

Ατομική Διπλωματική Εργασία

**ΟΝΤΟΛΟΓΙΑ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΜΙΣΑΛΛΟΔΟΞΟΥ
ΠΕΡΙΕΧΟΜΕΝΟΥ**

Έλλη Παναγή

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2018

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Οντολογία Αναπαράστασης Μισαλλόδοξου Περιεχομένου

Έλλη Παναγή

Επιβλέπων Καθηγητής

Δικαιάκος Μάριος

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

Μάιος 2018

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέπων καθηγητή μου Μάριο Δικαιάκο για την πολύτιμη βοήθεια και καθοδήγηση που μου παρείχε σε όλο αυτό το διάστημα της εκπόνησης της διπλωματικής μου εργασίας. Ακόμη ευχαριστώ θερμά τον Δημοσθένη που μου έχει δώσει μερικά κατηγοριοποιημένα tweets τα οποία μου ήταν απαραίτητα για να επικυρώσω την ορθότητα της οντολογίας μου. Επίσης, θα ήθελα να ευχαριστήσω τον συμφοιτητή μου Χρίστο ο οποίος με βοήθησε να ερμηνεύσω τα tweets εφόσον γνώριζε την κατομιλουμένη γλώσσα που ήταν διατυπωμένα στα tweets.

Περίληψη

Η Ρητορική Μίσους (hate speech) είναι ένα κοινωνικό πρόβλημα που εξαπλώθηκε πλέον και στο διαδίκτυο, και γίνονται πολλές προσπάθειες για την εντόπιση και καταπολέμηση του. Δυστυχώς όμως αυτό δεν είναι εύκολο ζήτημα, καθώς δεν υπάρχει ένας κοινώς αποδεκτός ορισμός της ρητορικής μίσους και η πληροφορία στο διαδίκτυο δεν είναι πάντοτε καλά δομημένη. Στα πλαίσια της διπλωματικής μου εργασίας παρουσιάζω μία οντολογία που αναπαριστά την ρητορική μίσους βάση νομοθετικού πλαισίου, έχοντας έτσι ένα πιο αντικειμενικό ορισμό. Ακόμη έχω υλοποιήσει μία διεπαφή που λειτουργεί ως μεσολαβητής μεταξύ χρηστών και της οντολογίας. Μέσω της διεπαφής γίνεται το annotation tweets τα οποία προσθέτουν την απαραίτητη γνώση στην οντολογία, για να βοηθήσουν τον reasoner να διεξάγει συμπεράσματα για το αν το επιλεγμένο tweet μπορεί να θεωρηθεί πιθανόν παράνομη ρητορική μίσους. Χρησιμοποιώντας την διεπαφή έκανα annotate μερικά tweets και εξέτασα κατά πόσο τα συμπεράσματα του reasoner σύγκλιναν με αυτά που έκαναν ειδικοί επιστήμονες.

Περιεχόμενα

Κεφάλαιο 1 Εισαγωγή	1
1.1 Πρόβλημα ρητορικής μίσους	1
1.2 Στόχοι και κίνητρα διπλωματικής εργασίας	2
1.3 Δομή διπλωματικής εργασίας	3
 Κεφάλαιο 2 Σημασιολογικός Ιστός	5
2.1 Τι είναι σημασιολογικός ιστός;.....	5
2.2 Αρχιτεκτονική Σημασιολογικού Ιστού	6
2.3 Semantic reasoner	11
 Κεφάλαιο 3 Οντολογίες	12
3.1 Τι είναι μια οντολογία (ορισμός).....	12
3.2 Τα είδη των οντολογιών	13
3.3 Βασικά συστατικά οντολογιών	16
3.4 Χρησιμότητα οντολογιών	19
3.5 Μεθοδολογία κατασκευής οντολογιών.....	19
3.6 Σημαντικά εργαλεία κατασκευής και επεξεργασίας οντολογιών	20
 Κεφάλαιο 4 Ρητορική μίσους.....	22
4.1 Ελευθερία του λόγου και έκφρασης	22
4.2 Πώς σχετίζεται η ελευθερία του λόγου με την ρητορική του μίσους;.....	23
4.3 Τί είναι η ρητορική μίσους;	23
4.4 Κριτήρια αναγνώρισης ρητορικής μίσους	23
4.5 Άτομα που μπαίνουν στο στόχαστρο.....	24
4.6 Αίτια ρητορικής μίσους	24

4.7 Επιπτώσεις ρητορικής μίσους.....	25
4.8 Τρόποι αντιμετώπισης	25
Κεφάλαιο 5 Ανάλυση Μεθοδολογίας	26
5.1 Πηγές πληροφοριών.....	27
5.2 Βήμα 1 - Καθορισμός πεδίου ενδιαφέροντος και σκοπός της οντολογίας	27
5.3 Βήμα 2 – Επαναχρησιμοποίηση υπαρχουσών οντολογιών	28
5.4 Βήμα 3 – Απαρίθμηση σημαντικών όρων οντολογίας	28
5.5 Βήμα 4 – Καθορισμός κλάσεων και ιεραρχία κλάσεων	29
5.6 Βήμα 5 – καθορισμός ιδιοτήτων των κλάσεων	31
5.7 Βήμα 6 – καθορισμός περιορισμών των ιδιοτήτων	31
5.8 Βήμα 7 – Δημιουργία στιγμιότυπων.....	32
Κεφάλαιο 6 Ανάλυση Οντολογίας Hate Speech	34
6.1 Κλάσεις - Υποκλάσεις	34
Κεφάλαιο 7 Ανάπτυξη διεπαφής	54
7.1 Εισαγωγή	54
7.2 Κύρια στοιχεία διεπαφής	55
7.3 Γραφική απεικόνιση των ερωτήσεων και επεξήγηση χρησιμότητας τους	55
Κεφάλαιο 8 Αποτελέσματα	61
8.1 Παραδείγματα εκτέλεσης.....	62
8.2 Σύγκριση αποτελεσμάτων.....	67
8.3 Επέκταση οντολογίας	68
Κεφάλαιο 9 Συμπεράσματα	69

9.1 Γενικά Συμπεράσματα	69
9.2 Βελτίωση Αποτελεσμάτων	69
9.3 Μελλοντική εργασία	70
Πηγές.....	71

Σχήμα 2-1 Αρχική αρχιτεκτονική σημασιολογικού ιστού (Πηγή Εικόνας: https://www.w3.org/Talks/2001/1102-semweb-fin/all.htm)	6
Σχήμα 2-2 Εναλλακτική αρχιτεκτονική σημασιολογικού ιστού (Πηγή Εικόνας: https://www.w3.org/2005/Talks/0517-boit-tbl/).....	10
Σχήμα 3-1 Κατηγορίες γλωσσολογικών οντολογιών (Πηγή: [13])	14
Σχήμα 3-2 Κατηγοριοποίηση οντολογιών με βάσει τον σκοπό (Πηγή: [8])	15
Σχήμα 3-3 Κλάσεις που περιέχουν στιγμιότυπα (Πηγή: [18])	16
Σχήμα 3-4 Αναπαράσταση υποκλάσης (Πηγή: [18])	17
Σχήμα 3-5 Κατηγορίες Ιδιοτήτων (Πηγή: [18]).....	18
Σχήμα 3-6 Στιγμιότυπα (Πηγή: [18]).....	18
Σχήμα 6-1 Οι πιο βασικές κλάσεις της οντολογίας μου και το πώς συνδέονται μεταξύ τους. Απαλείφονται οι υποκλάσεις τους.	35
Σχήμα 8-1 Screenshot από το Protégé, όπου εμφανίζω όλα τα Conduct τα οποία είναι πιθανώς παράνομα όπως τα έχει κατατάξει ο HermiT reasoner βάση των annotation που έχω κάνει.....	66
Σχήμα 8-2 Στο σχήμα αυτό φαίνονται δύο ομάδες: αριστερά βρίσκονται όλα τα tweets τα οποία κατηγοριοποιήθηκαν ως offensive από τους κοινωνικούς επιστήμονες, ενώ στα δεξιά βρίσκονται τα tweets τα οποία εκφράζουν κολάσιμο hate speech. Αυτά που είναι χρωματισμένα με πορτοκαλί χρώμα είναι αυτά που ο HermiT reasoner τα έχει κατηγοριοποιήσει διαφορετικά από τους κοινωνικούς επιστήμονες.....	67
Σχήμα 8-3 Η πιο πάνω εικόνα δείχνει μετρικές της οντολογίας από το Protégé. Αριστερά δείχνω τις μετρικές πριν να γίνει το annotation, και δεξιά δείχνω τις μετρικές μετά το annotation.....	68

Κεφάλαιο 1

Εισαγωγή

1.1 Πρόβλημα ρητορικής μίσους	1
1.2 Στόχοι και κίνητρα διπλωματικής εργασίας	2
1.3 Δομή διπλωματικής εργασίας	3

1.1 Πρόβλημα ρητορικής μίσους

Τα κοινωνικά δίκτυα και το διαδίκτυο γενικότερα ευνόησε να διαδοθούν οι σκέψεις και απόψεις των ανθρώπων σε όλα τα μήκη και πλάτη της Γης. Όλοι οι άνθρωποι έχουν το δικαίωμα να μπορούν να εκφράζονται ελεύθερα. Όμως αυτό το δικαίωμα πρέπει να χρησιμοποιείται με μέτρο γιατί όπως όλοι γνωρίζουμε «η ελευθερία μας σταματά εκεί που αρχίζει η ελευθερία των άλλων».

Δυστυχώς δεν αντιμετωπίζονται όλοι οι άνθρωποι ως ισότιμα μέλη μιας κοινωνίας. Αυτό συμβαίνει κυρίως από το φαινόμενο της ρητορικής μίσους, γνωστή αλλιώς ως hate speech η οποία έχει αυξηθεί αισθητά στο διαδίκτυο. Η ρητορική μίσους (hate speech) σαν έννοια δεν έχει καθολικό ορισμό και ο καθένας την αντιλαμβάνεται διαφορετικά. Είναι όμως κοινά αποδεκτό ότι η ρητορική μίσους σχετίζεται με την έκφραση έντονου μίσους και απέχθειας προς άτομα λόγω έμφυτων χαρακτηριστικών ή πεποιθήσεών τους. Τα θύματα

της ρητορικής μίσους βιώνουν σε καθημερινή βάση τον στιγματισμό και την περιθωριοποίηση από το περιβάλλον τους. Είναι αντιληπτό ότι το φαινόμενο αυτό δεν μπορεί να είναι κοινωνικά αποδεκτό.

Πολλοί ερευνητές προσπάθησαν να εντοπίσουν και να κατηγοριοποιήσουν κείμενα που εκφράζουν ρητορική μίσους χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων. Οι πλείστοι χρησιμοποίησαν μηχανική μάθηση για την αναγνώριση και την κατηγοριοποίηση μισαλλόδοξου περιεχομένου. Όμως για να σχηματιστεί το κατάλληλο μοντέλο απαιτείται προηγουμένως να αναλυθούν πολλά δεδομένα εκπαίδευσης από έναν αλγόριθμο κατηγοριοποίησης. Αν και έχουν γίνει αξιόλογες προσπάθειες εντοπισμού της ρητορικής μίσους στο διαδίκτυο, το πρόβλημα αυτό συνεχίζει να υφίσταται. Αυτό οφείλεται κυρίως από τη μη δομημένη σύνταξη των κειμένων. Έτσι δεν μπορεί κανείς να διακρίνει τη ρητορική μίσους μιας και δεν υπάρχουν κοινά πρότυπα (patterns) που να διευκολύνουν την ανάπτυξη ακριβών και αποτελεσματικών μοντέλων κατηγοριοποίησης.

1.2 Στόχοι και κίνητρα διπλωματικής εργασίας

Με την παρούσα διπλωματική εργασία επιδιώκω να προτείνω ένα κοινό πλαίσιο της ρητορικής μίσους ορίζοντας μια οντολογία. Είμαι πεπεισμένη ότι η οντολογία είναι η κατάλληλη δομή, για αυτό που θέλω να πετύχω. Χρησιμοποιείται κυρίως για την αναπαράσταση γνώσης και δομεί ιεραρχικά τις πληροφορίες. Η ανάπτυξη της οντολογίας είναι το πρώτο μέρος της διπλωματικής μου και το δεύτερο μέρος της είναι η ανάπτυξη της διεπαφής. Η διεπαφή χρησιμοποιείται κυρίως για να καθοδηγήσει και να κατευθύνει τους χρήστες που δεν είναι ειδήμονες του πεδίου της ρητορικής μίσους προκειμένου να βελτιώσουν την ποιότητα των annotations τους. Αυτό επιτυγχάνεται μέσω μιας σειράς ερωτήσεων κατανόησης. Οι ερωτήσεις κατανόησης πιστεύω ότι παίζουν σημαντικό ρόλο κατά τη διάρκεια των annotation καθώς παρακινούν τον χρήστη να εξετάσει πιο σχολαστικά το περιεχόμενο του κειμένου.

Η διπλωματική μου εργασία συμβάλλει στον σημασιολογικό ιστό όσο και στη μηχανική μάθηση. Η ανάπτυξη οντολογιών είναι το βασικό συστατικό του σημασιολογικού ιστού. Ακόμη συμβάλω στην μηχανική μάθηση με το να εμπλουτίσω και να βελτιώσω την

ποιότητα των δεδομένων εκπαίδευσης με μεταδεδομένα. Αυτό όπως είπα προηγουμένως είναι απαραίτητο για την παραγωγή καλών αποτελεσμάτων κατηγοριοποίησης από τα μοντέλα. Επιπρόσθετα, δίνω τη δυνατότητα σε άπειρους χρήστες να κάνουν annotate, αυξάνοντας έτσι το εργατικό δυναμικό και συνεπώς τα δεδομένα εκπαίδευσης.

1.3 Δομή διπλωματικής εργασίας

Στο κεφάλαιο 2 εξηγώ μερικές χρήσιμες πληροφορίες σχετικά με τον σημασιολογικό ιστό όπως τί είναι, τι αποσκοπεί να πετύχει καθώς επίσης αναλύω την αρχιτεκτονική του σημασιολογικού ιστού.

Στο κεφάλαιο 3 παρουσιάζω τον ορισμό της οντολογίας, τα είδη οντολογιών που υπάρχουν και τα βασικά συστατικά της. Στην συνέχεια εξηγώ τους λόγους για τους οποίους συνήθως χρησιμοποιείται. Ακολουθώς παρουσιάζω γενικά τη μεθοδολογία που έχω ακολουθήσει, καθώς επίσης αναφέρω κάποιες γενικές πληροφορίες σχετικά με τα πιο σημαντικά υφιστάμενα εργαλεία κατασκευής και επεξεργασίας οντολογιών.

Στο κεφάλαιο 4 παρουσιάζω γενικές πληροφορίες για τη ρητορική μίσους όπως τη σχέση του θεμελιώδους ανθρώπινου δικαιώματος της ελευθερίας του λόγου μαζί με τη ρητορική μίσους, τα άτομα που μπαίνουν στο στόχαστρο και πως θα μπορούσαμε να αναγνωρίσουμε τέτοια περιστατικά. Ακόμη αναλύω τα αίτια, τις επιπτώσεις που επιφέρει καθώς επίσης αναφέρω μερικούς τρόπους αντιμετώπισης αυτού του φαινομένου.

Στο κεφάλαιο 5 εξηγώ πως ακολούθησα τη μεθοδολογία για την κατασκευή της οντολογίας μου, δηλαδή αναλύω τι έκανα σε κάθε φάση.

Στο κεφάλαιο 6 παρουσιάζω την οντολογία μου, δηλαδή ερμηνεύω τι εκφράζουν οι κλάσεις που έκανα και πως συσχετίζονται μεταξύ τους.

Στο κεφάλαιο 7 εξηγώ για ποιόν σκοπό δημιούργησα τη διεπαφή, τα στοιχεία που περιέχει η διεπαφή μου καθώς και το πώς επηρεάζεται η οντολογία με τις απαντήσεις που έδωσε ο χρήστης.

Στο κεφάλαιο 8 παρουσιάζω τα αποτελέσματα που προέκυψαν από τα annotations 29 tweets χρησιμοποιώντας τη διεπαφή. Δίνω 2 παραδείγματα εκτέλεσης (1 offensive tweet και 1 hate speech tweet). Κατόπιν εξετάζω κατά πόσο τα αποτελέσματα μου συγκλίνουν με αυτά των κοινωνικών επιστημόνων οι οποίοι είχαν κατηγοριοποιήσει προηγουμένως τα tweets.

Στο κεφάλαιο 9 παρουσιάζω τα συμπεράσματα και κάποιες εισηγήσεις για μελλοντική εργασία.

Κεφάλαιο 2

Σημασιολογικός Ιστός

2.1 Τι είναι σημασιολογικός ιστός;	5
2.2 Αρχιτεκτονική Σημασιολογικού Ιστού	6
2.3 Semantic reasoner	11

2.1 Τι είναι σημασιολογικός ιστός;

Ο σημασιολογικός ιστός είναι το διαδίκτυο του μέλλοντος όπου οι ιστοσελίδες θα είναι οργανωμένες σύμφωνα με τη σημασιολογία του περιεχομένου τους. Ο Tim Berners Lee (εφευρέτης του παγκόσμιου ιστού) οραματίστηκε τον σημασιολογικό ιστό και είπε χαρακτηριστικά ότι ο σημασιολογικός ιστός θα είναι μια επέκταση του σημερινού ιστού όπου οι πληροφορίες θα είναι καλά δομημένες, με απώτερο σκοπό να γίνονται κατανοητές και επεξεργάσιμες από τις μηχανές. Αν αυτό το εγχείρημα πραγματοποιηθεί τότε θα βελτιώσει κατά πολύ την καθημερινότητα των ανθρώπων, καθώς θα δώσει τη δυνατότητα στους ανθρώπους να συνεργάζονται με τις μηχανές [1].

Ειδικότερα στις μέρες μας ο σημασιολογικός ιστός έγινε μια επιτακτική ανάγκη γιατί στο σημερινό παγκόσμιο ιστό επικρατεί ένα χάος από πληροφορίες και ο σημασιολογικός ιστός υπόσχεται να λύσει αυτό το πρόβλημα. Το βασικότερο πρόβλημα που αντιμετωπίζει ο σημερινός παγκόσμιος ιστός είναι ότι τα περιεχόμενά του (οι ιστοσελίδες) είναι διατυπωμένα με τέτοιο τρόπο που να τα καταλαβαίνουν και οι άνθρωποι και όχι τα

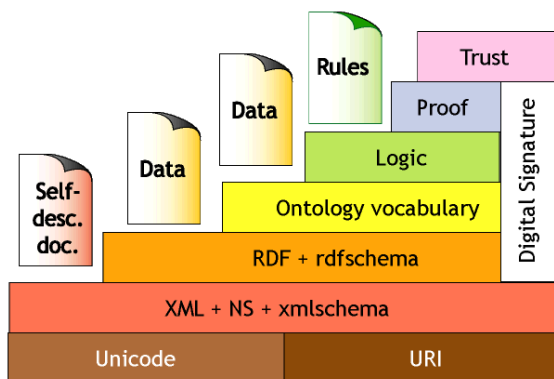
αυτοματοποιημένα εργαλεία (πράκτορες). Όπως γίνεται αντιληπτό οι μηχανές με τα σημερινά δεδομένα δεν είναι σε θέση να κατανοήσουν τη σημασία του περιεχομένου. Για αυτό τον λόγο όταν βάζουμε περίπλοκα ερωτήματα στις μηχανές αναζήτησης, λαμβάνουμε πολλά άσχετα αποτελέσματα, ενώ σχετικά αποτελέσματα που να αφορούν το συγκεκριμένο ερώτημα, παραλείπονται [2].

Ο σημασιολογικός ιστός θα πραγματοποιηθεί με την προσθήκη semantic annotations στις ιστοσελίδες. Τα semantic annotations είναι το αποτέλεσμα της περιγραφής που δόθηκε για να περιγράψει τη σημασία του περιεχομένου των ιστοσελίδων χρησιμοποιώντας μεταδεδομένα (“data about data”). Στη συνέχεια όλα τα annotations αναπαρίστανται σε ένα ενοποιημένο μοντέλο (RDF), το οποίο επιτρέπει την ενσωμάτωση δεδομένων και τη διαλειτουργικότητα. Με αυτό τον τρόπο οι μηχανές μπορούν να προσπελάσουν και να κατανοήσουν το περιεχόμενο των ιστοσελίδων [3].

2.2 Αρχιτεκτονική Σημασιολογικού Ιστού

2.2.1 Αρχική αρχιτεκτονική

Σύμφωνα με την αρχική πρόταση του Tim Berners Lee ο σημασιολογικός ιστός θα στηριζόταν μέσω της στοίβας τεχνολογιών που απεικονίζονται στο Σχήμα 2-1. Αυτή η στοίβα παρουσιάζει την αρχιτεκτονική του σημασιολογικού ιστού και τις τεχνολογίες που θα χρησιμοποιηθούν προκειμένου να πραγματοποιηθεί αυτό το όραμα. Η στοίβα του σημασιολογικού ιστού είναι γνωστή και ως Semantic Layer Cake.



Σχήμα 2-1 Αρχική αρχιτεκτονική σημασιολογικού ιστού (Πηγή Εικόνας: <https://www.w3.org/Talks/2001/1102-semweb-fin/all.htm>)

Στη βάση της στοίβας βρίσκονται τα Unicode και Uniform Resource Identifier (URI). Το Unicode είναι ένα παγκόσμιο σύστημα κωδικοποίησης χαρακτήρων. Είναι παρόμοιο με το ASCII TABLE με τη διαφορά ότι σε αυτό το σύστημα γίνεται κωδικοποίηση περισσότερων χαρακτήρων από διάφορες γλώσσες και όχι μόνο από τους χαρακτήρες που περιέχονται στο λατινικό αλφάβητο [4]. Η χρήση του URI είναι να δώσει ένα μοναδικό όνομα στα αντικείμενα που βρίσκονται στον ιστό με αυτό τον τρόπο αποφεύγονται οι συγκρούσεις ονομάτων [5].

Στο 2^ο επίπεδο της στοίβας βρίσκεται η XML, μια markup γλώσσα που επιτρέπει μια δομημένη αναπαράσταση των δεδομένων στον ιστό. Αυτή η γλώσσα κάνει χρήση μη προκαθορισμένων tags, τα οποία καθορίζονται από τους δημιουργούς τους ανάλογα με το περιεχόμενο του αντικειμένου που περιγράφουν. Μια περιγραφή ξεκινά από το άνοιγμα ενός tag και τελειώνει με το κλείσιμο του συγκεκριμένου tag. Αξίζει να επισημανθεί ότι αυτή η γλώσσα είναι κατανοητή από τους ανθρώπους, αλλά και προσπελάσιμη από μηχανές, γιατί κάθε πληροφορία περιγράφεται. Αυτό κάνει το XML κατάλληλο για να περιγράψει μετάδεδομένα στον σημασιολογικό ιστό. Παρακάτω ακολουθεί ένα απλό παράδειγμα δεδομένων που είναι διατυπωμένα σε XML.

```
<book>
  <title>A Semantic Web Primer</title>
  <author> Grigoris Antoniou</author>
  <author> Frank van Harmelen</author>
  <publisher>The MIT Press</publisher>
  <year>2008</year>
</book>
```

Στο πιο πάνω παράδειγμα το αντικείμενο που περιγράφεται είναι ένα βιβλίο στο οποίο υπάρχουν όλες οι χρήσιμες πληροφορίες που το αφορούν ανάμεσα σε tags. Η αρχή της περιγραφής φαίνεται από το άνοιγμα του tag (<book>) και το κλείσιμο της περιγραφής φαίνεται από το κλείσιμο του tag (</book>) [6].

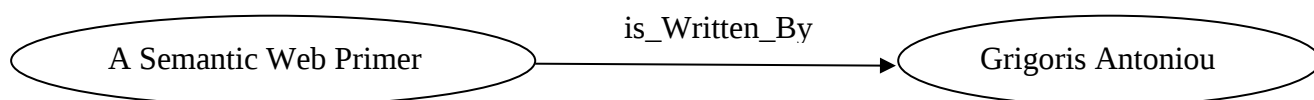
Στο 3^ο επίπεδο της στοίβας βρίσκεται το RDF (Resource Description Framework) ένα μοντέλο το οποίο προσθέτει τη σημασία στα δεδομένα και έχει XML σύνταξη. Είναι μια μέθοδος αναπαράστασης, ανταλλαγής και επαναχρησιμοποίησης των μεταδεδομένων [7]. Βασικό συστατικό του είναι η δημιουργία προτάσεων σε μορφή τριάδων (subject, predicate, object). Αξίζει να επισημανθεί ότι και τα 3 δομικά στοιχεία της τριάδας έχουν το δικό τους μοναδικό URI το οποίο μπορεί να παραπέμπει σε άλλη πηγή.

Subject: το πράγμα που περιγράφεται στην πρόταση.

Predicate: δείχνει την σχέση μεταξύ subject και object.

Object: μια τιμή η οποία ικανοποιεί την συσχέτιση.

Ακολουθεί ένα παράδειγμα που αναπαρίσταται γραφικά η έννοια της τριάδας:



Στο πιο πάνω παράδειγμα το subject είναι A Semantic Web Primer, το predicate είναι is_Written_By και το object στην προκειμένη περίπτωση είναι ο Grigoris Antoniou.

Το XML Schema περιορίζει τη δομή ενώ το RDF Schema ορίζει το κοινό λεξιλόγιο που χρησιμοποιείται στο RDF μοντέλο. Συγκεκριμένα στο RDF Schema γίνεται καθορισμός των ιδιοτήτων που ισχύουν στα αντικείμενα και τις μεταξύ τους συσχετίσεις. Επιπλέον ένα πρόβλημα που επιλύεται μέσω του RDF Schema είναι ότι κατανοεί γενικευμένες έννοιες το οποίο δεν είναι εφικτό με ένα XML – RDF μοντέλο. Για παράδειγμα μπορούμε να ορίσουμε ότι τα θηλαστικά είναι υποκλάση των σπονδυλωτών ζώων. Με αυτή την πρόταση δηλώνουμε ότι όλα τα θηλαστικά είναι επίσης και σπονδυλωτά ζώα. [6]

Στο 4^ο επίπεδο ορίζονται οι οντολογίες οι οποίες περιέχουν την κοινή συμφωνημένη γνώση για ένα συγκεκριμένο τομέα, έχοντας ως κύριο σκοπό να διευκολύνουν την επικοινωνία μεταξύ ανθρώπων και εφαρμογών [7]. Το RDF και RDF Schema από μόνα τους είναι πολύ περιοριστικά στην εκφραστικότητα. Για αυτό υπήρξε η ανάγκη εύρεσης μιας πιο εκφραστικής γλώσσας, της OWL που ο κύριος στόχος της είναι η μοντελοποίηση

οντολογιών. Συγκεκριμένα η OWL υποστηρίζει επιπλέον τον καθορισμό ξένων εννοιών/κλάσεων (πχ. Κορίτσι – Αγόρι), περιορισμούς πληθικότητας (πχ ένα άτομο έχει ακριβώς 2 γονείς), ειδικά χαρακτηριστικά (πχ. transitive, inverse) και άλλα. Ακόμη η OWL γλώσσα έχει RDF/XML σύνταξη [6]. Επιπλέον ανάλογα με τον βαθμό εκφραστικότητας που χρησιμοποιείται η OWL χωρίζεται σε τρεις κατηγορίες:

1. OWL Lite: αυτές οι οντολογίες είναι οι λιγότερο εκφραστικές και αποτελούνται από μια ιεραρχία και απλούς περιορισμούς όπως πληθικότητα.
2. OWL DL: αυτές οι οντολογίες είναι στο μέγιστο βαθμό εκφραστικές χωρίς όμως να στερούμαστε από υπολογιστική πληρότητα δηλαδή όλες οι συνεπαγωγές εκτελούνται σε πεπερασμένο χρόνο.
3. OWL Full: αυτές οι οντολογίες είναι στον μέγιστο βαθμό εκφραστικές και έχουν την ελευθερία της σύνταξης του RDF χωρίς όμως να υπάρχει εγγύηση στον υπολογισμό αποτελεσμάτων [8].

Στο 5^ο επίπεδο της λογικής χρησιμοποιείται για περισσότερη ενίσχυση των οντολογιών και επιτρέπει να προγραμματίσουμε δηλωτικά, δηλαδή περιγράφουμε τι θα κάνει το πρόγραμμα χωρίς να μας νοιάζει πως θα το κάνει. Αυτό επιτυγχάνεται μέσω κανόνων λογικής οι οποίοι έχουν τη συγκεκριμένη μορφή “If conditions, then conclusion”. Ένα απλό παράδειγμα που φαίνεται η δύναμη της λογικής είναι το ακόλουθο:

Έχουμε τα εξής δεδομένα στην βάση γνώση μας:

1. Όλοι οι άνθρωποι είναι θνητοί
2. Ο Σωκράτης είναι άνθρωπος

Επομένως μπορούμε να καταλήξουμε στο συμπέρασμα ότι ο Σωκράτης είναι θνητός. Είναι προφανές ότι οι κανόνες λογικής παίζουν σημαντικό ρόλο στην επέκταση των οντολογιών, καθώς δεδομένα που δεν ήταν ρητά στη βάση γνώσης, προθέτονται μέσω των κανόνων – αξιωμάτων που ορίστηκαν προηγουμένως [6].

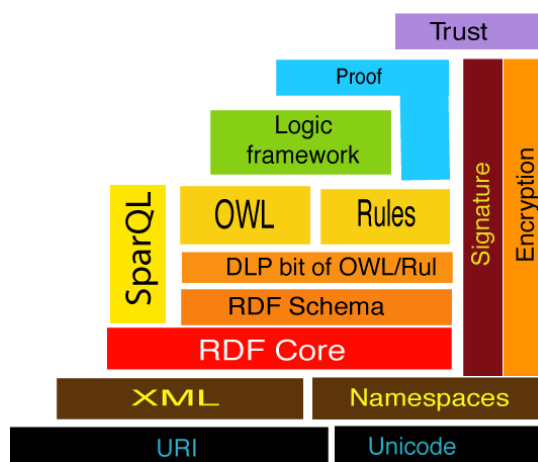
Το 6^ο επίπεδο (proof layer) έχει ως στόχο να εκτελέσει τους κανόνες που δόθηκαν από το προηγούμενο επίπεδο και να παρουσιάσει τις απαντήσεις σε μορφή αποδείξεων [6].

Στην κορυφή της στοίβας βρίσκεται το επίπεδο εμπιστοσύνης το οποίο αφορά στην εμπιστοσύνη των χρηστών του ιστού σε θέματα ασφάλειας και ποιότητας των δεδομένων. Αυτό επιτυγχάνεται μέσω ψηφιακών [6].

Στο επίπεδο digital signature είναι απαραίτητο για να εξασφαλίσει την εγκυρότητα των εγγράφων και συνεπώς να οδηγήσει στην εμπιστοσύνη των χρηστών. Με τις ψηφιακές υπογραφές γνωρίζουμε το άτομο που έγραψε την συγκεκριμένη πληροφορία (την προέλευση της). Έχει κριτική σημασία η γνώση της προέλευσης της γιατί από αυτό εξαρτάται αν θα εμπιστευτούμε την συγκεκριμένη πηγή [7].

2.2.2 Μια εναλλακτική αρχιτεκτονική

Η αρχική αρχιτεκτονική θεωρούσε ότι όλες οι γλώσσες του σημασιολογικού ιστού πρέπει να τοποθετούνται σε μια στοίβα και επέκταση της στοίβας από νεότερες τεχνολογίες γίνεται με την τοποθέτηση τους στην κορυφή της στοίβας. Αυτό προϋποθέτει όμως ότι οι νέες τεχνολογίες που θα εισάγονται πρέπει να είναι συμβατές με τα προηγούμενα επίπεδα της στοίβας. Όμως πολλοί ειδικοί αυτού του πεδίου υποστηρίζουν ότι η συγκεκριμένη αρχιτεκτονική δεν είναι ρεαλιστική. Εξάλλου ο σημασιολογικός ιστός είναι ακόμη στα πρώτα του βήματα και δε γνωρίζουμε τι προβλήματα ή απαιτήσεις θα προκύψουν στο μέλλον. Για αυτούς τους λόγους προτείνεται μια εναλλακτική αρχιτεκτονική που θεωρείται πιο εφικτή και ρεαλιστική από την αρχική. Η εναλλακτική αρχιτεκτονική του σημασιολογικού ιστού φαίνεται πιο κάτω στην Σχήμα 2-2 [9].



Σχήμα 2-2 Εναλλακτική αρχιτεκτονική σημασιολογικού ιστού (Πηγή Εικόνας: <https://www.w3.org/2005/Talks/0517-boit-tbl/>)

2.3 Semantic reasoner

Οι reasoners είναι εργαλεία που επιτρέπουν την διεξαγωγή συμπερασμάτων στον σημασιολογικό ιστό. Έχουν τη δυνατότητα να συμπεράνουν νέες σχέσεις μεταξύ αντικειμένων, καθώς και να ανιχνεύσουν λογικές ασυνέπειες σε μία βάση γνώσης. Πιο συγκεκριμένα ένας reasoner μπορεί να συνεργαστεί με μία οντολογία για να κατηγοριοποιήσει στιγμιότυπα, να ελέγξει για λάθη (αποσφαλμάτωση) και να απαντήσει ερωτήματα. Κάποιοι από τους πιο γνωστούς είναι : HermiT, Pellet, RACER, FACT++ [10].

Κεφάλαιο 3

Οντολογίες

3.1 Τι είναι μια οντολογία (ορισμός)	12
3.2 Τα είδη των οντολογιών	13
3.3 Βασικά συστατικά οντολογιών	16
3.4 Χρησιμότητα οντολογιών	19
3.5 Μεθοδολογία κατασκευής οντολογιών	19
3.6 Σημαντικά εργαλεία κατασκευής και επεξεργασίας οντολογιών	20

3.1 Τι είναι μια οντολογία (ορισμός)

Η οντολογία σαν έννοια προέρχεται από τη φιλοσοφία. Όμως ανάλογα με το πεδίο που χρησιμοποιείται αυτή η έννοια έχει διαφορετική σημασία. Από φιλοσοφική σκοπιά η οντολογία είναι ένας βασικός κλάδος της φιλοσοφίας που μελετά τη φυσική υπόσταση των υπαρκτών πραγμάτων. Όμως από την οπτική γωνία της πληροφορικής αυτός ο όρος χρησιμοποιείται για να αναπαραστήσει όλη τη γνώση που έχουμε γύρω από ένα συγκεκριμένο θέμα. Η γνώση αναπαρίσταται από μια ιεραρχία κλάσεων (έννοιών) και τις μεταξύ τους συσχετίσεις. Όμως για να επιτευχθεί μια οντολογία δεν αρκεί μόνο να χρησιμοποιηθεί το κατάλληλο λεξιλόγιο (έννοιες) χρειάζεται ακόμη οι έννοιες να είναι

καλά οργανωμένες μεταξύ τους για να παράγουν ένα σωστό εννοιολογικό αποτέλεσμα [11].

Σύμφωνα με τον Gruber η οντολογία είναι «an explicit specification of a conceptualization», όπου γίνεται μια αναπαράσταση του κόσμου [12]. Αργότερα ο Borst έδωσε τον δικό του ορισμό για την οντολογία ως “formal specification of a shared conceptualization.” Και ο τελικός ορισμός επήλθε από τον Studer το 1998 “An ontology is a formal, explicit specification of a shared conceptualization”. Αυτός ο ορισμός είναι ο πιο διαδεδομένος και υποδηλώνει ότι η οντολογία πρέπει να είναι επεξεργάσιμη και προσπελάσιμη από τις μηχανές (formal), να γίνεται ρητή περιγραφή/ αναπαράσταση του κόσμου δηλαδή του θέματος που αφορά η συγκεκριμένη οντολογία και αυτή η περιγραφή να είναι κοινά συμφωνημένη (shared) [13].

3.2 Τα είδη των οντολογιών

Υπάρχουν πάρα πολλά είδη οντολογιών. Οι οντολογίες διακρίνονται κυρίως από την εκφραστικότητα τους και από τον σκοπό τους [8].

3.2.1 Με βάσει την εκφραστικότητα

Παρακάτω παρουσιάζονται τα διάφορα είδη οντολογιών κατά αύξουσα σειρά (ξεκινώντας από τις λιγότερες εκφραστικές οντολογίες μέχρι τις πιο εκφραστικές) [8].

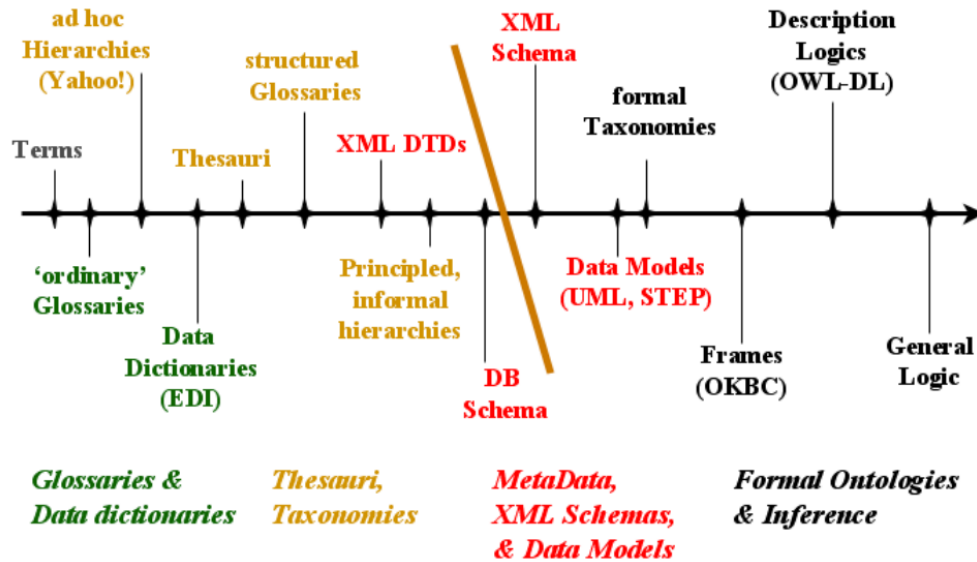
Information Ontologies

Η κύρια χρήση τους είναι για να παρουσιάσουν σχηματικά σε ποια φάση/κατάσταση βρίσκεται ένα οποιοδήποτε project. Αυτού του είδους οι οντολογίες γίνεται πιο εύκολο αντιληπτό για τους ανθρώπους να οργανώσουν και να ξεκαθαρίσουν τις ιδέες τους κατά την ανάπτυξη ενός project. [8]

Linguistic/Terminological Ontologies

Στην πιο κάτω εικόνα φαίνονται τα διάφορα είδη linguistic ontologies που μπορούμε να έχουμε. Όσο προχωράμε προς τα δεξιά ανεβαίνει ο βαθμός εκφραστικότητας και γίνονται πιο αυστηρές και λογικές οι οντολογίες. Οι οντολογίες αυτές υλοποιούν δύο στόχους:

1. Να ορίσουν ένα λεξιλόγιο
2. Να συμφωνηθεί μια ορολογία ούτως ώστε να αποφευχθεί η ασάφεια.



Σχήμα 3-1 Κατηγορίες γλωσσολογικών οντολογιών (Πηγή: [14])

Software Ontologies

Χρησιμοποιούνται για διαχείριση και επεξεργασία δεδομένων κατά την ανάπτυξη λογισμικών δραστηριοτήτων, έχοντας ως κύριο μέλημα τη συνέπεια των δεδομένων. Αυτού του είδους οι οντολογίες δημιουργούνται με τις εξής εννοιολογικές γλώσσες μοντελοποίησης: Entity-Relationship Model, Object Model Language, UML κ.ά. [8].

Formal Ontologies

Αυτές οι οντολογίες έχουν ξεκάθαρες έννοιες που η σημασιολογία τους βασίζεται σε αυστηρούς κανόνες. Έχουν ως στόχο να βγάλουν συμπεράσματα από τα δεδομένα που έχουν στη βάση γνώση τους. Αυτές οι οντολογίες περιγράφονται με διάφορες γλώσσες όπως: Description Logic, Conceptual Graphs, First Order Logic και OWL. [8]

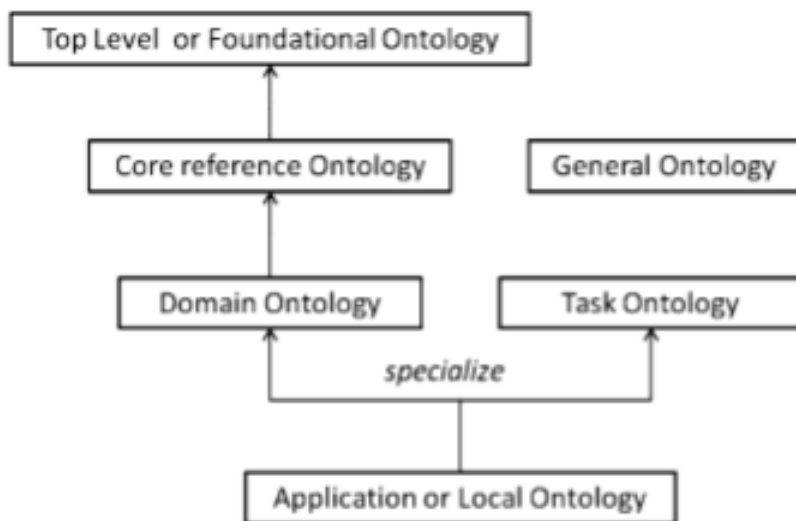
3.2.2 Με βάση τον σκοπό ή το επίπεδο αφαιρετικότητας της οντολογίας

Domain Ontologies

Αυτές οι οντολογίες περιέχουν την γνώση για ένα συγκεκριμένο πεδίο ενδιαφέροντος [8].

Local Ontologies/Application Ontologies

Αυτού του είδους οι οντολογίες είναι κατηγορίες της domain ontology, όμως έχουν σχεδιαστεί βάσει της γνώμης ενός ατόμου. Ακόμη οι application ontologies είναι ένα συνδυασμός domain ontology και task ontology [8]. Αυτές οι οντολογίες είναι κατάλληλες για κατευθείαν χρήση τόσο στον reasoner για εξαγωγή νέας γνώσης όσο και σε λογισμικά [13].



Σχήμα 3-2 Κατηγοριοποίηση οντολογιών με βάση τον σκοπό (Πηγή: [8])

Task Ontologies

Η συγκεκριμένη κατηγορία έχει τη γνώση που απαιτείται για να διεκπεραιωθεί μια εργασία [8].

Core Reference Ontologies

Η συγκεκριμένη κατηγορία είναι το αποτέλεσμα της ενσωμάτωσης πολλών domain οντολογιών και χρησιμεύει για να απεικονίσει τις κεντρικές έννοιες και συσχετίσεις του συγκεκριμένου πεδίου ενδιαφέροντος [8].

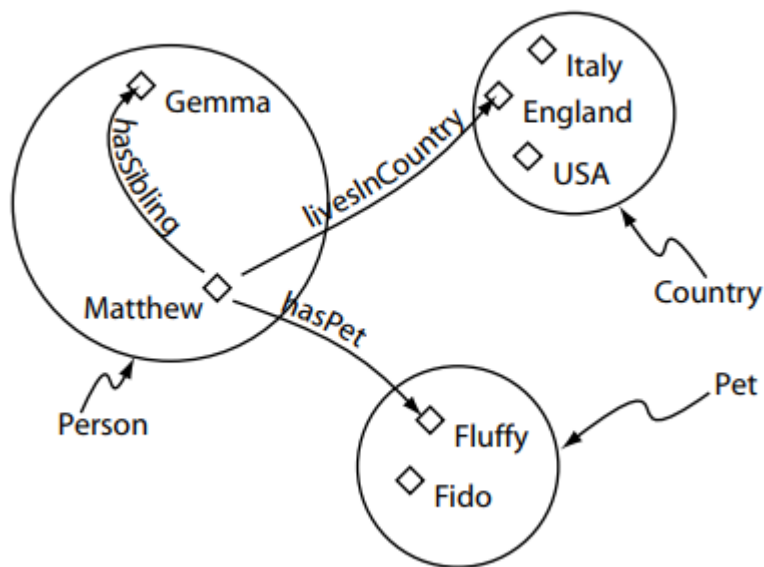
General Ontologies

Οι οντολογίες αυτές δεν έχουν ένα συγκεκριμένο πεδίο ενδιαφέροντος αλλά αναπαριστούν την γενική γνώση [8].

Foundational Ontologies/Top Level Ontologies/Upper Level Ontologies

Αυτές οι οντολογίες περιγράφουν πολύ γενικές έννοιες οι οποίες είναι ανεξάρτητες ενός συγκεκριμένου θέματος [15]. Επομένως είναι κατάλληλες σε πολλές διαφορετικές οντολογίες πεδίου. Έχουν ως κύριο σκοπό να παρέχουν μια ευρεία εικόνα του κόσμου [13].

3.3 Βασικά συστατικά οντολογιών

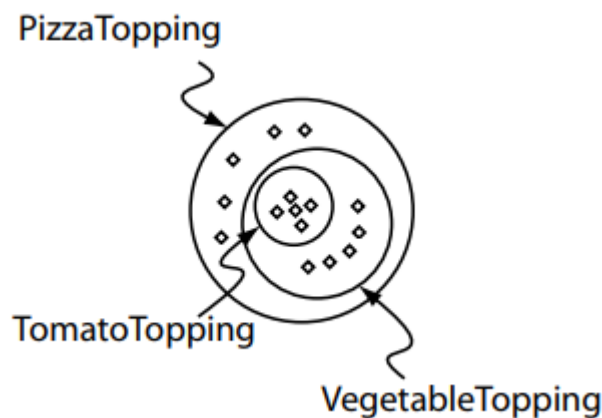


Σχήμα 3-3 Κλάσεις που περιέχουν στιγμιότυπα (Πηγή: [19])

Τα κύρια συστατικά που είναι απαραίτητα για να δημιουργηθεί μια οντολογία είναι τα ακόλουθα:

Κλάση (Class): περιγράφει ένα σύνολο αντικειμένων που έχουν παρόμοιες ιδιότητες. Τα αντικείμενα στον κόσμο ομαδοποιούνται σε διάφορες ομάδες με βάση τα κοινά χαρακτηριστικά τους. Αυτές οι ομάδες είναι οι γνωστές κλάσεις. Παράδειγμα μιας κλάσης είναι ο άνθρωπος [16]. Σε μια οντολογία οι κλάσεις περιγράφουν έννοιες σχετικές με τον συγκεκριμένο τομέα που μας απασχολεί [17].

Υποκλάση (Subclass): περιγράφει μερικά χαρακτηριστικά που δεν είναι κοινά σε όλη την ομάδα. Παράδειγμα μιας υποκλάσης για την κλάση «άνθρωπος» είναι η «γυναίκα» [18]. Περιγράφουν έννοιες πιο ειδικές από τις υπέρκλάσεις τους. Επομένως είναι λογικό να έχουν επιπρόσθετες ιδιότητες από τις υπέρκλάσεις τους. Οι υποκλάσεις κληρονομούν τις ιδιότητες των υπέρκλάσεις τους [17].



Σχήμα 3-4 Αναπαράσταση υποκλάσης (Πηγή: [19])

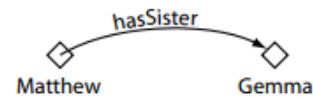
Στο Σχήμα 3-4 φαίνεται ότι το «TomatoTopping» είναι υποκλάση του «VegetableTopping» που είναι υποκλάση του «PizzaTopping». Επομένως όλα τα στιγμιότυπα της κλάσης «TomatoTopping» είναι «VegetableTopping» και «PizzaTopping» ταυτόχρονα.

Στιγμιότυπο (Individual): Είναι ένα αντικείμενο που ανήκει σε μια κλάση [18] και απεικονίζει ένα απτό πράγμα ή μια εικονική έννοια. Πολλές φορές είναι δύσκολο να διακρίνεις ένα στιγμιότυπο από μια κλάση. Τυπικά τα στιγμιότυπα είναι συγκεκριμένα πράγματα που έχουν μια ταυτότητα. Παράδειγμα ενός στιγμιότυπου για την κλάση «άνθρωπος» είναι η «Ελλη» [16]. Ακόμη τα στιγμιότυπα μιας κλάσης αποτελούν τη βάση γνώσης της οντολογίας [17].

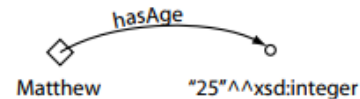


Σχήμα 3-5 Στιγμιότυπα (Πηγή: [19])

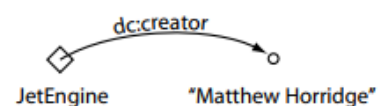
Ιδιότητα (property/slot): Οι ιδιότητες χωρίζονται στις εξής κατηγορίες: object properties, datatype properties και annotation properties. Τα object properties χρησιμοποιούνται για να συσχετίσουν δύο στιγμιότυπα μαζί. Τα datatype properties συσχετίζουν ένα στιγμιότυπο με μια τιμή. Τα annotation properties χρησιμοποιούνται για να προσθέσουν μεταδεδομένα για τις κλάσεις, στιγμιότυπα και object/datatype properties.



An object property linking the individual Matthew to the individual Gemma



A datatype property linking the individual Matthew to the data literal '25', which has a type of an xsd:integer.



An annotation property, linking the class 'JetEngine' to the data literal (string) "Matthew Horridge".

Σχήμα 3-6 Κατηγορίες Ιδιοτήτων (Πηγή: [19])

Περιορισμούς (property restriction/facet): ορίζονται οι κατηγορίες κλάσεων με βάση τις ιδιότητες που έχουν τα στιγμιότυπά τους. Υπάρχουν τρία διαφορετικά είδη περιορισμών: τα Quantifier Restrictions – περιορισμός ποσοδεικτών (υπαρξιακός και καθολικός), Cardinality

Restrictions (καθορισμός αριθμού συσχετίσεων) και hasValue Restrictions (περιορισμούς στις αποδεκτές τιμές σε αντικείμενα/στιγμιότυπα) [19].

3.4 Χρησιμότητα οντολογιών

3.4.1 Διαχείριση πληροφοριών:

Οργάνωση και δόμηση πληροφοριών συγκεκριμένου πεδίου ενδιαφέροντος της οντολογίας. Ορισμός του σχετικού λεξιλογίου που πρέπει να υπάρχει για την αναπαράσταση του κόσμου που αναλύει. Στην παρούσα χρήση μια οντολογία λειτουργεί ως χάρτης που καθορίζει τις διάφορες βασικές έννοιες που μπορούν να εντοπιστούν στον συγκεκριμένο τομέα. Ακόμη επιτρέπει επαναχρησιμοποίηση γνώσης, καθώς και εύκολη τροποποίησή της σε περίπτωση που δεν ισχύουν πλέον οι υφιστάμενες πληροφορίες [17].

3.4.2 Επίλυση προβλημάτων:

Μπορούμε να διατυπώσουμε ένα πρόβλημα σαν μια σειρά ερωτημάτων που δίνονται σε έναν reasoner. Ο reasoner χρησιμοποιεί τη γνώση που είναι δηλωμένη από την οντολογία για να δώσει τις ανάλογες απαντήσεις.

3.4.3 Παροχή κοινού πλαισίου

Παροχή μιας κοινής κατανόησης της πληροφορίας τόσο στους ανθρώπους όσο και στις μηχανές [17]. Αυτό θα έχει ως αποτέλεσμα πιο ακριβή και σαφή αναζήτηση, αφού πλέον θα αναζητούμε σημασιολογικά και έτσι θα αποφεύγονται τα προβλήματα που υπάρχουν από τις αναζητήσεις βάση λέξεων-κλειδιών (αντιμετώπιση πολυσημίας, αμφισημίας). Επιπλέον, με την υποστήριξη δια-λειτουργικότητας, οι εφαρμογές μπορούν να κάνουν ανταλλαγές πληροφοριών ηλεκτρονικά [20] .

3.5 Μεθοδολογία κατασκευής οντολογιών

Υπάρχουν πολλές μεθοδολογίες για την κατασκευή οντολογίας. Εναπόκειται στον κατασκευαστή αν θα χρησιμοποιήσει μια συγκεκριμένη μεθοδολογία ή όχι.

Όμως, στην παρούσα διπλωματική επέλεξα την μεθοδολογία που έχουν προτείνει η Natalya F. Noy και η Deborah L. McGuinness [17]. Αυτή η μεθοδολογία είναι απλή και συστήνεται για αρχάριους. Επιπρόσθετα αποτελείται από τις ακόλουθες 7 φάσεις:

1. Καθορισμός του σκοπού και του πεδίου που θα καλύψει η οντολογία
2. Επαναχρησιμοποίηση σχετικών οντολογιών
3. Απαρίθμηση σημαντικών όρων της οντολογίας

4. Καθορισμός κλάσεων και ιεραρχία κλάσεων
5. Καθορισμός ιδιοτήτων (properties - slots) των κλάσεων
6. Καθορισμός περιορισμών (facets) των ιδιοτήτων
7. Δημιουργία στιγμιότυπων

Σημείωση: για περισσότερες πληροφορίες μπορείτε να δείτε στο κεφάλαιο «Ανάλυσης μεθοδολογίας» στο οποίο αναλύω τι έκανα στην κάθε φάση ξεχωριστά.

3.6 Σημαντικά εργαλεία κατασκευής και επεξεργασίας οντολογιών

Υπάρχουν πολλά χρήσιμα εργαλεία που βοηθούν στην κατασκευή και επεξεργασία οντολογιών διαθέσιμα στο διαδίκτυο. Δύο από τα πιο γνωστά είναι το Protégé [21] και το Swoop [22].

3.6.1 Protégé

Το Protégé είναι ένα λογισμικό ανοιχτού κώδικα γραμμένο στην Java, το οποίο διατίθεται δωρεάν. Χρησιμοποιείται για την αναπαράσταση της γνώσης σε ένα πεδίο ενδιαφέροντος που έχει επιλέξει ο χρήστης. Αυτό το εργαλείο κατασκευάστηκε από το πανεπιστήμιο του Stanford και έχει χιλιάδες χρήστες παγκοσμίως που το χρησιμοποιούν. Εκτός από την μοντελοποίηση της γνώσης το Protégé είναι ένα χρήσιμο εργαλείο που παρέχει τη δυνατότητα να απαντά σε ερωτήσεις που θέτει ο χρήστης λαμβάνοντας υπόψη την βάση γνώση που είχε δημιουργηθεί προηγουμένως [21].

Η αρχιτεκτονική του Protégé χωρίζεται σε δύο τμήματα (model και view part). Το model part είναι ο εσωτερικός μηχανισμός αναπαράστασης οντολογιών και βάσεων γνώσεων, ενώ το view part είναι η διεπαφή που χρησιμοποιεί ο χρήστης για να δει και να επεξεργαστεί την οντολογία. Η αρχιτεκτονική του Protégé δίνει την δυνατότητα στους προγραμματιστές να αναπτύξουν τα δικά τους plugins. Επιπλέον το Protégé χρησιμοποιείται για να φορτώσει, να επεξεργαστεί και να αποθηκεύσει οντολογίες σε διάφορες μορφές όπως RDF, XML, OWL και άλλες [23].

3.6.2 Swoop

Είναι μια Java εφαρμογή που σκοπό έχει να κατασκευάζει και να τροποποιεί οντολογίες ιστού (OWL) παρέχοντας ταυτόχρονα όλα τα οφέλη του ιστού. Η αρχιτεκτονική του

βασίζεται στο Model-View-Controller (MVC). Τα χαρακτηριστικά που υποστηρίζει αυτό το εργαλείο είναι τα ακόλουθα [22]:

1. Οι οντολογίες που κατασκευάζονται βασίζονται στην αρχιτεκτονική του ιστού: Είναι εύκολο να επαναχρησιμοποιήσεις οντολογίες μέσω συνδέσεων ή συγχωνεύσεων. Ακόμη οι οντολογίες φορτώνονται μέσω της εισαγωγής του URL.
2. Επεξεργασία οντολογιών που είναι διαθέσιμες στον ιστό: οντολογίες σχετικές με το θέμα και ικανοποιείται η ανάγκη της δια λειτουργικότητας όπου τα δεδομένα μιας οντολογίας μπορούν να γίνουν δεδομένα μιας άλλης οντολογίας
3. Επιλύει το πρόβλημα της μερικής επαναχρησιμοποίησης δηλαδή ο χρήστης έχει τη δυνατότητα να επαναχρησιμοποιήσει μόνο το κομμάτι που τον ενδιαφέρει και να αποφύγει να ενσωματώσει όλη την οντολογία η οποία του είναι περιττή.
4. Έχει τις προδιαγραφές της OWL: υποστηρίζει διάφορες μορφές σύνταξης, reasoning, παρέχει την δυνατότητα για debugging και επιδιόρθωση μέσω glass box και black box τεχνικών.
5. Επιτρέπει μια ομάδα ατόμων να δουλεύουν μαζί σε διαφορετικούς υπολογιστές προκειμένου να κατασκευαστεί η οντολογία, καθώς τους παρέχει εύκολη ανταλλαγή δεδομένων, μπορούν να σχολιάσουν και να έχουν έλεγχο της έκδοσης και της επικύρωσης. Επομένως το Swoop είναι σχεδιασμένο για να διευκολύνει την συνεργασία κατά την διάρκεια ανάπτυξης της οντολογίας.

Κεφάλαιο 4

Ρητορική μίσους

4.1	Ελευθερία του λόγου και έκφρασης	22
4.2	Πώς σχετίζεται η ελευθερία του λόγου με την ρητορική του μίσους;	23
4.3	Τί είναι η ρητορική μίσους;	23
4.4	Κριτήρια αναγνώρισης ρητορικής μίσους	23
4.5	Άτομα που μπαίνουν στο στόχαστρο	24
4.6	Αίτια ρητορικής μίσους	24
4.7	Επιπτώσεις ρητορικής μίσους	25
4.8	Τρόποι αντιμετώπισης	25

4.1 Ελευθερία του λόγου και έκφρασης

Η ελευθερία του λόγου είναι ένα θεμελιώδες ανθρώπινο δικαίωμα. Είναι ζωτικής σημασίας να μπορούμε να εκφραζόμαστε ελεύθερα χωρίς να φοβόμαστε ότι θα υπάρξουν συνέπειες. Άλλωστε ζούμε σε μια δημοκρατική κοινωνία όπου ο διάλογος αποτελεί βασική προϋπόθεση της για να μπορεί να λειτουργήσει. Παρόλα αυτά το δικαίωμα αυτό μπορεί να περιοριστεί όταν θέτει σε κίνδυνο έναν άλλον άνθρωπο ή γενικότερα την κοινωνία [24].

4.2 Πώς σχετίζεται η ελευθερία του λόγου με την ρητορική του μίσους;

Το δικαίωμα της ελευθερίας της έκφρασης έχει άμεση σχέση με το φαινόμενο της ρητορικής μίσους. Καθώς μας θέτει κάποια όρια ως προς το τι θα επιτρέπεται να ειπωθεί χωρίς να γίνεται καταπάτηση άλλων ανθρωπίνων δικαιωμάτων [24].

4.3 Τί είναι η ρητορική μίσους;

Δεν υπάρχει ένας ενιαίος αποδεκτός ορισμός για το φαινόμενο της ρητορικής του μίσους. Αυτό οφείλεται στο γεγονός ότι οι νομοθεσίες διαφέρουν από χώρα σε χώρα. Μια τοξική έκφραση μπορεί να θεωρηθεί ως ρητορική μίσους σε μια χώρα, όμως η ίδια έκφραση μπορεί να μην θεωρείται ως ρητορική μίσους σε άλλη χώρα. Επομένως παίζει ρόλο το νομικό σύστημα κάθε χώρας για το αν μια έκφραση κρίνεται ως ποινικά κολάσιμη ρητορική μίσους ή όχι. Αυτό ισχύει και στο διαδίκτυο. Επίσης διάφορες διαδικτυακές υπηρεσίες (πχ facebook) έχουν αναπτύξει τις δικές τους πολιτικές για το τι θεωρείται μη αποδεκτό. Εμείς που κάνουμε χρήση αυτών των υπηρεσιών καλούμαστε να συμμορφωνόμαστε σε αυτές τις πολιτικές.

Ωστόσο αξίζει να λεχθεί τι θεωρείται ρητορική μίσους σύμφωνα με την επιτροπή υπουργών του συμβουλίου της Ευρώπης. “Ρητορική μίσους ή αλλιώς hate speech καλύπτει όλες τις μορφές έκφρασης που διαδίδουν, υποκινούν, προωθούν ή δικαιολογούν το φυλετικό μίσος, την ξеноφοβία, τον αντισημιτισμό ή άλλες μορφές μίσους που στηρίζονται στην μισαλλοδοξία ...” [25].

Με απλά λόγια ρητορική μίσους εκφράζει τον έντονο μίσος ή απέχθεια που τρέφει ένας άνθρωπος προς σε έναν άλλον άνθρωπο ή μια ομάδα ατόμων εξαιτίας ενός χαρακτηριστικού που τους προσδιορίζει. Ακόμη από τον ορισμό που δόθηκε πιο πάνω «κάθε μορφή έκφρασης», εννοείται ότι η ρητορική μίσους μπορεί να εκδηλωθεί είτε λεκτικά, είτε με εικόνες, βίντεο και οποιαδήποτε άλλη δραστηριότητα που γίνεται στο διαδίκτυο [24].

4.4 Κριτήρια αναγνώρισης ρητορικής μίσους

Υπάρχουν διάφορα κριτήρια που πρέπει να λαμβάνουμε υπόψιν προκειμένου να αξιολογήσουμε αν ένα μεμονωμένο περιστατικό είναι ρητορική μίσους. Καταρχήν πρέπει να δίνουμε βαρύτητα στο ύφος της έκφρασης και το περιεχόμενο που χρησιμοποιείται.

Είναι σημαντικό να γνωρίζουμε εάν αυτό που ειπώθηκε έχει προσβλητικό, επιθετικό ή υποτιμητικό τόνο. Ακόμη πρέπει να προσδιορίζεται η πρόθεση του ατόμου. Αν είχε σκοπό δηλαδή να βλάψει άλλα άτομα. Ακολούθως είναι απαραίτητο να αναγνωρίζεται αν υπάρχει μια συγκεκριμένη ομάδα που θίγεται στο συγκεκριμένο περιστατικό και τι αντίκτυπο επιφέρει [24].

4.5 Άτομα που μπαίνουν στο στόχαστρο

Οποιοσδήποτε μπορεί να αποτελέσει στόχο της ρητορικής μίσους. Υπάρχουν κάποια άτομα που είναι πιο ευάλωτα από άλλα λόγω του ότι ανήκουν σε κάποια μειονότητα και έχουν διαφορετικές ιδεολογίες, πιστεύω, στάσεις ζωής από την πλειονότητα ή ένα ιδιαίτερο χαρακτηριστικό που τους προσδιορίζει όπως το χρώμα του δέρματος, τη φυλή που ανήκουν, τον σεξουαλικό προσανατολισμό κλπ. Εξαιτίας αυτού του χαρακτηριστικού αντιμετωπίζονται διαφορετικά από την κοινωνία [25, 26].

4.6 Αίτια ρητορικής μίσους

Γενικά στην κοινωνία

Ο κύριος λόγος ύπαρξης της ρητορικής μίσους είναι τα αρνητικά στερεότυπα και οι κοινωνικές πεποιθήσεις. Δυστυχώς, πολλές φορές ασυνείδητα βάζουμε ετικέτες στους ανθρώπους χωρίς πραγματικά να τους γνωρίζουμε. Κάνουμε αυθαίρετες γενικεύσεις και επειδή κανένας δεν τις αμφισβητεί, σιγά σιγά γίνονται αποδέκτες στην κοινωνία ως «αυταπόδεκτες αλήθειες». Ακόμη η ρητορική μίσους μπορεί να οφείλεται στην άγνοια που έχουμε ή τον φόβο που νιώθουμε προς οτιδήποτε είναι διαφορετικό από εμάς [27].

Στο διαδίκτυο

Το διαδικτυακό μίσος αυξάνεται με την πάροδο του χρόνου. Αυτό οφείλεται κυρίως στην ατιμωρησία και στην ανωνυμία. Κάποιοι πιστεύουν ότι μπορούν να δημοσιεύσουν οτιδήποτε, χωρίς να υπάρχει οποιοδήποτε τίμημα, αφού θεωρούν ότι δεν μπορεί να αποκαλυφθεί η ταυτότητά τους [24].

4.7 Επιπτώσεις ρητορικής μίσους

Τα άτομα που βρίσκονται στο στόχαστρο αντιμετωπίζονται διαφορετικά από την κοινωνία. Δε θεωρούνται ως ισότιμα μέλη και δεν έχουν τις ίδιες ευκαιρίες. Αυτό έχει ως αποτέλεσμα, τα άτομα αυτά να βιώνουν περιθωριοποίηση και αποκλεισμό από την κοινωνία. Ακόμη η ρητορική μίσους μπορεί να επηρεάσει αρνητικά την ψυχική υγεία των θυμάτων της. Τα άτομα μπορούν να νιώσουν συναισθηματική δυσφορία όπως άγχος, θυμό και έντονο φόβο από την κακομεταχείριση που δέχονται τόσο λεκτικά όσο σωματικά. Επιπρόσθετα, η ρητορική μίσους μπορεί σε ακραίες περιπτώσεις να οδηγήσει σε εγκλήματα μίσους. Είναι ξεκάθαρο ότι η ρητορική μίσους πλήττει τον άνθρωπο που στοχοποιείται αλλά και γενικότερα την κοινωνία [27].

4.8 Τρόποι αντιμετώπισης

Μετά από τις διαστάσεις που πήρε η ρητορική μίσους ήταν αναμενόμενο να παρθούν κάποια μέτρα για να εκλείψει ή τουλάχιστον να ελαχιστοποιηθεί. Δημιουργήθηκαν πολλές νομοθεσίες προκειμένου να διαφυλάξουν τα ανθρώπινα δικαιώματα των ευάλωτων ομάδων. Όμως, για να αντιμετωπιστεί ριζικά αυτό το πρόβλημα είναι απαραίτητο να κατανοήσουμε ποια είναι τα ανθρώπινα δικαιώματα, ούτως ώστε να γνωρίζουμε πότε ξεπερνάμε τα όρια. Ακόμη είναι εξίσου σημαντικό να σεβόμαστε και να αποδεχόμαστε την διαφορετικότητα [26].

Θα ήταν παράλειψη μου να μην αναφέρω μερικούς τρόπους αντιμετώπισης της ρητορικής μίσους στα πλαίσια του διαδικτύου. Τα άτομα που χρησιμοποιούν πρέπει να προσέχουν τον τρόπο που το χρησιμοποιούν και να σκέφτονται πρώτα τι αντίκτυπα θα έχει η δημοσίευση απόψεων. Ακόμη οι χρήστες που τυγχάνει να παρατηρήσουν τέτοιου είδους καταχρηστικές συμπεριφορές θα ήταν καλό να καταγγείλουν τα συγκεκριμένα περιστατικά μέσω μηχανισμών αναφορών παρά να μένουν παθητικοί δέκτες.

Κεφάλαιο 5

Ανάλυση Μεθοδολογίας

5.1 Πηγές πληροφοριών	27
5.2 Βήμα 1 - Καθορισμός πεδίου ενδιαφέροντος και σκοπός της οντολογίας	27
5.3 Βήμα 2 – Επαναχρησιμοποίηση υπαρχουσών οντολογιών	28
5.4 Βήμα 3 – Απαρίθμηση σημαντικών όρων οντολογίας	28
5.5 Βήμα 4 – Καθορισμός κλάσεων και ιεραρχία κλάσεων	29
5.6 Βήμα 5 – καθορισμός ιδιοτήτων των κλάσεων	31
5.7 Βήμα 6 – καθορισμός περιορισμών των ιδιοτήτων	31
5.8 Βήμα 7 – Δημιουργία στιγμιότυπων	32

Για την κατασκευή της οντολογίας έχω χρησιμοποιήσει την Simple Knowledge Engineering Methodology όπως είναι διατυπωμένη στον οδηγό κατασκευής οντολογίας του Stanford University [17]. Όσο για τη συλλογή πληροφοριών έχω χρησιμοποιήσει την μέθοδο content analysis και πιο συγκεκριμένα από γραπτές πηγές. Οι πηγές που χρησιμοποίησα για την απόκτηση πληροφορίας έχουν βάση στη νομοθεσία στο πλαίσιο της ρητορικής μίσους. Πιο κάτω ακολουθούν περισσότερες λεπτομέρειες, αρχίζοντας από τις πηγές που χρησιμοποίησα.

5.1 Πηγές πληροφοριών

Το MANDOLA (Monitoring And OnLine hAte speech) είναι μια εργασία που χρηματοδοτήθηκε από το πρόγραμμα «δικαιωμάτων, ισότητας και ιθαγένειας» (REC) που έχει στόχο να κατανοήσει την ρητορική μίσους στο διαδίκτυο και να ενθαρρύνει πολίτες να καταγγέλλουν περιστατικά ρητορικής μίσους.

Κύρια πηγή πληροφοριών είναι από το D2.1b (final report) Definition of illegal hatred and implications το οποίο είναι deliverable του mandola project [28]. Αυτή η πηγή ήταν απαραίτητη για μένα καθώς οι κανόνες που δημιούργησα στην οντολογία είναι επηρεασμένοι από τις παράνομες συμπεριφορές, όπως ορίζονται στο κοινό πλαίσιο του mandola.

5.2 Βήμα 1 - Καθορισμός πεδίου ενδιαφέροντος και σκοπός της οντολογίας

Σύμφωνα με τον οδηγό κατασκευής οντολογία του Stanford το πρώτο βήμα είναι να καθοριστεί το πεδίο ενδιαφέροντος και σκοπός της οντολογίας. Αυτό γίνεται με το να απαντήσω μερικές ερωτήσεις. Παρακάτω ακολουθούν οι ερωτήσεις μαζί με τις απαντήσεις τους.

5.2.1 Ερωτήσεις:

1. “What is the domain that the ontology will cover?”
2. “For what we are going to use the ontology?”
3. “For what types of questions the information in the ontology should provide answers?”
4. “Who will use and maintain the ontology?”

5.2.2 Απαντήσεις:

1. Το πεδίο ενδιαφέροντος που καλύπτω στην παρούσα οντολογία είναι η ρητορική μίσους σύμφωνα με νομοθετικά πλαίσια.
2. Η οντολογία θα χρησιμοποιηθεί για να παρέχει ένα κοινό πλαίσιο στους χρήστες σχετικά με την ρητορική μίσους προκειμένου να τους καθοδηγήσει να σχολιάσουν / κατηγοριοποιήσουν tweets.
3. Η οντολογία πρέπει να είναι σε θέση να απαντά κατά πόσο ένα tweet μπορεί να θεωρηθεί πιθανώς παράνομο ή καθόλου παράνομο. Αν όμως θεωρηθεί ως παράνομο τότε πρέπει να μπορεί να μας εξηγήσει γιατί το έχει αναθέσει ως παράνομο και σε ποια υποκατηγορία παράνομης συμπεριφοράς το έχει κατατάξει.

Αξίζει να τονιστεί ότι η οντολογία από μόνη της δεν μπορεί να απαντήσει όταν ο χρήστης δεν έχει καταχωρίσει ορθά τις απαραίτητες πληροφορίες που χρειάζεται μέσω της διεπαφής.

4. Η οντολογία μπορεί να χρησιμοποιηθεί από τον οποιοδήποτε ο οποίος ενδιαφέρεται να κάνει annotate tweets σχετικά με τη ρητορική μίσους. Στα πλαίσια της παρούσας διπλωματικής δεν είναι αρμοδιότητά μου να συντηρήσω τη συγκεκριμένη οντολογία.

5.3 Βήμα 2 – Επαναχρησιμοποίηση υπαρχουσών οντολογιών

Στην παρούσα εργασία δεν έχω επαναχρησιμοποιήσει άλλες οντολογίες που είναι σχετικές με το πεδίο ενδιαφέροντος που εξετάζω.

5.4 Βήμα 3 – Απαρίθμηση σημαντικών όρων οντολογίας

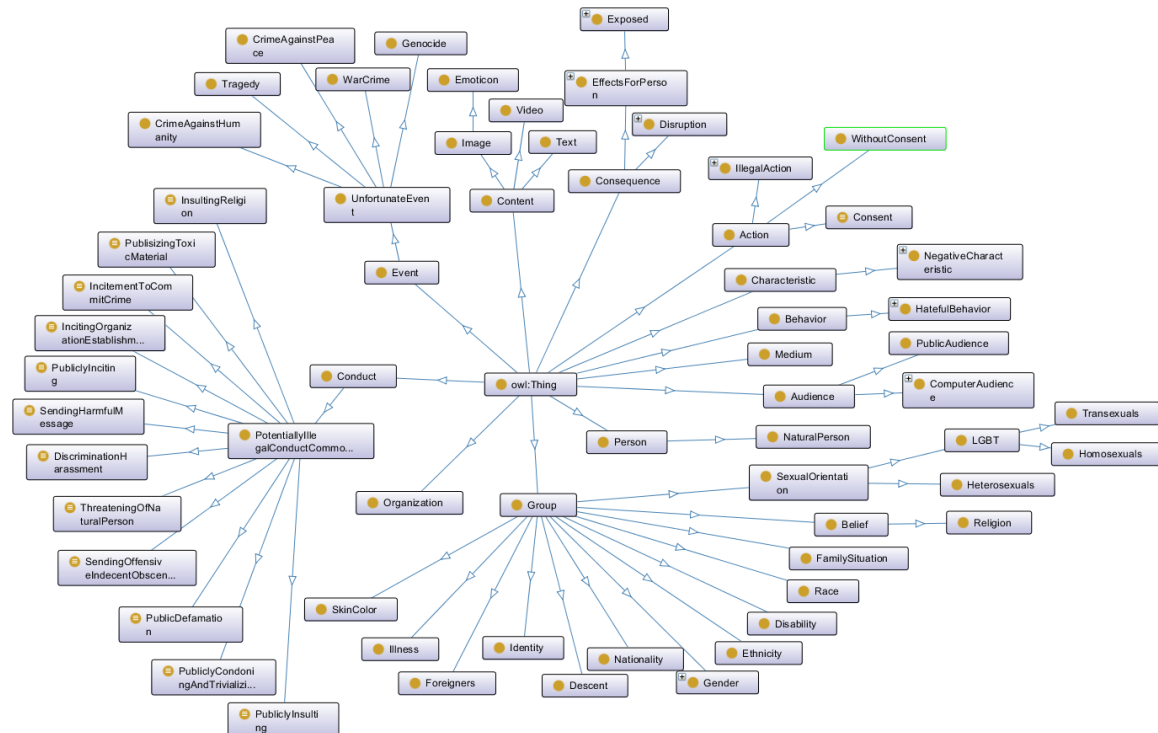
Action	Foreigners	Public Audience
Audience	Gender	Public Defamation
Behavior	Genocide	Public Order Disruption
Belief	Group	Public Peace Disruption
Characteristic	Hateful Behavior	Publicly Condoning And Trivializing
Computer Audience	Heterosexuals	Publicly Inciting
Conduct	Homosexuals	Publicly Insulting
Consent	Identity	Publisizing Toxic Material
Consequence	Illegal Action	Race
Content	Illness	Religion
Crime Against Humanity	Image	Sending Harmful Message
Crime Against Peace	Incitement To Commit Crime	Sending Offensive Indecent Obscene Content
Descent	Inciting Organization Establishment And Participation	Sexual Orientation
Disability	Insulting Religion	Skin Color
Discrimination Harassment	LGBT	Text
Disruption	Medium	Threatening Of Natural Person

Effects For Person	Nationality	Tragedy
Emoticon	Natural Person	Transexuals
Ethnicity	Negative Characteristic	Unfortunate Event
Event	Organization	Video
Exposed	Person	War Crime
Family Situation	Potentially Illegal Conduct	Without Consent

Σημείωση : Για περισσότερες λεπτομέρειες σχετικά με την οντολογία μπορείτε να δείτε στο κεφάλαιο “Ανάλυσης οντολογίας ” όπου σκοπός του συγκεκριμένου κεφαλαίου είναι να παρουσιάσει την οντολογία.

5.5 Βήμα 4 – Καθορισμός κλάσεων και ιεραρχία κλάσεων

Ακολούθησα a top-down approach κατά την κατασκευή της οντολογίας δηλαδή ξεκίνησα να ορίζω πρώτα τις γενικές κλάσεις και αργότερα τις πιο εξειδικευμένες κλάσεις. Πιο κάτω δείχνω την ιεραρχία των κλάσεων που έχω ορίσει:



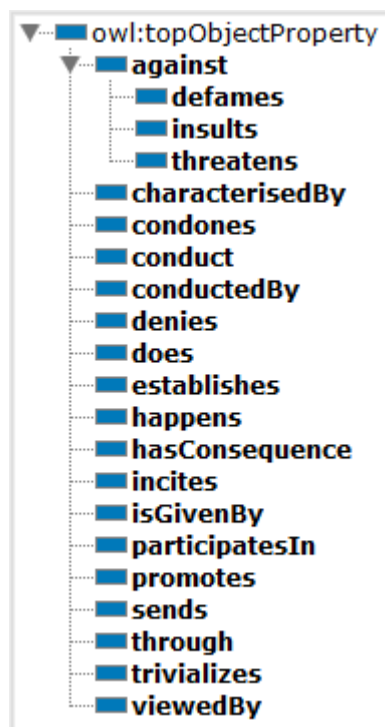
Σχετικά, με την ιεραρχία κλάσεων που όρισα πιστεύω ικανοποιεί τις ακόλουθες απαιτήσεις:

- Μία σχέση “is a”: κάθε φορά που όριζα μια κλάση A να είναι υποκλάση μιας άλλης κλάσης B, είχα υπόψιν μου ότι όλα τα στιγμιότυπα που ανήκουν στην κλάση A, πρέπει να ανήκουν υποχρεωτικά και στην κλάση B (αφού η B είναι υπερκλάση της A).
- Μια έννοια-κλάση που είναι διατυπωμένη στον ενικό δεν μπορεί να είναι υποκλάση της ίδιας έννοιας η οποία είναι στον πληθυντικό αριθμό: αυτό όπως μπορείτε να παρατηρήσετε δεν συμβαίνει στην ιεραρχία μου. Για παράδειγμα οι κλάσεις group και person μπορεί να αναπαριστούν την ίδια έννοια περίπου με την διαφορά ότι το group απεικονίζει πολλά άτομα που ανήκουν στην ίδια θρησκεία κ.α., ενώ η κλάση person απεικονίζει μεμονωμένα ένα άτομο. Παρόλα αυτά η κλάση Person δεν είναι υποκλάση του Group. Επομένως ικανοποιείται και αυτός ο περιορισμός.
- Transitivity of the hierarchical relation: ικανοποιεί και αυτόν τον περιορισμό η ιεραρχία μου καθώς τα στιγμιότυπα υποκλάσεων που δεν είναι απευθείας παιδιά πιο γενικευμένων υπερκλάσεων τους, θεωρούνται επίσης μέλη της συγκεκριμένης υπέρκλασης. Για παράδειγμα η κλάση Insulting Religion είναι υποκλάση τόσο της κλάσης Potentially Illegal Conduct Common (απευθείας υπερκλάσης-γονέα) όσο και της κλάσης Conduct που είναι η πιο γενική υπερκλάση. Θα ήταν παράλειψη μου να μην αναφέρω ότι όλες οι κλάσεις που έχω δημιουργήσει είναι υποκλάσεις της κλάσης owl: Thing (ισχύει πάντοτε, καθώς αποτελεί προεπιλογή του Protégé).
- Νομίζω, μπορεί εύκολα να συντηρηθεί και να επεκταθεί η ιεραρχία χωρίς να υπάρχουν συγκρούσεις με τους υπάρχον ορισμούς καθώς αυτό αποτελεί ένα από τα χαρακτηριστικά του RDF μοντέλου. Μπορεί πολύ απλά να προστεθούν νέες κλάσεις ή νέες ιδιότητες κλάσεων.
- Οι όροι που χρησιμοποίησα αναπαριστούν μια συγκεκριμένη έννοια και εκφράζουν είτε κάτι πολύ διαφορετικό είτε κάτι πολύ κοντινό, όμως σε καμία περίπτωση δεν εκφράζουν την ίδια έννοια.
- Δεν υπάρχουν κύκλοι στην ιεραρχία που έχω δημιουργήσει. Όταν λέω κύκλο εννοώ ότι δεν υπάρχουν ισοδυναμίες μεταξύ των κλάσεων.

Αφού ικανοποιούνται οι πιο πάνω απαιτήσεις, μπορώ να θεωρήσω ότι είναι ορθή η ιεραρχία κλάσεών μου. Όμως σύμφωνα με τη μεθοδολογία του Stanford η οντολογία μου δεν είναι καλά δομημένη, γιατί δεν ικανοποιεί τον κανόνα των 2 με 12 απευθείας υποκλάσεων που μπορεί να έχει η κάθε κλάση.

5.6 Βήμα 5 – καθορισμός ιδιοτήτων των κλάσεων

Στην παρούσα οντολογία έχω ορίσει μόνο μια συγκεκριμένη κατηγορία ιδιοτήτων. Καθόρισα μόνο πως συσχετίζονται τα στιγμιότυπα μιας κλάσης με τα στιγμιότυπα μιας άλλης κλάσης. Πιο κάτω σας δείχνω τα object properties που έχω δηλώσει.



5.7 Βήμα 6 – καθορισμός περιορισμών των ιδιοτήτων

Ο μοναδικός περιορισμός ιδιοτήτων που έχω ορίσει στην συγκεκριμένη οντολογία είναι domain και range για το κάθε slot (στην προκειμένη περίπτωση για το κάθε object property). Στον πιο κάτω πίνακα μπορείτε να δείτε το domain και range που έχει το κάθε slot της οντολογίας μου

Object Property	Domain	Range
against	Conduct	Group or Person
defames	Conduct	Group or Person
insults	Conduct	Group or Person
threatens	Conduct	Person

characterisedBy	Conduct	Characteristic
condones	Conduct	Event
conduct	Organization or Person	Conduct
conductedBy	Conduct	Organization or Person
denies	Conduct	Event
does	Conduct	Behavior
establishes	Person	Organization
happens	Conduct	Action
hasConsequence	Conduct	Consequence
incites	Conduct	Hateful behavior
isGivenBy	Consent	Person
participatesIn	Person	Organization
promotes	Conduct	Hateful behavior
sends	Conduct	Content
through	Conduct	Medium
trivializes	Conduct	Event
viewedBy	Conduct	Audience

5.8 Βήμα 7 – Δημιουργία στιγμιότυπων

Έχω δημιουργήσει μερικά στιγμιότυπα και τα έχω αναθέσει στις κλάσεις που ανήκουν. Τα στιγμιότυπα αυτά χρησιμεύουν για να ορίσουν τις προ συνθήκες που πρέπει να ικανοποιούνται σε μια κλάση. Αν ένα στιγμιότυπο τηρεί όλες τις προ συνθήκες μιας κλάσης τότε αυτόματα γίνεται μέλος αυτής της κλάσης. Παρακάτω σας δείχνω τα στιγμιότυπα που έχω ορίσει:

Individual	Types / Classes
Annoyance	EffectsForPerson
CommitCrime	IllegalAction
CommitOffense	IllegalAction

ComputerSystem	Medium
Discrimination	HatefulBehavior
ExposedToContempt	Exposed
ExposedToHatred	Exposed
ExposedToRidicule	Exposed
FacebookUsers	ComputerAudience , PublicAudience
Harassment	EffectsForPerson , HatefulBehavior
Hatred	HatefulBehavior
Indecent	NegativeCharacteristic
Intolerance	HatefulBehavior
Islam	Religion
Menacing	NegativeCharacteristic
NeedlessAnxiety	EffectsForPerson
NewspaperReaders	PublicAudience
Obscene	NegativeCharacteristic
Offensive	NegativeCharacteristic
TwitterUsers	ComputerAudience , PublicAudience
Violence	HatefulBehavior
YoutubeUsers	ComputerAudience , PublicAudience

Μπορούν να επεκτείνουν την οντολογία με την προσθήκη νέων στιγμιότυπων τόσο τα άτομα που χρησιμοποιούν τη διεπαφή όσο και τα άτομα που θα συντηρήσουν την οντολογία.

Κεφάλαιο 6

Ανάλυση Οντολογίας Hate Speech

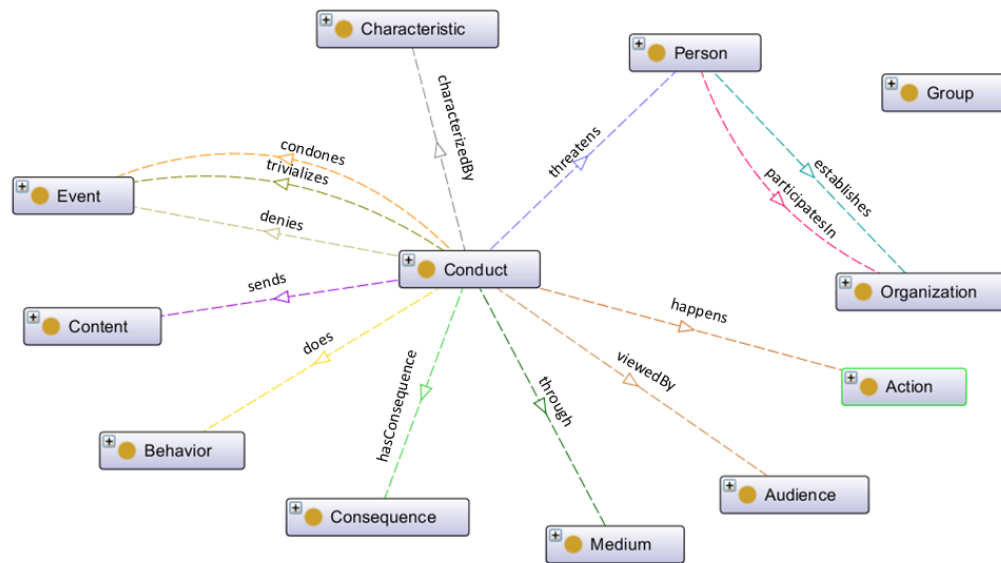
6.1 Κλάσεις - Υποκλάσεις

34

Για την ανάπτυξη της οντολογίας Hate Speech χρησιμοποίησα το εργαλείο Protégé. Η οντολογία αποτελείται από συνολικά 68 κλάσεις, μερικές από αυτές είναι υποκλάσεις άλλων κλάσεων. Επιπλέον υπάρχουν οι ιδιότητες αντικειμένων (object property) οι οποίες περιγράφουν τον τρόπο που σχετίζονται οι κλάσεις μεταξύ τους. Επιπρόσθετα, υπάρχουν δύο κύριοι λόγοι που με προώθησαν να δημιουργήσω στιγμιότυπα. Καταρχήν, για να περιγράψω τις κλάσεις μου, δηλαδή τι πρέπει να ισχύει προκειμένου ένα στιγμιότυπο να είναι μέλος μιας συγκεκριμένης κλάσης. Ο δεύτερος λόγος είναι για να επικυρώσω την ορθότητα της οντολογίας με μερικά παραδείγματα.

6.1 Κλάσεις - Υποκλάσεις

Η οντολογία αποτελείται από τις ακόλουθες κύριες κλάσεις: Action, Audience, Behavior, Characteristic, Conduct, Consequence, Content, Event, Group, Medium, Organization και Person. Ακολουθεί μια σύνοψη για τις προαναφερθέντες κλάσεις.



Σχήμα 6-1 Οι πιο βασικές κλάσεις της οντολογίας μου και το πώς συνδέονται μεταξύ τους. Απαλείφονται οι υποκλάσεις τους.

Conduct: Η κυριότερη κλάση της οντολογίας, με την οποία ορίζονται πιθανές παράνομες εκφράσεις της ρητορικής μίσους.

Person: Εκφράζει τα θύματα και τους θύτες της ρητορικής μίσους.

Organization: Αναπαριστά οργανισμούς που προωθούν και υποκινούν την βία ή την μισαλλοδοξία.

Group: Αναπαριστά μία ομάδα ατόμων που προσδιορίζεται από κάποιο χαρακτηριστικό και δέχεται μίσος για αυτό.

Content: Εκφράζει κάποιο υλικό που μπορεί να περιέχει μισαλλόδοξο περιεχόμενο.

Characteristic: Εκφράζει κατά πόσο έχει εχθρικό, απειλητικό ή προσβλητικό τόνο ένα υλικό.

Audience: Εκφράζει το κοινό που παρακολουθεί μια συγκεκριμένη συμπεριφορά.

Medium: Ορίζει το μέσο μετάδοσης του υλικού.

Consequence: Εκφράζει τις συνέπειες που έχει ένα συγκεκριμένο Conduct σε κοινωνικό ή ατομικό επίπεδο.

Event: Ανήκουν όλα τα δυσάρεστα γεγονότα που συνέβησαν (όπως γενοκτονίες, εγκλήματα πολέμου) τα οποία κάποιος μπορεί να τα υποτιμά, να αρνείται ή να παραβλέπει.

Behavior: εκφράζει κάποια συμπεριφορά.

Action: εκφράζει μια στιγμιαία εγκληματική ενέργεια ή μια ενέργεια που γίνεται με ή χωρίς την συγκατάθεση ενός ατόμου.

Στα επόμενα υπό-κεφάλαια θα αναφερθώ αναλυτικά για την κάθε κλάση ξεχωριστά καθώς επίσης και το πώς συνδέονται μεταξύ τους.

6.1.1 Κλάση Action

Η κλάση Action αναπαριστά μια άμεση ενέργεια ενός ατόμου. Στην κλάση αυτή υπάρχουν τρεις υποκλάσεις: 1) Consent, 2)Illegal Action και 3)Without Consent. Στην κλάση αυτή απεικονίζονται οι ενέργειες που δόθηκαν είτε με την συγκατάθεση κάποιου ατόμου, είτε χωρίς την συγκατάθεση του. Επιπλέον περιγράφονται και οι ενέργειες οι οποίες θεωρούνται παράνομες. Η κλάση Consent είναι disjoint με την κλάση Without Consent, δηλαδή τα στιγμιότυπα της κλάσης Consent δεν μπορούν να είναι και μέλη της κλάσεως Without Consent.



Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή και στις υποκλάσεις της είναι η

happens η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Action. Αυτή η σχέση είναι απαραίτητη για να οριστούν οι ενέργειες που έγιναν από μια συμπεριφορά (πχ αν διάπραξαν κάποιο έγκλημα/αδίκημα).

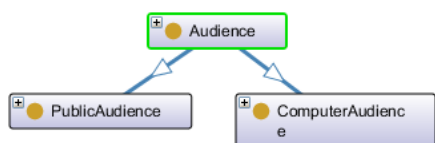
Μια επιπρόσθετη ιδιότητα αντικειμένων που αντιστοιχεί μόνο στην υποκλάση Consent είναι η ιδιότητα που ακολουθεί:

isGivenBy η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Consent με μερικά στιγμιότυπα της κλάσης Person. Με αυτή την σχέση ορίζονται όλες οι ενέργειες που έδωσε την συγκατάθεση του ένα συγκεκριμένο άτομο. Για παράδειγμα ConsentA isGivenBy PersonA, όπου ConsentA είναι στιγμιότυπο της κλάσης Consent και το PersonA είναι στιγμιότυπο της κλάσης Person.

6.1.2 Κλάση Audience

Η κλάση Audience αναπαριστά το κοινό/ τους θεατές που είδαν την συγκεκριμένη συμπεριφορά. Στην κλάση αυτή υπάρχουν δύο υποκατηγορίες θεατών, αυτών που ενημερώνονται μέσω του ηλεκτρονικών υπολογιστών (Computer Audience) και αυτών που ενημερώνονται από άλλα μέσα επικοινωνίας και όχι κατά ανάγκη από τον ηλεκτρονικό υπολογιστή (Public Audience).

Ένα στιγμιότυπο μπορεί να ανήκει και στις δύο υποκλάσεις, όπως στην περίπτωση χρηστών μέσων κοινωνικής δικτύωσης (facebook, twitter, youtube κα.) όπου γίνεται χρήση υπολογιστή και το περιεχόμενο είναι δημόσια ορατό.

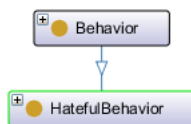


Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή και στις υποκλάσεις της είναι η **viewedBy** η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά

στιγμιότυπα της κλάσης Audience. Με αυτή την σχέση ορίζεται ποιοι βλέπουν αυτό το Conduct. Για παράδειγμα ConductA viewedBy AudienceA.

6.1.3 Κλάση Behavior

Η κλάση Behavior αναπαριστά ένα σύνολο ενεργειών που διεξάγονται κατά τη διάρκεια ενός σχετικά μεγάλου χρονικού πλαισίου, σε αντίθεση με μια ενέργεια που διαρκεί στιγμιαία. Για παράδειγμα η διάκριση, η βία, η παρενόχληση, το μίσος και η μισαλλοδοξία θεωρώ ότι έχουν αυτή την ιδιότητα. Στην κλάση αυτή υπάρχει μόνο μια υποκλάση η Hateful Behavior όπου καταγράφει τις τοξικές συμπεριφορές έναντι ενός ατόμου ή μια συγκεκριμένης ομάδας ατόμων.



Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή και στις υποκλάσεις της είναι η **does** η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα της κλάσης Behavior. Με αυτή την σχέση μπορούμε να ορίσουμε την τοξική συμπεριφορά (Hateful Behavior) που σχετίζεται με μία πιθανώς παράνομη συμπεριφορά (Conduct) έναντι κάποιας ομάδας ανθρώπων ή ατόμων.

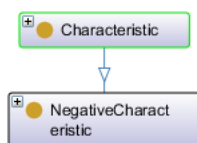
Οι επιπρόσθετες ιδιότητες αντικειμένων που αντιστοιχούν μόνο στην υποκλάση Hateful Behavior είναι οι εξής:

incites η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Hateful Behavior. Αυτή η σχέση είναι απαραίτητη για να οριστούν οι συμπεριφορές που υποκινούνται εναντίον μιας συγκεκριμένης κατηγορίας ανθρώπων. Για παράδειγμα όταν κάποιος υποκινεί βία με τη συμπεριφορά του (ConductA incites Violence).

promotes μία παρόμοια σχέση με την προηγούμενη (incites) αλλά με διαφορετική σημασία. Αντί να υποκινεί κάποιος μια τοξική συμπεριφορά, την προωθεί.

6.1.4 Κλάση Characteristic

Η κλάση Characteristic χρησιμοποιείται για να προσδιορίσει τις έννοιες οι οποίες χαρακτηρίζουν μια συμπεριφορά που έγινε με βάση το νομοθετικό πλαίσιο (Conduct). Στην προκειμένη περίπτωση χαρακτηρίζεται το περιεχόμενο ενός κειμένου ως προς τον τόνο του (πχ «αυτό το κείμενο είναι προσβλητικό»). Σε αυτή την κλάση υπάρχει μόνο μια υποκλάση η Negative Characteristic στην οποία φυλάγονται όλα τα δηλωμένα αρνητικά χαρακτηριστικά.



Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή και στην υποκλάση της είναι η **characterisedBy** η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα της κλάση Characteristic.

6.1.5 Κλάση Conduct

Η κλάση Conduct αναπαριστά συμπεριφορές ατόμων που παρουσιάζουν “ενδιαφέρον” με βάση κάποιο νομικό πλαίσιο. Δηλαδή σε αντίθεση με την κλάση Action ή Behavior, μιλάμε για συμπεριφορές που μπορεί να θεωρηθούν παράνομες, όπως όταν ένα άτομο προωθεί την βία έναντι μιας συγκεκριμένης ομάδας που προσδιορίζεται από ένα ιδιαίτερο χαρακτηριστικό (π.χ χρώμα του δέρματος). Σε αυτή την κλάση υπάρχει η υποκλάση Potentially Illegal Conduct Common στην οποία ορίζονται οι συμπεριφορές που είναι πιθανόν παράνομες. Η Potentially Illegal Conduct Common χωρίζεται στις εξής υποκατηγορίες:

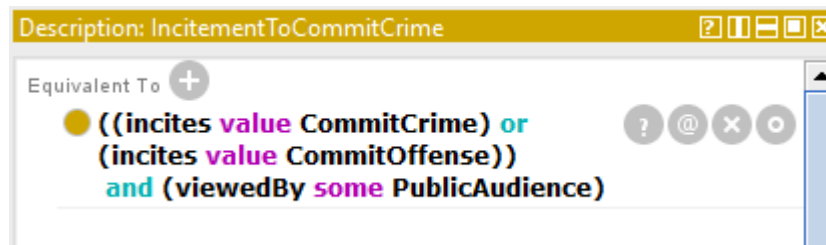
Discrimination Harassment

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να γίνονται διακρίσεις ή παρενοχλήσεις σε άλλα άτομα εξαιτίας της εθνικότητας ή της ράτσας που ανήκουν.



Incitement To Commit Crime

Για να είναι ένα στιγμιότυπο μέλος της κλάσης Incitement To Commit Crime είναι απαραίτητο να υποκινείς κάποιον δημόσια να διαπράξει ένα έγκλημα ή ένα αδίκημα.



Inciting Organization Establishment And Participation

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να γίνεται από ένα άτομο το οποίο έχει ιδρύσει έναν οργανισμό που με τη δράση του ενεργοποιείται η κλάση Publicly Inciting (αυτή η κλάση θα αναλυθεί αργότερα πότε ενεργοποιείται). Όμως υπάρχει άλλη μία περίπτωση που ένα στιγμιότυπο θα άνηκε σε αυτή την κλάση αρκεί να γίνεται από ένα άτομο το οποίο συμμετέχει σε έναν οργανισμό που με τη δράση του ενεργοποιείται η κλάση Publicly Inciting.

Description: IncitingOrganizationEstablishmentAndParticipation

Equivalent To 

- **conductedBy some**
 (Person
 and (establishes some
 (Organization
 and (conduct some PubliclyInciting))))
- **conductedBy some**
 (Person
 and (participatesIn some
 (Organization
 and (conduct some PubliclyInciting))))

Insulting Religion

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να προσβάλλεται μια θρησκεία με αυτήν την συμπεριφορά.

Description: InsultingReligion

Equivalent To 

- **insults some Religion**

Public Defamation

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να δυσφημείται δημόσια ένα συγκεκριμένο άτομο ή μια ομάδα ατόμων εξαιτίας ενός χαρακτηριστικού που τους προσδιορίζει (πχ εθνικότητα).

Description: PublicDefamation

Equivalent To 

- **(defames some**
 (Group or Person))
 and (viewedBy some PublicAudience)

Publicly Condoning And Trivializing

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να ικανοποιούνται οι εξής προϋποθέσεις: να αρνούνται ή να παραβλέπουν ή ακόμα και να υποτιμούν εγκλήματα/ γενοκτονίες που συνέβησαν κατά της ανθρωπότητας τόσο εν καιρώ ειρήνης όσο και εν καιρώ πολέμου.

Description: PubliclyCondoningAndTrivializing

Equivalent To

- (against some (Descent or Ethnicity or Nationality or Race or Religion or SkinColor)) and (denies some (CrimeAgainstHumanity or CrimeAgainstPeace or Genocide or WarCrime))
- (against some (Descent or Ethnicity or Nationality or Race or Religion or SkinColor)) and (condones some (CrimeAgainstHumanity or CrimeAgainstPeace or Genocide or WarCrime))
- (against some (Descent or Ethnicity or Nationality or Race or Religion or SkinColor)) and (trivializes some (CrimeAgainstHumanity or CrimeAgainstPeace or Genocide or WarCrime))

Publicly Inciting

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο αυτή η συμπεριφορά να έχει ως αποτέλεσμα τη διαταραχή της δημόσιας τάξης / ειρήνης και η καθώς επίσης και να υποκινεί μια εχθρική συμπεριφορά. Μια άλλη περίπτωση όπου ένα στιγμιότυπο γίνεται μέλος αυτής της κλάσης είναι όταν υποκινεί δημόσια μια εχθρική συμπεριφορά εναντίον ενός ατόμου ή μιας ομάδας εξαιτίας κάποιων από τα ακόλουθα χαρακτηριστικά που τους προσδιορίζουν, όπως εθνικότητας, φύλου, ράτσας, θρησκείας, σεξουαλικού προσανατολισμού ή να έχουν κάποια μορφή αναπηρίας.

Description: PubliclyInciting

Equivalent To

- (hasConsequence some (PublicOrderDisruption or PublicPeaceDisruption)) and (incites some HatefulBehavior)
- (against some (Disability or Ethnicity or Gender or Nationality or Race or Religion or SexualOrientation)) and (incites some HatefulBehavior) and (viewedBy some PublicAudience)

Publicly Insulting

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο με αυτήν την συμπεριφορά να προσβάλλει δημόσια κάποιο άτομο ή μια ομάδα ατόμων εξαιτίας είτε της

εθνικότητας, είτε της ράτσας ή ακόμη λόγω της θρησκείας τους. Μια άλλη περίπτωση που ένα στιγμιότυπο γίνεται μέλος αυτής της κλάσης είναι όταν προσβάλλεται δημόσια ένα συγκεκριμένο άτομο ή μια ομάδα ατόμων, αφήνοντας τους έτσι εκτεθειμένους στην γελοιοποίηση, περιφρόνηση ή ακόμη στην έχθρα εξαιτίας ενός οποιουδήποτε χαρακτηριστικού που έχουν.

Description: PubliclyInsulting

Equivalent To 

- (against some (Ethnicity or Race or Religion))
and (insults some (Group or Person))
and (viewedBy some PublicAudience)
- (against some Group)
and (hasConsequence some Exposed)
and (insults some (Group or Person))
and (viewedBy some PublicAudience)

Publisizing Toxic Material

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο αυτή η συμπεριφορά να γίνεται μέσω ενός υπολογιστικού συστήματος και να υποκινεί / προωθεί μια εχθρική συμπεριφορά εναντίον ατόμων με ειδικές ανάγκες ή λόγω της εθνικότητας, φύλου, ράτσας, θρησκείας ή σεξουαλικού προσανατολισμού που έχουν.

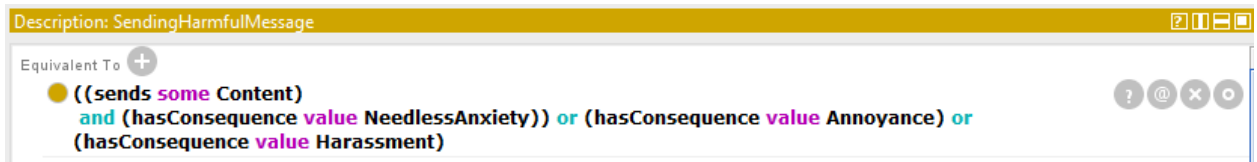
Description: PublisizingToxicMaterial   

Equivalent To 

- ((incites some HatefulBehavior) or (promotes some HatefulBehavior))
and (against some (Disability or Ethnicity or Gender or Nationality or Race or Religion or SexualOrientation))
and (sends some Content)
and (viewedBy some ComputerAudience)
and (viewedBy some PublicAudience)
- ((incites some HatefulBehavior) or (promotes some HatefulBehavior))
and (against some (Disability or Ethnicity or Gender or Nationality or Race or Religion or SexualOrientation))
and (sends some Content)
and (viewedBy some PublicAudience)
and (through value ComputerSystem)

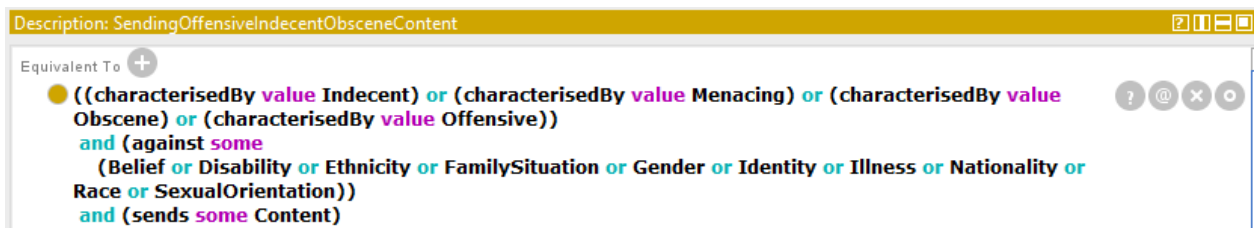
Sending Harmful Message

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να σταλθεί ένα υλικό το οποίο προκαλεί είτε αχρείαστο άγχος είτε ενόχληση.



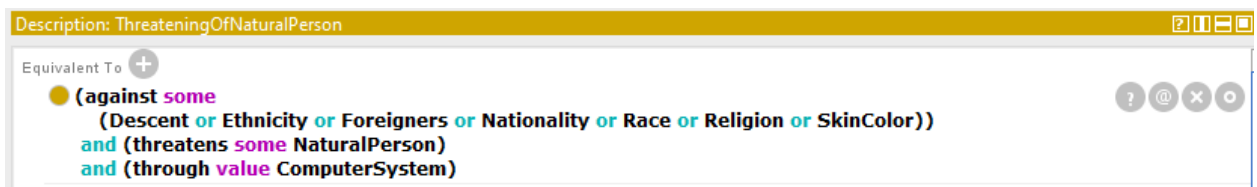
Sending Offensive Indecent Obscene Content

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να σταλεί υλικό το οποίο να έχει χαρακτηριστεί ως απρεπές ή απειλητικό ή άσεμνο ή προσβλητικό και να επιτίθεται σε άτομα βάσει των πεποιθήσεων τους, της αναπηρίας τους, της ιθαγένειας, της εθνικότητας, της οικογενειακής τους κατάστασης, του φύλου τους, της μεταδοτική τους ασθένειας, της ράτσα τους ή λόγω του σεξουαλικού τους προσανατολισμού.

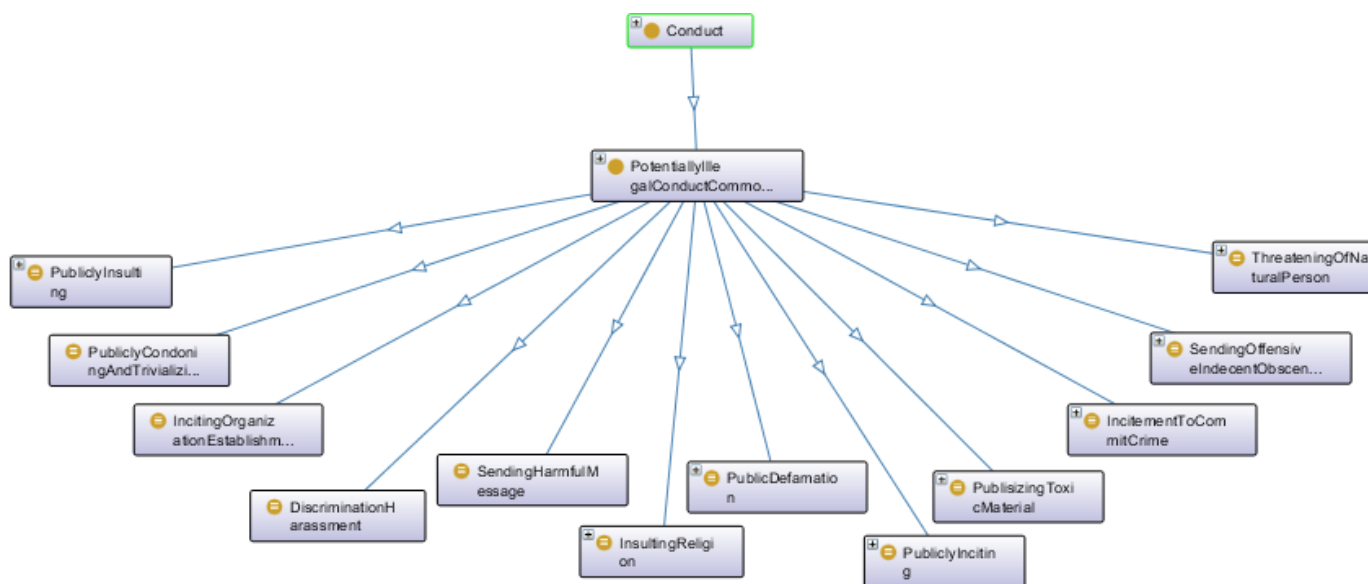


Threatening of Natural Person

Για να είναι ένα στιγμιότυπο μέλος αυτής της κλάσης είναι απαραίτητο να γίνεται μέσω ενός υπολογιστικού συστήματος και να απειλείται ένα υπαρκτό άτομο εξαιτίας των πιο κάτω χαρακτηριστικών: γενεαλογικές καταβολές, ιθαγένεια / εθνικότητα, αλλοδαπός, ράτσα, θρησκεία και το χρώμα του δέρματος.



Στην πιο κάτω εικόνα αναπαρίστανται γραφικά όλες οι υποκλάσεις της κλάσης Conduct.



Οι ιδιότητες αντικειμένων που αντιστοιχούν στην κλάση αυτή και στις υποκλάσεις της είναι οι εξής:

Against : συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Group / Person. Η συγκεκριμένη ιδιότητα αντικειμένων έχει τις εξής 3 υποιδιότητες:

- **Defames :** Η σχέση αυτή είναι απαραίτητη για να ορίσεις ποια ομάδα ατόμων δυσφημεί η συγκεκριμένη συμπεριφορά.
- **Insults :** Η σχέση αυτή είναι απαραίτητη για να ορίσεις ποια ομάδα ατόμων προσβάλλει η συγκεκριμένη συμπεριφορά.
- **Threatens :** Η σχέση αυτή είναι απαραίτητη για να ορίσεις ποια ομάδα ατόμων απειλεί η συγκεκριμένη συμπεριφορά.

Αξίζει να τονιστεί ότι έχει λόγο που τοποθέτησα αυτές τις ιδιότητες ως υποιδιότητες της against, γιατί προσθέτει επιπλέον στη σημασιολογία της οντολογίας μου. Με αυτό θέλω να πω ότι αν ένα στιγμιότυπο συσχετίζεται με ένα άλλο στιγμιότυπο μέσω των σχέσεων

defames/ insults/ threaten σημαίνει ότι το συγκεκριμένο στιγμιότυπο συσχετίζεται με το άλλο στιγμιότυπο μέσω της σχέσης against αλλά δεν σημαίνει το αντίστροφο. Η σχέση against είναι απαραίτητη για να ορίσεις σε ποια ομάδα ατόμων επιτίθεται η συγκεκριμένη συμπεριφορά.

characterisedBy : συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Characteristic. Το έχω αναλύσει προηγουμένως στην κλάση Characteristic.

Condone : συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Event. Με αυτήν την σχέση μπορείς να ορίσεις ότι το άτομο ή ένας οργανισμός παραβλέπει ένα γεγονός που έχει συμβεί.

conduct: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Organization / Person με τα στιγμιότυπα της κλάσης Conduct. Δεν πρέπει να συγχυστεί με την κλάση που έχει το ίδιο όνομα. Ο σκοπός αυτής της σχέσης είναι να οριστούν οι διάφορες συμπεριφορές που κάνουν οι οργανισμοί και οι άνθρωποι και οι οποίες παρουσιάζουν ενδιαφέρον όταν αναλυθούν μέσα στα νομοθετικά πλαίσια.

conductedBy: είναι αντίστροφη (inverse) της ιδιότητας conduct. Δηλαδή αν υπάρχει μια σχέση που συνδέει ένα στιγμιότυπο της κλάσης Organization / Person με ένα στιγμιότυπο της κλάσης Conduct τότε θα υπάρξει μια άλλη σχέση η conductedBy που θα ενώνει αυτά τα στιγμιότυπα αλλά με την αντίθετη φορά.

Denies: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Event. Με αυτήν την σχέση μπορείς να ορίσεις ότι ένα άτομο ή οργανισμός αρνείται ότι έχει συμβεί ένα γεγονός.

Does: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Behavior. Το έχω αναλύσει προηγουμένως στην κλάση Behavior.

Happens: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Action. Το έχω αναλύσει προηγουμένως στην κλάση Action.

hasConsequence: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Consequence. Με αυτήν τη σχέση μπορείς να ορίσεις ότι η συγκεκριμένη συμπεριφορά έχει κάποιες επιπτώσεις είτε σε ένα συγκεκριμένο άτομο είτε στην κοινωνία.

Incites: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης HatefulBehavior. Το έχω αναλύσει προηγουμένως στην κλάση Behavior.

Promotes: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης HatefulBehavior. Το έχω αναλύσει προηγουμένως στην κλάση Behavior.

Sends: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Content. Με αυτήν τη σχέση μπορείς να ορίσεις το είδος του υλικού που στέλνεται από κάποιο άτομο ή οργανισμό.

Through: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Medium. Με αυτήν τη σχέση μπορείς να ορίσεις μέσω ποιου μέσου πραγματοποιήθηκε η συγκεκριμένη συμπεριφορά. Δηλαδή αν στάλθηκε κάποιο υλικό μέσω ενός υπολογιστικού συστήματος τότε θα είχαμε την σχέση ConductA through ComputerSystem.

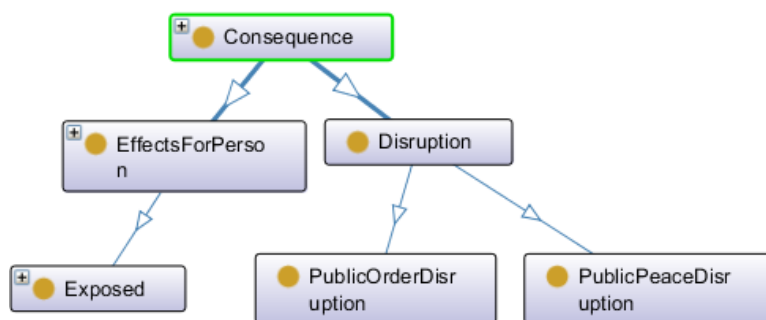
Trivializes: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Event. Με αυτήν τη σχέση μπορείς να ορίσεις ότι η συγκεκριμένη συμπεριφορά υποτιμά τη σημαντικότητα και τις διαστάσεις που είχε ένα γεγονός.

viewedBy: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με τα στιγμιότυπα της κλάσης Audience. Το έχω αναλύσει προηγουμένως στην κλάση Audience.

6.1.6 Κλάση Consequence

Η κλάση Consequence αναπαριστά τα επακόλουθα που προκλήθηκαν από συγκεκριμένες συμπεριφορές ατόμων ή οργανισμών. Τα επακόλουθα τα βλέπουμε από την οπτική γωνία ενός ανθρώπου ή της κοινωνίας γενικότερα: τι επιδράσεις είχε στο άτομο που δέχτηκε αυτή

τη συμπεριφορά ως προς την ψυχολογική του υγεία πχ. ένιωσε περιττό άγχος ή δυσφορία. Όσο αφορά τις επιδράσεις που είχε στην κοινωνία είναι να φέρει αναταραχές είτε στην δημόσια τάξη είτε στην δημόσια ειρήνη. Στην πιο κάτω εικόνα απεικονίζεται η ιεραρχία που αφορά στην κλάση Consequence. Η υποκλάση Exposed χρησιμοποιείται για να εκφράσει ότι το συγκεκριμένο άτομο που δέχτηκε αυτήν τη συμπεριφορά αφήνεται εκτεθειμένο σε έχθρα / γελοιοποίηση / περιφρόνηση.

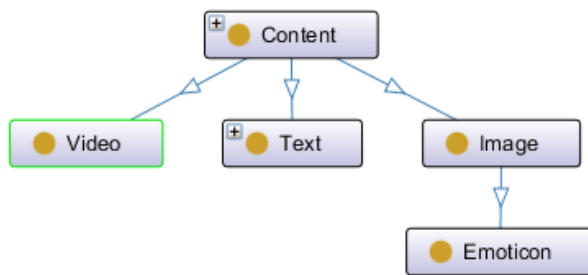


Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση Consequence και στις υποκλάσεις της είναι η ακόλουθη:

hasConsequence η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Consequence.

6.1.7 Κλάση Content

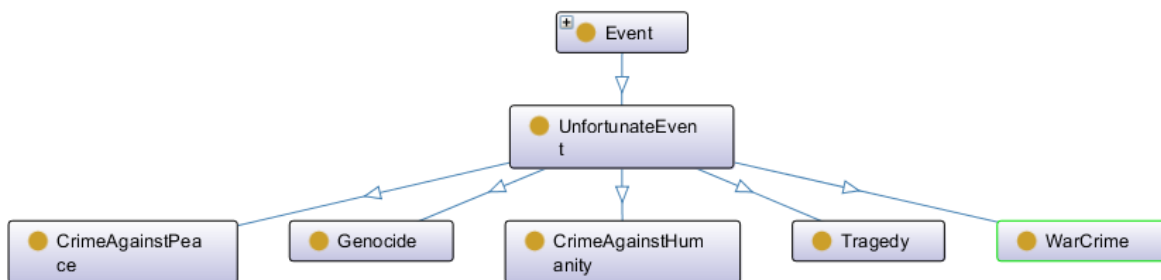
Η κλάση Content αναπαριστά συγκεκριμένα το είδος του μέσου που χρησιμοποιήθηκε για την πραγματοποίηση μιας συμπεριφοράς. Για παράδειγμα όταν κάποιος στέλνει ηλεκτρονικό μήνυμα που περιέχει ρητορική μίσους. Αυτό το μήνυμα μπορεί να είναι ένα απλό κείμενο, μία εικόνα ή ακόμη ένα βίντεο.



Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή και στις υποκλάσεις της είναι η **sends** η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Content.

6.1.8 Κλάση Event

Η κλάση Event αναπαριστά κάποιο γεγονός, που στα πλαίσια αυτής της οντολογίας έχουν δυσάρεστο τόνο. Αναφέρονται σε τραγικά περιστατικά που συνέβησαν στο παρελθόν εναντία μιας ομάδας ατόμων όπως γενοκτονίες και εγκλήματα πολέμου.



Οι ιδιότητες αντικειμένων που αντιστοιχούν στην κλάση αυτή και στις υποκλάσεις της είναι οι εξής:

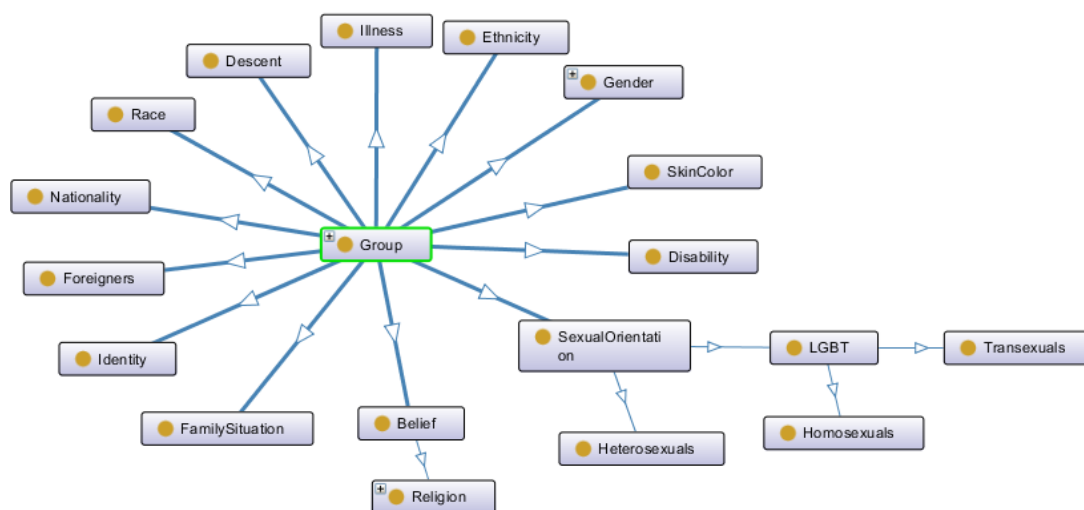
condones η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Event.

denies η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Event.

trivializes η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Event.

6.1.9 Κλάση Group

Η κλάση Group ομαδοποιεί άτομα με βάσει κάποιων χαρακτηριστικών που τα προσδιορίζουν. Αυτά τα χαρακτηριστικά ευθύνονται για τις διακρίσεις και άλλες μορφές μισαλλοδοξίας που γίνονται εναντίον αυτών των ατόμων. Πιο κάτω παρουσιάζονται οι διάφορες κατηγορίες ομάδων που έχω ορίσει.



Οι ιδιότητες αντικειμένων που αντιστοιχούν στην κλάση αυτή και στις υποκλάσεις της είναι οι εξής:

against: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Group.

defames: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Group.

insults: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Group.

6.1.10 Κλάση Medium

Η κλάση Medium αναπαριστά το μέσο από το οποίο πραγματοποιήθηκε μια συμπεριφορά. Στην παρούσα φάση το μόνο στιγμιότυπο που είναι μέλος αυτής της κλάσης είναι το υπολογιστικό σύστημα. Δεν χρειάστηκε να ορίσω υποκλάσεις για αυτήν την κλάση.

Η ιδιότητα αντικειμένων που αντιστοιχεί στην κλάση αυτή είναι η thought η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Medium. Το έχω αναλύσει προηγουμένως στην κλάση Conduct.

6.1.11 Κλάση Organization

Αυτή η κλάση αναπαριστά έναν οργανισμό στον πραγματικό κόσμο. Εμάς συγκεκριμένα μας ενδιαφέρει να γνωρίζουμε αν ο οργανισμός αυτός προωθεί την βία και το μίσος έναντι συγκεκριμένων ομάδων. Επίσης μας ενδιαφέρει εάν ένα άτομο έχει ιδρύσει ή λαμβάνει μέρος σε έναν από αυτούς τους οργανισμούς. Δεν χρειάστηκε να ορίσω υποκλάσεις για αυτήν την κλάση.

Οι ιδιότητες αντικειμένων που αντιστοιχούν στην κλάση αυτή και στις υποκλάσεις της είναι οι εξής:

Establishes: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Person με μερικά στιγμιότυπα που ανήκουν στην κλάση Organization. Αυτή η σχέση είναι απαραίτητη για να εκφράσει αν ένα άτομο έχει ιδρύσει έναν οργανισμό.

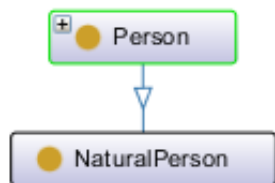
participatesIn: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Person με μερικά στιγμιότυπα που ανήκουν στην κλάση Organization. Αυτή η σχέση είναι απαραίτητη για να εκφράσει αν ένα άτομο συμμετέχει σε έναν οργανισμό.

Conduct: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Organization με τα στιγμιότυπα της κλάσης Conduct.

conductedBy: Το έχω αναλύσει προηγουμένως στην κλάση Conduct.

6.1.12 Κλάση Person

Αυτή η κλάση αναπαριστά ένα άτομο στην οντολογία και χρησιμοποιείται κυρίως για να αποτυπώσω κανόνες που σχετίζονται με επιθέσεις ενάντια ατόμων μειονοτήτων ή άλλων προστατευμένων ομάδων.



Οι ιδιότητες αντικειμένων που αντιστοιχούν στην κλάση αυτή και στις υποκλάσεις της είναι οι εξής:

Establishes: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Person με μερικά στιγμιότυπα που ανήκουν στην κλάση Organization. Αυτή η σχέση είναι απαραίτητη για να εκφράσει αν ένα άτομο έχει ιδρύσει έναν οργανισμό.

participatesIn: η οποία συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Person με μερικά στιγμιότυπα που ανήκουν στην κλάση Organization. Αυτή η σχέση είναι απαραίτητη για να εκφράσει αν ένα άτομο συμμετέχει σε έναν οργανισμό.

Conduct: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Person με τα στιγμιότυπα της κλάσης Conduct.

conductedBy: Το έχω αναλύσει προηγουμένως στην κλάση Conduct.

against: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Person.

defames: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Person.

insults: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Person.

threatens: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Conduct με μερικά στιγμιότυπα που ανήκουν στην κλάση Person.

isGivenBy: συνδέει τα στιγμιότυπα που ανήκουν στην κλάση Consent με μερικά στιγμιότυπα της κλάσης Person. Το έχω αναλύσει προηγουμένως στην κλάση Action.

Κεφάλαιο 7

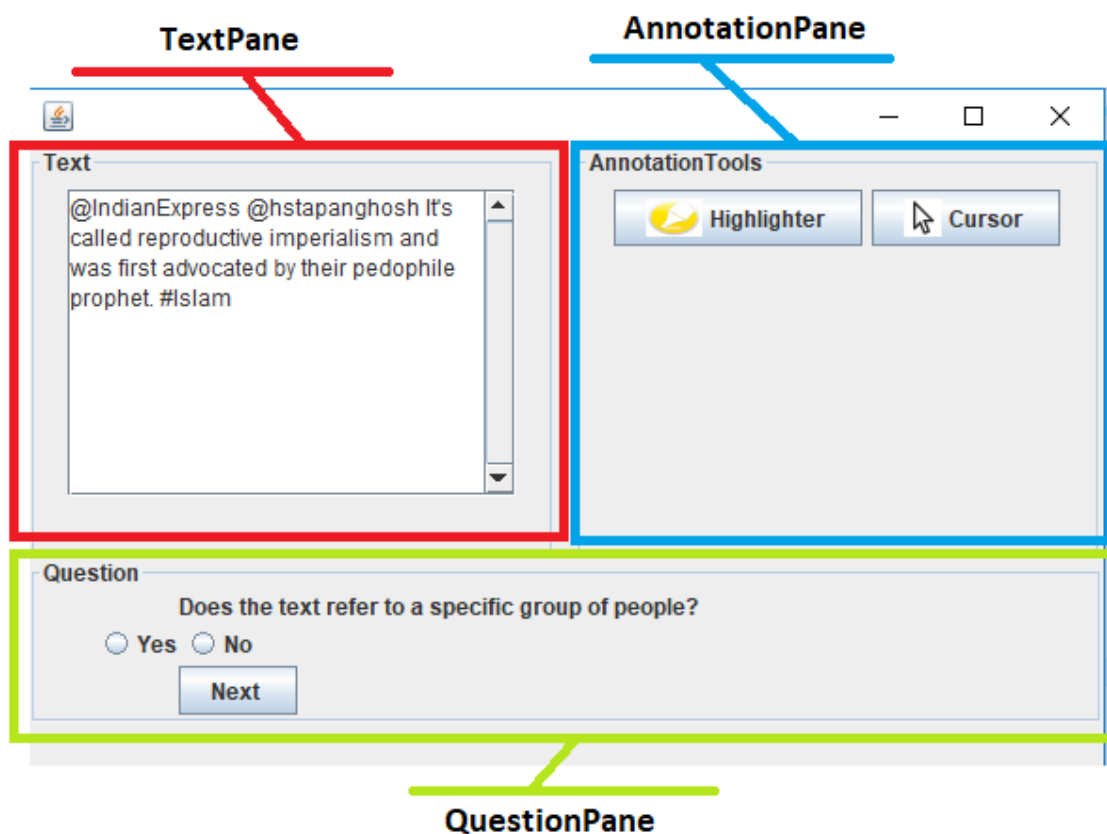
Ανάπτυξη διεπαφής

7.1 Εισαγωγή	54
7.2 Κύρια στοιχεία διεπαφής	55
7.3 Γραφική απεικόνιση των ερωτήσεων και επεξήγηση χρησιμότητας τους	55

7.1 Εισαγωγή

Στην προηγούμενη φάση ανέπτυξα μια οντολογία με βάσει κάποιες υφιστάμενες νομοθεσίες που αφορούν την ρητορική μίσους. Σκοπός μου δεν είναι να ορίσω τι είναι παράνομο ή όχι. Σκοπός μου είναι να κατηγοριοποιήσω τέτοιες συμπεριφορές βοηθώντας έτσι απλούς χρήστες να επεξηγήσουν καλύτερα γιατί το συγκεκριμένο κείμενο το θεωρούν ως hate speech. Θεώρησα ότι μια οντολογία από μόνη της δεν θα ήταν εύκολη στην χρήση για τους χρήστες καθώς θα τους ανάγκαζα να μάθουν πως να χρησιμοποιούν το protégé και όλες τις βασικές πληροφορίες που χρειάζονται να ξέρουν για να μπορέσουν να επεκτείνουν την οντολογία. Προκειμένου να λύσω αυτό το πρόβλημα, αποφάσισα ότι η καλύτερη επιλογή ήταν να αναπτύξω μια απλή διεπαφή για τους χρήστες οι οποίοι το μόνο που θα έχουν να κάνουν είναι να απαντήσουν μερικές ερωτήσεις κατανόησης που σχετίζονται με το κείμενο.

7.2 Κύρια στοιχεία διεπαφής



Η διεπαφή αποτελείται από τρία στοιχεία το TextPane, το questionPane και το annotationPane. Το textPane είναι ένα container που περιέχει μέσα το κείμενο. Το κείμενο είναι απαραίτητο για τους χρήστες προκειμένου να μπορούν να απαντήσουν στις ερωτήσεις που το αφορούν. Το questionPane περιέχει μια ερώτηση κάθε φορά και όταν επιλεγθεί το κουμπί “Next” εμφανίζεται η επόμενη ερώτηση. Το annotationPane δίνει την επιλογή στον χρήστη να μπορεί να επιλέξει το κείμενο και να χρωματίσει τις λέξεις-κλειδιά.

7.3 Γραφική απεικόνιση των ερωτήσεων και επεξήγηση χρησιμότητάς τους

Το πιο σημαντικό στοιχείο της διεπαφής είναι το questionPane που περιέχει τις ερωτήσεις. Παρακάτω, φαίνεται παραστατικά η λογική σειρά με την οποία εμφανίζονται οι ερωτήσεις. Όπως μπορείτε να διαπιστώσετε και από μόνοι σας, η επόμενη ερώτηση που θα εμφανιστεί στον χρήστη εξαρτάται από την απάντηση που έδωσε προηγουμένως.

οντολογία, όμως καθορίζει την επόμενη ερώτηση που θα εμφανιστεί στον χρήστη. Ο χρήστης μπορεί να απαντήσει με ναι ή όχι. Απαντώντας ναι θα εμφανιστεί η ερώτηση 2, αλλιώς θα εμφανιστεί η ερώτηση 3.

7.3.2 Ερώτηση 2

“To which group does the text refer to?”

Αφού ο χρήστης απάντησε προηγουμένως θετικά στην ερώτηση 1, καλείται να καθορίσει συγκεκριμένα σε ποια κατηγορία ομάδας ατόμων απευθύνεται το συγκεκριμένο κείμενο. Για παράδειγμα πρέπει να απαντήσει “Religion” όταν γίνεται αναφορά σε θρησκευτική ομάδα, και “Race” όταν γίνεται αναφορά σε κάποια φυλετική ομάδα. Η απάντηση της ερώτησης αυτής καθορίζει τον τύπο-κλάση του “TargetIndividual” που πρόκειται να δημιουργηθεί.

7.3.3 Ερώτηση 2b

“Which targetGroup is being targeted?”

Όπου targetGroup είναι η κατηγορία που επιλέχθηκε προηγουμένως από την ερώτηση 2. Για παράδειγμα αν προηγουμένως ο χρήστης επέλεξε την κατηγορία Race τότε με αυτήν την ερώτηση θέλουμε να μας απαντήσει συγκεκριμένα σε ποια φυλετική ομάδα επιτίθεται το κείμενο. Ο χρήστης μπορεί να απαντήσει σε αυτήν την ερώτηση επιλέγοντας μια από τις απαντήσεις στην λίστα. Αν στην προκειμένη φάση δεν υπάρχει μια απάντηση που να ικανοποιεί τον χρήστη τότε μπορεί να επιλέξει ως απάντηση το “Other”.

7.3.4 Ερώτηση 2c

“Please name the targetGroup that is being targeted?”

Αν έχει επιλέξει προηγουμένως την απάντηση “Other” τότε ο χρήστης καλείται να πληκτρολογήσει συγκεκριμένα ποια ομάδα ατόμων στοχοποιείται στο υφιστάμενο κείμενο. Αφού απαντήσει δημιουργείται ένα στιγμιότυπο στην οντολογία ανάλογο με τις απαντήσεις στον χρήστη. Το όνομα του στιγμιότυπου καθορίζεται στην ερώτηση 2b ή 2c (αν επιλέχθηκε ως απάντηση το other στην ερώτηση 2b). Η κλάση στην οποία ανήκει το στιγμιότυπο που μόλις δημιουργήθηκε καθορίζεται από την ερώτηση 2. Ακόμη προστίθεται στην οντολογία ως αξίωμα ότι το συγκεκριμένο κείμενο είναι εναντίον του

συγκεκριμένου στιγμιότυπου-ομάδας που δημιουργήθηκε σε αυτό το ερώτημα. Ακολουθώντας προχωράμε στην ερώτηση 3.

7.3.5 Ερώτηση 3

“What is the aim of this text?”

Με αυτήν την ερώτηση μαθαίνω από τον χρήστη τον σκοπό του υφιστάμενου κειμένου. Ο χρήστης μπορεί να απαντήσει επιλέγοντας ένα από τα στοιχεία της λίστας. Αν καμία από τις απαντήσεις δεν αντιστοιχεί στο συγκεκριμένο κείμενο ο χρήστης μπορεί να επιλέξει ως απάντηση: “None of the above”. Όμως με αυτήν την επιλογή δεν γίνεται καμία αλλαγή στην οντολογία. Στην περίπτωση που επιλέξει μια από τις ακόλουθες απαντήσεις η οντολογία θα αλλάξει καθώς θα προστεθούν νέα αξιώματα:

“to insult” τότε δημιουργείται η εξής σχέση στην οντολογία: tweet insults TargetGroup

“to defame” τότε δημιουργείται η εξής σχέση στην οντολογία: tweet defames TargetGroup

“to does discrimination” τότε δημιουργείται η εξής σχέση στην οντολογία: tweet does Discrimination (αφού προηγουμένως όρισα ότι το Discrimination είναι ένα στιγμιότυπο της κλάσης “HatefulBehavior”)

“to incite hate / discrimination / violence” τότε δημιουργείται η εξής σχέση στην οντολογία: tweet incites Hatred / Discrimination / Violence.

Αν έχει επιλεγεί ως απάντηση οτιδήποτε άλλο εκτός από το incite τότε η επόμενη ερώτηση είναι η 4^η διαφορετικά είναι η 7^η.

7.3.6 Ερώτηση 4

“Does this text have any consequence / impact in a society, or a specific person?”

Αυτή η ερώτηση δεν επιφέρει καμία αλλαγή στην οντολογία. Όμως μαθαίνουμε από τον χρήστη αν το συγκεκριμένο κείμενο έχει κάποιο αντίκτυπο στην κοινωνία ή στον άνθρωπο. Αν επιλέξει ο χρήστης ότι το συγκεκριμένο κείμενο επηρεάζει γενικότερα την κοινωνία τότε η επόμενη ερώτηση που θα εμφανιστεί είναι η 5^η, ενώ αν επηρεάζει τον άνθρωπο τότε θα είναι η 6^η.

7.3.7 Ερώτηση 5

“What sort of impact on society does it have?”

Με αυτήν την ερώτηση μαθαίνω αν το συγκεκριμένο κείμενο διαταράσσει τη δημόσια ειρήνη / τάξη ή όχι. Αν όντως τη διαταράζει τότε δημιουργείται το αντίστοιχο στιγμιότυπο (DisruptionOfPublicPeace / DisruptionOfPublicOrder) το οποίο συσχετίζεται με το tweet με την σχέση “hasConsequence” δηλαδή tweet hasConsequence (DisruptionOfPublicPeace / DisruptionOfPublicOrder). Ακολούθως εμφανίζεται η ερώτηση 9.

7.3.8 Ερώτηση 6

“What sort of impact does it have on the person?”

Με αυτήν την ερώτηση μαθαίνω πως το συγκεκριμένο κείμενο επηρεάζει τον άνθρωπο. Δηλαδή αν του προκαλεί ενόχληση, άγχος ή τον αφήνει εκτεθειμένο στο μίσος / περιφρόνηση / γελοιοποίηση. Ανάλογα με την απάντηση του χρήστη προστίθεται στην οντολογία ως αξίωμα ότι το tweet+ hasConsequence + την απάντηση του χρήστη. Ακολούθως εμφανίζεται η ερώτηση 9.

7.3.9 Ερώτηση 7

“Was the text publicized by a person which is part of an organization that promotes violence / hate / animosity?”

Με την παρούσα ερώτηση μαθαίνω από τον χρήστη αν το άτομο που έγραψε το συγκεκριμένο κείμενο είναι μέλος ενός οργανισμού που ο κύριος σκοπός του είναι η υποκίνηση της βίας και του μίσους. Αν απαντήσει θετικά ο χρήστης τότε γίνονται οι ακόλουθες ενέργειες στην οντολογία :

α) Δημιουργία ενός στιγμιότυπου τύπου ανθρώπου

β) Καθορισμός ότι το συγκεκριμένο στιγμιότυπο (άνθρωπος) δημοσίευσε το tweet που εμφανίζεται στη διεπαφή

Στην περίπτωση όμως που ο χρήστης απαντήσει αρνητικά τότε δεν επηρεάζεται η οντολογία και μεταβαίνει στην ερώτηση 4. Όμως στην περίπτωση που έχει απαντήσει θετικά τότε η επόμενη ερώτηση που θα εμφανιστεί είναι η 7b.

7.3.10 Ερώτηση 7b

“Which organization is attacking the group?”

Σε αυτήν την ερώτηση υπάρχει μια λίστα που θα περιέχει ονόματα οργανισμών που ευθύνονται για το συγκεκριμένο tweet. Αρχικά αυτή η λίστα περιέχει ως μόνο στοιχείο το “Other” έτσι καλείται ο χρήστης να το επιλέξει και να προχωρήσει στην ερώτηση 7c.

7.3.11 Ερώτηση 7c

“Please name the organization that is attacking the group?”

Σε αυτό το σημείο ο χρήστης πρέπει να πληκτρολογήσει το όνομα του οργανισμού που βρίσκεται πίσω από το συγκεκριμένο κείμενο. Αφότου πληκτρολογηθεί το όνομα του οργανισμού, δημιουργείται το αντίστοιχο στιγμιότυπο A στην οντολογία (ότι ο A είναι ένας οργανισμός) και ότι το στιγμιότυπο – άνθρωπος που δημιουργήθηκε προηγουμένως στην ερώτηση 7 συμμετέχει στον συγκεκριμένο οργανισμό. Η επόμενη ερώτηση που θα εμφανιστεί είναι η 4^η.

7.3.12 Ερώτηση 9

“How would describe the tone of the text?”

Αυτή είναι η τελευταία ερώτηση που πρέπει να απαντήσει ο χρήστης. Με αυτήν την ερώτηση μαθαίνω αν το συγκεκριμένο κείμενο έχει απειλητικό / προσβλητικό / απρεπές τόνο. Αν όντως έχει κάποιο από τα προαναφερθέντα (απειλητικό / προσβλητικό / απρεπές) τότε δημιουργείται μια σχέση στην οντολογία που συσχετίζει τον τόνο με το κείμενο.

Στην συνέχεια καλείται ο hermit reasoner ο οποίος αναθέτει το συγκεκριμένο κείμενο σε κάποιες υποκλάσεις του conduct ανάλογα με τις απαντήσεις που έδωσε προηγουμένως ο χρήστης. Η ανάθεση ενός στιγμιότυπου σε μια συγκεκριμένη κλάση γίνεται από τον reasoner μόνο όταν το συγκεκριμένο στιγμιότυπο ικανοποιεί όλες τις προδιαγραφές / απαιτήσεις της συγκεκριμένης κλάσης.

Κεφάλαιο 8

Αποτελέσματα

8.1 Παραδείγματα εκτέλεσης	62
8.2 Σύγκριση αποτελεσμάτων	67
8.3 Επέκταση οντολογίας	68

Προκειμένου να αξιολογήσω κατά πόσο η οντολογία και η διεπαφή μου λειτουργούν σωστά, έκανα annotate 29 tweets με την χρήση της διεπαφής. Αυτά τα tweets ήταν προηγουμένως κατηγοριοποιημένα από ειδικούς επιστήμονες οι οποίοι τα έχουν κατατάξει στις κατηγορίες offensive και hate speech. Σύμφωνα με τους ειδικούς επιστήμονες τα πρώτα 15 ήταν παραδείγματα offensive tweets ενώ τα υπόλοιπα 14 tweets ήταν παραδείγματα hate speech.

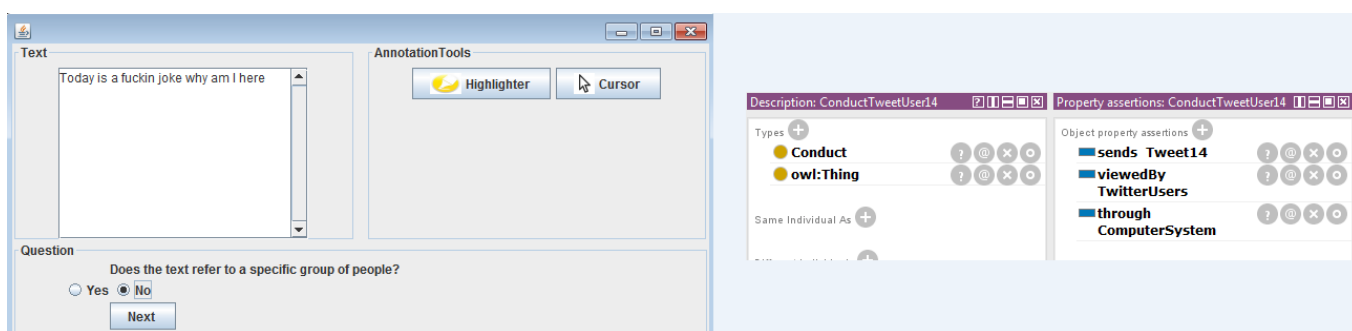
Offensive	Hate Speech
0-14 tweets	15-28 tweets

Σημείωση : Λόγω ακατάλληλου περιεχομένου των περισσότερων tweets, προτίμησα να μην τα παρουσιάσω στο γραπτό μέρος της Διπλωματικής μου. Εξαίρεση αποτελούν 2 tweets (ένα offensive, και ένα hatespeech) τα οποία χρειάζομαι για παραδείγματα χρήσης. Όμως τα tweets καθώς και τα αποτελέσματα των annotations συμπεριλαμβάνονται μέσα στα παραδοτέα της διπλωματικής μου.

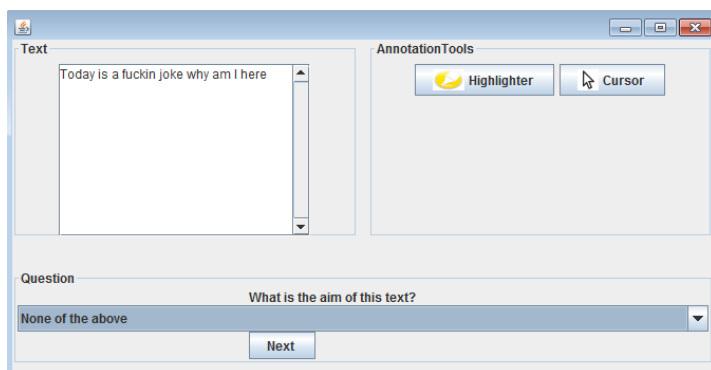
8.1 Παραδείγματα εκτέλεσης

Σε αυτήν την ενότητα θα σας δείξω οπτικά πώς επηρεάζεται η οντολογία ανάλογα με τις απαντήσεις που έδωσα. Πρώτο παράδειγμα εκτέλεσης είναι το εξής tweet “Today is a f*** joke why am I here” το οποίο είναι ένα offensive tweet.

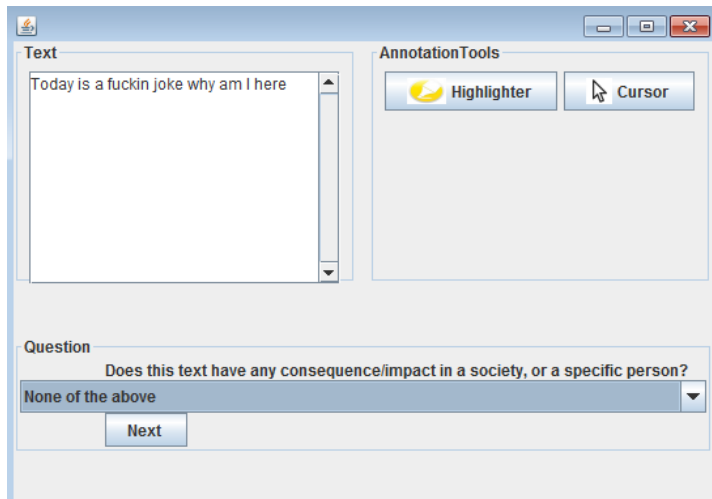
Αρχικά μου δίνεται η ερώτηση, αν το συγκεκριμένο tweet αναφέρεται σε συγκεκριμένη ομάδα ατόμων. Το κείμενο του tweet όμως δεν φαίνεται να αναφέρεται σε οποιοδήποτε άτομο, πόσο μάλλον σε μία ομάδα ατόμων. Γι’ αυτό επέλεξα την επιλογή όχι. Κάνοντας αυτό δημιουργήθηκαν μερικά νέα individuals και αξιώματα στην οντολογία.



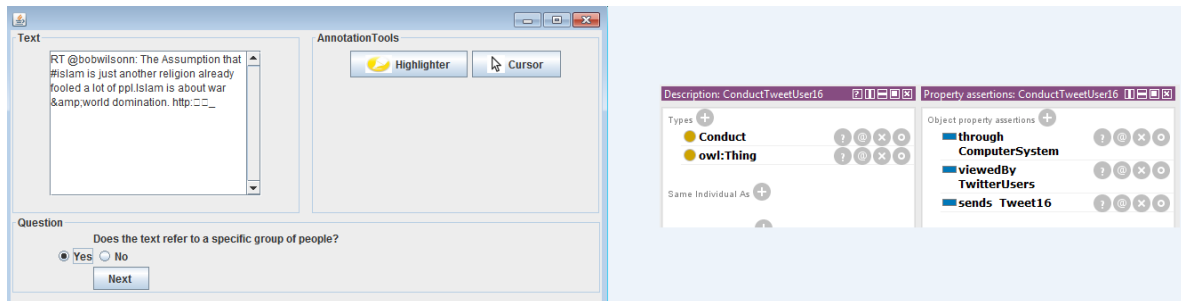
Μετά εμφανίστηκε η ερώτηση για τον στόχο του κειμένου. Το συγκεκριμένο tweet θεώρησα πως δεν είχε στόχο ούτε να προσβάλει αλλά ούτε να κάνει διάκριση. Είναι απλώς η έκφραση ενός παραπόνου. Επομένως επέλεξα την επιλογή “κανένα από τα πιο πάνω”, η οποία δεν πρόσθεσε κάτι στην οντολογία.



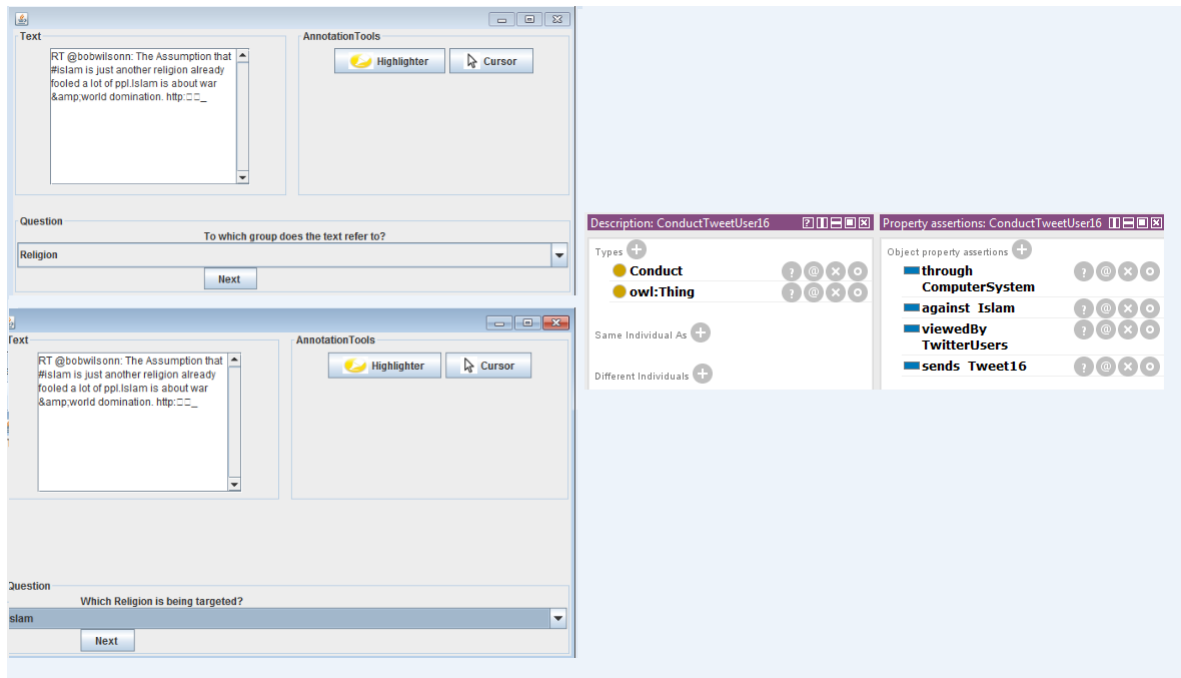
Εφόσον το κείμενο δεν έχει κάποιο φανερό αντίκτυπο στην κοινωνία ή σε κάποιο άτομο, επέλεξα και στην επόμενη ερώτηση την επιλογή “κανένα από τα πιο πάνω”.



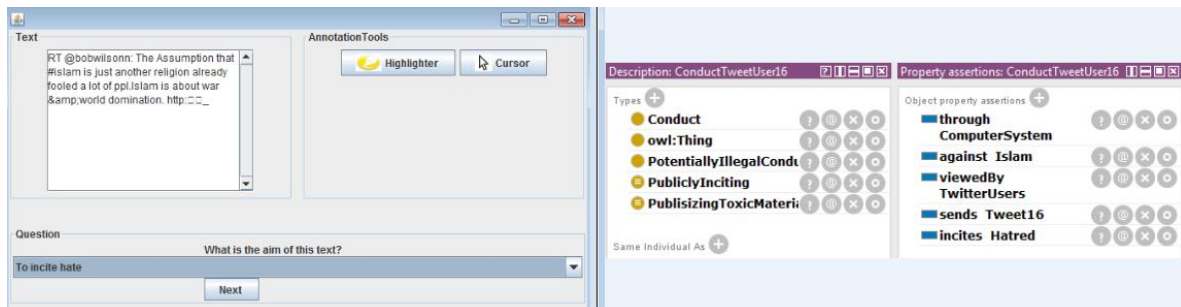
Το δεύτερο παράδειγμα εκτέλεσης είναι ένα tweet κατηγοριοποιημένο ως hate speech στο οποίο εκφράζεται η άποψη ότι το Ισλάμ είναι μια θρησκεία που έχει ως στόχο την πολιορκία του κόσμου. Συνεπώς μέσω του συγκεκριμένου tweet θίγεται μια συγκεκριμένη κοινότητα ατόμων, αυτών που πιστεύουν στο Ισλάμ.



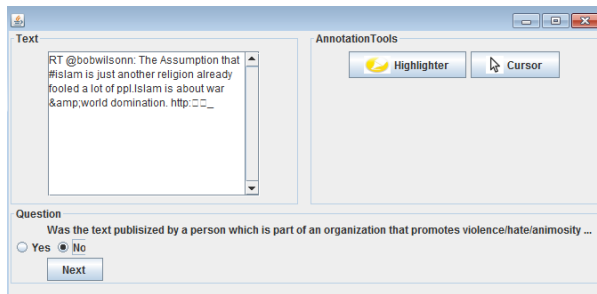
Στην δεύτερη ερώτηση απάντησα σε ποια συγκεκριμένη κοινότητα ατόμων απευθύνεται το κείμενο. Το κείμενο θίγει το Islam σαν θρησκεία και γι' αυτό επέλεξα τις ακόλουθες απαντήσεις. Παρακάτω βλέπουμε ότι οι συγκεκριμένες απαντήσεις επηρέασαν την οντολογία καθώς προστέθηκε σαν γνώση στην οντολογία ότι το συγκεκριμένο conduct είναι εναντίον του Ισλάμ.



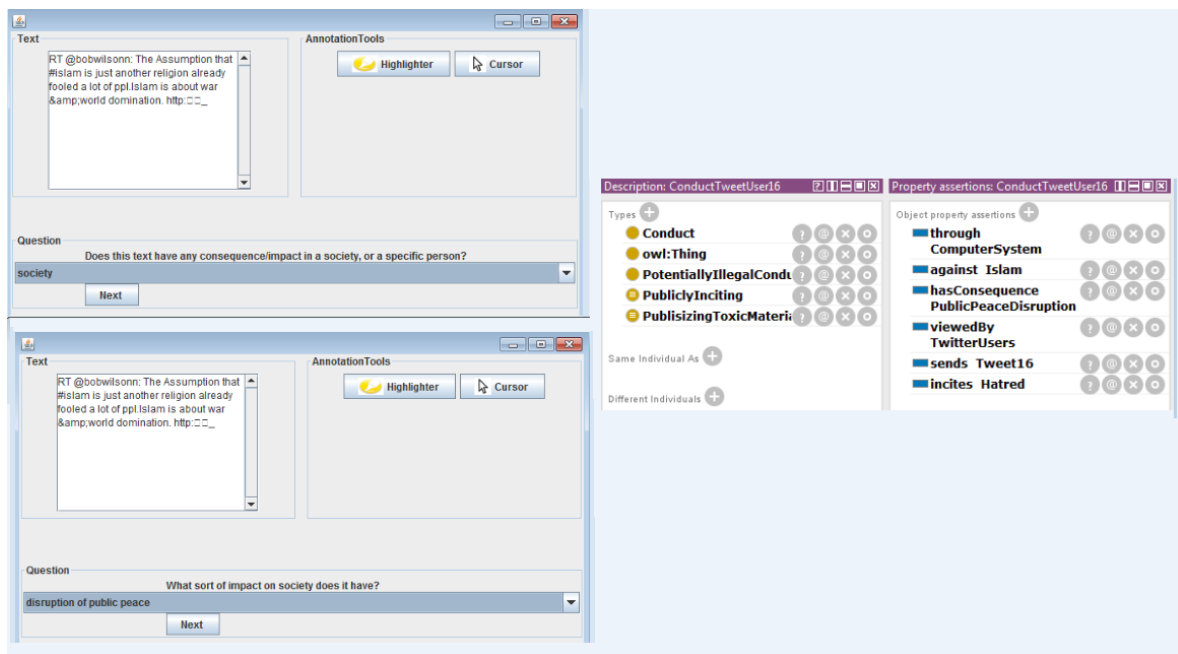
Ακολούθως, εμφανίστηκε η ερώτηση 3 που έπρεπε να απαντήσω ποιος είναι ο σκοπός του κειμένου. Επέλεξα ως απάντηση την επιλογή «προώθηση μίσους» γιατί θεώρησα ότι το άτομο που δημοσίευσε αυτό το κείμενο, είχε ως πρόθεση να κάνει τα άτομα να μισήσουν το Ισλάμ σαν θρησκεία. Όπως ήταν αναμενόμενο, η οντολογία άλλαξε με την συγκεκριμένη ενέργεια, αφού προστέθηκε το αξίωμα ότι το συγκεκριμένο conduct προωθεί το μίσος.



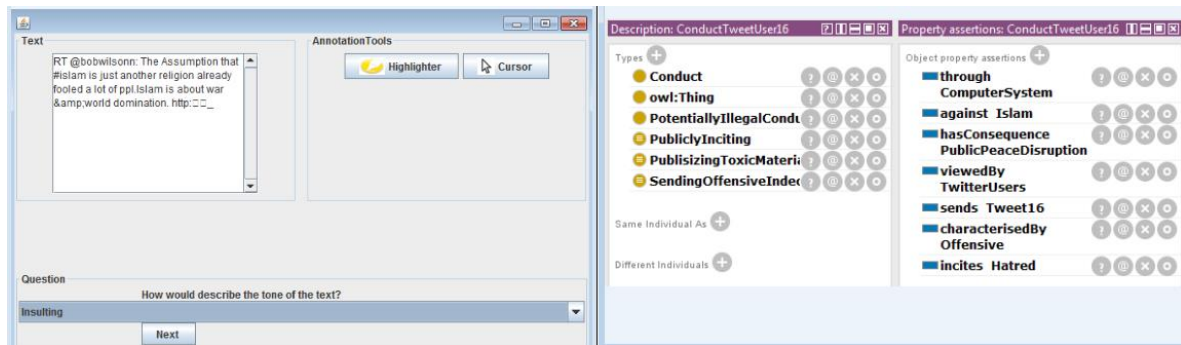
Στην συνέχεια με ρωτούσε αν το συγκεκριμένο κείμενο δημοσιεύτηκε από κάποιον ο οποίος είναι μέλος ενός οργανισμού που προωθεί τέτοιου είδους συμπεριφορές. Δεν φαίνεται να ισχύει κάτι τέτοιο για αυτό απάντησα αρνητικά. Όμως αυτή η απάντηση, δεν άλλαξε την οντολογία.



Έπειτα απάντησα ποιο αντίκτυπο προκαλεί το συγκεκριμένο κείμενο. Κατά τη γνώμη μου, το κείμενο δεν επηρεάζει μόνο ένα άτομο, αλλά ολόκληρη την κοινωνία. Πιο συγκεκριμένα διαταράσσει τη δημόσια ειρήνη. Προστέθηκε σαν γνώση το αντίστοιχο αξίωμα στην οντολογία.

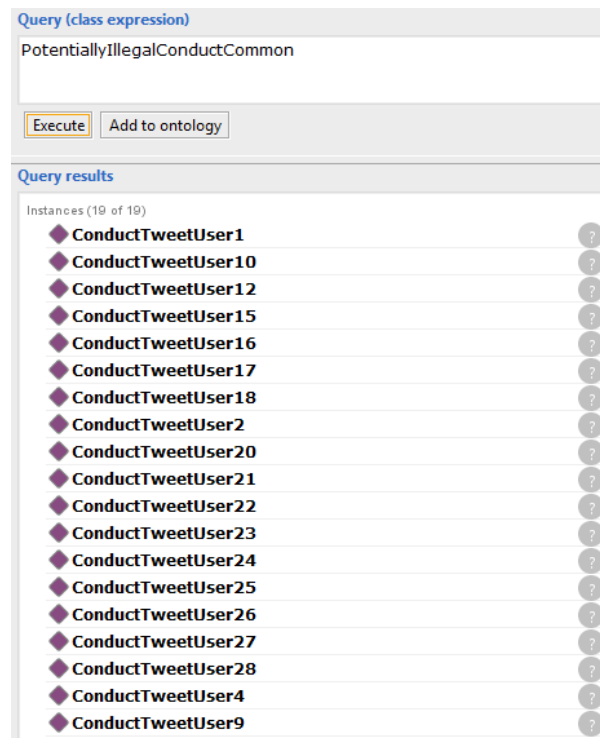


Κλείνοντας, μου ζητήθηκε να απαντήσω πως θα περίγραφα τον τόνο του συγκεκριμένου κειμένου. Προφανώς, έχει προσβλητικό τόνο αφού προσβάλλεται το Ισλάμ σαν θρησκεία. Παρακάτω, φαίνεται η αλλαγή στην οντολογία.

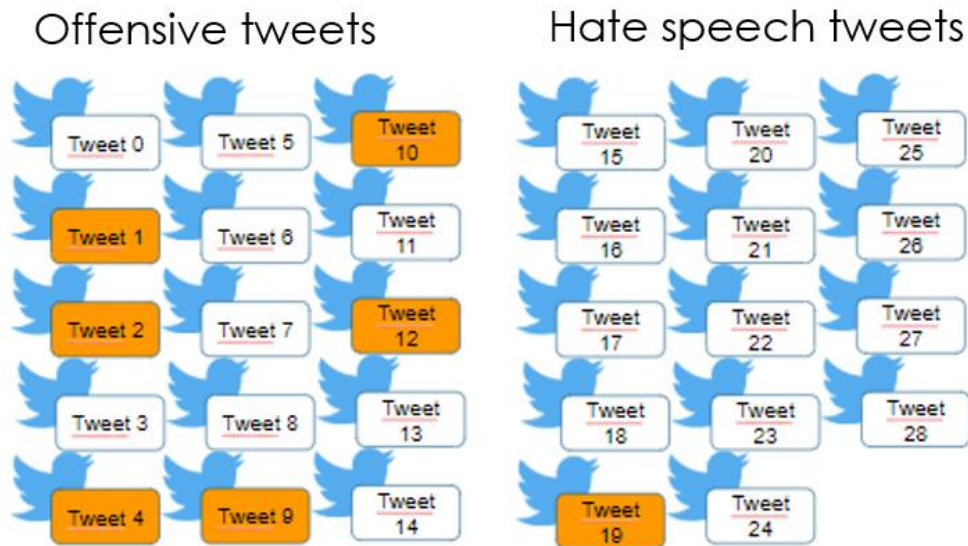


Επανέλαβα την διαδικασία για τα υπόλοιπα 27 tweets και άφησα τον hermit reasoner να βγάλει τα δικά του συμπεράσματα, δηλαδή να αποφασίσει αν το συγκεκριμένο tweet είναι potentially illegal και αν είναι σε ποια πιθανών παράνομη συμπεριφορά το κατατάσσει στα πλαίσια της ρητορικής μίσους. Το έκανα αυτό προκειμένου να δω αν οι απαντήσεις του reasoner συγκλίνουν με τις απαντήσεις των ειδικών επιστημόνων.

Όπως μπορείτε να δείτε από την πιο κάτω εικόνα δεν συγκλίνουν πλήρως τα αποτελέσματα.



Σχήμα 8-1 Screenshot από το Protégé, όπου εμφανίζω όλα τα Conduct τα οποία είναι πιθανώς παράνομα όπως τα έχει κατατάξει ο Hermit reasoner βάση των annotation που έχω κάνει.



Σχήμα 8-2 Στο σχήμα αυτό φαίνονται δύο ομάδες: αριστερά βρίσκονται όλα τα tweets τα οποία κατηγοριοποιήθηκαν ως offensive από τους κοινωνικούς επιστήμονες, ενώ στα δεξιά βρίσκονται τα tweets τα οποία εκφράζουν κολάσιμο hate speech. Αυτά που είναι χρωματισμένα με πορτοκαλί χρώμα είναι αυτά που ο Hermit reasoner τα έχει κατηγοριοποιήσει διαφορετικά από τους κοινωνικούς επιστήμονες.

8.2 Σύγκριση αποτελεσμάτων

False positive error: είναι όταν βγαίνει θετικό αποτέλεσμα ενώ θα έπρεπε να βγει αρνητικό.

False negative error: είναι όταν βγαίνει αρνητικό αποτέλεσμα ενώ θα έπρεπε να βγει θετικό.

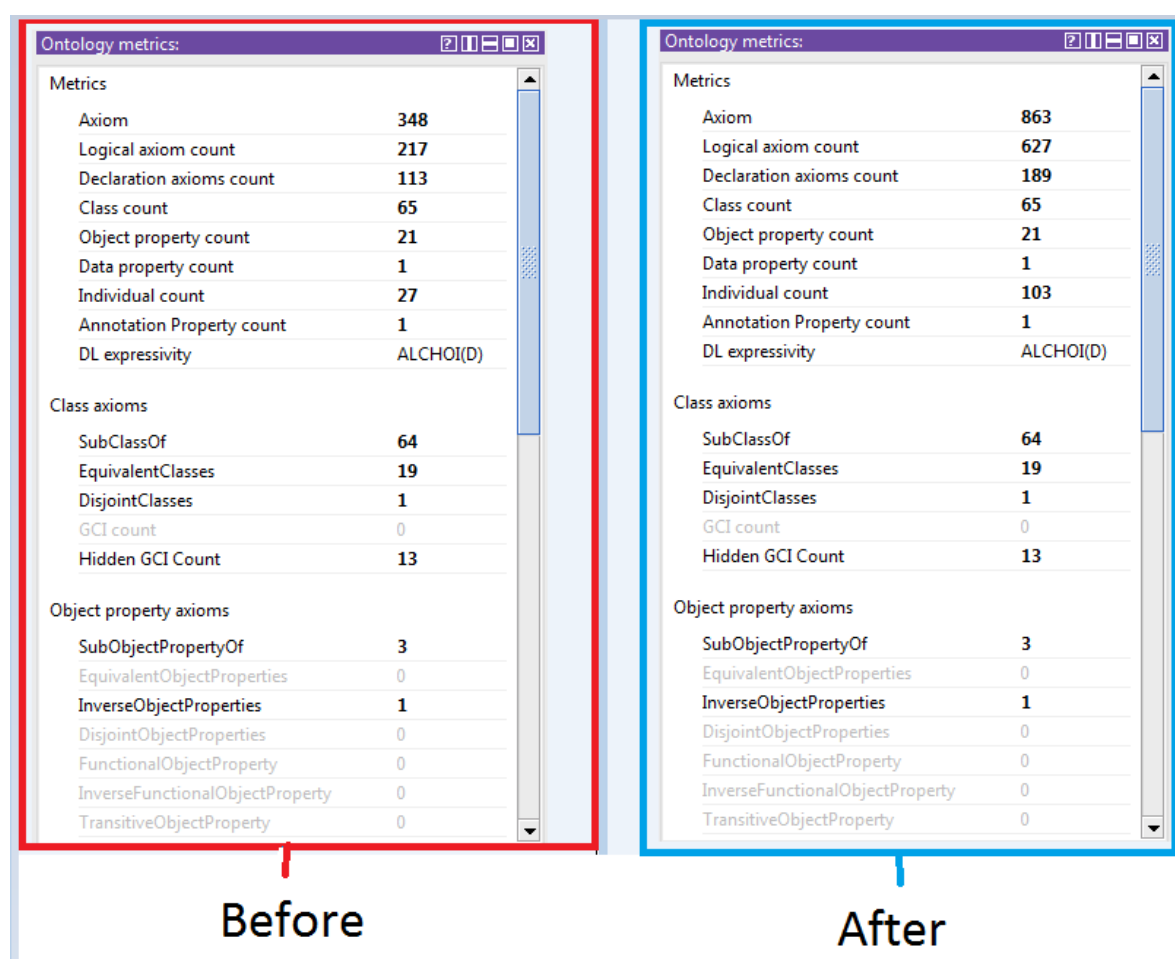
Στην προκειμένη περίπτωση false positive error εννοώ τα tweets που έχουν κατηγοριοποιηθεί από τον reasoner ως hate speech ενώ οι κοινωνικοί επιστήμονες το έχουν κατατάξει ως offensive tweet. Με false negative error εννοώ τα tweets το οποίο ο reasoner δεν βρήκε κανένα λόγο να το κατατάξει ως hate speech ενώ οι κοινωνικοί επιστήμονες το έχουν ορίσει ως hate speech tweet.

Για να εξετάσω την ορθότητα των αποτελεσμάτων μου έτρεξα ένα query μέσω του Protégé για να βρω όλα τα στιγμιότυπα που ανήκαν στην κλάση “PotentiallyIllegalConduct” (Σχήμα 8-1). Να υπενθυμίσω ότι η κλάση “PotentiallyIllegalConduct” αντιπροσωπεύει μία κολάσιμη συμπεριφορά που εκφράζει hate speech, επομένως αντιστοιχεί με την έννοια hate speech των κοινωνικών επιστημόνων. Συγκρίνοντας τα αποτελέσματα μου με αυτά των

κοινωνικών επιστημόνων διαπίστωσα ότι έχω ποσοστό απόκλισης 24,1%. Πιο συγκεκριμένα έχω κάνει 6 σφάλματα false positive και 1 false negative σφάλμα όπως μπορείτε να δείτε και στο Σχήμα 8-2.

8.3 Επέκταση οντολογίας

Μία από τις συνεισφορές της εργασίας μου είναι το ότι οι χρήστες οι οποίοι κάνουν annotate, επεκτείνουν και εμπλουτίζουν την οντολογία μέσω της γραφικής διεπαφής. Όπως μπορείτε να δείτε από το Σχήμα 8-3, αφού έγινε το annotation έχουν προστεθεί αξιώματα και στιγμιότυπα.



Σχήμα 8-3 Η πιο πάνω εικόνα δείχνει μετρικές της οντολογίας από το Protégé. Αριστερά δείχνω τις μετρικές πριν να γίνει το annotation, και δεξιά δείχνω τις μετρικές μετά το annotation.

Κεφάλαιο 9

Συμπεράσματα

9.1 Γενικά Συμπεράσματα	69
9.2 Βελτίωση Αποτελεσμάτων	69
9.3 Μελλοντική εργασία	70

9.1 Γενικά Συμπεράσματα

Με ένα ποσοστό σύγκλισης 75,9% μπορώ να πω ότι το annotation μου δεν ήτανε και τόσο άσχημο. Επίσης υπενθυμίζω πως τα σφάλματα τύπου False Positive είναι περισσότερα από τα False Negative. Αυτό κατά την γνώμη μου είναι καλύτερο καθώς η ζημιά των False Negative errors είναι μεγαλύτερη. Στην περίπτωση ενός False Positive το κείμενο θα περάσει και από δεύτερο ανθρώπινο έλεγχο που θα εξεταστεί κατά πόσο το συγκεκριμένο tweet εκφράζει ρητορική μίσους. Όμως στην περίπτωση ενός False Negative το κείμενο θα παραμείνει ορατό, προωθώντας έτσι hate speech μέχρις ότου αναφερθεί από χρήστες.

Επίσης πρόσεξα ότι η διεπαφή μου είναι λίγο περιοριστική καθώς δεν επιτρέπει στον χρήστη να επιλέξει περισσότερες από μία απαντήσεις σε μία ερώτηση. Αυτό όμως περιορίζει τον χρήστη στην εκφραστικότητα του, αφού ένα κείμενο μπορεί να έχει πολλές ερμηνείες, συνεπώς να αντιστοιχούν περισσότερες από μια απαντήσεις.

9.2 Βελτίωση Αποτελεσμάτων

Παρόλο που τα αποτελέσματα ήτανε αρκετά ικανοποιητικά, δεδομένου ότι δεν είχαν γίνει annotated από κάποιο ειδικό, υπάρχει περιθώριο για βελτίωση. Ένα από τα λάθη που είχα

κάνει, είναι το ότι είχα κάνει μερικές υποθέσεις κατά την διάρκεια του annotation που ήτανε σε κάποιες περιπτώσεις αβάσιμες. Αυτό το πρόβλημα προκύπτει από την έλλειψη περισσότερων πληροφοριών για το συγκεκριμένο tweet, έτσι δεν είναι εύκολο να λυθεί. Παρόλα αυτά πιστεύω πώς η καλύτερη λύση είναι να αυξηθεί ο αριθμός των χρηστών που κάνουν annotate το συγκεκριμένο tweet, ώστε να έχουμε μία πιο σφαιρική και αντικειμενική άποψη σχετικά με αυτό το ζήτημα.

Ακόμη ένα λάθος που προέκυψε είναι λόγω του ότι η οντολογία δεν έχει κάποιο αξίωμα το οποίο συσχετίζει το χρώμα δέρματος με την αντίστοιχη φυλή. Αυτό το λάθος ανήκει σε ένα πιο γενικό πρόβλημα το οποίο λύνεται με τη συνεχή συντήρηση και βελτίωση της οντολογίας.

Μία τελευταία κατηγορία λαθών οφείλονταν στην περιορισμένη εκφραστικότητα που προσφέρει τόσο η διεπαφή όσο και η οντολογία. Ως αποτέλεσμα της περιορισμένης εκφραστικότητας, ο χρήστης επιλέγει απαντήσεις που κοντεύουν σε αυτό που ήθελε να εκφράσει αλλά αυτό στο τέλος οδηγεί σε λανθασμένα συμπεράσματα από τον reasoner. Λύση αυτού του προβλήματος έρχεται και πάλι με την συντήρηση της οντολογίας, όμως απαιτείται και η διεπαφή να γίνει πιο ευέλικτη (π.χ με περισσότερες επιλογές).

9.3 Μελλοντική εργασία

Προτείνω τις εξής αλλαγές και βελτιώσεις σαν μελλοντική εργασία:

Αρχικά θα ήτανε καλό να δοκιμαστούν διαφορετικές μεθοδολογίες σχεδίασης οντολογιών, ώστε να βρεθεί η κατάλληλη για την λύση του συγκεκριμένου προβλήματος.

Δεύτερον θα πρέπει να γίνει η διεπαφή πιο ευέλικτη ως προς την εκφραστικότητα των annotations, καθώς και να βελτιωθεί η ευχρηστία της. Ιδανικά η διεπαφή θα πρέπει να επανασχεδιαστεί ως διαδικτυακή εφαρμογή, ούτως ώστε να μπορούνε οι χρήστες να κάνουν annotate μέσω του browser χωρίς να απαιτείται η εγκατάσταση στον υπολογιστή τους. Με αυτό τον τρόπο θα αυξηθεί και ο αριθμός των ατόμων που θα κάνουν annotate. Όσο αφορά το τεχνικό κομμάτι της σχεδίασης της εφαρμογής, ο reasoner και η οντολογία θα βρίσκονται στο backend.

Πηγές

- [1] T. Berners-Lee, J. Hendler και O. Lassila, «The Semantic Web,» *Scientific american*, pp. 28-37, 17 May 2001.
- [2] I. Horrocks και S. Bechhofer, «Semantic Web,» σε *Web Accessibility*, Springer, 2008, pp. 315-330.
- [3] E. Oren, K. H. Moller, S. Scerri, S. Handschuh και M. Sintek, «What are Semantic Annotations?,» σε *Relatório técnico*, DERI Galway, 2006, p. 62.
- [4] «The Unicode® Standard: A Technical Introduction,» 23 6 2017. [Ηλεκτρονικό]. Available: <http://www.unicode.org/standard/principles.html>.
- [5] J. Hebel, M. Fisher, R. Blace και A. Perez-Lopez, σε *Semantic Web Programming*, Wiley, 2009, p. 11.
- [6] G. Antoniou και F. v. Harmelen, σε *A Semantic Web Primer*, MIT Press, 2008, pp. 13-15, 18, 19, 25-28, 65-69, 114-119.
- [7] A. Medic και A. Golubovic, «Making secure Semantic Web,» *Universal Journal of Computer Science and Engineering Technology*, pp. 99-104, 2010.
- [8] C. Roussey, F. Pinet, M. A. Kang και O. Corcho, «An introduction to ontologies and ontology engineering,» σε *Ontologies in Urban Development Projects*, London, Springer, 2011, pp. 9-38.
- [9] M. Kifer, J. d. Bruijn, H. Boley και D. Fensel, «A Realistic Architecture for the Semantic Web,» σε *First International Conference, RuleML*, Galway, 2005.
- [10] A. Sunitha, «A Survey on Ontology Reasoners and Comparison,» *International Journal of Computer Applications*, pp. 33-38, November 2012.

- [11] B. Chandrasekaran, J. John R και B. V. Richard, «What are ontologies, and why do we need them?,» *IEEE Intelligent Systems and their applications*, pp. 20-26, 1999.
- [12] T. R. Gruber, «A translation approach to portable ontology specifications,» σε *Knowledge acquisition*, Elsevier, 1993, pp. 199-220.
- [13] N. Guarino, D. Oberle και S. Stabb, «What is an Ontology?,» σε *Handbook on ontologies*, Berlin Heidelberg, Springer, 2009, pp. 1-17.
- [14] M. Uschold και M. Gruninger, «Ontologies and Semantics for Seamless Connectivity,» *SIGMOD*, pp. 58-64, 2004.
- [15] N. Guarino, «Formal ontology and information systems,» σε *Formal Ontology in Information Systems*, Amsterdam, 1998.
- [16] L. Lee W., σε *OWL: Representing information using the web ontology language*, Victoria, British Columbia, Canada, Trafford Publishing, 2005, pp. 33-34.
- [17] N. Natalya F. και M. Deborah L., «Ontology Development 101: A Guide to Creating Your First Ontology,» 2001. [Ηλεκτρονικό]. Available: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf. [Πρόσβαση 6 Ιουνίου 2018].
- [18] M. Synak, M. Dabrowski και S. Ryszard Kruk, «Semantic Web and Ontologies,» σε *Semantic Digital Libraries*, Berlin Heidelberg, 2009.
- [19] M. Horridge, *A Practical Guide To Building OWL Ontologies. Using Protege 4 and CO-ODE Tools Edition 1.3*, University of Manchester, 2011.
- [20] A. Valente, «Types and Roles of Legal Ontologies,» σε *Law and the Semantic Web*, Berlin Heidelberg, Springer, 2005, pp. 69-71.
- [21] «protege,» 2016. [Ηλεκτρονικό]. Available: <https://protege.stanford.edu/>. [Πρόσβαση

10 1 2018].

- [22] A. Kalyanpur, B. Parsia, E. Sirin, B. Grau και J. Hendler, «Swoop: A web ontology editing browser,» *Web Semantics: Science, Services and Agents on the World Wide Web*, pp. 144-153, 2006.
- [23] H. Knublauch, R. W. Fergerson, N. F. Noy και M. A. Musen, «The Protégé OWL plugin: An open development environment for semantic web applications,» σε *The Semantic Web – ISWC 2004*, Hiroshima, Japan, 2004.
- [24] σε *Σύνδεσμοι Ένα εγχειρίδιο για την καταπολέμηση της ρητορικής μίσους στο διαδίκτυο μέσα από την εκπαίδευση στα ανθρώπινα δικαιώματα*, No Hate Speech Movement, 2015, pp. 168, 182-189, 196, 211.
- [25] «MANDOLA - Monitoring and Detecting Online Hate Speech,» σε *D4.1b: FAQ on Responding to online hate speech*, Rights, Equality and Citizenship Programme of the European Commission, 2017, pp. 20,24.
- [26] «Επίτροπος Διοικήσεως και Προστασίας Ανθρωπίνων Δικαιωμάτων,» 21 Μάιος 2018. [Ηλεκτρονικό]. Available: <http://www.ombudsman.gov.cy/Ombudsman/ombudsman.nsf/All/3FE93DE941BC57EFC2257E8C002F458B?OpenDocument>.
- [27] «AEQUITAS,» 21 Μάιος 2018. [Ηλεκτρονικό]. Available: http://www.aequitas-humanrights.org/wp-content/uploads/2017/01/booklet_hate-speech.pdf.
- [28] «MANDOLA - Monitoring And Detecting Online Hate Speech,» σε *D2.1b (final report) Definition of illegal hatred and implications*, Rights, Equality and Citizenship Programme of the EU Commission, 2017, pp. 76-161.