

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

Interface

USER'S GUIDE

* * * * * * * This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

November 2020

http://pharmproject.usal.es

This report constitutes Deliverable 19, for Work Package 4 of the PHARM project.

October 2020

© 2020 – PHARM, Preventing Hate Against Refugees and Migrants, – GA number 875217.

General contact: pharm.project.eu@gmail.com

For more information type the e-mail address of the corresponding author

Information may be quoted provided the source is stated accurately and clearly. This publication is also available via https://pharmproject.usal.es/

This project is funded by the European Union's Rights, Equality and Citizenship Programme (2014-2020). REC-RRAC-RACI-AG-2019. Grant Agreement number 875217

The content of this report represents the views of the author only and is his/her sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

Copyright © Pharm Consortium, 2020

All rights reserved. No part of the report may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, and may not be quoted or cited, without prior permission in writing from the project coordinator.

The views expressed during the execution of the project, in whatever form and or by whatever medium, are the sole responsibility of the authors. The European Union is not liable for any use that may be made of the information contained therein.



VERSION CONTROL SHEET

Deliverable number + title	D10. Guide of use of the interface
Prepared by	AUTH TEAM
Work package number	2
Work package leader	Andreas Veglis
Dissemination level (PU, CO)	PU
Delivery date	30/11/2020
Submission date to EC	30/11/2020
Main authors	Lazaros Vrysis / Nikolaos Vryzas
Reviewers	Carlos Arcila Calderón / Patricia Sánchez Holgado

REVISION HISTORY

Version	Date	Summary of changes	Initials	Changes marked
1				

εγχειρίδιο χρησης

Για τη διαδικτυακή εφαρμογή PHARM

Ελληνικά έκδοση 1.0

Πίνακας περιεχομένων

1	Γενι	κές πλ	ηροφορίες	2
	1.1	Η εφα	αρμογή PHARM	2
	1.2	Τεχνι	κές πληροφορίες	3
	1.2.3	1 9	Scraping	3
	1.2.2	2 1	Προσδιορισμός χρόνου	3
	1.2.3	3 I	Προσδιορισμός τοποθεσίας	4
	1.2.4	4 I	Εντοπισμός γλώσσας	4
	1.2.	5 I	Εντοπισμός ρητορικής μίσους	4
	1.2.0	6	Ανάλυση συναισθήματος	4
	1.3	Χρήσ ⁻	τες	4
	1.4	Λειτο	ουργίες	5
	1.4.3	1 /	Αναζήτηση δεδομένων	5
	1.4.2	2	Ανάλυση δεδομένων	5
	1.4.3	3 2	Συλλογή δεδομένων	6
	1.4.4	4 E	Επισημείωση δεδομένων	7
	1.4.	5 I	Εισαγωγή δεδομένων	7
	1.5	Моро	φή δεδομένων	7
2	Οδη	γίες χρ	ρήσης της εφαρμογής PHARM1	0
	2.1	Εισαγ	γωγή1ι	0
	2.2	Η αρχ	χική οθόνη της εφαρμογής1	0
	2.3	Αναζή	ήτηση δεδομένων1	0
	2.4	Ανάλι	υση δεδομένων1	1
	2.5	Συλλα	ογή νέων δεδομένων1	3
	2.6	Επιση	ημείωση δεδομένων1	3
	2.7	Εισαγ	γωγή δεδομένων1	3
	2.8	Είσοδ	δος στην εφαρμογή1	5
3	Οδη	γίες χρ	ρήσης της εφαρμογής doccano1	6
	3.1	Εισαγ	<i>γ</i> ωγή1	6
	3.2	Επισr	ημείωση καταχώρησης1	6
	3.3	Έγκρι	ση επισημείωσης1	8

1 Γενικές πληροφορίες

To Preventing Hate Against Refugees and Migrants (PHARM) αποτελεί ένα ερευνητικό έργο που χρηματοδοτείται από την Ευρωπαϊκή Ένωση, στο πλαίσιο της πράξης Rights, Equality, and Citizenship REC-RRAC-RACI-AG-2019 (GA n. 875217). Ο κύριος στόχος του PHARM είναι η παρακολούθηση και η μοντελοποίηση της ρητορικής μίσους εναντίον προσφύγων και μεταναστών στην Ελλάδα, την Ιταλία και την Ισπανία, με σκοπό την πρόβλεψη και την καταπολέμηση των εγκλημάτων μίσους και την αντιμετώπιση των επιπτώσεών τους. Γι' αυτόν τον σκοπό χρησιμοποιούνται τεχνολογίες αιχμής, όπως η δημοσιογραφία δεδομένων. Το κύριο αποτέλεσμα του έργου θα είναι ο εντοπισμός και η μείωση της ρητορικής μίσους στο διαδίκτυο και η πρόβλεψη πιθανών εγκλημάτων μίσους.

1.1 Η εφαρμογή PHARM

Η εφαρμογή *PHARM* είναι μια διαδικτυακή διεπαφή για την αναζήτηση και την αποθήκευση δεδομένων κειμένου από πολλαπλές διαδικτυακές πηγές. Η εφαρμογή αποτελείται από μια βάση δεδομένων και τη γραφική διεπαφή που τη συνοδεύει, η οποία εξασφαλίζει τη λειτουργικότητα για τη διαχείριση των δεδομένων. Η πρώτη οθόνη που παρουσιάζεται στον χρήστη κατά την είσοδό του στην εφαρμογή είναι η αρχική οθόνη, η οποία απεικονίζεται στο Figure 1. Η αρχική οθόνη, πέρα από την αποτύπωση κάποιων βασικών πληροφοριών του έργου, αποτελεί και το εναρκτήριο σημείο για να πραγματοποιηθούν αναλύσεις σε κείμενο.



Figure 1 Η αρχική οθόνη της εφαρμογής

Οι βασικές λειτουργίες που προσφέρει η εφαρμογή είναι:

- **Αναζήτηση:** Αφορά την υποβολή ερωτημάτων στη βάση δεδομένων, την παρουσίαση των σχετικών αποτελεσμάτων και τη μεταφόρτωση τους στον Η/Υ του χρήστη.
- Ανάλυση: Αυτή η λειτουργία σχετίζεται με τα αποτελέσματα αναλύσεων από ποικίλους αλγόριθμους επεξεργασίας φυσικής γλώσσας (Natural Language Processing – NLP), όπως εντοπισμός ρητορικής μίσους, ανάλυση συναισθήματος, προσδιορισμός της τοποθεσίας στην οποία αναφέρεται κάποιο κείμενο.

- Συλλογή: Αφορά τη μαζική συλλογή δεδομένων κειμένου (scraping).
- **Επισημείωση:** Αναφέρεται στη δυνατότητα για επισημείωση δεδομένων.
- **Εισαγωγή:** Σχετίζεται με την εισαγωγή νέων καταγραφών στη βάση δεδομένων.

Οι λειτουργίες αναζήτησης και ανάλυσης προσφέρονται ελεύθερα σε όλους τους χρήστες που χρησιμοποιούν την εφαρμογή, ενώ, αντίθετα, απαιτείται έλεγχος ταυτότητας για τις λειτουργίες της συλλογής, της επισημείωσης και της εισαγωγής. Εάν ένας χρήστης θέλει να συνεισφέρει στη βάση δεδομένων του PHARM, θα πρέπει να επικοινωνήσει με την ομάδα ανάπτυξης της εφαρμογής, ώστε να αιτηθεί έναν τέτοιο ρόλο.

1.2 Τεχνικές πληροφορίες

Για την ανάπτυξη της εφαρμογής, χρησιμοποιήθηκε το Python framework Flask. Η επιλογή αυτή έγινε γιατί οι ρουτίνες επεξεργασίας φυσικής γλώσσας και οι ανάλυσης δεδομένων αναπτύχθηκαν επίσης σε Python. Με αυτήν την προσέγγιση, όλες οι λειτουργίες της εφαρμογής μπορούν να συμπεριληφθούν σε ένα κοινό έργο λογισμικού. Το γραφικό περιβάλλον έχει σχεδιαστεί κυρίως στο Bootstrap, μια δημοφιλή βιβλιοθήκη HTML, CSS και JavaScript ενώ πρόσθετα μπλοκ HTML, CSS και JavaScript έχουν συμπεριληφθεί όπου χρειάζεται. Η εφαρμογή έχει αναπτυχθεί και εξυπηρετείται από μια εικονική μηχανή (Virtual Machine - VM) κάνοντας χρήση του πακέτου λογισμικού waitress. Αυτό αφορά έναν διακομιστή παραγωγής σε Python με αποδεκτή απόδοση.

Διάφορες βιβλιοθήκες έχουν χρησιμοποιηθεί για την ανάπτυξη των μεθόδων ανάλυσης, μαζί με προσαρμοσμένους αλγόριθμους που έχουν σχεδιαστεί ειδικά για το έργο PHARM. Μερικές από αυτές τις μεθόδους εξασφαλίζουν δυνατότητες όπως τον προσδιορισμό ημερομηνίας και ώρας, την εκτίμηση της γεωγραφικής τοποθεσίας, την ανίχνευση γλώσσας ή την ανίχνευση μίσους. Επιπλέον λειτουργίες, όπως η ανάλυση συναισθήματος, θα είναι διαθέσιμες μελλοντικά, σύμφωνα με το χρονοδιάγραμμα του έργου. Οι θεμελιώδεις σχεδιαστικές επιλογές αυτών των μεθόδων παρουσιάζονται παρακάτω. Για περισσότερες τεχνικές πληροφορίες, επισκεφθείτε το <u>αποθετήριο του PHARM στο GitHub</u>.

1.2.1 Scraping

Η εφαρμογή υποστηρίζει τη συλλογή κειμένων από το *Twitter* μέσω του σχετικού *API* και της μεθόδου ροής (stream). Η βιβλιοθήκη tweepy χρησιμοποιείται για την πρόσβαση στο *Twitter API* μέσω της *Python*. Έχουν αναπτυχθεί τέσσερα λεξικά (lexicons) για φιλτράρισμα των tweets, τα οποία περιλαμβάνουν επιλεγμένους ελληνικούς, αγγλικούς, ισπανικούς και ιταλικούς όρους. Η συλλογή σχολίων από το *YouTube* υποστηρίζεται μέσω του *Google API*. Σε αυτή την περίπτωση απαιτείται ο προσδιορισμός ενός ερωτήματος αναζήτησης (π.χ. "πρόσφυγες, μετανάστευση") από τον χρήστη για να συλλεχθούν σχόλια από τα σχετικά βίντεο.

1.2.2 Προσδιορισμός χρόνου

Έχει αναπτυχθεί μια μέθοδος για τον εντοπισμό και την τυποποίηση πληροφοριών ημερομηνίας και ώρας από μεταδεδομένα και κείμενο. Εκτός από την τοποθεσία και τη γλώσσα, όταν μεταδεδομένα είναι διαθέσιμα, το *PHARM* χρησιμοποιεί επιπλέον πληροφορίες για τον εντοπισμό ρητορικής μίσους. Ορισμένες από αυτές τις επιπλέον πληροφορίες, όπως ημερομηνία και ώρα, ενδέχεται να είναι διαθέσιμες σε πολλές διάφορες μορφές. Αυτό καθιστά αναγκαία την τυποποίηση. Για τις ανάγκες αυτής της απαίτησης, χρησιμοποιούνται κατά σειρά οι βιβλιοθήκες λογισμικού *dateparser, datefinder* και *parsedatetime*, με προτεραιότητα στη μέθοδο που εξασφαλίζει υψηλότερη ακρίβεια. Εάν η πιο ακριβής μέθοδος δεν επιστρέψει αποτέλεσμα, καλείται η επόμενη. Η ανίχνευση βασίζεται στα μεταδεδομένα, στα οποία συνήθως εντοπίζεται πληροφορία χρονικής σήμανσης. Εάν η ανίχνευση αποτύχει για όλες τις υπηρεσίες στα μεταδεδομένα, η ίδια ροή εργασίας εφαρμόζεται στο ίδιο το κείμενο.

1.2.3 Προσδιορισμός τοποθεσίας

Στην εφαρμογή ενσωματώνεται μια μέθοδος για τον εντοπισμό της γεωγραφικής τοποθεσίας από δεδομένα κειμένου Η βιβλιοθήκη geopy, μαζί με το γεωκωδικοποιητή nominatim έχουν αξιοποιηθεί. Ονοματισμένες οντότητες (named entities) εξάγονται από τα κείμενα, σύμφωνα με την ακόλουθη κατάταξη: GPE (χώρες, πολιτείες, πόλεις κ.λπ.), LOC (βουνά, ωκεανοί κ.λπ.), FAC (κτίρια, αεροδρόμια, αυτοκινητόδρομοι κ.λπ.), ORG (εταιρείες, ιδρύματα κ.λπ.), και μετέπειτα, εφαρμόζεται σε αυτές αντίστροφη γεωκωδιοποίηση για τον χωρικό προσδιορισμό τους.

1.2.4 Εντοπισμός γλώσσας

Αντικείμενο το έργου είναι η ανάλυση κειμένων στην ελληνική, την ιταλική και την ισπανική γλώσσα. Όμως, πολλές καταγραφές ενδέχεται να περιέχουν κείμενο σε άλλες γλώσσες ή τοπικές διαλέκτους. Για να καταστεί δυνατή η επεξεργασία κειμένου στις παραπάνω τρεις εθνικές γλώσσες, έχει κωδικοποιηθεί μια διαδικασία για την ανίχνευση της γλώσσας σε κείμενο. Υιοθετείται μια προσέγγιση με αξιοποίηση πολλών εφεδρικών υπηρεσιών ανάλυσης για αυξημένη ευρωστία. Γι' αυτόν τον σκοπό χρησιμοποιούνται οι βιβλιοθήκες *textblob*, googletrans και langdetect. Εάν μια υπηρεσία αποτύχει, ζητείται το αποτέλεσμα από την επόμενη.

1.2.5 Εντοπισμός ρητορικής μίσους

Στο πλαίσιο του έργου έχουν αναπτυχθεί μερικές μέθοδοι για την ανίχνευση ρητορικής μίσους σε κείμενα. Σε αυτές περιλαμβάνονται η απλή αντιστοίχιση συμβολοσειρών (strings) και η κατά προσέγγιση αντιστοίχιση συμβολοσειρών. Η διαδικασία αντιστοίχισης όρων στοχεύει στο να μην εξαρτάται από τις διαφορετικές καταλήξεις για το ίδιο λήμμα, φαινόμενο που μπορεί να συμβαίνει σε πολλές γλώσσες, ανάλογα το πρόσωπο, το φύλο, ή τον ενικό/πληθυντικό αριθμό. Επίσης έχει αναπτυχθεί μια προσέγγιση με χρήση διανυσμάτων λέξεων, η οποία λαμβάνει υπόψη τη σημασιολογική έννοια των όρων. Μια υβριδική προσέγγιση που βασίζεται σε λεξικά με προκαθορισμένες φράσεις, μαζί με δυναμικούς συνδυασμούς όρων (κυρίως επίθετα με ουσιαστικά) έχει υλοποιηθεί και βρίσκεται υπό αξιολόγηση. Η διαδικασία έχει ως εξής: φορτώνεται ένα μοντέλο γλώσσας, ακολουθούνται κοινές πρακτικές τυποποίησης κειμένου (lowercasing, lemmatization, stop-word removal) και οι όροι-κλειδιά αναζητούνται στο κείμενο. Εάν βρεθεί ένας όρος (ή συνδυασμός όρων), το κείμενο επιεθόδους μηχανικής μάθησης, σχεδιάζεται και θα ενσωματωθεί επίσης στη εφαρμογή.

1.2.6 Ανάλυση συναισθήματος

Δυο μέθοδοι για τον εντοπισμό ρητορικής μίσους είναι υπό ανάπτυξη και θα ενσωματωθούν στην εφαρμογή, σύμφωνα με το χρονοδιάγραμμα του έργου. Αυτές αφορούν μια μη-επιβλεπόμενη προσέγγιση με λεξικά, και μια που στηρίζεται σε αλγορίθμους μηχανικής μάθησης.

1.3 Χρήστες

Η εφαρμογή προδιαγράφει δύο τύπους χρηστών: τον συνεργάτη (επικυρωμένος) και τον φιλοξενούμενο (μη επικυρωμένος). Ο φιλοξενούμενος έχει πρόσβαση στις λειτουργίες της αναζήτησης και της ανάλυσης δεδομένων, ενώ ο συνεργάτης έχει πλήρη πρόσβαση σε όλες τις διαθέσιμες λειτουργίες της εφαρμογής. Οι βασικές ροές ενεργειών και για τους δύο τύπους χρηστών δίνονται παρακάτω.

Συνεργάτης

- Αναζήτηση -> Ανάλυση
- Συλλογή -> Επισημείωση -> Εισαγωγή

Φιλοξενούμενος

• Αναζήτηση -> Ανάλυση

1.4 Λειτουργίες

Οι βασικές λειτουργίες που υποστηρίζονται από την εφαρμογή είναι: Αναζήτηση, Ανάλυση, Συλλογή, Επισημείωση και Εισαγωγή. Ακολουθεί μια πιο αναλυτική περιγραφή αυτών των λειτουργιών.

1.4.1 Αναζήτηση δεδομένων

Η πιο σημαντική λειτουργία είναι η πλοήγηση σε δεδομένα κειμένου που χαρακτηρίζονται από ρητορική μίσους. Αυτή η λειτουργία υποστηρίζεται από τη δυνατότητα εφαρμογής μιας σειράς φίλτρων (όπως πηγή, γλώσσα, ημερομηνία, κ.α.). Συγκεκριμένα, προς το παρόν, τα διαθέσιμα φίλτρα είναι τα εξής:

- Επιλογή πηγής (Twitter, YouTube Facebook, Website Articles, Website Comments).
- Επιλογή γλώσσας (αγγλική, ελληνική, ιταλική, ισπανική).
- Προσδιορισμός χρονικού εύρους (ημερομηνία και ώρα).
- Αναζήτηση με λέξεις κλειδιά στο κείμενο.

Ο χρήστης μπορεί να κάνει προεπισκόπηση των αποτελεσμάτων αναζήτησης, να επιλέξει μεταξύ απλής (Simple) και εκτεταμένης (Scientific) προβολής, να μεταφορτώσει τα δεδομένα ως αρχείο CSV ή JSON, ή να προβάλει περισσότερες πληροφορίες για κάθε καταχώρηση.

PHARM I	nterface Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
	Analysis Enter text for hate speech and sentiment analysis	
	"RT @GerryConnolly: Bill Barr has sold his soul to the president and sullied the reputation of the Department of Justice. He is a disgrace t"	
	Perform Analysis	
	Results	
	Unsupervised Hate Speech Detection	
	Supervised Hate Speech Detection	
	Unsupervised Sentiment Analysis	
	Supervised Sentiment Analysis	

Figure 2 Η οθόνη προβολής αποτελεσμάτων ανάλυσης κειμένου

1.4.2 Ανάλυση δεδομένων

Όταν μια καταχώρηση επιλέγεται (ή γίνεται εισαγωγή κειμένου μέσω της αρχικής οθόνης), φορτώνεται η οθόνη πληροφοριών καταχώρησης (Figure 2). Η τοποθεσία στην οποία εκτιμάται ότι αναφέρεται το κείμενο σημειώνεται πάνω σε χάρτη, ενώ διάφορα αποτελέσματα αλγοριθμικής ανάλυσης παρουσιάζονται με κείμενο και γραφικά.

PHARM Interface	Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
Scrape Scrape comments from social Youtube Scraper Twitter Scraper	nedia	
Youtube scraper		
Enter search query		
Max results		
100		•
Scrape		

Figure 3 Η οθόνη συλλογής δεδομένων

1.4.3 Συλλογή δεδομένων

Αυτή η λειτουργία αναφέρεται στη μαζική συλλογή δεδομένων κειμένου από τις δημοφιλείς πλατφόρμες κοινωνικής δικτύωσης Twitter και YouTube. Ο χρήστης μπορεί εύκολα να συλλέξει κείμενα με ρητορική μίσους από το Twitter, επιλέγοντας τη γλώσσα που επιθυμεί (αγγλική, ελληνική, ιταλική ή ισπανική). Γι' αυτή τη διαδικασία χρησιμοποιούνται λεξικά που έχουν αναπτυχθεί στο πλαίσιο του έργου και τα αποτελέσματα της σάρωσης μπορούν να μεταφερθούν στον Η/Υ του χρήστη μέσω αρχείου JSON. Τα δεδομένα αυτά μπορούν μετέπειτα να αξιοποιηθούν για οποιαδήποτε εργασία επεξεργασίας φυσικής γλώσσας.

PHARM Interface	Home Search Submit Scrape Annotate Instructions Contact About	Welcome admin Logout
PHARM Interface About 6 results About 6 results Barack Hussen Obarna II (born August 4, 1961) is an American The Whate House is the official residence and workplace of L The Democratic Party is one of the	D annotate the data D annotate the data Live Demo V GitHub Login 1/6 Person p Loc 1 Org o Event o Date d Other z Barack Hussein Obama II • (born August 4, 1961 •) is an American • attorney and politician who served as the 44th President of	Logout
two major contemporary po Stanford University (officially Leland Stanford Junior Unive Donald John Trump (dom June 14, 1946) is the 45th and curre Silicon Valley (abbreviated as SV) is a region in the southe	the United States • from January 20, 2009 • , to January 20, 2017 • . A member of the Democratic Party • , he was the first African American • to serve as president. He was previously a United States Senator • from Illinois • and a member of the Illinois State Senate •.	

Figure 4 Η οθόνη επισημείωσης δεδομένων

Στην περίπτωση του YouTube, αντί να γίνεται επιλογή της γλώσσας, πρέπει να προσδιοριστεί ένα ερώτημα αναζήτησης, το οποίο να περιλαμβάνει φράσεις ή λέξεις-κλειδιά. Η εργασία αυτή ολοκληρώνεται σε κλάσματα του δευτερολέπτου και τα αποτελέσματα προβάλλονται

άμεσα. Εξυπακούεται πως τα δεδομένα μπορούν να μεταφορτωθούν ως αρχείο CSV ή JSON. Για περισσότερες πληροφορίες σχετικά με τα λεξικά-φίλτρα που χρησιμοποιούνται για τη συλλογή δεδομένων από το Twitter, μπορείτε να μελετήσετε το <u>αποθετήριο του PHARM στο</u> <u>GitHub</u>.

1.4.4 Επισημείωση δεδομένων

Η λειτουργία της επισημείωσης υποστηρίζεται από το εργαλείο doccano. Το doccano είναι μια εφαρμογή επισημείωσης δεδομένων κειμένου για την ανάπτυξη datasets. Για τους σκοπούς του συγκεκριμένου έργου, το doccano χρησιμοποιείται για την ανάθεση ετικετών στις καταχωρήσεις, προκειμένου να σημανθούν κατάλληλα, σύμφωνα με το προσδιορισμένο σχήμα ταξινόμησης (Hate/No Hate, Positive/Neutral/Negative sentiment).

1.4.5 Εισαγωγή δεδομένων

Η εισαγωγή νέων δεδομένων στη βάση δεδομένων γίνεται μέσω της οθόνης "Submit" (Figure 5). Η εισαγωγή μπορεί να γίνει είτε χειροκίνητα (μία-προς-μία) ή μαζικά. Όσον αφορά την πρώτη μέθοδο, ο χρήστης πρέπει να ορίσει όλα τα πεδία δεδομένων (κείμενο) και μεταδεδομένων (πηγή, γλώσσα, ημερομηνία, κ.α.) μέσω της ειδικά διαμορφωμένης φόρμας. Στην περίπτωση που πολλά δεδομένα είναι στη μορφή που υποστηρίζεται από την εφαρμογή (Παράγραφος 1.5), τότε μπορούν να εισαχθούν ως αρχείο JSON.

PHARM Interface	Home Search Submi	t Scrape Ar	notate Instructions	Contact About		Welcome admin Logout
Submit Submit a single or a batch o	of records to the PHARN	1 database				
Single Text Batch File						
Submit single hate-speech related	text to the PHARM database					
Enter text here						¹
refugees ×						
Choose text language:			Choose date (Optior	nal):		
English		~	mm / dd / yyyy			
Choose text source: O Twitter C) YouTube O Facebook Comi	ment O Faceb	oook Post 🔿 Website	Article O Website Com	ment	
Submit						

Figure 5 Η οθόνη εισαγωγής δεδομένων

1.5 Μορφή δεδομένων

Λαμβάνοντας υπόψη τις προδιαγραφές για τη μορφή δεδομένων στο πλαίσιο του έργου (δηλαδή τη χρήση επιπλέον πληροφοριών -μεταδεδομένων- που ενδέχεται να συνοδεύουν μια καταχώρηση κειμένου), τις πολλαπλές πηγές που χρησιμοποιούνται για τη συλλογή των δεδομένων (tweets, άρθρα και σχόλια ειδησεογραφικά websites), δυνατότητες διαλειτουργικότητας και συμβατότητας για την εισαγωγή και την εξαγωγή δεδομένων σε άλλες εφαρμογές λογισμικού (όπως την πλατφόρμα doccano), ορίστηκαν συγκεκριμένες προδιαγραφές που αφορούν τη μορφή των δεδομένων.

εγχειρίδιο χρήσης

Το θεμελιώδες πεδίο περιγραφής για κάθε καταχώρηση είναι αυτό του κειμένου (text), ενώ συνοδεύεται από τα πεδία τύπου (type), επισημείωσης (annotations) και μεταδεδομένων (meta). Ένα βασικό σύνολο πεδίων περιγραφής είναι κοινό για δεδομένα από όλες τις πηγές (όπως type, plang, phate, pdate, phate, psent, pterms, ploc). Στον πίνακα που ακολουθεί αποτυπώνεται ο τίτλος κάθε πεδίου και η σύντομη περιγραφή του.

Πεδίο	Περιγραφή
id	μοναδικό αναγνωριστικό
annotations	επισημειώσεις
meta	μεταδεδομένα
text	περιεχόμενο
type	τύπος καταγραφής (tweet, άρθρο, κτλ.)
plang	εντοπισμός γλώσσας μέσω αλγοριθμικής ανάλυσης
pdate	προσδιορισμός χρονικής σήμανσης μέσω αλγοριθμικής ανάλυσης
phate	εντοπισμός ρητορικής μίσους μέσω αλγοριθμικής ανάλυσης
psent	προσδιορισμός συναισθήματος μέσω αλγοριθμικής ανάλυσης
pterms	εντοπισμός συχνών όρων μέσω αλγοριθμικής ανάλυσης
ploc	Προσδιορισμός τοποθεσίας μέσω αλγοριθμικής ανάλυσης

Table 1 Τα βασικά πεδία περιγραφής της μορφής δεδομένων PHARM

Στα δεδομένα που προέρχονται από ιστότοπους (κυρίως ειδησεογραφικούς), τα μεταδεδομένα εξαρτώνται από τη διαθεσιμότητά τους από ιστότοπο σε ιστότοπο, ενώ στο YouTube και στο Tweeter είναι σαφώς προσδιορισμένα (Table 2**Error! Not a valid bookmark self-ref**erence.).

Twitter	YouTube
tweet_id	comment_id
is_retweet	reply_count
is_quote	like_count
user_id	video_id
username	video_title
scr_name	channel
location	video_description
followers	author_id
friends	author_name
date	date

Table 2 Τα επιπλέον πεδία μεταδεδομένων για τις καταχωρήσεις από YouTube και Twitter

Για μια πιο πλήρη άποψη σχετικά με τη μορφή των δεδομένων, ακολουθούν ενδεικτικές καταγραφές για κάθε πηγή.

Facebook

{"id": "1", "annotations": [{"label": "hate", "user": 2}, {"label': "negative", 'user': 2}], "meta": {"id": "80056833", "type": "facebook_comment", "source": "https://www.facebook.com/provinciale.rivista", "plang": "it", "pdate": "1571-10-15 00:00:00", "phate": "(nero cinesi)", "pterms": "", "ploc": "Italy, United States of America"}, "text": "andrebbe anche fatta con Venezia (principale flotta a Lepanto). 1571: le mie potentissime galeazze spaccheranno il culo all infedele ottomano. 2020: per favore basta quarantena, devo vendere lasagne sur-gelate in nero ai cinesi a 50 euro"}

Twitter

{"id": "2", "annotations": [{"label": "hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"type": "twitter_comment", "date": "10/13/2020", "tweet_id": 1315981643111432192, "is_retweet": true, "is_quote": false, "user_id": 1025701121749340160,

"username": "Christina Dim", "scr_name": "ChristinaDim31", "location": "Αθήνα ", "followers": 3178, "friends": 4756, "quoted_text": "", "pid": 27454420, "plang": "el", "pdate": "", "phate": "(Τουρκία μετανάστες)", "pterms": "", "ploc": "Turkey"}, "text": "RT @kanekos69: Εν τω μεταξύ αν γίνει στραβή με Τουρκία βλέπω μετανάστες εθελοντές στο μέτωπο και δεξιούς προσφυγες στα Παρίσια"}

Youtube

{"id": "3", "annotations": [{"label": "no_hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"type": "youtube_comment", "comment_id": "Ugy-SPKz3HGo4OohJnfR4AaABAg", "reply_count": 3, "like_count": 30, "video_id": "Gaz6UvRW0G8", "channel": "Mηδέν Ένα Μηδέν 010", "video_title": "0 1 0 ~ Πρόσφυγες", "video_desc": "Lyrics/Raps - 0 1 0 Beat by Apo (Aισθήσεις) Recorded @ Blackspot Studio Mix/Master by Sativa Cover by SpyOne (Baseline Co.) 0 1 0 IG: ...", "author_id": {"value": "UCPP-OugMmE8pNbWRbFv4YCA"}, "au-thor_name": "GATE21QNZ", "rating": "none", "date": "2020-09-03T21:16:31Z", "plang": "el", "pdate": "2020-10-25 00:00:00", "phate": "(Αφγανιστάν ισλαμιστές)", "pterms": "", "ploc": "Afghanistan"}, "text": "Είμαι 25 ετών, ονομάζομαι Μήτσος ο μαλάκας, Μαζεύω ισλαμιστές και λιποτάκτες από μια τούρκο βάρκα, Το παίζω ανοιχτόκαρδος με τις τσέπες του μπαμπάκαα, Και όταν ο Μήτσος πήγε να κάνει φίλους στο Αφγανιστάν, Τον σφάξαν σαν τους προγόνους του, για χάρη του Ισλαμ."}

Website Article

{"id": "4", "annotations": [{"label": "no_hate", "user": 2}, {"label': "positive", 'user': 2}], "meta": {"id": "92860318", "type": "article", "source": "http://defencereview.gr/gnorizontas-ta-gallika-ploia-oi-fremm-kai-o/", "meta": "Άμυνα Ελλάδα 9 Σεπτεμβρίου 2020 18:22 ", "title": "Γνωρίζοντας τα γαλλικά πλοία: Οι FREMM και οι (Video)", "lang": "el", "date": "2020-09-09 00:00:00", "hate": "", "terms": "", "loc": ""}, "text": Συχνά πυκνά αναφερόμαστε στις γαλλικές ναυπηγικές σχεδιάσεις. Τα παρακάτω βίντεο που βρήκα-με είναι αντιπροσωπευτικά για τις δυνατότητες των πλοίων με πολύ καλά σκηνοθετημένα βίντεο και ενδιαφέροντα πλάνα. Αξίζει να τα δείτε: "}

Website Comment

{"id": "5", "annotations": [{"label": "no_hate", "user": 2}, {"label': "neutral", 'user': 2}], "meta": {"id": "92860318", "type": "comment", "source": "http://defencereview.gr/gnorizontas-ta-gallika-ploia-oi-fremm-kai-o/", "lang": "el", "date": "", "hate": "", "terms": "", "loc": ""}, "text": "Το μπαραζ επεκτεινεται. Ελπιζω αυριο τετοια ωρα πανω κατω να μην κλαιμε"}

Σε κάθε περίπτωση, πληρέστερη περιγραφή σχετικά με τη μορφή δεδομένων που χρησιμοποιείται στο έργο PHARM δίνεται στο σχετικό αποθετήριο στο GitHub.

2 Οδηγίες χρήσης της εφαρμογής PHARM

2.1 Εισαγωγή

Όπως προαναφέρθηκε, η εφαρμογή PHARM προσφέρει δυνατότητες αναζήτησης, προβολής, ανάλυσης, συλλογής, επισημείωσης και εισαγωγής δεδομένων, μέσω μιας φιλικής γραφικής διεπαφής. Η εφαρμογή είναι προσβάσιμη μέσω του συνδέσμου: <u>pharm-interface.usal.es</u>.

2.2 Η αρχική οθόνη της εφαρμογής

Η αρχική οθόνη μπορεί να χρησιμοποιηθεί για την εισαγωγή και την ανάλυση κειμένου. Ο χρήστης μπορεί να πληκτρολογήσει (ή να κάνει επικόλληση) το κείμενο που επιθυμεί στο πλαίσιο κειμένου και έπειτα πρέπει να επιλέξει το κουμπί "Perform Analysis", όπως παρουσιάζεται στο Figure 6.



Figure 6 Ανάλυση κειμένου μέσω της αρχικής οθόνης

Με αυτόν τον τρόπο ανοίγει μια νέα καρτέλα στον περιηγητή διαδικτύου και παρουσιάζονται τα αποτελέσματα της ανάλυσης (Figure 9). Περισσότερες πληροφορίες σχετικά με αυτήν την Οθόνη Ανάλυσης υπάρχουν στην Παράγραφο 2.4. Σημειώνεται, πως η περιήγηση σε όλες τις επιμέρους οθόνες της εφαρμογής γίνεται από το κεντρικό μενού, το οποίο βρίσκεται στο πάνω μέρος της οθόνης.

2.3 Αναζήτηση δεδομένων

Μία από τις πιο σημαντικές λειτουργίες της εφαρμογής PHARM είναι η πλοήγηση στις διάφορες καταχωρήσεις κειμένου που φυλάσσονται στη βάση δεδομένων της.

Ο χρήστης επιλέγει "Search" από το κεντρικό μενού. Έπειτα, επιλέγει τα φίλτρα της αναζήτησης, όπως πηγή, γλώσσα, ημερομηνία (από-έως) και λέξεις-κλειδιά. Έπειτα επιλέγει το κουμπί "Search" (Figure 7). Τα φίλτρα της πηγής και της γλώσσας είναι υποχρεωτικά, ενώ ο ορισμός ημερομηνίας και λέξεων-κλειδιών γίνεται προαιρετικά.

εγχειρίδιο χρησης

Με την ολοκλήρωση του παραπάνω βήματος φορτώνει η οθόνη με τα αποτελέσματα της αναζήτησης, όπως φαίνεται στο Figure 8. Σε αυτή την οθόνη, πέρα από την προεπισκόπηση των δεδομένων, δίνεται η δυνατότητα για την μεταφόρτωση τους σε CSV ή JSON μορφή, όπως και για διαγραφή ή την εμφάνιση των πληροφοριών ανάλυσης κάθε εγγραφής. Επιπλέον, ο χρήστης μπορεί να επιλέξει μεταξύ της "Simple" (απλής) ή της "Scientific" (εκτεταμένης) προεπισκόπησης.

Σημειώνεται, πως δυνατότητα διαγραφής έχουν μόνο οι επικυρωμένοι χρήστες.

PHARM Interface		Home Search Su				
Searc Search for I	h nate speech content	using multiple f	ilters			
Choose source	e: ∕ouTube □Facebook Po aqe:	st 🛛 Facebook Con	nment 🗌 Website /	Article 🗌 Website Comment 🗌 A	ll Sources	
Italian						
Posted from:				until:		
11/02/202	0		\otimes	11/19/2020		3
Filter results w	ith query:					
Enter search	query					
Search						



PHARM Interface Home Search	Submit Scrape Annot	ate Instructions	Contact Abo	out	Welcome admin Logout
Search View the results of your query or go back for	a new search				
82 results found with the selected filters					
Simple View Scientific View			Ų	 Download results in csv (simple Download results in json (scientific 	view) view)
.@MoniquePariatEU (Διευθύντρια Μετανάστευσης Κ μεταναστευτικό https://t.co/7VrOQWdHbz https://t.c	ομισιόν): Ικανοτοίος o/fEb9iHEDui more	ους χειρισμούς	του υπ. Ναυτιλ	λίας και του Λιμενικού στο	
Φ Υ Λ Α Κ Η. Χωρίς δικαιώματα χωρίς παροχές χωρ	ίς τίποτα. Όλοι οι λαθρομ	ετανάστες να μπο	ουν φυλακή. m	iore 👕	
RT @giorgosptk: Λαθρομετανάστες που πήραν άσυλ φαντ more	ιο και έφυγαν από την κό	λαση της Μόριας	για Αθήνα επι	στρέφουν και πάλι στη Λέσβο.Ρε μ	λκες
RT @stoxos: ΔΕΙΤΕ ΤΟΥΣ ΣΙΧΑΜΕΝΟΥΣ!!! Σάμος: Λα more	θρομετανάστες έγλειφαν	τρόφιμα σε σούπι	ερ μάρκετ και γ	γελούσαν - Δείτε βίντεο https://t.co	o/2
Είμαστε σπόροι, που θα λουλουδιασουν. Η ευωδία μ από 400 χρόνια ζυγού. Δε θα χαθούμε από μία ξεφτ	ιας θα μυρίζει λιβάνι και ιλισμενη παγκόσμια ελίτ.	θα διώξει τους σα Είμαστε το μέλλοι	τανάδες. Το Έθ ν. more 👕	θνος των Ελλήνων δε χάθηκε κάτι	Ĵ
RT @hellene_bc: Είμαστε σπόροι, που θα λουλουδια more	σουν. Η ευωδία μας θα μι	ρίζει λιβάνι και θ	α διώξει τους ο	σατανάδες. Το Έθνος των Ελλήνων	ν δ
Σάμος: Λαθρομετανάστες έγλειφαν τρόφιμα σε σούτ	τερ μάρκετ και γελούσαν	- Δείτε βίντεο http	os://t.co/FFLJJ5	2WxM more 👕	

Figure 8 Εμφάνιση αποτελεσμάτων αναζήτησης

2.4 Ανάλυση δεδομένων

Με την επιλογή εμφάνισης των πληροφοριών ανάλυσης μιας καταχώρησης (ή μέσω της εισαγωγής κειμένου στην αρχική οθόνη), γίνεται φόρτωση της οθόνης πληροφοριών ανάλυσης

(Figure 9). Εκεί παρουσιάζονται τα αποτελέσματα της ανάλυσης κειμένου από τους αλγορίθμους που έχουν αναπτυχθεί στο πλαίσιο του έργου. Οι αναλύσεις αφορούν τον εντοπισμό ρητορικής μίσους, την ανάλυση συναισθήματος, τον εντοπισμό συχνών όρων/λέξεων, όπως και τον προσδιορισμό της γεωγραφικής θέσης.

Ο χρήστης έχει τη δυνατότητα να πραγματοποιήσει μια νέα ανάλυση, χωρίς να γυρίσει στην προηγούμενη οθόνη, τροποποιώντας απλώς το κείμενο και επιλέγοντας το κουμπί "Perform Analysis".

Στο κάτω μέρος της οθόνης, υπάρχουν συντομεύσεις για τη σήμανση λαθών/προτάσεων για μία καταχώρηση και τον έλεγχο της από την ερευνητική ομάδα του PHARM, όπως και για το αποθετήριο λογισμικού του έργου στο GitHub.

M Interface	Home Search Submit Scrape Annotate Instructions Contact About	Welco
Analysis	l sentiment analysis	
".@MoniquePariatEU (Διευθ μεταναστευτικό https://t.co	ντρια Μετανάστευσης Κομισιόν): Ικανοποίηση για τους χειρισμούς του υπ. Ναυτιλίας και του Λιμενικού στο '7VrOQWdHbz https://t.co/fEb9iHEDun"	
Perform Analysis		
Unsupervised Hate Spe	ech Detection	
Supervised Hate Speech	1 Detection	
Unsupervised Sentimen	t Analysis	
Supervised Sentiment A	nalysis	
Geographical estimation	n of hate speech references	1
business Consulting none, Lagina 572 00 View larger map	y Evangelia Directions VPCPATERA UDBAADAKH	
	Scholio Laginon Electrical supply store	
	Passias Cafe Bar Control Contr	Anny Pet sto Beorgia 5 'Dom 5 by Dol 00PTIAA
	Google Map data 82020 Terms of Use Report a ma	t IC
Do you disagree with the ha Consider contributing your of For more information about	te-speech and sentiment analysis evaluation? pinion to help us improve. the analysis algorithms please visit our GitHub page.	

Figure 9. Αποτελέσματα ανάλυσης κειμένου

2.5 Συλλογή νέων δεδομένων

Η εφαρμογή παρέχει την κατάλληλη λειτουργικότητα για τη μαζική συλλογή δεδομένων κειμένου από δύο δημοφιλείς πλατφόρμες κοινωνικής δικτύωσης, το Twitter και το YouTube. Με την επιλογή "YouTube Scraper" φορτώνεται η οθόνη που παρουσιάζεται στο Figure 10. Εκεί ο χρήστης ορίζει της λέξεις-κλειδιά, βάσει των οποίων θα γίνει συλλογή σχολίων από το YouTube. Επιπρόσθετα, μπορεί να οριστεί και ο αριθμός των σχολίων που θέλει να συλλεχθούν. Τέλος, με την επιλογή του κουμπιού "Scrape" γίνεται η φόρτωση των αποτελεσμάτων, όπως φαίνεται στο Figure 8. Τα δεδομένα μπορούν να μεταφορτωθούν στον Η/Υ του χρήστη σε αρχείο μορφής CSV ή JSON.

Στην περίπτωση του "Twitter Scraper", η οθόνη έχει παραπλήσια διάταξη (Figure 11). Σε αυτήν την περίπτωση όμως, επιλέγεται η γλώσσα της αναζήτησης και ο χρόνος για τον οποίο θα διαρκέσει η συλλογή των δεδομένων (tweets). Τέλος, πρέπει να επιλεχθεί το κουμπί "Scrape" για να ξεκινήσει η διαδικασία, ενώ ένας υπερσύνδεσμος γίνεται διαθέσιμος για την μεταφόρτωση των σχετικών δεδομένων.

PHARM Interface		n Submit Scr a	a pe Annotate			Welcome admin Logout
Scrape Scrape comments from social	media					
Youtube Scraper Twitter Scraper	<u>l</u>					
Youtube scraper						
refugees						
Max results						
100					B	8
Scrape						

Figure 10 Συλλογή κειμένων από το YouTube

PHARM Interface	Home Se	arch Submit	Scrape	Annotate	Instructions	Contact	About	Welcome admin Logout
Scrape Scrape comments from social Youtube Scraper Twitter Scraper	media							
Italian Streaming time (minutes): 10								×

Figure 11 Συλλογή κειμένων από το Twitter

2.6 Επισημείωση δεδομένων

Για την επισημείωση των δεδομένων και τη μετάβαση στην πλατφόρμα doccano, ο χρήστης πρέπει να επιλέξει "Annotate" από το κεντρικό μενού. Για περισσότερες πληροφορίες μεταβείτε στο Κεφάλαιο 3.

2.7 Εισαγωγή δεδομένων

Για την εισαγωγή νέων δεδομένων, ο χρήστης πρέπει να επιλέξει "Submit" από το κεντρικό μενού. Η εισαγωγή μπορεί να γίνει είτε καταχώρηση-προς-καταχώρηση, είτε μαζικά.

Όσον αφορά την πρώτη μέθοδο, ο χρήστης πρέπει να κάνει εισαγωγή του κειμένου, των μεταδεδομένων (πηγή, γλώσσα, ημερομηνία, επισημείωση ρητορικής μίσους και συναισθήματος, κτλ.), κάνοντας χρήση των κατάλληλων φορμών. Η εισαγωγή μιας καταχώρησης ολοκληρώνεται αφού επιλεγεί το κουμπί "Submit" (Figure 12).

Αν τα δεδομένα ακολουθούν τη μορφή δεδομένων που προτείνεται στο πλαίσιο του έργου PHARM (Παράγραφος 1.5), τότε μπορούν να εισαχθούν στο σύστημα μαζικά, μέσω αρχείου JSON. Ο χρήστης πρέπει να επιλέξει "Batch File", <u>bva επιλέξει το αρχείο από τον H/Y του</u>, και να πατήσει το κουμπί "Submit".

Με την ολοκλήρωση της διαδικασίας, παρουσιάζεται η οθόνη σύνοψης της εισαγωγής (Figure 14). Αυτή η οθόνη δίνει χρήσιμες πληροφορίες σχετικά με τον αριθμό των καταχωρήσεων που επεξεργάστηκαν, ενδεχόμενες διπλότυπες καταγραφές ή κάποιες που υπάρχουν ήδη στη βάση δεδομένων, όπως και από καταχωρήσεις που έχουν προβλήματα μορφοποίησης ή δεν έχουν όλα τα απαραίτητα πεδία.

HARM Interface	Home Search Submit	Scrape Annotate Instructions Contact About	Welcome admin Logout
Submit Submit a single or a batch of	f records to the PHARM (latabase	
Submit single hate-speech related t	ext to the PHARM database		
Enter text here			
			ji.
Insert tags (Optional)			
refugees ×			
Choose text language:		Choose date (Optional):	
English		mm / dd / yyyy	
Choose text source: O Twitter O	YouTube O Facebook Comme	nt O Facebook Post O Website Article O Website Comment	
Insert source url (Optional)			
Submit			

Figure 12 Εισαγωγή μιας καταχώρησης στη βάση δεδομένων

PHARM Interface	Home Search Subm	it Scrape Annotat	e Instructions	Contact About	Welcome admin Logout
Submit Submit a single or a batch of	records to the PHARI	VI database			
Single Text Batch File					
Submit batch hate-speech texts to th Instructions page.	ne PHARM database. The re	ecords should be in a	JSON format. Fo	or the JSON data format	configuration, please refer to the
Browse No file selected.					

Figure 13 Εισαγωγή πολλαπλών καταχωρήσεων στη βάση δεδομένων

ΕΓΧΕΙΡΙΔΙΟ ΧΡΗΣΗΣ

PHARM Interface	Home Search :	Submit Scraper	Annotate	Instructions	Contact	About	Welcome admin Logout
Data succesfully added!							
OK							
60 records were processed in total.							
0 records were added successfully.							
1 not valid records were found and	rejected. Take into c	onsideration that	t the "type",	"text" and "an	notations	" attributes are mandatory.	
0 duplicates were found and rejecte	d.						
59 records were found already in da	tabase.						

Figure 14 Προβολή των αποτελεσμάτων εισαγωγής στη βάση δεδομένων

2.8 Είσοδος στην εφαρμογή

Όταν ο χρήστης προσπαθήσει να μεταβεί σε κάποια λειτουργία που απαιτεί είσοδο στο σύστημα (λειτουργίες συλλογής, επισημείωσης και εισαγωγής δεδομένων) τότε εμφανίζεται η οθόνη εισόδου (Figure 15), στην οποία πρέπει να γίνει εισαγωγή του ονόματος χρήση και του κωδικού πρόσβασης.



Figure 15 Η οθόνη εισόδου επικυρωμένου χρήστη

Για την απόκτηση ενός επικυρωμένου λογαριασμού, ο χρήστης πρέπει να επικοινωνήσει με την ερευνητική ομάδα του έργου μέσω της σχετικής φόρμας επικοινωνίας (επιλέγοντας "Contact" από το κεντρικό μενού).

3 Οδηγίες χρήσης της εφαρμογής doccano

Η εφαρμογή PHARM φιλοξενεί εγκατάσταση της πλατφόρμας doccano, για την εκτέλεση όλων των απαιτούμενων εργασιών επισημείωσης κειμένου.

3.1 Εισαγωγή

To doccano αποτελεί ένα σύστημα διαχείρισης επισημειώσεων για δεδομένα κειμένου. Μπορεί να χρησιμοποιηθεί για την ανάπτυξη συνόλων δεδομένων (datasets) με σκοπό την ταξινόμηση, την ανάλυση συναισθήματος, ή την μετάφραση δεδομένων κειμένου. Στην προκειμένη περίπτωση αξιοποιείται για εργασίες ταξινόμησης. Έτσι, σε κάθε καταχώρηση πρέπει ανατεθούν ετικέτες, ώστε να ενταχθεί στις απαιτούμενες κατηγορίες (κλάσεις).

Ως πρώτο βήμα, ο χρήστης μπορεί να δει το online demo του doccano για ανάλυση συναισθήματος (sentiment analysis): <u>doccano.herokuapp.com/demo/text-classification</u>. Η ίδια διαδικασία επισημείωσης ακολουθείται και για εργασίες ταξινόμησης (classification). Η πρόσβαση στην εγκατάσταση του doccano για το PHARM είναι γίνεται με χρήση του ακόλουθου συνδέσμου: <u>pharm-doccano.usal.es</u>. Φυσικά, υπάρχει η κατάλληλη συντόμευση (υπερσύνδεσμος) και μέσω της εφαρμογής (Figure 4). Τέλος, εν συντομία, η εργασία ενός χρήστη επισημείωσης (annotator) είναι να επιλέξει τις κατάλληλες ετικέτες ("Hate" ή "No Hate", "Positive", "Neutral" ή "Negative") για κάθε καταχώρηση, καθώς προηγείται σε αυτές.

🕞 doccano	Live Demo 🗸	GitHub	Login
Please login to see this page. Username			
<u> </u>			
Password			
Login			
Not registered yet ? Sign up			
			Ų

Figure 16 Η οθόνη εισόδου του εργαλείου doccano

3.2 Επισημείωση καταχώρησης

Αρχικά, ο χρήστης κάνει είσοδο στην πλατφόρμα ως χρήστης επισημείωσης, όπως φαίνεται στο Figure 16. Αμέσως μετά πρέπει να επιλέξει το project στο οποίο έχει κληθεί να συνεισφέρει (Figure 17). Έπειτα, αφού διαβάσει το κείμενο κάθε καταχώρησης, πρέπει να επιλέξει μία εκ των ετικετών "Hate" και "No Hate", ανάλογα αν το περιεχόμενο του μηνύματος χαρακτηρίζεται από ρητορική μίσους ή όχι, και να το σημειώσει ως "Positive", "Neutral" ή "Negative", σύμφωνα με το συναίσθημα που προκαλεί. Είναι υποχρεωτικό να γίνει επιλογή μιας μπλε

εγχειρίδιο χρήσης

και μιας μωβ ετικέτας, καθώς η επισημείωση αφορά δύο διαφορετικά σχήματα ταξινόμησης ταυτόχρονα (Figure 18). Οι μπλε ετικέτες χρησιμοποιούνται για τη σήμανση ή όχι ρητορικής μίσους, ενώ οι μωβ είναι απαραίτητες για τον χαρακτηρισμό του συναισθήματος.

loccano 🔕		Projects Logout	
	Hello, El_annotator_0 hope you are having a great day!		
	1 Projects	All Project 🗸	
	Wave 01 - EL Hate speech and sentiment annotation for Greek texts. updated0 days ago	DocumentClassification	
			Ŷ

Figure 17 Επιλογή project στο doccano

😂 doccano		Projects Log
Search docum	1/13	8
About 13 results (page 1 of 3)	Hate No Hate Positive <u>Neutral</u>	Negative
 RT @MakisHellas1: @thedukeoriginal @rapidis Μια σημείωση δεν 	No Hate × Neutral ×	
RT @thedukeoriginal: Υπουργός Εξωτερικών Ην. Αραβικών Εμιράτ	RT @MakisHellas1: @thedukeoriginal @rapidis είναι μετανάστευση αλλά εισβολή μισθοφόρων μα	Μια σημείωση δεν αχητών του ISIS FSA
RT @oulosP: Germany QUICKLY offered Greece support "in anywa	€ < 1/13	> >
RT @CimbrianMarius: Η Τουρκια διαφωνει με το #ΑΠΕΛΑΣΗ .Στα π		
RT @thedukeoriginal: #ΑΠΕΛΑΣΗ https://t.co/HhHyf2YuzU		



Στις καταχωρήσεις που έχει γίνει επισημείωση εμφανίζεται το εικονίδιο "√". Οι πλοήγηση στις καταχωρήσεις γίνεται με χρήση των κουμπιών "<" και ">".

3.3 Έγκριση επισημείωσης

Για την έγκριση των επισημειώσεων απαιτείται είσοδος στο doccano με χρήση κατάλληλου λογαριασμού χρήστη (approver). Αφού επιλεγεί το project (όπως περιγράφηκε προηγουμένως), η έγκριση μιας επισημείωσης γίνεται με χρήση του κουμπιού έγκρισης, όπως φαίνεται στο Figure 19. Οι εγκεκριμένες καταχωρήσεις σημειώνονται με το εικονίδιο "

I doccano		Projects
Search document	1/13	8
About 13 results (page 1 of 3) Sort by	Hate No Hate Positive Neutral	Negative
 RT @MakisHellas1: @thedukeoriginal @rapidis Μια αυτείρωση δεν 	No Hate 💌 Neutral 💌	
RT @thedukeoriginal: Υπουργός Εξωτερικών Ην. Αραβικών Εμιράτ	RT @MakisHellas1: @thedukeoriginal @ra είναι μετανάστευση αλλά εισβολή μισθοφόρ	pidis Μια σημείωση δεν ων μαχητών του ISIS FSA
RT @oulosP: Germany QUICKLY offered Greece support "in anywa	€ (> >
RT @CimbrianMarius: Η Τουρκια διαφωνει με το #ΑΠΕΛΑΣΗ .Στα π		
RT @thedukeoriginal: #ΑΠΕΛΑΣΗ https://t.co/HhHyf2YuzU		

Figure 19 Έγκριση επισημείωσης στο doccano

PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

ABOUT PHARM (2020-2022)

Preventing Hate against Refugees and Migrants (PHARM)

Migration to Europe has grown in the last years in scale and complexity. The so called 'refugee crisis' and the migratory pressure is particularly acute in southern EU countries as the main entrance to the EU.

The main goal of PHARM project is to monitor and model hate speech against refugees and migrants in Greece, Italy and Spain in order to predict and combat hate crime and also counter its effects using cutting-edge techniques, such as data journalism and narrative persuasion. The activities distributed in 5 coordinated work packages include:

(i) Implementation of a conceptual and methodological common framework for large-scale analysis and detection of hate speech; (ii) Implementation and evaluation of machine learning approaches to model and predict hate crimes against refugees and migrants based on hate speech features; (iii) Survey journalists to understand how they inform and raise awareness about hate speech and how they can help building and disseminating counter-narratives based in data-driven news pieces; (iv) Creation, evaluation and dissemination of counter-narrative fictional stories adapted to different characteristics of citizens using large-scale narrative persuasion.

COORDINATOR

University of Salamanca, Faculty of Social Sciences (ES)

PARTNERS

University of Milan, (IT) Aristotle University of Thessaloniki (GR)

> VISIT: http://pharmproject.usal.es CONTACT US: pharm.project.eu@gmail.com FOLLOW US:

> > @Pharm_project

Pharm_project

Pharm Project

Pharm Project