πλατφόρμες, όπως αναγραμματισμοί και παραλλαγές των όρων που μελετήθηκαν στο Γενικό Κανονισμό για τη Προστασία Δεδομένων. Η μελέτη των διαδικτυακών πλατφορμών ήταν σημαντική για την επέκταση της λίστας όρων για πολιτικές βάσει του GDPR. Στη συνέχεια έγινε μελέτη τεχνολογιών ανάπτυξης ενός προγράμματος που περιηγείται συστηματικά στον Παγκόσμιο Ιστό για να δημιουργήσει ένα ευρετήριο δεδομένων γνωστό και ως crawler, καθώς και για ανάπτυξη ενός προγράμματος που εντοπίζει αντιστοιχίες λέξεων μέσα σε κείμενα (parser).

Το δεύτερο μέρος πραγματεύεται την υλοποίηση του crawler και του parser ώστε να γίνεται αυτοματοποιημένη ανίχνευση της σελίδας που περιέχει την πολιτική (privacy policy) και ανάλυση του κειμένου της με τους ακόλουθους δύο τρόπους:

- Σε σύγκριση με τη λίστα όρων για πολιτικές βάσει του GDPR που δημιουργήθηκε στα πλαίσια αυτής της διπλωματικής.

- Σε σύγκριση με μια έτοιμη οντολογία βασισμένη στο GDPR.

Τα αποτελέσματα της ανάλυσης παρουσιάζονται στο χρήστη.

Στη συνέχεια στα πλαίσια αξιολόγησης του εργαλείου αυτού, έγινε εφαρμογή της αυτοματοποιημένης ανάλυσης μέσω του εργαλείου αυτού, σε υπάρχουσες πλατφόρμες χρησιμοποιώντας και μετά συγκρίνοντας τους δύο τρόπους που αναφέρονται πιο πάνω.

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Γενική Περιγραφή

Είναι γεγονός ότι στην εποχή που ζούμε, λόγω της ταχείας ανάπτυξης της τεχνολογίας, οι άνθρωποι εξαρτώνται όλο και περισσότερο από διάφορα πληροφοριακά συστήματα και ηλεκτρονικές συσκευές για την διεκπεραίωση των καθημερινών τους εργασιών. Η συμμετοχή σε μεγάλο αριθμό κοινωνικών δικτύων, ιστοτόπων, πλατφορμών και εφαρμογών οδηγεί τους χρήστες στην παροχή πολλών προσωπικών πληροφοριών οι οποίες αποθηκεύονται και τυγχάνουν επεξεργασίας με συνέπεια να παρακολουθούνται και να καταγράφονται οι ενέργειες τους. Γι’ αυτό το λόγο η Ευρωπαϊκή Ένωση προχώρησε στη ψήφιση και εφαρμογή του Γενικού Κανονισμού Προστασίας Δεδομένων (General Data Protection Regulation). Λόγω της σημασίας της προστασίας των προσωπικών και ευαίσθητων δεδομένων, θα ήταν απαραίτητο να ενισχυθεί οποιαδήποτε πλατφόρμα με λειτουργίες που θα την καθιστούσαν όσο το δυνατόν πιο συμβατή με το GDPR. Οι πληροφορίες που συλλέγονται από τέτοια συστήματα όπως δηλώνονται και στο GDPR συνήθως σχετίζονται με το χρήστη. Κάποιες από αυτές τις πληροφορίες είναι προσωπικά δεδομένα όπως όνομα, ημερομηνία γέννησης, αριθμός τηλεφώνου, χώρα διαμονής, αριθμός κοινωνικής ασφάλισης, αριθμός ταυτότητας, οικογενειακή κατάσταση, ετήσιο εισόδημα κ.τ.λ.

Τα συστήματα πρέπει να κατασκευάζονται με τέτοιο τρόπο ώστε να διασφαλίζουν την προστασία και το απόρρητο των προσωπικών πληροφοριών των χρηστών, καθώς και να ζητούν μόνο δεδομένα που είναι πράγματι απαραίτητα για τη λειτουργικότητα του συστήματος. Με την εισαγωγή του Γενικού Κανονισμού Προστασίας Δεδομένων της ΕΕ

(GDPR), η ανάγκη αυτή αντιμετωπίζεται και οι οργανισμοί και οι εταιρείες εντός ή και εκτός των χωρών της Ευρωπαϊκής Ένωσης αναγκάζονται να προστατεύσουν τα δεδομένα των χρηστών τους.

Ταυτόχρονα, κάθε σύστημα υποχρεούται να περιλαμβάνει μια πολιτική, η οποία θα περιέχει τα σημεία που καθορίζει ο γενικός κανονισμός για τη προστασία δεδομένων (GDPR).

1.2 Στόχος

Στόχος αυτής της εργασίας είναι να μελετηθούν οι κανονισμοί του GDPR, πως εφαρμόζονται στις πλατφόρμες και πως διατυπώνονται στο privacy policy σε αυτές τις διαδικτυακές πλατφόρμες. Ο πυρήνας αυτής της εργασίας είναι η υλοποίηση ενός συστήματος/λογισμικού που στόχο έχει να εντοπίζει αυτοματοποιημένα τα διάφορα σημεία του GDPR που καλύπτει η πολιτική μιας πλατφόρμας και να τους παρουσιάζει στον χρήστη. Τα αποτελέσματα της ανάλυσης που αναφέρεται πιο πάνω μπορούν να βοηθήσουν ένα οργανισμό ή κάποιο προγραμματιστή να καταλάβει τις αδυναμίες της διαδικτυακής πλατφόρμας που υλοποίησε σε επίπεδο προστασίας προσωπικών δεδομένων των χρηστών του αλλά και οποιοδήποτε χρήστη θελήσει να δει αν μια διαδικτυακή πλατφόρμα που χρησιμοποιεί έχει σωστή πολιτική.

1.3 Μεθοδολογία

Προκειμένου να επιτευχθούν οι παραπάνω στόχοι, έχουμε πραγματοποιήσει πρώτα μια βιβλιογραφική έρευνα σχετικά με το απόρρητο και την προστασία, στις διαδικτυακές πλατφόρμες διαχείρισης χρηστών, και τους κανονισμούς και τα δικαιώματα του χρήστη των δεδομένων εντός του GDPR. Αυτή η βιβλιογραφική έρευνα μας βοήθησε να εξοικειωθούμε με τις ορολογίες που χρησιμοποιούνται στο Γενικό Κανονισμό Προστασίας Δεδομένων και πως αυτές οι ορολογίες διατυπώνονται στις διαδικτυακές πλατφόρμες. Επίσης από τη πιο πάνω έρευνα πήραμε τους κανονισμούς και τα δικαιώματα για τα προσωπικά δεδομένα του χρήστη και δημιουργήσαμε μια λίστα όρων για πολιτικές βάσει του γενικού κανονισμού για τη προστασία δεδομένων (GDPR). Στη συνέχεια μελετήσαμε ένα αριθμό από διαδικτυακές πλατφόρμες. Με αυτή τη μελέτη αναλύσαμε τη δομή της πολιτικής απορρήτου και πως οι διαδικτυακές πλατφόρμες διατυπώνουν τους όρους της λίστας που δημιουργήσαμε με τη μελέτη πάνω στο γενικό κανονισμό για τη προστασία δεδομένων (GDPR). Από αυτή τη μελέτη βρέθηκαν παραλλαγές και αναγραμματισμοί στη διατύπωση των όρων στις πολιτικές απορρήτου των διαδικτυακών πλατφορμών. Αυτό οδήγησε στην επέκταση της υπάρχον

λίστας όρων για πολιτικές βάσει του GDPR. Ακολούθως η έρευνα συνεχίστηκε πάνω στις πλατφόρμες με διερεύνηση στο πηγαίο κώδικα για να παρατηρηθεί το πως γίνεται η περιήγηση στη σελίδα από ένα crawler. Στη περιήγηση από τη κύρια σελίδα στο privacy policy, contact form, registration form και μαζί με τα στοιχεία που πάρθηκαν από τη front-end επισκόπηση συλλέχθηκαν πληροφορίες για την δημιουργία μιας λίστας όρων για περιήγηση. Αυτή η λίστα όρων για περιήγηση μας βοήθησε με την έρευνα και μεθοδολογία που έγινε σε μεταγενέστερο στάδιο για το crawler.

Ακολούθως έγινε η υλοποίηση του crawler χρησιμοποιώντας τη γλώσσα προγραμματισμού Python και τον IDE PyCharm. Με διευκόλυνε στην υλοποίηση η συγκεκριμένη γλώσσα γιατί είχε πολλές βοηθητικές βιβλιοθήκες και ένα user friendly περιβάλλον. Επόμενο στάδιο της υλοποίησης ήταν η αυτόματη ανάλυση του κειμένου που συλλέχθηκε βάσει της λίστας ως επίσης αυτόματη ανάλυση του κειμένου βάση της οντολογίας και παρουσίαση και σύγκριση των αποτελεσμάτων.

1.4 Δομή

Στο επόμενο κεφάλαιο παρουσιάζονται διάφορες έννοιες του σχετικού έργου που έχω μελετήσει, όπως το απόρρητο, τα προσωπικά και ευαίσθητα δεδομένα, εισαγωγή στον Γενικό Κανονισμό για τη Προστασία Δεδομένων, μια γενική επεξήγηση της τεχνολογίας του Crawler, μια γενική επεξήγηση της Οντολογίας και επεξήγηση Συναφών Εργασιών και σύγκριση του με τη δική μας Διπλωματική Εργασία.
Στο Κεφάλαιο 3, εξηγούμε τη διαδικασία δημιουργίας της Λίστας όρων για πολιτικές Απορρήτου Βάσει του GDPR. Επίσης εξηγούμε τη διαδικασία δημιουργίας της Λίστας Όρων για Περιήγηση.
Στο Κεφάλαιο 4, η μεθοδολογία ανάπτυξης του εργαλείου αυτοματοποιημένης αξιολόγησης πολιτικών απορρήτου βάσει του Γενικού Κανονισμού για Προστασία Δεδομένων παρουσιάζεται με πλήρη λεπτομέρεια. Πρώτα εισάγεται η προδιαγραφή και η ανάλυση των απαιτήσεων, η αρχιτεκτονική του λογισμικού και τα επιλεγμένα εργαλεία. Επιπλέον, γίνεται εκτενής επεξήγηση της υλοποίησης του λογισμικού και του γραφικού περιβάλλοντος χρήστη.
Στο Κεφάλαιο 5, παρουσιάζω τη συλλογή και ανάλυση των αποτελεσμάτων, εξάγοντας μερικές πληροφορίες για την αξιολόγηση μιας διαδικτυακής πλατφόρμας. Σε αυτό το κεφάλαιο παρουσιάζονται διάφορα γραφήματα και συγκρίσεις των αποτελεσμάτων.
Τέλος, στο Κεφάλαιο 6 δίνουμε μια εικόνα για την ανάλυση των συμπερασμάτων από τη μελέτη του θέματος και την εφαρμογή του εργαλείου και ταυτόχρονα προτείνουμε κάποιες προτάσεις για μελλοντική εργασία.

Κεφάλαιο 2

Υπόβαθρο/ Συναφείς Εργασίες

2.1 Υπόβαθρο

2.1.1 Προστασία Προσωπικών Δεδομένων

Το απόρρητο (privacy) είναι ένα θεμελιώδες ανθρώπινο δικαίωμα, το οποίο διακηρύχθηκε από τη Γενική Συνέλευση των Ηνωμένων Εθνών στο Παρίσι στις 10 Δεκεμβρίου 1948 με το άρθρο 12 [1] και στη συνέχεια εισήχθη από την Ευρωπαϊκή Σύμβαση για τα Ανθρώπινα Δικαιώματα στη Ρώμη στις 4 Νοεμβρίου 1950 με το άρθρο 8 [2]. Θεωρείται ένα από τα πιο σημαντικά ανθρώπινα δικαιώματα της σύγχρονης εποχής στην οποία ζούμε, και πολλά έργα από διαφορετικούς τομείς μελέτης εξετάζουν και συζητούν την ιδιωτική ζωή των πληροφοριών εδώ και πολλά χρόνια.

Σύμφωνα με το [3] και τους τότε μελλοντικούς δικαστές του Ανώτατου Δικαστηρίου των ΗΠΑ Samuel Warren και Louis Brandeis το 1890, ιδιωτική ζωή περιγράφεται ως το δικαίωμα κάποιου να μείνει μόνος του. Το όρισαν ως το δικαίωμα που επιτρέπει στα άτομα να έχουν προσωπική αυτονομία, ελευθερία του συνεταιρίζεσθαι, στιγμές επιφύλαξης, μοναξιά, οικειότητα και ανεξαρτησία. Σχεδόν 100 χρόνια αργότερα, καθώς ξεκίνησε η εποχή του υπολογιστή, ένα νέο υποσύνολο της ιδιωτικής ζωής εμφανίστηκε το «πληροφοριακό απόρρητο», το οποίο διατυπώθηκε και ορίστηκε πως σε βαθμό ατομικό ή ομαδικό να καθορίζουν για τον εαυτό τους πότε, πώς και σε ποιο βαθμό οι πληροφορίες γι' αυτούς

κοινοποιούνται σε άλλους. Ως έννοια, βασίζεται στην υπόθεση ότι όλες οι πληροφορίες για ένα άτομο είναι καθολικά δικές του, για να μοιραστεί ή να διατηρήσει, όπως αυτός/αυτή θεωρεί κατάλληλα.

2.1.2 Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR)

Η Ευρωπαϊκή Ένωση αποφάσισε να αντιμετωπίσει αυτήν την ανάγκη προστασίας προσωπικών δεδομένων, ψηφίζοντας και εφαρμόζοντας το EU General Data Protection Regulation (GDPR) που τέθηκε σε ισχύ στις 24 Μαΐου 2016 και εφαρμόζεται άμεσα σε όλα τα κράτη μέλη της ΕΕ από τις 25 Μαΐου, 2018. Ο GDPR επιβάλλει σε όλους τους οργανισμούς και εταιρείες που κατοικούν στην ΕΕ ή έχουν τους κατοίκους της ΕΕ ως χρήστες τους να συμμορφώνονται με τους κανονισμούς του. Το GDPR αποτελείται από 99 άρθρα και 173 απαγγελίες, όπως Lawfulness of processing (άρθρο 6), Conditions for consent (άρθρο 7), Right to rectification (άρθρο 16), Right to erasure (άρθρο 17), Right to restriction of processing (άρθρο 18) και πολλά άλλα. Στόχος της είναι να προστατεύσει όλους τους πολίτες της ΕΕ από όλα τα στάδια της επεξεργασίας δεδομένων και των παραβιάσεων δεδομένων στον σημερινό κόσμο.

Υπάρχουν πολλοί κανονισμοί και πολλές αρχές που πρέπει ένας οργανισμός να προσέξει ώστε να μπορέσει να δημιουργήσει μια διαδικτυακή πλατφόρμα που είναι συνεπής με το Γενικό Κανονισμό Προστασίας Δεδομένων (GDPR). Γι’ αυτό πιστεύω πως ένα εργαλείο αξιολόγησης πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR θα φανεί πολύ χρήσιμο σε αυτούς τους οργανισμούς για να πετύχουν την συνέπια που αναφέρεται πιο πάνω αλλά και οποιοδήποτε χρήστη θελήσει να δει αν μια διαδικτυακή πλατφόρμα που χρησιμοποιεί έχει σωστή πολιτική.

2.1.3 Crawler

Ένας Web Crawler, που μερικές φορές ονομάζεται spider ή spiderbot και συχνά συντομευμένο σε crawler, είναι ένα Internet bot που περιηγείται συστηματικά στον Παγκόσμιο Ιστό, συνήθως για σκοπούς ευρετηρίασης Ιστού (web spidering).

Οι μηχανές αναζήτησης και ορισμένοι άλλοι ιστότοποι χρησιμοποιούν Web Crawling για να ενημερώσουν το περιεχόμενο ιστού τους ή δείκτες στο περιεχόμενο άλλων ιστοσελίδων. Οι Web crawlers αντιγράφουν σελίδες για επεξεργασία από μια μηχανή αναζήτησης που σημαδεύει τις εισερχόμενες σελίδες, ώστε οι χρήστες να μπορούν να αναζητούν πιο αποτελεσματικά. Μπορούν να χρησιμοποιηθούν για επικύρωση υπερσυνδέσμων (hyperlinks), κώδικα HTML και για web scrapping.

Οι λειτουργίες αυτών των προγραμμάτων πρέπει να καθοριστούν πριν ξεκινήσει η ανίχνευση. Κάθε παραγγελία ορίζεται εκ των προτέρων. Στη συνέχεια, το πρόγραμμα ανίχνευσης εκτελεί αυτές τις οδηγίες αυτόματα. Δημιουργείται ευρετήριο με τα αποτελέσματα του προγράμματος ανίχνευσης, στο οποίο μπορεί κάποιος να έχει πρόσβαση μέσω λογισμικού εξόδου. Οι πληροφορίες που θα συλλέξει ένας crawler εξαρτάται από τις οδηγίες που θα λάβει.

Για την εκπόνηση της Διπλωματικής Εργασίας μας ήταν ζωτικής σημασίας η κατανόηση της λειτουργικότητας του crawler και η εύρεση των λειτουργιών που θέλουμε να εκτελούνται από τον crawler αυτόματα.

2.1.4 Parser

Ένας Parser είναι η διαδικασία ανάλυσης μιας σειράς συμβόλων, είτε σε φυσική γλώσσα, σε γλώσσες υπολογιστών είτε σε δομές δεδομένων, σύμφωνα με τους κανόνες μιας τυπικής γραμματικής.

Ο όρος έχει ελαφρώς διαφορετικές έννοιες σε διαφορετικούς κλάδους της γλωσσολογίας και της επιστήμης των υπολογιστών. Η παραδοσιακή ανάλυση των προτάσεων εκτελείται συχνά ως μέθοδος κατανόησης της ακριβούς έννοιας μιας πρότασης ή λέξης, μερικές φορές με τη βοήθεια συσκευών όπως διαγράμματα προτάσεων (sentence diagrams). Τονίζει συνήθως τη σημασία των γραμματικών διαιρέσεων όπως το θέμα και το κατηγορηματικό.

Μέσα στην υπολογιστική γλωσσολογία, ο όρος χρησιμοποιείται για να αναφέρεται στην τυπική ανάλυση από έναν υπολογιστή μιας φράσης ή άλλης σειράς λέξεων στα συστατικά του, με αποτέλεσμα ένα δέντρο ανάλυσης να δείχνει τη συντακτική τους σχέση μεταξύ τους, η οποία μπορεί επίσης να περιέχει σημασιολογικές (semantic) και άλλες πληροφορίες.

Ορισμένοι αλγόριθμοι ανάλυσης ενδέχεται να δημιουργήσουν ένα δάσος ανάλυσης ή μια λίστα με δέντρα ανάλυσης για μια συντακτικά ασαφή είσοδο.
Στην επιστήμη των υπολογιστών, ο όρος χρησιμοποιείται στην ανάλυση των διαφόρων γλωσσών προγραμματισμού, αναφερόμενος στη συντακτική ανάλυση του κωδικού εισαγωγής στα συστατικά του μέρη προκειμένου να διευκολυνθεί η σύνταξη μεταγλωττιστών και διερμηνέων. Ο όρος μπορεί επίσης να χρησιμοποιηθεί για να περιγράψει μια διάσπαση ή διαχωρισμό κειμένου ή κώδικα.

Στο πρακτικό κομμάτι της Διπλωματικής Εργασίας που είναι αξιολόγησης πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR η ανάλυση (parsing) γίνεται με δύο τρόπους. Με τη χρήση μιας λίστας όρων για πολιτικές βάσει του γενικού κανονισμού για τη προστασία δεδομένων και με τη χρήση μιας έτοιμης οντολογίας βασισμένης στο GDPR. Η ανάλυσης χρησιμοποιήθηκε για τη σύγκριση των όρων της λίστας μας και της οντολογίας με το πηγαίο κώδικα της πολιτικής απορρήτου. Διερευνήθηκε κατά πόσων υπάρχουν αυτοί οι όροι μέσα στην πολιτική.

2.1.4 Οντολογίες

Στην επιστήμη των υπολογιστών, η οντολογία είναι μια επίσημη αναπαράσταση της γνώσης από ένα σύνολο εννοιών σε έναν τομέα και τις σχέσεις μεταξύ αυτών. Χρησιμοποιείται για να αιτιολογήσει τις ιδιότητες αυτού του τομέα και μπορεί να χρησιμοποιηθεί για την περιγραφή του τομέα. Θεωρητικά, η οντολογία είναι μια επίσημη, ρητή προδιαγραφή μιας κοινής σύλληψης (”formal, explicit specification of a shared conceptualization”). Η οντολογία παρέχει ένα κοινό λεξιλόγιο, το οποίο μπορεί να χρησιμοποιηθεί για τη μοντελοποίηση ενός τομέα που είναι, τον τύπο αντικειμένων ή/ και τις έννοιες που υπάρχουν, καθώς και τις ιδιότητες και τις σχέσεις τους.

Οι οντολογίες χρησιμοποιούνται στην Τεχνητή Νοημοσύνη, τον Σημασιολογικό Ιστό, τη Μηχανική Συστημάτων, τη Μηχανική Λογισμικού, τη Βιοϊατρική Πληροφορική, την Επιστήμη της Βιβλιοθήκης και την Αρχιτεκτονική Πληροφοριών ως μια μορφή αναπαράστασης της γνώσης για τον κόσμο ή μέρος αυτού. Υπάρχουν τέσσερις κατηγορίες οντολογίας: στατική, δυναμική, σκόπιμη και κοινωνική.

- Static: Περιγράφει πράγματα που υπάρχουν, τα χαρακτηριστικά και τη σχέση τους.
- Dynamic: Περιγράφει τον κόσμο σε όρους καταστάσεων, μεταβάσεων κατάστασης και διαδικασιών.

- Intentional: Περιλαμβάνει τον κόσμο των πρακτόρων, πράγματα που πιστεύουν, θέλουν, αποδεικνύουν ή διαψεύδουν και διαφωνούν.
- Social: Καλύπτει κοινωνικές ρυθμίσεις, μόνιμες οργανωτικές δομές ή μεταβαλλόμενα δίκτυα συμμαχιών και ανεξαρτησίας.

Στο πρακτικό κομμάτι της Διπλωματικής Εργασίας που είναι αξιολόγησης πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR, η αξιολόγηση γίνεται με δύο τρόπους ο πρώτος είναι με τη χρήση της λίστας όρων για πολιτικές βάσει του GDPR δικής μας δημιουργίας και ο δεύτερος τρόπος είναι βάσει μιας έτοιμης οντολογίας βασισμένη στο GDPR. Ο δεύτερος τρόπος προστέθηκε για να υπάρξει μια σύγκριση με την δική μας λίστα όρων.

2.2 Συναφείς Εργασίες

Σε αυτό το υποκεφάλαιο θα μιλήσω για σχετικές εργασίες οι οποίες ασχολήθηκαν με ίδιο ή παρόμοιο θέμα με την αξιολόγηση πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR.

Όπως αναφέρεται στη συναφή εργασία [4] και όπως το επιβάλλει ο νόμος οποιοσδήποτε οργανισμός δεν ακολουθεί το GDPR και τους κανονισμούς του θα υπόκειται σε σημαντικά πρόστιμα και νομικά προβλήματα. Στο [4] επίσης παρουσίασαν με δική τους αρίθμηση όλα τα σημεία του GDPR που πρέπει να περιέχει μια πολιτική:

- GDPR1: Καθορισμός δεδομένων που συλλέγονται: Οι χρήστες πρέπει να γνωρίζουν πια στοιχεία συλλέγονται από του οργανισμούς για εκείνους. Επιπρόσθετα οι οργανισμοί πρέπει να καταγράφουν τα δεδομένα που μαζεύονται το οποίο συνδέεται με τη λογοδοσία που απαιτείται από το GDPR [5].
- GDPR2: Αιτιολόγηση για τη συλλογή δεδομένων: Οι οργανισμοί πρέπει να εξηγούν τα δικαιώματα τους για τη περισυλλογή δεδομένων [5], πρέπει επίσης να είναι ξεκάθαρο και στους ίδιους ο λόγος για τον οποίο χρειάζεται να συλλέξουν αυτά τα δεδομένα [6].
- GDPR3: Πώς θα επεξεργαστούν τα δεδομένα: Ο οργανισμός πρέπει να ενημερώσει το χρήστη για τα νόμιμα δικαιώματα που έχει για την επεξεργασία

προσωπικών δεδομένων [5]. Το GDPR υπογραμμίζει τους τρόπους όπου η επεξεργασία θεωρείται νόμιμη (ένα από τα ακόλουθα πρέπει να ισχύει):

1. Ο χρήστης έχει δώσει τη συγκατάθεση (consent) του, χρησιμοποιείται για να δημιουργηθεί σύμβαση με τον χρήστη.
2. Ο υπεύθυνος επεξεργασίας δεδομένων (data controller) συμμορφώνεται με μια νομική υποχρέωση, χρησιμοποιείται για την προστασία των συμφερόντων ενός ατόμου.
3. Απαιτείται για μια εργασία που περιλαμβάνει το δημόσιο συμφέρον
4. Απαιτείται για νόμιμο σκοπό από τον υπεύθυνο επεξεργασίας υπό την προϋπόθεση ότι δεν παραβιάζονται δικαιώματα και ελευθερίες [7].
5. Επιπλέον, το άτομο έχει το δικαίωμα να εξαιρεθεί από την επεξεργασία των δεδομένων του (the right to opt out of processing) με αλγόριθμο ή οποιοδήποτε άλλο προφίλ.

- GDPR4: Πόσο καιρό θα διατηρηθούν τα δεδομένα: Το GDPR υπαγορεύει ότι τα δεδομένα πρέπει να διατηρούνται για το μικρότερο δυνατό χρονικό διάστημα και οι οργανισμοί πρέπει να αναφέρουν πόσο καιρό θα διατηρηθούν [5][8]
- GDPR5: Με πιο μπορεί να επικοινωνήσει ο χρήστης για την κατάργηση ή την παραγωγή δεδομένων: Οι χρήστες έχουν το δικαίωμα να ζητήσουν για όλα τα δεδομένα τους να ξεχαστούν και να τους τα παρέχουν, είτε αυτά είναι δεδομένα που δόθηκαν από τον οργανισμό είτε δημιουργήθηκαν από το σύστημα. Για να διευκολυνθεί αυτό, τα στοιχεία επικοινωνίας πρέπει να παρέχονται στην πολιτική.
- GDPR6: Επικοινωνία πληροφοριών απορρήτου (Privacy Information): Η τεκμηρίωση σχετικά με τη νομοθεσία σημειώνει ότι “requires the information to be provided in concise, easy to understand and clear language” [5].

Στη συναφή εργασία [4] παρουσιάζουν τα σημεία που πρέπει να περιέχει μια πολιτική απορρήτου. Στην Διπλωματική μας εργασία γίνεται αξιολόγηση αυτών των πολιτικών απορρήτου από την άποψη προστασίας προσωπικών δεδομένων βάσει του Γενικού Κανονισμού για την Προστασία Δεδομένων (GDPR) και η εργασία [4] μας βοήθησε να κατανοήσουμε περισσότερο τη δομή μιας πολιτικής.

Η εργασία [9] επηρεάζεται κυρίως από το GDPR, στόχο έχει την υποστήριξη των χρηστών του Διαδικτύου απλοποιώντας την αναγνωσιμότητα των πολιτικών απορρήτου. Λαμβάνουν τις απαιτήσεις που αναφέρονται στο GDPR ως βάση και παρέχουν ένα εργαλείο συγκριτικής αξιολόγησης της πολιτικής απορρήτου που λαμβάνει υπόψη πτυχές απορρήτου και μέσω της

χρήσης δεικτών κινδύνου (risk indicators) ελκύουν την προσοχή των χρηστών στα πιο συναφή μέρη της πολιτικής απορρήτου. Για να το επιτύχουν, εισάγουν πρώτα μια κατηγοριοποίηση δεδομένων με βάση την ερμηνεία του GDPR. Δεύτερον, προτείνουν μια προσέγγιση βάσει κινδύνου για τη συγκριτική αξιολόγηση της πολιτικής απορρήτου που χρησιμοποιούν εποπτευόμενες τεχνικές μηχανικής μάθησης για να εξάγουν το πιο σχετικό περιεχόμενο των μεγάλων πολιτικών απορρήτου. Στη συνέχεια, χρησιμοποιούν μια ταξινόμηση βάσει κινδύνου σε τρεις κλίμακες (Πράσινο, Κίτρινο, Κόκκινο) προκειμένου να απλοποιήσουν την ερμηνεία της πολιτικής απορρήτου και να ελκύουν την προσοχή των χρηστών στους δείκτες κινδύνου.

Στην εργασία [9] προσπαθούν να ενθαρρύνουν το χρήστη να διαβάσει την πολιτική απορρήτου δείχνοντας του χρήστη τα σημαντικά σημεία. Στην Διπλωματική μας εργασία γίνεται αξιολόγηση αυτών των πολιτικών απορρήτου, εμείς βοηθάμε τον οργανισμό ώστε να καταλάβει τα αδύνατα και δυνατά στοιχεία της διαδικτυακής πλατφόρμας του σε επίπεδο προστασίας προσωπικών δεδομένων του χρήστη. Κοινό στοιχείο είναι ότι λαμβάνονται οι απαιτήσεις που αναφέρονται στο GDPR ως βάση αλλά εμείς χρησιμοποιούμε και παραλλαγές των όρων που βρεθήκαν στις διαδικτυακές πλατφόρμες όπου έγινε η περιήγηση.

Η εργασία [10] πρότεινε το Privee το οποίο συνδυάζει την ταξινόμηση κανόνων και μηχανικής μάθησης (ML) με την πολιτική απορρήτου και crowdsourcing για απρόσκοπτη ενσωμάτωση στο υπάρχον καθεστώς απορρήτου στον Ιστό. Υλοποιήθηκε με τη βοήθεια ενός extension που προσφέρει το Google Chrome το οποίο αλληλεπιδρά με τις ιστοσελίδες των Privacy Policies και ToS · DR αποθετήριο αποτελεσμάτων από τις crowdsourced πολιτικές απορρήτου. Επίσης πραγματοποίησαν μια στατιστική ανάλυση των πειραματικών τους αποτελεσμάτων που δείχνει ότι η ασάφεια των πολιτικών απορρήτου τις καθιστά δύσκολες στην κατανόηση τόσο για τους ανθρώπους όσο και για τα automatic classifiers.

Στην εργασία [10] έγινε έρευνα κατά πόσον είναι δύσκολες στη κατανόηση οι πολιτικές απορρήτου στον χρήστη αλλά και στα automatic classifiers. Στην Διπλωματική μας εργασία γίνεται αξιολόγηση αυτών των πολιτικών απορρήτου από την άποψη προστασίας προσωπικών δεδομένων βάσει του Γενικού Κανονισμού για την Προστασία Δεδομένων (GDPR).

Η εργασία [11] χρησιμοποιούν την domain-specific γλώσσα RSL-IL4Privacy ως τον απαραίτητο μηχανισμό για τον καθορισμό των πολιτικών, παρέχοντας παράλληλα δυνατότητες για καλύτερη ανάλυση και επικύρωση των αντίστοιχων πολιτικών. Η υιοθέτηση αυτής της γλώσσας ορίστηκε στην προσέγγιση RSLingo4Privacy με την ενσωμάτωση των ακόλουθων βασικών διαδικασιών : αυτόματη ταξινόμηση, εξαγωγή και μετάφραση των δηλώσεων από πολιτικές που γράφτηκαν στο NL σε προδιαγραφές RSLIL4Privacy, απεικόνιση και συγγραφή, ανάλυση και επικύρωση και δημοσίευση σε δομημένο και αναγνώσιμο από άνθρωπο και μηχανή. Επίσης προσθέτει τις ακόλουθες νέες συνεισφορές: περιγραφή ενός εργαλείου το RSLingo4Privacy Studio που υλοποιεί κατά προσέγγιση την RSLingo4Privacy και χρησιμοποιεί το RSLIL4Privacy ως ενδιάμεση γλώσσα για τον καθορισμό των πολιτικών απορρήτου και μια εκτενή συζήτηση για το πώς αυτό το εργαλείο υποστηρίζει αρκετούς μετασχηματισμούς για την αναπαράσταση της πληθώρας των πολιτικών.

Στην εργασία [11] χρησιμοποιήθηκε μια domain-specific γλώσσα η RSL-IL4Privacy και δημιούργησε ένα εργαλείο για να βελτιώσει τις προδιαγραφές και τις αναλύσεις των πολιτικών απορρήτου. Στην Διπλωματική μας εργασία χρησιμοποιήθηκε η γλώσσα python και δημιουργήσαμε ένα εργαλείο που αξιολογά πολιτικές απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR.

Στην εργασία [12] αναπτύξαν ένα αυτοματοποιημένο αγωγό για τη συλλογή και ανάλυση 6.278 μοναδικών αγγλικών πολιτικών απορρήτου συγκρίνοντας τις εκδόσεις πριν από το GDPR και μετά το GDPR. Αυτές οι πολιτικές καλύπτουν τις πρακτικές απορρήτου των ιστοσελίδων από διαφορετικά θέματα και περιοχές. Προσεγγίζουμε το πρόβλημα μελετώντας την αλλαγή που προκαλείται σε ολόκληρη την εμπειρία των χρηστών στην αλληλεπίδραση τους με τις πολιτικές απορρήτου.

Έκαναν crawl το κώδικα HTML χρησιμοποιώντας το Selenium framework5 και ένα headless Chrome Browser. Προσδιόρισαν ένα σύνολο συνδέσμων από υποψήφιες πολιτικές της αρχικής σελίδας βασιζόμενοι κανονικές εκφράσεις. Χρησιμοποίησαν το Wayback Machine6 από το Internet Archive για να συλλέξουν μια σειρά από αποκόμματα HTML για κάθε πολιτική απορρήτου. Χρησιμοποίησαν τη βιβλιοθήκη της python την Waybackpack7 για να

ζητήσουν κάθε μοναδική αρχειοθετημένη παρουσία από τον Ιανουάριο του 2016 έως τον Μάιο 2019, με μηνιαία ανάλυση.

Η μεθοδολογία τους καταλήγει σε δύο συμπεράσματα σχετικά με τον αρχικό αντίκτυπο του GDPR στο τοπίο της πολιτικής απορρήτου:

- Ο Μάιος 2018 ήταν ένα σημείο καμπής περισσότερο από το 45% των πολιτικών που μελέτησαν έχουν αλλάξει μεταξύ Μαρτίου και Ιουλίου 2018.
- Φαίνεται ότι το GDPR είχε υψηλότερη επίδραση στην ΕΕ από ό, τι στο εξωτερικό.

Στην εργασία [12] έγινε συλλογή και ανάλυση ενός υψηλού αριθμού πολιτικών απορρήτου και έγινε σύγκριση στις εκδόσεις αυτών των πολιτικών πριν και μετά την καθιέρωση του Γενικού Κανονισμού για την Προστασία Δεδομένων (GDPR). Στην Διπλωματική μας εργασία χρησιμοποιήθηκε η γλώσσα python και δημιουργήσαμε ένα εργαλείο που αξιολογεί πολιτικές απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR.

Κεφάλαιο 3

Περιήγηση και Ανάλυση Πολιτικών

3.1 Εισαγωγή

Το επίκεντρο αυτού του κεφαλαίου είναι η επεξήγηση της διαδικασίας που ακολουθήσαμε για τη δημιουργία της λίστας όρων για πολιτικές με βάσει το Γενικό Κανονισμό Προσωπικών Δεδομένων (GDPR). Για το πρώτο βήμα της δημιουργίας της λίστας όρων βάσει το GDPR, έχουμε πραγματοποιήσει μια βιβλιογραφική έρευνα σχετικά με τους κανονισμούς και τα δικαιώματα του χρήστη των δεδομένων εντός του GDPR. Αυτή η βιβλιογραφική έρευνα μας βοήθησε να εξοικειωθούμε με τις ορολογίες που χρησιμοποιούνται στο Γενικό Κανονισμό Προστασίας Δεδομένων και πως αυτές οι ορολογίες διατυπώνονται στις διαδικτυακές πλατφόρμες. Από τη πιο πάνω έρευνα πήραμε τους κανονισμούς και τα δικαιώματα για τα προσωπικά δεδομένα του χρήστη και

δημιουργήσαμε την λίστα όρων για πολιτικές βάσει του γενικού κανονισμού για τη προστασία δεδομένων (GDPR). Για το δεύτερο βήμα της δημιουργίας της λίστας όρων βάσει το GDPR μελετήσαμε ένα αριθμό από διαδικτυακές πλατφόρμες. Με αυτή τη μελέτη αναλύσαμε τη δομή της πολιτικής απορρήτου και πως οι διαδικτυακές πλατφόρμες διατυπώνουν τους όρους της λίστας που δημιουργήσαμε με τη μελέτη πάνω στο γενικό κανονισμό για τη προστασία δεδομένων (GDPR). Από αυτή τη μελέτη βρέθηκαν παραλλαγές και αναγραμματισμοί στη διατύπωση των όρων στις πολιτικές απορρήτου των διαδικτυακών πλατφορμών. Αυτό οδήγησε στην επέκταση της λίστας όρων για πολιτικές βάσει του GDPR που δημιουργήθηκε στο πρώτο βήμα.

Το δεύτερο κομμάτι αυτού του κεφαλαίου είναι η επεξήγηση της διαδικασίας δημιουργίας μιας λίστας όρων για περιήγηση. Αυτό επιτεύχθηκε με μια έρευνα πάνω στις πλατφόρμες με διερεύνηση στο πηγαίο κώδικα για να παρατηρηθεί το πως γίνεται η περιήγηση στη σελίδα από ένα crawler. Στη περιήγηση από τη κύρια σελίδα στο privacy policy, contact form, registration form και μαζί με τα στοιχεία που πάρθηκαν από τη front-end επισκόπηση συλλέχθηκαν πληροφορίες όπως φράσεις-κλειδιά για την δημιουργία μιας λίστας όρων για περιήγηση. Αυτοί οι όροι που αποτελούν τη λίστα είναι μέρος του συνδέσμου στον οποίο βρισκόμασταν κατά την περιήγηση.

3.2 Δημιουργία Λίστας Όρων για Πολιτικές Βάσει του GDPR

3.2.1 Γενικός Κανονισμός Προσωπικών Δεδομένων

Σε αυτό το υποκεφάλαιο παρουσιάζεται η επεξήγηση των δικαιωμάτων και αρχών που φάνηκαν χρήσιμα στη Διπλωματική Εργασία και πως αυτά συνέβαλαν στη δημιουργία της λίστας όρων βάσει του γενικού κανονισμού προστασίας δεδομένων (GDPR).

3.2.1.1 “Lawfulness, fairness and transparency”

Όπως ορίζεται στο GDPR άρθρο 5 παράγραφος a [13]

Για την ανάπτυξη διαδικτυακής πλατφόρμας, αυτό σημαίνει ότι πριν από οποιαδήποτε επεξεργασία ή αποθήκευση προσωπικών δεδομένων, η πλατφόρμα πρέπει να διασφαλίσει ότι έχει αποδειχθεί η συναίνεση του χρήστη. Η επεξεργασία θεωρείται νόμιμη μόνο εάν το σύστημα έχει νόμιμη βάση (lawful basis) σύμφωνα με το άρθρο 6. Επίσης, η πλατφόρμα θα

πρέπει να διατηρεί αρχείο επιβεβαίωσης της συγκατάθεσης που αναφέρει πότε και πώς έλαβε τη συγκατάθεση από τον χρήστη και τι ακριβώς είπε ο χρήστης τη στιγμή της παρέχοντας τη συγκατάθεση.

3.2.1.2 Lawfulness of Processing – Άρθρο 6

Όπως ορίζεται στο GDPR άρθρο 6 [13]

Υπάρχουν έξι νόμιμες βάσεις (lawful bases). Κατά την ανάπτυξη μιας διαδικτυακής πλατφόρμας ένα από αυτά τουλάχιστον πρέπει να εφαρμόζεται κάθε φορά που ένα σύστημα επεξεργάζεται προσωπικά δεδομένα. Η νόμιμη βάση πρέπει να καθοριστεί πριν από την έναρξη της επεξεργασίας δεδομένων και επίσης να τεκμηριωθεί στο σύστημα. Η επιλεγμένη νόμιμη βάση καθώς και ο σκοπός της επεξεργασίας πρέπει να παρουσιάζονται με σαφήνεια στην Πολιτική Απορρήτου του συστήματος.

3.2.1.3 Conditions for Consent – Άρθρο 7

Όπως ορίζεται στο GDPR άρθρο 7 [13]

Όταν η επεξεργασία βασίζεται στη συγκατάθεση, ο υπεύθυνος επεξεργασίας θα μπορεί να αποδείξει ότι το υποκείμενο των δεδομένων έχει συναινέσει στην επεξεργασία των προσωπικών του δεδομένων.

Εάν η συγκατάθεση του υποκειμένου των δεδομένων παρέχεται στο πλαίσιο γραπτής δήλωσης που αφορά και άλλα θέματα, η αίτηση συγκατάθεσης υποβάλλεται κατά τρόπο σαφώς διακριτό από τα άλλα θέματα, σε κατανοητή και εύκολα προσβάσιμη μορφή, χρησιμοποιώντας σαφή και απλή γλώσσα Οποιοδήποτε μέρος μιας τέτοιας δήλωσης που συνιστά παράβαση του παρόντος κανονισμού δεν είναι δεσμευτικό.

3.2.1.4 Right of Access by the Data Subject – Άρθρο 15

Όπως ορίζεται στο GDPR άρθρο 15 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να λάβει από την επιβεβαίωση του υπευθύνου επεξεργασίας σχετικά με το κατά πόσον υποβάλλονται σε επεξεργασία προσωπικά δεδομένα που τον αφορούν.

Όταν τα προσωπικά δεδομένα διαβιβάζονται σε τρίτη χώρα ή σε διεθνή οργανισμό, το υποκείμενο των δεδομένων έχει το δικαίωμα να ενημερώνεται για τις κατάλληλες διασφαλίσεις σύμφωνα με το άρθρο 46 σχετικά με τη διαβίβαση.

Ο υπεύθυνος επεξεργασίας παρέχει αντίγραφο των προσωπικών δεδομένων που υποβάλλονται σε επεξεργασία. Για τυχόν αντίγραφα που ζητούνται από το υποκείμενο των δεδομένων, ο υπεύθυνος επεξεργασίας μπορεί να χρεώσει εύλογο τέλος βάσει των διοικητικών δαπανών. Όταν το υποκείμενο των δεδομένων υποβάλλει το αίτημα με ηλεκτρονικά μέσα, και εκτός εάν ζητηθεί διαφορετικά από το υποκείμενο των δεδομένων, οι πληροφορίες παρέχονται σε ηλεκτρονική μορφή που χρησιμοποιείται συνήθως.

3.2.1.5 Right to Rectification – Άρθρο 16

Όπως ορίζεται στο GDPR άρθρο 16 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να λάβει από τον υπεύθυνο επεξεργασίας χωρίς αδικαιολόγητη καθυστέρηση την διόρθωση ανακριβών προσωπικών δεδομένων που τον αφορούν. Λαμβάνοντας υπόψη τους σκοπούς της επεξεργασίας, το υποκείμενο των δεδομένων έχει το δικαίωμα να έχει συμπληρώσει ελλιπή προσωπικά δεδομένα, συμπεριλαμβανομένης της παροχής συμπληρωματικής δήλωσης.

3.2.1.6 Right to Erasure (Right to Be Forgotten) – Άρθρο 17

Όπως ορίζεται στο GDPR άρθρο 17 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να λάβει από τον υπεύθυνο επεξεργασίας τη διαγραφή των προσωπικών δεδομένων που τον αφορούν χωρίς αδικαιολόγητη καθυστέρηση και ο υπεύθυνος επεξεργασίας έχει την υποχρέωση να διαγράψει προσωπικά δεδομένα χωρίς αδικαιολόγητη καθυστέρηση.

Όταν ο υπεύθυνος επεξεργασίας έχει δημοσιοποιήσει τα προσωπικά δεδομένα, είναι υποχρεωμένος να διαγράψει τα προσωπικά δεδομένα. Ο υπεύθυνος επεξεργασίας, λαμβάνοντας υπόψη τη διαθέσιμη τεχνολογία και το κόστος εφαρμογής, λαμβάνει εύλογα μέτρα, συμπεριλαμβανομένων τεχνικών μέτρων, για να ενημερώσει τους υπευθύνους επεξεργασίας των προσωπικών δεδομένων ότι το υποκείμενο των δεδομένων ζήτησε τη διαγραφή από αυτούς τους υπευθύνους επεξεργασίας τυχόν συνδέσμων ή αντιγραφής ή αναπαραγωγής αυτών των προσωπικών δεδομένων.

3.2.1.7 Right to Restriction of Processing – Άρθρο 18

Όπως ορίζεται στο GDPR άρθρο 18 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να λάβει από τον ελεγκτή τον περιορισμό της επεξεργασίας των δεδομένων αυτών.

Όταν η επεξεργασία έχει περιοριστεί, αυτά τα προσωπικά δεδομένα, εκτός από την αποθήκευση, θα υποβάλλονται σε επεξεργασία μόνο με τη συγκατάθεση του υποκειμένου των δεδομένων ή για τη σύσταση, άσκηση ή υπεράσπιση νομικών αξιώσεων ή για την προστασία των δικαιωμάτων άλλου φυσικού ή νομικού προσώπου ή για λόγους σημαντικού δημοσίου συμφέροντος της Ευρωπαϊκής Ένωσης ή ενός κράτους μέλους.

Ένα υποκείμενο δεδομένων που έχει επιτύχει περιορισμό της επεξεργασίας ενημερώνεται από τον υπεύθυνο επεξεργασίας πριν αρθεί ο περιορισμός της επεξεργασίας.

3.2.1.8 Right to Data Portability – Άρθρο 20

Όπως ορίζεται στο GDPR άρθρο 20 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να λαμβάνει τα προσωπικά δεδομένα που τον αφορούν, τα οποία έχει παράσχει σε υπεύθυνο επεξεργασίας, σε δομημένη, κοινώς χρησιμοποιούμενη και αναγνώσιμη από μηχανή μορφή και έχει το δικαίωμα να διαβιβάζει αυτά τα δεδομένα σε άλλο ελεγκτή χωρίς εμπόδιο από τον ελεγκτή στον οποίο έχουν παρασχεθεί τα προσωπικά δεδομένα όπου:

- Η επεξεργασία βασίζεται στη συγκατάθεση σύμφωνα με το άρθρο 6 παράγραφος 1 σημείο α ή στο άρθρο 9 παράγραφος 2 σημείο α ή σε σύμβαση σύμφωνα με το άρθρο 6 παράγραφος 1 σημείο β.
- Η επεξεργασία πραγματοποιείται με αυτοματοποιημένα μέσα.

3.2.1.9 Right to object – Άρθρο 21

Όπως ορίζεται στο GDPR άρθρο 21 [13]

Το υποκείμενο των δεδομένων έχει το δικαίωμα να προβάλλει ένσταση, για λόγους που σχετίζονται με την συγκεκριμένη κατάστασή του, ανά πάσα στιγμή σε επεξεργασία δεδομένων προσωπικού χαρακτήρα που τον αφορούν, η οποία βασίζεται στο άρθρο 6 παράγραφος 1 σημεία e και f , συμπεριλαμβανομένου του προφίλ βάσει αυτών των διατάξεων. Ο υπεύθυνος επεξεργασίας δεν θα επεξεργάζεται πλέον τα προσωπικά δεδομένα, εκτός εάν ο υπεύθυνος επεξεργασίας αποδείξει επιτακτικούς νόμιμους λόγους για την επεξεργασία που υπερισχύουν των συμφερόντων, των δικαιωμάτων και των ελευθεριών του υποκειμένου των δεδομένων ή για τη θέσπιση, την άσκηση ή την υπεράσπιση νομικών αξιώσεων.

Όταν τα δεδομένα προσωπικού χαρακτήρα υποβάλλονται σε επεξεργασία για σκοπούς άμεσου μάρκετινγκ, το υποκείμενο των δεδομένων έχει το δικαίωμα να υποβάλει ένσταση ανά πάσα στιγμή στην επεξεργασία προσωπικών δεδομένων που τον αφορούν για τέτοιο μάρκετινγκ, το οποίο περιλαμβάνει τη δημιουργία προφίλ στο βαθμό που σχετίζεται με αυτό το άμεσο μάρκετινγκ.

Με αργότερη την πρώτη επικοινωνία με το υποκείμενο των δεδομένων, το δικαίωμα που αναφέρεται στις παραγράφους 1 και 2 να γνωστοποιείται ρητά στο υποκείμενο των δεδομένων και πρέπει να παρουσιάζεται σαφώς και χωριστά από οποιαδήποτε άλλη πληροφορία.

Η πιο πάνω έρευνα των αρχών και δικαιωμάτων με βοήθησε στο να κατανοήσω καλύτερα το Γενικό Κανονισμό για την Προστασία Δεδομένων (GDPR) και οι αρχές και δικαιώματα που αναλύονται πιο πάνω αποτέλεσαν την αρχή σύνταξης της λίστας όρων για πολιτικές βάσει του GDPR. Από τη πιο πάνω έρευνα πήραμε τους κανονισμούς και τα δικαιώματα για τα προσωπικά δεδομένα του χρήστη και δημιουργήσαμε την λίστα όρων για πολιτικές βάσει του

γενικού κανονισμού για τη προστασία δεδομένων (GDPR). Η αρχική λίστα παρατίθεται στον Πίνακα 3.2.1.1.

Lawfulness of Processing
Right to erasure
Right of access by the data subject
Right to data portability
Right to Rectification
Right to restriction of processing
Right to object

Πίνακας 3.2.1.1 Αρχική Λίστα Όρων Βάσει του GDPR

3.2.2 Μελέτη Διαδικτυακών Πλατφορμών

Μελετήθηκε ένας αριθμός από διαδικτυακές πλατφόρμες και βρέθηκαν διάφοροι όροι βασιζόμενοι στο GDPR. Με αυτή τη μελέτη αναλύσαμε τη δομή της πολιτικής απορρήτου και πως οι διαδικτυακές πλατφόρμες διατυπώνουν τους όρους της λίστας που δημιουργήσαμε με τη μελέτη πάνω στο γενικό κανονισμό για τη προστασία δεδομένων (3.2.1). Μεταξύ των διαδικτυακών πλατφορμών που μελετήθηκαν παρατηρήθηκαν ομοιότητες και διαφορές ως προς τους όρους και τη διατύπωση.

Από αυτή τη μελέτη βρέθηκαν παραλλαγές και αναγραμματισμοί στη διατύπωση των όρων της λίστας για πολιτικές βάσει του GDPR κατά την περιήγηση μας στις πολιτικές απορρήτου των διαδικτυακών πλατφορμών.

Παραδείγματα:

Παραλλαγές:
«Right of Erasure» ή «The Right to Request Deletion» ή «Right To be Forgotten»
«Right to restriction of processing» ή «Requests the Restriction of Their Use»
«Right to Correction» ή «Right to Correct» ή « The Right to Correct and Update»

Αναγραμματισμοί:
«Right to Object to Processing» ή «Processing Objection»
«Right Of Access» ή «Access Personal Data»

Η μελέτη ενός αριθμού από διαδικτυακές πλατφόρμες βοήθησε στην επέκταση της υπάρχον λίστας όρων για πολιτικές βάσει του GDPR που δημιουργήθηκε με τη μελέτη αρχών και κανονισμών του Γενικού Κανονισμού για Προστασία Δεδομένων που παρουσιάζεται στο υποκεφάλαιο 3.2.1.

3.2.3 Παρουσίαση Λίστας Όρων για Πολιτικές Βάσει του GDPR

Με τη βοήθεια των κανονισμών και αρχών σύμφωνα με το Γενικό Κανονισμό για τη Προστασία Δεδομένων που περιγράφονται στο υποκεφάλαιο 3.2.1 και με τη μελέτη ενός αριθμού από διαδικτυακές πλατφόρμες καταλήξαμε σε μια λίστα όρων για πολιτικές απορρήτου. Για παράδειγμα το δικαίωμα να αλλάξει ή να διορθώσει τα προσωπικά του δεδομένα που έχει ένας χρήστης σε μια διαδικτυακή πλατφόρμα όπως ορίζεται στο GDPR είναι Right to Rectification ενώ στη διαδικτυακή πλατφόρμα DiFens αναφέρεται ως “the right to correct and update”. Έτσι αυτός ο επιπρόσθετος όρος προστέθηκε στη λίστα όρων για πολιτικές βάσει του GDPR ως μια παραλλαγή του αρχικού όρου. Αυτή η λίστα βρίσκεται σε ένα αρχείο απλού κειμένου και είναι ανοικτή για επεξεργασία στο μέλλον. Πιο κάτω παρατίθεται η λίστα στον Πίνακα 3.2.3.1:

Lawfulness of Processing
Consent
Contract
right to withdraw consent
Withdraw consent
Right of Erasure
The Right to Request Deletion
Right To be Forgotten
Erase your Information
Request erasure of your personal data
Erase the Personal data
To Erase Your Data
Erase any personal data
Right Of Access
The Right To Access
Access Personal Data
Access Your Personal Information
The Right to Lodge a Complaint
Right to complaint
Right to File a Complaint
The Right to Obtain a Copy
Request a Copy of your Information
Request access to a copy of your personal data
Right to Information
Request access
Right to Data Portability
Right of Portability
Right to Transmit Those Data
The Right to Transmit Those Data
Right to Transmit Data
To Transmit Your Data
Right to Transmit Personal Data
Request the transfer of your personal data

Right to Receive the Personal Data
Right to Receive a Subset of the Personal Data
Right to receive a copy of your personal information
Right To Rectification
right to have incomplete personal data
Right to Request Proper Rectification
To Rectify Your Data
The Right to Correct and Update
Update or Correct your Information
Right to request the correction
Right to Correction
Right to Correct
Rectify
Right to demand processing restrictions
Right to restriction of processing
Restriction of Processing
To restrict Your data
Right to Restrict
Requests the Restriction of Their Use
Right of data subjects to be informed about the restriction
Right to Object
Right to Object at any time to Processing of Personal data
Right to Object at any time to Processing
Right to Object to Processing
Processing Objection
Object to processing

Πίνακας 3.2.3.1.Λίστα Όρων για Πολιτικές βάσει του GDPR

3.2.4 Ομαδοποίηση Λίστας Όρων για Πολιτικές Βάσει του GDPR

Λόγω των διαφορετικών εκδοχών κάθε κανονισμού διαχωρίσαμε σε ομάδες τη λίστα όρων βάσει του GDPR στις κατηγορίες που φαίνονται πιο κάτω:

Τα ονόματα της κάθε κατηγορίας είναι οι όροι όπως αναγράφονται στο Γενικό Κανονισμό για τη Προστασία Δεδομένων. Τα περιεχόμενα της κάθε κατηγορίας είναι οι παραλλαγές και αναγραμματισμοί του ονόματος αυτής.

1. Lawfulness of Processing:

Lawfulness of Processing
Consent
Contract
Right to withdraw Consent

Withdraw consent

Πίνακας 3.2.4.1. Κατηγορία Lawfulness of Processing

2. Right of Erasure:

Right of Erasure
The Right to Request Deletion
Right To be Forgotten
Erase your Information
Request erasure of your personal data
Erase the Personal data
To Erase Your Data
Erase any personal data

Πίνακας 3.2.4.2. Κατηγορία Right of Erasure

3. Right Of Access:

Right Of Access
The Right To Access
Access Personal Data
Access Your Personal Information
The Right to Lodge a Complaint
Right to complaint
Right to File a Complaint
The Right to Obtain a Copy
Request a Copy of your Information
Request access to a copy of your personal data
Right to Information
Request access

Πίνακας 3.2.4.3. Κατηγορία Right Of Access

4. Right to Data Portability:

Right to Data Portability
Right of Portability
Right to Transmit Those Data
The Right to Transmit Those Data
Right to Transmit Data
To Transmit Your Data
Right to Transmit Personal Data
Request the transfer of your personal data
Right to Receive the Personal Data

Right to Receive a Subset of the Personal Data
Right to receive a copy of your personal information

Πίνακας 3.2.4.4. Κατηγορία Right to Data Portability

5. Right to Rectification:

Right To Rectification
right to have incomplete personal data
Right to Request Proper Rectification
To Rectify Your Data
The Right to Correct and Update
Update or Correct your Information
Right to request the correction
Right to Correction
Right to Correct
Rectify

Πίνακας 3.2.4.5. Κατηγορία Right to Rectification

6. Right to restriction of processing:

Restriction of Processing
To restrict Your data
Right to Restrict
Right to demand processing restrictions
Right to restriction of processing
Requests the Restriction of Their Use
Right of data subjects to be informed about the restriction

Πίνακας 3.2.4.6. Κατηγορία Right to restriction of processing

7. Right to Object:

Right to Object
Right to Object at any time to Processing of Personal data
Right to Object at any time to Processing
Right to Object to Processing
Processing Objection
Object to processing

Πίνακας 3.2.4.7. Κατηγορία Right to Object

3.3 Περιήγηση στις σελίδες του privacy policy

data-processing
data-privacy
Login
about
Privacy
Policy
Register
privacy-policy
legal
privacy-policy-link
privacy_check
guidelines
loginlink
Login_link
sign_up
sign_up
register-form
en_eu

Πίνακας 3.3.1. Λίστα Όρων για Περιήγηση

Στη Εικόνα 3.5.1 παρουσιάζεται η Λίστα Όρων για Περιήγηση. Αυτή η λίστα προέκυψε από την μελέτη ενός αριθμού διαδικτυακών πλατφορμών με διερεύνηση στο πηγαίο κώδικα για να παρατηρηθεί πως γίνεται η περιήγηση στη σελίδα από ένα crawler. Στη περιήγηση μας από τη κύρια σελίδα στους συνδέσμους του privacy policy, contact form και registration form συλλέχθηκαν πληροφορίες όπως φράσεις-κλειδιά για να δημιουργηθεί μια λίστα όρων για περιήγηση. Αυτοί οι όροι που μαζεύτηκαν είναι μέρος του συνδέσμου στον οποίο βρισκόμασταν κατά την περιήγηση. Για παράδειγμα στο σύνδεσμο της διαδικτυακής πλατφόρμας DiFens (DIGITAL SECURITY FOR YOUNG ENTREPRENEURS) http://www.difens.eu/**data-processing**/ ο όρος που πάρθηκε είναι data-processing, αυτό μας δείχνει ότι βρίσκεται στη σελίδα της πολιτικής απορρήτου της διαδικτυακής πλατφόρμας. Κατά την περιήγηση υπήρξαν διάφορες εκδοχές των όρων, πχ data-processing, privacy-policy και αυτό βοήθησε στον εμπλουτισμό της Λίστας Όρων για Περιήγηση. Η δημιουργία της Λίστας μας βοήθησε με την έρευνα και μεθοδολογία που έγινε σε μεταγενέστερο στάδιο για την ανάπτυξη του crawler.

3.4 Λίστα Όρων Οντολογίας

Η λίστα όρων οντολογίας δημιουργήθηκε με τη βοήθεια μιας προϋπάρχων οντολογίας [15], η οποία είναι βασισμένη στο Γενικό Κανονισμό Προστασίας Δεδομένων (GDPR). Η διαδικασία που ακολουθήσαμε για να καταλήξουμε στη πιο κάτω λίστα είναι η εξής: πρώτα μαζέψαμε όλα τα ονόματα των κλάσεων και των υποκλάσεων που αποτελούσαν την οντολογία και τις εισαγάγαμε σε ένα αρχείο κειμένου. Στη συνέχεια αφαιρέσαμε τυχόν επαναλήψεις μεταξύ των όρων που συλλέχθηκαν από το προηγούμενο βήμα και καταλήξαμε στη τελική μορφή της λίστας όρων της οντολογίας. Αυτή η λίστα αποθηκεύτηκε σε ένα αρχείο κειμένου. Ένα κομμάτι της λίστας όρων της οντολογίας παρατίθεται στον Πίνακα 3.4.1. Ολόκληρος ο πίνακας παρουσιάζεται στο Παράρτημα Α.

Legal Resource Subdivision
Article
Chapter
Citation
Point
Recital
Section
Sub Point
Legal Resource
Concept
Principle
Accountability
Accuracy
Data Minimisation
Lawfulness Fairness And Transparency
Integrity And Confidentiality
Purpose Limitation
Storage Limitation
Compliance
Monitor Compliance
Demonstrating Consent
Report Data Breach
Maintain Record Of Breach
Notify Data Subject Of Breach
Notify Data Subject About DPO For Data Breach
Notify Data Subject About Consequences Of Data Breach
Notify Data Subject Of Measures Taken For Data Breach
Report Data Breach To DP AWithin72 Hours
Report Data Breach To Controller
Activity
Processing
Automated Processing
Automated Decision Making With Significant Effect
Confirming Or Matching Datasets

Large Scale Processing
Processing Affected Vulnerable Individuals
Unlawful Processing
Processing Sensitive Data
Processing Using Untested Technologies
Appointment Of Processors
Data Activity
Erase Data
Rectify Data
Collection Of Personal Data
Collection Mechanism
Store Data
Archive Data
Cross Border Transfer
Share Data With Third Party
Use Data
Provide Copy Of Personal Data
Should Be Commonly Used Format
Should Be Machine Readable
Should Be Structured
Should Support Reuse
Consent Activity
Withdrawing Consent
Obtaining Consent
Security Of Personal Data
Protection Against Accidental Loss
Protection Against Damage
Protection Against Destruction
Protection Against Unlawful Processing
Marketing
Direct Marketing
Identification Of Data Subject
Exercise Rights
Impact Assessment
Propogate Rights To Third Parties
Systematic Monitoring
Data
Personal Data
Sensitive Personal Data
Criminal Data
Genetic Data
Health Data
Racial Data
Anonymous Data
Pseudo Anonymous Data
Rights
Right To Restrict Processing
Accuracy Is Contested
Data No Longer Needed For Original Purpose
Right Of Erasure

Πίνακας 3.4.1. Λίστα Όρων της Οντολογίας

Κεφάλαιο 4

Υλοποίηση

4.1 Εισαγωγή

Στο πλαίσιο αυτής της διπλωματικής εργασίας, σχεδιάστηκε και υλοποιήθηκε ένα εργαλείο. Είναι ένα πρόγραμμα που αξιολογεί αυτοματοποιημένα πολιτικές απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR. Δίνει τη δυνατότητα σε όποιο έχει δημιουργήσει μια διαδικτυακή πλατφόρμα να δει πόσο νόμιμη αυτή η πλατφόρμα είναι, αλλά και ένας χρήστης μιας διαδικτυακής πλατφόρμας να δει αν η διαδικτυακή πλατφόρμα που χρησιμοποιεί έχει σωστή πολιτική. Αυτή η αξιολόγηση γίνεται είτε με τη χρήση της λίστας

που δημιουργήθηκε στα πλαίσια αυτής της διπλωματικής είτε με τη χρήση μια έτοιμης οντολογίας που βασίζεται στο GDPR.

Όπως εξηγήθηκε προηγουμένως, η προστασία της ιδιωτικής ζωής των προσωπικών δεδομένων σε συστήματα πληροφοριών δεν είναι μόνο κρίσιμη, αλλά υποβάλλεται και από τον νέο Γενικό Κανονισμό Προστασίας Δεδομένων που είναι υποχρεωτικός να εφαρμοστεί σε οποιοδήποτε σύστημα.

Όπως και πολλά άλλα συστήματα, αρκετές διαδικτυακές πλατφόρμες αποθηκεύουν και επεξεργάζονται προσωπικά στοιχεία των εγγεγραμμένων χρηστών τους. Πλατφόρμες που ανήκουν στην κατηγορία συστημάτων διαχείρισης προφίλ χρηστών, πρέπει να σχεδιαστούν και να αναπτυχθούν με τρόπο που θα τις καθιστά συμβατές με τον νέο κανονισμό προστασίας δεδομένων.

4.2 Απαιτήσεις

Οι απαιτήσεις για το πρόγραμμα που υλοποιήθηκε ανανεώνονταν καθώς η διπλωματική εργασία προχωρούσε για το λόγο ότι, ενώ αρχικά είχαμε μια μόνο κύρια απαίτηση που ήταν η αυτοματοποιημένη αξιολόγηση πολιτικών απορρήτου βάσει του GDPR, με τη πάροδο του χρόνου οι απαιτήσεις γίνονταν όλο και πιο ξεκάθαρες. Στο πίνακα πιο κάτω αναφέρονται όλες οι απαιτήσεις του συστήματος.

	Απαιτήσεις	Τύπος Απαίτησης
(1)	Ο οργανισμός/χρήστης να δώσει τον σύνδεσμο της διαδικτυακής πλατφόρμας	Απαίτηση Οργανισμού/Χρήστη
(2)	Ο crawler να παίρνει ως είσοδο το σύνδεσμο της διαδικτυακής πλατφόρμας	Απαίτηση Crawler
(3)	Ο crawler να παίρνει ως είσοδο τη λίστα όρων για περιήγηση	Απαίτηση Crawler
(4)	Δημιουργία φακέλου όπου θα μπουν τα αρχεία κειμένου	Απαίτηση Crawler
(5)	Γίνεται αναζήτηση της πλατφόρμας χρησιμοποιώντας τους όρους της λίστας για περιήγηση	Απαίτηση Crawler

(6)	Ανίχνευση μέχρι την εύρεση του συνδέσμου της πολιτικής απορρήτου της διαδικτυακής πλατφόρμας	Απαίτηση Crawler
(7)	Αποθήκευση πηγαίου κώδικα της σελίδας της πολιτικής απορρήτου σε ένα αρχείο κειμένου	Απαίτηση Crawler
(8)	Εμφάνιση μηνύματος λάθους αν δεν βρει συνδέσμους να ανιχνεύσει	Απαίτηση Crawler
(9)	Ο list parser παίρνει ως είσοδο την λίστα όρων βάσει του GDPR	Απαίτηση List Parser
(10)	Ο ontology parser παίρνει ως είσοδο την έτοιμη οντολογία βασισμένη στο GDPR	Απαίτηση Ontology Parser
(11)	Οι parsers παίρνουν ως είσοδο το πηγαίο κώδικα της πολιτικής απορρήτου	Απαίτηση Parser
(12)	Σύγκριση όρων με πολιτική απορρήτου	Απαίτηση Parser
(13)	Αποθήκευση αποτελεσμάτων σε αρχεία κειμένου	Απαίτηση Parser
(14)	Παίρνει ως είσοδο τον σύνδεσμο της διαδικτυακής πλατφόρμας από τον οργανισμό/χρήστη	Απαίτηση GUI
(15)	Διαχωρισμός της λίστας όρων βάσει του GDPR στις ομάδες που εξηγήθηκαν στο υποκεφάλαιο 3.2.4	Απαίτηση GUI
(16)	Δημιουργία δύο πινάκων για τη λίστα όρων βάσει το GDPR. Ο πρώτος πίνακας περιέχει όλους τους όρους και ο δεύτερος περιέχει μόνο τις ομάδες	Απαίτηση GUI
(17)	Διαχωρισμός των όρων της οντολογίας σε δύο πίνακες. Ο πρώτος πίνακας είναι οι όροι που βρέθηκαν και ο δεύτερος πίνακας είναι οι όροι που δεν βρέθηκαν στην πολιτική απορρήτου κατά τη σύγκριση	Απαίτηση GUI
(18)	Υπολογισμός score της διαδικτυακής πλατφόρμας	Απαίτηση GUI
(19)	Παρουσίαση αποτελεσμάτων στον οργανισμό/χρήστη	Απαίτηση GUI
(20)	Εμφάνιση αποτελεσμάτων σε εύλογο χρονικό διάστημα	Απαίτηση Συστήματος

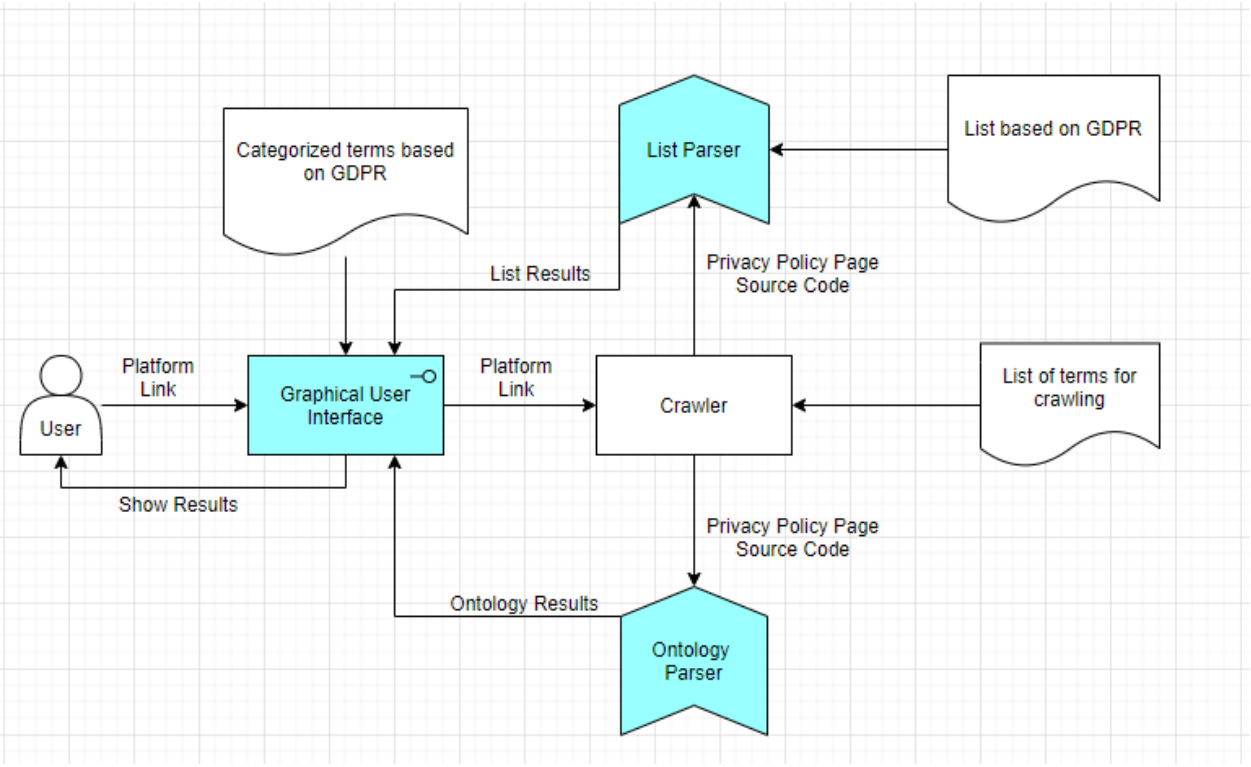
Πίνακας 4.2.1. Απαιτήσεις εργαλείου

Οι απαιτήσεις λογισμικού παρουσιάζονται στον Πίνακα 4.2.1 και ο τύπος των απαιτήσεων βασίζεται σε κομμάτια της υλοποίησης. Στη πρώτη στήλη του πίνακα ("Απαιτήσεις"), παρουσιάζονται οι απαιτήσεις του εργαλείου, ενώ στη δεύτερη στη στήλη του πίνακα ("Τύπος") παρουσιάζεται το αντίστοιχο κομμάτι της υλοποίησης. Κάθε απαίτηση αντιστοιχεί στο συγκεκριμένο κομμάτι της υλοποίησης που συνεπάγεται η διάταξη του πίνακα. Οι

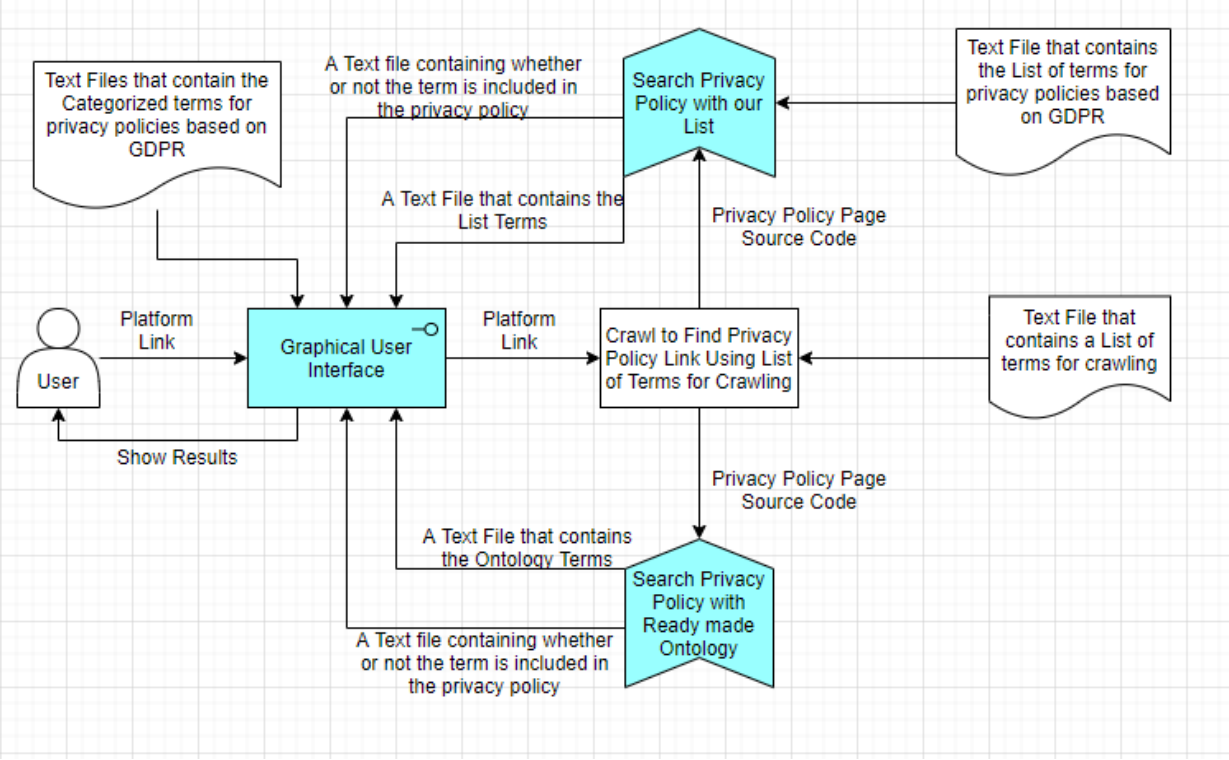
απαιτήσεις του συστήματος ανανεώνονταν καθ' όλη τη διάρκεια της υλοποίησης του εργαλείου.

4.3 Σχεδίαση Εργαλείου

4.3.1 Αρχιτεκτονική



Εικόνα 4.3.1.1. Αρχιτεκτονική Συστήματος



Εικόνα 4.3.1.2. Αναλυτική Αρχιτεκτονική Συστήματος

Στην εικόνα 4.3.1.1 παρουσιάζεται η αρχιτεκτονική του συστήματος. Ο χρήστης δίνει ώς είσοδο στο σύστημα το σύνδεσμο της διαδικτυακής πλατφόρμας. Ο crawler παίρνει το σύνδεσμο της διαδικτυακής πλατφόρμας. Επίσης παίρνει ως είσοδο και ένα αρχείο που περιέχει μια λίστα όρων για περιήγηση, έτσι ώστε να γίνει η περιήγηση στη διαδικτυακή πλατφόρμα βάσει αυτής της λίστας. Ο crawler χρησιμοποιώντας την λίστα όρών για περιήγηση διαπερνά το σύνδεσμο που δόθηκε ως είσοδος μέχρι να βρει το link της πολιτικής απορρήτου, ανεβάζει το κώδικα html της πολιτικής απορρήτου και σταματά. Στη συνέχεια εκτελούνται οι διαδικασίες List Parser και Ontology Parser. Στη List Parser αναζητά στην πολιτική απορρήτου να βρει αν περιέχει τους όρους της λίστας για πολιτικές βάσει του GDPR, Η Ontology Parser αναζητά στην πολιτική απορρήτου να βρει αν περιέχει τους όρους της οντολογίας. Μετά το γραφικό περιβάλλον χρήστη παίρνει από τη διαδικασία List Parser ένα αρχείο κειμένου που περιέχει την λίστα όρων για πολιτικές βάσει του GDPR και ένα αρχείο που περιέχει αν εμπεριέχεται ή όχι ο κάθε όρος στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας. Από τη διαδικασία Ontology Parser παίρνει ένα αρχείο κειμένου που περιέχει τους όρους της οντολογίας και και ένα αρχείο που περιέχει αν εμπεριέχεται ή όχι ο κάθε όρος στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας. Επίσης παίρνει ως είσοδο και τα αρχεία κειμένου που περιέχουν τους όρους για πολιτικές απορρήτου βάσει του Γενικού Κανονισμού για την Προστασία Δεδομένων ομαδοποιημένα όπως αναφέρεται και στο υποκεφάλαιο 3.2.4. Στη συνέχεια το γραφικό περιβάλλον χρήστη επεξεργάζεται τα δεδομένα, υπολογίζει τις βαθμολογίες για την ανάλυση της πολιτικής απορρήτου βάσει του

List Parser και βάσει του Ontology Parser και τους παρουσιάζει στο χρήστη. Στην εικόνα 4.3.1.2 παρουσιάζεται η αρχιτεκτονική του συστήματος με περισσότερη επεξήγηση. Σε αυτή την εικόνα για κάθε είσοδο, διεργασία και αποτελέσματα δίνεται μια περεταίρω εξήγηση έτσι ώστε ο αναγνώστης να πάρει μια ιδέα για τις αλληλεπιδράσεις μεταξύ των μερών που αποτελούν την αρχιτεκτονική του συστήματος. Η επεξήγηση αυτής της εικόνας είναι η ίδια με την επεξήγηση που βρίσκεται πιο πάνω. Στο Πίνακα 4.3.1.3 παρουσιάζονται οι αντιστοιχίες των αλλαγών στα στοιχεία στις δύο εικόνες, όσα δεν περιέχονται στον πιο κάτω πίνακα σημαίνει πως παραμένουν τα ίδια και στις δύο εικόνες:

Αρχιτεκτονική Συστήματος	Αναλυτική Αρχιτεκτονική Συστήματος
Crawler	Crawl to Find Privacy Policy Link Using List of Terms for Crawling
List Parser	Search Privacy Policy with our List
Ontology Parser	Search Privacy Policy with Ready made Ontology
List based on GDPR	Text File that contains the List of terms for privacy policies based on GDPR
List of terms for crawling	Text File that contains a List of terms for crawling
Categorized terms based on GDPR	Text Files that contain the Categorized terms for privacy policies based on GDPR
List Results	• A Text File that contains the List Terms • A Text file containing whether or not the term is included in the privacy policy
Ontology Results	• A Text File that contains the Ontology Terms • A Text file containing whether or not the term is included in the privacy policy

Πίνακας 4.3.1.3. Αντιστοιχία Διαφορετικών Στοιχείων

4.3.2 Εργαλεία

Σε αυτό το υποκεφάλαιο θα μιλήσουμε για τα εργαλεία που φάνηκαν χρήσιμα για το κομμάτι της υλοποίησης της διπλωματικής εργασίας.

4.3.2.1 PyCharm

Το PyCharm είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης (integrated development environment-IDE) που χρησιμοποιείται στον προγραμματισμό υπολογιστών, ειδικά για τη γλώσσα Python. Αναπτύχθηκε από μια τσεχική εταιρεία την JetBrains. Παρέχει ανάλυση κώδικα, ένα πρόγραμμα εντοπισμού σφαλμάτων γραφικών, μια ενσωματωμένη μονάδα δοκιμής, ενοποίηση με συστήματα ελέγχου εκδόσεων (VCSes) και υποστηρίζει την ανάπτυξη ιστού με το Django καθώς και την Επιστήμη δεδομένων με την Anaconda.

Το PyCharm υποστηρίζεται από πολλαπλές πλατφόρμες (cross-platform), με εκδόσεις σε Windows, macOS και Linux. Η κοινοτική έκδοση κυκλοφόρησε με την άδεια Apache, υπάρχει επίσης η επαγγελματική έκδοση με επιπλέον δυνατότητες η οποία κυκλοφόρησε με άδεια χρήσης.

4.3.2.2 Python

Η Python είναι μια ερμηνευμένη (interpreted), αντικειμενοστρεφής (object-oriented), γλώσσα προγραμματισμού υψηλού επιπέδου με δυναμική σημασιολογία (dynamic semantics). Η ενσωματωμένη δομή δεδομένων υψηλού επιπέδου, σε συνδυασμό με τη δυναμική πληκτρολόγηση και τη δυναμική δέσμευση, το καθιστούν κατάλληλο για την ταχεία ανάπτυξη εφαρμογών, καθώς και για χρήση ως γλώσσα δέσμης ενεργειών ή κόλλα για σύνδεση των υπαρχόντων στοιχείων. Η απλή, εύχρηστη σύνταξη της Python δίνει έμφαση στην αναγνωσιμότητα και συνεπώς μειώνει το κόστος συντήρησης του προγράμματος. Η Python υποστηρίζει λειτουργικές μονάδες και πακέτα, τα οποία ενθαρρύνουν τη διαμόρφωση του προγράμματος και την επαναχρησιμοποίηση κώδικα. Ο διερμηνέας (interpreter) Python και η εκτεταμένη τυπική βιβλιοθήκη διατίθενται σε πηγή ή δυαδική μορφή χωρίς χρέωση για όλες τις μεγάλες πλατφόρμες και μπορούν να διανεμηθούν ελεύθερα.

Δεδομένου ότι δεν υπάρχει βήμα μεταγλώττισης, ο κύκλος επεξεργασίας-δοκιμής-εντοπισμού σφαλμάτων είναι απίστευτα γρήγορος. Ο εντοπισμός σφαλμάτων των προγραμμάτων Python είναι εύκολος, ένα σφάλμα ή κακή είσοδος δεν θα προκαλέσει ποτέ σφάλμα segmentation. Αντ' αυτού, όταν ο διερμηνέας ανακαλύπτει ένα σφάλμα, δημιουργεί μια εξαίρεση. Όταν το πρόγραμμα δεν έχει την εξαίρεση, ο διερμηνέας εκτυπώνει ένα ίχνος στοίβας. Ένα πρόγραμμα εντοπισμού σφαλμάτων επιτρέπει την επιθεώρηση τοπικών και δημόσιων μεταβλητών, την αξιολόγηση αυθαίρετων εκφράσεων, τον καθορισμό σημείων διακοπής, τον έλεγχο του κώδικα μια γραμμή κάθε φορά και ούτω καθεξής.

4.3.2.3 Qt Designer

Το Qt Designer είναι το εργαλείο Qt για το σχεδιασμό και τη δημιουργία γραφικής διαπροσωπίας χρήστη (GUI) με Qt Widgets. Μπορείς να συνθέσεις και να προσαρμόσεις τα παράθυρα ή τους διαλόγους σου με τον τρόπο ότι βλέπετε-είναι-τι-παίρνετε (what-you-see-is-what-you-get-WYSIWYG) και να τα δοκιμάσεις χρησιμοποιώντας διαφορετικά στυλ και αναλύσεις.

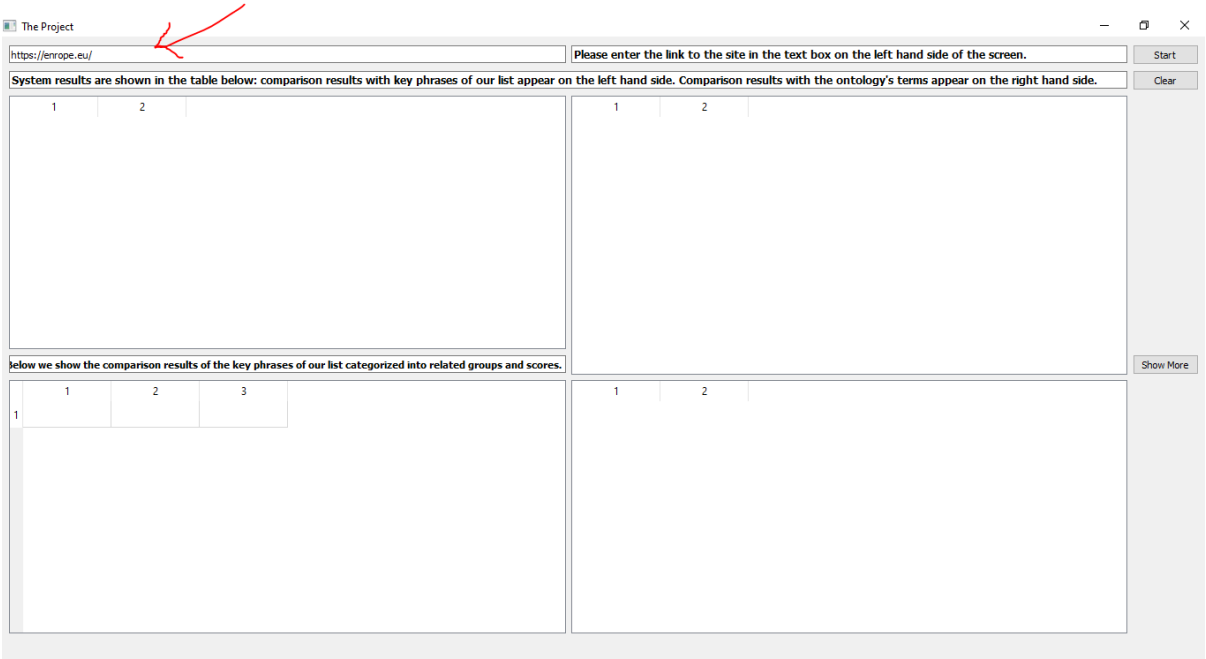
Τα γραφικά στοιχεία και οι φόρμες που δημιουργήθηκαν με το Qt Designer ενσωματώνονται απρόσκοπτα με προγραμματισμένο κώδικα, χρησιμοποιώντας τον μηχανισμό signals και slots Qt, έτσι ώστε να μπορεί κάποιος να αντιστοιχίζει εύκολα τη συμπεριφορά σε γραφικά στοιχεία. Όλες οι ιδιότητες που έχουν οριστεί στο Qt Designer μπορούν να αλλάξουν δυναμικά μέσα στον κώδικα. Επιπλέον, λειτουργίες όπως η προώθηση widget και τα προσαρμοσμένα πρόσθετα σάς επιτρέπουν να χρησιμοποιείτε τα δικά σας στοιχεία με το Qt Designer.

Στη Διπλωματική μας το εργαλείο Qt Designer ήταν μια πολύ χρήσιμη εφαρμογή γιατί σου έδινε την επιλογή να προσθέσεις όσα αντικείμενα θέλεις στο παράθυρο με drag and drop και μετά να μετατρέψεις αυτό το αρχείο σε κώδικα python. Έχοντας έτοιμο το πρότυπο σχεδίασης, προσθέσαμε τις διάφορες λειτουργίες που θα εκτελούνται στο γραφικό περιβάλλον χρήστη.

4.4 Υλοποίηση

4.4.1 Είσοδος Χρήστη

Η είσοδος γίνεται από το χρήστη ο οποίος δίνει το σύνδεσμο της διαδικτυακής πλατφόρμας στο γραφικό περιβάλλον όπως φαίνεται και στην Εικόνα 4.4.1.1.



Εικόνα 4.4.1.1. Είσοδος Συνδέσμου στο Σύστημα

Η καταχώρηση στο γραφικό περιβάλλον της εισόδου του χρήστη γίνεται με την εντολή: input = self.Link_Input.text(). Το δεξί κομμάτι της εντολής είναι η είσοδος του χρήστη και το αριστερά κομμάτι είναι μια μεταβλητή όπου θα αποθηκευτεί ο σύνδεσμος της διαδικτυακής πλατφόρμας.

4.4.2 Επεξήγηση Crawler

Μετά την είσοδο του συνδέσμου της διαδικτυακής πλατφόρμας από τον χρήστη στο γραφικό περιβάλλον καλείται μια μέθοδος τύπου constructor που αρχικοποιεί τις απαραίτητες μεταβλητές, η εντολή είναι: main(input) όπου το input είναι ο σύνδεσμος της διαδικτυακής πλατφόρμας. Ακολουθεί εξήγηση μεταβλητών και αρχείων που δημιουργούνται κατά την εκτέλεση της πιο πάνω εντολής:

- HOMEPAGE = Είναι ο σύνδεσμος της διαδικτυακής πλατφόρμας
- PROJECT_NAME = Είναι το όνομα του φακέλου όπου θα εισαχθούν τα αρχεία crawled.txt, queue.txt και htmlfile.txt.
- DOMAIN_NAME = Είναι το domain name του συνδέσου της διαδικτυακής πλατφόρμας , πχ https://enrope.eu/ το domain name θα είναι enrope.eu.

- QUEUE_FILE = Είναι ένα αρχείο κειμένου που περιέχει τους συνδέσμους που θα διαπεράσει ο crawler.
- CRAWLED_FILE = Είναι ένα αρχείο κειμένου που περιέχει τους συνδέσμους που έχει διαπεράσει ο crawler.
- HTML_FILE = Στο τέλος της περιήγησης στη πλατφόρμα αυτό το αρχείο θα περιέχει τον πηγαίο κώδικα της σελίδας της πολιτικής απορρήτου.
- gdpr_ontology = Είναι ένα αρχείο κειμένου που περιέχει όλους τους όρους της οντολογίας
- gdpr_mine = Είναι ένα αρχείο κειμένου που περιέχει τη λίστα όρων για πολιτικές απορρήτου βάσει του GDPR
- gdpr_mine_yes_no = Είναι ένα αρχείο κειμένου που περιέχει είτε YES αν ένας συγκεκριμένος όρος της λίστας συμπεριλαμβάνεται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας, είτε NO αν ένας συγκεκριμένος όρος της λίστας δεν συμπεριλαμβάνεται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας.
- gdpr_onto_words_yes = Είναι ένα αρχείο κειμένου που περιέχει τους όρους της οντολογίας που συμπεριλαμβάνονται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας.
- gdpr_onto_words_no = Είναι ένα αρχείο κειμένου που περιέχει τους όρους της οντολογίας που δεν συμπεριλαμβάνονται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας.
- gdpr_onto_yes = Είναι ένα αρχείο κειμένου που περιέχει YES το οποίο σημαίνει ότι ένας συγκεκριμένος όρος της οντολογίας συμπεριλαμβάνεται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας
- gdpr_onto_no = Είναι ένα αρχείο κειμένου που περιέχει NO το οποίο σημαίνει ότι ένας συγκεκριμένος όρος της οντολογίας δεν συμπεριλαμβάνεται στη πολιτική απορρήτου της διαδικτυακής πλατφόρμας.
- flag_group = Είναι ένα αρχείο κειμένου που περιέχει 0 και 1. Αν είναι 0 σημαίνει ότι στην ομάδα Lawfulness of Processing για παράδειγμα δεν βρήκε κάποιο από τους όρους αυτής της ομάδας στην πολιτική απορρήτου της διαδικτυακής πλατφόρμας, αν είναι 1 σημαίνει ότι βρήκε τουλάχιστο ένα από τους όρους αυτής της ομάδας.
- Καλείται η μέθοδος Spider που είναι μέθοδος τύπου constructor για τη κλάση Spider η οποία παίρνει ως είσοδο το PROJECT_NAME, HOMEPAGE και DOMAIN_NAME και με τη σειρά της καλεί τη μέθοδο boot που δημιουργεί το φάκελο που θα μπουν τα αρχεία κειμένου, επίσης δημιουργεί τα αρχεία εφόσον δεν υπάρχουν. Ακόμα καλεί τη μέθοδο crawl_page που παίρνει ως είσοδο τον σύνδεσμο

της σελίδας που βρίσκεται εκείνη τη στιγμή και αν αυτός ο σύνδεσμος δεν είναι μέσα στο αρχείο crawled τότε γίνεται η αναζήτηση σε αυτόν το σύνδεσμο. Μαζεύονται οι σύνδεσμοι που περιέχονται στο πηγαίο κώδικα του συνδέσμου που περιέχουν ένα από τους όρους που βρίσκονται στη λίστα που εξηγήθηκε στο υποκεφάλαιο 3.5 και προθέτονται στην ουρά αναμονής. Ταυτόχρονα ο αρχικός σύνδεσμος αφαιρείται από την ουρά αναμονής, προστίθεται στο αρχείο crawled και ενημερώνονται τα αρχεία για τις αλλαγές. Πιο κάτω παραθέτω τον κώδικα για τις μεθόδους Spider(), boot() και crawl_page() στην Εικόνα 4.4.2.1:

```
def __init__(self, project_name, base_url, domain_name):
    Spider.project_name = project_name
    Spider.base_url = base_url
    Spider.domain_name = domain_name
    Spider.queue_file = Spider.project_name + '/queue.txt'
    Spider.crawled_file = Spider.project_name + '/crawled.txt'
    Spider.html_file = Spider.project_name + '/htmlfile.txt'
    self.boot()
    self.crawl_page('First spider', Spider.base_url)

@staticmethod
def boot():
    create_project_dir(Spider.project_name)
    create_data_files(Spider.project_name,Spider.base_url)
    Spider.queue = file_to_set(Spider.queue_file)
    Spider.crawled = file_to_set(Spider.crawled_file)
    Spider.html_text = file_to_set(Spider.html_file)


@staticmethod
def crawl_page(thread_name, page_url):
    if page_url not in Spider.crawled:
        Spider.add_links_to_queue(Spider.gather_links(page_url))
        Spider.queue.remove(page_url)
        Spider.crawled.add(page_url)
        update_files()
```

Εικόνα 4.4.2.1. Κώδικας Μεθόδων Spider, boot και crawl_page

Στη συνέχεια καλείται η μέθοδος create_workers() και ο κώδικας αυτής παρατίθεται στην Εικόνα 4.4.2.2:

```
# Create worker threads
@staticmethod
def create_workers():
    for _ in range(main.NUMBER_OF_THREADS):
        t = threading.Thread(target=main.work)
        t.daemon = True
        t.start()
```

Εικόνα 4.4.2.2. Κώδικας Μεθόδου create_workers()

Στη κλάση main χρησιμοποιούνται νήματα για τη διεξαγωγή των εργασιών με περισσότερη ταχύτητα, ο αριθμός των νημάτων βρίσκεται στη μεταβλητή NUMBER_OF_THREADS και είναι 8. Αυτή η μέθοδος δημιουργεί τους «εργάτες» που αντιπροσωπεύουν αυτά τα νήματα και καλείται η μέθοδος work η οποία αναθέτει δουλειά που έχει σειρά τη συγκεκριμένη χρονική στιγμή.

```
    # Do the next Job in the queue
    @staticmethod
    def work():
        while True:
            url = main.queue.get()
            Spider.crawl_page(threading.current_thread().name, url)
            main.queue.task_done()
```

Εικόνα 4.4.2.3. Κώδικας Μεθόδου work()

Η μέθοδος work() όπως φαίνεται στην Εικόνα 4.4.2.3 εφόσον υπάρχουν σύνδεσμοι στην ουρά καλεί την διαδικασία crawl_page() δίνοντας είσοδο τον σύνδεσμο και το όνομα του νήματος και ειδοποιεί όταν τελειώσει τη δουλειά του κάποιο νήμα την ουρά.

Στη συνέχεια καλείται η μέθοδος crawl() και ο κώδικας αυτής παρατίθεται στην Εικόνα 4.4.2.4:

```
    # Check if there are items in the queue if yes crawl
    @staticmethod
    def crawl():
        queued_links = file_to_set(main.QUEUE_FILE)
        if len(queued_links) > 0:
            # print(str(len(queued_links)) + 'links in the queue')
            main.create_jobs()
```

Εικόνα 4.4.2.4. Κώδικας Μεθόδου crawl()

Η μέθοδος crawl() όπως φαίνεται στην Εικόνα 4.4.2.4 ελέγχει αν υπάρχουν σύνδεσμοι στην ουρά, αν ναι τότε καλείται η μέθοδος create_jobs().

```
    # Each queued link is a new job
    @staticmethod
    def create_jobs():
        for link in file_to_set(main.QUEUE_FILE):
            main.queue.put(link)
        main.queue.join()
        main.crawl()
```

Εικόνα 4.4.2.5. Κώδικας Μεθόδου create_jobs()

Η μέθοδος create_jobs() όπως φαίνεται στην Εικόνα 4.4.2.5 βάζει όλους τους συνδέσμους από το αρχείο QUEUE_FILE στην ουρά. Στη συνέχεια καλεί τη μέθοδο crawl(). Τέλος, τερματίζει όταν φτάσει στον σύνδεσμο της πολιτικής απορρήτου της διαδικτυακής πλατφόρμας και αποθηκεύσει σε ένα αρχείο κειμένου το πηγαίο κώδικα της πολιτικής.

4.4.3 Ανάλυση Με τη Χρήση Λίστας Όρων για Πολιτικές Βάσει του GDPR

Μετά από το τερματισμό του Crawler καλείται η List Parser και η μέθοδος που την αντιπροσωπεύει είναι η search_policy_my_list() η οποία παίρνει ως είσοδο το αρχείο κειμένου που περιέχει το πηγαίο κώδικα της σελίδας της πολιτικής απορρήτου και το αρχείο κειμένου που περιέχει τη λίστα όρων βάσει του GDPR. Στην Εικόνα 4.4.3.1 πιο κάτω παρατίθεται το κομμάτι κώδικα για τη διεργασία search_policy_my_list():

```
    @staticmethod
    def search_policy_my_list():
        count_YES = 0
        f = open("gdpr.txt", "r")
        gdpr_words = f.readline().rstrip()
        while gdpr_words != '':
            text = 0
            for line in str(Spider.html_text).splitlines():
                result = re.findall(gdpr_words.lower(), str(line).lower())
                if result:
                    text = 1
                    break
            if text == 1:
                append_to_file(main.gdpr_mine, gdpr_words)
                append_to_file(main.gdpr_mine_yes_no, "YES")
                count_YES += 1
            elif text == 0:
                append_to_file(main.gdpr_mine, gdpr_words)
                append_to_file(main.gdpr_mine_yes_no, "NO")
            gdpr_words = f.readline().rstrip()
        append_to_file(main.gdpr_mine, "Number of Key Terms Found:")
        append_to_file(main.gdpr_mine_yes_no, repr(count_YES))
```

Εικόνα 4.4.3.1. Κώδικας Μεθόδου search_policy_my_list()

Η μέθοδος search_policy_my_list() δηλώνει μια μεταβλητή που μετρά πόσοι όροι από τη λίστα περιέχονται στη πολιτική απορρήτου (count_YES). Στη συνέχεια διαβάζει μία γραμμή κάθε επανάληψη από το αρχείο κειμένου που περιέχει τη λίστα όρων βάσει του GDPR. Ακολούθως ορίζει μια μεταβλητή text με τη τιμή 0 και αρχίζει να συγκρίνει κάθε όρο της λίστας με όλο το πηγαίο κώδικα της σελίδας της πολιτικής απορρήτου χωρίς να λαμβάνει υπόψη αν τα γράμματα είναι κεφαλαία ή μικρά (case insensitive). Αν βρει τον όρο στην πολιτική απορρήτου τότε η μεταβλητή text γίνεται 1 και τερματίζει την αναζήτηση για το συγκεκριμένο όρο. Μετά ελέγχουμε αν η μεταβλητή text είναι 0 ή 1, αν είναι 0 τότε βάζουμε στο αρχείο κειμένου gdpr_mine τον όρο που ελέγχουμε εκείνη τη στιγμή και στο αρχείο κειμένου gdpr_mine_yes_no βάζουμε τη λέξη NO, δηλαδή δεν συμπεριλαμβάνεται ο συγκεκριμένος όρος στη πολιτική απορρήτου. Αν είναι 1 τότε βάζουμε στο αρχείο κειμένου gdpr_mine τον όρο που ελέγχουμε εκείνη τη στιγμή και στο αρχείο κειμένου gdpr_mine_yes_no βάζουμε τη λέξη YES, δηλαδή ο συγκεκριμένος όρος συμπεριλαμβάνεται στη πολιτική απορρήτου και αυξάνει την μεταβλητή count_YES κατά 1.

Διαβάζει την επόμενη γραμμή στη λίστα όρων για πολιτικές βάσει του GDPR και επαναλαμβάνει τη διαδικασία μέχρι να έχει ελέγξει όλους τους όρους. Τέλος εισαγάγει στο τέλος του αρχείου gdpr_mine_yes_no τον αριθμό των όρων που βρέθηκαν στη πολιτική απορρήτου. Στη συγκεκριμένη μέθοδο φάνηκε χρήσιμη η βιβλιοθήκη re για την εντολή re.findall(gdpr_words.lower(), str(line).lower()) που έκανε την σύγκριση του όρου της λίστας με την πολιτική απορρήτου. Την εισαγωγή των δεδομένων στα αρχεία κειμένου γινόταν με τη διεργασία append_to_file().

```
#Add data onto an existing file
def append_to_file(path, data):
    with open(path, 'a') as file:
        file.write(data + '\n')
```

Εικόνα 4.4.3.2. Κώδικας Μεθόδου append_to_file()

Η μέθοδος append_to_file() όπως φαίνεται στην Εικόνα 4.4.3.2 προσθέτει δεδομένα σε ένα είδη υπάρχον αρχείο και αλλάζει γραμμή. Παίρνει ως είσοδο το μονοπάτι που οδηγεί στο αρχείο και τα δεδομένα που θα εισαχθούν στο συγκεκριμένο αρχείο.

4.4.4 Ανάλυση με τη Χρήση Έτοιμης Οντολογίας

Μετά από το τερματισμό του Crawler και της List Parser καλείται η Ontology Parser και η μέθοδος που την αντιπροσωπεύει είναι η search_policy_ontology() η οποία παίρνει ως είσοδο το αρχείο κειμένου που περιέχει το πηγαίο κώδικα της σελίδας της πολιτικής απορρήτου και το αρχείο κειμένου που περιέχει τη λίστα όρων της οντολογίας. Για τη διαχείριση και επεξεργασία της οντολογίας χρησιμοποιούμε τη βιβλιοθήκη owlready2. Πριν την εκτέλεση της μεθόδου search_policy_ontology(), ανεβάζουμε την έτοιμη οντολογία σε μορφή xml και την αποθηκεύουμε σε μια μεταβλητή με την εντολή:

```
onto = get_ontology("C:/Users/user/PycharmProjects/CrawlerProject/ontology_gdpr.xml").load()
```

Στη συνέχεια παίρνουμε τους όρους από την οντολογία αφαιρούμε της επαναλήψεις των όρων και τους βάζουμε στο αρχείο gdpr_ontology_no_duplicates.

Στην Εικόνα 4.4.4.1 πιο κάτω παρατίθεται το κομμάτι κώδικα για τη διεργασία search_policy_ontology():

```
@staticmethod
def search_policy_ontology():
    count_YES = 0
    count_NO = 0
    f = open("gdpr_ontology_no_duplicates.txt", "r")
    gdpr_words_ontology = f.readline().rstrip()
    while gdpr_words_ontology != '':
        text = 0
        for line in str(Spider.html_text).splitlines():
            result = re.findall(gdpr_words_ontology.lower(), str(line).lower())
            if result:
                text = 1
                break
        if text == 1:
            append_to_file(main.gdpr_onto_words_yes, gdpr_words_ontology)
            append_to_file(main.gdpr_onto_yes, "YES")
            count_YES += 1
        elif text == 0:
            append_to_file(main.gdpr_onto_words_no, gdpr_words_ontology)
            append_to_file(main.gdpr_onto_no, "NO")
            count_NO += 1
        gdpr_words_ontology = f.readline().rstrip()
    append_to_file(main.gdpr_onto_words_yes, "Number of Key Terms Found:")
    append_to_file(main.gdpr_onto_yes, repr(count_YES))
    append_to_file(main.gdpr_onto_words_no, "Number of Key Terms NOT Found:")
    append_to_file(main.gdpr_onto_no, repr(count_NO))
```

Εικόνα 4.4.4.1. Κώδικας Μεθόδου search_policy_ontology()

Η μέθοδος search_policy_ontology() δηλώνει μια μεταβλητή που μετρά πόσοι όροι από τη λίστα συμπεριλαμβάνονται στη πολιτική απορρήτου (count_YES) και μία μεταβλητή που μετρά πόσοι όροι από τη λίστα δεν συμπεριλαμβάνονται στη πολιτική απορρήτου (count_NO). Στη συνέχεια διαβάζει μία γραμμή κάθε επανάληψη από το αρχείο κειμένου που περιέχει τη λίστα όρων της οντολογίας. Ακολούθως ορίζει μια μεταβλητή text με τη τιμή 0 και αρχίζει να συγκρίνει κάθε όρο της λίστας με όλο το πηγαίο κώδικα της σελίδας της πολιτικής απορρήτου χωρίς να λαμβάνει υπόψη αν τα γράμματα είναι κεφαλαία ή μικρά (case insensitive). Αν βρει τον όρο στην πολιτική απορρήτου τότε η μεταβλητή text γίνεται 1 και τερματίζει την αναζήτηση για το συγκεκριμένο όρο. Μετά ελέγχουμε αν η μεταβλητή text είναι 0 ή 1, αν είναι 0 τότε βάζουμε στο αρχείο κειμένου gdpr_onto_words_no τον όρο που ελέγχουμε εκείνη τη στιγμή και στο αρχείο κειμένου gdpr_onto_no βάζουμε τη λέξη NO, δηλαδή δεν συμπεριλαμβάνεται ο συγκεκριμένος όρος στη πολιτική απορρήτου και αυξάνει την μεταβλητή count_NO κατά 1. Αν είναι 1 τότε βάζουμε στο αρχείο κειμένου gdpr_onto_words_yes τον όρο που ελέγχουμε εκείνη τη στιγμή και στο αρχείο κειμένου gdpr_onto_yes βάζουμε τη λέξη YES, δηλαδή ο συγκεκριμένος όρος συμπεριλαμβάνεται στη πολιτική απορρήτου και αυξάνει την μεταβλητή count_YES κατά 1.

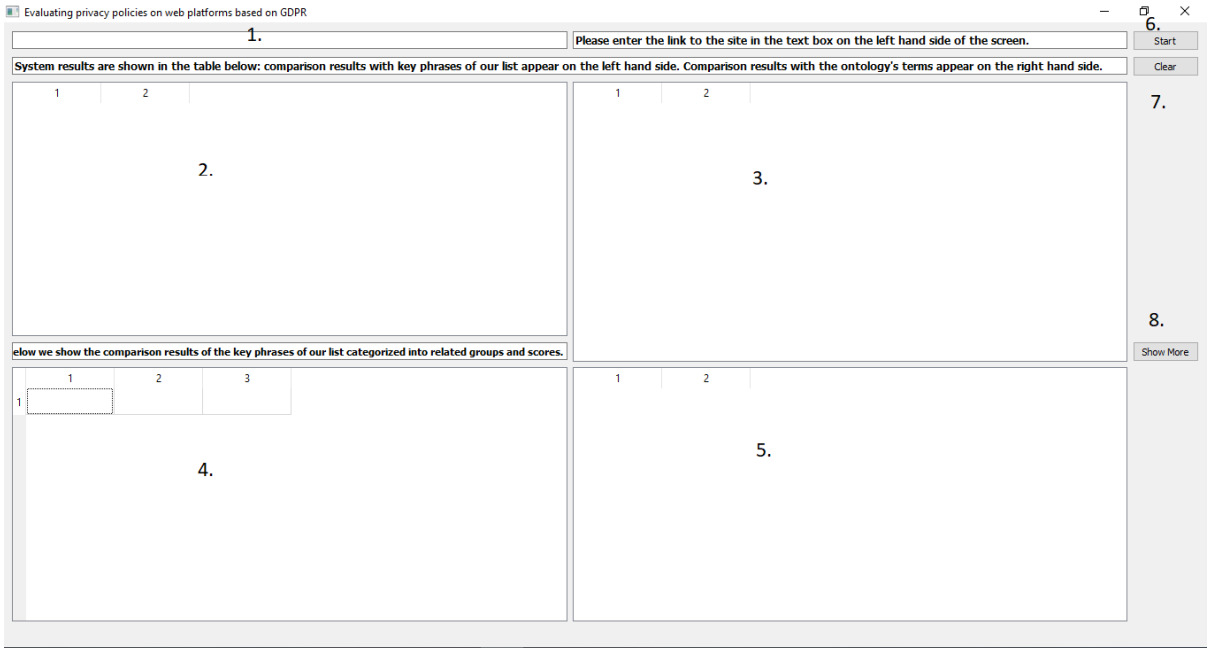
Διαβάζει την επόμενη γραμμή στη λίστα όρων της οντολογίας και επαναλαμβάνει τη διαδικασία μέχρι να έχει ελέγξει όλους τους όρους. Τέλος εισαγάγει στο τέλος του αρχείου gdpr_onto_yes τον αριθμό των όρων που βρέθηκαν στη πολιτική απορρήτου και στο αρχείο gdpr_onto_no τον αριθμό των όρων που δεν βρέθηκαν στη πολιτική απορρήτου. Στη συγκεκριμένη μέθοδο φάνηκε χρήσιμη η βιβλιοθήκη re για την εντολή re.findall(gdpr_words.lower(), str(line).lower()) που έκανε την σύγκριση του όρου της λίστας με την πολιτική απορρήτου. Την εισαγωγή των δεδομένων στα αρχεία κειμένου γινόταν με τη διεργασία append_to_file() όπου η λειτουργία της εξηγείτε στο υποκεφάλαιο 4.4.3.

4.4.5 Γραφικό Περιβάλλον Χρήστη

Σε αυτό το υποκεφάλαιο θα δώσουμε μια αναλυτική περιγραφή στο σχεδιασμό του γραφικού περιβάλλοντος χρήστη , πως γίνεται η επεξεργασία των δεδομένων που παίρνει από τις συναρτήσεις List Parser και Ontology Parser και πως τα αποτελέσματα παρουσιάζονται στο χρήστη. Όλες οι λειτουργίες του γραφικού περιβάλλοντος χρήστη υλοποιούνται στη κλάση Design_Template.py και όλες οι εικόνες κώδικα που ακολουθούν είναι από αυτή τη κλάση.

4.4.5.1 Περιγραφή Γραφικού Περιβάλλοντος Χρήστη

Για τη δημιουργία του γραφικού περιβάλλοντος χρήστη μας φάνηκε χρήσιμη η βιβλιοθήκη PyQt5. Στην Εικόνα 4.4.5.1.1 παρουσιάζεται το γραφικό περιβάλλον χρήστη και είναι αριθμημένα όλα τα σημεία τα οποία θα επεξηγηθούν σε αυτό το υποκεφάλαιο.



Εικόνα 4.4.5.1.1. Γραφικό Περιβάλλον Χρήστη

Επεξήγηση σημείων:

- Το σημείο 1. είναι μια ετικέτα εισαγωγής κειμένου όπου ο χρήστης/οργανισμός αποθέτει τον σύνδεσμο της διαδικτυακής πλατφόρμας το οποίο είναι τύπου QLineEdit.
- Τα σημεία 2,3,4 και 5 είναι πίνακες τύπου QTableWidget
- Το σημείο 2. είναι ο πίνακας που θα εμφανιστούν τα αποτελέσματα που πάρθηκαν από τη διεργασία List Parser όπως εξηγούνται στο υποκεφάλαιο 4.4.3
- Το σημείο 3. και 5. είναι οι πίνακες που θα εμφανιστούν τα αποτελέσματα που πάρθηκαν από τη διεργασία Ontology Parser όπως εξηγούνται στο υποκεφάλαιο 4.4.4. Το σημείο 3 θα περιέχει τους όρους που συμπεριλαμβάνονται στη πολιτική απορρήτου και το σημείο 5 θα περιέχει τους όρους που δεν συμπεριλαμβάνονται στη πολιτική απορρήτου.

- Το σημείο 4. εμφανίζονται οι ομάδες όπως ορίζονται στο υποκεφάλαιο 3.2.4, τους όρους που περιέχει η κάθε ομάδα και το σκορ της κάθε ομάδας. Στο τέλος της στήλης σκορ εμφανίζεται η συνολική βαθμολογία της σελίδας.
- Τα σημεία 6,7,8 είναι τύπου κουμπιά τύπου QPushButton
- Το σημείο 6. είναι το κουμπί 'Start' το οποίο όταν ενεργοποιηθεί ξεκινά η εκτέλεση του προγράμματος
- Το σημείο 7. είναι το κουμπί 'Clear' το οποίο όταν ενεργοποιηθεί διαγράφει τα στοιχεία που περιέχονται στα σημεία 1,2,3,4,5
- Το σημείο 8. είναι το κουμπί 'Show More' το οποίο όταν ενεργοποιηθεί εμφανίζει τα δεδομένα του πίνακα στο σημείο 5.

4.4.5.2 Επεξεργασία δεδομένων

Το Γραφικό περιβάλλον χρήστη παίρνει ως είσοδο τα αποτελέσματα των μεθόδων List Parser και Ontology Parser. Τα αποτελέσματα για τη διαδικασία List Parser είναι η λίστα όρων για πολιτικές βάσει του GDPR και ένα αρχείο που περιέχει αν συμπεριλαμβάνονται ή όχι οι όροι στη πολιτική απορρήτου (YES/NO). Τα αποτελέσματα για τη διαδικασία Ontology Parser είναι η λίστα όρων της οντολογίας, ένα αρχείο που περιέχει τους όρους που συμπεριλαμβάνονται στη πολιτική απορρήτου (YES) και ένα αρχείο που περιέχει τους όρους που δεν συμπεριλαμβάνονται στη πολιτική απορρήτου (NO). Επίσης παίρνει τα αρχεία κειμένου που περιέχουν τις ομάδες και τους όρους που τις αποτελούν.

Για την επεξήγηση της επεξεργασίας δεδομένων θα χρησιμοποιήσω την αρίθμηση από την Εικόνα 4.4.5.1.1 στο υποκεφάλαιο 4.4.5.1.

Αφού ο χρήστης/οργανισμός εισάγει το σύνδεσμο της διαδικτυακής πλατφόρμας και ενεργοποιήσει το σημείο 6. που είναι το κουμπί Start καλείται η μέθοδος Run με την εντολή: self.Start_Button.clicked.connect(self.Run). Στη συνέχεια για το γέμισμα του πίνακα στο σημείο 2. εκτελείται ο κώδικας που δίνεται στην Εικόνα 4.4.5.2.1 πιο κάτω:

```
f1 = open(main.gdpr_mine, "r")
f2 = open(main.gdpr_mine_yes_no, "r")
mine = f1.readline().rstrip()
mine_yes_no = f2.readline().rstrip()
while mine != '' and mine_yes_no != '':
    list.insert(c, mine)
    list_yes_no.insert(c, mine_yes_no)
    mine = f1.readline().rstrip()
    mine_yes_no = f2.readline().rstrip()
    c += 1
header_labels = ['KEY TERMS', 'INCLUDED']
self.tableWidget.setHorizontalHeaderLabels(header_labels)
while i != len(list):
    self.tableWidget.insertRow(i)
    self.tableWidget.setItem(i, 0, QTableWidgetItem(list[i]))
    self.tableWidget.setItem(i, 1, QTableWidgetItem(list_yes_no[i]))
    i += 1
```

Εικόνα 4.4.5.2.1. Γέμισμα Πίνακα Σημείου 2.

Αρχικά διαβάζει τα δεδομένα που πήρε από τη διαδικασία List Parser. Είναι το αρχείο gdpr_mine και το αρχείο gdpr_mine_yes_no, επαναλαμβάνει την ακόλουθη διαδικασία μέχρι να διαπεράσει όλα τα δεδομένα και στα δύο αρχεία. Εισάγει τα στοιχεία των δύο αρχείων κειμένου σε δύο λίστες (list, list_yes_no). Αφού ολοκληρωθεί η πιο πάνω διαδικασία δημιουργεί τις δύο στήλες του πίνακα ('KEY TERMS', 'INCLUDED'). Η πρώτη στήλη θα περιέχει του όρους της λίστας για πολιτικές βάσει το GDPR και η δεύτερη στήλη θα περιέχει αν ο κάθε όρος συμπεριλαμβάνεται ή όχι στη πολιτική (YES/NO). Παίρνει κάθε στοιχείο από τις λίστες που δημιουργήθηκαν πιο πάνω και τα προσθέτει στο πίνακα.

Στη συνέχεια για το γέμισμα του πίνακα στο σημείο 3. εκτελείται ο κώδικας που δίνεται στην Εικόνα 4.4.5.2.2 πιο κάτω:

```
f3 = open(main.gdpr_onto_words_yes, "r")
f4 = open(main.gdpr_onto_yes, "r")
on_data = f3.readline().rstrip()
on_data_yes = f4.readline().rstrip()
while on_data != '' and on_data_yes != '':
    list_onto.insert(d, on_data)
    list_onto_yes.insert(d, on_data_yes)
    on_data = f3.readline().rstrip()
    on_data_yes = f4.readline().rstrip()
    d += 1
header_labels = ['KEY TERMS', 'INCLUDED']
self.tableWidget_2.setHorizontalHeaderLabels(header_labels)
while j != len(list_onto):
    self.tableWidget_2.insertRow(j)
    self.tableWidget_2.setItem(j, 0, QTableWidgetItem(list_onto[j]))
    self.tableWidget_2.setItem(j, 1, QTableWidgetItem(list_onto_yes[j]))
    j += 1
```

Εικόνα 4.4.5.2.2. Γέμισμα Πίνακα Σημείου 3.

Διαβάζει πρώτα τα δεδομένα που πήρε από τη διαδικασία Ontology Parser. Είναι το αρχείο gdpr_onto_words_yes και το αρχείο gdpr_onto_yes, επαναλαμβάνει την ακόλουθη διαδικασία μέχρι να διαπεράσει όλα τα δεδομένα και στα δύο αρχεία. Εισάγει τα στοιχεία των δύο αρχείων κειμένου σε δύο λίστες (list_onto, list_onto_yes). Αφού ολοκληρωθεί η πιο πάνω διαδικασία δημιουργεί τις δύο στήλες του πίνακα ('KEY TERMS', 'INCLUDED'). Η πρώτη στήλη θα περιέχει όλους του όρους της οντολογίας που συμπεριλαμβάνονται στη πολιτική και η δεύτερη στήλη θα περιέχει τη λέξη YES. Παίρνει κάθε στοιχείο από τις λίστες που δημιουργήθηκαν πιο πάνω και τα προσθέτει στο πίνακα.

Στη συνέχεια για το γέμισμα του πίνακα στο σημείο 4. εκτελείται ο κώδικας που δίνεται στις Εικόνες 4.4.5.2.3, 4.4.5.2.4, 4.4.5.2.5 πιο κάτω:

```
f3 = open("Lawfulness_of_Processing.txt", "r")
f4 = open("Right_to_erasure.txt", "r")
f5 = open("Right_of_access.txt", "r")
f6 = open("Right_to_data_portability.txt", "r")
f7 = open("Right_To_Rectification.txt", "r")
f8 = open("Right_to_restriction.txt", "r")
f9 = open("Right_to_object.txt", "r")
lawfulness = f3.readline().rstrip()
while lawfulness != '':
    flag_law = 0
    for i in range(len(main.gdpr_mine)):
        result = re.search(lawfulness, list[i])
        result_yes = re.search(list_yes_no[i], 'YES')
        if result and result_yes:
            flag_law = 1
            break
    if flag_law == 1:
        break
    lawfulness = f3.readline().rstrip()
f3.close()
f11 = open("Lawfulness_of_Processing.txt", "r")
law = f11.readline().rstrip()
while law != '':
    s_l += 1
    law = f11.readline().rstrip()
```

Εικόνα 4.4.5.2.3. Γέμισμα Πίνακα Σημείου 4.

Αρχικά διαβάζει από τα αρχεία κειμένου που περιέχουν τις ομάδες και τους όρους που τις αποτελούν. Στη συνέχεια ορίζει μια μεταβλητή flag_law ίση με 0, εντοπίζει το συγκεκριμένο όρο που βρίσκεται στο αρχείο Lawfulness_of_Processing.txt μέσα στη λίστα όρων για πολιτικές βάσει του GDPR και βλέπει αν αυτός ο όρος βρισκόταν στην πολιτική, αν ναι τότε η μεταβλητή flag_law γίνεται 1 και τερματίζει την επανάληψη. Αν είναι 0 τότε συνεχίζεται η επανάληψη μέχρι να τελειώσουν τα στοιχεία. Στη συνέχεια καταμετρά τους όρους της κάθε

ομάδας. Η πιο πάνω διαδικασία συνεχίζεται για όλα τα αρχεία κειμένου που περιέχουν τις ομάδες και τους όρους που τις αποτελούν.

```
    append_to_file(main.flag_group, str(flag_law))
    append_to_file(main.flag_group, str(flag_erase))
    append_to_file(main.flag_group, str(flag_access))
    append_to_file(main.flag_group, str(flag_port))
    append_to_file(main.flag_group, str(flag_rect))
    append_to_file(main.flag_group, str(flag_rest))
    append_to_file(main.flag_group, str(flag_obj))
    f18 = open(main.flag_group, "r")
    f19 = open("groups.txt", "r")
    flag = f18.readline().rstrip()
    name = f19.readline().rstrip()
    while flag != '' and name != '':
        list_name.insert(h, name)
        list_flag.insert(h, flag)
        if flag == '1':
            count_ones += 1
        flag = f18.readline().rstrip()
        name = f19.readline().rstrip()
        h += 1
```

Εικόνα 4.4.5.2.4. Γέμισμα Πίνακα Σημείου 4. Συνέχεια

Αφού ολοκληρώσει η διαδικασία στην Εικόνα 4.4.5.2.3 για όλα τα αρχεία κειμένου των ομάδων, όλες οι μεταβλητές τύπου flag (0 ή 1) τοποθετούνται στο αρχείο κειμένου flag_group. Ακολούθως διαβάζει από τα αρχεία flag_group και groups (περιέχει το όνομα κάθε ομάδας) και τοποθετεί τα στοιχεία τους σε δύο λίστες list_name και list_flag. Ταυτόχρονα υπολογίζει πόσες φορές εμφανίστηκε ο αριθμός 1 στο αρχείο flag_group.

```
    header_labels_group = ['GROUP NAME', 'INCLUDING TERMS', 'SCORE']
    self.tableWidget_3.setHorizontalHeaderLabels(header_labels_group)
    while e != len(list_flag) + 1:
        if e == len(list_flag):
            self.tableWidget_3.setItem(e, 2, QTableWidgetItem(repr(count_ones)+"/"+repr(len(list_name))+"("+repr(round(((count_ones
                                                                                                                / len(list_flag)) * 100), 2)) + "%)"))
        else:
            self.tableWidget_3.insertRow(e)
            self.tableWidget_3.setItem(e, 0, QTableWidgetItem(list_name[e]))
            self.tableWidget_3.setItem(e, 2, QTableWidgetItem(list_flag[e]))
        e += 1
    self.tableWidget_3.setItem(0, 1, QTableWidgetItem(repr(1) + " to " + repr(s_l)))
    self.tableWidget_3.setItem(1, 1, QTableWidgetItem(repr(s_l + 1) + " to " + repr(s_e + s_l)))
    self.tableWidget_3.setItem(2, 1, QTableWidgetItem(repr(s_e + s_l + 1) + " to " + repr(s_e + s_l + s_a)))
    self.tableWidget_3.setItem(3, 1, QTableWidgetItem(repr(s_e + s_l + s_a + 1) + " to " + repr(s_e+s_l+s_a+s_p)))
    self.tableWidget_3.setItem(4, 1, QTableWidgetItem(repr(s_e+s_l+s_a+s_p + 1) + " to " + repr(s_e+s_l+s_a+s_p+s_rec)))
    self.tableWidget_3.setItem(5, 1, QTableWidgetItem(repr(s_e+s_l+s_a+s_p+s_rec + 1) + " to " + repr(s_e+s_l+s_a+s_p+s_rec+s_res)))
    self.tableWidget_3.setItem(6, 1, QTableWidgetItem(repr(s_e+s_l+s_a+s_p+s_rec+s_res + 1) + " to " + repr(s_e+s_l+s_a+s_p+s_rec+s_res+s_obj)))
```

Εικόνα 4.4.5.2.5. Γέμισμα Πίνακα Σημείου 4. Συνέχεια

Στη συνέχεια ορίζει τα ονόματα των στηλών ('GROUP NAME', 'INCLUDING TERMS', 'SCORE') τα οποία είναι το όνομα της ομάδας, οι αριθμοί θέσεων των όρων που περιέχει η κάθε ομάδα και το σκορ της ομάδας (0 ή 1). Κατά την επανάληψη όπου τοποθετούνται τα δεδομένα στον πίνακα του σημείου 4 ελέγχει αν έχει φτάσει στο τέλος της λίστας, αν ναι τότε

υπολογίζει το σκορ και το ποσοστό (πχ 4/7(57.14%)) της διαδικτυακής πλατφόρμας με ακρίβεια δύο δεκαδικών ψηφίων.

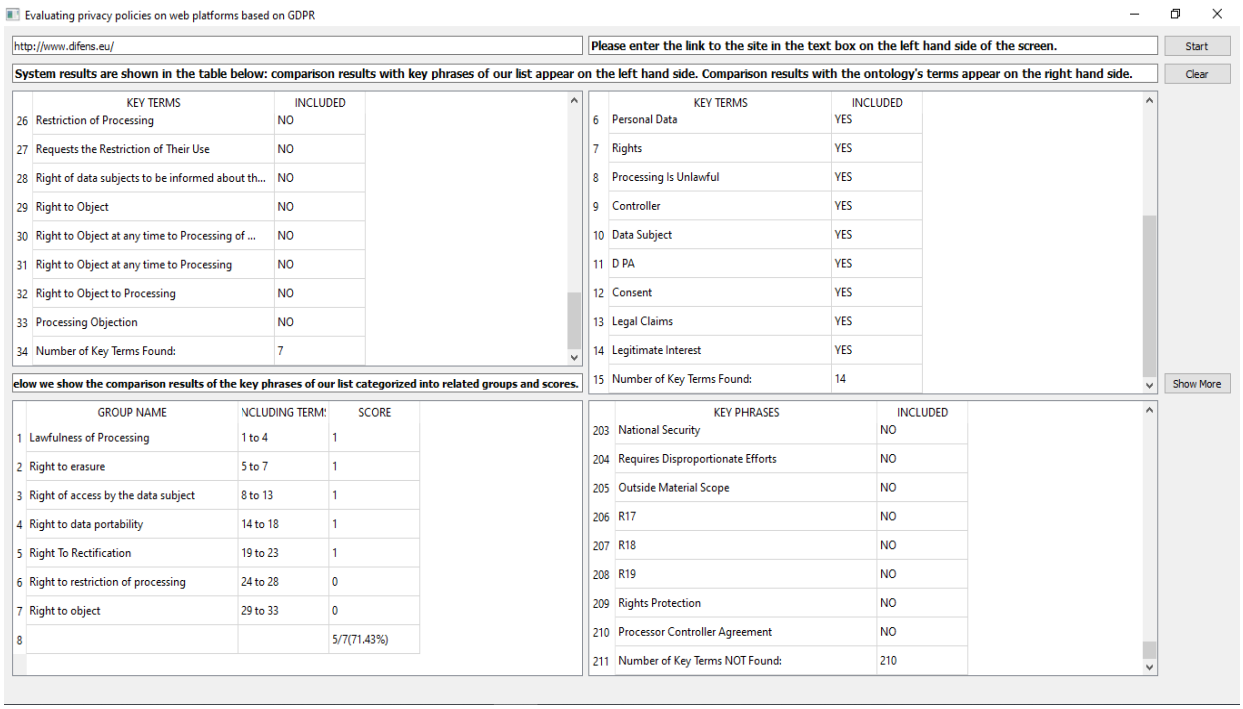
Αφού τελειώσει η εκτέλεση της μεθόδου Run με την ενεργοποίηση του κουμπιού Start, ο χρήστης/οργανισμός αν θελήσει να δει περεταίρω πληροφορίες για το ποιοι όροι της οντολογίας δεν βρέθηκαν στην πολιτική μπορεί να ενεργοποιήσει το σημείο 8. που είναι το κουμπί Show More και αυτό με τη σειρά του καλεί τη μέθοδο Show με την εντολή: self.ShowMore.clicked.connect(self.Show). Για το γέμισμα του πίνακα στο σημείο 5. εκτελείται ο κώδικας που δίνεται στην Εικόνα 4.4.5.2.6 πιο κάτω:

```
def Show(self):
    list_onto = []
    list_onto_no = []
    d = 0
    j = 0
    f3 = open(main.gdpr_onto_words_no, "r")
    f4 = open(main.gdpr_onto_no, "r")
    on_data = f3.readline().rstrip()
    on_data_no = f4.readline().rstrip()
    while on_data != '' and on_data_no != '':
        list_onto.insert(d, on_data)
        list_onto_no.insert(d, on_data_no)
        on_data = f3.readline().rstrip()
        on_data_no = f4.readline().rstrip()
        d += 1
    header_labels = ['KEY PHRASES', 'INCLUDED']
    self.tableWidget_4.setHorizontalHeaderLabels(header_labels)
    while j != len(list_onto):
        self.tableWidget_4.insertRow(j)
        self.tableWidget_4.setItem(j, 0, QTableWidgetItem(list_onto[j]))
        self.tableWidget_4.setItem(j, 1, QTableWidgetItem(list_onto_no[j]))
        j += 1
```

Εικόνα 4.4.5.2.6. Γέμισμα Πίνακα Σημείου 5.

Διαβάζει πρώτα τα δεδομένα που πήρε από τη διαδικασία Ontology Parser. Είναι το αρχείο gdpr_onto_words_no και το αρχείο gdpr_onto_no, επαναλαμβάνει την ακόλουθη διαδικασία μέχρι να διαπεράσει όλα τα δεδομένα και στα δύο αρχεία. Εισάγει τα στοιχεία των δύο αρχείων κειμένου σε δύο λίστες (list_onto, list_onto_no). Αφού ολοκληρωθεί η πιο πάνω διαδικασία δημιουργεί τις δύο στήλες του πίνακα ('KEY TERMS', 'INCLUDED'). Η πρώτη στήλη θα περιέχει όλους του όρους της οντολογίας που δεν συμπεριλαμβάνονται στη πολιτική και η δεύτερη στήλη θα περιέχει τη λέξη NO. Παίρνει κάθε στοιχείο από τις λίστες που δημιουργήθηκαν πιο πάνω και τις προσθέτει στο πίνακα.

4.4.5.3 Παρουσίαση Αποτελεσμάτων στο Χρήστη



Εικόνα 4.4.5.3.1. Παρουσίαση Αποτελεσμάτων

Στην Εικόνα 4.4.5.3.1 φαίνεται η παρουσίαση αποτελεσμάτων στο χρήστη. Στο πίνακα κάτω αριστερά στη τελευταία γραμμή εμφανίζεται η «βαθμολογία» της διαδικτυακής πλατφόρμας DiFens και είναι 5/7(71,43%) ως παράδειγμα, αυτό σημαίνει ότι από τις επτά ομάδες στις πέντε από αυτές βρέθηκε τουλάχιστον ένας όρος στη πολιτική από κάθε ομάδα. Επίσης ο χρήστης βλέπει ποιοι όροι βρίσκονται μέσα σε κάθε ομάδα στη δεύτερη στήλη του πίνακα. Στο πίνακα πάνω αριστερά βρίσκονται οι όροι της λίστας για πολιτικές βάσει του GDPR και στη τελευταία γραμμή του πίνακα εμφανίζεται ο αριθμός των όρων που εντοπίστηκαν στην πολιτική. Στο πίνακα πάνω δεξιά βρίσκονται οι όροι της οντολογίας οι οποίοι εντοπίστηκαν στην πολιτική και στο κάτω μέρος του πίνακα εμφανίζεται ο συνολικός αριθμός τους. Στο πίνακα κάτω δεξιά βρίσκονται οι όροι της οντολογίας οι οποίοι δεν εντοπίστηκαν στην πολιτική και στο κάτω μέρος του πίνακα εμφανίζεται ο συνολικός αριθμός τους.

Κεφάλαιο 5

Αξιολόγηση και Σύγκριση Αποτελεσμάτων

5.1.Μέθοδος Αξιολόγησης

Η μέθοδος αξιολόγησης που χρησιμοποιήθηκε στο λογισμικό αυτοματοποιημένης ανάλυσης πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR είναι με τη χρήση ενός κατωφλίου (threshold) για τη σύγκριση των διαδικτυακών πλατφορμών ως προς την βαθμολογία τους. Το κατώφλι είναι μία τιμή όριο, που μπορεί να θεωρηθεί και ως ένα μέτρο σύγκρισης. Το κατώφλι που χρησιμοποιήσαμε ως μέτρο σύγκρισης, δημιουργήθηκε παίρνοντας τέσσερις γνωστές διαδικτυακές πλατφόρμες βασιζόμενοι στον αριθμό χρηστών που τις χρησιμοποιούν. Αυτές οι διαδικτυακές πλατφόρμες εντοπίστηκαν στο [14], είναι ένα άρθρο που παρουσιάζει τις καλύτερες 20 πλατφόρμες διαδικτυακής αγοράς που εδρεύουν στην Ευρωπαϊκή Ένωση. Επιλέξαμε αυτές τις πλατφόρμες λόγω του ότι δημιουργήθηκαν σε χώρες της ΕΕ οι οποίες ήταν περισσότερο εξοικειωμένες με το γενικό κανονισμό για τη προστασία δεδομένων.

Οι διαδικτυακές πλατφόρμες που χρησιμοποιήθηκαν είναι:

- Zalando, https://www.zalando.co.uk/
- Fruugo, https://www.fruugo.co.uk/
- Coolshop, https://www.coolshop.co.uk/
- Iamfy, https://www.iamfy.co/

Εκτελέσαμε το λογισμικό μας χρησιμοποιώντας ως είσοδο αυτές τις πλατφόρμες διαδικτυακής αγοράς και τα αποτελέσματα που πάρθηκαν από τη λίστα όρων για πολιτικές

βάσει του GDPR και από την οντολογία υπολογίστηκε ο μέσος όρος μεταξύ τους, το οποίο είναι το κατώφλι. Η βαθμολογία των πιο πάνω πλατφορμών φαίνονται στο Πίνακα 5.1.1:

Όνομα Πλατφόρμας	Σκορ Λίστας	Ποσοστό (%) Λίστας	Όροι Οντολογίας
Zalando	6/7	85,71	25
Fruugo	4/7	57,14	25
Coolshop	7/7	100	14
Iamfy	3/7	42,86	14

Πίνακας 5.1.1 Αποτελέσματα γνωστών διαδικτυακών πλατφορμών

Το κατώφλι που θα χρησιμοποιηθεί για τη σύγκριση φαίνεται το Πίνακα 5.1.2:

	Ποσοστό Λίστας	Μέσος όρος Οντολογίας
Κατώφλι (Threshold)	71,43%	19,5

Πίνακας 5.1.2 Κατώφλι

5.2 Αξιολόγηση Διαδικτυακών Πλατφορμών

Στη Διπλωματική Εργασία επιλέξαμε να αξιολογήσουμε διαδικτυακές πλατφόρμες που δημιουργήθηκαν στα πλαίσια ερευνητικών έργων του προγράμματος Erasmus Plus. Τέτοιου είδους πλατφόρμες κρίθηκαν ως ένα κατάλληλο σύνολο για την αξιολόγηση του εργαλείου επειδή αποτελούν ένα ομοιόμορφο σύνολο και θέλαμε να επιλέξουμε πλατφόρμες που είναι στα πλαίσια της Ευρωπαϊκής Ένωσης και που έχουν δημιουργηθεί για σκοπούς έρευνας και για ευρωπαϊκά έργα. Για την αξιολόγηση του εργαλείου επιλέξαμε οκτώ από αυτές τις πλατφόρμες οι οποίες είναι:

- ENROPE, https://enrope.eu/
- ENRICH, http://enrichproject.eu/
- EUROPEAN HISTORY MOVING, http://www.europeanhistorymoving.eu/
- ALL INCLUSIVE SCHOOL, https://allinclusiveschool.eu/
- Galway Community Circus, https://www.galwaycommunitycircus.com/
- InterMediaKT, http://intermediakt.org/home/
- SMART SKILLS DEVELOPMENT, http://smarties.e-code.sk/
- Di Cult Youth, https://dicultyouth.eu/en/

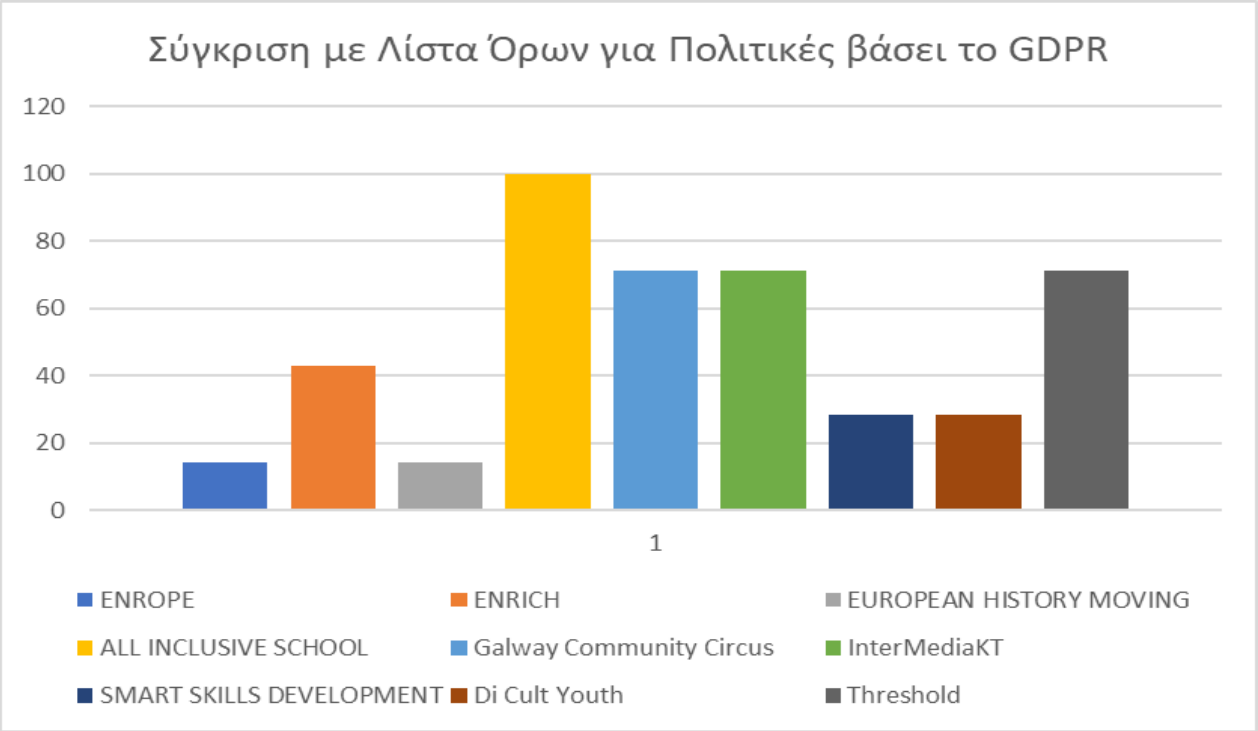
Έγινε εκτέλεση του εργαλείου χρησιμοποιώντας ως είσοδο τις πιο πάνω διαδικτυακές πλατφόρμες και πάρθηκαν αποτελέσματα από τη λίστα όρων για πολιτικές βάσει του GDPR και από την λίστα όρων της οντολογίας τα αποτελέσματα τους. Η βαθμολογία των πιο πάνω πλατφορμών φαίνονται στο Πίνακα 5.2.1:

Όνομα Πλατφόρμας	Σκορ Λίστας	Ποσοστό (%) Λίστας	Όροι Οντολογίας
ENROPE	1/7	14.29	13
ENRICH	3/7	42.86	5
EUROPEAN HISTORY MOVING	1/7	14.29	6
ALL INCLUSIVE SCHOOL	7/7	100	15
Galway Community Circus	5/7	71.43	12
InterMediaKT	5/7	71.43	9
SMART SKILLS DEVELOPMENT	2/7	28.57	4
Di Cult Youth	2/7	28.57	4

Πίνακας 5.2.1 Αποτελέσματα διαδικτυακών πλατφορμών Erasmus Plus

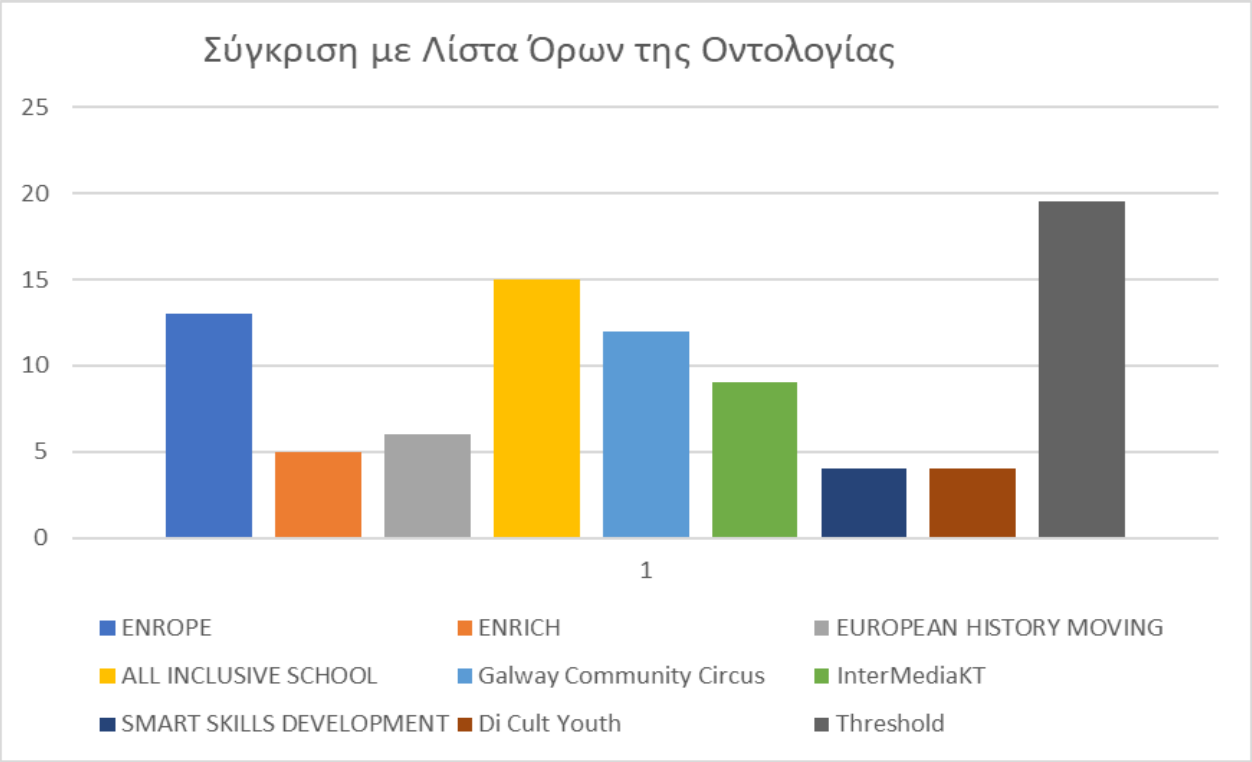
5.3 Σύγκριση Αποτελεσμάτων

Σε αυτό το υποκεφάλαιο θα συγκρίνουμε τα αποτελέσματα που πάρθηκαν από τα υποκεφάλαια 5.1 και 5.2. Στη Γραφική 5.3.1 και 5.3.2 φαίνονται τα αποτελέσματα για τους δύο τρόπους αξιολόγησης:



Γραφική 5.3.1 Σύγκριση με Λίστα Όρων για Πολιτικές βάσει το GDPR

Στη πιο πάνω γραφική φαίνεται η σύγκριση των διαδικτυακών πλατφορμών με το κατώφλι μετά από την εκτέλεση του προγράμματος με τη χρήση της λίστας όρων για πολιτικές βάσει του GDPR. Από την πιο πάνω γραφική παράσταση βλέπουμε ότι οι πέντε από τις διαδικτυακές πλατφόρμες που αναλύθηκαν έχουν πιο χαμηλό ποσοστό από το κατώφλι, οι δύο έχουν ίσο ποσοστό και οι τρεις έχουν πιο ψηλό ποσοστό από το κατώφλι. Αυτό οφείλεται σε δύο λόγους, ο πρώτος είναι ότι η πλειοψηφία των διαδικτυακών πλατφορμών δεν διατυπώνουν σωστά τις ορολογίες του GDPR στις πολιτικές απορρήτου και αυτό έχει ως αποτέλεσμα να μην εντοπίζονται από το πρόγραμμα μας κατά την εκτέλεση. Ο δεύτερος λόγος είναι ότι απλά οι διαδικτυακές πλατφόρμες δεν εφάρμοσαν τους όρους που προϋποθέτει ο γενικός κανονισμός προστασίας δεδομένων (GDPR) στη πολιτική απορρήτου.



Γραφική 5.3.2 Σύγκριση με Λίστα της Οντολογίας

Στη πιο πάνω γραφική φαίνεται η σύγκριση των διαδικτυακών πλατφορμών με το κατώφλι μετά από την εκτέλεση του προγράμματος με τη χρήση της λίστας όρων της οντολογίας. Από την πιο πάνω γραφική παράσταση βλέπουμε ότι όλες οι διαδικτυακές πλατφόρμες που αναλύθηκαν έχουν πιο χαμηλό αριθμό όρων από το κατώφλι. Αυτό οφείλεται στο γεγονός ότι μια οντολογία αποτελείται από πάρα πολλούς όρους και δεν απευθύνονται όλοι οι όροι σε πολιτικές, είναι γενικότερη. Ακόμα οι διαδικτυακές πλατφόρμες που χρησιμοποιήθηκαν για τον υπολογισμό του κατωφλίου έχουν καλύτερη πολιτική σε σχέση με τις πλατφόρμες που δημιουργήθηκαν στα πλαίσια ερευνητικών έργων του προγράμματος Erasmus Plus. Επίσης μία οντολογία δεν περιέχει παραλλαγές και αναγραμματισμούς ενός όρου, ένας όρος αναφέρεται στην οντολογία όπως ορίζεται στο γενικό κανονισμό προστασίας δεδομένων.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντική Εργασία

6.1 Εισαγωγή

Κατά την ανάπτυξη αυτής της διπλωματικής εργασίας και μέσω της ανάλυσης και περαιτέρω μελέτης του GDPR και ενός αριθμού διαδικτυακών πλατφορμών, καθώς και του σχεδιασμού, υλοποίησης και της εφαρμογής του εργαλείου αυτοματοποιημένης αξιολόγησης πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR , καταλήξαμε σε ορισμένα συμπεράσματα. Σε αυτό το κεφάλαιο αυτά τα συμπεράσματα θα παρουσιαστούν καθώς και μελλοντικές εργασίες που μπορούν να γίνουν για τη βελτίωση του εργαλείου.

6.2 Συμπεράσματα

Πραγματοποιήθηκε πρώτα μια βιβλιογραφική έρευνα σχετικά με το GDPR και πως ορίζει τους κανονισμούς και τα δικαιώματα του χρήστη των δεδομένων. Αυτή η βιβλιογραφική έρευνα μας βοήθησε να εξοικειωθούμε με τις ορολογίες που χρησιμοποιούνται στο Γενικό Κανονισμό Προστασίας Δεδομένων και πως αυτές οι ορολογίες διατυπώνονται στις διαδικτυακές πλατφόρμες. Επίσης από την έρευνα αυτή πήραμε τους κανονισμούς και τα δικαιώματα για τα προσωπικά δεδομένα του χρήστη και δημιουργήσαμε μια λίστα όρων για πολιτικές βάσει του γενικού κανονισμού για τη προστασία δεδομένων (GDPR).

Στη συνέχεια μελετήσαμε ένα αριθμό από διαδικτυακές πλατφόρμες. Με αυτή τη μελέτη αναλύσαμε τη δομή της πολιτικής απορρήτου και πως οι διαδικτυακές πλατφόρμες διατυπώνουν τους όρους της λίστας που δημιουργήσαμε με τη μελέτη πάνω στο γενικό κανονισμό για τη προστασία δεδομένων (GDPR). Από αυτή τη μελέτη βρέθηκαν παραλλαγές και αναγραμματισμοί στη διατύπωση των όρων στις πολιτικές απορρήτου των διαδικτυακών

πλατφορμών. Αυτό οδήγησε στην επέκταση της υπάρχον λίστας όρων για πολιτικές βάσει του GDPR.

Ακολούθως η έρευνα συνεχίστηκε πάνω στις πλατφόρμες με διερεύνηση στο πηγαίο κώδικα για να παρατηρηθεί το πως γίνεται η περιήγηση στη σελίδα από ένα crawler. Στη περιήγηση αυτή συλλέχθηκαν πληροφορίες για την δημιουργία μιας λίστας όρων για περιήγηση. Αυτή η λίστα όρων για περιήγηση μας βοήθησε με την έρευνα και μεθοδολογία που έγινε σε μεταγενέστερο στάδιο για το crawler.

Ακολούθως έγινε η υλοποίηση του crawler χρησιμοποιώντας τη γλώσσα προγραμματισμού Python. Επόμενο στάδιο της υλοποίησης ήταν η αυτόματη ανάλυση του κειμένου που συλλέχθηκε βάσει του λεξιλογίου ως επίσης αυτόματη ανάλυση του κειμένου βάση της οντολογίας και παρουσίαση και σύγκριση των αποτελεσμάτων.

Για σκοπούς αξιολόγησης χρησιμοποιήσαμε ένα κατώφλι ως μέτρο σύγκρισης, δημιουργήθηκε παίρνοντας τέσσερις γνωστές διαδικτυακές πλατφόρμες βασιζόμενοι στον αριθμό χρηστών που τις χρησιμοποιούν. Επιλέξαμε να αξιολογήσουμε διαδικτυακές πλατφόρμες που δημιουργήθηκαν στα πλαίσια ερευνητικών έργων του προγράμματος Erasmus Plus. Το 62,5% αυτών των διαδικτυακών πλατφορμών ήταν κάτω από το κατώφλι στα αποτελέσματα που πάρθηκαν με τη χρήση της λίστας όρων για πολιτικές βάσει του GDPR και το 100% αυτών των διαδικτυακών πλατφορμών ήταν κάτω από το κατώφλι στα αποτελέσματα που πάρθηκαν με τη χρήση της οντολογίας.

Καταλήξαμε στο συμπέρασμα ότι μεγάλος αριθμός από τις διαδικτυακές πλατφόρμες που αξιολογήσαμε δεν έχουν κατάλληλες πολιτικές απορρήτου και αυτό σημαίνει ότι η πλειοψηφία των διαδικτυακών πλατφορμών δεν διατυπώνουν σωστά τις ορολογίες του GDPR στις πολιτικές απορρήτου και αυτό έχει ως αποτέλεσμα να μην εντοπίζονται από το πρόγραμμα μας κατά την εκτέλεση. Επίσης, οι διαδικτυακές πλατφόρμες δεν εφάρμοσαν τους όρους που προϋποθέτει ο γενικός κανονισμός προστασίας δεδομένων (GDPR) στη πολιτική απορρήτου. Η πλειοψηφία των διαδικτυακών πλατφορμών είτε δεν έδωσαν αρκετή προσοχή για να εναρμονιστούν με τον γενικό κανονισμό προστασίας δεδομένων, είτε δεν είναι ενημερωμένες για τη σωστή διατύπωση της πολιτικής απορρήτου σύμφωνα με το GDPR.

6.3 Μελλοντική Εργασία

Για βελτίωση της υλοποίησης θα μπορούσαμε να δώσουμε επιλογή στο χρήστη ή οργανισμό να εισάγει ως είσοδο στο πρόγραμμα περισσότερους από ένα συνδέσμους με την εισαγωγή αρχείου που να τους εμπεριέχει.

Θα μπορούσε η αυτοματοποιημένη ανάλυση πολιτικών απορρήτου σε διαδικτυακές πλατφόρμες βάσει του GDPR να γίνεται με τη χρήση μηχανικής μάθησης, δηλαδή με αναγνώριση διαφόρων φράσεων ή pattern στο κείμενο της πολιτικής απορρήτου.

Για μελλοντική εργασία θα μπορούσε να γίνει επέκταση της λίστας όρων για πολιτικές βάσει του GDPR έτσι ώστε να μπορεί το πρόγραμμα να αναγνωρίζει μεγαλύτερο αριθμό όρων στην πολιτική.

Βιβλιογραφία

[1] United Nations, "Universal Declaration of Human Rights", (1948), https://www.un.org/en/universal-declaration-human-rights/index.html

[2] European Convention on Human Rights, "Convention for the Protection of Human Rights and Fundamental Freedoms", (1950), https://www.echr.coe.int/Pages/home.aspx?p=basictexts&c

[3] Ann Cavoukian, Ph.D., "Privacy by Design in Law, Policy and Practice", Information and Privacy Commissioner, Ontario, Canada, 2011.

[4] Karen Renaud, Lynsay A. Shepherd, "How to Make Privacy Policies both GDPR-Compliant and Usable", Abertay University, Dundee, United Kingdom

[5] Information Commissioner's Office, "Preparing for the General Data Protection Regulation (GDPR) - 12 Steps to Take Now," 2018, https://ico.org.uk/media/1624219/preparing-for-the-gdpr-12-steps.pdf

[6] A. Cormack, "GDPR: What's your justification?" 2017, https://community.jisc.ac.uk/blogs/regulatory-developments/article/gdpr-whats-your-justification

[7] Intersoft Consulting, "Art. 6 GDPR Lawfulness of processing," 2016, https://gdprinfo.eu/art-6-gdpr/

[8] Data Protection Network, "GDPR Data Retention Quick Guide," 2017, https://www.dpnetwork.org.uk/gdpr-data-retention-guide/.

[9] Welderufael B. Tesfay, Peter Hofmann, Toru Nakamura, Jetzabel Serna, Shinsaku Kiyomoto, "PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation", Goethe University Frankfurt, KDDI R&D Laboratories Inc., 2018

[10] Sebastian Zimmeck, Steven M. Bellovin, “Privee: An Architecture for Automatically Analyzing Web Privacy Policies”, Department of Computer Science, Columbia University, 2014

[11] André Ribeiro, Alberto Rodrigues da Silva, “RSLingo4Privacy Studio: A Tool to Improve the Specification and Analysis of Privacy Policies”, University of Lisbon, Lisbon, Portugal, 2017

[12] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, Kassem Fawaz, “The Privacy Policy Landscape After the GDPR”, 2019

[13] “General Data Protection Regulation GDPR”, https://gdpr-info.eu/

[14] “Top 20 of cross-border marketplaces from Europe”, https://ecommercenews.eu/top-20-of-cross-border-marketplaces-from-europe/, 2019

[15] “GDPRtEXT”,https://openscience.adaptcentre.ie/ontologies/GDPRtEXT/deliverables/docs/ontology, 2020

Παράρτημα Α

Λίστα Όρων Οντολογίας

Legal Resource Subdivision
Article
Chapter
Citation
Point
Recital
Section
Sub Point
Legal Resource
Concept
Principle
Accountability
Accuracy
Data Minimisation
Lawfulness Fairness And Transparency
Integrity And Confidentiality
Purpose Limitation
Storage Limitation
Compliance
Monitor Compliance
Demonstrating Consent
Report Data Breach
Maintain Record Of Breach
Notify Data Subject Of Breach
Notify Data Subject About DPO For Data Breach
Notify Data Subject About Consequences Of Data Breach
Notify Data Subject Of Measures Taken For Data Breach
Report Data Breach To DP AWithin72 Hours
Report Data Breach To Controller
Activity
Processing
Automated Processing
Automated Decision Making With Significant Effect
Confirming Or Matching Datasets
Large Scale Processing
Processing Affected Vulnerable Individuals
Unlawful Processing
Processing Sensitive Data
Processing Using Untested Technologies
Appointment Of Processors
Data Activity
Erase Data
Rectify Data
Collection Of Personal Data
Collection Mechanism

Store Data
Archive Data
Cross Border Transfer
Share Data With Third Party
Use Data
Provide Copy Of Personal Data
Should Be Commonly Used Format
Should Be Machine Readable
Should Be Structured
Should Support Reuse
Consent Activity
Withdrawing Consent
Obtaining Consent
Security Of Personal Data
Protection Against Accidental Loss
Protection Against Damage
Protection Against Destruction
Protection Against Unlawful Processing
Marketing
Direct Marketing
Identification Of Data Subject
Exercise Rights
Impact Assessment
Propogate Rights To Third Parties
Systematic Monitoring
Data
Personal Data
Sensitive Personal Data
Criminal Data
Genetic Data
Health Data
Racial Data
Anonymous Data
Pseudo Anonymous Data
Rights
Right To Restrict Processing
Accuracy Is Contested
Data No Longer Needed For Original Purpose
Processing Is Unlawful
Right Of Erasure
Erase When Consent Was Withdrawn
Erase When Data Is No Longer Needed For Original Purpose
Right To Access Personal Data
If And Where Controller Is Processing Data
Info About Automated Processing With Significant Effects On Data Subject
Info About Categories Of Recipients Data Shared With
Info About Categories Of Data Being Processed
Info About Existence Of Rights
Info About Processing
Info About Source Of Data
Info About Storage Period

No Charge Levied For Right To Access
Right To Basic Information About Processing
Information About Third Parties
Right To Transparency
Information Should Be Concise
Information Should Be Easily Accessible
Information Should Be Intelligible
Information Should Be Transparent
Right Of Data Portability
Right To Not Be Evaluated Through Automated Processing
Right To Object For Direct Marketting
Right To Object To Processing
Right To Rectification
Factors For Impact Assessment
Evaluation Of Data Subject
Seals And Certification
Conditions For Seals And Certifications
Adherence To Seal Certification
Maximum Validity3 Years
Voluntary System Of Accredition
Certification
Seal
Entity
Processor
Sub Processor
Controller
Joint Controller
Data Subject
Certification Body
Controller Representative
D PA
D PO
Processor Representative
Regulatory Authority
Code Of Conduct
Consent
Given Consent
Valid Consent
Freely Given Consent Obligation
Informed Consent Obligation
Specific Consent Obligation
Voluntary Opt In Consent Obligation
Data Breach
Obligation
Lawful Basis For Processing
Purpose Of New Processing
Context Of Data Collection
Existence Of Safeguards
Link Between New And Old Purpose
Nature Of Personal Data
Possible Consequence For Data Subject

Contract With Data Subject
Employment Law
Exempted By National Law
Historic Statistic Scientific Purposes
Legal Claims
Legal Obligation
Legitimate Interest
Made Public By Data Subject
Medical Diagnosis Treatment
Not For Profit Org
Public Interest
Vital Interest
Obligation For Collection Of Personal Data
Accurate Collection
Explicit Purpose
Specified Purpose
Legitimate Purpose
Not Further Processed Than Original Purpose
Retention Of Personal Data
Kept Up To Date
Rectify Inaccuracy
Adequate For Processing
Identifiable For Required Processing
Limited For Processing
Relevant For Processing
Processor Obligation
Appointing Sub Processors
Requires Written Consent Of Controller To Appoint Sub Processor
Sub Processor Must Follow Same Terms As Processor Controller Agreement
Assist In Complying With Rights
Compliance With Controllers Instructions
Cooperate With DP A
Data Security
Impose Confidentiality Obligation On Personnel
Inform Controller Of Conflict With Law
Maintain Records Of Processing Activities
Record Categories Of Recipients Where Data Shared
Record Categories Of Data Subjects And Personal Data
Record Cross Border Data Transfer
Record Data Retention Period
Record Purpose Of Processing
Record Security Measures
Only Act On Documented Instruction From Controller
Provide Controller With Info For Compliance
Restrictions On Cross Border Transfers
Return Or Destroy Personal Data At End Term
Controller Obligation
Controller Responsibility
Implement Technical Measures For Compliance
Controller Accountability
Data Protection By Design And By Default

Liability Of Joint Controller
Privacy By Design
Obligation For Obtaining Consent
Can Be Withdrawn Easily Consent Obligation
Clear Explanation Of Processing Consent Obligation
Not From Silence Or Inactivity Consent Obligation
Should Be Demonstrable
Should Be Distinguishable From Other Matters
D PO Obligation
Exclusion Exception
Exceptions On Reporting Data Subjects Of Breach
Controller Has Taken Action
Harm Was Remote
Notification Requires Disproportionate Efforts
Data Was Inferred Or Derived
Exempted Without Proof Of Data Subject Identity
Freedoms Protection
Is Impossible
National Security
Requires Disproportionate Efforts
Outside Material Scope
R17
R18
R19
Rights Protection
Processor Controller Agreement