



ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

**ΑΝΑΚΤΗΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΑΝΑΛΗΘΩΝ
ΔΗΜΟΣΙΕΥΜΑΤΩΝ ΚΑΙ ΨΕΥΔΩΝ
ΙΣΧΥΡΙΣΜΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ
ΔΙΑΔΙΚΤΥΟ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Εισηγητής: Elizaveta Zabalueva, ΔΕ20197

Επιβλέπων: Παύλος Φαφαλιός, Επίκουρος Καθηγητής

©
2024



HELLENIC MEDITERRANEAN UNIVERSITY

**SCHOOL OF MANAGEMENT AND ECONOMICS
SCIENCE**

**DEPARTMENT OF MANAGEMENT SCIENCE AND
TECHNOLOGY**

**RETRIEVAL AND ANALYSIS OF FALSE
PUBLICATIONS AND FALSE CLAIMS ON
THE GREEK INTERNET**

DIPLOMA THESIS

Student: Elizaveta Zabalueva, ΔE20197

Supervisor: Pavlos Fafalios, Assistant Professor

©
2024

Υπεύθυνη Δήλωση : Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του ΕΛ.ΜΕ.ΠΑ.

ΠΕΡΙΛΗΨΗ

Οι ψευδείς ειδήσεις είναι ένα φαινόμενο που έγινε δημοφιλές στις ΗΠΑ κατά τις προεδρικές εκλογές του 2016 και γνωστό παγκοσμίως με την εξάπλωση του Covid-19. Ο όρος αυτός χρησιμοποιείται για να περιγράψει αναληθή δημοσιεύματα, ψευδούς ισχυρισμούς, παραποιήσεις πληροφοριών ή γεγονότων και άλλες σχετικές μορφές παραπληροφόρησης. Από το 2016 και μετά έχει παρατηρηθεί μια τεράστια αύξηση των επιστημονικών και μη δημοσιεύσεων για αυτό το θέμα, ωστόσο οι περισσότερες από αυτές επικεντρώνονται στις ψευδείς ειδήσεις στα αγγλικά. Η παρούσα εργασία τίθεται στις ψευδείς ειδήσεις στο ελληνικό Διαδίκτυο και χωρίζεται σε δυο τμήματα. Στο πρώτο τμήμα, γίνεται η ανάπτυξη του κώδικα που ανακτά τα άρθρα από τον ιστότοπο της Ellhnikahoaxes.gr, όπου αναλύονται οι πιο επίκαιρες ψευδείς ειδήσεις στην Ελλάδα, και επεξεργάζεται ειλημμένα δεδομένα. Στο δεύτερο τμήμα, πραγματοποιείται η ανάλυσή των δεδομένων. Τέλος, παρουσιάζονται τα αποτελέσματα και τα συμπεράσματα για τις ψευδείς ειδήσεις στο ελληνικό Διαδίκτυο και θέτονται προτάσεις για μελλοντική εργασία.

Λέξεις Κλειδιά : ψευδείς ειδήσεις, ανάκτηση δεδομένων, ανάλυση δεδομένων, προγραμματισμός, Java

ABSTRACT

Fake news is a phenomenon that became popular in the US during the 2016 presidential election and became known worldwide during the spread of Covid-19. This term is used to describe false reports, false claims, misrepresentations of information or facts, and other relevant forms of disinformation. Since 2016 there has been a huge increase in scientific and non-scientific publications on this topic, however most of them focus on fake news in English. This paper is about fake news on the Greek Internet, and it is divided into two parts. In the first part, the development of the code that helps retrieve the articles from the EllhnikaHoaxes.gr website, where the most current fake news in Greece is analyzed, and process the retrieved data. In the second part, the analysis of the data is conducted. Finally, the results and conclusions about fake news on the Greek Internet are presented and suggestions for future work are made.

Key Words: fake news, data retrieval, data analysis, programming, Java

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΛΙΣΤΑ ΠΙΝΑΚΩΝ	vi
ΛΙΣΤΑ ΕΙΚΟΝΩΝ	vii
ΛΙΣΤΑ ΣΧΗΜΑΤΩΝ	xiii
ΛΙΣΤΑ ΔΙΑΓΡΑΜΜΑΤΩΝ	xiv
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ	xix
ΕΥΧΑΡΙΣΤΙΕΣ	xx
	ΚΕΦΑΛΑΙΟ 1
	1
ΕΙΣΑΓΩΓΗ	1
	ΚΕΦΑΛΑΙΟ 2
	3
ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗΣ	3
2.1 ΙΣΤΟΡΙΚΟ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ	3
2.1.1 ΕΝΝΟΙΑ ΤΟΥ ΟΡΟΥ «ΨΕΥΔΕΙΣ ΕΙΔΗΣΕΙΣ»	3
2.1.2 ΕΜΦΑΝΙΣΗ ΦΑΙΝΟΜΕΝΟΥ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ	3
2.1.3 ΣΗΜΕΡΙΝΗ ΕΠΙΡΡΟΗ ΤΟΥ ΔΙΑΔΙΚΤΥΟΥ ΣΤΗ ΔΙΑΔΟΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ	4
2.2 ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΤΥΠΟΙ	6
2.3 ΕΥΡΟΣ ΠΡΟΒΛΗΜΑΤΟΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ	12
2.3.1 ΣΤΑΤΙΣΤΙΚΑ ΔΕΔΟΜΕΝΑ	12
2.3.2 ΕΜΠΟΔΙΑ ΕΝΤΟΠΙΣΜΟΥ ΚΑΙ ΤΕΡΜΤΙΣΜΟΥ ΔΙΑΔΟΣΗΣ	16
2.4 ΣΥΝΕΠΕΙΕΣ	16
2.5 ΤΡΟΠΟΙ ΑΝΙΧΝΕΥΣΗΣ	18
2.5.1 ΧΕΙΡΩΝΑΚΤΙΚΗ Ή ΠΡΑΚΤΙΚΗ ΑΝΙΧΝΕΥΣΗ	20
2.5.2 ΑΝΙΧΝΕΥΣΗ ΜΕ ΤΗ ΧΡΗΣΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ Ή ΥΠΟΛΟΓΙΣΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ	20
	ΚΕΦΑΛΑΙΟ 3
	21
ΕΛΕΓΧΟΣ ΕΓΚΥΡΟΤΗΤΑΣ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ	21
3.1 ΥΠΗΡΕΣΙΕΣ ΚΑΙ ΟΡΓΑΝΙΣΜΟΙ	21
3.2 ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΓΑΣΙΕΣ, ΣΥΣΤΗΜΑΤΑ ΚΑΙ DATASETS	23
	ΚΕΦΑΛΑΙΟ 4
	27
Η ΥΠΗΡΕΣΙΑ ELLINIKAHOAXES.GR	27
4.1 ΙΔΡΥΣΗ ΚΑΙ ΛΕΙΤΟΥΡΓΙΑ	27

4.2	ΧΡΗΜΑΤΟΔΟΤΗΣΗ	28
4.3	ΔΙΑΔΙΚΑΣΙΑ ΕΠΙΛΟΓΗΣ ΚΑΙ ΕΞΕΤΑΣΗΣ ΕΙΔΗΣΕΩΝ	28
4.4	ΜΕΘΟΛΟΓΙΑ ΕΛΕΓΧΟΥ ΕΥΚΥΡΟΤΗΤΑΣ ΜΙΑΣ ΕΙΔΗΣΗΣ	29
4.5	ΧΡΗΣΙΜΟΠΟΙΗΜΕΝΕΣ ΚΑΤΗΓΟΡΙΕΣ	30
	ΚΕΦΑΛΑΙΟ 5	35
	ΑΝΑΛΥΣΗ ΤΩΝ ΙΣΥΡΙΣΜΩΝ ΣΤΟΝ ELLINKAHOAXES.GR	35
5.1	ΣΚΟΠΟΣ	35
5.2	ΜΕΘΟΔΟΛΟΓΙΑ	35
5.3	ΑΝΑΚΤΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟΝ ΙΣΤΟΤΟΠΟ	36
5.3.1	ΑΝΑΚΤΗΣΗ ΤΩΝ ΔΙΕΥΘΥΝΣΕΩΝ ΤΩΝ ΑΡΘΡΩΝ	37
5.3.2	ΑΝΑΚΤΗΣΗ ΤΩΝ ΑΠΑΡΑΙΤΗΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΩΝ ΣΥΝΔΕΣΜΩΝ	47
5.4	ΤΑ ΔΕΔΟΜΕΝΑ	75
5.5	ΠΡΟΕΤΟΙΜΑΣΙΑ ΑΝΑΛΥΣΗΣ	81
5.5.1	ΕΞΑΓΩΓΗ ΗΜΕΡΟΜΗΝΙΩΝ ΔΗΜΟΣΙΕΥΜΑΤΟΣ	81
5.5.2	ΕΞΑΓΩΓΗ ΚΑΤΗΓΟΡΙΩΝ	86
5.5.3	ΕΞΑΓΩΓΗ ΛΕΞΕΩΝ – ΚΛΕΙΔΙΩΝ ΑΠΟ ΙΣΧΥΡΙΣΜΟΥΣ	93
5.5.4	ΕΞΑΓΩΓΗ ΤΩΝ ΑΡΘΡΩΝ ΠΟΥ ΑΦΟΡΟΥΝ ΕΝΑ ΣΥΓΚΕΚΡΙΜΕΝΟ ΘΕΜΑ 103	
5.6	ΑΠΟΤΕΛΕΣΜΑΤΑ	109
5.6.1	ΟΙ ΚΑΤΗΓΟΡΙΕΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ ΔΙΑΔΙΚΤΥΟ	110
5.6.2	ΤΑ ΘΕΜΑΤΑ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ ΔΙΑΔΙΚΤΥΟ 120	
5.6.3	ΠΕΡΙΟΔΟΙ ΕΞΑΡΣΗΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ ΔΙΑΔΙΚΤΥΟ	135
	ΚΕΦΑΛΑΙΟ 6	145
	ΕΠΙΛΟΓΟΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	145
6.1	ΕΠΙΛΟΓΟΣ	145
6.2	ΙΔΕΕΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	146
	ΒΙΒΛΙΟΓΡΑΦΙΑ	148
7.1	ΞΕΝΟΓΛΩΣΣΗ	148
7.2	ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ	152
	ΠΑΡΑΡΤΗΜΑ Α	153

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

Πίνακας 2.2.1 Κατηγορίες ψευδών ειδήσεων.....	10
Πίνακας 3.2.1 Προτεινόμενες προσεγγίσεις ανίχνευσης ψευδών ειδήσεων	24
Πίνακας 4.5.1 Κατηγορίες που χρησιμοποιούνται από την υπηρεσία EllhnikaHoaxes.gr.....	31
Πίνακας 5.4.1 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (σε αριθμούς).....	77
Πίνακας 5.4.2 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (σε ποσοστά)	78
Πίνακας 5.5.1 Δεδομένα, αποθηκευμένα στο αρχείο keywordsDictionary.csv ως στοιχεία του HashMap dict	101
Πίνακας 5.5.2 Στοιχεία του HashMap themeHashMap	105
Πίνακας 5.5.3 Σχετική δομή του HashMap catHashMap	109
Πίνακας 5.5.4 Σχετική δομή του HashMap dateHashMap	109

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

Εικόνα 3.2.1 Επιστημονικά δημοσιεύματα πριν το 2016, που περιλαμβάνουν τον όρο «fake news»	23
Εικόνα 3.2.2 Επιστημονικά δημοσιεύματα από το 2016 που περιλαμβάνουν τον όρο «fake news»	23
Εικόνα 5.3.1 Μέθοδος <code>getOldLinks()</code> της κλάσης <code>LinkRetrieval</code>	38
Εικόνα 5.3.2 Σελιδοαρίθμηση στον <code>EllhnikaHoaxes.gr</code>	39
Εικόνα 5.3.3 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (1).....	39
Εικόνα 5.3.4 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (2).....	39
Εικόνα 5.3.5 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (3).....	39
Εικόνα 5.3.6 Μέθοδος <code>getLastPageNum()</code> της κλάσης <code>LinkRetrieval</code>	40
Εικόνα 5.3.7 Επεξεργασία του αριθμού της τελευταίας σελίδας (<code>maxPage</code>)	40
Εικόνα 5.3.8 Μέθοδος <code>getURLContent()</code> της κλάσης <code>LinkRetrieval</code>	41
Εικόνα 5.3.9 Σύνδεσμοι των άρθρων στον <code>EllhnikaHoaxes.gr</code>	42
Εικόνα 5.3.10 Ανάκτηση συνδέσμου.....	42
Εικόνα 5.3.11 Διαχείριση του προβλήματος διπλών εγγραφών	43
Εικόνα 5.3.12 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (1).....	44
Εικόνα 5.3.13 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (2).....	44
Εικόνα 5.3.14 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (3).....	45
Εικόνα 5.3.15 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (4).....	45
Εικόνα 5.3.16 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (5).....	46
Εικόνα 5.3.17 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (6).....	46
Εικόνα 5.3.18 Μέθοδος <code>addLinks()</code> της κλάσης <code>LinkRetrieval</code> (1).....	46
Εικόνα 5.3.19 Μέθοδος <code>addLinks()</code> της κλάσης <code>LinkRetrieval</code> (2).....	47
Εικόνα 5.3.20 <code>HashSet</code> και <code>HashMap</code> για τη μετάφραση των συνδέσμων των εικόνων στο HTML-κώδικα σε μια συμβολοσειρά, εδώ: «ΚΙΝΔΥΝΟΛΟΓΙΑ».....	48
Εικόνα 5.3.21 <code>HashMap</code> για τη μετάφραση των συνδέσμων των κατηγοριών στο HTML-κώδικα σε μια συμβολοσειρά, εδώ: «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ», «ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ» και «ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ»	48
Εικόνα 5.3.22 <code>HashMap</code> για την αλλαγή μορφής των ημερομηνιών	48
Εικόνα 5.3.23 Ανάκτηση ή δημιουργία του αρχείου <code>fakeLinksData.csv</code>	49
Εικόνα 5.3.24 <code>HashSet</code> <code>validCatVerdictValues</code>	50

Εικόνα 5.3.25 Μέθοδος processNewLink() της κλάσης Link (1)	50
Εικόνα 5.3.26 Μέθοδος processNewLink() της κλάσης Link (2)	52
Εικόνα 5.3.27 Μέθοδος processNewLink() της κλάσης Link (3)	53
Εικόνα 5.3.28 Βρόγχος, στον οποίο καλείται η μέθοδος processNewLink() της κλάσης Link	53
Εικόνα 5.3.29 Κατηγορία «HOAXES» πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου)	55
Εικόνα 5.3.30 Κατηγορία του άρθρου , που βρίσκεται πάνω από τον τίτλο (σε μορφή κειμένου), στο HTML-κώδικα (1)	55
Εικόνα 5.3.31 Κατηγορία του άρθρου , που βρίσκεται πάνω από τον τίτλο (σε μορφή κειμένου), στο HTML-κώδικα (2)	55
Εικόνα 5.3.32 Κατηγορία του άρθρου , που βρίσκεται πάνω από τον τίτλο, (σε μορφή κειμένου) στο HTML-κώδικα (3)	56
Εικόνα 5.3.33 Εξαγωγή της κατηγορίας, που βρίσκεται πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου)	56
Εικόνα 5.3.34 Κατηγορία «ΨΕΥΔΕΣ» σε μορφή εικόνας	56
Εικόνα 5.3.35 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (1)	57
Εικόνα 5.3.36 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (2)	57
Εικόνα 5.3.37 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (3)	57
Εικόνα 5.3.38 Εξαγωγή της κατηγορίας σε μορφή εικόνας (1).....	58
Εικόνα 5.3.39 Εξαγωγή της κατηγορίας σε μορφή εικόνας (2).....	59
Εικόνα 5.3.40 Εξαγωγή της κατηγορίας σε μορφή εικόνας (3).....	60
Εικόνα 5.3.41 Εξαγωγή της κατηγορίας σε μορφή εικόνας (4).....	60
Εικόνα 5.3.42 Τίτλος	61
Εικόνα 5.3.43 Τίτλος στο HTML-κώδικα (1).....	61
Εικόνα 5.3.44 Τίτλος στο HTML-κώδικα (2).....	61
Εικόνα 5.3.45 Εξαγωγή του τίτλου.....	61
Εικόνα 5.3.46 Ημερομηνία	62
Εικόνα 5.3.47 Ημερομηνία στο HTML-κώδικα (1)	62
Εικόνα 5.3.48 Ημερομηνία στο HTML-κώδικα (2)	62
Εικόνα 5.3.49 Εξαγωγή της ημερομηνίας	62
Εικόνα 5.3.50 Συντάκτης (ελεγκτής) του άρθρου	63

Εικόνα 5.3.51 Όνομα του συντάκτη στο HTML-κώδικα (1)	63
Εικόνα 5.3.52 Όνομα του συντάκτη στο HTML-κώδικα (2)	63
Εικόνα 5.3.53 Εξαγωγή του ονόματος του συντάκτη.....	63
Εικόνα 5.3.54 Παράδειγμα άρθρου χωρίς τα στοιχεία «Ισχυρισμός» και «Συμπέρασμα»	64
Εικόνα 5.3.55 Πρώτη δομή της σελίδας.....	65
Εικόνα 5.3.56 Δεύτερη δομή της σελίδας.....	65
Εικόνα 5.3.57 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (1)	66
Εικόνα 5.3.58 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (2)	66
Εικόνα 5.3.59 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (3)	66
Εικόνα 5.3.60 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (1).....	67
Εικόνα 5.3.61 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (2).....	67
Εικόνα 5.3.62 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (3).....	67
Εικόνα 5.3.63 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (1).....	68
Εικόνα 5.3.64 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (2).....	68
Εικόνα 5.3.65 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (3).....	68
Εικόνα 5.3.66 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (1)	69
Εικόνα 5.3.67 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (2)	69
Εικόνα 5.3.68 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (3)	69
Εικόνα 5.3.69 Εξαγωγή του ισχυρισμού, πρώτη περίπτωση.....	70
Εικόνα 5.3.70 Εξαγωγή του ισχυρισμού, δεύτερη περίπτωση	70
Εικόνα 5.3.71 Εξαγωγή του συμπεράσματος, πρώτη περίπτωση.....	71
Εικόνα 5.3.72 Εξαγωγή του συμπεράσματος, δεύτερη περίπτωση.....	71
Εικόνα 5.3.73 Αρχείο fakeLinksData.csv στο MS Excel	72
Εικόνα 5.3.74 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (1).....	73
Εικόνα 5.3.75 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (2).....	74
Εικόνα 5.3.76 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (3).....	74
Εικόνα 5.4.1 Αποκωδικοποιημένα δεδομένα, που αποκτήθηκαν με τη βοήθεια του κώδικα στη Java, έτοιμα για περαιτέρω επεξεργασία και ανάλυση.....	75
Εικόνα 5.4.2 Άρθρο με πολλούς ισχυρισμούς και συμπεράσματα στον ιστότοπο EllhnikaHoaxes.gr.....	80
Εικόνα 5.4.3 Άρθρο με πολλούς ισχυρισμούς και συμπεράσματα μετά την ανάκτηση.....	81
Εικόνα 5.5.1 Μετατροπή ημερομηνίας από «dd-mm-yyyy» σε «mmmm-yyyy»	82

Εικόνα 5.5.2 Υπολογισμός των άρθρων που δημοσιεύτηκαν ένα συγκεκριμένο μήνα (Μάιος, 2023)	82
Εικόνα 5.5.3 Φιλτράρισμα των άρθρων ανά έτος (2020)	83
Εικόνα 5.5.4 Φιλτράρισμα των άρθρων ανά περίοδο πάνω από ένα έτος (2020 - τώρα)	84
Εικόνα 5.5.5 Διαγραφή κενών γραμμών (1)	85
Εικόνα 5.5.6 Διαγραφή κενών γραμμών (2)	85
Εικόνα 5.5.7 Διαγραφή κενών γραμμών (3)	86
Εικόνα 5.5.8 Αποτέλεσμα διαγραφής κενών γραμμών	86
Εικόνα 5.5.9 Άρθρο με πολλούς ισχυρισμούς πριν από την επεξεργασία των κατηγοριών του	87
Εικόνα 5.5.10 Άρθρο με πολλούς ισχυρισμούς μετά την επεξεργασία των κατηγοριών του	88
Εικόνα 5.5.11 Επεξεργασία των κατηγοριών (1)	88
Εικόνα 5.5.12 Επεξεργασία των κατηγοριών (2)	89
Εικόνα 5.5.13 Επεξεργασία των κατηγοριών (3)	90
Εικόνα 5.5.14 Επεξεργασία των κατηγοριών (4)	90
Εικόνα 5.5.15 Επεξεργασία των κατηγοριών (5)	91
Εικόνα 5.5.16 Επεξεργασία των κατηγοριών (6)	91
Εικόνα 5.5.17 Επεξεργασία των κατηγοριών (7)	92
Εικόνα 5.5.18 Επεξεργασία των κατηγοριών (8)	92
Εικόνα 5.5.19 Επεξεργασία των κατηγοριών (9)	93
Εικόνα 5.5.20 Αρχείο keywordsDictionary.csv	94
Εικόνα 5.5.21 Επεξεργασία των ισχυρισμών (1)	95
Εικόνα 5.5.22 Επεξεργασία των ισχυρισμών (2)	95
Εικόνα 5.5.23 Επεξεργασία των ισχυρισμών (3)	96
Εικόνα 5.5.24 Επεξεργασία των ισχυρισμών (4)	97
Εικόνα 5.5.25 Επεξεργασία των ισχυρισμών (5)	97
Εικόνα 5.5.26 Επεξεργασία των ισχυρισμών (6)	98
Εικόνα 5.5.27 Επεξεργασία των ισχυρισμών (7)	98
Εικόνα 5.5.28 Επεξεργασία των ισχυρισμών (8)	99
Εικόνα 5.5.29 Μέθοδος chooseFile() της κλάσης DictionaryReader	100
Εικόνα 5.5.30 Μέθοδος retrieveDictionary() της κλάσης DictionaryReader	101
Εικόνα 5.5.31 Μέθοδος changeKeywords() της κλάσης DictionaryReader	102

Εικόνα 5.5.32 Αρχείο claimDate.csv με όλους τους ισχυρισμούς και τις ημερομηνίες δημοσίευσής τους.....	103
Εικόνα 5.5.33 Αρχείο claimCat.csv με όλους τους ισχυρισμούς και οι κατηγορίες τους.....	104
Εικόνα 5.5.34 Δημιουργία των HashMap των κατηγοριών Covid, Vaccination, Russia, Ukraine, Greece, Turkey, Politics, the USA.....	105
Εικόνα 5.5.35 Μέθοδος extractClaimDate().....	107
Εικόνα 5.5.36 Μέθοδος extractClaimCategories()	108
Εικόνα 5.6.1 Word-cloud των ισχυρισμών στον Ellhnikahoaxes.gr	122
Εικόνα 9.1 Κλάση LinkRetrieval (1)	157
Εικόνα 9.2 Κλάση LinkRetrieval (2)	158
Εικόνα 9.3 Κλάση LinkRetrieval (3)	159
Εικόνα 9.4 Κλάση LinkRetrieval (4)	160
Εικόνα 9.5 Κλάση PageDataRetrieval	161
Εικόνα 9.6 Κλάση PageDataRetrieval (2)	162
Εικόνα 9.7 Κλάση PageDataRetrieval (3)	163
Εικόνα 9.8 Κλάση PageDataRetrieval (4)	164
Εικόνα 9.9 Κλάση PageDataRetrieval (5)	165
Εικόνα 9.10 Κλάση PageDataRetrieval (6)	166
Εικόνα 9.11 Κλάση PageDataRetrieval (7)	167
Εικόνα 9.12 Κλάση PageDataRetrieval (8)	167
Εικόνα 9.13 Κλάση PageDataRetrieval (9)	168
Εικόνα 9.14 Κλάση PageDataRetrieval (10)	169
Εικόνα 9.15 Κλάση PageDataRetrieval (11)	170
Εικόνα 9.16 Κλάση Link (1).....	171
Εικόνα 9.17 Κλάση Link (2).....	172
Εικόνα 9.18 Κλάση Link (3).....	173
Εικόνα 9.19 Κλάση Link (4).....	174
Εικόνα 9.20 Κλάση Link (5).....	175
Εικόνα 9.21 Κλάση Link (6).....	176
Εικόνα 9.22 Κλάση Link (7).....	177
Εικόνα 9.23 Κλάση Link (8).....	178
Εικόνα 9.24 Κλάση Link (9).....	179

Εικόνα 9.25 Κλάση Link (10).....	180
Εικόνα 9.26 Κλάση Link (11).....	181
Εικόνα 9.27 Κλάση MinorClass (1).....	182
Εικόνα 9.28 Κλάση MinorClass (2).....	183
Εικόνα 9.29 Κλάση DictionaryReader (1).....	184
Εικόνα 9.30 Κλάση DictionaryReader (2).....	185
Εικόνα 9.31 Κλάση DictionaryReader (3).....	186
Εικόνα 9.32 Κλάση SubjectSelectionMenu.....	186
Εικόνα 9.33 Κλάση ClaimDateRetrieval (1)	187
Εικόνα 9.34 Κλάση ClaimDateRetrieval (2)	188
Εικόνα 9.35 Κλάση ClaimDateRetrieval (3)	189
Εικόνα 9.36 Κλάση ClaimDateRetrieval (4)	190
Εικόνα 9.37 Κλάση ClaimCategoryRetrieval (1)	191
Εικόνα 9.38 Κλάση ClaimCategoryRetrieval (2)	192
Εικόνα 9.39 Κλάση ClaimCategoryRetrieval (3)	193
Εικόνα 9.40 Κλάση ClaimCategoryRetrieval (4)	194

ΛΙΣΤΑ ΣΧΗΜΑΤΩΝ

Σχήμα 5.2.1 Διαδικασία εξαγωγής, επεξεργασίας και ανάλυσης δεδομένων από τον ιστότοπο EllhnikaHoaxes.gr.....	36
---	----

ΛΙΣΤΑ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 5.4.1 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (γραφικά).....	79
Διάγραμμα 5.6.1 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που δημοσιεύτηκαν το 2013 - 2023	111
Διάγραμμα 5.6.2 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που δημοσιεύτηκαν το 2020 - 2023	112
Διάγραμμα 5.6.3 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν το COVID-19 και δημοσιεύτηκαν το 2013 - 2023	113
Διάγραμμα 5.6.4 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν το COVID-19 και δημοσιεύτηκαν το 2013 - 2023	113
Διάγραμμα 5.6.5 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τον εμβολιασμό και δημοσιεύτηκαν το 2013 - 2023	114
Διάγραμμα 5.6.6 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τον εμβολιασμό και δημοσιεύτηκαν το 2013 - 2023	114
Διάγραμμα 5.6.7 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τη Ρωσία και δημοσιεύτηκαν το 2013 - 2023	115
Διάγραμμα 5.6.8 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τη Ρωσία και δημοσιεύτηκαν το 2013 - 2023	115
Διάγραμμα 5.6.9 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ουκρανία και δημοσιεύτηκαν το 2013 - 2023	116
Διάγραμμα 5.6.10 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ουκρανία και δημοσιεύτηκαν το 2013 - 2023	116
Διάγραμμα 5.6.11 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ελλάδα και δημοσιεύτηκαν το 2013 - 2023	117
Διάγραμμα 5.6.12 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ελλάδα και δημοσιεύτηκαν το 2013 - 2023	117
Διάγραμμα 5.6.13 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Τουρκία και δημοσιεύτηκαν το 2013 - 2023	118
Διάγραμμα 5.6.14 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Τουρκία και δημοσιεύτηκαν το 2013 - 2023	118
Διάγραμμα 5.6.15 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τις ΗΠΑ και δημοσιεύτηκαν το 2013 - 2023	119

Διάγραμμα 5.6.16 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τις ΗΠΑ και δημοσιεύτηκαν το 2013 - 2023.....	119
Διάγραμμα 5.6.17 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την πολιτική και δημοσιεύτηκαν το 2013 - 2023	120
Διάγραμμα 5.6.18 Κατανομή των κατηγοριών των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την πολιτική και δημοσιεύτηκαν το 2013 - 2023	120
Διάγραμμα 5.6.19 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2013 - 2023.....	121
Διάγραμμα 5.6.20 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2020 - 2023.....	122
Διάγραμμα 5.6.21 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2020	123
Διάγραμμα 5.6.22 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2021	124
Διάγραμμα 5.6.23 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2022	124
Διάγραμμα 5.6.24 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2023	125
Διάγραμμα 5.6.25 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανακοινώσεις» της EllhnikaHoaxes.gr το 2013 - 2023.....	126
Διάγραμμα 5.6.26 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανακυκλωμένη είδηση» της EllhnikaHoaxes.gr το 2013 - 2023	127
Διάγραμμα 5.6.27 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανεπιβεβαίωτο» της EllhnikaHoaxes.gr το 2013 - 2023.....	127
Διάγραμμα 5.6.28 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Απάτη» της EllhnikaHoaxes.gr το 2013 - 2023.....	127
Διάγραμμα 5.6.29 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Clickbait» της EllhnikaHoaxes.gr το 2013 - 2023.....	128
Διάγραμμα 5.6.30 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Διάφορα» της EllhnikaHoaxes.gr το 2013 - 2023	128
Διάγραμμα 5.6.31 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Επικαιρότητα» της EllhnikaHoaxes.gr το 2013 - 2023	128

Διάγραμμα 5.6.32 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Έρευνες» της EllhnikaHoaxes.gr το 2013 - 2023	129
Διάγραμμα 5.6.33 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Fake news» της EllhnikaHoaxes.gr το 2013 - 2023	129
Διάγραμμα 5.6.34 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Hoaxes» της EllhnikaHoaxes.gr το 2013 - 2023	129
Διάγραμμα 5.6.35 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Κινδυνολογία» της EllhnikaHoaxes.gr το 2013 - 2023	130
Διάγραμμα 5.6.36 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Λείπει θεματικό περιεχόμενο» της EllhnikaHoaxes.gr το 2013 - 2023.....	130
Διάγραμμα 5.6.37 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Like farming» της EllhnikaHoaxes.gr το 2013 - 2023	130
Διάγραμμα 5.6.38 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Μίξη γεγονότων και παραποιήσεων» της EllhnikaHoaxes.gr το 2013 - 2023	131
Διάγραμμα 5.6.39 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Μύθοι και ιστορία» της EllhnikaHoaxes.gr το 2013 - 2023.....	131
Διάγραμμα 5.6.40 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Παραπληροφόρηση» της EllhnikaHoaxes.gr το 2013 - 2023.....	131
Διάγραμμα 5.6.41 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδές» της EllhnikaHoaxes.gr το 2013 - 2023.....	132
Διάγραμμα 5.6.42 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδείς ισχυρισμός» της EllhnikaHoaxes.gr το 2013 - 2023.....	132
Διάγραμμα 5.6.43 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδοεπιστήμη» της EllhnikaHoaxes.gr το 2013 - 2023	132
Διάγραμμα 5.6.44 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψεύτικα ρητά και αποφθέγματα» της EllhnikaHoaxes.gr το 2013 - 2023	133
Διάγραμμα 5.6.45 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Σάτιρα» της EllhnikaHoaxes.gr το 2013 - 2023	133
Διάγραμμα 5.6.46 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Συνωμοσιολογία» της EllhnikaHoaxes.gr το 2013 - 2023	133
Διάγραμμα 5.6.47 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Τροποποιημένη εικόνα» της EllhnikaHoaxes.gr το 2013 - 2023	134

Διάγραμμα 5.6.48 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερα στα άρθρα της κατηγορίας «Τροποποιημένο βίντεο» της EllhnikaHoaxes.gr το 2013 - 2023	134
Διάγραμμα 5.6.49 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερα στα άρθρα της κατηγορίας «Υγεία» της EllhnikaHoaxes.gr το 2013 - 2023	134
Διάγραμμα 5.6.50 Αριθμός των δημοσιευμένων άρθρων της EllhnikaHoaxes.gr ανά μήνα το 2013 - 2023	135
Διάγραμμα 5.6.51 Μήνες με τα περισσότερα δημοσιευμένα άρθρα στον EllhnikaHoaxes.gr το 2013 - 2023.....	136
Διάγραμμα 5.6.52 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν το COVID-19, ανά μήνα από το 2020 έως 2023	137
Διάγραμμα 5.6.53 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τον εμβολιασμό, ανά μήνα από το 2020 έως 2023	138
Διάγραμμα 5.6.54 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τη Ρωσία, ανά μήνα από το 2020 έως 2023	139
Διάγραμμα 5.6.55 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τη Ρωσία, ανά μήνα το από τον Ιανουάριο 2022 έως το Μάιο 2023.....	139
Διάγραμμα 5.6.56 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ουκρανία, ανά μήνα από το 2020 έως 2023	140
Διάγραμμα 5.6.57 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ουκρανία, ανά μήνα από τον Ιανουάριο 2022 έως το Μάιο 2023	141
Διάγραμμα 5.6.58 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ελλάδα, ανά μήνα από το 2020 έως 2023.....	141
Διάγραμμα 5.6.59 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Τουρκία, ανά μήνα από το 2020 έως 2023.....	142
Διάγραμμα 5.6.60 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τις ΗΠΑ, ανά μήνα από το 2020 έως 2023.....	143
Διάγραμμα 5.6.61 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την πολιτική, ανά μήνα από το 2020 έως 2023.....	143
Διάγραμμα 5.6.62 Αριθμός των άρθρων στον EllhnikaHoaxes.gr ανά θέμα ανά μήνα από το 2020 έως 2023.....	144
Διάγραμμα 8.1 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν το Covid-19, ανά μήνα από το 2013 έως 2023.....	153

Διάγραμμα 8.2 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τον εμβολιασμό, ανά μήνα από το 2013 έως 2023	153
Διάγραμμα 8.3 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τη Ρωσία, ανά μήνα από το 2013 έως 2023	154
Διάγραμμα 8.4 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ουκρανία, ανά μήνα από το 2013 έως 2023	154
Διάγραμμα 8.5 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Ελλάδα, ανά μήνα από το 2013 έως 2023	155
Διάγραμμα 8.6 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την Τουρκία, ανά μήνα από το 2013 έως 2023	155
Διάγραμμα 8.7 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν τις ΗΠΑ, ανά μήνα από το 2013 έως 2023	156
Διάγραμμα 8.8 Αριθμός των άρθρων στον EllhnikaHoaxes.gr, που αφορούν την πολιτική, ανά μήνα από το 2013 έως 2023	156

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

ΒΔ – η Βάση Δεδομένων

ΗΠΑ – οι Ηνωμένες Πολιτείες της Αμερικής

ΕΕ – η Ευρωπαϊκή Ένωση

ΕΕ – το Ερώτημα Εργασίας

ΜΜΕ – τα Μέσα Μαζικής Ενημέρωσης

CSV – Comma Separated Value

HTML – Hypertext Markup Language

HTTP – Hypertext Transfer Protocol

IP – Internet Protocol

XLSX – Microsoft Excel Spreadsheet Extended

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας την πτυχιακή μου εργασία και τις σπουδές μου στο Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας του ΕΛΜΕΠΑ, θα ήθελα να εκφράσω ένα μεγάλο ευχαριστώ στους καθηγητές του τμήματος, χάρη στους οποίους αντιλήφθηκα τι είναι το αληθινό μου ενδιαφέρον: τον **κ. Παναγιωτάκη Κωνσταντίνο** (Δομημένος Προγραμματισμός), τον **κ. Βάρδα Ιωάννη** (Μαθηματική Ανάλυση, Πιθανότητες και Στατιστική, Εφαρμοσμένη Στατιστική, Μέθοδοι Βελτιστοποίησης), τον **κ. Φαφαλιό Παύλο** (Αλγόριθμοι και Δομές Δεδομένων, Τεχνολογία Λογισμικού, Ηλεκτρονικό Εμπόριο και Εφαρμογές Διαδικτύου), τον **κ. Βασιλειάδη Γεώργιο** (Αντικειμενοστραφής Προγραμματισμός, Ανάπτυξη Διαδικτυακών και Νεφούπολογιστικών Εφαρμογών), τον **κ. Καπανταϊδάκη Ιωάννη** (Πληροφοριακά Συστήματα, Σχεδίαση και Ανάπτυξη Κινητών Εφαρμογών) και τον **κ. Παπουτσάκη Κωνσταντίνο** (Ανάλυση Χρονοσειρών).

Στον **κ. Φαφαλιό Παύλο** θέλω να εκφράσω μια ιδιαίτερη ευχαριστία ως στον επιβλέπον μου καθηγητή, που με καθοδηγούσε στη συγγραφή αυτής της εργασίας και ήταν πάντα διαθέσιμος για οποιαδήποτε μου απορία.

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Οι ψευδείς ισχυρισμοί και τα αναληθή δημοσιεύματα που διαδίδονται ως ειδήσεις είναι ευρέως γνωστά ως ψευδείς ειδήσεις. (Στην συνέχεια για την ευκολία θα χρησιμοποιείται αυτός ο όρος.) Το φαινόμενο των ψευδών ειδήσεων δεν είναι ένα νέο φαινόμενο που εμφανίστηκε λόγω ραγδαίας ανάπτυξης της τεχνολογίας και του Διαδικτύου. Το θέμα αυτό εμφανίστηκε εκατοντάδες χρόνια πριν αλλά έγινε πραγματική απειλή για την κοινωνία μας μερικά χρόνια πριν. Πλέον, είναι ένα παγκόσμιο πρόβλημα που επηρεάζει όλους μας.

Πριν το 2016 ο ορός «fake news» (ελλην. ψευδείς ειδήσεις) αναφέρθηκε μόνο σε περίπου 6000 επιστημονικές μελέτες, ενώ από το 2016 μέχρι σήμερα ο αριθμός αυτός ξεπερνάει τις 180 000. Οι εργασίες στο θέμα των ψευδών ειδήσεων χωρίζονται σε ανάπτυξη εργαλείων ανίχνευσης ψευδών ειδήσεων, σύγκριση και ανάλυση ψευδών ειδήσεων ως φαινόμενο.

Αυτή η εργασία εστιάζεται στην ανάκτηση και ανάλυση ψευδών ειδήσεων στο ελληνικό Διαδίκτυο κατά τη διάρκεια μιας δεκαετίας (2013-2023). Οι παρόμοιες εργασίες έχουν πραγματοποιηθεί από διάφορους μελετητές από όλο τον κόσμο. Go και Lee (2020)¹ αναλύουν τις ψευδείς ειδήσεις στη Βόρεια Κορέα κατά τη διάρκεια των προεδρικών εκλογών του 2017, Awan, Yasin, Nobanee, Ali, Shahzad, Nabeel, Zain και Shahzad (2021)² εξετάζουν τις ψευδείς ειδήσεις στις ΗΠΑ προς τα πιο συχνά θέματα παραπληροφόρησης, Gutiérrez-Coba, Coba-Gutiérrez και Gómez-Díaz (2020)³ αναλύουν το φαινόμενο του ψευδών ειδήσεων για το COVID-19 προς τις κατηγορίες παραπληροφόρησης και τη μορφή των ψευδών ειδήσεων (βίντεο, εικόνα, κείμενο κ.α.) στην Κολομβία, την Περού, την Αργεντινή, το Εκουαδόρ, την Ισπανία και τη Βολιβία. Nakon, Alam, Shaar, Da San Martino και Zhang (2021)⁴ ερευνούν τις ψευδείς ειδήσεις για τον εμβολιασμό για το COVID-19 στο Κατάρ. Berduygina, Vladimirova και Chernyaeva (2019)⁵ αναλύουν τις ψευδείς ειδήσεις προς τα πιο συχνά θέματα παραπληροφόρησης και τις πηγές της διάδοσης των ψευδών ειδήσεων. Fourney, Racz, Ranade,

¹ <https://koreascience.kr/article/JAKO202022449681023.pdf> (προσπελάστηκε στις 26/01/2024)

² <https://www.mdpi.com/2079-9292/10/19/2326> (προσπελάστηκε στις 26/01/2024)

³ <https://nuevaepoca.revistalatinacs.org/index.php/revista/article/view/375/830> (προσπελάστηκε στις 26/01/2024)

⁴ <https://acl-bg.org/proceedings/2021/RANLP%202021/pdf/2021.ranlp-1.114.pdf> (προσπελάστηκε στις 26/01/2024)

⁵ <https://hcommons.org/deposits/objects/hc:39636/datastreams/CONTENT/content> (προσπελάστηκε στις 26/01/2024)

Mobius και Horvitz (2017)⁶ εξετάζουν τις ψευδείς ειδήσεις στις ΗΠΑ που δημοσιεύτηκαν κατά τη διάρκεια των προεδρικών εκλογών του 2016 προς τις πηγές διάδοσης και τις τάσεις στην παραπληροφόρηση. Ceron, de-Lima-Santos και Quiles (2021)⁷ αναλύουν τις τάσεις στις ψευδείς ειδήσεις κατά τη διάρκεια της πανδημίας του Covid-19 σε διαφορετικές χώρες.

Η δομή αυτής της εργασίας είναι η εξής: στο **Κεφάλαιο 2** εξηγείται τι είναι οι ψευδείς ειδήσεις και γιατί είναι η παγκόσμια απειλή και περιγράφονται οι πιο δημοφιλείς τρόποι ανίχνευσης των ψευδών ειδήσεων, στο **Κεφάλαιο 3** αναφέρονται οι πιο γνωστοί οργανισμοί ελέγχου γεγονότων, ο τρόπος λειτουργίας τους και οι ερευνητικές εργασίες που έχουν πραγματοποιηθεί για να βρεθούν νέοι τρόποι εντοπισμού ψευδών ειδήσεων, στο **Κεφάλαιο 4** παρουσιάζεται η πρώτη υπηρεσία ελέγχου γεγονότων στην Ελλάδα – EllhnikaHoaxes.gr, στο **Κεφάλαιο 5** περιγράφονται η διαδικασία της ανάκτησης και επεξεργασίας των άρθρων στον ιστότοπο της EllhnikaHoaxes.gr, η ανάλυση αυτών των άρθρων και τα αποτελέσματα της ανάλυσης και στο **Κεφάλαιο 6** εκφράζονται το συμπέρασμα και οι προτάσεις για μελλοντική εργασία.

⁶ <https://racz.statistics.northwestern.edu/FRRMH17CIKM.pdf> (προσπελάστηκε στις 26/01/2024)

⁷ <https://www.sciencedirect.com/sdfe/reader/pii/S2468696420300562/pdf> (προσπελάστηκε στις 26/01/2024)

ΚΕΦΑΛΑΙΟ 2

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗΣ

2.1 ΙΣΤΟΡΙΚΟ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

2.1.1 ΕΝΝΟΙΑ ΤΟΥ ΟΡΟΥ «ΨΕΥΔΕΙΣ ΕΙΔΗΣΕΙΣ»

Οι ψευδείς ειδήσεις (αγγλ. fake news) δεν είναι πρόσφατο φαινόμενο, ωστόσο οι προθέσεις και τα αποτελέσματά τους αλλάζουν με τον καιρό. Υπάρχουν πολλές γνώμες για τη χρονολογία της εμφάνισης πρώτων ψευδών ειδήσεων διότι η έννοια του όρου «ψευδείς ειδήσεις» παραμένει όχι απολύτως ορισμένη. Σύμφωνα με Golbeck, Mauriello, Auxier, Bhanushali, Bonk, Bouzaghane, Buntain, Chanduka, Cheakalos, Everett και Falak (2018)⁸, οι ψευδείς ειδήσεις είναι οι πληροφορίες που παρουσιάζονται ως είδηση παρότι είναι ανακριβείς και έχουν σχεδιαστεί για να εξαπατήσουν τον παραλήπτη να πιστέψει ότι είναι αληθινή η είδηση. Ο Higdon (2020)⁹ ορίζει τις ψευδείς ειδήσεις ως ψευδές ή παραπλανητικό περιεχόμενο που παρουσιάζεται ως είδηση και μεταδίδεται σε προφορική, γραπτή, έντυπη, ηλεκτρονική ή ψηφιακή μορφή. Dey, Rafi, Hasan και Kundu (2018)¹⁰ προτείνουν παρόμοιο ορισμό αλλά επίσης διευκρινίζουν ότι οι ψευδείς ειδήσεις δεν προέρχονται από καθιερωμένες πηγές ειδήσεων. Bounegru, Gray, Venturini και Mauri (2018)¹¹ θεωρούν ότι ένα ψέμα γίνεται ψευδής είδηση όταν «συλλαμβάνεται από δεκάδες άλλα ιστολόγια, αναμεταδίδεται από εκατοντάδες ιστοχώροι, δημοσιεύεται ταυτόχρονα από χιλιάδες λογαριασμούς στα μέσα κοινωνικής δικτύωσης και διαβάζεται από εκατοντάδες χιλιάδες χρήστες».

Αν και οι ερμηνείες του όρου «ψευδείς ειδήσεις» διαφέρουν από ερευνητή σε ερευνητή, υπάρχουν κοινά χαρακτηριστικά μεταξύ τους. Συνοψίζοντας αυτούς τους ορισμούς, οι ψευδείς ειδήσεις ή fake news είναι οι πληροφορίες που παρότι είναι ανακριβείς ή ψευδείς μεταδίδονται ως αληθινή είδηση.

2.1.2 ΕΜΦΑΝΙΣΗ ΦΑΙΝΟΜΕΝΟΥ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

Υπάρχουν τουλάχιστον τρεις εκδοχές πρώτης διάδοσης ψευδών ειδήσεων.

⁸ <https://dl.acm.org/doi/pdf/10.1145/3201064.3201100> (προσπελάστηκε στις 26/01/2024)

⁹ https://en.wikipedia.org/wiki/Fake_news#cite_ref-Higdon_15-0 (προσπελάστηκε στις 26/01/2024)

¹⁰

https://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/10139/13201055%2c13201070%2c13201030%2c14101233_CSE.pdf?sequence=1&isAllowed=y (προσπελάστηκε στις 26/01/2024)

¹¹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3097666 (προσπελάστηκε στις 26/01/2024)

Σύμφωνα με την πρώτη θεωρία, οι ψευδείς ειδήσεις εμφανίστηκαν στην αρχαιότητα, τον 13^ο αιώνα π.Χ., όταν ο Ραμσής ο Μέγας, ο Φαραώ της Αρχαίας Αιγύπτου, διέδωσε την ψευδή είδηση για τη νίκη του στη Μάχη του Καντές (Weir, 2009)¹².

Σύμφωνα με τη δεύτερη θεωρία, που είναι αποδεκτή από πιο πολλούς ερευνητές, οι ψευδείς ειδήσεις εμφανίστηκαν το 1453, μόλις οι γραπτές ειδήσεις ξεκίνησαν να κυκλοφορούν, όταν εφευρέθηκε του επίπεδου πιεστηρίου από τον Γουτεμβέργιο (Collins, Hoang, Nguyen και Hwang, 2021)¹³ και (Soll, 2016)¹⁴.

Σύμφωνα με την τρίτη θεωρία, οι ψευδείς ειδήσεις εμφανίστηκαν στο τέλος του 19^{ου} αιώνα. Οι υποστηρικτές αυτής της εκδοχής τονίζουν ότι, αφού ο όρος «fake news» χρησιμοποιήθηκε επισήμως για πρώτη φορά το 1890, εκείνη την εποχή αυτό το φαινόμενο και άρχισε να υπάρχει. Παρ' όλες διαφωνίες για την εμφάνιση των ψευδών ειδήσεων, αναμφισβήτητα το φαινόμενο αυτό υπήρχε πολύ καιρό πριν μπουν το Internet και τα μέσα κοινωνικής δικτύωσης στην καθημερινότητά μας.

2.1.3 ΣΗΜΕΡΙΝΗ ΕΠΙΡΡΟΗ ΤΟΥ ΔΙΑΔΙΚΤΥΟΥ ΣΤΗ ΔΙΑΔΟΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

Οι εμφάνιση του Internet και των μέσων κοινωνικής δικτύωσης, όπως το Facebook, το Twitter ή το YouTube, διευκόλυνε τη ζωή μας, προσφέροντας πρόσβαση στην πληθώρα πληροφοριών και σβήνοντας σύνορα μεταξύ χωρών. Ωστόσο, αυτές οι αλλαγές διευκόλυναν και τη διάδοση ψευδών ειδήσεων παγκοσμίως με ελάχιστο κόστος.

Ακόμα και 50 χρόνια πριν υπήρχαν μόνο τρία επίσημα μέσα διάδοσης ειδήσεων: η ραδιοφωνική εκπομπή, η τηλεοπτική εκπομπή ή η γραπτή δημοσίευση, πχ μια εφημερίδα. Για να γίνει κάποιος ανταποκριτής έπρεπε να έχει σπουδάσει τη δημοσιογραφία. Για να μεταδίδει ειδήσεις στο κοινό, έπρεπε είτε να έχει την επαγγελματική άδεια ως τηλεοπτικός ή ραδιοφωνικός σταθμός ή εκδοτικός οίκος. Υπήρχαν αρκετά εμπόδια εισόδου σε αυτό το κλάδο.

Παλιά όλες οι ειδήσεις, τόσο αληθείς, όσο και αναληθείς είχαν τοπικό χαρακτήρα και συνήθως διαδιδόταν ανάμεσα στους κατοίκους μιας κοινωνίας όπως μια πόλη ή μια επαρχία, τώρα, όμως, δεν υπάρχουν σύνορα, θέματα μετάφρασης ή υψηλά κόστη μετάδοσης.

¹² https://en.wikipedia.org/wiki/Fake_news#cite_ref-148 (προσπελάστηκε στις 26/01/2024)

¹³ <https://www.tandfonline.com/doi/epdf/10.1080/24751839.2020.1847379?needAccess=true> (προσπελάστηκε στις 26/01/2024)

¹⁴ <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/> (προσπελάστηκε στις 26/01/2024)

Με την εμφάνιση του Διαδικτύου ο καθένας απέκτησε την ευκαιρία να γίνει ακουστός από εκατοντάδες χιλιάδες ανθρώπους παγκοσμίως σε μερικά λεπτά. Ένας χρήστης μπορεί να γίνει «πηγή ψευδούς ειδήσεως» ακόμα και ακούσια, αν δημοσιεύσει κάτι που θα παρεξηγηθεί από άλλους χρήστες του Διαδικτύου.

Ωστόσο, οι περισσότερες ειδήσεις σήμερα δημιουργούνται σκοπίμως και οι προθέσεις της δημιουργίας τους διαφέρουν πολύ από το πρόσωπο που τους μεταδίδει. Σύμφωνα με Berduygina, Vladimirova και Chernyaeva (2019)¹⁵, οι προθέσεις είναι οι εξής:

- Τα ΜΜΕ:
 - Αύξηση δημοτικότητας και φήμης
 - Αύξηση κίνησης χρηστών (αγγλ. user traffic)
 - Πιο πολλή ζήτηση για τοποθέτηση διαφημίσεων
 - Αύξηση τιμής για τοποθέτηση διαφημίσεων
- Οι πολιτικοί:
 - Δημιουργία της θετικής εικόνας του υποψήφιου ή του κόμματος
 - Πλήγμα στην εικόνα του ανταγωνιστή
 - Διαμόρφωση της κοινής γνώμης
- Οι εταιρείες:
 - Χειρισμός της αγοράς
 - Επιρροή στους καταναλωτές
 - Δημιουργία δημοτικότητας και φήμης
 - Επιρροή στη ζήτηση
- Τα φυσικά πρόσωπα:
 - Κλοπή προσοχής
 - Απόκτηση δημοτικότητας και φήμης
 - Εκδίκηση
 - Απάτη

Το σημείο εκκίνησης των ψευδών ειδήσεων, με τη μορφή με την οποία είμαστε συνηθισμένοι, είναι η εκλογική εκστρατεία στις ΗΠΑ το 2016, κατά την οποία έγινε διάδοση ενός μεγάλου αριθμού ψευδών ειδήσεων και ισχυρισμών από πολιτικά πρόσωπα, με σκοπό την επικράτηση στις επικείμενες εκλογές. Από τότε ό,τι σχετίζεται με παραπληροφόρηση θεωρείται «fake

¹⁵ <https://hcommons.org/deposits/objects/hc:39636/datastreams/CONTENT/content> (προσπελάστηκε στις 26/01/2024)

news» ή ψευδής είδηση, αν και υπάρχει η σημασιολογική διαφορά μεταξύ των όρων «παραπληροφόρηση», «ψευδής είδηση» και «ψευδής ισχυρισμός», αυτοί χρησιμοποιούνται σχεδόν ως ισοδύναμοι.

Πριν το 2016 ο όρος «fake news» δε χρησιμοποιούταν από μελετητές σχεδόν καθόλου (με 3-10 μελέτες τη χρονιά). Ο αριθμός των επιστημονικών δημοσιευμάτων που αφορούν ψευδές ειδήσεις ξεκίνησε να μεγαλώνει το 2016, αυξάνοντας κατά 200 δημοσιεύσεις κάθε χρόνο (Righetti, 2021)¹⁶.

Η επόμενη περίοδος στη σημερινή μας ιστορία, κατά την οποία ο αριθμός των ψευδών ειδήσεων αυξανόταν με ραγδαίους ρυθμούς ήταν η πανδημία του Covid-19. Σε σύγκριση με τις ψευδείς ειδήσεις του 2016 που επηρέασαν κυρίως τους πολίτες των ΗΠΑ, οι παραπλανητικές ειδήσεις για το Covid-19 έγιναν παγκόσμιο πρόβλημα, βάζοντας σε ρίσκο τις ζωές εκατοντάδων εκατομμυρίων ανθρώπων σε όλο τον κόσμο. Κατά την περίοδο αυτή, οι ψευδές ειδήσεις έγιναν απειλή για όλα τα κράτη, συμπεριλαμβανομένης της Ελλάδας.

2.2 ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΤΥΠΟΙ

Μιας και δεν υπάρχει ένας διαδεδομένος ορισμός του τι είναι οι ψευδείς ειδήσεις, ο κάθε ερευνητής τις κατηγοριοποιεί διαφορετικά. Rubin, Chen και Conroy (2015)¹⁷ γενικεύουν το πλήθος κατηγοριών έως τις τρεις βάση τη σοβαρότητα της είδησης: σοβαρές σκευωρίες, απάτες μεγάλης κλίμακας και χιουμοριστικά κίβδηλα. Golbeck, Mauriello, Auxier, Bhanushali, Bonk, Bouzaghrane, Buntain, Chanduka, Cheakalos, Everett και Falak (2018)¹⁸, αντιθέτως, προσπάθησαν να είναι πιο σχολαστικοί και κατέταξαν τα fake news σε επτά κατηγορίες, δίνοντας έμφαση όχι στον τρόπο παραπληροφόρησης και προσέλκυσης προσοχής αλλά πιο πολύ στον σκοπό της ψευδούς είδησης. Οι μελετητές δίνουν αρκετά παραδείγματα αλλά είναι πιο κατανοητά για Αμερικάνους και όχι για τους Έλληνες. Για καλύτερη κατανόηση, ακολουθούν οι κατηγορίες που προτάθηκαν στην εργασία τους συμπεριλαμβανομένων παραδειγμάτων από το ελληνικό Διαδίκτυο (η ορθογραφία και το στυλ των τίτλων διατηρούνται):

- Υπερβολική θέση έναντι ενός ατόμου ή μιας ομάδας:

¹⁶ <https://firstmonday.org/ojs/index.php/fm/article/view/11645/11217> (προσπελάστηκε στις 26/01/2024)

¹⁷ <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/pra2.2015.145052010083> (προσπελάστηκε στις 26/01/2024)

¹⁸ <https://dl.acm.org/doi/pdf/10.1145/3201064.3201100> (προσπελάστηκε στις 26/01/2024)

- «ΣΕΙΣΜΟΣ ΣΤΑ SOCIAL MEDIA ΜΕ ΤΟΝ ΤΣΙΠΡΑ!!! Πανικός με την φωτογραφία του Τσίπρα που κρατά ανάποδα και κοιτά με τα κιάλια!!!»¹⁹
- «Πάνω από 2 εκατομμύρια άρπαξε το βρωμοτόμαρο ο Τζανάκης από την αρχή της πανδημίας για να τρομοκρατήσει τους Έλληνες»²⁰
- Υπερβολική θέση υπέρ ενός ατόμου ή μιας ομάδας:
 - «Δεν είναι τρολιά! Υποψήφιος για Όσκαρ ο Σάκης Ρουβάς!»²¹
 - «Οι Παροχές Τσίπρα Στη ΔΕΘ : Μειώσεις Φόρων, Παροχές Σε Ευπαθείς Ομάδες, Αύξηση Μισθών, Προσλήψεις»²²
- Απαξίωση μιας συνήθως αξιόπιστης πηγής
 - «ΟΤΑΝ Ο ΕΟΔΥ ΟΝΟΜΑΖΕ «ΓΡΙΠΗ» ΤΟΝ ΚΟΡΩΝΟΪΟ ΤΟΥ 2014! Ντοκουμέντα και ερωτήματα»²³
- Τραβηγμένα εγκλήματα και βία
 - «ΦΡΙΚΗ – ΑΓΡΙΑ ΔΟΛΟΦΟΝΙΑ 17ΧΡΟΝΗΣ ΣΤΗΝ ΘΕΣΣΑΛΟΝΙΚΗ!!!!ΒΡΕΘΗΚΕ ΣΤΟ ΔΑΣΟΣ ΤΟΥ ΣΕΙΧ ΣΟΥ ΣΦΑΓΜΕΝΗ ΚΥΡΙΟΛΕΚΤΙΚΑ!!!ΕΛΕΙΠΑΝ ΟΛΑ ΤΑ ΟΡΓΑΝΑ ΤΗΣ!!!»²⁴
 - ΕΚΤΑΚΤΟ: Ελεύθερος ο δολοφόνος της φοιτήτριας στην Ρόδο!²⁵
- Ρατσιστικά μηνύματα:
 - «Ο Παπαδόπουλος απέλασε τους Πακιστανούς επειδή ο πρόεδρος του Πακιστάν είπε ότι σε πόλεμο Ελλάδας Τουρκίας θα πάνε με τους Τούρκους»²⁶
 - «ΒΙΝΤΕΟ ΣΟΚ! ΣΤΟ ΣΥΝΤΑΓΜΑ ΠΑΚΙΣΤΑΝΟΙ ΜΕ ΣΤΡΑΤΙΩΤΙΚΑ ΔΕΙΧΝΟΥΝ ΠΩΣ ΕΚΠΑΙΔΕΥΟΝΤΑΙ...»²⁷
- Παραφυσικές ιστορίες

¹⁹ <https://archive.ph/20160531174909/www.iokh.gr/2016/05/seismos-sta-social-media.html#selection-549.0-549.122> (προσπελάστηκε στις 26/01/2024)

²⁰ <https://archive.ph/20221012043453/https://karditsastakra.com/2022/10/02πάνω-από-2-εκατομμύρια-άρπαξε-το-βρωμοτ/#selection-429.0-429.116> (προσπελάστηκε στις 26/01/2024)

²¹ <https://archive.is/RbV7T#selection-1315.0-1315.53> (προσπελάστηκε στις 26/01/2024)

²² <https://archive.fo/K1XzJ#selection-369.0-369.97> (προσπελάστηκε στις 26/01/2024)

²³ <https://archive.fo/IjzgO#selection-59689.1-59693.70> (προσπελάστηκε στις 26/01/2024)

²⁴ <https://www.ellinikahoaxes.gr/2013/08/12/φρικη-αγρια-δολοφονια-17χρονησ-στην-θε/> (προσπελάστηκε στις 26/01/2024)

²⁵ <https://archive.is/2bb5u#selection-8487.0-8487.56> (προσπελάστηκε στις 26/01/2024)

²⁶ <https://archive.ph/5pL0Q> (προσπελάστηκε στις 26/01/2024)

²⁷ <https://archive.ph/2022.01.23-220737/https://ilovepatrida.com/βιντεο-σοκ-στο-συνταγμα-πακιστανοι-με/#selection-4211.0-4211.78> (προσπελάστηκε στις 26/01/2024)

- «Αυτή είναι η μυστική καταγωγή των Ελλήνων - Τι κρύβει η μυθολογία μας...»²⁸
- Θεωρίες συνωμοσίας
 - «Λιβύη> Τα σημάδια από το πολεμικό χτύπημα στο λεωφορείο της ελληνικής στρατιωτικής αποστολής [VIDEO]»²⁹

Οι κατηγορίες που ορίζουν Tandoc, Lim και Ling (2018)³⁰ είναι

- Σάτιρα ειδήσεων είναι μια από πιο αφελείς κατηγορίες των ψευδών ειδήσεων, γιατί συνήθως έχει να κάνει με κωμικές εκπομπές που μεταδίδουν ειδήσεις με χιουμοριστικό τρόπο και υπερβολές. Παρότι τέτοιου είδους εκπομπές επικεντρώνονται στην επικαιρότητα, τα σόου αυτά προβάλλονται ως διασκεδαστικές εκπομπές όπου οι παρουσιαστές είναι κωμικοί και όχι ως ενημερωτική εκπομπή ή δελτίο ειδήσεων όπου οι παρουσιαστές είναι δημοσιογράφοι.
- Παρωδία ειδήσεων είναι οι ψευδείς ειδήσεις που ναι μεν έχει πολλές ομοιότητες με τη σάτιρα ειδήσεων, έχει τα δικά τις χαρακτηριστικά. Η παρουσίαση των εκπομπών παρωδίας ειδήσεων μιμείται δελτία ειδήσεων αλλά σε υπερβολικό βαθμό και το περιεχόμενο που μεταδίδεται είναι συνήθως μην πραγματικό, γελοίο και σχεδόν αδύνατο.
- Σκευωρία ειδήσεων είναι οι ψευδείς ειδήσεις που σε σύγκριση με την παρωδία ειδήσεων και τη σάτιρα ειδήσεων δε μεταδίδονται ως κάτι διασκεδαστικό και ο παραλήπτης του ψευδούς περιεχομένου δε γνωρίζει εξαρχής ότι πρόκειται για κάτι αναληθές. Ο στόχος του παραγωγού αυτού του είδους ειδήσεων είναι κατά κανόνα να παραπλανήσει το κοινό.
- Παραποίηση εικόνων είναι μια από τις πιο συνηθισμένες μορφές των ψευδών ειδήσεων και περιλαμβάνει διάφορες τεχνικές επεξεργασίας εικόνων που χρησιμοποιούνται για την παραγωγή παραποιημένου οπτικού υλικού, ξεκινώντας από μια απλή αύξηση του κορεσμού χρωμάτων στη φωτογραφία μέχρι την αφαίρεση ή την επικόλληση ενός προσώπου στη φωτογραφία.

²⁸ https://web.archive.org/web/20190105191951/https://www.pronews.gr/istoria/740223_ayti-einai-i-mystiki-katagogi-ton-ellinon-ti-kryvei-i-mythologia-mas (προσπελάστηκε στις 26/01/2024)

²⁹ <https://archive.ph/20230920205356/https://kourdistoportocali.com/news-desk/livyi-gt-idoy-poy-chtypithike-to-leoforeio-tis-ellinikis-stratiotikis-apostolis/> (προσπελάστηκε στις 26/01/2024)

³⁰ <https://doi.org/10.1080/21670811.2017.1360143> (προσπελάστηκε στις 26/01/2024)

- Διαφήμιση και δημόσιες σχέσεις είναι οι ψευδείς ειδήσεις που προβάλλονται ως αληθείς ενώ στην πραγματικότητα ο στόχος του δημιουργού αυτών των ειδήσεων δεν είναι παραπλάνηση αλλά η διαφήμιση ενός προσώπου η ενός προϊόντος.
- Προπαγάνδα είναι οι ειδήσεις που δημιουργούνται από μια πολιτική οντότητα ώστε να επηρεάσει την αντίληψη των κοινού στο οποίο στοχεύει. Μέσω συνεχούς προπαγάνδας, όπου εμπλέκονται όλα τα ΜΜΕ, επιτυγχάνεται η πλήση εγκεφάλου (αγγλ. brainwashing) που αποτρέπει τους ανθρώπους από το σκέφτονται κριτικά και ανεξάρτητα.

Σύμφωνα με Sousa-Silva (2022)³¹, οι κατηγορίες είναι οι έξι:

- Παραπληροφόρηση είναι οι ειδήσεις των οποίων η πραγματικότητα είναι παραποιημένη με σκοπό να εξαπατήσει το κοινό ώστε να αποκτήσει οικονομικό ή/και πολιτικό κέρδος (Sousa-Silva, 2022)³²³¹.
- Προκατάληψη είναι η δημιουργία μιας κοινής άποψης για κάτι ή κάποιον μέσω γενίκευσης μιας ομάδας ανθρώπων ή αντικειμένων βασίζοντας σε παρατηρήσεις ή/και χαρακτηριστικά ενός ή μερικούς εκπροσώπους της ομάδας αυτής.
- Κλικοδόλωμα είναι σκευωρίες που μεταδίδονται στο Διαδίκτυο με πιασάρικους τίτλους με αποτέλεσμα ο ιστότοπος να αποκτάει αυξημένη επισκεψιμότητα και συνεπώς μπορεί να αυξήσει τις τιμές τοποθέτησης διαφήμισης.
- Κιτρινισμός είναι η κατηγορία ψευδών ειδήσεων που ελάχιστα μοιάζει με το κλικοδόλωμα γιατί επίσης χρησιμοποιεί πιασάρικους τίτλους. Στις ιστορίες που ανήκουν σε αυτή την κατηγορία οι επιλεγμένες λέξεις είναι εξαιρετικά εντυπωσιακές και συχνά υπερβάλλουσες.
- Σάτιρα, όπως και προαναφερθείς σάτιρα ειδήσεων, δεν έχει σκοπό να εξαπατήσει και να παραπλανήσει το κοινό αλλά να το διασκεδάσει γελοιοποιώντας κάποιο γεγονός

Collins, Hoang, Nguyen και Hwang (2021)³³ μοιάζει αρκετά με τις κατηγοριοποιήσεις άλλων ερευνητών και αποτελείται από

- Κλικοδόλωμα
- Προπαγάνδα
- Σάτιρα και παρωδία

³¹ <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

³² <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

³³ <https://www.tandfonline.com/doi/epdf/10.1080/24751839.2020.1847379?needAccess=true> (προσπελάστηκε στις 26/01/2024)

- Κλοπή ονόματος είναι η κατηγορία που μοιάζει με την κατηγορία «Απαξίωση μιας συνήθως αξιόπιστης πηγής», και είναι συνήθως η δημιουργία ενός ιστότοπου που μοιάζει πολύ με τον ιστότοπο της πρωτότυπης πηγής για περαιτέρω εξαπάτηση υπό το όνομα της πρωτότυπης πηγής.
- Πλαισίωση είναι η απόκρυψη της αλήθειας τονίζοντας σε εξαιρετικό βαθμό κάποιες σχετικά ασήμαντες πληροφορίες
- Δημοσιογραφική εξαπάτηση είναι η ψευδής είδηση όπου χρησιμοποιείται η ίδια τεχνική όπως και στην πλαισίωση με μόνη διαφορά ότι η πλαισίωση μπορεί να δημιουργηθεί από οποιονδήποτε, αλλά η δημοσιογραφική εξαπάτηση μόνο από τον επαγγελματία δημοσιογράφο (Collins, Hoang, Nguyen και Hwang, 2021)³⁴.

H Wardle (2017)³⁵ χωρίζει τις ψευδείς ειδήσεις σε επτά κατηγορίες: σάτιρα ή παρωδία, λάθος σχέση, παραπλάνηση, λάθος συγκριμένο, απάτη, χειραγώγηση και σκευωρία.

Πίνακας 2.2.1 Κατηγορίες ψευδών ειδήσεων

ΕΡΕΥΝΗΤΗΣ	ΚΑΤΗΓΟΡΙΕΣ (ΜΕΤΑΦΡΑΣΜΕΝΕΣ)	ΚΑΤΗΓΟΡΙΕΣ (ΠΡΩΤΟΤΥΠΗ ΓΛΩΣΣΑ)
Rubin, Chen και Conroy (2015)	◦ Σοβαρές σκευωρίες	Serious fabrications
	◦ Απάτες μεγάλης κλίμακας	Large scale hoaxes
	◦ Χιουμοριστικά κίβδηλα	Humorous fakes
Golbeck, Mauriello, Auxier, Bhanushali, Bonk, Bouzaghrane, Buntain, Chanduka, Cheakalos, Everett και Falak (2018)	◦ Υπερβολική θέση έναντι ενός ατόμου ή μιας ομάδας	Hyperbolic position against one person or group
	◦ Υπερβολική θέση υπέρ ενός ατόμου ή μιας ομάδας	Hyperbolic position in favor of one person or group
	◦ Απαξίωση μιας συνήθως αξιόπιστης πηγής	Discredit a normally credible source
	◦ Τραβηγμένα εγκλήματα και βία	Sensationalist crimes and violence
	◦ Ρατσιστικά μηνύματα	Racist messaging
	◦ Παραφυσικές ιστορίες	Paranormal theories
	◦ Θεωρίες συνωμοσίας	Conspiracy theories

³⁴ <https://www.tandfonline.com/doi/epdf/10.1080/24751839.2020.1847379?needAccess=true> (προσπελάστηκε στις 26/01/2024)

³⁵ <https://firstdraftnews.org/fake-news-complicated/> (προσπελάστηκε στις 26/01/2024)

Sousa-Silva (2022)	◦ Παραπληροφόρηση	Disinformation
	◦ Προκατάληψη	Bias
	◦ Κλικοδόλωμα	Clickbait
	◦ Κιτρινισμός	Sensationalism
	◦ Σάτιρα (συμπλ. παρωδία)	Satire (incl. parody)
Tandoc, Lim και Ling (2018)	◦ Σάτιρα ειδήσεων	News satire
	◦ Παρωδία ειδήσεων	News parody
	◦ Σκευωρία ειδήσεων	News fabrication
	◦ Παραποίηση εικόνων	Photo manipulation
	◦ Διαφήμιση	Advertising and public relations
	◦ Προπαγάνδα	Propaganda
Collins, Hoang, Nguyen και Hwang (2021)	◦ Κλικοδόλωμα	Clickbait
	◦ Προπαγάνδα	Propaganda
	◦ Σάτιρα και παρωδία	Satire and parody
	◦ Απάτες	Hoaxes
	◦ Κλοπή ονόματος	Name theft
	◦ Πλαισίωση	Framing
	◦ Δημοσιογραφική εξαπάτηση	Journalistic deception
Wardle (2017)	◦ Σάτιρα και παρωδία	Satire or parody
	◦ Λάθος σχέση	False connection
	◦ Παραπλάνηση	Misleading content
	◦ Λάθος συγκείμενο	False context
	◦ Απάτη	Impostor content
	◦ Χειραγώγηση	Manipulated content
	◦ Σκευωρία	Fabricated content

Όμως, η προσέγγιση των Berduygina, Vladimirova και Chernyaeva (2019)³⁶ διαφέρει πολύ από τις υπόλοιπες. Οι μελετητές κατηγοριοποιούν ψευδείς ειδήσεις βάσει τα έξι κριτήρια:

- Αναλογία αληθών– αναληθών πληροφοριών
 - Απόλυτο ψέμα
 - Κυρίως ψέμα με στοιχεία αλήθειας

³⁶ <https://hcommons.org/deposits/objects/hc:39636/datastreams/CONTENT/content> (προσπελάστηκε στις 26/01/2024)

- Ένα πραγματικό γεγονός με μεμονωμένα παραμορφωμένα αποσπάσματα
- Αυθεντικότητα των περιστάσεων, χαρακτηριστικών χρόνου και τόπου του γεγονότος
 - Παλιές πληροφορίες μεταμφιεσμένες ως επίκαιρες ειδήσεις
 - Το σκηνικό του γεγονότος άλλαξε
- Εμπλεκόμενα στην είδηση άτομα
 - Παράθεση ή συγγραφή για ένα δημόσιο πρόσωπο από το ψεύτικο λογαριασμό
 - Ένα δευτερεύον πρόσωπο που έρχεται στο προσκήνιο
- Οι ειδήσεις βασίζονται στην ανεπιβεβαίωτη κατάθεση
- Κίνητρα δημιουργίας και διάδοσης
 - Διασκέδαση
 - Πολιτικό πλεονέκτημα
 - Διάκριση
 - Αύξηση του Internet κίνησης (αγγλ. traffic)
 - Απάτη
 - Καταστροφή πληροφοριών αποθηκευμένων στον υπολογιστή
 - Σύλληψη της προσοχής σε ένα συγκεκριμένο πρόσωπο
 - Χειραγώγηση αγοράς
- Αντίληψη της αυθεντικότητας της είδησης
 - Προφανώς ψευδής
 - Ειδήσεις που προκαλούν αμφιβολίες
 - Πειστικά παραποιημένα

2.3 ΕΥΡΟΣ ΠΡΟΒΛΗΜΑΤΟΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

2.3.1 ΣΤΑΤΙΣΤΙΚΑ ΔΕΔΟΜΕΝΑ

Από το 2016, όταν το φαινόμενο των ψευδών ειδήσεων ξεκίνησε να γίνεται δημοφιλές, συντάχτηκαν πολλές μελέτες και διενεργήθηκε αριθμός δημοσκοπήσεων από πολλά ινστιτούτα παγκοσμίως για τον βαθμό επιρροής των μέσων κοινωνικής δικτύωσης και των ψευδών ειδήσεων στους ανθρώπους διάφορης ηλικίας και υπηκοότητας.

Σύμφωνα με την έρευνα που διενεργήθηκε από Go και Lee (2020)³⁷ και τη δημοσκόπηση υπ' ονομασία «The Korean Academic Multimode Open Survey.(KAMOS)»³⁸ από το Center for

³⁷ <https://koreascience.kr/article/JAKO202022449681023.pdf> (προσπελάστηκε στις 26/01/2024)

³⁸ http://cnukamos.com/eng/sub4/menu_1.php (προσπελάστηκε στις 26/01/2024)

Asian Public Opinion Research & Collaboration Initiative (CAPORCI)³⁹ που έγινε τον Απρίλιο – Ιούνιο 2019^{40,41}, στην Κορέα,

- Το 65.9% των ερωτηθέντων έχει έρθει σε επαφή με ψευδείς ειδήσεις τους τελευταίους τρεις μήνες.
- Το 78.1% των ερωτηθέντων ηλικίας 18-29 και το 41.2% των ερωτηθέντων ηλικίας από 60 και άνω έχει έρθει σε επαφή με ψευδείς ειδήσεις γενικώς.
- Το 63.9% των ερωτηθέντων μαθαίνει ψευδείς ειδήσεις από τους ανθρώπους γύρω τους.
- Το 57.1% των ερωτηθέντων μαθαίνει ψευδείς ειδήσεις από διάφορους ιστότοπους.
- Το 56.9% των ερωτηθέντων μαθαίνει ψευδείς ειδήσεις από τα μέσα κοινωνικής δικτύωσης.
- Ανάμεσα στους ερωτηθείς ηλικίας 60 και άνω οι ψευδείς ειδήσεις μεταδίδονται από στόμα σε στόμα, ενώ οι νεότεροι ερωτηθείς έρχονται συνήθως σε επαφή με ψευδές ειδήσεις στο Διαδίκτυο.

Σύμφωνα με Dey, Rafi, Hasan και Kundu (2018)⁴², το ποσοστό των ενηλίκων πολιτών των ΗΠΑ, που εμπιστεύονται τις ειδήσεις στα μέσα κοινωνικής δικτύωσης, αυξήθηκε από το 49% το 2012 έως σχεδόν το 62% το 2016.

Βασίζοντας στα αποτελέσματα της Gallup⁴³, της αξιόπιστης και πιστευτής αμερικάνικης αναλυτικής εταιρείας, η Riffkin (2015)⁴⁴ αναφέρει ότι το 40% τους Αμερικάνους εμπιστεύονται τα ΜΜΕ. Το ποσοστό εμπιστοσύνης μειώθηκε σε σύγκριση με τα προηγούμενα έτη αλλά ακόμα παραμένει σχετικά υψηλό, ιδιαιτέρως ανάμεσα στους ανθρώπους ηλικίας 50 και άνω (45%).

Στην έρευνα των Go και Lee (2020)⁴⁵ οι ερευνητές συμπέραναν ότι

³⁹ <http://cnukamos.com/eng/main/> (προσπελάστηκε στις 26/01/2024)

⁴⁰

http://cnukamos.com/eng/ezs_board/ezsboard_filedownload.php?nfile=1_1564979920.pdf&ofile=KAMOS+4-1+%28E%29.pdf&b_path2=../ezs_data/board/table_301&b_code=301&idx=37 (πρωτότυπη δημοσκόπηση, προσπελάστηκε στις 26/01/2024)

⁴¹

http://cnukamos.com/eng/ezs_board/ezsboard_filedownload.php?nfile=1_1571298519.sav&ofile=KAMOS+4-1+%28E%29.sav&b_path2=../ezs_data/board/table_301&b_code=301&idx=38 (SPSS αποτελέσματα, προσπελάστηκε στις 26/01/2024)

⁴²

https://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/10139/13201055%2c13201070%2c13201030%2c14101233_CSE.pdf?sequence=1&isAllowed=y (προσπελάστηκε στις 26/01/2024)

⁴³ <https://news.gallup.com/> (προσπελάστηκε στις 26/01/2024)

⁴⁴ <https://news.gallup.com/poll/185927/americans-trust-media-remains-historical-low.aspx> (προσπελάστηκε στις 27/01/2024)

⁴⁵ <https://koreascience.kr/article/JAKO202022449681023.pdf> (προσπελάστηκε στις 26/01/2024)

- Το 38.2% των ερωτηθέντων, που έχουν εκτεθεί σε ψευδείς ειδήσεις, ελέγχει άλλες πηγές για να επιβεβαιώσει την εγκυρότητα αυτών των πληροφοριών.
- Μόνο το 6.7% των ερωτηθέντων έχει έρθει σε επαφή με τις ειδήσεις ελέγχου γεγονότων (αγγλ. fact-checking news). Επιπλέον, το 57.4% από αυτούς έχει έρθει σε επαφή με τις ειδήσεις ελέγχου γεγονότων μόνο μια φορά.

Σύμφωνα με τη δημοσκόπηση που διενεργήθηκε το 2018 από την Ευρωπαϊκή Ένωση⁴⁶,

- Το 10% των πολιτών της ΕΕ εμπιστεύεται και το 53% συνήθως εμπιστεύεται τις ειδήσεις στις έντυπες εφημερίδες και τα ενημερωτικά περιοδικά, με το 6% και το 43%, αντίστοιχα στην Ελλάδα, και το 12% και το 47%, αντίστοιχα στην Κύπρο. Οι κάτοικοι της Φινλανδίας εμπιστεύονται έντυπα μέσα ενημέρωσης πιο πολύ από άλλες χώρες της ΕΕ με το 16% και το 74%, αντίστοιχα.
- Το 5% των πολιτών της ΕΕ εμπιστεύεται και το 42% συνήθως εμπιστεύεται τις ειδήσεις στις διαδικτυακές εφημερίδες και τα ενημερωτικά περιοδικά, με το 4% και το 40%, αντίστοιχα στην Ελλάδα, και το 9% και το 46% αντίστοιχα στην Κύπρο. Οι κάτοικοι της Δανίας εμπιστεύονται διαδικτυακά μέσα ενημέρωσης πιο πολύ από άλλες χώρες της ΕΕ με το 11% και το 60%, αντίστοιχα.
- Το 2% των πολιτών της ΕΕ εμπιστεύεται και το 24% συνήθως εμπιστεύεται τις ειδήσεις στα μέσα κοινωνικής δικτύωσης και τις εφαρμογές ανταλλαγής άμεσων μηνυμάτων, με το 3% και το 29%, αντίστοιχα στην Ελλάδα, και το 5% και το 33%, αντίστοιχα στην Κύπρο. Οι κάτοικοι της Κύπρου και της Ρουμανίας εμπιστεύονται τις ειδήσεις στα μέσα κοινωνικής δικτύωσης πιο πολύ από άλλες χώρες της ΕΕ (5%).
- Το 12% των πολιτών της ΕΕ εμπιστεύεται και το 54% συνήθως εμπιστεύεται τις ειδήσεις στην τηλεόραση, με το 6% και το 34%, αντίστοιχα στην Ελλάδα, και το 18% και το 55%, αντίστοιχα στην Κύπρο. Οι κάτοικοι της Σουηδίας και του Βελγίου εμπιστεύονται τις ειδήσεις στην τηλεόραση πιο πολύ από άλλες χώρες της ΕΕ (19%).
- Το 15% των πολιτών της ΕΕ εμπιστεύεται και το 55% συνήθως εμπιστεύεται τις ειδήσεις στο ραδιόφωνο, με το 10% και το 47%, αντίστοιχα στην Ελλάδα, και το 17% και το 52%, αντίστοιχα στην Κύπρο. Οι κάτοικοι της Δανίας εμπιστεύονται τις ειδήσεις στο ραδιόφωνο πιο πολύ από άλλες χώρες της ΕΕ (29%).
- Το 2% των πολιτών της ΕΕ εμπιστεύεται και το 25% συνήθως εμπιστεύεται τις ειδήσεις στους ιστότοπους παροχής πρόσβασης σε βίντεο και μουσική, με το 2% και το 27%,

⁴⁶ <https://europa.eu/eurobarometer/surveys/detail/2183> (προσπελάστηκε στις 26/01/2024)

αντίστοιχα στην Ελλάδα, και το 3% και το 31% αντίστοιχα στην Κύπρο. Οι κάτοικοι της Δανίας, της Μάλτας και της Ρουμανίας εμπιστεύονται τις ειδήσεις στους ιστότοπους παροχής πρόσβασης σε βίντεο και μουσική πιο πολύ από άλλες χώρες της ΕΕ (4%).

- Το 37% των ερωτηθέντων έρχεται σε επαφή με ψευδές ειδήσεις καθημερινά ή σχεδόν καθημερινά, στην Ελλάδα είναι το 55% και στην Κύπρο είναι το 45%.
- Το 31% των ερωτηθέντων έρχεται σε επαφή με ψευδείς ειδήσεις τουλάχιστον μια φορά τη βδομάδα, στην Ελλάδα είναι το 19% και στην Κύπρο είναι το 16%.
- Το 15% των ερωτηθέντων θεωρεί πως μπορεί να καταλάβει αν μια είδηση είναι ψευδής ή όχι, στην Ελλάδα είναι το 12% και στην Κύπρο είναι το 30%.
- Το 56% των ερωτηθέντων είναι σχετικά σίγουρο πως μπορεί να καταλάβει αν μια είδηση είναι ψευδής ή όχι, στην Ελλάδα είναι το 52% και στην Κύπρο είναι το 48%.
- Το 44% των ερωτηθέντων θεωρεί ψευδείς ειδήσεις πραγματικό πρόβλημα και το 41% τείνει να συμφωνεί ότι είναι πραγματικό πρόβλημα, στην Ελλάδα είναι το 63% και το 27%, αντίστοιχα, και στην Κύπρο είναι το 69% και το 22%, αντίστοιχα.
- Το 45% θεωρεί ψευδείς ειδήσεις πραγματικό πρόβλημα και το 38% τείνει να συμφωνεί ότι είναι πραγματικό πρόβλημα για δημοκρατία γενικώς, στην Ελλάδα είναι το 60% και το 27%, αντίστοιχα, και στην Κύπρο είναι το 67% και το 22%, αντίστοιχα.
- Οι περισσότερες ερωτηθείς θεωρούν ότι οι ενημερωτικές πηγές που πρέπει να σταματήσουν να μεταδίδουν ψευδές ειδήσεις είναι οι δημοσιογράφοι (45%) και οι εθνικές αρχές (39%).

Συνολικά, το 68% των πολιτών της ΕΕ έρχεται σε επαφή με ψευδές ειδήσεις τέσσερις ή περισσότερες φορές το μήνα, ενώ μόνο το 15% θεωρεί πως μπορεί να καταλάβει αν μια είδηση είναι αληθής ή όχι. Επίσης, σχεδόν ένα μισό των πολιτών της ΕΕ πιστεύει πως οι ψευδές ειδήσεις είναι ένα πραγματικό πρόβλημα τόσο για τη χώρα τους, όσο και για δημοκρατία γενικώς.

Όταν οι άνθρωποι έρχονται σε επαφή με νέα και εξαιρετικά εκπληκτικά γεγονότα και πληροφορίες, προκαλείται η υψηλή διέγερση στον εγκέφαλο (Berger και Milkman, 2012)⁴⁷. Η διέγερση αυτή κάνει τους ανθρώπους να μοιράζονται τις ψευδείς ειδήσεις που μόλις είδαν με τους άλλους, χωρίς να ελέγχουν την εγκυρότητα της είδησης⁴⁸.

⁴⁷ <https://jonahberger.com/wp-content/uploads/2013/02/ViralityB.pdf> (προσπελάστηκε στις 26/01/2024)

⁴⁸ https://en.wikipedia.org/wiki/Fake_news (προσπελάστηκε στις 27/01/2024)

2.3.2 ΕΜΠΟΔΙΑ ΕΝΤΟΠΙΣΜΟΥ ΚΑΙ ΤΕΡΜΤΙΣΜΟΥ ΔΙΑΔΟΣΗΣ

Το πρόβλημα που αντιμετωπίζεται στην ανίχνευση ψευδών ειδήσεων είναι τα ερωτήματα το τι είναι μια ψευδής είδηση και εάν όλες οι κατηγορίες των αναληθών ειδήσεων είναι εξίσου βλαβερές.

Δευτέρων, κάθε χρόνο ο κλάδος της παραγωγής και διάδοσης ψευδών ειδήσεων εξελίσσεται όλο και πιο πολύ, αξιοποιώντας νέα εργαλεία και μέσα διάδοσης, πχ bots και deepfakes.

Τρίτων, η πρόσβαση στο Διαδίκτυο δεν είναι πια ένα προνόμιο αλλά κάτι υποχρεωτικό που αποκτάνε ολοένα και περισσότεροι άνθρωποι κάθε χρόνο. Η πλέον εύκολη προσβασιμότητα κάνει το κόστος διάδοσης των ψευδών ειδήσεων πολύ χαμηλό.

Ο όγκος των δεδομένων στο Internet αυξάνεται συνεχώς, με αποτέλεσμα ο εντοπισμός ψευδών ειδήσεων να γίνεται πιο χρονοβόρος (Sousa-Silva, 2022)⁴⁹.

Σύμφωνα με τα προαναφερθέντα αποτελέσματα της δημοσκόπησης που έλαβε χώρα στην ΕΕ το 45% των πολιτών πιστεύει ότι για να βελτιωθεί η κατάσταση με τις ψευδείς ειδήσεις οι ίδιοι οι δημοσιογράφοι πρέπει να σταματήσουν να μεταδίδουν παραπλανητικές ειδήσεις, συνεπώς η δραστηριότητα κάποιων δημοσιογράφων είναι επίσης ένα εμπόδιο εντοπισμού, διότι πλέον όχι όλοι οι δημοσιογράφοι ακολουθούν τις βασικές αρχές της δημοσιογραφίας: την αλήθεια, την αντικειμενικότητα και την ουδετερότητα.

2.4 ΣΥΝΕΠΕΙΕΣ

Αναμφισβήτητα, ο όρος «fake news» ήδη μπήκε στην καθημερινότητά μας. Επιπλέον, λόγω της ραγδαίας ανάπτυξης του διαδικτύου και της τεχνητής νοημοσύνης γίνεται πιο δύσκολο να ελέγχεται το φαινόμενο των ψευδών ειδήσεων.

Μα γιατί οι ψευδείς ειδήσεις είναι τόσο βλαβερές; Μήπως πρέπει να τις αντιμετωπίζουμε σαν ένα φυσικό γεγονός; Εν συντομία, όχι, διότι οι συνέπειες μαζικής διάδοσης ψευδών ειδήσεων μπορεί να είναι πολύ σοβαρές.

Σύμφωνα με Paul και Matthews (2016)⁵⁰, οι ψευδείς ειδήσεις δημιουργούν την προκατάληψη επιβεβαίωσης, με αποτέλεσμα οι άνθρωποι να αντιλαμβάνονται τις πληροφορίες που επιβεβαιώνουν τα πιστεύω τους ως πιο έγκυρες από αυτές, που παρότι είναι ποιοτικές και αξιόπιστες, αντικρούονται με τα πιστεύω τους. Shu, Bernard και Lui (2019)⁵¹ προσθέτουν ότι η επικράτηση των αναληθών ειδήσεων στα μέσα μαζικής ενημέρωσης και στο Διαδίκτυο έχει

⁴⁹ <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

⁵⁰ https://cdn.locals.com/documents/682652/682652_tp78i7ma7l3xmnp.pdf (προσπελάστηκε στις 26/01/2024)

⁵¹ <https://www.academia.edu/download/56015294/1.pdf> (προσπελάστηκε στις 26/01/2024)

ενδεχόμενο ρίσκο να καταστρέψει την αξιοπιστία ολοκλήρου του οικοσυστήματος των ειδήσεων. Berduygina, Vladimirova και Chernyaeva (2019)⁵² κάνουν έμφαση στις συνέπειες που μπορεί να αντιμετωπίσει ο διασπορέας των ειδήσεων και χωρίζουν τις συνέπειες σε δύο κατηγορίες: τις θετικές και τις αρνητικές. Οι θετικές είναι η διασκέδαση, η αύξηση του Internet κίνησης (αγγλ. traffic), η αύξηση της διασημότητας, η αύξηση του χρηματικού κέρδους και ο υγιής ανταγωνισμός, ενώ οι αρνητικές είναι οι παραπληροφόρηση χρηστών, ο κίνδυνος φήμης, η ηθική βλάβη, οι υλικές ζημιές, η χειραγώγηση αγοράς και ο κίνδυνος δικαστικής προσφυγής. Αν και η προσέγγιση αυτή είναι αρκετά αμφιλεγόμενη, είναι απαραίτητο να σημειωθεί ότι οι ψευδείς ειδήσεις πράγματι είναι ωφέλιμες αλλά για ένα μικρό κύκλο ατόμων – για αυτούς που παράγουν και μεταδίδουν αυτές τις ειδήσεις.

Ωστόσο, για τους καταναλωτές των ψευδών ειδήσεων δεν υπάρχουν θετικές συνέπειες. Επιπροσθέτως, οι επιπτώσεις κάποιων ψευδών πληροφοριών, που προέρχονται από υποτιθέμενες αξιόπιστες πηγές, μπορεί να είναι και μοιραίες.

Το 2019 ξέσπασε η πανδημία του Covid-19 που φόβιζε όλο τον κόσμο τουλάχιστον δυο χρόνια. Επειδή ήταν κάτι που ο σύγχρονος κόσμος δεν είχε ξαναζήσει, πχ η απαγόρευση κυκλοφορίας, η αδυναμία να ταξιδέψεις κ.α., ο αριθμούς των ψευδών ειδήσεων και συνωμοσιών για το Covid -19 έσπαγε όλα τα ρεκόρ. Επίσης, οι άνθρωποι ήταν κλεισμένοι στα σπίτια τους και το Διαδίκτυο έγινε ένα εργαλείο όχι μόνο για την εργασία τους και για τις σπουδές τους αλλά και για την τακτική τους ενημέρωση. Ένας αμέτρητος αριθμός συνωμοσιών και ψεύτικων θεραπειών για τον κορονοϊό δημοσιευόταν στο Internet κάθε μέρα κατά τη διάρκεια της πανδημίας, θέτοντας τη ζωή εκατομμυρίων ανθρώπων σε κίνδυνο. Για να προστατέψουν τους πολίτες τους, πολλές χώρες έχουν δημιουργήσει νόμους για να ελέγχουν παραπλανητικές πληροφορίες. Σε πολλές χώρες οι διασπορείς των ψευδών ειδήσεων συνελήφθησαν για τη διάδοση των ψευδών ειδήσεων για το Covid -19⁵³. Κατά τη διάρκεια της πανδημίας το πρόβλημα των αναληθών ειδήσεων έχει γίνει το Πρώτο Παγκόσμιο Infodemic (αγγλ. the First Global Infodemic)⁵⁴ σύμφωνα με το Παγκόσμιο Οργανισμό Υγείας (Nakov, Alam, Shaar, Da San Martino και Zhang, 2021)⁵⁵.

⁵² <https://hcommons.org/deposits/objects/hc:39636/datastreams/CONTENT/content> (προσπελάστηκε στις 26/01/2024)

⁵³ https://en.wikipedia.org/wiki/Fake_news#cite_note-bans-122 (προσπελάστηκε στις 28/01/2024)

⁵⁴ <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> (προσπελάστηκε στις 26/01/2024)

⁵⁵ <https://acl-bg.org/proceedings/2021/RANLP%202021/pdf/2021.ranlp-1.114.pdf> (προσπελάστηκε στις 26/01/2024)

2.5 ΤΡΟΠΟΙ ΑΝΙΧΝΕΥΣΗΣ

Για να βρεθεί ο πιο κατάλληλος και ο πιο ακριβής τρόπος ανίχνευσης ψευδών ειδήσεων, έχουν γίνει πολλές προσπάθειες. Πολλές προσεγγίσεις έχουν προταθεί, δοκιμαστεί και αναλυθεί.

Υπάρχουν πολλοί τρόποι ανίχνευσης αλλά ακόμα δεν έχει βρεθεί ο τρόπος που μπορεί να χρησιμοποιηθεί για να παράγει ποιοτικά αποτελέσματα ανεξαρτήτως από τα δεδομένα εισόδου. Ωστόσο, οι πιο συχνοί τρόποι ανίχνευσης είναι η χειρωνακτική προσέγγιση (αγγλ. manual approach) και η υπολογιστική προσέγγιση (αγγλ. computational approach). Sousa-Silva (2022)⁵⁶ πέρα από τη χειρωνακτική και την υπολογιστική προσεγγίσεις, αναφέρει τον τρόπο ανίχνευσης μέσω προσδιορισμού της προέλευσης της είδησης και την προσέγγιση εγκληματολογικής γλωσσολογίας. Iwendi, Mohan, Ibeke, Ahmadian και Ciano (2022)⁵⁷ αναφέρουν μόνο τη χρήση μηχανικής μάθησης ως τον τρόπο ανίχνευσης ψευδών ειδήσεων, παραπέμποντας τις ερευνητικές εργασίες των Bhoir (2020)⁵⁸, Das και Clark (2019)⁵⁹, Cui, Wang και Lee (2019)⁶⁰, Mayakul, Srisawat και Kiattisin (2019)⁶¹, Han και Mehta (2019)⁶² και Iwendi, Mohan, Ibeke, Ahmadian και Ciano (2022)⁶³ ως παραδείγματα των αλγορίθμων μηχανικής μάθησης για εντοπισμό ψευδών ειδήσεων (Πίνακα 3.2.1). Awan, Yasin, Nobanee, Ali, Shahzad, Nabeel, Zain και Shahzad (2021)⁶⁴ επίσης κάνουν έμφαση στη χρήση μηχανικής μάθησης για ανίχνευση αναληθών ειδήσεων. Ceron, de-Lima-Santos και Quiles (2021)⁶⁵ κάνουν λόγο για τη χρήση τεχνικών επεξεργασίας φυσικής γλώσσας (αγγλ. natural language processing) και ενσωμάτωσης (αγγλ. embedding) ως το καταλληλότερο τρόπο εντοπισμού αποκλειστικά ψευδών αξιολογήσεων και κουτσομπολιού στο Διαδίκτυο. Collins, Hoang, Nguyen και Hwang (2021)⁶⁶ διακρίνουν δέκα μεθόδους ανίχνευσης, που είναι οι εξής:

1. Έλεγχος από ειδικούς
2. Εφαρμογή τεχνικής επεξεργασίας φυσικού γλώσσας
3. Χρήση μηχανικής μάθησης
4. Χρήση συστήματος συστάσεων (αγγλ. recommendation system)

⁵⁶ <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

⁵⁷ <https://www.sciencedirect.com/science/article/pii/S0045790622002439> (προσπελάστηκε στις 26/01/2024)

⁵⁸ <https://ieeexplore.ieee.org/document/9104177/> (προσπελάστηκε στις 26/01/2024)

⁵⁹ <https://cs.pomona.edu/~ajc/pdf/Das.2019.TA.SatireVsFake.pdf> (προσπελάστηκε στις 26/01/2024)

⁶⁰ <https://dl.acm.org/doi/pdf/10.1145/3341161.3342894> (προσπελάστηκε στις 26/01/2024)

⁶¹ <https://ieeexplore.ieee.org/document/9024427/> (προσπελάστηκε στις 26/01/2024)

⁶² <https://ieeexplore.ieee.org/abstract/document/9065138/> (προσπελάστηκε στις 26/01/2024)

⁶³ <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

⁶⁴ <https://www.mdpi.com/2079-9292/10/19/2326> (προσπελάστηκε στις 26/01/2024)

⁶⁵ <https://www.sciencedirect.com/sdfe/reader/pii/S2468696420300562/pdf> (προσπελάστηκε στις 26/01/2024)

⁶⁶ <https://www.tandfonline.com/doi/epdf/10.1080/24751839.2020.1847379?needAccess=true> (προσπελάστηκε στις 26/01/2024)

5. Χρήση βαθιάς μάθησης (αγγλ. deep learning)
6. Μέθοδος ομαδοποίησης (αγγλ. group-based method)
7. Μέθοδος πληθοπορισμού (αγγλ. crowdsourcing)
8. Μέθοδος πληθοπορισμού και ελέγχου από ειδικούς (αγγλ. expert-crowdsourcing approach)
9. Μέθοδος αλληλεπίδρασης ανθρώπου και μηχανής (αγγλ. human-machine approach)
10. Υβριδική τεχνική

Σύμφωνα με το χωρισμό των Zhang και Ghorbani (2020)⁶⁷, οι τρόποι ανίχνευσης ψευδών ειδήσεων μπορεί να είναι είτε πρακτικές, είτε ερευνητικές. Στην πρώτη κατηγορία συμπεριλαμβάνονται οι ιστότοποι που ασχολούνται με τον έλεγχο γεγονότων επαγγελματικά. Ενώ η δεύτερη κατηγορία αποτελείται από ερευνητικές προσεγγίσεις που βασίζονται σε:

1. Στοιχεία (αγγλ. component-based category)
2. Εξόρυξη δεδομένων (αγγλ. data-mining category)
3. Υλοποιήσεις (αγγλ. implement-based category)

Τελικά, Zhou και Zafarani (2020)⁶⁸ χωρίζουν τους τρόπους ανίχνευσης ψευδών ειδήσεων στους πέντε:

1. Βάση τις γνώσεις με δυο κατηγορίες:
 - a. Χειρωνακτική προσέγγιση με δύο κατηγορίες
 - i. Βάση τους ειδικούς
 - ii. Βάση του μεγάλου κοινού που ελέγχει την εγκυρότητα των γεγονότων (collective intelligence)
 - b. Αυτοματοποιημένη προσέγγιση, πχ με χρήση μηχανικής μάθησης, επεξεργασίας φυσικής γλώσσας κ.α.
2. Βάση το στυλ
3. Βάση την παροχή
4. Βάση την αξιολόγηση της αξιοπιστίας της πηγής
 - a. Βάση τους συγγραφείς και τους εκδότες ειδήσεων
 - b. Βάση τους χρήστες των μέσων κοινωνικής δικτύωσης

⁶⁷ <https://www.sciencedirect.com/sdfe/reader/pii/S0306457318306794/pdf> (προσπελάστηκε στις 26/01/2024)

⁶⁸ <https://dl.acm.org/doi/fullHtml/10.1145/3395046> (προσπελάστηκε στις 26/01/2024)

Ο λόγος ύπαρξης τόσο πολλών μεθόδων και της συνεχούς προσπάθειας των ερευνητών να βρουν τον καλύτερο τρόπο ανίχνευσης ψευδών ειδήσεων είναι το γεγονός ότι δεν έχει βρεθεί ακόμα ένας τέλειος τρόπος ανίχνευσης για όλες τις κατηγορίες των ψευδών ειδήσεων. Ένας τρόπος μπορεί να εφαρμόζεται άψογα για τον εντοπισμό της σάτιρας στο Διαδίκτυο, αλλά να παράγει μην ικανοποιητικά αποτελέσματα όταν πρόκειται για συνωμοσίες.

Παρακάτω ακολουθεί η σύντομη περιγραφή των πιο δημοφιλών προσεγγίσεων.

2.5.1 ΧΕΙΡΩΝΑΚΤΙΚΗ Ή ΠΡΑΚΤΙΚΗ ΑΝΙΧΝΕΥΣΗ

Η μέθοδος αυτή βασίζεται στην κρίση των ειδικών που αναλύουν και ελέγχουν την εγκυρότητα κάθε είδησης. Αυτός ο τρόπος ανίχνευσης είναι ο πιο παλιός και ο πιο δημοφιλής. Τα πλεονεκτήματα της χειρωνακτικής ανίχνευσης είναι το ότι υπάρχουν ακόμα κάποια χαρακτηριστικά, τα οποία οι έξυπνες αλγόριθμοι δεν μπορούν ακόμα να αποκωδικοποιήσουν σωστά, πχ ειρωνεία, κάτι που μπορεί εύκολα να εντοπίσει ένας έμπειρος αναλυτής-δημοσιογράφος. Ωστόσο, υπάρχουν και πολλά μειονεκτήματα, τα μεγαλύτερα από τα οποία είναι τα ανθρώπινα λάθη και ο χρόνος, ο οποίος απαιτείται για τον έλεγχο ενός άρθρου από ένα ειδικό. Δυστυχώς, όταν αυξάνεται ο αριθμός των ειδήσεων που πρέπει να ελεγχθούν (είναι κάτι που συμβαίνει συνεχώς), αυξάνεται και ο φόρτος εργασίας των ειδικών, που σημαίνει μεγαλύτερη πιθανότητα λαθών.

2.5.2 ΑΝΙΧΝΕΥΣΗ ΜΕ ΤΗ ΧΡΗΣΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ Ή ΥΠΟΛΟΓΙΣΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

Αυτή η μέθοδος βασίζεται σε αλγορίθμους μηχανικής μάθησης που εκπαιδεύουν ένα σύστημα να ανιχνεύει ψευδείς ειδήσεις στο Διαδίκτυο. Οι περισσότεροι υπολογιστικές προσεγγίσεις χρησιμοποιούν νευρωνικά δίκτυα και την τεχνική επεξεργασίας φυσικής γλώσσας. Μερικά από τα πλεονεκτήματα της ανίχνευσης ψευδών ειδήσεων με τη χρήση μηχανικής μάθησης είναι η δυνατότητα γρήγορης επεξεργασίας μεγάλου όγκου δεδομένων και μηδενική πιθανότητα ανθρωπίνων λαθών. Το μεγαλύτερο μειονέκτημα είναι το ότι η φύση και η μορφή των ψευδών ειδήσεων αλλάζει συνεχώς άρα κάθε εκπαίδευσης του συστήματος πρέπει να προηγείται ο ανθρώπινος έλεγχος για να «βοηθάει» το σύστημα να μαθαίνει να διαχωρίζει τα νέα είδη ψευδών ειδήσεων στο Διαδίκτυο.

ΚΕΦΑΛΑΙΟ 3

ΕΛΕΓΧΟΣ ΕΓΚΥΡΟΤΗΤΑΣ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

3.1 ΥΠΗΡΕΣΙΕΣ ΚΑΙ ΟΡΓΑΝΙΣΜΟΙ

Τα τελευταία χρόνια έχουν εμφανιστεί πολλές υπηρεσίες και οργανισμοί που ασχολούνται με τον έλεγχο και τον εντοπισμό αναληθών ειδήσεων στο Internet. Σύμφωνα με Nakon, Alam, Shaar, Da San Martino και Zhang (2021)⁶⁹, μόνο το 2021 στον κόσμο υπήρχαν ήδη 200 ενεργοί οργανισμοί ελέγχου γεγονότων. Οι πιο δημοφιλείς από αυτούς είναι

- FactCheck.org⁷⁰ είναι ο αμερικάνικος μην κερδοσκοπικός οργανισμός, το έργο του Πανεπιστημίου της Πενσυλβάνια (ΗΠΑ), που ασχολείται κυρίως με τις ψευδείς ειδήσεις για την πολιτική στην Αμερική. Επίσης, ελέγχει πληροφορίες για τον κορονοϊό (SciCheck's COVID-19/Vaccination Project)⁷¹ και δημοσιεύει τα αποτελέσματα στα αγγλικά και τα ισπανικά, δηλώσεις που αφορούν επιστήμη (SciCheck)⁷² και viral-δηλώσεις στο Facebook (Facebook Initiative)⁷³. Είναι «ένας μην κομματικός, μην κερδοσκοπικός “συνήγορος των καταναλωτών” για τους ψηφοφόρους που στοχεύει στη μείωση του βαθμού εξαπάτησης και σύγχυσης στην πολιτική των ΗΠΑ» (FactCheck.org, 2024)⁷⁴.
- PolitiFact.com⁷⁵ είναι ο αμερικάνικος μην κερδοσκοπικός οργανισμός που ιδρύθηκε το 2007 στη Φλόριντα, ΗΠΑ. Επικεντρώνεται σε ισχυρισμούς πολιτικών και τους αξιολογεί προς την ακρίβεια τους. Ανήκει στο Poynter Institute for Media Studies.
- Snopes.com⁷⁶ ιδρύθηκε το 1995 στις ΗΠΑ και είναι ο παλιότερος ιστότοπος ελέγχου γεγονότων. Είναι μια ανεξάρτητη έκδοση υπό την εταιρεία Snopes Media Group Inc. Ασχολείται με τον έλεγχο εγκυρότητας ειδήσεων. Επίσης, είναι μέλος του IFCN (International Fact-Checking Network)⁷⁷.

⁶⁹ <https://acl-bg.org/proceedings/2021/RANLP%202021/pdf/2021.ranlp-1.114.pdf> (προσπελάστηκε στις 26/01/2024)

⁷⁰ <https://www.factcheck.org/> (προσπελάστηκε στις 26/01/2024)

⁷¹ <https://www.factcheck.org/covid-misconceptions/> (προσπελάστηκε στις 26/01/2024)

⁷² <https://www.factcheck.org/scicheck/> (προσπελάστηκε στις 26/01/2024)

⁷³ <https://www.factcheck.org/fake-news/> (προσπελάστηκε στις 26/01/2024)

⁷⁴ <https://www.factcheck.org/about/our-mission/> (προσπελάστηκε στις 26/01/2024)

⁷⁵ <https://www.politifact.com/> (προσπελάστηκε στις 26/01/2024)

⁷⁶ <https://www.snopes.com/> (προσπελάστηκε στις 26/01/2024)

⁷⁷ <https://ifcncodeofprinciples.poynter.org/> (προσπελάστηκε στις 26/01/2024)

- TruthOrFiction.com⁷⁸ είναι ο αμερικάνικος μην κομματικός ιστότοπος που ιδρύθηκε το 1999 και ελέγχει ψευδείς ειδήσεις διάφορων κατηγοριών.
- FullFact⁷⁹ είναι το βρετανικό επίσημα εγγεγραμμένο φιλανθρωπικό ίδρυμα που υποστηρίζεται οικονομικά από τις εταιρείες όπως η Google και η Facebook. Είναι μέλος του IFCN.
- The Washington Post Fact Checker⁸⁰ είναι η στήλη της Washington Post που πρωτοεμφανίστηκε το 2007 και βασίζεται όσο στην αμερικάνικη, τόσο και στην πολιτική στο εξωτερικό.

Οι οργανισμοί αυτοί, όμως, βασίζονται κυρίως στις αμερικάνικες ειδήσεις και οι ιστότοποί τους είναι στα αγγλικά, άρα οι πηγές αυτές δεν είναι πάντα εφαρμόσιμες στο ελληνικό Διαδίκτυο. Εντούτοις, στην Ελλάδα επίσης υπάρχουν οργανισμοί ελέγχου γεγονότων, από τους οποίους ο πιο γνωστός και αξιόπιστος είναι ο EllinikaHoaxes.gr⁸¹. Ο οργανισμός αυτός είναι πιστοποιημένος από τον IFCN και συνεργάζεται με την πλατφόρμα Facebook ως ελεγκτής γεγονότων. (Η περαιτέρω ανάλυση και περιγραφή αυτού του οργανισμού γίνεται στο Κεφάλαιο 4.) Οι περισσότεροι οργανισμοί είναι μην κερδοσκοπικοί και υπάρχουν χάρη σε δωρεές από άλλους οργανισμούς ή φυσικά πρόσωπα.

Συνήθως, ένας τέτοιος οργανισμός αποτελείται από σχετικά μικρή ομάδα ειδικών σε διάφορους κλάδους, πχ δημοσιογραφία, ανάλυση κ.α., που ελέγχει κάθε είδηση στο Διαδίκτυο χειρωνακτικά. Σε κάποιους οργανισμούς χρησιμοποιείται μερικώς υπολογιστική προσέγγιση αλλά δεν είναι πολύ δημοφιλής πρακτική. Αυτό το γεγονός δημιουργεί ένα μεγάλο θέμα καθυστέρησης ελέγχου γεγονότων. Σύμφωνα με Shao, Ciampaglia, Flammini και Menczer (2016)⁸² υπάρχει η 13ωρη καθυστέρηση μεταξύ της εμφάνισης της ψευδής είδησης και του ελέγχου της εγκυρότητάς της. Αυτό το χρονικό κενό γίνεται ακόμα πιο μεγάλο κατά τη διάρκεια των εκλογών όταν ο όγκος των παραγόμενων παραπλανητικών ειδήσεων φτάνει στην κορυφή. Σε μια ακόμα δυσκολότερη κατάσταση βρισκόταν οι οργανισμοί ελέγχου γεγονότων κατά τη διάρκεια της πανδημίας της Covid-19.

⁷⁸ <https://www.truthorfiction.com/> (προσπελάστηκε στις 26/01/2024)

⁷⁹ <https://fullfact.org/> (προσπελάστηκε στις 26/01/2024)

⁸⁰ <https://www.washingtonpost.com/politics/fact-checker/> (προσπελάστηκε στις 26/01/2024)

⁸¹ <https://www.ellinikahoaxes.gr/> (προσπελάστηκε στις 26/01/2024)

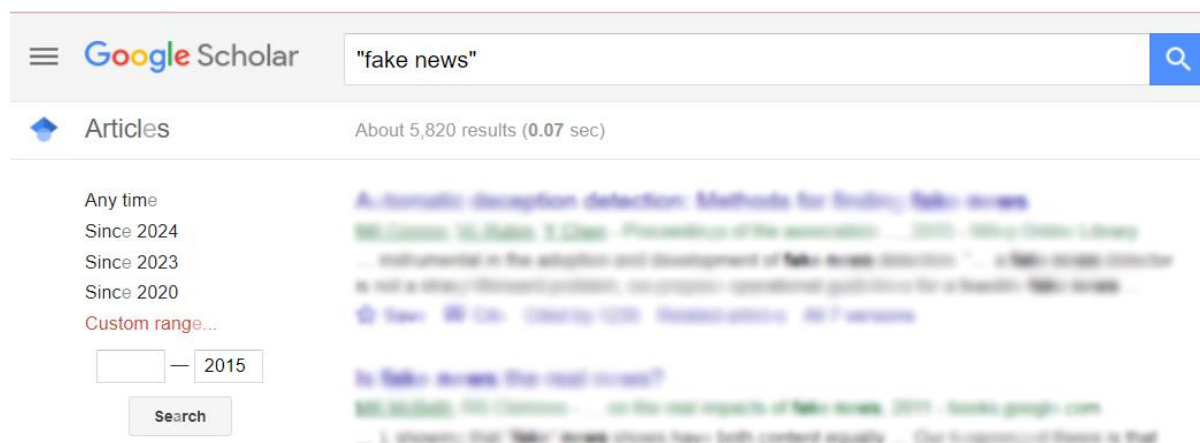
⁸² <https://dl.acm.org/doi/pdf/10.1145/2872518.2890098> (προσπελάστηκε στις 26/01/2024)

3.2 ΕΡΕΥΝΗΤΙΚΕΣ ΕΡΓΑΣΙΕΣ, ΣΥΣΤΗΜΑΤΑ ΚΑΙ DATASETS

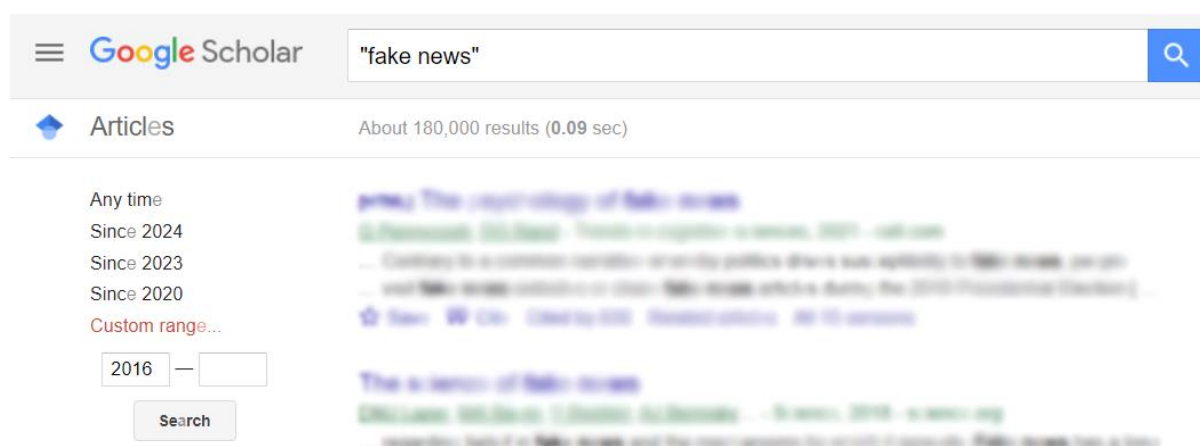
Από το 2016 έχουν δημοσιευτεί 180 000 επιστημονικά άρθρα όπου αναφέρεται ο όρος «fake news», ενώ πριν το 2016 ήταν μόνο 5 820, που σημαίνει ότι σήμερα αυτό το θέμα είναι πολύ πιο δημοφιλές από παλιά (Εικόνες 3.2.1 και 3.2.2). Οι ερευνητικές εργασίες πάνω σε αυτό το θέμα χωρίζονται συνήθως σε δύο κατηγορίες:

1. Ανάλυση του προβλήματος των ψευδών ειδήσεων
2. Ανάλυση διάφορων μεθόδων ανίχνευσης ή/κα ανάπτυξη νέου/νέων μεθόδων ανίχνευσης (Πίνακα 3.2.1).

Εικόνα 3.2.1 Επιστημονικά δημοσιεύματα πριν το 2016, που περιλαμβάνουν τον όρο «fake news»



Εικόνα 3.2.2 Επιστημονικά δημοσιεύματα από το 2016 που περιλαμβάνουν τον όρο «fake news»



Πίνακας 3.2.1 Προτεινόμενες προσεγγίσεις ανίχνευσης ψευδών ειδήσεων

ΕΡΕΥΝΗΤΗΣ	ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΤΡΟΠΟΣ
Bhoir (2020) ⁸³	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ επαναλαμβανόμενα νευρωνικά δίκτυα GRU ◦ Naïve Bayes ◦ δέντρο απόφασης
Das και Clark (2019) ⁸⁴	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ επαναλαμβανόμενα νευρωνικά δίκτυα GRU ◦ νευρωνικά δίκτυα LSTM ◦ Android software
Han και Mehta (2019) ⁸⁵	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ υβριδικά συνελεκτικά νευρωνικά δίκτυα (CNN) ◦ επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) ◦ Naïve Bayes
Iwendi, Mohan, Ibeke, Ahmadian και Ciano (2022) ⁸⁶	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ πολυκλαδικά συνελεκτικά νευρωνικά δίκτυα (CNN) ◦ επαναλαμβανόμενα νευρωνικά δίκτυα (RNN)
Agudelo, Parra και Velandia (2018) ⁸⁷	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ Naïve Bayes με τις 2 μορφές του tokenization - CountVectorizer και TfidfVectorizer
Klyuev (2019) ⁸⁸	Επεξεργασία φυσικής γλώσσας
Oshikawa, Qian και Wang (2018) ⁸⁹	Επεξεργασία φυσικής γλώσσας
Braşoveanu και Andonie (2019) ⁹⁰	Επεξεργασία φυσικής γλώσσας και σημασιολογία

⁸³ <https://ieeexplore.ieee.org/document/9104177/> (προσπελάστηκε στις 26/01/2024)

⁸⁴ <https://cs.pomona.edu/~ajc/pdf/Das.2019.TA.SatireVsFake.pdf> (προσπελάστηκε στις 26/01/2024)

⁸⁵ <https://ieeexplore.ieee.org/abstract/document/9065138/> (προσπελάστηκε στις 26/01/2024)

⁸⁶ <https://www.sciencedirect.com/science/article/pii/S0045790622002439> (προσπελάστηκε στις 26/01/2024)

⁸⁷ https://inria.hal.science/hal-02274166/file/474698_1_En_52_Chapter.pdf (προσπελάστηκε στις 26/01/2024)

⁸⁸ <https://ieeexplore.ieee.org/document/8748506/> (προσπελάστηκε στις 26/01/2024)

⁸⁹ <http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.747.pdf> (προσπελάστηκε στις 26/01/2024)

⁹⁰ https://link.springer.com/chapter/10.1007/978-3-030-20521-8_54 (προσπελάστηκε στις 26/01/2024)

Awan, Yasin, Nobanee, Ali, Shahzad, Nabeel, Zain και Shahzad (2021) ⁹¹	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ random forest classifier ◦ logistic regression ◦ term frequency-inverse document frequency (TF-IDF) vectorizer
Mayakul, Srisawat και Kiattisin (2019) ⁹²	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ αυτόματο μοντέλο συμπερασμάτων
Cui, Wang και Lee (2019) ⁹³	Μηχανική μάθηση <ul style="list-style-type: none"> ◦ νευρωνικά δίκτυα ◦ bag of words
Sousa-Silva (2022) ⁹⁴	Εγκληματολογική γλωσσολογία
Jang, Geng, Li, Xia, Huang, Kim και Tang (2018) ⁹⁵	Προσδιορισμός πηγής προέλευσης

Jang, Geng, Li, Xia, Huang, Kim και Tang (2018)⁹⁶ εξάγουν 307 738 tweets που περιλαμβάνουν τριάντα ψευδείς και τριάντα αληθινές ιστορίες ως ένα dataset και εφαρμόζουν την προσέγγιση προσδιορισμού πηγής προέλευσης. Η παρατήρησή τους μετά την εφαρμογή αυτής της μεθόδου στα δεδομένα τους είναι ότι το 87% των ψευδών ιστοριών πρωτοδημοσιεύτηκε από απλούς ανθρώπους και όχι από δημοσιογράφους, πολιτικούς ή διάσημα πρόσωπα.

Iwendi, Mohan, Ibeke, Ahmadian και Ciano (2022)⁹⁷ κάνουν έμφαση στις ψευδείς ειδήσεις εν μέσω της πανδημίας του Covid-19 και χρησιμοποιούν ένα ήδη έτοιμο dataset «Explore Covid-19 infodemic» από Li Susan (2020)⁹⁸. Το dataset αυτό αποτελείται από 586 αληθείς ειδήσεις, 578 ψευδείς ειδήσεις και πάνω από 1100 ενημερωτικά άρθρα που αφορούν τον κορονοϊό. Πριν

⁹¹ <https://www.mdpi.com/2079-9292/10/19/2326> (προσπελάστηκε στις 26/01/2024)

⁹² <https://ieeexplore.ieee.org/document/9024427/> (προσπελάστηκε στις 26/01/2024)

⁹³ <https://dl.acm.org/doi/pdf/10.1145/3341161.3342894> (προσπελάστηκε στις 26/01/2024)

⁹⁴ <https://link.springer.com/article/10.1007/s11196-022-09901-w#citeas> (προσπελάστηκε στις 26/01/2024)

⁹⁵ <https://www.sciencedirect.com/sdfe/reader/pii/S0747563218300906/pdf> (προσπελάστηκε στις 26/01/2024)

⁹⁶ <https://www.sciencedirect.com/sdfe/reader/pii/S0747563218300906/pdf> (προσπελάστηκε στις 26/01/2024)

⁹⁷ <https://www.sciencedirect.com/science/article/pii/S0045790622002439> (προσπελάστηκε στις 26/01/2024)

⁹⁸ <https://towardsdatascience.com/explore-covid-19-infodemic-2d1ceaae2306> (προσπελάστηκε στις 26/01/2024)

εφαρμοστεί η προτεινομένη από τους ερευνητές μέθοδος ανίχνευσης, γίνεται η επεξεργασία δεδομένων και το tokenization. Όταν τα απαραίτητα χαρακτηριστικά εξάχθηκαν, γίνεται η εφαρμογή της μεθόδου. Τα αποτελέσματα της εφαρμογής του αλγορίθμου ανίχνευσης τους είναι:

1. Στις αληθινές ειδήσεων οι πιο συχνά χρησιμοποιημένες λέξεις είναι covid, people και virus με σχεδόν ίση συχνότητα.
2. Στις ψευδείς ειδήσεις η πιο συχνά χρησιμοποιημένη λέξη είναι coronavirus.

Awan, Yasin, Nobanee, Ali, Shahzad, Nabeel, Zain και Shahzad (2021)⁹⁹ χρησιμοποιούν ένα dataset με αναληθείς ειδήσεις όπου κάθε αρχείο περιέχει πάνω από 20 000 ψευδείς και αληθείς ειδήσεις. Μετά την ανάλυση των δεδομένων γίνεται η προ-επεξεργασία, τα δεδομένα χωρίζονται στα σύνολα για την εκπαίδευση και το τεστάρισμα. Έπειτα, η νέος μέθοδος εφαρμόζεται. Οι παρατηρήσεις των ερευνητών είναι ότι στους τίτλους των αληθών ειδήσεων επικρατούν οι λέξεις όπως trump, say, new, house. Ενώ στους τίτλους των ψευδών ειδήσεων πιο συχνά αναφέρονται οι λέξεις όπως trump, video, watch, obama. Αυτό το γεγονός υποστηρίζει την ιδέα ότι στις ψευδείς ειδήσεις χρησιμοποιούνται διάφοροι τρόποι για να τραβήξουν την προσοχή του αναγνώστη, πχ οι λέξεις όπως video και watch.

Karusta, Benko και Munk (2020)¹⁰⁰ χρησιμοποίησαν τρία datasets: ένα dataset των πραγματικών ειδήσεων με 15 707 άρθρα για ανάλυση, ένα dataset των ψευδών ειδήσεων με 12 761 άρθρα για ανάλυση, ένα dataset των ψευδών και αληθών ειδήσεων με 402 άρθρα για επικύρωση. Έπειτα οι παρακάτω τεχνικές εφαρμόστηκαν σε αυτά τα datasets:

1. Υπολογίστηκε ο μέσος αριθμός λέξεων σε κάθε άρθρο.
2. Τα stop-words απομονώθηκαν και ο μέσος αριθμός λέξεων σε κάθε άρθρο χωρίς stop-words υπολογίστηκε.
3. Υπολογίστηκε το ποσοστό συναισθήματος για κάθε άρθρο.

Τα αποτελέσματα της ανάλυσής τους είναι ότι στα άρθρα με ψευδείς ειδήσεις χρησιμοποιήθηκαν λιγότερες λέξεις και οι προτάσεις είναι πιο σύνθετες σε σύγκριση με τα άρθρα με αληθινές ειδήσεις.

⁹⁹ <https://www.mdpi.com/2079-9292/10/19/2326> (προσπελάστηκε στις 26/01/2024)

¹⁰⁰ https://link.springer.com/chapter/10.1007/978-3-030-36778-7_44 (προσπελάστηκε στις 26/01/2024)

ΚΕΦΑΛΑΙΟ 4

Η ΥΠΗΡΕΣΙΑ ELLINIKAHOAXES.GR

4.1 ΙΔΡΥΣΗ ΚΑΙ ΛΕΙΤΟΥΡΓΙΑ

Ο μην κερδοσκοπικός οργανισμός Ellhnikahoaxes.gr είναι ο πρώτη υπηρεσία ελέγχου αναληθών δημοσιευμάτων στην Ελλάδα. Ιδρύθηκε το 2013 στη Θεσσαλονίκη για να αντιμετωπίσει τεράστιο όγκο πληροφοριών στο Διαδίκτυο και τα ΜΜΕ. Είναι μέλος του Παγκόσμιου Δικτύου Ελέγχου Γεγονότος IFCN, Ευρωπαϊκού Παρατηρητηρίου κατά της Παραπληροφόρησης (SOMA)¹⁰¹, του Ευρωπαϊκού Παρατηρητηρίου Ψηφιακών Μέσων (EDMO)¹⁰², είναι συνεργάτης της πλατφόρμας Facebook στον τομέα ελέγχου γεγονότων, συμμετείχε στο πρόγραμμα FactCheckEU.info¹⁰³ και η αξιοπιστία του οργανισμού ελέγχεται κάθε χρόνο.

Το όραμα του οργανισμού είναι η αναλυτική και εύληπτη αποδόμηση παραπλανητικών ισχυρισμών που διαμοιράζονται στο Διαδίκτυο.

Η αποστολή του οργανισμού είναι η παροχή διασταυρωμένης πληροφόρησης και, συγχρόνως, των κατάλληλων εργαλείων που θα βοηθήσουν τους πολίτες στην ανεπηρέαστη, συνειδητή και ορθά ενημερωμένη λήψη αποφάσεων.

Κάθε ελεγκτής γεγονότων του οργανισμού έχει το δικό του πεδίο απασχόλησης. Η ομάδα ελεγκτών αποτελείται από τον Γιώργο Γιώτη με μεταπτυχιακό στη Διοίκηση Επιχειρήσεων, τον Στυλιανό Πουρνή με μεταπτυχιακό στη Δημοσιογραφία, τον Αντώνη Παπαθανασίου (ανεξάρτητος δημοσιογράφος στους τομείς της επιστήμης και της τεχνολογίας), τον Ιωάννη Δελημάρη με προπτυχιακό στη Δημοσιογραφία, την Αλίκη Βασιλείου με προπτυχιακό στη Δημοσιογραφία, τον Αλέξανδρο Τάσσο με μεταπτυχιακό στη Δημοσιογραφία, την Έλενα Βατάλα με δυο μεταπτυχιακούς τίτλους στη Γλωσσολογία και τη Σημειωτική και στην Επικοινωνία, τον Άγγελο Παλιοκούδη με μεταπτυχιακό στην Ευρωπαϊκή Δημοσιογραφία και τον Ορφέα Τσαυταρίδη με προπτυχιακό στη Δημοσιογραφία.

Επίσης, η υπηρεσία αυτή συνεργάζεται με τους εξής επιστημονικούς συμβούλους: τον ιστορικό-διεθνολόγο Ευστάθιο Βασιλείου, τη δικηγόρο Μαριάννα Βασιλείου, τον ορθοπεδικό Νικόλαο Γεωργόπουλο, τον ηλεκτρονικό μηχανικό και τεχνικό υπολογιστικών συστημάτων

¹⁰¹ <https://www.disinfobservatory.org/the-observatory/> (προσπελάστηκε στις 26/01/2024)

¹⁰² <https://edmo.eu/list-of-institutions-connected-to-edmo/> (προσπελάστηκε στις 26/01/2024)

¹⁰³ <https://web.archive.org/web/20201012171836/https://factcheckeu.info/el/about> (προσπελάστηκε στις 26/01/2024)

και δικτύων Κωνσταντίνο Γωνιάδη, τον παιδίατρο-βρεφολόγο Αντώνιο Δαρζέντα, τον αεροναυπηγό Φαίδωνα Καραιωσηφίδη, τον μικροβιολόγο Κίμωνα Ανδρέα Καρατζά, τον κλινικό φαρμακοποιό Διαμαντή Κλιμεντίδη, τον τεχνικό τροφίμων Στυλιανό Κουλούρη, τον διατροφολόγο-διαιτολόγο Εμμανουήλ Μανωλαράκη, τον καθηγητή της πολιτικής της υγείας Ηλία Μόσιαλο, τον ψυχίατρο-ψυχοθεραπευτή Αθανάσιο Μπασδρά, τον παιδίατρο Κωνσταντίνο Νταλούκα, τον ερευνητή του Center for Cancer Research (National Cancer Institute) Γεώργιο Παυλάκη, τον βιολόγο Κωνσταντίνο Φλωρίδη και τον ιστορικό Ιάσονα Χανδρινό.

4.2 ΧΡΗΜΑΤΟΔΟΤΗΣΗ

Ο οργανισμός είναι μην κερδοσκοπικός και από το 2013 μέχρι 2018 λειτουργούσε χωρίς καμία χρηματοδότηση εξαιρετικά σε εθελοντική βάση. Το 2019 ξεκίνησε να λαμβάνει τις πρώτες χρηματοδοτήσεις. Αφού ξεκίνησε να συνεργάζεται με το Facebook το 2019, το μεγαλύτερο ποσοστό των χρηματοδοτήσεων τα τελευταία χρόνια αποτελείται από τις χρηματοδοτήσεις της πλατφόρμας Facebook. Επίσης, η υπηρεσία λαμβάνει οικονομική υποστήριξη από το IFCN.

4.3 ΔΙΑΔΙΚΑΣΙΑ ΕΠΙΛΟΓΗΣ ΚΑΙ ΕΞΕΤΑΣΗΣ ΕΙΔΗΣΕΩΝ

Πριν ξεκινήσει η διαδικασία ελέγχου μιας είδησης, αυτή η είδηση πρέπει να επιλεγεί και να εξεταστεί από ένα μέλος της ομάδας ελεγκτών, τον ιδρυτή, τον αρχισυντάκτη και τον συντονιστή της ιστότοπου της υπηρεσίας.

Ο οργανισμός λαμβάνει από δεκάδες έως εκατοντάδες αιτήματα για τον έλεγχο ειδήσεων και ισχυρισμών από τους αναγνώστες του κάθε βδομάδα. Έπειτα, επιλέγεται το υποψήφιο θέμα, το οποίο εξετάζεται εάν ταιριάζει στον ρόλο του οργανισμού. Την προτεραιότητα έχουν οι ειδήσεις που μπορούν να επηρεάσουν μεγάλο κοινό ή οι ειδήσεις που στοχοποιούν ευαίσθητες κοινωνικές ομάδες. Μετά, γίνεται μια προκαταρκτική έρευνα, η οποία έπειτα ελέγχεται για τυχόν λάθη και παραλήψεις. Αφού κριθεί κατάλληλη για την περαιτέρω ανάλυση, ξεκινάει η κύρια έρευνα, κατά την οποία ο ελεγκτής προσπαθεί επικοινωνήσει με την αρχική αρχή του ισχυρισμού. Αν η επικοινωνία αυτή δεν είναι δυνατή ή οι ειλημμένες πληροφορίες δεν είναι αρκετές, γίνεται προσπάθεια επικοινωνίας με άτομα ή οργανισμούς που ενδέχεται να γνωρίζουν αυτό το θέμα καλύτερα. Επίσης, κατά τη διάρκεια της εξέτασης της είδησης ο ελεγκτής μπορεί να επικοινωνήσει με τους επιστημονικούς συμβούλους που συνεργάζονται με την υπηρεσία. Πριν δημοσιευθεί, το άρθρο πρέπει να περάσει από τουλάχιστον ένα από τους αρχαιότερους συντάκτες.

4.4 ΜΕΘΟΛΟΓΙΑ ΕΛΕΓΧΟΥ ΕΥΚΥΡΟΤΗΤΑΣ ΜΙΑΣ ΕΙΔΗΣΗΣ

Επειδή κάθε ισχυρισμός που ελέγχεται από το EllhnikaHoaxes.gr είναι μια διαφορετική περίπτωση που μπορεί να προέρχεται από μια πληθώρα πηγών, δεν υπάρχει μια ορισμένη μεθοδολογία ελέγχου. Ωστόσο, σε γενικευμένη μορφή η μεθοδολογία ελέγχου ακολουθεί τα παρακάτω βήματα:

1ο. Εντοπισμός δυνητικά ύποπτου υλικού:

Σε αυτό το βήμα το υλικό εξετάζεται οπτικά για ό,τι μπορεί να χρησιμοποιείται για να τραβήξει την προσοχή του αναγνώστη, όπως οι τίτλοι με κεφαλαία γράμματα και έντονα συναισθηματική γλώσσα, η απουσία αρθρογράφου ή η πρόταση για άμεση κοινοποίηση. Δεν είναι πάντα δυνατό να κριθεί αν το άρθρο είναι αληθές ή ψευδές σε αυτό το βήμα, αλλά αυτό το βήμα βοηθάει τον ελεγκτή να καταλάβει τι και που πρέπει να ψάχνει να βρει στοιχεία για να αποδείξει ή διαψεύσει την εγκυρότητα του υλικού.

2ο. Ανάλυση του περιεχομένου:

Όλοι οι ισχυρισμοί που υπάρχουν σε αυτό το υλικό καταγράφονται και ελέγχονται ξεχωριστά. Ο ελεγκτής προσπαθεί να επικοινωνήσει με την αρχική πηγή.

3ο. Έρευνα στο οπτικοακουστικό υλικό:

Αν το εξεταζόμενο υλικό περιέχει εικόνες, κάθε εικόνα ελέγχεται μέσω της υπηρεσίας Google Images. Με τη χρήση αυτού του εργαλείου ελέγχεται εάν η εικόνα, που χρησιμοποιείται στο εξεταζόμενο υλικό, έχει ήδη δημοσιευτεί στο Διαδίκτυο συμπεριλαμβανομένου κειμένου που δεν έχει καμία σχέση με τον ισχυρισμό στον εξεταζόμενο υλικό. Καμιά φορά οι φωτογραφίες μπορεί να είναι αλλοιωμένες ή η υπηρεσία Google Images να μη δίνει αποτέλεσμα. Σε αυτή την περίπτωση ο ελεγκτής απευθύνεται σε εμπειρογνώμονες στον κλάδο, με τον οποίον σχετίζεται ο ισχυρισμός, και επιστημονικούς συμβούλους.

Αν το εξεταζόμενο υλικό περιέχει βίντεο, η διαδικασία γίνεται πιο δύσκολη, γιατί ο ελεγκτής πρέπει να βγάλει πολλά στιγμιότυπα από διαφορετικά σημεία του βίντεο και να κάνει αναζήτηση των στιγμιότυπων αυτών με τη χρήση της Google Images.

Αν το εξεταζόμενο υλικό περιέχει κάποιες συγκεκριμένες διευθύνσεις, για να επιβεβαιώσει ή διαψεύσει τον ελεγχόμενο ισχυρισμό, μπορούν να χρησιμοποιηθούν οι υπηρεσίες Google Maps και Google Street View.

4ο. Εξέταση επιστημονικών μελετών σε περιπτώσεις ψευδοεπιστημονικών ισχυρισμών:

Ο ελεγκτής εξετάζει κάθε ισχυρισμό και η μεθοδολογία αυτή εξαρτάται από το πώς αυτός ο ισχυρισμός αναφέρεται στο εξεταζόμενο υλικό.

- Αν στον ισχυρισμό αυτό αναφέρεται η έρευνα, στην οποία βασίζεται ο ισχυρισμός, τότε είναι εύκολο για τον ελεγκτή να βρει αυτήν την έρευνα. Ωστόσο, όχι όλοι οι δημοσιεύσεις που παρουσιάζονται ως «έρευνες» είναι έγκυρες. Λαμβάνοντας υπόψη αυτό, το κείμενο της έρευνας περνάει από μια διπλά τυφλή εξέταση από ομότιμους, δηλαδή το κείμενο ελέγχεται από δυο ή περισσότερους ελεγκτές και ο κανένας από αυτούς δεν ξέρει ούτε ποιος άλλος ελέγχει αυτή την έρευνα, ούτε ποιανού αυτή η έρευνα είναι. Επίσης, υπάρχουν έτοιμες λίστες με περιοδικά που δεν έχουν καμία εγκυρότητα και δημοσιεύουν έρευνες επί πληρωμή.
- Αν στον ισχυρισμό δεν αναφέρεται ούτε μια συγκεκριμένη μελέτη, ούτε το όνομα του ερευνητή, ο ελεγκτής προσπαθεί να βρει το πρωτότυπο άρθρο χρησιμοποιώντας λέξεις-κλειδιά του ισχυρισμού. Για παράδειγμα, ο ισχυρισμός είναι «Οι επιστήμονες του Πανεπιστημίου Γέιλ λένε ότι οι άνθρωποι που καταναλώνουν πολλή ζάχαρη, ζουν περισσότερο!», τότε ο ερευνητής θα αναζητάει επιστημονικές εργασίες που περιλαμβάνουν τις λέξεις «Yale University», «sugar», «live longer», «longer life». Αν η έρευνα που περιλαμβάνει αυτές τις λέξεις-κλειδιά όντως υπάρχει, ελέγχεται αν αυτός ο ισχυρισμός πράγματι υπάρχει σε αυτή τη μελέτη

Στο τέλος, εάν είναι εφικτό ο ελεγκτής βρίσκει περισσότερες μελέτες για αυτό το θέμα για να επιβεβαιωθεί, ότι τα συμπεράσματα που έβγαλε, συσχετίζονται με τα συμπεράσματα άλλων ερευνητών.

5ο. Επικοινωνία με άλλες ομάδες ελέγχου ειδήσεων:

Αφού ο ελεγκτής εξαντλήσει όλα τα άλλα εργαλεία ελέγχου, μπορεί να επικοινωνήσει με άλλους οργανισμούς που επίσης έχουν πιστοποιηθεί από το IFCN για να μάθει αν αυτοί έχουν έρθει σε επαφή με αυτή την είδηση στη χώρα τους.

4.5 ΧΡΗΣΙΜΟΠΟΙΗΜΕΝΕΣ ΚΑΤΗΓΟΡΙΕΣ


Η υπηρεσία EllinikaHoaxes.gr αυτή τη στιγμή χρησιμοποιεί δύο τρόπους κατηγοριοποίησης των άρθρων τους: μέσω χωρισμού σε κατηγορίες στην αρχή της σελίδας (σε μορφή κειμένου) και μέσω χρήσης εικόνων-σφραγίδων (σε μορφή εικόνας). Καμιά φορά αυτές οι δυο κατηγοριοποιήσεις δείχνουν διαφορετικά αποτελέσματα, πχ στο άρθρο <https://www.ellinikahoaxes.gr/2018/12/21/nato-turkey-greece/> στην αρχή της ιστοσελίδας

αναφέρεται η κατηγορία «Hoaxes», ενώ οι εικόνες-σφραγίδες που χρησιμοποιούνται εκεί είναι «Μίξη γεγονότων και παραποιήσεων» και «Σάτιρα». Σε κάποια άρθρα η πρώτη κατηγοριοποίηση είναι κενή, σε κάποια είναι κενή η δεύτερη, σε κάποια και οι δυο δείχνουν τα ίδια και στα υπόλοιπα οι κατηγορίες δεν αναφέρονται καθόλου. (Αυτό το αντικείμενο εξετάζεται λεπτομερώς στο Κεφάλαιο 5.)

Όλες οι κατηγορίες που χρησιμοποιούνται από τον EllnikaHoaxes.gr παρουσιάζονται στον Πίνακα 4.5.1.

Πίνακας 4.5.1 Κατηγορίες που χρησιμοποιούνται από την υπηρεσία EllnikaHoaxes.gr

ΚΑΤΗΓΟΡΙΑ	ΕΙΚΟΝΑ-ΣΦΡΑΓΙΔΑ
ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	 <u>Παραπληροφόρηση</u>
ΨΕΥΔΕΣ	 <u>Ψευδές</u>
FAKE NEWS	Δεν έχει εικόνα
ΗΟΑΧΕΣ	Δεν έχει εικόνα
ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	 <u>Συνωμοσιολογία</u>
ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	 <u>Λείπει θεματικό περιεχόμενο</u>

<p>ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ</p>		<p>Ψευδής Ισχυρισμός</p>
<p>ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ</p>		<p>Τροποποιημένη εικόνα</p>
<p>ΕΠΗΚΑΙΡΟΤΗΤΑ</p>		<p>Δεν έχει εικόνα</p>
<p>ΜΙΞΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ</p>		<p>Μίξη γεγονότων και παραποιήσεων</p>
<p>ΨΕΥΔΟΕΠΙΣΤΗΜΗ</p>		<p>Ψευδοεπιστήμη</p>
<p>ΤΡΟΠΟΠΟΙΗΜΕΝΟ ΒΙΝΤΕΟ</p>		<p>Τροποποιημένο βίντεο</p>
<p>ΣΑΤΙΡΑ</p>		<p>Σάτιρα</p>
<p>ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ</p>		<p>Δεν έχει εικόνα</p>
<p>ΥΓΕΙΑ</p>		<p>Δεν έχει εικόνα</p>

ΚΙΝΔΥΝΟΛΟΓΙΑ		<u>Κινδυνολογία</u>
ΑΠΑΤΗ		<u>Απάτη</u>
LIKE FARMING		<u>Like farming</u>
ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ		Δεν έχει εικόνα
ΑΝΑΚΟΙΝΩΣΕΙΣ		Δεν έχει εικόνα
CLICKBAIT		<u>Clickbait</u>
ΕΡΕΥΝΕΣ		Δεν έχει εικόνα
ΑΝΑΚΥΚΛΩΜΕΝΗ ΕΙΔΗΣΗ		<u>Ανακυκλωμένη είδηση</u>
ΔΙΑΦΟΡΑ		Δεν έχει εικόνα

<p>ΕΙΚΟΝΑ ΑΠΟ ΠΡΟΓΡΑΜΜΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ</p>	 <p>Εικόνα από πρόγραμμα τεχνητής νοημοσύνης</p>
<p>ΑΝΕΠΙΒΕΒΑΙΩΤΟ</p>	 <p>Ανεπιβεβαίωτο</p>
<p>ΠΡΑΓΜΑΤΙΚΕΣ ΙΣΤΟΡΙΕΣ</p>	<p>Δεν έχει εικόνα</p>
<p>ΟΔΗΓΟΙ ΚΑΙ ΣΥΜΒΟΥΛΕΣ</p>	<p>Δεν έχει εικόνα</p>

ΚΕΦΑΛΑΙΟ 5

ΑΝΑΛΥΣΗ ΤΩΝ ΙΣΥΡΙΣΜΩΝ ΣΤΟΝ ELLINKAHOAXES.GR

5.1 ΣΚΟΠΟΣ

Η EllhnikaHoaxes.gr ιδρύθηκε το 2013 και σε αυτά τα ένδεκα χρόνια έχει γίνει μια πληθώρα αλλαγών τόσο στην Ελλάδα, όσο και σε όλο τον κόσμο. Επίσης, κατά αυτή την περίοδο η τεχνολογία εξελισσόταν σε ραγδαίους ρυθμούς με αποτέλεσμα το Internet έγινε κάτι συνηθισμένο και εύκολα προσβάσιμο από σχεδόν οποιοδήποτε σημείο του πλανήτη μας. Ο τομέας των ψευδών ειδήσεων επίσης αναπτυσσόταν όλα αυτά τα έτη. Οι εκλογές, τα δημοψηφίσματα και οι διαδικασίες μομφής, οι πόλεμοι και οι συρράξεις, οι επιδημίες και οι πανδημίες – όλα αυτά τα γεγονότα επηρέαζαν, επηρεάζουν και θα επηρεάζουν τις ζωές μας όχι μόνο επειδή αυτό μπορεί να μας επηρεάσει στον πραγματικό κόσμο αλλά και επειδή κάθε σημαντικό γεγονός είναι μια νέα πηγή για παραπληροφόρηση που μπορεί να μας επηρεάσει διαδικτυακά. Ο σκοπός της εργασίας μου είναι η ανάκτηση των δεδομένων από τον ιστόχωρο της EllhnikaHoaxes.gr και η ανάλυση των τάσεων των ψευδών ειδήσεων στην Ελλάδα με την πάροδο του χρόνου.

5.2 ΜΕΘΟΔΟΛΟΓΙΑ

Η εργασία χωρίζεται σε δύο μέρη:

- Τεχνικό: συγγραφή κώδικα για την ανάκτηση, την αποθήκευση και την επεξεργασία των δεδομένων, χρήση συναρτήσεων και άλλων εργαλείων του MS Excel για την επεξεργασία δεδομένων.
- Αναλυτικό: ανάλυση των στατιστικών δεδομένων και εξαγωγή σωστών πληροφοριών από αυτά τα δεδομένα.

Για να εξάγω, επεξεργαστώ και αναλύσω τα δεδομένα από τον EllhnikaHoaxes.gr, έσπασα την εργασία μου στα εξής στάδια:

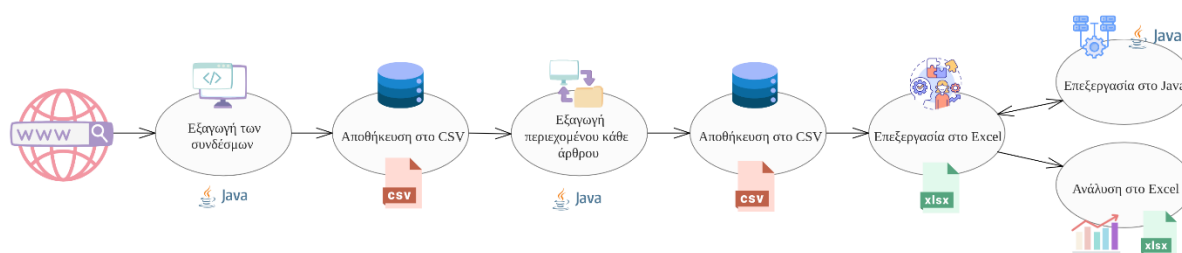
- 1ο. Εξαγωγή όλων των συνδέσμων των ψευδών ειδήσεων που ελέχθηκαν από τον EllhnikaHoaxes.gr. Για τον σκοπό αυτό χρησιμοποιείται ο κώδικας στη γλώσσα προγραμματισμού Java και οι σύνδεσμοι αποθηκεύονται στο CSV-αρχείο «fakeLinks.csv».
- 2ο. Εξαγωγή των κύριων στοιχείων κάθε συνδέσμου αποθηκευμένου στο CSV αρχείο «fakeLinks.csv», όπως ο τίτλος, ο ελεγκτής, η ημερομηνία δημοσίευσης, ο

εξεταζόμενος ισχυρισμός, το συμπέρασμα του ελεγκτή, η κατηγορία / οι κατηγορίες της ψευδούς είδησης και οι κατηγορίες-σφραγίδες της ψευδούς είδησης. Για τον σκοπό αυτό χρησιμοποιείται ο κώδικας στη γλώσσα προγραμματισμού Java και τα δεδομένα αυτά μαζί με το σύνδεσμο, με τον οποίο σχετίζονται, αποθηκεύονται στο CSV-αρχείο «fakeLinksData.csv».

- 3ο. Γίνεται επεξεργασία στο MS Excel. Για περαιτέρω ανάλυση γίνεται ένωση δυο τύπου κατηγοριών (που αναφέρονται στην Ενότητα 4.5), για κάποιους σκοπούς διαγράφονται διπλές εγγραφές βάση το σύνδεσμο που σχετίζονται με δύο ή περισσότερους ισχυρισμούς, οι εγγραφές φιλτράρονται και ομαδοποιούνται βάση διαφορά χαρακτηριστικά, stemming κ.α.
- 4ο. Σε κάποιες περιπτώσεις γίνεται ξανά η επεξεργασία με τη βοήθεια της γλώσσας Java (αντικατάσταση λέξεων κλειδιών βάση το stemming).
- 5ο. Τα δεδομένα ξανά επεξεργάζονται στο MS Excel, εάν είναι απαραίτητο.
- 6ο. Τα δεδομένα αναλύονται στο MS Excel (δημιουργούνται τα vocabulary corpora και τα γραφήματα).

Ολόκληρη η διαδικασία που προηγείται λήψης τελικού αποτελέσματος παρουσιάζεται στο Σχήμα 5.2.1 και εξηγείται λεπτομερώς στις Ενότητες 5.3 – 5.6.

Σχήμα 5.2.1 Διαδικασία εξαγωγής, επεξεργασίας και ανάλυσης δεδομένων από τον ιστότοπο EllhnikaHoaxes.gr



5.3 ΑΝΑΚΤΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟΝ ΙΣΤΟΤΟΠΟ

Πριν ξεκινήσω τη διαδικασία εξαγωγής δεδομένων, έπρεπε να εξετάσω τη δομή του ιστότοπου της EllhnikaHoaxes.gr σχολαστικά. Για να ανακτήσω όλα αυτά τα δεδομένα από τον ιστότοπο χρησιμοποίησα τις κατάλληλες μεθόδους μέσα στη γλώσσα Java που θα διαβάζουν τον κώδικα σε HTML.

5.3.1 ΑΝΑΚΤΗΣΗ ΤΩΝ ΣΥΝΔΕΣΜΩΝ ΤΩΝ ΑΡΘΡΩΝ

Για την ανάκτηση και την αποθήκευση των δεδομένων δημιουργήθηκε η κλάση `LinkRetrival`, όπου, αρχικά, διαβάζονται οι σύνδεσμοι ήδη αποθηκευμένοι στη βάση δεδομένων (η συνάρτηση `getOldLinks()`), έπειτα, διαβάζεται ο αριθμός των σελίδων με δημοσιευμένα άρθρα (η συνάρτηση `getLastPageNum()`) και, τελικά, προστίθενται οι σύνδεσμοι που ανιχνεύθηκαν (η συνάρτηση `addLinks()`) με τη βοήθεια της συνάρτησης `getURLContent()`, που εξάγει το περιεχόμενο της σελίδας.

5.3.1.1 ΔΙΑΒΑΣΜΑ ΤΩΝ ΗΔΗ ΑΠΟΘΗΚΕΥΜΕΝΩΝ ΣΤΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ ΣΥΝΔΕΣΜΩΝ

Ο στόχος αυτού του κώδικα δεν είναι απλά το διάβασμα και αποθήκευση των συνδέσμων αλλά και η αποδοτική λειτουργία. Αυτή τη στιγμή (στις 5 Φεβρουαρίου 2024) ο αριθμός των δημοσιευμένων στον ιστότοπο `Ellhnikahoaxes.gr` άρθρων είναι σχεδόν 6000 και το να επαναλαμβάνουμε τη διαδικασία ανάκτησης 6000 ή περισσότερων άρθρων κάθε φορά για να προστεθούν άλλοι σύνδεσμοι είναι αναποτελεσματικό. Για να αποφευχθεί το ανούσιο σκανάρισμα όλων των σελίδων συγγράφηκε μέθοδος `getOldLinks()`, που διαβάζει τους ήδη αποθηκευμένους στο CSV-αρχείο συνδέσμους `fakeLinks.csv` (σταθερή συμβολοσειρά `HOAXES_LINKS_FILENAME`) και τους αποθηκεύει ένα-ένα στο `HashSet oldRetrievedLinks`. Αν αυτό το αρχείο δεν υπάρχει ακόμα, η μέθοδος παύει να εκτελείται και το `HashSet oldRetrievedLinks` παραμένει κενό.

Εικόνα 5.3.1 Μέθοδος getOldLinks() της κλάσης LinkRetrieval

```
/*Reads the file with old links and adds these links
to the hashset "alreadyRetrievedLinks".
*/
public static void getOldLinks() throws IOException {
    System.out.println(
        "# Reading the links in the text file '"
        + HOAXES_LINKS_FILENAME + "'");
    File file = new File(HOAXES_LINKS_FILENAME);

    try (BufferedReader in = new BufferedReader(
        new InputStreamReader(
            new FileInputStream(file),
            "UTF8"))
    ) {

        String str;

        while ((str = in.readLine()) != null) {

            if (!str.trim().equals("")) {
                oldRetrievedLinks.add(str);
            }
        }
    } catch (Exception e) {
        System.out.println(e.getMessage());
    }
    System.out.println(
        " - Total number of links in the file: "
        + oldRetrievedLinks.size());
}
```

5.3.1.2 ΑΝΑΚΤΗΣΗ ΤΟΥ ΑΡΙΘΜΟΥ ΤΗΣ ΤΕΛΕΥΤΑΙΑΣ ΣΕΛΙΔΑΣ

Σε κάθε σελίδα υπάρχουν 20 άρθρα (μπορεί να είναι λιγότερα στην τελευταία σελίδα) και η ανάκτηση του HTML κώδικα κάθε σελίδας γίνεται με τη βοήθεια του βρόγχου, άρα είναι εξαιρετικά σημαντικό να εκτελείται αυτός ο βρόγχος τόσες φορές, όσες είναι οι σελίδες στο website.

Η πρώτη σελίδα με τα άρθρα βρίσκεται στη διεύθυνση <https://www.ellinikahoaxes.gr/category/kathgories/> και στο κάτω μέρος της σελίδας υπάρχει η σελιδοαρίθμηση (αγγλ. pagination), όπου ο χρήστης μπορεί να κλικάρει και να μεταφερθεί στην πρώτη, στην προηγούμενη, στην επόμενη και στην τελευταία σελίδες (Εικόνα 5.3.2). Για τη σωστή εκτέλεση του κώδικα χρειάζεται ο αριθμός των σελίδων ή απλά ο αριθμός της τελευταίας σελίδας.

Εικόνα 5.3.2 Σελιδοαρίθμηση στον Ellhnikahoaxes.gr



Για να εξάγω αυτό το στοιχείο από τη σελίδα έπρεπε να βρω που ακριβώς στον κώδικα βρίσκεται αυτό το στοιχείο και να εξάγω κομμάτι-κομμάτι το μέρος του κώδικα, όπου είναι το στοιχείο. Το αποτέλεσμα αποθηκεύτηκε στη μεταβλητή (συμβολοσειρά) lastPageNumStr. Επειδή η συνάρτηση πρέπει να επιστρέφει ένα αριθμό και όχι τη γραμματοσειρά, χρησιμοποιήθηκε η συνάρτηση parseInt() της κλάσης Integer (Εικόνες 5.3.3 – 5.3.5).

Εικόνα 5.3.3 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (1)

```
</div>
  <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-proka...
  <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-proka...
  </div>
  </div>
  </div>
  </div>
  <div class="col-md-6 col-12 front-news archive-post post-227362 post type-post status-publish format-standard has-post-thumbnail hentry c...
  4376 tag-axiomatouchos tag-pezonaftes-2 tag-chikwe-ihekweazu" id="post-227362"> </div>
</div>
<div class="pagination-list text-center" style>
  <span aria-current="page" class="page-numbers current" style>1</span>
  <a class="page-numbers" href="https://www.ellinikahoaxes.gr/category/kathgories/page/2/" data-wpel-link="internal" style>2</a>
  <a class="page-numbers" href="https://www.ellinikahoaxes.gr/category/kathgories/page/3/" data-wpel-link="internal" style>3</a>
  <span class="page-numbers dots" style>...</span>
  <a class="page-numbers" href="https://www.ellinikahoaxes.gr/category/kathgories/page/275/" data-wpel-link="internal" style>275</a> == $0
  <a class="next page-numbers" href="https://www.ellinikahoaxes.gr/category/kathgories/page/2/" data-wpel-link="internal" style>⏪</a>
</div>
</div>
  <div class="col-md-3 col-sm-12 col-sm-12 sidebar" style="height: auto !important;"> </div>
</div>
</div>
</div>
<section id="footer" style> </section>
<section id="copyright"> </section>
<script id="google_cse_v2-js-extra"> </script>
<script src="https://www.ellinikahoaxes.gr/wp-content/plugins/wp-google-search/assets/js/google_cse_v2.js?ver=1" id="google_cse_v2-js"></script>
<script id="collapseomatic-js-js-before"> </script>
<script src="https://www.ellinikahoaxes.gr/wp-content/plugins/jquery-collapse-o-matic/js/collapse.js?ver=1.7.2" id="collapseomatic-js-js"></script>
<script id="rocket-browser-checker-js-after"> </script>
<script id="rocket-preload-links-js-extra"> </script>
html body.archive.category.category-kathgories.category-2724.wp-embed-responsive.multiple-domain-www-ellinikahoaxes-gr div.white-row div.container div.row div.col-lg-9.col-
```

Εικόνα 5.3.4 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (2)

```
<span class="page-numbers dots" style>...</span>
<a class="page-numbers" href="https://www.ellinikahoaxes.gr/category/kathgories/page/275/" data-wpel-link="internal" style>275</a> == $0
```

Εικόνα 5.3.5 Αριθμός της τελευταίας σελίδας στο HTML-κώδικα (3)

```
kathgories/page/275/" data-wpel-link="internal"
```

Εικόνα 5.3.6 Μέθοδος `getLastPageNum()` της κλάσης `LinkRetrieval`

```
/*Reads the content of the first page and retrieves the number of the last
page from the bottom of the page.
*/
public static int getLastPageNum() throws IOException{
    String firstPage = "https://www.ellinikahoaxes.gr/category/kathgories/";
    String firstPageContent = getURLContent(firstPage);
    int lastPageNumElementStart = firstPageContent.indexOf(
        "\"page-numbers dots\"");
    int lastPageNumElementEnd = firstPageContent.indexOf(
        "</a>", lastPageNumElementStart);
    String lastPageNumElement = firstPageContent.substring(
        lastPageNumElementStart, lastPageNumElementEnd);
    int lastPageNumStart = lastPageNumElement.indexOf(
        "kathgories/page/");
    int lastPageNumEnd = lastPageNumElement.indexOf(
        "/\" data-wpel-link=\"internal\"");
    String lastPageNumStr = lastPageNumElement.substring(
        lastPageNumStart+16, lastPageNumEnd);
    System.out.println("Total number of pages: " + lastPageNumStr);
    int lastPage = Integer.parseInt(lastPageNumStr);
    return lastPage;
}
```

Επίσης, λαμβάνοντας υπόψη τον τρόπο μέτρησης σε βρόγχους (η τελευταία τιμή δεν υπολογίζεται), στον αριθμό αυτό προστίθεται ένα.

Εικόνα 5.3.7 Επεξεργασία του αριθμού της τελευταίας σελίδας (`maxPage`)

```
int maxPage = getLastPageNum();
int numberOfNewLinks = addLinks(linkFile,maxPage+1);
```

5.3.1.3 ΑΝΑΚΤΗΣΗ ΠΕΡΙΕΧΟΜΕΝΟΥ ΜΙΑΣ ΣΕΛΙΔΑΣ

Πρώτον, δημιουργείται το αντικείμενο της κλάσης `URL`. Με τη χρήση της μεθόδου `openConnction()` πραγματοποιείται η σύνδεση μεταξύ του συνδέσμου (`URL`) και της εφαρμογής. Μόλις πραγματοποιηθεί η σύνδεση, κάθε γραμμή του `HTML`-κώδικα διαβάζεται με τη βοήθεια της μεθόδου `getInputStream()` της κλάσης `URLConnection` και προστίθεται στη συμβολοσειρά `pageContent`.

Εικόνα 5.3.8 Μέθοδος `getURLContent()` της κλάσης `LinkRetrieval`

```
/* Scans the HTML code of the webpage and returns the HTML code as a string
*/
public static String getURLContent(String urlPath) throws
    MalformedURLException, IOException {
    URL fakeNewsPage = new URL(urlPath);
    URLConnection fakeNewsCon = fakeNewsPage.openConnection();

    String pageContent;

    try (BufferedReader in = new BufferedReader(
        new InputStreamReader(
            fakeNewsCon.getInputStream(), "UTF-8")
    )) {
        pageContent = "";
        String inputLine;

        while ((inputLine = in.readLine()) != null) {
            pageContent += (inputLine + "\n");
        }
    }

    return pageContent;
}
```

5.3.1.4 ΑΝΑΚΤΗΣΗ ΚΑΙ ΑΠΟΘΗΚΕΥΣΗ ΕΝΟΣ ΣΥΝΔΕΣΜΟΥ

Για να ανακτήσω συνδέσμους πρώτον έπρεπε να δημιουργήσω ένα βρόγχο που εκτελείται όσες είναι οι σελίδες με τα άρθρα (η έξοδος της συνάρτησης `getLastPageNum() + 1`).

Η πρώτη σελίδα έχει τη διεύθυνση <https://www.ellinikahoaxes.gr/category/kathgories/>, ενώ οι επόμενες έχουν τη διεύθυνση <https://www.ellinikahoaxes.gr/category/kathgories/page/2/>, <https://www.ellinikahoaxes.gr/category/kathgories/page/3/> και ούτω καθεξής. Άρα η αλλαγή της σελίδας στην επομένη γίνεται με τη βοήθεια της μεταβλητής `n`, η οποία αυξάνεται με την κάθε επανάληψη του βρόγχου.

Το περιεχόμενο κάθε σελίδας ανακτάται μέσω της συνάρτησης `getURLContent()`.

Κάθε σελίδα έχει την ίδια δομή: είκοσι άρθρα, κάθε από τα οποία έχει την εικόνα, τον τίτλο, την ημερομηνία δημοσίευματος, ένα κομμάτι του κειμένου και το κουμπί «Περισσότερα» (Εικόνα 5.3.9), και η σελιδοαρίθμηση. Η εικόνα, ο τίτλος, ένα κομμάτι του κειμένου του άρθρου και το κουμπί μπορούν να πατηθούν, συνεπώς αυτά τα στοιχεία περιέχουν το σύνδεσμο που οδηγεί στην ιστοσελίδα με το άρθρο (Εικόνα 5.3.10).

Εικόνα 5.3.9 Σύνδεσμοι των άρθρων στον EllhnikaHoaxes.gr



Ο FDA ΔΕΝ παραδέχτηκε ότι οι αντικαρκινικές Θεραπείες στην πραγματικότητα προκαλούν καρκίνο
📅 10 Ιανουαρίου 2024
Στα μέσα Δεκεμβρίου ένας ισχυρισμός διαμοιράστηκε ευρέως στα ελληνικά κοινωνικά δίκτυα και σε ιστοσελίδες, με τον ισχυρισμό πως ο Οργανισμός [...]

Αξιωματούχος του Π.Ο.Υ. ΔΕΝ συνελήφθη από πεζοναύτες των ΗΠΑ
📅 9 Ιανουαρίου 2024
Στα τέλη Δεκεμβρίου 2023 άρχισε να διαμοιράζεται σε μέσα κοινωνικής δικτύωσης ο ισχυρισμός ότι πεζοναύτες των ΗΠΑ συνέλαβαν τον αξιωματούχο [...]

ΠΕΡΙΣΣΟΤΕΡΑ **ΠΕΡΙΣΣΟΤΕΡΑ**

Εικόνα 5.3.10 Ανάκτηση συνδέσμου

```
String fakeNewsContent = getURLContent(pageAddress);
int i1 = fakeNewsContent.indexOf(
    "<div class=\"col-lg-9 col-md-9 col-sm-12 col-sm-12\"");
int i2 = fakeNewsContent.indexOf(
    "<div class=\"pagination-list text-center\"");
String newsPart = fakeNewsContent.substring(i1, i2);
int divStartPos = newsPart.indexOf(
    "<div class=\"blog-post-content\"");

while (divStartPos != -1) {
    int divEndPos = newsPart.indexOf("</div", divStartPos);
    String divContent = newsPart.substring(divStartPos, divEndPos);
    System.out.println("divContent: " + divContent);
    int aStartPos = divContent.indexOf("<a ");
    int aEndPost = divContent.indexOf(
        "\" data-wpel-link=\"internal\">", aStartPos);
    String hrefValue = divContent.substring(
        aStartPos + 9, aEndPost);
```

Κάτι που εντοπίστηκε κατά τον έλεγχο της λειτουργίας του κώδικα είναι τα διπλά δημοσιεύματα, δηλαδή τα άρθρα που δημοσιεύτηκαν κατά λάθος δύο φορές. Αγνόηση αυτού του θέματος οδηγεί στον πρόωρο τερματισμό του βρόγχου. Το πρόβλημα λύθηκε με τη δημιουργία δυο διαφορετικών HashSets: το oldRetrievedLinks (οι σύνδεσμοι που ανιχνεύτηκαν

κατά τις προηγούμενες εκτελέσεις του κώδικα και είναι ήδη αποθηκευμένοι στη ΒΔ) και το `newRetrievedLinks` (οι σύνδεσμοι που ανιχνεύτηκαν κατά την τρέχουσα εκτέλεση του κώδικα) (Εικόνα 5.3.11).

Εικόνα 5.3.11 Διαχείριση του προβλήματος διπλών εγγραφών

```
if (oldRetrievedLinks.contains(hrefValue)) {
    System.out.println(
        " - The link '" + hrefValue
            + "' already exists in the list! No need to check the remaining...");
    canStopReadingPages = true;
    break;
} else {

    if (newRetrievedLinks.contains(hrefValue) == false) {
        System.out.println(" - Writing the link: " + hrefValue);
        textWriter.append(hrefValue).append("\n");
        newRetrievedLinks.add(hrefValue);
        n++;
    }
}
```

Η συνάρτηση θα σταματήσει σε δύο περιπτώσεις: είτε ο αριθμός της σελίδας που πρέπει να διαβαστεί (`pageNum`) είναι μεγαλύτερος του αριθμού των διαθέσιμων σελίδων στον ιστότοπο, είτε ο σύνδεσμος που ανιχνεύτηκε τώρα ήδη υπάρχει στο αρχείο `fakeLinks.csv`.

Επίσης, μέσω στο βρόγχο χρησιμοποιείται η μέθοδος `sleep()` της κλάσης `Thread` για να ελαχιστοποιηθεί η πιθανότητα μπλοκαρίσματος του IP λόγω συνεχόμενων HTTP-αιτημάτων.

Η λεπτομερής ανάλυση του HTML κώδικα παρουσιάζεται στις Εικόνες 5.3.12 – 5.3.17.

Εικόνα 5.3.12 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (1)

```
<!DOCTYPE html>
<html lang="el" data-lt-installed="true">
  <head> </head>
  <body class="archive category category-kathories category-2724 wp-embed-responsive multiple-domain-www-ellinikahoaxes-gr">
    <div class="header-menu d-xl-block d-lg-block d-none fixed-header"> </div>
    <div class="header-menu d-md-block d-lg-none d-xl-none fixed-header"> </div>
    <div class="header-menu d-lg-none d-md-block d-xl-none d-none fixed-header"> </div>
    <div class="mobile-side-menu"> </div>
    <div class="mobile-menu-overlay2"></div>
    <div id="header-widget-area" class="chw-widget-area widget-area" role="complementary"> </div>
    <div class="white-row" style="height: auto !important;">
      <div class="container" style="height: auto !important;">
        <div class="row"> </div> <flex>
        <div class="row" style="height: auto !important;"> <flex>
          <div class="col-lg-9 col-md-9 col-sm-12 col-sm-12">
            <div class="row"> </div> <flex == $0
            <div class="pagination-list text-center" style="> </div>
          </div>
          <div class="col-md-3 col-sm-12 col-sm-12 sidebar" style="height: auto !important;"> </div>
        </div>
      </div>
    </div>
    <section id="footer" style="> </section>
    <section id="copyright"> </section>
    <script id="google_cse_v2-js-extra"> </script>
    <script src="https://www.ellinikahoaxes.gr/wp-content/plugins/wp-google-search/assets/js/google_cse_v2.js?ver=1" id="google_c"> </script>
    <script id="collapseomatic-js-js-before"> </script>
    <script src="https://www.ellinikahoaxes.gr/wp-content/plugins/jquery-collapse-o-matic/js/collapse.js?ver=1.7.2" id="collapse"> </script>
    <script id="rocket-browser-checker-js-after"> </script>
    <script id="rocket-preload-links-js-extra"> </script>
    <script id="rocket-preload-links-js-after"> </script>
  </body>
</html>
```

Εικόνα 5.3.13 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (2)

```
<div class="mobile-menu-overlay2"></div>
<div id="header-widget-area" class="chw-widget-area widget-area" role="complementary"> </div>
<div class="white-row" style="height: auto !important;">
  <div class="container" style="height: auto !important;">
    <div class="row"> </div> <flex>
    <div class="row" style="height: auto !important;"> <flex>
      <div class="col-lg-9 col-md-9 col-sm-12 col-sm-12">
        <div class="row"> <flex>
          <div class="col-md-6 col-12 front-news archive-post post-229233 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ag-erythra-thalassa-2 tag-2522 tag-2533 tag-flegomeno tag-1220 tag-chouthi" id="post-229233"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228757 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ex-offenders" id="post-228757"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228868 post type-post status-publish format-standard has-post-thumbnail hentry ca
            kos" id="post-228868"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228535 post type-post status-publish format-standard has-post-thumbnail hentry ca
            -hasan tag-prodosia" id="post-228535"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-229266 post type-post status-publish format-standard has-post-thumbnail hentry ca
            g-4679 tag-apothikes tag-niou-tzersei tag-new-jersey tag-elizabeth" id="post-229266"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-229261 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ag-lionel-mesi tag-ai tag-tzefri-epstain-2" id="post-229261"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228786 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ir-zelenski-2 tag-mohammed-al-alawi" id="post-228786"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-227343 post type-post status-publish format-standard has-post-thumbnail hentry ca
            tag-emvolia-mrna tag-tzozeif-lantapo tag-florida tag-joseph-ladapo tag-genikos-cheirourgos tag-thrafsmata-dna tag-ron-desantis tag-joe-lada
          <div class="col-md-6 col-12 front-news archive-post post-228588 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ia-kata-tou-koronoiou tag-real-raw-news" id="post-228588"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228067 post type-post status-publish format-standard has-post-thumbnail hentry ca
            4311 tag-villa-potenza tag-marche" id="post-228067"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-228538 post type-post status-publish format-standard has-post-thumbnail hentry ca
            ag-612 tag-airbus tag-1449 tag-techniti-noimosyni tag-ypogeies-siranges tag-tropopoiimeni-fotografia tag-syntonistis tag-eboria-paidion ta
            228538"> </div>
          <div class="col-md-6 col-12 front-news archive-post post-227420 post type-post status-publish format-standard has-post-thumbnail hentry ca
            news.archive-post.post-227863.post.type-post.status-publish.format-standard.has-post-thumbnail.hentry.category-paraplhroforhsh.tag-70.tag-fda.tag-t-kyttara.tag-the-peoples-voice.tag
```

Εικόνα 5.3.14 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (3)

```
...
tag-ehs-iei-emf tag-emf tag-annals-of-clinical-and-medical-case-reports tag-beall tag-bealls-list tag-predatory-journal tag-doaj tag-foren
tag-chd tag-the-defender tag-ionizousa tag-mi-ionizousa" id="post-226408"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-227257 post type-post status-publish format-standard has-post-thumbnail hentry c
02" id="post-227257"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228536 post type-post status-publish format-standard has-post-thumbnail hentry c
d-austin tag-ypoungos-amynas tag-pentagono" id="post-228536"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228218 post type-post status-publish format-standard has-post-thumbnail hentry c
astirio tag-roberto-sperantza" id="post-228218"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228229 post type-post status-publish format-standard has-post-thumbnail hentry c
shnikov tag-lahore" id="post-228229"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228248 post type-post status-publish format-standard has-post-thumbnail hentry c
tag-33chronos tag-viastis" id="post-228248"></div>
▼ <div class="col-md-6 col-12 front-news archive-post post-227863 post type-post status-publish format-standard has-post-thumbnail hentry c
eoples-voice tag-car-t tag-organismos-trofimon-kai-farmakon" id="post-227863">
  ▶ <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-prokalou
  ▶ <div class="blog-category"></div>
  ▼ <div class="blog-post-content">
    ▼ <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-prokalou
      ▶ <h2></h2>
      </a>
      ▶ <div class="blog-excerpt"></div>
    </div>
  </div>
▶ <div class="col-md-6 col-12 front-news archive-post post-227362 post type-post status-publish format-standard has-post-thumbnail hentry c
4376 tag-axiomatouchos tag-pezonaftes-2 tag-chikwe-ihekweazu" id="post-227362"></div>
</div>
▶ <div class="pagination-list text-center" style></div>
</div>
▶ <div class="col-md-3 col-sm-12 col-sm-12 sidebar" style="height: auto !important;"></div>
</div>
</div>
```

Εικόνα 5.3.15 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (4)

```
...
d-austin tag-ypoungos-amynas tag-pentagono" id="post-228536"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228218 post type-post status-publish format-standard has-post-thumbnail hentry c
astirio tag-roberto-sperantza" id="post-228218"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228229 post type-post status-publish format-standard has-post-thumbnail hentry c
shnikov tag-lahore" id="post-228229"></div>
▶ <div class="col-md-6 col-12 front-news archive-post post-228248 post type-post status-publish format-standard has-post-thumbnail hentry c
tag-33chronos tag-viastis" id="post-228248"></div>
▼ <div class="col-md-6 col-12 front-news archive-post post-227863 post type-post status-publish format-standard has-post-thumbnail hentry c
eoples-voice tag-car-t tag-organismos-trofimon-kai-farmakon" id="post-227863">
  ▶ <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-pr
  ▶ <div class="blog-category"></div>
  ▼ <div class="blog-post-content">
    ▼ <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-
      ▶ <h2>
        "Ο FDA DEN παραδέχτηκε ότι οι αντικαρκινικές θεραπείες στην πραγματικότητα προκαλούν καρκίνο"
      </h2>
      </a>
      ▼ <div class="blog-excerpt">
        ▼ <div class="blog-date" style>
          :before
          "10 Ιανουαρίου 2024"
        </div>
        ▶ <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotit
        <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechthike-oti-oi-antikarkinikes-therapeies-stin-pragmatikotit
          nal" style>Περισσότερα</a>
        </div>
      </div>
    </div>
  </div>
▶ <div class="col-md-6 col-12 front-news archive-post post-227362 post type-post status-publish format-standard has-post-thumbnail hentry c
4376 tag-axiomatouchos tag-pezonaftes-2 tag-chikwe-ihekweazu" id="post-227362"></div>
</div>
▶ <div class="pagination-list text-center" style></div>
```

Εικόνα 5.3.16 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (5)

```
<div class="blog-post-content">
  <a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechti-ke-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-prokaloun-karkino/" data-wpel-link="internal"> == $0
  <h2>
    "Ο FDA ΔΕΝ παραδέχτηκε ότι οι αντικαρκινικές θεραπείες στην πραγματικότητα προκαλούν καρκίνο"
  </h2>
</a>
<div class="blog-excerpt">
  <div class="blog-date" style=
    ::before
    "10 Ιανουαρίου 2024"
  </div>
```

Εικόνα 5.3.17 Σύνδεσμος ενός άρθρου στο HTML-κώδικα (6)

```
<a href="https://www.ellinikahoaxes.gr/2024/01/10/o-fda-den-paradechti-ke-oti-oi-antikarkinikes-therapeies-stin-pragmatikotita-prokaloun-karkino/" data-wpel-link="internal"> == $0
```

Εικόνα 5.3.18 Μέθοδος addLinks() της κλάσης LinkRetrieval (1)

```
133      /* Retrieves links from the website
134      1. If a retrieved link has already been retrieved and saved in the file
135      with retrieved links during of the previous executions, the loop stops and
136      the execution is over.
137      2. If a retrieved link has not already been retrieved and saved in the file
138      with retrieved links during one of the previous executions, it is added
139      to the file with retrieved links and the executions continues.
140      3. OIf a retrieved link has already been retrieved during this current
141      execution but it is not in the file with retrieved links (duplicate
142      publications), this link is just ignored and the execution continues.
143      */
144      public static int addLinks(File f, int maxP) throws
145      IOException, InterruptedException {
146          int n;
147
148          try (Writer textWriter = new BufferedWriter(
149              new OutputStreamWriter(
150                  new FileOutputStream(f, true),
151                  StandardCharsets.UTF_8)
152          )) {
153
154              boolean canStopReadingPages = false;
155              n = 0;
156              for (int pageNum = 1; pageNum < maxP; pageNum++) {
157                  String pageAddress =
158                      "https://www.ellinikahoaxes.gr/category/kathgories/";
159                  String pageNumStr = String.valueOf(pageNum);
160
161                  if (pageNum >= 2) {
162                      pageAddress = pageAddress + "page/" + pageNumStr + "/";
163                  }
164
165                  System.out.println(
166                      "\n# Reading the contents of the page: " + pageAddress);
167                  String fakeNewsContent = getURLContent(pageAddress);
168                  int i1 = fakeNewsContent.indexOf(
169                      "<div class=\"col-lg-9 col-md-9 col-sm-12 col-sm-12\"");
170                  int i2 = fakeNewsContent.indexOf(
171                      "<div class=\"pagination-list text-center\"");
172                  String newsPart = fakeNewsContent.substring(i1, i2);
173                  int divStartPos = newsPart.indexOf(
174                      "<div class=\"blog-post-content\"");
175
176                  while (divStartPos != -1) {
177                      int divEndPos = newsPart.indexOf("</div", divStartPos);
178                      String divContent = newsPart.substring(divStartPos, divEndPos);
179                      System.out.println("divContent: " + divContent);
180                      int aStartPos = divContent.indexOf("<a ");
```

Εικόνα 5.3.19 Μέθοδος addLinks() της κλάσης LinkRetrieval (2)

```
181         int aEndPost = divContent.indexOf(
182             "\"" data-wpel-link=\"internal\">", aStartPos);
183         String hrefValue = divContent.substring(
184             aStartPos + 9, aEndPost);
185
186         if (oldRetrievedLinks.contains(hrefValue)) {
187             System.out.println(
188                 " - The link '" + hrefValue
189                 + "' already exists in the list! No need to check the remaining...");
190             canStopReadingPages = true;
191             break;
192         } else {
193
194             if (newRetrievedLinks.contains(hrefValue) == false) {
195                 System.out.println(" - Writing the link: " + hrefValue);
196                 textWriter.append(hrefValue).append("\n");
197                 newRetrievedLinks.add(hrefValue);
198                 n++;
199             }
200         }
201
202         divStartPos = newsPart.indexOf(
203             "<div class=\"blog-post-content\"", divEndPos);
204     }
205
206     if (canStopReadingPages == true) {
207         break;
208     }
209
210     Thread.sleep(2000);
211 }
212 textWriter.flush();
213 }
214 return n;
215 }
```

5.3.2 ΑΝΑΚΤΗΣΗ ΤΩΝ ΑΠΑΡΑΙΤΗΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΩΝ ΣΥΝΔΕΣΜΩΝ

Για την ανάκτηση και την αποθήκευση των δεδομένων, όπως ο τίτλος, η ημερομηνία, το όνομα του συντάκτη-ελεγκτή, τον ισχυρισμό, το συμπέρασμα, οι κατηγορίες των άρθρων δημιουργήθηκαν δυο κλάσεις: PageDataRetrival και Link.

Στην πρώτη κλάση

1. Δημιουργούνται τα απαραίτητα HashSets και HashMaps που βοηθάνε να «μεταφράζουν» τις κατηγορίες των άρθρων σε μορφή συνδέσμων και εικόνων (Πίνακας 4.5.1) σε κατανοητές συμβολοσειρές (Εικόνες 5.3.20 και 5.3.21).

Εικόνα 5.3.20 HashSet και HashMap για τη μετάφραση των συνδέσμων των εικόνων στο HTML-κώδικα σε μια συμβολοσειρά, εδώ: «ΚΙΝΔΥΝΟΛΟΓΙΑ»

```
String kindynolog_images[] = {
    "https://media.ellinikahoaxes.gr/uploads/2020/08/kindinologia-full.png",
    "https://media.ellinikahoaxes.gr/uploads/2019/06/Fear-Mongering-300x108.png",
    "https://media.ellinikahoaxes.gr/uploads/2019/06/Fear-Mongering-1024x369.png",
    "https://media.ellinikahoaxes.gr/uploads/2019/05/Fear-Mongering-1-300x108.png",
    "https://media.ellinikahoaxes.gr/uploads/2019/05/Fear-Mongering-1024x369.png",
    "https://media.ellinikahoaxes.gr/uploads/2019/04/Fear-Mongering-1-300x108.png"
};
kindynolog_imageLinks.addAll(Arrays.asList(kindynolog_images));
imageCategory = MinorClass.addDictValues(
    kindynolog_imageLinks,
    imageCategory,
    "ΚΙΝΔΥΝΟΛΟΓΙΑ");
```

Εικόνα 5.3.21 HashMap για τη μετάφραση των συνδέσμων των κατηγοριών στο HTML-κώδικα σε μια συμβολοσειρά, εδώ: «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ», «ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ» και «ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ»

```
linkCategory.put (
    "https://www.ellinikahoaxes.gr/category/kathgories/paraplhroforhsh/",
    "ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ");
linkCategory.put (
    "https://www.ellinikahoaxes.gr/category/kathgories/leipei-thematiko-periechomeno/",
    "ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ");
linkCategory.put (
    "https://www.ellinikahoaxes.gr/category/kathgories/tropopoiimeni-eikona/",
    "ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ");
```

2. Δημιουργούνται τα απαραίτητα HashSets και HashMaps που βοηθάνε να αλλάζουν τις ημερομηνίες στα ελληνικά σε αριθμητική μορφή, πχ «3 Νοεμβρίου 2017» σε «3-11-2017». Αυτή η αλλαγή είναι απαραίτητη για την περαιτέρω επεξεργασία και την ανάλυση των δεδομένων.

Εικόνα 5.3.22 HashMap για την αλλαγή μορφής των ημερομηνιών

```
monthConvert.put ("Ιανουαρίου", "1");
monthConvert.put ("Φεβρουαρίου", "2");
monthConvert.put ("Μαρτίου", "3");
monthConvert.put ("Απριλίου", "4");
monthConvert.put ("Μαΐου", "5");
monthConvert.put ("Ιουνίου", "6");
monthConvert.put ("Ιουλίου", "7");
monthConvert.put ("Αυγούστου", "8");
monthConvert.put ("Σεπτεμβρίου", "9");
monthConvert.put ("Οκτωβρίου", "10");
monthConvert.put ("Νοεμβρίου", "11");
monthConvert.put ("Δεκεμβρίου", "12");
```

3. Διαβάζεται το αρχείο fakeLinksData.csv που περιέχει τα δεδομένα που συλλέχθηκαν κατά τις προηγούμενες εκτελέσεις του κώδικα. Αν αυτό το αρχείο δεν υπάρχει, τότε δημιουργείται.

Εικόνα 5.3.23 Ανάκτηση ή δημιουργία του αρχείου fakeLinksData.csv

```
System.setOut(new PrintStream(System.out, true, "UTF8"));

//Reads a file with already retrieved links data
File readFile = new File(allData);

/* If the file exists, all the links in the file are added to the
hashset "accessedLinks"
*/
try (BufferedReader in = new BufferedReader(
    new InputStreamReader(
        new FileInputStream(readFile), "UTF8")
    )) {
    String str;

    while ((str = in.readLine()) != null) {

        if (!str.trim().equals("")) {
            String[] strSplit;
            strSplit = str.split("\t");
            accessedLinks.add(strSplit[0].replace("\\", ""));
        }
    }
}

//If the file does not exist, it is created and a header is added
} catch (Exception e) {
    System.out.println(e.getMessage());
    File linkFile = new File(allData);
    try (Writer textWriter = new BufferedWriter(
        new OutputStreamWriter(
            new FileOutputStream(linkFile, true),
            StandardCharsets.UTF_8)
    )) {
        System.setOut(new PrintStream(System.out, true, "UTF8"));
        String fieldName =
            "Link\tTitle\tDate\tAuthor\tClaim\tConclusion\tCategory\tVerdict\n";
        textWriter.append(fieldName);
        textWriter.flush();
    }
}
```

4. Εκτελείται ο βρόγχος για ανάκτηση δεδομένων κάθε συνδέσμου στο αρχείο fakeLinks.csv με καθυστέρηση 1000 χιλιοστών του δευτερολέπτου (1000 ms) για την αποφυγή μπλοκαρίσματος του IP. Ωστόσο, υπάρχουν μερικές οδηγίες εκτέλεσης που πρέπει να ακολουθούνται για την ομαλή εκτέλεση του κώδικα, που περιγράφονται στην Ενότητα 5.3.2.6. Η εξαγωγή δεδομένων του κάθε άρθρου γίνεται με το κάλεσμα της συνάρτησης processNewLink() της κλάσης Link μέσα στο βρόγχο.

Επειδή στον EllhnikaHoaxes.gr υπάρχουν άρθρα που δεν έχουν να κάνουν με ψευδείς ειδήσεις, πχ https://www.ellinikahoaxes.gr/2016/04/22/dawkins_richard-david_hume/, κάποια άρθρα αγνοούνται. Για να ανακτώνται τα δεδομένα των άρθρων που περιέχουν την ανάλυση μιας ψευδούς είδησης (ή έστω να μεγιστοποιηθεί αυτή η πιθανότητα) δημιουργήθηκε ένα HashSet validCatVerdictValues που αποτελείται από όλες τις κατηγορίες που συνήθως περιλαμβάνουν κάποιες ψευδείς ειδήσεις. Οι κατηγορίες αυτές είναι οι δικές μου παρατηρήσεις μετά τη χειρωνακτική ανάλυση πολλών άρθρων. Δυστυχώς, αυτή η προσέγγιση δεν παράγει αποτελέσματα με απόλυτη εγκυρότητα, διότι οι κατηγορίες ανατίθενται από τους συντάκτες και όχι αυτόματα, άρα υπάρχουν ανακρίβειες, αλλά στο 95% των περιπτώσεων τα αποτελέσματα είναι έγκυρα. Συνήθως το άρθρο έχει περισσότερες από μια κατηγορίες αλλά υπάρχουν οι περιπτώσεις όταν η κατηγορία είναι μόνο μία. Αν καμία κατηγορία του άρθρου δε βρίσκεται στο HashSet validCatVerdictValues με μεγάλη πιθανότητα αυτό το άρθρο δεν περιέχει ψευδείς ειδήσεις άρα δεν αποθηκεύεται στο αρχείο fakeLinksData.csv. Αν το άρθρο δεν ανήκει σε καμία κατηγορία, αυτό το άρθρο επίσης αποθηκεύεται. Ο έλεγχος αυτός πραγματοποιείται στη συνάρτηση processNewLink της κλάσης Link, η οποία επιστρέφει true, αν το άρθρο έχει σχέση με την ανάλυση ενός ψευδούς υλικού, και false, αν το αυτό το άρθρο δεν έχει να κάνει με κάποια ψευδή είδηση.

Εικόνα 5.3.24 HashSet validCatVerdictValues

```

/*Adds values of all existing categories and verdicts (text values) of
fake news to the hashset "validCatVerdictValues"
*/
String validCatVerdictValuesStr [] = {"ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ",
    "ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ", "ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ", "ΨΕΥΔΕΣ",
    "ΑΝΑΚΥΚΛΩΜΕΝΗ ΕΙΔΗΣΗ", "ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ", "ΤΡΟΠΟΠΟΙΗΜΕΝΟ ΒΙΝΤΕΟ",
    "ΜΙΣΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ", "ΨΕΥΔΟΕΠΙΣΤΗΜΗ", "ΑΠΑΘΗ",
    "LIKE FARMING", "ΚΙΝΔΥΝΟΛΟΓΙΑ", "ΑΝΕΠΙΒΕΒΑΙΩΤΟ", "CLICKBAIT", "ΣΑΤΙΡΑ",
    "ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ", "HOAXES", "ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ",
    "FAKE NEWS", "ΥΓΕΙΑ", "ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ", "ΔΙΑΦΟΡΑ"};
validCatVerdictValues.addAll(Arrays.asList(validCatVerdictValuesStr));

```

Εικόνα 5.3.25 Μέθοδος processNewLink() της κλάσης Link (1)

```

public static boolean processNewLink(String str) throws IOException {

```

Για να πάρει η μεταβλητή noRecord την τιμή false πρέπει να τηρείται μια από τις δυο περιπτώσεις:

- Έστω μια από τις κατηγορίες του άρθρου είναι στο HashSet validCatVerdictValues. Σε αυτή την περίπτωση στον βρόγχο ελέγχεται αν έστω μια κατηγορία (ένα στοιχείο του HashSet catVerdicts) είναι στο HashSet validCatVerdictValues. Αν αυτό το στοιχείο υπάρχει, η μεταβλητή noRecord παίρνει την τιμή false και ο βρόγχος τερματίζεται
- Το άρθρο δεν έχει καθόλου κατηγορίες. Αν και αυτό φαίνεται παράλογο τα άρθρα χωρίς κατηγορία καμιά φορά περιέχουν την ανάλυση κάποιου ύποπτου ισχυρισμού ή είδησης, πχ <https://www.ellinikahoaxes.gr/2020/08/15/hagia-sophia-fake-prophecy/>, <https://www.ellinikahoaxes.gr/2020/09/14/ochi-afto-to-vinteo-den-einai-apo-tourkia/> ή <https://www.ellinikahoaxes.gr/2020/10/15/ochi-afti-i-eikona-den-deichnei-tourkiko-f-16-pou-katarrifthike-apo-armenikous-s-300/>, άρα είναι χρήσιμο να αποθηκεύουμε αυτά τα δεδομένα. Σε αυτή την περίπτωση ελέγχεται αν το HashSet catVerdicts είναι κενό. Αν είναι κενό, τότε η μεταβλητή noRecord παίρνει την τιμή false.

Μετά από αυτό, αν η μεταβλητή noRecord έχει την τιμή false, γίνεται η περαιτέρω εξαγωγή των στοιχείων, όπως ο τίτλος, η ημερομηνία κ.α. (Εικόνα 5.3.26).

Εικόνα 5.3.26 Μέθοδος processNewLink() της κλάσης Link (2)

```
Boolean noRecord = true;

if (catVerdicts.isEmpty() == false){

    for (String cV : catVerdicts) {

        if (validCatVerdictValues.contains(cV)) {
            noRecord = false;
            break;
        }
    }
}

if (catVerdicts.isEmpty()){
    noRecord = false;
}

if (noRecord == false) {

    //Retrieves the title
    int titleStart = openedLinkContent.indexOf("<h1");
    int titleEnd = openedLinkContent.indexOf("</h1");
    String openedLinkTitle = openedLinkContent.substring(
        titleStart+4, titleEnd);

    if (openedLinkTitle.contains("<strong>")){
        openedLinkTitle = openedLinkTitle.replace(
            "</strong>", "").replace("<strong>", "");
    }

    //Retrieves the date
    int dateStart = openedLinkContent.indexOf(
        "<div class=\"blog-date\"");
    int dateEnd = openedLinkContent.indexOf("</div>", dateStart);
    String openedLinkDate = openedLinkContent.substring(
        dateStart+23, dateEnd);
}
```

Στο τέλος της συνθήκης if (noRecord == false) η συνάρτηση επιστρέφει true. Αν αυτή η συνθήκη δεν εκτελείται, η συνάρτηση επιστρέφει false και τα δεδομένα του συνδέσμου δεν αποθηκεύονται (Εικόνα 5.3.27).

Εικόνα 5.3.27 Μέθοδος processNewLink() της κλάσης Link (3)

```
if (noRecord == false) {
    new OutputStreamWriter(
        new FileOutputStream(
            linkFile, true), StandardCharsets.UTF_8)
    )) {
        System.setOut(new PrintStream(System.out, true, "UTF8"));
        addNewLine (numOfRecords, claims, conclusions,
            openedLinkTitle, dateNumber, openedLinkAuthor,
            categoryString, imageValueString, textWriter);
        textWriter.flush();
    }
    return true;
} else {
    return false;
}
}
```

Εικόνα 5.3.28 Βρόγχος, στον οποίο καλείται η μέθοδος processNewLink() της κλάσης Link

```
/* Runs the hashset of the links that were retrieved from the file
"FakeData.csv (function getOldLinks())
*/
for (String l : links) {
    System.out.println("\n-----\n");

    //If the link had been already accessed we ignore it
    if (accessedLinks.contains(l)) {
        System.out.println(
            " - The link '" + l + "' already exists in the list!");
    } else {
        Link link = new Link (l);

        if (Link.processNewLink(l)) {
            System.out.println("Link: " + link.getLink());
            System.out.println("Title: " + link.getTitle());
            System.out.println("Date: " + link.getDate());
            System.out.println("Author: " + link.getAuthor());
            System.out.println("Claims: ");

            for (String cl : link.getClaimsAll()) {
                System.out.println(cl);
            }

            System.out.println("Conclusions: ");

            for (String con : link.getConclusionsAll()) {
                System.out.println(con);
            }

            System.out.println("Categories: " + link.getCategory());
            System.out.println("Verdicts: " + link.getVerdict());
        } else {
            System.out.println("Link: " + l);
            System.out.println("does not contain any fakenews");
        }

        Thread.sleep(1000);
    }
}
```

5. Επίσης, εκτελούνται οι συναρτήσεις `getOldLink()` και `getURLContent()` που δε διαφέρουν από τις συναρτήσεις υπό τα ίδια ονόματα στην κλάση `LinkRetrieval` και αναλύονται στις Ενότητες 5.3.1.1 και 5.3.1.3, αντίστοιχα.

Στη δεύτερη κλάση `Link` γίνεται η επεξεργασία του συνδέσμου. Η κλάση αυτή καλείται μέσα στον βρόγχο της κλάσης `PageDataRetrieval` για κάθε σύνδεσμο που ήδη υπάρχει στο αρχείο `fakeLinks.csv` αλλά ακόμα δεν υπάρχει στο αρχείο `fakeLinksData.csv`.

Η υπηρεσία `EllinikaHosxes.gr` ιδρύθηκε το 2013 και η δομή των σελίδων το 2013 διαφέρουν σημαντικά από τη δομή των σελίδων το 2023. Για αυτό το λόγο, για να ανακτηθούν κάποια δεδομένα εφαρμόζονται διάφοροι τρόποι. Με αυτή την προσέγγιση επιτυγχάνεται η ανάκτηση περισσότερων δεδομένων που παίζουν μεγάλο ρόλο στα αποτελέσματα της περαιτέρω ανάλυσης.

5.3.2.1 ΑΝΑΚΤΗΣΗ ΤΩΝ ΚΑΤΗΓΟΡΙΩΝ ΕΝΟΣ ΑΡΘΡΟΥ

Στα άρθρα οι κατηγορίες αναφέρονται με διάφορους τρόπους, όπως

- Μόνο πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου), πχ εδώ <https://www.ellinikahoaxes.gr/2016/05/14/israeli-jews-dont-have-cancer-hoax/> ή εδώ <https://www.ellinikahoaxes.gr/2017/06/08/gardasil/>.
- Μόνο σε μορφή εικόνας (Δείτε τον Πίνακα 4.5.1.), πχ εδώ <https://www.ellinikahoaxes.gr/2020/08/06/leoforio-mask/> ή εδώ <https://www.ellinikahoaxes.gr/2020/08/22/bitcoin-scam-returns/>.
- Και πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου) και σε μορφή εικόνας, αλλά οι κατηγορίες αυτές δεν συμφωνούν, πχ εδώ <https://www.ellinikahoaxes.gr/2021/03/14/hoax-jonathan-galindo/> ή εδώ <https://www.ellinikahoaxes.gr/2020/01/11/jews-cancer-conspiracy-theory-debunked/>.
- Και πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου) και σε μορφή εικόνας, και οι κατηγορίες αυτές συμφωνούν, πχ εδώ <https://www.ellinikahoaxes.gr/2021/09/09/denmark-drops-mass-vaccination-and-covid-19-restrictions-missing-context/> ή εδώ <https://www.ellinikahoaxes.gr/2020/08/11/psevdis-titlos-anastellontai-oles-oi-thriskeftikes-teletes/>.
- Δεν αναφέρονται καθόλου, πχ εδώ <https://www.ellinikahoaxes.gr/2016/08/10/ψεκάζουν-με-coca-cola-τα-σπαρτά/>.

Στην αρχή, ο κώδικας προσπαθεί να εξάγει τις κατηγορίες από μέρος πάνω από τον τίτλο, δηλαδή τις κατηγορίες σε μορφή κειμένου (Εικόνα 5.3.29). Στον Java-κώδικα αυτές οι κατηγορίες λέγονται «categories».

Εικόνα 5.3.29 Κατηγορία «HOAXES» πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου)



Το κομμάτι του HTML-κώδικα, όπου βρίσκεται αυτό το στοιχείο παρουσιάζεται στις Εικόνες 5.3.30 – 5.3.32, ο Java-κώδικας που χρησιμοποιήθηκε παρουσιάζεται στην Εικόνα 5.3.33.

Εικόνα 5.3.30 Κατηγορία του άρθρου, που βρίσκεται πάνω από τον τίτλο (σε μορφή κειμένου), στο HTML-κώδικα (1)

```

<div class="container" style="height: auto !important;">
  <div data-bg="https://media.ellinikahoaxes.gr/uploads/2018/01/20150311_cancerresearch1.jpg" class="single-header single-img rocket-lazyload entered lazyload edia.ellinikahoaxes.gr/uploads/2018/01/20150311_cancerresearch1.jpg&quot;);" data-ll-status="loaded">
    ::before
    <div class="content-bottom">
      <div class="row"> (flex)
        <div class="col-sm-12">
          <span class="inner-text-img">
            
            <noscript></noscript>
            <strong>Ευαίσθητο περιεχόμενο</strong>
            <p></p>
          </span>
          <span class="header-category">
            <span>
              <span class="breadcrumb_last" aria-current="page">Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</span> == $0
            </span>
            <a href="https://www.ellinikahoaxes.gr/category/kathgories/ygeia/" rel="category tag" data-wpel-link="internal"></a>
            <a href="https://www.ellinikahoaxes.gr/category/kathgories/psevdoepisthmh/" rel="category tag" data-wpel-link="internal" style="</a>
          </span>
          <h1 style="<h1 style="Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1>
          <div class="blog-date" style="5 Ιανουαρίου 2018</div>
          <p class="header-author"></p>
        </div>
      </div>
    </div>
  <div class="row" style="height: auto !important;"> </div> (flex)
</div>
</div>
<section id="footer"></section>

```

Εικόνα 5.3.31 Κατηγορία του άρθρου, που βρίσκεται πάνω από τον τίτλο (σε μορφή κειμένου), στο HTML-κώδικα (2)

```

<span class="breadcrumb_last" aria-current="page">Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</span>
</span>
<a href="https://www.ellinikahoaxes.gr/category/kathgories/ygeia/" rel="category tag" data-wpel-link="internal"></a> == $0
<a href="https://www.ellinikahoaxes.gr/category/kathgories/psevdoepisthmh/" rel="category tag" data-wpel-link="internal" style="</a>
</span>
<h1 style="Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1>

```

Εικόνα 5.3.32 Κατηγορία του άρθρου , που βρίσκεται πάνω από τον τίτλο, (σε μορφή κειμένου) στο HTML-κώδικα (3)

```
<a href="https://www.ellinikahoaxes.gr/category/kathgories/ygeia/" rel="category tag" data-wpel-link="internal" style></a> == $0
```

Εικόνα 5.3.33 Εξαγωγή της κατηγορίας, που βρίσκεται πάνω από τον τίτλο του άρθρου (σε μορφή κειμένου)

```
//Extracts categories from H1 field
ArrayList<String> catVerdicts = new ArrayList<>();
int spanCategoryStart = openedLinkContent.indexOf(
    "<span class=\"breadcrumb_last\"");
int spanCategoryEnd = openedLinkContent.indexOf(
    "<h1", spanCategoryStart);
String spanCategory = openedLinkContent.substring(
    spanCategoryStart, spanCategoryEnd);
String[] categoryList= spanCategory.split("<a href=\"");
rawCategories.addAll(Arrays.asList(categoryList));
rawCategories.remove(0);
String categoryString = "";

//Normalizes links in the hashset
for (String rawCat : rawCategories) {
    int rawCatEnd = rawCat.indexOf("\" rel");
    String categoryKey = rawCat.substring(0, rawCatEnd);
    String categoryValue = linkCategory.get(categoryKey);

    /* If the category exists in linkCategory dictionary,
    adds the category to the "categories" hashset,
    expands the string "categoryString",
    and adds the category to the "catVerdicts" hashset
    */
    if (categoryValue != null){
        categories.add(categoryValue);
        categoryString = categoryString + "," + categoryValue;
        catVerdicts.add(categoryValue);
    }
}

if (!categories.isEmpty()) {
    categoryString = categories.toString();
    categoryString = categoryString.replace("[", "");
    categoryString = categoryString.replace("]", "");
}
```

Στο τέλος, ο κώδικας προσπαθεί να εξάγει κατηγορίες από τις εικόνες στην ιστοσελίδα (Εικόνα 5.3.34). Στον Java-κώδικα αυτές οι κατηγορίες λέγονται «verdicts».

Εικόνα 5.3.34 Κατηγορία «ΨΕΥΔΕΣ» σε μορφή εικόνας



Το κομμάτι του HTML-κώδικα, όπου βρίσκεται αυτό το στοιχείο παρουσιάζεται στις Εικόνες 5.3.35 – 5.3.37, ο Java-κώδικας που χρησιμοποιήθηκε παρουσιάζεται στις Εικόνες 5.3.38 – 5.3.41.

Εικόνα 5.3.35 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (1)

```
<!DOCTYPE html>
<html lang="el" data-ll-installed="true">
  <head>
  </head>
  <body class="post-template-default single single-post postid-66233 single-format-standard wp-embed-responsive multiple-domain-www-ellinikahoaxes-gr">
    <div class="header-menu d-xl-block d-lg-block d-none fixed-header">
    </div>
    <div class="header-menu d-block d-md-none d-lg-none d-xl-none fixed-header">
    </div>
    <div class="header-menu d-lg-none d-md-block d-xl-none d-none fixed-header">
    </div>
    <div class="mobile-side-menu">
    </div>
    <div class="mobile-menu-overlay2">
    </div>
    <div id="header-widget-area" class="chw-widget-area widget-area" role="complementary">
    </div>
    <div class="white-row" style="height: auto !important;">
    <div class="container" style="height: auto !important;">
      <div data-bg="https://media.ellinikahoaxes.gr/uploads/2020/07/31.png" class="single-header single-img rocket-lazyload entered lazyloaded" style="background-image: url(https://media.ellinikahoaxes.gr/uploads/2020/07/31.png);" data-ll-status="loaded">
      </div>
      <div class="row" style="height: auto !important;">
        <div class="col-md-3 col-12 order-md-1 order-3 sidebar-container" style="height: auto !important;">
        </div>
        <div class="col-md-9 col-12 order-md-2 order-2 main-content-body">
          <div class="alert" style="display: none;">
          </div>
          <div class="row" style="height: auto !important;">
            <div class="row" style="height: auto !important;">
              <div class="col-12 text-center stams" style="height: auto !important;">
                
                <noscript>
                </noscript>
              </div>
            </div>
          </div>
          <div class="single-article post-66233 post-type-post status-publish format-standard has-post-thumbnail hentry category-hoax tag-430 tag-2451">
          <div class="author row padding-author">
          </div>
          <script async src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js"></script>
          <!-- Custom ad 1 -->
          <ins class="adsbygoogle" style="display: block; height: 280px;" data-ad-client="ca-pub-1938692651700370" data-ad-slot="7769959131" data-ad-format="auto">
          </ins>
        </div>
      </div>
    </div>
  </div>
</body>
</html>
```

Εικόνα 5.3.36 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (2)

```
<div class="col-12 text-center stams" style="height: auto !important;">
  
  <noscript>
  </noscript>
</div>
```

Εικόνα 5.3.37 Κατηγορία του άρθρου σε μορφή εικόνας στο HTML-κώδικα (3)

```

```


Εικόνα 5.3.38 Εξαγωγή της κατηγορίας σε μορφή εικόνας (1)

```
185     int divImageStart;
186     int divImageEnd;
187     String divImage;
188     int figImageStart;
189     int figImageEnd;
190     String figImage;
191     int imageStart;
192     int imageEnd;
193     String openedLinkImage;
194     String openedLinkImageValue;
195     String imageValueString;
196
197     divImageStart = openedLinkContent.indexOf(
198         "<div class=\"col-12 text-center stams\"");
199     divImageEnd = openedLinkContent.indexOf("</div",divImageStart);
200     divImage = openedLinkContent.substring(divImageStart, divImageEnd);
201     imageStart = divImage.indexOf("data-lazy-src");
202
203     //Tries to extract values from the images on the webpage (if exist)
204     System.out.println("1st try");
205
206     while (imageStart != -1) {
207         imageEnd = divImage.indexOf("\" ",imageStart);
208         openedLinkImage = divImage.substring(imageStart+15, imageEnd);
209         openedLinkImageValue = imageCategory.get(openedLinkImage);
210
211         /* If this value exists in the "imageCategory" dictionary,
212         adds the value to the hashmap "catVerdicts"
213         */
214         if (openedLinkImageValue != null) { //if
215             catVerdicts.add(openedLinkImageValue);
216             rawImageValues.add(openedLinkImageValue);
217         }
218
219         imageStart = divImage.indexOf("data-lazy-src", imageEnd);
220
221     }
222
223     imageValueString = rawImageValues.toString();
224     imageValueString = imageValueString.replace("[", "");
225     imageValueString = imageValueString.replace("]", "");
226
```

Εικόνα 5.3.39 Εξαγωγή της κατηγορίας σε μορφή εικόνας (2)

```
227 //If the previous if-clause had not been executed
228 if (imageValueString.isEmpty()) {
229
230     try {
231         System.out.println("2nd try");
232         divImageStart = openedLinkContent.indexOf(
233             "<div class=\"wp-block-columns is-layout-flex wp-container-6\"");
234         divImageEnd = openedLinkContent.indexOf("<p", divImageStart);
235         divImage = openedLinkContent.substring(
236             divImageStart, divImageEnd);
237         figImageStart = divImage.indexOf(
238             "<figure class=\"wp-block-image size-large is-resized\"");
239
240         while (figImageStart != -1) {
241             figImageEnd = divImage.indexOf("</figure", figImageStart);
242             figImage = divImage.substring(figImageStart, figImageEnd);
243             imageStart = figImage.indexOf("data-lazy-src");
244             imageEnd = figImage.indexOf("<noscript", imageStart);
245             openedLinkImage = figImage.substring(
246                 imageStart+15, imageEnd-3).replace("\\"", "");
247             openedLinkImageValue = imageCategory.get(openedLinkImage);
248
249             if (openedLinkImageValue != null) {
250                 rawImageValues.add(openedLinkImageValue);
251                 catVerdicts.add(openedLinkImageValue);
252             }
253
254             figImageStart = divImage.indexOf(
255                 "<figure class=\"wp-block-image size-large is-resized\"",
256                 figImageEnd);
257
258         }
259
260         imageValueString = rawImageValues.toString();
261         imageValueString = imageValueString.replace("[", "");
262         imageValueString = imageValueString.replace("]", "");
263
264     } catch (Exception e1){
265     }
266 }
267
268 int mainImageDivStart;
269 int mainImageDivEnd;
270 String mainDivImage;
271
```

Εικόνα 5.3.40 Εξαγωγή της κατηγορίας σε μορφή εικόνας (3)

```
272 //If the previous if-clause had not been executed
273 if (imageValueString.isEmpty()) {
274
275     try {
276         System.out.println("3rd try");
277         mainImageDivStart = openedLinkContent.indexOf(
278             "<div class=\"single-article post\"");
279         mainImageDivEnd = mainImageDivStart + 5000;
280         mainDivImage = openedLinkContent.substring(
281             mainImageDivStart, mainImageDivEnd);
282         int divClassStart = mainDivImage.indexOf(
283             "<div class=\"wp-block-image\"");
284         System.out.println(divClassStart);
285         int figureClassStart= mainDivImage.indexOf(
286             "figure class=\"wp-block-image\"");
287         System.out.println(figureClassStart);
288
289         if
290         (
291             (mainDivImage.contains("figure class=\"wp-block-image\")&&
292             (mainDivImage.contains("<div class=\"wp-block-image\"") &&
293             (divClassStart < figureClassStart)
294             ){
295             divImageStart = mainDivImage.indexOf(
296                 "<div class=\"wp-block-image\"");
297             System.out.println("IF clause is TRUE");
298         } else {
299             divImageStart = mainDivImage.indexOf(
300                 "class=\"wp-block-image\"");
301             System.out.println("IF clause is FALSE");
302         }
303
304         while (divImageStart != -1 ) {
305             divImageEnd = mainDivImage.indexOf(
306                 "</figure\"", divImageStart);
307
308             if (divImageEnd == -1) {
309                 break;
310             }
311             divImage = mainDivImage.substring(
312                 divImageStart, divImageEnd);
313             imageStart = divImage.indexOf("data-lazy-src");
314             imageEnd = divImage.indexOf("<noscript\"", imageStart);
315             openedLinkImage = divImage.substring(
316                 imageStart+15, imageEnd-3).replace("\\"", "");
317             openedLinkImageValue = imageCategory.get(openedLinkImage);
318         }
```

Εικόνα 5.3.41 Εξαγωγή της κατηγορίας σε μορφή εικόνας (4)

```
319         if (openedLinkImageValue != null) {
320             rawImageValues.add(openedLinkImageValue);
321             catVerdicts.add(openedLinkImageValue);
322         }
323
324         divImageStart = mainDivImage.indexOf(
325             "class=\"wp-block-image\"", divImageEnd);
326
327     }
328
329     imageValueString = rawImageValues.toString();
330     imageValueString = imageValueString.replace("[", "");
331     imageValueString = imageValueString.replace("]", "");
332 } catch (Exception e2) {
333     imageValueString = "";
334 }
335 }
```

5.3.2.2 ΑΝΑΚΤΗΣΗ ΤΟΥ ΤΙΤΛΟΥ ΕΝΟΣ ΑΡΘΡΟΥ

Εικόνα 5.3.42 Τίτλος



Το κομμάτι του HTML-κώδικα, όπου βρίσκεται ο τίτλος παρουσιάζεται στις Εικόνες 5.3.43 – 5.3.44, ο Java-κώδικας που χρησιμοποιήθηκε παρουσιάζεται στην Εικόνα 5.3.45.

Εικόνα 5.3.43 Τίτλος στο HTML-κώδικα (1)

```
<!DOCTYPE html>
<html lang="el" data-lt-installed="true">
  <head>
  </head>
  <body class="post-template-default single single-post postid-27100 single-format-standard wp-embed-responsive multiple-domain-www-ellinikahoaxes-gr">
    <div class="header-menu d-xl-block d-lg-block d-none">
    </div>
    <div class="header-menu d-block d-md-none d-lg-none d-xl-none">
    </div>
    <div class="header-menu d-lg-none d-md-block d-xl-none d-none">
    </div>
    <div class="mobile-side-menu">
    </div>
    <div class="mobile-menu-overlay2">
    </div>
    <div id="header-widget-area" class="chw-widget-area widget-area" role="complementary" style="width: 100%; height: auto; margin: 0 0 0 0;">
    </div>
    <div class="white-row" style="height: auto; margin: 0 0 0 0;">
    <div class="container" style="height: auto; margin: 0 0 0 0;">
      <div data-bg="https://media.ellinikahoaxes.gr/uploads/2018/01/20150311_cancerresearch1.jpg" class="single-header single-img rocket-lazyload entered lazy">
      </div>
      <div class="content-bottom">
        <div class="row">
          <div class="col-sm-12">
            <span class="inner-text-img">
            </span>
            <span class="header-category">
            </span>
            <h1 style="font-size: 1.2em; margin: 0 0 0 0;">Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1>
            <div class="blog-date" style="font-size: 0.8em; margin: 0 0 0 0;">5 Ιανουαρίου 2018</div>
            <p class="header-author" style="font-size: 0.8em; margin: 0 0 0 0;">
            </p>
          </div>
        </div>
      </div>
    </div>
  </div>
  <div class="row" style="height: auto; margin: 0 0 0 0;">
  </div>
</div>
<section id="footer">
</section>
```

Εικόνα 5.3.44 Τίτλος στο HTML-κώδικα (2)

```
<h1 style="font-size: 1.2em; margin: 0 0 0 0;">Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1> == $0
```

Εικόνα 5.3.45 Εξαγωγή του τίτλου

```
//Retrieves the title
int titleStart = openedLinkContent.indexOf("<h1");
int titleEnd = openedLinkContent.indexOf("</h1>");
String openedLinkTitle = openedLinkContent.substring(
    titleStart+4, titleEnd);

if (openedLinkTitle.contains("<strong>")) {
    openedLinkTitle = openedLinkTitle.replace(
        "</strong>", "").replace("<strong>", "");
}
```

5.3.2.3 ΑΝΑΚΤΗΣΗ ΤΗΣ ΗΜΕΡΟΜΗΝΙΑΣ ΕΝΟΣ ΑΡΘΡΟΥ

Εικόνα 5.3.46 Ημερομηνία



Το κομμάτι του HTML-κώδικα, όπου βρίσκεται η ημερομηνία παρουσιάζεται στις Εικόνες 5.3.47 – 5.3.48, ο Java-κώδικας που χρησιμοποιήθηκε παρουσιάζεται στην Εικόνα 5.3.49.

Εικόνα 5.3.47 Ημερομηνία στο HTML-κώδικα (1)

```
<!DOCTYPE html>
<html lang="el" data-1t-installed="true">
<head>
</head>
<body class="post-template-default single single-post postid-27100 single-format-standard wp-embed-responsive multiple-domain-www-ellinikahoaxes-gr">
<div class="header-menu d-xl-block d-lg-block d-none">
</div>
<div class="header-menu d-block d-md-none d-lg-none d-xl-none">
</div>
<div class="mobile-side-menu">
</div>
<div class="mobile-menu-overlay2">
</div>
<div id="header-widget-area" class="chw-widget-area widget-area" role="complementary" style="
</div>
<div class="white-row" style="height: auto !important;">
<div class="container" style="height: auto !important;">
<div data-bg="https://media.ellinikahoaxes.gr/uploads/2018/01/20150311_cancerresearch1.jpg" class="single-header single-img rocket-lazyload enter
edia.ellinikahoaxes.gr/uploads/2018/01/20150311_cancerresearch1.jpg&quot;);" data-1l-status="loaded">
::before
<div class="content-bottom">
<div class="row">
<div class="col-sm-12">
<span class="inner-text-img">
</span>
<span class="header-category">
</span>
<h1 style="font-size: 24px; margin: 0 0 10px 0;">Όχι, ο γιατρός Ευάγγελος Μιχαλάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1>
<div class="blog-date" style="font-size: 18px; font-weight: bold; margin-top: 5px 0 0 0;">5 Ιανουαρίου 2018</div>
<p class="header-author">
<a href="https://www.ellinikahoaxes.gr/author/thanos80w/" class="author" rel="author" data-wpel-link="internal" style="font-size: 18px; font-weight: bold;">Θάνος Σιτίτσας</a>
</p>
</div>
</div>
</div>
</div>
<div class="row" style="height: auto !important;">
</div>
</div>
```

Εικόνα 5.3.48 Ημερομηνία στο HTML-κώδικα (2)

```
<div class="blog-date" style">5 Ιανουαρίου 2018</div> == $0
```

Εικόνα 5.3.49 Εξαγωγή της ημερομηνίας

```
//Retrieves the date
int dateStart = openedLinkContent.indexOf(
    "<div class=\"blog-date\"");
int dateEnd = openedLinkContent.indexOf("</div>", dateStart);
String openedLinkDate = openedLinkContent.substring(
    dateStart+23, dateEnd);

//Splits the date in Greek into 3: date, month, and year
String[] dateName = openedLinkDate.split(" ");
dateName[1] = monthConvert.get(dateName[1]);
String dateNumber = dateName[0] + "-" + dateName[1] + "-"
    + dateName[2];
```

5.3.2.4 ΑΝΑΚΤΗΣΗ ΤΟΥ ΣΥΝΤΑΚΤΗ (ΕΛΕΓΚΤΗ) ΕΝΟΣ ΑΡΘΡΟΥ

Εικόνα 5.3.50 Συντάκτης (ελεγκτής) του άρθρου



Το κομμάτι του HTML-κώδικα, όπου βρίσκεται το όνομα του συντάκτη παρουσιάζεται στις Εικόνες 5.3.51 – 5.3.52, ο Java-κώδικας που χρησιμοποιήθηκε παρουσιάζεται στην Εικόνα 5.3.53.

Εικόνα 5.3.51 Όνομα του συντάκτη στο HTML-κώδικα (1)

```

<noscript></noscript>
<strong>Ευαίσθητο περιεχόμενο</strong>
<p></p>
</span>
<span class="header-category">
  <span>
    <span></span>
    <span class="breadcrumb_last" aria-current="page">Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</span>
  </span>
  <a href="https://www.ellinikahoaxes.gr/category/kathories/ygeia/" rel="category tag" data-wpel-link="internal"></a>
  <a href="https://www.ellinikahoaxes.gr/category/kathories/psevdoepisthmi/" rel="category tag" data-wpel-link="internal" style></a>
</span>
<h1 style>Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου</h1>
<div class="blog-date" style>5 Ιανουαρίου 2018</div>
<p class="header-author">
  <a href="https://www.ellinikahoaxes.gr/author/thanos80w/" class="author" rel="author" data-wpel-link="internal" style>Θάνος Σιτίστας</a> -- $0
</p>
</div>
</div>
</div>
</div>
<div class="row" style="height: auto !important;"> </div>
</div>
</div>
<section id="footer"></section>
<section id="copyright"></section>
<style id="core-block-supports-inline-css"></style>
<script id="google_cse_v2-js-extra"></script>
<script src="https://www.ellinikahoaxes.gr/wp-content/plugins/wp-google-search/assets/js/google_cse_v2.js?ver=1" id="google_cse_v2-js"></script>
<script id="collapseomatic-js-before"></script>
<script src="https://www.ellinikahoaxes.gr/wp-content/plugins/collapseomatic/js/collapseomatic.js?ver=1.2.2" id="collapseomatic-js"></script>
</pre>
```

Εικόνα 5.3.52 Όνομα του συντάκτη στο HTML-κώδικα (2)

```
rel="author" data-wpel-link="internal" style>Θάνος Σιτίστας</a>
```

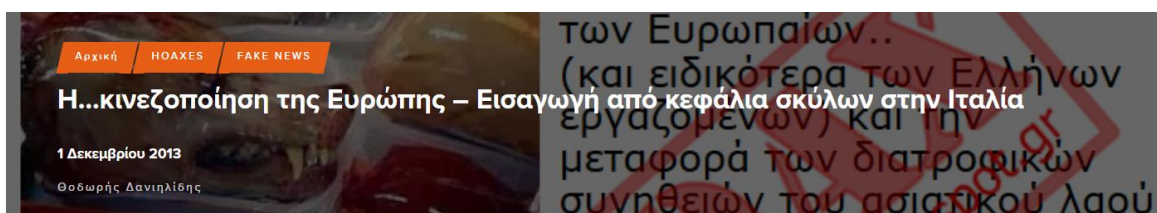
Εικόνα 5.3.53 Εξαγωγή του ονόματος του συντάκτη

```
//Retrieves the author
int authorStart = openedLinkContent.indexOf(
    "rel=\"author\" data-wpel-link=\"internal\"");
int authorEnd = openedLinkContent.indexOf("</a>", authorStart);
String openedLinkAuthor = openedLinkContent.substring(
    authorStart+39, authorEnd);
```

5.3.2.5 ΑΝΑΚΤΗΣΗ ΤΩΝ ΙΣΧΥΡΙΣΜΩΝ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΩΝ ΕΝΟΣ ΑΡΘΡΟΥ

Η ανάκτηση των ισχυρισμών και των συμπερασμάτων αποτελούσε ένα από τα δυσκολότερα κομμάτια της εργασίας διότι τα άρθρα είναι πάρα πολλά και η δομή τους καμιά φορά διαφέρει πολύ. Για να εντοπιστεί ένα άρθρο με διαφορετική δομή έπρεπε να συγγραφεί ο κώδικας έστω για μια περίπτωση και μετά να αναλυθεί σε ποια από τα άρθρα η ανάκτηση του ισχυρισμού και του συμπεράσματος δεν ήταν αποτελεσματική. Δυστυχώς, σε μερικές από τις παλιές δημοσιεύσεις η ανάκτηση δεν ήταν δυνατή γιατί δεν υπήρχαν τα στοιχεία «Ισχυρισμός» και «Συμπέρασμα» στην ουσία. Μια από αυτές τις περιπτώσεις¹⁰⁴ παρουσιάζεται στην Εικόνα 5.3.54.

Εικόνα 5.3.54 Παράδειγμα άρθρου χωρίς τα στοιχεία «Ισχυρισμός» και «Συμπέρασμα»



Τα Ελληνικά Hoaxes είναι μέλος του International Fact-Checking Network Το σήμα αυτό αποδεικνύει ότι τα ellinikahoaxes.gr ελέγχθηκαν και είναι σύμφωνα με τις αρχές του IFCN, για πολιτική ανεξαρτησία, διαφάνεια των πηγών, της χρηματοδότησης, της μεθοδολογίας, και

i Το συγκεκριμένο άρθρο δημοσιεύτηκε πριν 10 έτη.

Ένα ακόμη ανυπόστατο τρομολαγνικό, άρθρο από τα "γνωστά" και δυστυχώς πολυπληθή site αυτού του τύπου...

Κυριακή, 3 Νοεμβρίου 2013

Η...κινεζοποίηση της Ευρώπης - Εισαγωγή από κεφάλια σκύλων στην Ιταλία



Η...κινεζοποίηση της Ευρώπης έχει σαν αποτέλεσμα πέρα από τις εξωπικές τάσεις ανάμεσα στις αμοιβές των κινέζων και των Ευρωπαίων... (και ειδικότερα των Ελλήνων εργαζομένων) και την

Τα πιο πολλά έχουν δομή όπου είναι δυνατή η ανάκτηση των ισχυρισμών και των συμπερασμάτων. Αυτά τα άρθρα έχουν μια από τις δυο δομές (Εικόνες 5.3.55 και 5.3.56).

¹⁰⁴ <https://www.ellinikahoaxes.gr/2013/12/01/ηκινεζοποίηση-της-ευρώπης-εισαγω/> (προσπελάστηκε στις 26/01/2024)

Εικόνα 5.3.55 Πρώτη δομή της σελίδας

Αρχική HOAXES

Όχι, δεν υπάρχουν λουλούδια με τη μορφή γυναίκας

17 Ιουλίου 2020
Σπυριδούλα Μάρκου

Τα Ελληνικά Hoaxes είναι μέλος του International Fact-Checking Network. Το σήμα αυτό αποδεικνύει ότι τα ellinikahoaxes.gr ελέγχθηκαν και είναι σύμφωνα με τις αρχές του IFCN, για πολιτική ανεξαρτησία, διαφάνεια των πηγών, της χρηματοδότησης, της μεθοδολογίας, και δέσμευση για ανοικτές και ειλικρινείς διορθώσεις.

! Το συγκεκριμένο άρθρο δημοσιεύτηκε πριν 4 έτη.

Ισχυρισμός:
“ Δέντρο σε περιοχή των Ιμαλαίων έχει λουλούδια σε μορφή γυναίκας. ”

Συμπέρασμα:
Πρόκειται για αναπαράσταση μυθικού δέντρου σε πάρκο στην Ταϊλάνδη.

Εικόνα 5.3.56 Δεύτερη δομή της σελίδας

Αρχική ΥΓΕΙΑ ΨΕΥΔΟΕΠΙΣΤΗΜΗ

Όχι, ο γιατρός Ευάγγελος Μιχελάκης ΔΕΝ βρήκε το φάρμακο εναντίον του καρκίνου

5 Ιανουαρίου 2018
Θάνος Σιτίτσας

Τα Ελληνικά Hoaxes είναι μέλος του International Fact-Checking Network. Το σήμα αυτό αποδεικνύει ότι τα ellinikahoaxes.gr ελέγχθηκαν και είναι σύμφωνα με τις αρχές του IFCN, για πολιτική ανεξαρτησία, διαφάνεια των πηγών, της μεθοδολογίας, και δέσμευση για ανοικτές και ειλικρινείς διορθώσεις.

! Το συγκεκριμένο άρθρο δημοσιεύτηκε πριν 6 έτη.

Στην ελληνική πλευρά του διαδικτύου κυκλοφορεί ο ισχυρισμός ότι ο γιατρός Ευάγγελος Μιχελάκης στην Αλμπέρτα του Καναδά, ανακάλυψε ότι η χορήγηση της ουσίας DCA σε ασθενείς που πάσχουν από καρκίνο βοηθά δραστικά στην υποχώρηση της ασθένειας, αλλά επειδή είναι φθηνό, οι φαρμακοβιομηχανίες δεν το χορηγούν. Ο εν λόγω ισχυρισμός είναι ψευδής.

Ισχυρισμός:
Ο γιατρός Ευάγγελος Μιχελάκης ανακάλυψε ένα φάρμακο για την καταπολέμηση του καρκίνου αλλά επειδή είναι πολύ φθηνό οι φαρμακοβιομηχανίες δε το χορηγούν.

Συμπέρασμα:
Το DCA έχει εξεταστεί ως πιθανός αντικαρκινικός παράγοντας για συγκεκριμένους τύπους καρκίνου αλλά μετά από πολλά χρόνια μελετών δεν τεκμηριώθηκε η αποτελεσματικότητά του για χρήση σε ανθρώπους. Επιπρόσθετα η τιμή ενός φαρμάκου περιλαμβάνει το κόστος επένδυσης της φαρμακευτικής εταιρίας για την ανάπτυξη του επομένως, αν το εν λόγω φάρμακο κυκλοφορούσε η τιμή του δε θα ήταν χαμηλή.

Είναι φανερό πώς το κομμάτι του HTML-κώδικα, όπου βρίσκονται οι ισχυρισμοί και τα συμπεράσματα, σε αυτές τις δυο περιπτώσεις διαφέρει πολύ. Οι Εικόνες 5.3.57 – 5.3.62 αποδεικνύουν τις δυο διαφορετικές τοποθεσίες των ισχυρισμών στο HTML-κώδικα και οι

Εικόνες 5.3.63 – 5.3.68 αποδεικνύουν τις δυο διαφορετικές τοποθεσίες των συμπερασμάτων στο HTML-κώδικα.

Εικόνα 5.3.57 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (1)

```
<!DOCTYPE html>
<html lang="el" data-lt-installed="true">
  <head> </head>
  <body class="post-template-default single single-post postid-66233 single-format-standard wp-embed-responsive multiple-domain-www-ellin...
    <div class="header-menu d-xl-block d-lg-block d-none fixed-header"> </div>
    <div class="header-menu d-lg-none d-md-block d-xl-none d-none fixed-header"> </div>
    <div class="mobile-side-menu"> </div>
    <div class="mobile-menu-overlay2"></div>
    <div id="header-widget-area" class="chw-widget-area widget-area" role="complementary"> </div>
    <div class="white-row" style="height: auto !important;">
      <div class="container" style="height: auto !important;">
        <div data-bg="https://media.ellinikahoaxes.gr/uploads/2020/07/31.png" class="single-header single-img rocket-lazyload entered lazy
r/uploads/2020/07/31.png&quot;;" data-ll-status="loaded"> </div>
        <div class="row" style="height: auto !important;"> <flex>
          <div class="col-md-3 col-12 order-md-1 order-3 sidebar-container" style="height: auto !important;"> </div>
          <div class="col-md-9 col-12 order-md-2 order-2 main-content-body">
            <div class="alert" style="</div>
            <div class="row" style="flex>
              <div class="col-12 text-md-left" style>
                <div class="claim_conclusion" style>
                  <div class="true-false" style>
                    <div class="Claim" style>
                      <h4 style>Ισχυρισμός:</h4>
                      <p style>
                        ::before
                        "Δέντρο σε περιοχή των Ιμαλαΐων έχει λουλούδια σε μορφή γυναίκας." == $0
                        ::after
                      </p>
                    </div>
                  <div class="conclusion" style> </div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>
  </body>
</html>
```

Εικόνα 5.3.58 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (2)

```
<div class="Claim" style>
  <h4 style>Ισχυρισμός:</h4>
  <p style>
    ::before
    "Δέντρο σε περιοχή των Ιμαλαΐων έχει λουλούδια σε μορφή γυναίκας." == $0
    ::after
  </p>
</div>
```

Εικόνα 5.3.59 Ισχυρισμός στο HTML-κώδικα, πρώτη περίπτωση (3)

```
<p style>
  ::before
  "Δέντρο σε περιοχή των Ιμαλαΐων έχει λουλούδια σε μορφή γυναίκας." == $0
  ::after
</p>
```

Εικόνα 5.3.60 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (1)

```
    <p class="header-author"></p>
  </div>
</div>
</div>
</div>
<div class="row" style="height: auto !important;"> flex
  <div class="col-md-3 col-12 order-md-1 order-3 sidebar-container" style="height: auto !important;"></div>
  <div class="col-md-9 col-12 order-md-2 order-2 main-content-body" style="height: auto !important;">
    <div class="alert" style"></div>
    <div class="row"> flex
  <div class="single-article post-27100 post type-post status-publish format-standard has-post-thumbnail hentry category-ygeia cate
id="post-27100" style>
  <p></p>
  <div class="wp-block-columns is-layout-flex wp-container-core-columns-layout-1 wp-block-columns-is-layout-flex" style"> flex
    <div class="wp-block-column is-layout-flow wp-block-column-is-layout-flow" style>
      <p class="has-background has-luminous-vivid-orange-background-color" style>
        <strong></strong>
        <br>
        <br>
... "Ο γιατρός Ευάγγελος Μιχελάκης ανακάλυψε ένα φάρμακο για την καταπολέμηση του καρκίνου αλλά επειδή είναι πολύ φθηνό οι φαρ
      </p>
    </div>
    <div class="wp-block-column is-layout-flow wp-block-column-is-layout-flow" style>
      <p class="has-background has-vivid-green-cyan-background-color" style"></p>
    </div>
  </div>
  <hr class="wp-block-separator">
  <div class="wp-block-image" style"></div>
  <div class="wp-block-image" style"></div>
  <hr class="wp-block-separator">
  <p></p>
</div>
</div>
```

Εικόνα 5.3.61 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (2)

```
<p class="has-background has-luminous-vivid-orange-background-color" style>
  <strong></strong>
  <br>
  <br>
  "Ο γιατρός Ευάγγελος Μιχελάκης ανακάλυψε ένα φάρμακο για την καταπολέμηση του καρκίνου αλλά επειδή είναι πολύ φθηνό οι φαρμακοβιομηχανίες δε το χορηγούν." == $0
</p>
```

Εικόνα 5.3.62 Ισχυρισμός στο HTML-κώδικα, δεύτερη περίπτωση (3)

```
<br>
  "Ο γιατρός Ευάγγελος Μιχελάκης ανακάλυψε ένα φάρμακο για την καταπολέμηση του καρκίνου αλλά επειδή είναι πολύ φθηνό οι φαρμακοβιομηχανίες δε το χορηγούν." == $0
</p>
```

Εικόνα 5.3.63 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (1)

```
</div class="alert" style"></div>
▼ <div class="row" style"> flex
  ▼ <div class="col-12 text-md-left" style">
    ▼ <div class="claim_conclusion" style">
      ▼ <div class="true-false" style">
        ▼ <div class="Claim" style">
          <h4 style">Ισχυρισμός:</h4>
          ▼ <p style">
            ::before
            "Δέντρο σε περιοχή των Ιμαλαΐων έχει λουλούδια σε μορφή γυναίκας."
            ::after
          </p>
        </div>
        ▼ <div class="conclusion" style">
          <h4 style">Συμπέρασμα:</h4>
          <p style">Πρόκειται για αναπαράσταση μυθικού δέντρου σε πάρκο στην Ταϊλάνδη.</p> == $0
        </div>
      </div>
    </div>
  </div>
</div>
▼ <div class="row" style"> flex
  ▼ <div class="col-12 text-center stams" style">
    
    ▶ <noscript> </noscript>
  </div>
  <div class="single-article post-66233 post type-post status-publish format-standard has-post-thumbnail hentry category-hoax tag-430 tag-24" style">
  <div class="author row padding-author"> </div> flex
  <script async src="https://pagead2.googlesyndication.com/pagead/js/adsbygoogle.js"></script>
  <!-- Custom ad 1 -->
  </div>
```

Εικόνα 5.3.64 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (2)

```
<div class="conclusion" style">
  <h4 style">Συμπέρασμα:</h4>
  <p style">Πρόκειται για αναπαράσταση μυθικού δέντρου σε πάρκο στην Ταϊλάνδη.</p> == $0
</div>
```

Εικόνα 5.3.65 Συμπέρασμα στο HTML-κώδικα, πρώτη περίπτωση (3)

```
<p style">Πρόκειται για αναπαράσταση μυθικού δέντρου σε πάρκο στην Ταϊλάνδη.</p> == $0
```

Εικόνα 5.3.66 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (1)

```
><div class="alert" style"> </div>
><div class="row"> </div> flex
▼<div class="single-article post-27100 post type-post status-publish format-standard has-post-thumbnail hentry category-ygeia category-psevdoepis
id="post-27100" style">
  <p> </p>
  <div class="wp-block-columns is-layout-flex wp-container-core-columns-layout-1 wp-block-columns-is-layout-flex" style"> flex
    <div class="wp-block-column is-layout-flow wp-block-column-is-layout-flow" style">
      <p class="has-background has-luminous-vivid-orange-background-color" style">
        <strong> </strong>
        <br>
        <br>
        "Ο γιατρός Ευάγγελος Μιχελάκης ανακάλυψε ένα φάρμακο για την καταπολέμηση του καρκίνου αλλά επειδή είναι πολύ φθηνό οι φαρμακοβιομηχανίες
      </p>
    </div>
    <div class="wp-block-column is-layout-flow wp-block-column-is-layout-flow" style">
      <p class="has-background has-vivid-green-cyan-background-color" style">
        <strong> </strong>
        <br>
        <br>
        "Το DCA έχει εξεταστεί ως πιθανός αντικαρκινικός παράγοντας για συγκεκριμένους τύπους καρκίνου αλλά μετά από πολλά χρόνια μελετών δεν τεκ
        ανθρώπους. Επιπρόσθετα η τιμή ενός φαρμάκου περιλαμβάνει το κόστος επένδυσης της φαρμακευτικής εταιρίας για την ανάπτυξη του επομένως, αν
        χαμηλή." == $0
      </p>
    </div>
  </div>
  <hr class="wp-block-separator">
  <div class="wp-block-image" style"> </div>
  <div class="wp-block-image" style"> </div>
  <hr class="wp-block-separator">
  <p> </p>
  <p> </p>
  <p> </p>
  <p> </p>
```

Εικόνα 5.3.67 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (2)

```
<p class="has-background has-vivid-green-cyan-background-color" style">
  <strong> </strong>
  <br>
  <br>
  "Το DCA έχει εξεταστεί ως πιθανός αντικαρκινικός παράγοντας για συγκεκριμένους τύπους καρκίνου αλλά μετά από πολλά χρόνια μελετών δεν τεκμηριώθηκε η αποτελεσματικότητα του για χρήση σε
  ανθρώπους. Επιπρόσθετα η τιμή ενός φαρμάκου περιλαμβάνει το κόστος επένδυσης της φαρμακευτικής εταιρίας για την ανάπτυξη του επομένως, αν το εν λόγω φάρμακο κυκλοφορούσε η τιμή του δε θα ήταν
  χαμηλή." == $0
</p>
```

Εικόνα 5.3.68 Συμπέρασμα στο HTML-κώδικα, δεύτερη περίπτωση (3)

```
<br>
"Το DCA έχει εξεταστεί ως πιθανός αντικαρκινικός παράγοντας για συγκεκριμένους τύπους καρκίνου αλλά μετά από πολλά χρόνια μελετών δεν τεκμηριώθηκε η αποτελεσματικότητα του για χρήση σε
ανθρώπους. Επιπρόσθετα η τιμή ενός φαρμάκου περιλαμβάνει το κόστος επένδυσης της φαρμακευτικής εταιρίας για την ανάπτυξη του επομένως, αν το εν λόγω φάρμακο κυκλοφορούσε η τιμή του δε θα ήταν
χαμηλή." == $0
</p>
```

Επίσης, δεν υπάρχει ένας ορισμένος αριθμός των ισχυρισμών και των συμπερασμάτων που μπορεί να περιέχει ένα άρθρο, συνεπώς η ανάκτηση αυτών των δεδομένων από τη μια ιστοσελίδα γίνεται μέσω βρόγχου και τα δεδομένα προσωρινά αποθηκεύονται στα HashSets. Στις Εικόνες 5.3.69 και 5.3.70 παρουσιάζεται η διαδικασία ανάκτησης ισχυρισμών και στις Εικόνες 5.3.71 και 5.3.72 παρουσιάζεται η διαδικασία ανάκτησης συμπερασμάτων.

Εικόνα 5.3.69 Εξαγωγή του ισχυρισμού, πρώτη περίπτωση

```
String openedLinkClaim = "";
int divClaimStart;
int divClaimEnd;
int pClaimStart;
int pClaimEnd;
String pClaim;
String divClaim;
int claimStart;
int claimEnd;

//Retrieves the claim/claims
divClaimStart = openedLinkContent.indexOf("<div class=\"Claim\"");

while (divClaimStart != -1) {
    System.out.println("Claim - 1st try");
    divClaimEnd = openedLinkContent.indexOf(
        "</div\"", divClaimStart);
    divClaim = openedLinkContent.substring(
        divClaimStart, divClaimEnd);
    claimStart = divClaim.indexOf("<p");
    claimEnd = divClaim.indexOf("</p");
    openedLinkClaim = divClaim.substring(
        claimStart + 3, claimEnd);
    //Adds all found claims to the hashset "claims"
    claims.add(openedLinkClaim);
    divClaimStart = openedLinkContent.indexOf(
        "<div class=\"Claim\"", divClaimEnd);
}
}
```

Εικόνα 5.3.70 Εξαγωγή του ισχυρισμού, δεύτερη περίπτωση

```
/* If claims had not been retrieved earlier,
The webpage has a different structure, need to try again
*/
if (openedLinkClaim.isEmpty() ) {

    try {
        pClaimStart = openedLinkContent.indexOf(
            "<p class=\"has-background has-luminous-vivid-orange-background-color\"");
        pConclusionStart = openedLinkContent.indexOf(
            "<p class=\"has-background has-vivid-green-cyan-background-color\"");

        while (pConclusionStart != -1) {

            //Tries to retrieve a claim
            System.out.println("Claim - 2nd try");
            pClaimEnd = openedLinkContent.indexOf(
                "</p\"", pClaimStart);
            pClaim = openedLinkContent.substring(
                pClaimStart, pClaimEnd);
            claimStart = pClaim.indexOf("<br");
            openedLinkClaim = pClaim.substring(
                claimStart+4).replace("<br>", "");
            claims.add(openedLinkClaim);
            pClaimStart = openedLinkContent.indexOf(
                "<p class=\"has-background has-luminous-vivid-orange-background-color\"",
                pClaimEnd);
        }
    }
}
```

Εικόνα 5.3.71 Εξαγωγή του συμπεράσματος, πρώτη περίπτωση

```
String openedLinkConclusion = "";
int pConclusionStart;
int pConclusionEnd;
String pConclusion;
int conclusionStart;
int divConclusionStart;
int divConclusionEnd;
String divConclusion;
int conclusionEnd;

//Retrieves the conclusion/conclusions
divConclusionStart = openedLinkContent.indexOf("<div class=\"conclusion\"");

while (divConclusionStart != -1) {
    System.out.println("Conclusion - 1st try");
    divConclusionEnd = openedLinkContent.indexOf(
        "</div", divConclusionStart);
    divConclusion = openedLinkContent.substring(
        divConclusionStart, divConclusionEnd);
    conclusionStart = divConclusion.indexOf("<p");
    conclusionEnd = divConclusion.indexOf("</p");
    openedLinkConclusion = divConclusion.substring(
        conclusionStart + 3, conclusionEnd);
    //Adds all found conclusions to the hashset "conclusions"
    conclusions.add(openedLinkConclusion);
    divConclusionStart = openedLinkContent.indexOf(
        "<div class=\"conclusion\"", divConclusionEnd);
}
}
```

Εικόνα 5.3.72 Εξαγωγή του συμπεράσματος, δεύτερη περίπτωση

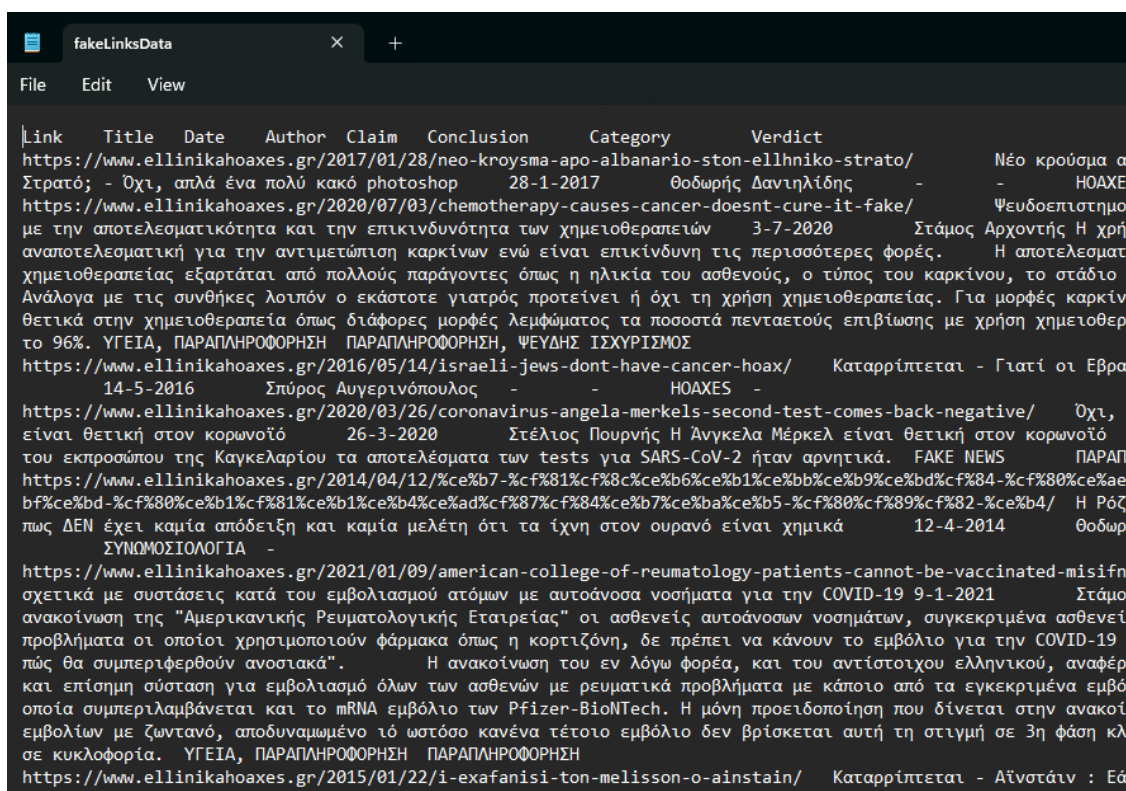
```
    //Tries to retrieve a conclusion
    System.out.println("Conclusion - 2nd try");
    pConclusionEnd = openedLinkContent.indexOf(
        "</p", pConclusionStart);
    pConclusion = openedLinkContent.substring(
        pConclusionStart, pConclusionEnd);
    conclusionStart = pConclusion.indexOf("<br");
    openedLinkConclusion = pConclusion.substring(
        conclusionStart + 4).replace("<br>", "");
    conclusions.add(openedLinkConclusion);
    pConclusionStart = openedLinkContent.indexOf(
        "< p class=\"has-background has-vivid-green-cyan-background-color\"",
        pConclusionEnd);
}
} catch (Exception e) {
}
}
```


Το αρχείο fakeLinksData.csv χρησιμοποιείται ως μια βάση δεδομένων την όποια μπορεί να διαχειρίζεται μόνο μέσω του κώδικα της κλάσης PageDataRetrieval και σε καμία περίπτωση αυτό το αρχείο δεν πρέπει να σώζεται μέσω οποιασδήποτε άλλης εφαρμογής, πχ MS Excel, γιατί η επόμενη σωστή εγγραφή των νέων άρθρων δε θα είναι δυνατή. Όλες οι πράξεις και επεξεργασίες πρέπει να γίνονται πάνω σε αντιγραφή αυτών των δεδομένων μετά την αποκωδικοποίηση.

Για την καλύτερη κατανόηση τα βήματα περιγράφονται παρακάτω:

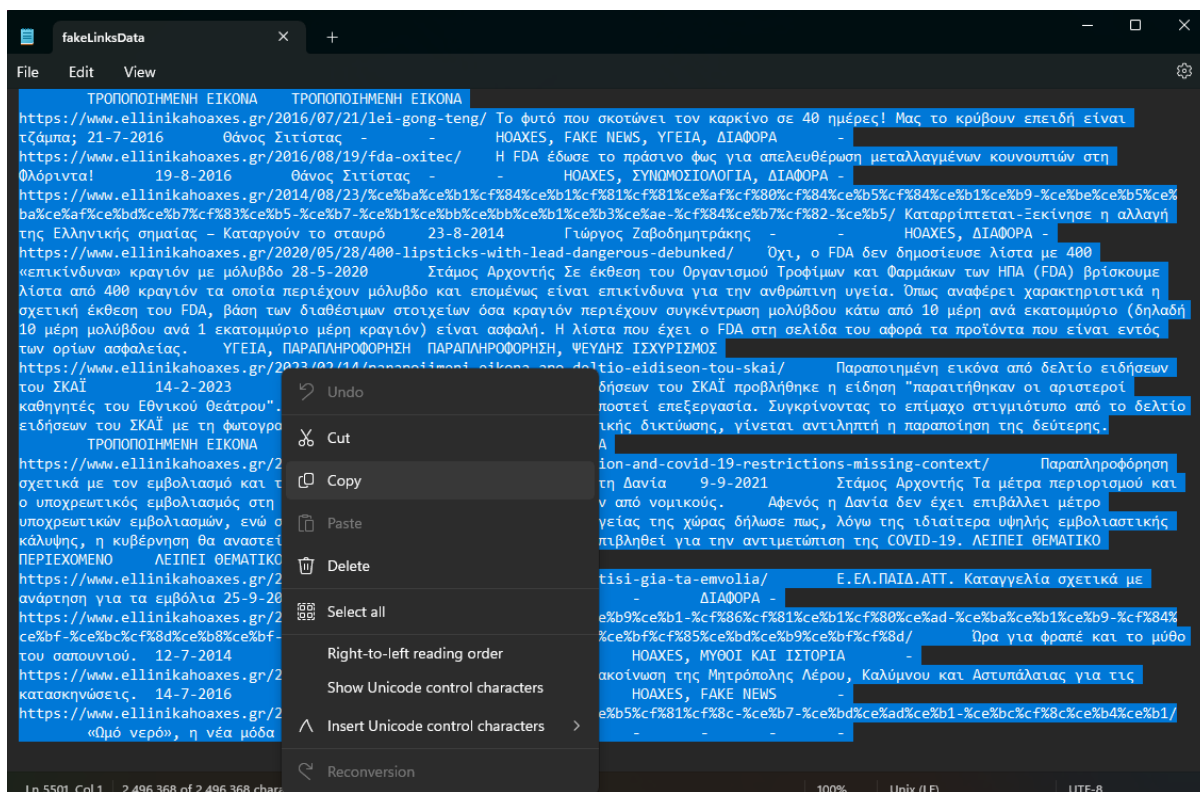
1. Το αρχείο ανοίγεται μέσω Notepad (Εικόνα 5.3.74).

Εικόνα 5.3.74 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (1)



2. Το περιεχόμενο αυτού του αρχείου αντιγράφεται (Εικόνα 5.3.75) και επικολλάται σε ένα νέο αρχείο CSV UTF-8 (Εικόνα 5.3.76).

Εικόνα 5.3.75 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (2)



Εικόνα 5.3.76 Βήματα αποκωδικοποίησης του αρχείου fakeLinksData.csv (3)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Link	Title	Date	Author	Claim	Conclusio	Category	Verdict					
2	https://w	Νέο κρού	28-01-17	Θοδωρής -	-	-	HOAXES, F-						
3	https://w	Ψευδοεπι	03-07-20	Στάμος Αρ	Η χρήση χ	Η αποτελε	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ					
4	https://w	Καταρρίπ	14-05-16	Σπύρος Αι	-	-	HOAXES -						
5	https://w	Όχι, η Καγ	26-03-20	Στέλιος Πς	Η Άνγκελς	Σύμφωνα	FAKE NEW	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ					
6	https://w	Η Ρόζαλιν	12-04-14	Θοδωρής -	-	-	ΣΥΝΩΜΟΣ-						
7	https://w	Παραπληρ	09-01-21	Στάμος Αρ	Σύμφωνα	Η ανακοίν	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ					
8	https://w	Καταρρίπ	22-01-15	Γιώργος Γι	-	-	ΨΕΥΤΙΚΑ F-						
9	https://w	Θέλετε να	08-06-17	Θάνος Σιτ	-	-	HOAXES, F-						
10	https://w	Καταρρίπ	14-01-15	Γιώργος Ζι	-	-	ΣΥΝΩΜΟΣ-						
11	https://w	Η ψευδής	27-04-20	Στέλιος Πς	Ο Μελ Γκί	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ					
12	https://w	Η ψευδής	27-04-20	Στέλιος Πς	ο γνωστός	Ο ηθοποι	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ					
13	https://w	Γαλλία: Πς	19-05-21	Στέλιος Πς	Εκατοντά	Τα εικονι	ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ					
14	https://w	Όχι, ο καρ	21-07-19	Στάμος Αρ	-	-	ΥΓΕΙΑ ΨΕΥΔΟΕΠΙΣΤΗΜΗ						
15	https://w	Διαδήλωσ	14-05-18	Ανδρόνικς	-	-	ΠΑΡΑΠΛΗ-						
16	https://w	Καταρρίπ	22-10-15	Θοδωρής -	-	-	HOAXES, Σ-						
17	https://w	Να γιατί π	09-10-14	Γιώργος Γι	-	-	ΔΙΑΦΟΡΑ -						
18	https://w	Πήγε στην	09-02-17	Θοδωρής -	-	-	HOAXES, F-						
19	https://w	Ξαναεμφ	24-12-15	Θοδωρής -	-	-	ΑΠΑΘΗ, LI -						
20	https://w	Παραπληρ	03-09-20	Ανδρόνικς	Στη διαδή	Η αστυνο	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ					
21	https://w	Παραπληρ	03-09-20	Ανδρόνικς	Άρματα μ	Η μετακί	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ					
22	https://w	Ο Αμερικ	31-01-17	Θοδωρής -	-	-	HOAXES, F-						

3. Το αρχείο αυτό σώζεται υπό οποιαδήποτε όνομα εκτός fakeLinksData με τύπο CSV UTF-8 (όχι το απλό CSV) ή XLSX, αν σκοπεύετε να πράξετε περαιτέρω επεξεργασία και ανάλυση.

Εγώ το αρχείο το ονόμασα fakeLinksDataAnalysis.xlsx.

Κάθε φορά που αυτός ο κώδικας εκτελείται ξανά, τα βήματα αυτά πρέπει να επαναλαμβάνονται. Επίσης, για γρηγορότερη εκτέλεση και μειωμένη πιθανότητα του σφάλματος 524 (πάρα πολλά αιτήματα από την ίδια διεύθυνση IP), συστήνεται να μη διαγράφονται τα αρχεία fakeLinks.csv και fakeLinksData.csv και να γίνεται απλά επέκταση αυτών των αρχείων.

5.4 ΤΑ ΔΕΔΟΜΕΝΑ

Μετά από την επιτυχή αποκωδικοποίηση των δεδομένων (Δείτε την Ενότητα 5.3.2.6.), τα δεδομένα έχουν τη μορφή όπως στην Εικόνα 5.4.1.

Όλα τα δεδομένα που αξιοποιούνται στην περαιτέρω ανάλυση είναι διαθέσιμα στο αρχείο fakeLinksDataAnalysis.xlsx.

Εικόνα 5.4.1 Αποκωδικοποιημένα δεδομένα, που αποκτήθηκαν με τη βοήθεια του κώδικα στη Java, έτοιμα για περαιτέρω επεξεργασία και ανάλυση

1	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8
2	Link	Title	Date	Author	Claim	Conclusion	Category	Verdict
5468	https://www.ε	Καίνε τα δάση	27-7-2018	Ανδρόνικος Κο	-	-	ΣΥΝΩΜΟΣΙΟΛ	ΠΑΡΑΠΛΗΡΟ
5469	https://www.ε	Φυλακίστηκε	17-11-2018	Θάνος Σιτίστα	-	-	ΠΑΡΑΠΛΗΡΟ	-
5470	https://www.ε	Δεν γνωρίζουμ	3-3-2019	Γιώργος Γιώτης	Η φράση «ευγ	Με βάση όλα τ	ΨΕΥΤΙΚΑ ΡΗΤΑ	ΑΝΕΠΙΒΕΒΑΙΩ
5471	https://www.ε	Όχι η φλούδα	9-10-2019	Στάμος Αρχοντ	Η φλούδα του	Ο ισχυρισμός π	ΠΑΡΑΠΛΗΡΟ	ΠΑΡΑΠΛΗΡΟ
5472	https://www.ε	Ανυπόστατες	29-4-2020	Γιώργος Γιώτης	Ζουν 120 χρόν	Δεν υπάρχει κ	HOAXES	ΨΕΥΔΕΣ
5473	https://www.ε	Ψευδές δημοσ	29-3-2020	Στάμος Αρχοντ	Το πανεπιστήμ	Παρότι αρκετέ	ΠΑΡΑΠΛΗΡΟ	ΠΑΡΑΠΛΗΡΟ
5474	https://www.ε	Όχι, το φάρμα	13-4-2016	Θοδωρής Μαγ	-	-	-	-
5475	https://www.ε	Βίντεο που κυ	19-8-2021	Στέλιος Πουρν	Ρούχα που προ	Πρόκειται για	HOAXES, ΠΑΡΑ	ΨΕΥΔΕΣ
5476	https://www.ε	Εξαγορά της G	26-2-2019	Θάνος Σιτίστα	-	-	FAKE NEWS	ΨΕΥΔΕΣ, CLICK
5477	https://www.ε	Όχι, ο Τζαβέντ	15-5-2023	Έλενα Βατάλα	Φωτογραφία	ο Τζαβέντ Ασλ	HOAXES	ΠΑΡΑΠΛΗΡΟ
5478	https://www.ε	Μήνυση εναντ	31-5-2017	Θάνος Σιτίστα	-	-	HOAXES, FAKE	-
5479	https://www.ε	Καταρρίπτετα	14-9-2015	Θοδωρής Δανι	-	-	HOAXES, FAKE	-
5480	https://www.ε	Καταρρίπτετα	13-2-2015	Γιώργος Ζαβο	-	-	HOAXES, ΔΙΑΦ	-
5481	https://www.ε	Όχι, δικαστήρι	31-1-2022	Σπυριδούλα Μ	Το "Διεθνές Δι	Το "Διεθνές Δι	HOAXES, ΣΥΝΩ	ΨΕΥΔΕΣ, ΣΥΝΩ
5482	https://www.ε	Παραπληροφό	19-5-2023	Έλενα Βατάλα	Σημαντική αύξ	Η αυξομείωση	ΜΙΞΗ ΓΕΓΟΝΟΤ	ΜΙΞΗ ΓΕΓΟΝΟΤ
5483	https://www.ε	Η πριγκίπισσα	23-8-2016	Γιώργος Ζαβο	-	-	ΔΙΑΦΟΡΑ	-
5484	https://www.ε	Καταρρίπτετα	18-1-2016	Θοδωρής Δανι	-	-	ΣΥΝΩΜΟΣΙΟΛ	-
5485	https://www.ε	Ψηφιακά επεξ	8-3-2022	Σπυριδούλα Μ	Οι φωτογραφί	Χρήστης του Τ	ΤΡΟΠΟΠΟΙΗΜ	ΠΑΡΑΠΛΗΡΟ
5486	https://www.ε	Το βίντεο του	6-1-2015	Γιώργος Ζαβο	-	-	ΔΙΑΦΟΡΑ	-
5487	https://www.ε	Μαθήματα προ	1-11-2017	Θάνος Σιτίστα	-	-	HOAXES, FAKE	-
5488	https://www.ε	Καταρρίπτετα	11-7-2015	Θοδωρής Δανι	-	-	HOAXES, ΜΥΘ	-
5489	https://www.ε	Όχι, δυο Αμερι	16-5-2020	Στάμος Αρχοντ	Δυο Αμερικαν	Οι εν λόγω για	ΥΓΕΙΑ, ΠΑΡΑΠ	ΠΑΡΑΠΛΗΡΟ
5490	https://www.ε	Καταρρίπτοντ	6-1-2018	Θοδωρής Δανι	-	-	FAKE NEWS	-
5491	https://www.ε	Παραποιημέν	25-10-2020	Θοδωρής Δανι	Φωτογραφία	Η φωτογραφία	ΤΡΟΠΟΠΟΙΗΜ	ΤΡΟΠΟΠΟΙΗΜ
5492	https://www.ε	Το φυτό που σ	21-7-2016	Θάνος Σιτίστα	-	-	HOAXES, FAKE	-
5493	https://www.ε	Η FDA έδωσε τ	19-8-2016	Θάνος Σιτίστα	-	-	HOAXES, ΣΥΝΩ	-

Επειδή τα άρθρα δεν είχαν και δεν έχουν μια καθορισμένη δομή, υπάρχουν στοιχεία που κάποια άρθρα δεν περιέχουν. Με τη βοήθεια της συνάρτησης COUNTIFS() της Excel, συμπεραίνεται ότι

- Όλα τα άρθρα έχουν τον τίτλο, την ημερομηνία δημοσιεύματος, το όνομα του συντάκτη.
- Σχεδόν όλα τα άρθρα έχουν έστω μια κατηγορία σε μορφή εικόνας ή κειμένου (99.36%) και το 99.08% των άρθρων έχει την κατηγορία σε μορφή κειμένου.
- Σε λιγότερα από τα μισά άρθρα αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα (42.37%).
- Σχεδόν τα μισά άρθρα περιέχουν έστω μια εικόνα κατηγοριοποίησης (46.81%).
- Το 2013 όλα τα άρθρα έχουν έστω μια κατηγορία σε μορφή κειμένου αλλά μόνο ένα άρθρο (0.58%) περιέχει έστω ένα ισχυρισμό και ένα συμπέρασμα και στα δύο άρθρα υπάρχει έστω μια εικόνα-κατηγορία (1.17%).
- Το 2014, όλα τα άρθρα έχουν έστω μια κατηγορία σε μορφή κειμένου, το 2.23% των οποίων περιέχει έστω ένα ισχυρισμό και ένα συμπέρασμα και στο 3.04% των άρθρων υπάρχει έστω μια εικόνα-κατηγορία.
- Το 2015 το 99.5% των άρθρων έχει έστω μια κατηγορία. Μόνο στο 1.26% των άρθρων αναφέρεται έστω ένας ισχυρισμός, ένα συμπέρασμα και μια κατηγορία.
- Το 2016 το 95.7% των άρθρων έχει έστω μια κατηγορία. Στο 2.37% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 2.8% έστω μια εικόνα-κατηγορία.
- Το 2017 το 99.72% των άρθρων έχει έστω μια κατηγορία. Στο 2.54% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 2.68% έστω μια εικόνα-κατηγορία.
- Το 2018 το 98.98% των άρθρων έχει έστω μια κατηγορία. Στο 2.28% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 3.3% έστω μια εικόνα-κατηγορία.
- Το 2019 η κατάσταση ξεκίνησε να αλλάζει, ιδιαιτέρως στην κατηγοριοποίηση, και όλα τα άρθρα έχουν έστω μια κατηγορία. Στο 30.15% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 89.46% έστω μια εικόνα-κατηγορία.

- Το 2020 ή κατάσταση βελτιώθηκε σημαντικά. Το 99.35% των άρθρων έχει έστω μια κατηγορία. Στο 95.59%% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 94.12% έστω μια εικόνα-κατηγορία.
- Το 2021 όλα τα άρθρα έχουν έστω μια κατηγορία. Στο 98.84% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 98.5% έστω μια εικόνα-κατηγορία.
- Το 2022 όλα τα άρθρα έχουν έστω μια κατηγορία. Στο 99.62% των άρθρων αναφέρεται έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 97.53% έστω μια εικόνα-κατηγορία.
- Τέλος, το 2023 σε όλα τα άρθρα αναφέρεται έστω μια κατηγορία και έστω ένας ισχυρισμός και ένα συμπέρασμα και στο 96.36% έστω μια εικόνα-κατηγορία.

Τα δεδομένα απεικονίζονται στους Πίνακες 5.4.1 και 5.4.2 και το Διάγραμμα 5.4.1.

Πίνακας 5.4.1 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (σε αριθμούς)

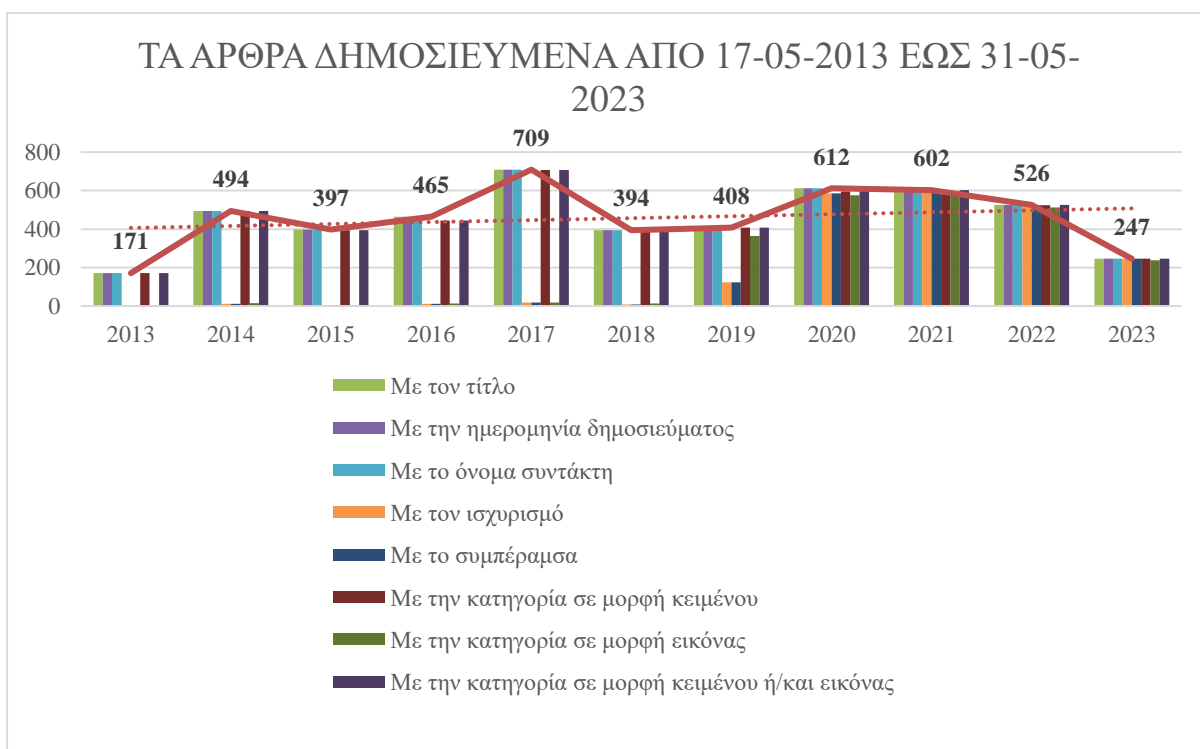
ΕΤΟΣ	ΑΡΙΘΜΟΣ ΑΡΘΡΩΝ, ΑΠΟ ΤΑ ΟΠΟΙΑ	ΜΕ ΤΟΝ ΤΙΤΛΟ	ΜΕ ΤΗΝ ΗΜΕΡΟΜΗΝΙΑ ΔΗΜΟΣΙΕΥΜΑΤΟΣ	ΜΕ ΤΟ ΟΝΟΜΑ ΤΟΥ ΣΥΝΤΑΚΤΗ	ΜΕ ΤΟΝ ΙΣΧΥΡΙΣΜΟ	ΜΕ ΤΟ ΣΥΜΠΕΡΑΣΜΑ	ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΚΕΙΜΕΝΟ)	ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΕΙΚΟΝΑ)	ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΚΕΙΜΕΝΟ Ή/ΚΑΙ ΕΙΚΟΝΑ)
ΣΥΝ.	5025	5025	5025	5025	2129	2129	4979	2352	4993
2013	171	171	171	171	1	1	171	2	171
2014	494	494	494	494	11	11	494	15	494
2015	397	397	397	397	5	5	395	5	395
2016	465	465	465	465	11	11	445	13	445
2017	709	709	709	709	18	18	707	19	707
2018	394	394	394	394	9	9	390	13	390
2019	408	408	408	408	123	123	408	365	408
2020	612	612	612	612	585	585	596	576	608
2021	602	602	602	602	595	595	602	593	602
2022	526	526	526	526	524	524	524	513	526
2023	247	247	247	247	247	247	247	238	247

Πίνακας 5.4.2 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (σε ποσοστά)

ΕΤΟΣ	ΕΤΟΣ	ΣΥΝ.	2013	2014	2015	2016	2017	2018	2019	2020	2021
ΑΡΙΘΜΟΣ ΑΡΘΡΩΝ, ΑΠΟ ΤΑ ΟΠΟΙΑ	5023	171	494	397	465	709	394	408	612	602	
ΜΕ ΤΟΝ ΤΙΤΛΟ	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
ΜΕ ΤΗΝ ΗΜΕΡΟΜΗΝΙΑ ΔΗΜΟΣΙΕΥΜΑΤΟΣ	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
ΜΕ ΤΟ ΟΝΟΜΑ ΤΟΥ ΣΥΝΤΑΚΤΗ	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
ΜΕ ΤΟΝ ΙΣΧΥΡΙΣΜΟ	42.37%	0.58%	2.23%	1.26%	2.37%	2.54%	2.28%	30.15%	95.59%	98.84%	
ΜΕ ΤΟ ΣΥΜΠΕΡΑΣΜΑ	42.37%	0.58%	2.23%	1.26%	2.37%	2.54%	2.28%	30.15%	95.59%	98.84%	
ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΚΕΙΜΕΝΟ)	99.08%	100.00%	100.00%	99.50%	95.70%	99.72%	98.98%	100.00%	97.39%	100.00%	100.00%
ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΕΙΚΟΝΑ)	46.81%	1.17%	3.04%	1.26%	2.80%	2.68%	3.30%	89.46%	94.12%	98.50%	
ΜΕ ΤΗΝ ΚΑΤΗΓΟΡΙΑ (ΚΕΙΜΕΝΟ Ή/ΚΑΙ ΕΙΚΟΝΑ)	99.36%	100.00%	100.00%	99.50%	95.70%	99.72%	98.98%	100.00%	99.35%	100.00%	100.00%

2022	526	100.00%	100.00%	100.00%	99.62%	99.62%	99.62%	97.53%	100.00%
2023	247	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	96.36%	100.00%

Διάγραμμα 5.4.1 Άρθρα δημοσιευμένα από 17-05-2013 έως 31-05-2023 (γραφικά)



Κάτι που μπορεί να παρατηρήσει κάποιος είναι ο συνολικός αριθμός άρθρων (5023) και ο αριθμός των άρθρων στο αρχείο με όλες τις εγγραφές και στην Εικόνα 5.4.1, που είναι πολύ μεγαλύτερος από 5023. Αυτό συμβαίνει λόγω των άρθρων με πολλούς ισχυρισμούς και συμπεράσματα, όπως στην Εικόνα 5.4.2.

Εικόνα 5.4.2 Άρθρο με πολλούς ισχυρισμούς και συμπεράσματα στον ιστότοπο EllhnikaHoaxes.gr

Ισχυρισμός 1:

“ Ο διευθύνων σύμβουλος της Pfizer Α. Μπουρλά αρνήθηκε να κάνει το εμβόλιο κατά του κορωνοϊού. ”

Συμπέρασμα 1:

Ο κ. Μπουρλά είχε δηλώσει ότι θα έκανε το εμβόλιο όταν ερχόταν η σειρά του, και πλέον έχει κάνει και τις δύο δόσεις.

Ισχυρισμός 2:

“ Τα εμβόλια κατά του κορωνοϊού περιέχουν γενετικά τροποποιημένους οργανισμούς και τροποποιούν γενετικά τους εμβολιασμένους. ”

Συμπέρασμα 2:

Μόνο 2 εκ των 4 εμβολίων που έχουν εγκριθεί στην ΕΕ χρησιμοποιούν γενετικά τροποποιημένους οργανισμούς, αλλά ανεξαρτήτως, κανένα εκ των εγκεκριμένων εμβολίων δεν διαθέτει καν τα «εργαλεία» για την τροποποίηση του ανθρωπίνου γονιδιώματος.

Ισχυρισμός 3:

“ Τα παιδιά δεν μεταδίδουν, δεν κολλάνε, και δε χρειάζεται να κάνουν το εμβόλιο του κορωνοϊού. ”

Συμπέρασμα 3:

Στα αρχικά δεδομένα (το αρχείο fakeLinksDataAnalysis.xlsx, το φύλλο fakeLinksDataRAW) είναι όλα τα δεδομένα που ανακτήθηκαν από τον ιστοχώρο EllhnikaHoaxes.gr, συμπεριλαμβανομένων των άρθρων με πολλούς ισχυρισμούς και συμπεράσματα.

Για την περαιτέρω ανάλυση ήταν σημαντικό να σωθούν όλοι οι ισχυρισμοί, γι' αυτό και υπάρχουν περιπτώσεις, όπου ένας σύνδεσμος συσχετίζεται με πολλές εγγραφές (πολλούς ισχυρισμούς), όπως στην Εικόνα 5.4.3.

Εικόνα 5.4.3 Άρθρο με πολλούς ισχυρισμούς και συμπεράσματα μετά την ανάκτηση

5432	ht	Παρανοήσεις περί «μεταλλάξεων στα πα	12	Αν	Ο διευθύνων σύμβουλος της Pfiz	Ο κ. Μπουρλά είχε δηλώσει ότι	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
5433	ht	Παρανοήσεις περί «μεταλλάξεων στα πα	12	Αν	Τα εμβόλια κατά του κορωνοϊού	Μόνο 2 εκ των 4 εμβολίων που ε	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
5434	ht	Παρανοήσεις περί «μεταλλάξεων στα πα	12	Αν	Τα παιδιά δεν μεταδίδουν, δεν κ	Η μετάδοση από παιδιά σε ενήλ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
5435	ht	Παρανοήσεις περί «μεταλλάξεων στα πα	12	Αν	Η χρήση μάσκας από τα παιδιά ε	Φορείς υγείας όπως το Κέντρο Ε	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
5436	ht	Παρανοήσεις περί «μεταλλάξεων στα πα	12	Αν	Τα εμβολιασμένα παιδιά κινδυν	Τα εμβόλια είναι αδύνατο να π	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ

Ωστόσο, για την περαιτέρω ανάλυση ήταν επίσης σημαντικό να υπάρχει πρόσβαση μόνο σε μοναδικές εγγραφές, δηλαδή αν και με ένα σύνδεσμο συσχετίζονται πολλοί ισχυρισμοί, κάθε σύνδεσμος εμφανίζεται μόνο μια φορά (συσχετίζεται μόνο με μια εγγραφή). Επειδή στη συνέχεια θα χρειαστούν και τα αρχικά δεδομένα και τα δεδομένα με μοναδικούς συνδέσμους, στο αρχείο fakeLinksDataAnalysis.csv δημιουργήθηκε ένα ακόμα φύλλο fakeLinksDataUNIQUE.

5.5 ΠΡΟΕΤΟΙΜΑΣΙΑ ΑΝΑΛΥΣΗΣ

Προκειμένου η ανάλυση να είναι όσο το δυνατόν πιο σχολαστική, πρέπει να γίνουν οι απαραίτητες επεξεργασίες δεδομένων (filtering, stemming, αντικατάσταση δεδομένων με τη βοήθεια του κώδικα στη Java κ.α.), που θα επιτρέπουν να εξάγουμε πληροφορίες.

Όλη η ανάλυση βασίζεται μόνο σε τρία χαρακτηριστικά: στους ισχυρισμούς, τις κατηγορίες και τις ημερομηνίες δημοσιεύματος – για τον λόγο αυτό στην προετοιμασία της ανάλυσης δίνεται έμφαση στην επεξεργασία αυτών των χαρακτηριστικών.

5.5.1 ΕΞΑΓΩΓΗ ΗΜΕΡΟΜΗΝΙΩΝ ΔΗΜΟΣΙΕΥΜΑΤΟΣ

Αρχικά, οι ημερομηνίες έχουν τη μορφή «dd-mm-yyyy», αλλά επειδή στη συνέχεια τα άρθρα θα φιλτράρονται βάση το έτος και θα αναλύονται ανά μήνα, θα είναι χρήσιμο να παρουσιάσουμε την ημερομηνία σε μορφή «mmmm-yyyy», δηλαδή αγνοώντας τις ημέρες. Για αυτό τον σκοπό χρησιμοποιήθηκε η συνάρτηση TEXT που μετατρέπει τη μορφή της ημερομηνίας (Εικόνα 5.5.1).

Εικόνα 5.5.1 Μετατροπή ημερομηνίας από «dd-mm-yyyy» σε «mmmm-yyyy»

Column2	Column3	Column4	Column5	Column6	Column7	Column8
Title	Date	Author	Claim	Conclusion	Category	Verdict
Νέο κρούσμα από Αλβανό...	28-1-2017	Θοδωρής Δ...	-	-	HOAX	January-2017
Ψευδοεισητημονική παραπ...	3-7-2020	Στάμος Αρχ...	Η χρήση	Η αποτελεσμ...	ΥΓΕΙΑ	ΠΑΡΑΠ/ July-2020
Καταρρίπτεται - Γιατί οι Εβρ...	14-5-2016	Σπύρος Αυγ...	-	-	HOAX	May-2016
Όχι, η Καγκελάριας Μέρκελ...	26-3-2020	Στέλιος Που...	Η Άνγκελ...	Σύμφωνα με	FAKE	ΠΑΡΑΠ/ March-2020
Η Ρόζαλιντ Πήτερσον παραβ...	12-4-2014	Θοδωρής Δ...	-	-	ΣΥΝΩ	April-2014
Παραπληροφόρηση σχετικά...	9-1-2021	Στάμος Αρχ...	Σύμφωνη	Η ανακοίνωσ...	ΥΓΕΙΑ	ΠΑΡΑΠ/ January-2021
Καταρρίπτεται - Αϊνστάιν: Ε...	22-1-2015	Γιώργος Γιά...	-	-	ΨΕΥΤ	January-2015
Θέλετε να μάθετε για τα εμ...	8-6-2017	Θάνος Σιτί...	-	-	HOAX	June-2017
Καταρρίπτεται-ΠΑΓΚΟΣΜΙΟ...	14-1-2015	Γιώργος Ζα...	-	-	ΣΥΝΩ	January-2015
Η ψευδής ιστορία τελετών π...	27-4-2020	Στέλιος Που...	Ο Μελ Γκ...	Πρόκειται για	ΣΥΝΩ	ΣΥΝΩΜ/ April-2020
Η ψευδής ιστορία τελετών π...	27-4-2020	Στέλιος Που...	ο γνωστός	Ο ηθοποιός κ...	ΣΥΝΩ	ΣΥΝΩΜ/ April-2020
Γαλλία: Παραπληροφόρηση...	19-5-2021	Στέλιος Που...	Εκατοντά	Τα εικονιζόμ...	ΛΕΙΠΗ	ΠΑΡΑΠ/ May-2021
Όχι, ο καρκίνος ΔΕΝ είναι ιά...	21-7-2019	Στάμος Αρχ...	-	-	ΥΓΕΙΑ	ΨΕΥΔΟΕ July-2019
Διαδήλωσαν μουσουλμάνοι...	14-5-2018	Ανδρόνικος...	-	-	ΠΑΡΑ	May-2018
Καταρρίπτεται - ΝΙΓΗΡΙΑΝΑ...	22-10-2015	Θοδωρής Δ...	-	-	HOAX	October-2015
Να γιατί πιάνεσαι κορόιδο...	9-10-2014	Γιώργος Γιά...	-	-	ΔΙΑΦ	October-2014
Πήγε στην κηδεία του γιου τ...	9-2-2017	Θοδωρής Δ...	-	-	HOAX	February-2017
Ξαναεμφανίστηκε η απάτη...	24-12-2015	Θοδωρής Δ...	-	-	ΑΠΑΤ	December-2015
Παραπληροφόρηση για τη θ...	3-9-2020	Ανδρόνικος...	Στη διαδ...	Η αστυνομία	ΠΑΡΑ	ΨΕΥΔΗΣ September-2020
Παραπληροφόρηση για τη θ...	3-9-2020	Ανδρόνικος...	Άρματα	Η μετακίνησ...	ΠΑΡΑ	ΨΕΥΔΗΣ September-2020
Ο Αμερικανός πρόεδρος προ...	31-1-2017	Θοδωρής Δ...	-	-	HOAX	January-2017
ΟΧΙ, επιτήδευτοι ΔΕΝ κλέβου...	27-9-2019	Στέλιος Που...	-	-	HOAX	September-2019
«Οι διασώστες της Αλ Κάιντ...	21-12-2016	Θοδωρής Δ...	-	-	HOAX	December-2016

Εν συνέχεια, για να υπολογιστεί ο αριθμός άρθρων ανά μήνα,

1. Η ίδια συνάρτηση TEXT εφαρμόστηκε στα δεδομένα στο φύλλο με μοναδικούς συνδέσμους fakeLinksDataUNIQUE.
2. Ο υπολογισμός έγινε με τη βοήθεια της συνάρτησης COUNTIF (Εικόνα 5.5.2).

Εικόνα 5.5.2 Υπολογισμός των άρθρων που δημοσιεύτηκαν ένα συγκεκριμένο μήνα (Μάιος, 2023)

AA	AB	AC	AD
Month of publication			
January-2017	May-2023	=COUNTIF(AA3:AA5027, AB3)	
July-2020	April-2023	COUNTIF(range, criteria)	
May-2016	March-2023		
March-2020	February-2023		
April-2014	January-2023		
January-2021	December-2022		
January-2015	November-2022		
June-2017	October-2022		
January-2015	September-2022		
April-2020	August-2022		
May-2021	July-2022		
July-2019	June-2022		

Προκειμένου να φιλτραριστούν τα δεδομένα ανά έτη, χρησιμοποιήθηκε η συνάρτηση $IF(YEAR(fakeLinksDataUNIQUE!\$AA\mathbf{X})=Y, fakeLinksDataUNIQUE!A3, "")$, όπου το \mathbf{X} είναι ο αριθμός της γραμμής μιας εγγραφής και το \mathbf{Y} είναι το έτος, που μας ενδιαφέρει (Εικόνα 5.5.3). Για να αποκτηθούν τα δεδομένα για την περίοδο 2020-2023 χρησιμοποιήθηκε η συνάρτηση $IF(YEAR(fakeLinksDataUNIQUE!\$AA\mathbf{X})>2019, fakeLinksDataUNIQUE!\mathbf{AX}, "")$, όπου το \mathbf{X} είναι ο αριθμός της γραμμής μιας εγγραφής (Εικόνα 5.5.4).

Εικόνα 5.5.3 Φιλτράρισμα των άρθρων ανά έτος (2020)

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	https://w	Ψευδοεπι	3-7-2020	Στάμος Αρ	Η χρήση χι	Η αποτελε	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑΡ	ΥΓΕΙΑ, ΠΑ
3												
4	https://w	Όχι, η Καγ	26-3-2020	Στέλιος Πι	Ανγκελς	Σύμφωνα	FAKE NEW	ΠΑΡΑΠΛΗ	FAKE NEW	ΠΑΡΑΠΛΗ	FAKE NEW	FAKE NE
5												
6												
7												
8												
9												
10	https://w	Η ψευδής	27-4-2020	Στέλιος Πι	Ο Μελ Γκί	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟ
11												
12												
13												
14												
15												
16												
17												
18	https://w	Παραπλη	3-9-2020	Ανδρόνικ	Στη διαδή	Η αστυνο	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												
32	https://w	Η συνταγ	11-4-2020	Σπυριδού	Η "μαγική	Είναι μία	ΨΕΥΔΟΕΠΙ	ΨΕΥΔΟΕΠΙ	ΨΕΥΔΟΕΠΙ	ΨΕΥΔΟΕΠΙ	ΨΕΥΔΟΕΠΙ	ΨΕΥΔΟΕ
33	https://w	Η ψευδής	9-4-2020	Στέλιος Πι	Σε υπόγει	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟ
34												
35												
36												
37												

Εικόνα 5.5.4 Φιλτράρισμα των άρθρων ανά περίοδο πάνω από ένα έτος (2020 - τώρα)

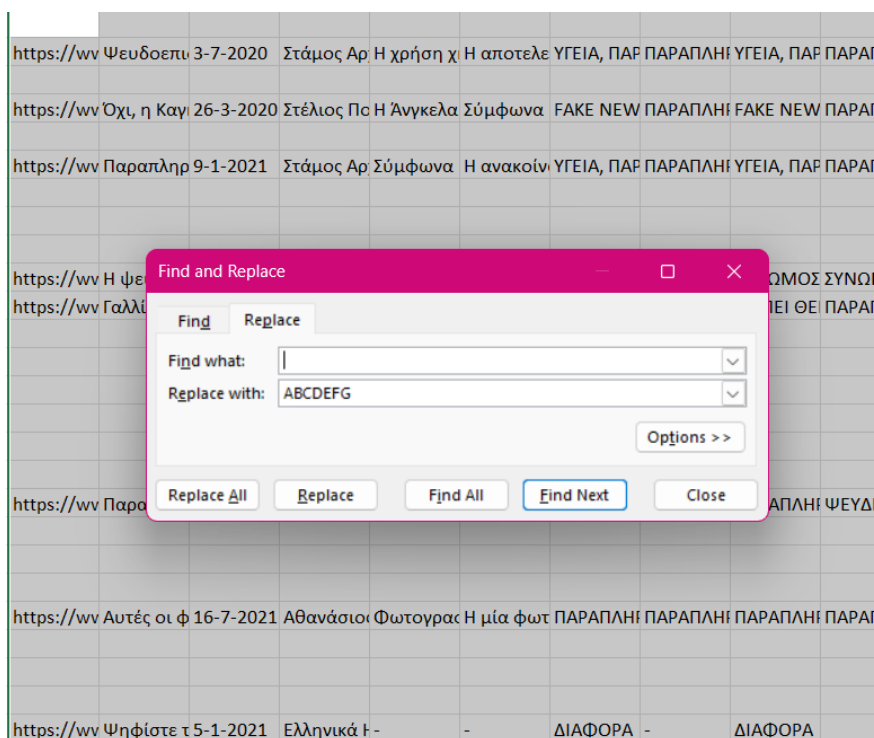
	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	https://w	Ψευδοεπι	3-7-2020	Στάμος Αρ	Η χρήση χι	Η αποτελε	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑ	
3												
4	https://w	Όχι, η Καγ	26-3-2020	Στέλιος Πρ	Η Άνγκελς	Σύμφωνα	FAKE NEW	ΠΑΡΑΠΛΗ	FAKE NEW	ΠΑΡΑΠΛΗ	FAKE NEW	FAKE NEW
5												
6	https://w	Παραπληρ	9-1-2021	Στάμος Αρ	Σύμφωνα	Η ανακοίν	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑΙ	ΠΑΡΑΠΛΗ	ΥΓΕΙΑ, ΠΑ	
7												
8												
9												
10	https://w	Η ψευδής	27-4-2020	Στέλιος Πρ	Ο Μελ Γκί	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟ
11	https://w	Γαλλία: Πρ	19-5-2021	Στέλιος Πρ	Εκατοντά	Τα εικονιζ	ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗ	ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ
12												
13												
14												
15												
16												
17												
18	https://w	Παραπληρ	3-9-2020	Ανδρόνικ	Στη διαδή	Η αστυνο	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ
19												
20												
21												
22	https://w	Αυτές οι φ	16-7-2021	Αθανάσιο	Φωτογρα	Η μία φωτ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ
23												
24												
25												
26	https://w	Ψηφίστε τ	5-1-2021	Ελληνικά	Ι-	-	ΔΙΑΦΟΡΑ	-	ΔΙΑΦΟΡΑ	ΔΙΑΦΟΡΑ	ΔΙΑΦΟΡΑ	ΔΙΑΦΟΡΑ
27												
28												
29	https://w	Στη φωτογ	21-6-2021	Στέλιος Πρ	Σε φωτογ	Στην υπό	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ
30	https://w	Φωτογρα	8-3-2023	Αθανάσιο	Φωτογρα	Η φωτογρ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ	ΛΕΙΠΕΙ ΘΕ
31	https://w	Όχι, ΔΕΝ σ	4-7-2021	Ανδρόνικ	Οι ΗΠΑ σ	Το σχετικ	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ	ΙΣ
32	https://w	Η συνταγ	11-4-2020	Σπυριδού	Η "μαγική	Είναι μία	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ
33	https://w	Η ψευδής	9-4-2020	Στέλιος Πρ	Σε υπόγει	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ

5.5.1.1 ΔΙΑΓΡΑΦΗ ΚΕΝΩΝ ΓΡΑΜΜΩΝ

Για να διαγραφούν οι κενές γραμμές η απευθείας χρήση του Editing → Find & Select → Go to Special → Blanks, δε θα δώσει κανένα αποτέλεσμα, διότι τα κελιά δεν είναι στην ουσία κενά. Η διαδικασία διαγραφής κενών γραμμών, που χρησιμοποιείται πολλές φορές κατά την επεξεργασία αυτών των δεδομένων είναι η εξής:

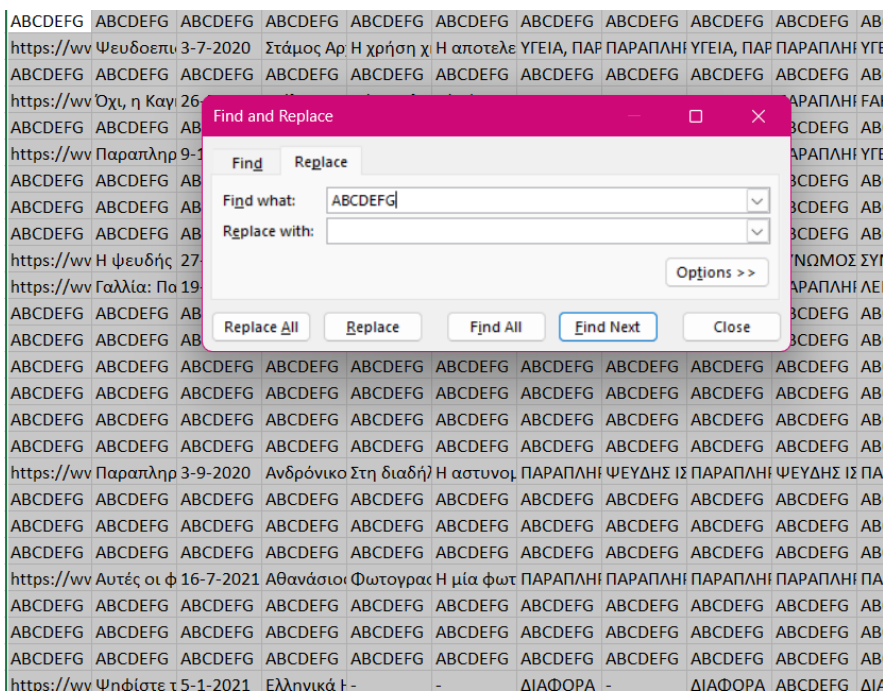
1. Όλες οι τιμές των κελιών αντιγράφονται (Copy → Paste Values).
2. Όλα τα κελιά επιλέγονται και γίνεται η αντικατάσταση του κενού (όχι του SPACE) με ένα κείμενο, πχ ABCDEFG (Editing → Find & Select → Replace → Replace All) (Εικόνα 5.5.5).

Εικόνα 5.5.5 Διαγραφή κενών γραμμών (1)



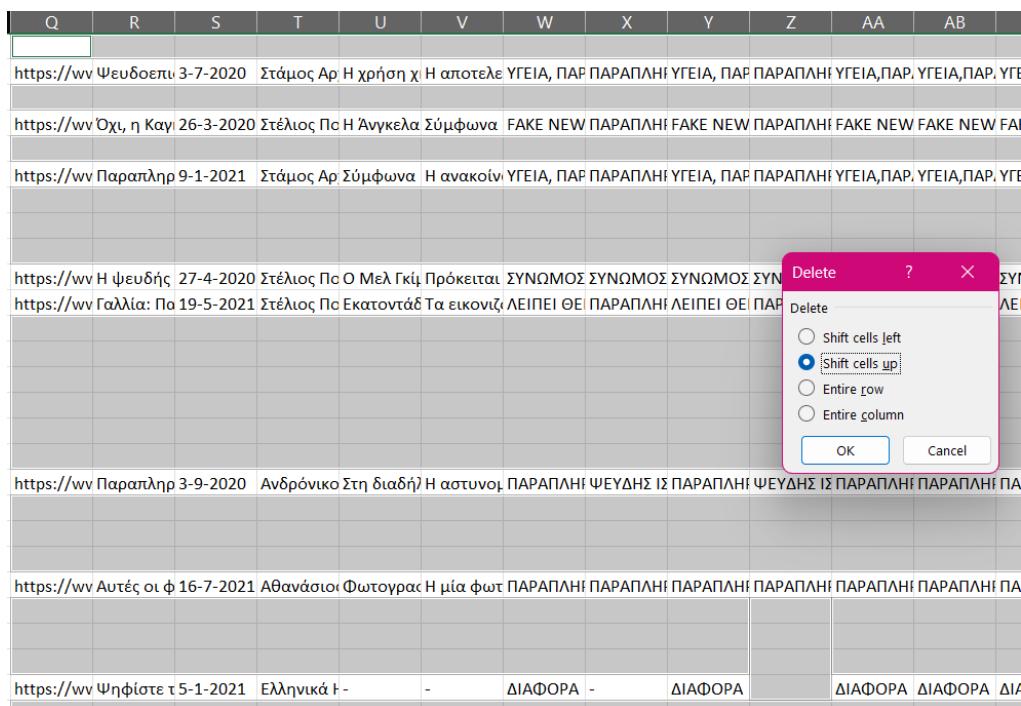
3. Αντικαθιστούμε το ABCDEFG με κενό (όχι με SPACE) (Εικόνα 5.5.6).

Εικόνα 5.5.6 Διαγραφή κενών γραμμών (2)



4. Επιλέγουμε όλα τα κενά (Editing → Find & Select → Go to Special → Blanks) και με το δεξί κλικ τα διαγράφουμε (Delete → Shift cells up) (Εικόνα 5.5.7).

Εικόνα 5.5.7 Διαγραφή κενών γραμμών (3)



Το αποτέλεσμα επεξεργασίας παρουσιάζεται στην Εικόνα 5.5.8.

Εικόνα 5.5.8 Αποτέλεσμα διαγραφής κενών γραμμών

https://wn Ψευδοεπι 3-7-2020	Στάμος Αρ	Η χρήση χι	Η αποτελε	ΥΓΕΙΑ, ΠΑΡ	ΠΑΡΑΠΛΗΓ	ΥΓΕΙΑ, ΠΑΡ	ΠΑΡΑΠΛΗΓ	ΥΓΕΙΑ,ΠΑΡ	ΥΓΕΙΑ,ΠΑΡ	Υ
https://wn Όχι, η Καγί 26-3-2020	Στέλιος Πο	Η Άνγκελα	Σύμφωνα	FAKE NEW	ΠΑΡΑΠΛΗΓ	FAKE NEW	ΠΑΡΑΠΛΗΓ	FAKE NEW	FAKE NEW	F
https://wn Παραπληρ 9-1-2021	Στάμος Αρ	Σύμφωνα	Η ανακοίν	ΥΓΕΙΑ, ΠΑΡ	ΠΑΡΑΠΛΗΓ	ΥΓΕΙΑ, ΠΑΡ	ΠΑΡΑΠΛΗΓ	ΥΓΕΙΑ,ΠΑΡ	ΥΓΕΙΑ,ΠΑΡ	Υ
https://wn Η ψευδής 27-4-2020	Στέλιος Πο	Ο Μελ Γκί	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	Σ
https://wn Γαλλία: Πα 19-5-2021	Στέλιος Πο	Εκατοντάδ	Τα εικονιζ	ΛΕΙΠΕΙ ΘΕΙ	ΠΑΡΑΠΛΗΓ	ΛΕΙΠΕΙ ΘΕΙ	ΠΑΡΑΠΛΗΓ	ΛΕΙΠΕΙ ΘΕΙ	ΛΕΙΠΕΙ ΘΕΙ	Λ
https://wn Παραπληρ 3-9-2020	Ανδρόνικο	Στη διαδήλ	Η αστυνο	ΠΑΡΑΠΛΗΓ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗΓ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗΓ
https://wn Αυτές οι φ 16-7-2021	Αθανάσιοι	Φωτογρας	Η μία φωτ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑ
https://wn Ψηφίστε τ 5-1-2021	Ελληνικά Ι-	-	-	ΔΙΑΦΟΡΑ	-	ΔΙΑΦΟΡΑ	ΠΑΡΑΠΛΗΓ	ΔΙΑΦΟΡΑ	ΔΙΑΦΟΡΑ	ΔΙ
https://wn Στη φωτογ 21-6-2021	Στέλιος Πο	Σε φωτογρ	Στην υπό ε	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	ΛΕΙΠΕΙ ΘΕΙ	ΠΑΡΑΠΛΗΓ	ΠΑΡΑΠΛΗΓ	Γ
https://wn Φωτογρας 8-3-2023	Αθανάσιοι	Φωτογρας	Η φωτογρ	ΛΕΙΠΕΙ ΘΕΙ	ΛΕΙΠΕΙ ΘΕΙ	ΛΕΙΠΕΙ ΘΕΙ	ΨΕΥΔΗΣ	ΙΣ	ΛΕΙΠΕΙ ΘΕΙ	Λ
https://wn Όχι, ΔΕΝ σ 4-7-2021	Ανδρόνικο	Οι ΗΠΑ σ	Το σχετικ	ΠΑΡΑΠΛΗΓ	ΨΕΥΔΗΣ	ΙΣ	ΠΑΡΑΠΛΗΓ	ΨΕΥΔΟΕΠ	ΠΑΡΑΠΛΗΓ	Γ
https://wn Η συνταγή 11-4-2020	Σπυριδούλ	Η "μαγική"	Είναι μία	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	ΣΥΝΩΜΟΣ	ΨΕΥΔΟΕΠ	ΨΕΥΔΟΕΠ	Ψ
https://wn Η ψευδής 9-4-2020	Στέλιος Πο	Σε υπόγει	Πρόκειται	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	ΤΡΟΠΟΠΟ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣ	Σ
https://wn Όχι, ρωσικ 7-3-2023	Αλέξανδρος	Ως ένδειξη	Οι εικόνες	ΤΡΟΠΟΠΟ	ΤΡΟΠΟΠΟ	ΤΡΟΠΟΠΟ	ΨΕΥΔΗΣ	ΙΣ	ΤΡΟΠΟΠΟ	Τ

5.5.2 ΕΞΑΓΩΓΗ ΚΑΤΗΓΟΡΙΩΝ

Επειδή κάθε άρθρο μπορεί να έχει διαφορετική δομή (Στην Ενότητα 5.3.2.1 αυτό το γεγονός περιγράφεται με λεπτομέρεια.), ήταν σημαντικό να αξιοποιηθούν όλες οι μορφές

κατηγοριοποίησης. Ωστόσο, επειδή υπάρχουν άρθρα, όπου και σε μορφή κειμένου, και σε μορφή εικόνας οι κατηγορίες συμφωνούν, ήταν σημαντικό να μην υπολογιστούν αυτές οι κατηγορίες δύο φορές. Για παράδειγμα, το άρθρο <https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/> έχει την κατηγορία «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ» σε μορφή κειμένου και τις κατηγορίες «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ» και «ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ» σε μορφή εικόνας. Στη συνέχεια, πρέπει να υπολογιστεί η συχνότητα κάθε κατηγορίας παραπληροφόρησης και αν δεν επεξεργαστούν αυτά τα δεδομένα με ένα σωστό τρόπο, αντί για να υπολογιστεί ένα για την κατηγορία «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ» θα υπολογιστεί δυο, άρα τα αποτελέσματα δε θα είναι έγκυρα.

Επίσης, έπρεπε να ληφθούν υπόψιν και τα άρθρα με πολλές εγγραφές, γιατί αν, για παράδειγμα, με ένα άρθρο συσχετίζονται πέντε εγγραφές (πέντε ισχυρισμοί), όλες οι κατηγορίες του άρθρου επίσης είναι αποθηκευμένες πέντε φορές. Ακόμα πιο άκυρα αποτέλεσμα θα υπάρχουν αν ένα άρθρο έχει πολλούς ισχυρισμούς και οι κατηγορίες και σε μορφή εικόνας και σε μορφή κειμένου είναι οι ίδιες. Ένα από τα παραδείγματα παρουσιάζεται στην Εικόνα 5.5.9. Σε αυτό το παράδειγμα, η κατηγορία «ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ» θα υπολογιστεί δέκα φορές αντί για μία και η κατηγορία «ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ» επίσης θα υπολογιστεί δέκα φορές αντί για μία.

Ωστόσο, μετά τη σωστή επεξεργασία αυτό το άρθρο θα έχει μόνο μια εγγραφή που περιλαμβάνει αυτές τις δυο κατηγορίες χωρίς επαναλήψεις (Εικόνα 5.5.10).

Εικόνα 5.5.9 Άρθρο με πολλούς ισχυρισμούς πριν από την επεξεργασία των κατηγοριών του

Link	Title	Date	Author	Claim	Conclusion	Category	Verdict
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Ο Bill Gates ΔΕΝ παρευρέθηκ	26-11	Γιάννη	Ο Bill Gat	Ο Gates δεν π	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Εφαρμογή που βρίσκει σε δε	19-3	Στάμος	-	-	HOAXES, FAKE NEWS	-
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Καταρρίπτεται - ΠΙΛΟΤΟΣ ΕΛ	20-7	Θοδωρ	-	-	HOAXES, FAKE NEWS	-
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Ψευδής είδηση με υποτιθέμε	25-4	Θοδωρ	Ανακοίν	Η προειδοποι	HOAXES	ΨΕΥΔΕΣ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Καταρρίπτεται - Φορτηγάκι Σ	8-11	Γιώργα	-	-	FAKE NEWS	-
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Όχι, βίντεο που κυκλοφορεί	5-8-2	Σπυριδ	Ρώσος χρ	Σύμφωνα με	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Στο Τόκιο ανοίγει εστιατόριο	8-12	Θάνος	-	-	HOAXES	-
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Καταρρίπτεται-Θεία Δίκη; Μ	21-1	Γιώργα	-	-	HOAXES	-
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Συνωμοσιολογικό βίντεο ανα	5-11	Ανδρόν	Ο πρώην	Ο κ. Νετανιάχ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Συνωμοσιολογικό βίντεο ανα	5-11	Ανδρόν	Οι Μακκ	Ιστορικά, οι Μ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Συνωμοσιολογικό βίντεο ανα	5-11	Ανδρόν	Η ελληνικ	Η κρίση ήταν	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Συνωμοσιολογικό βίντεο ανα	5-11	Ανδρόν	Το 1974,	Στην πραγματ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Συνωμοσιολογικό βίντεο ανα	5-11	Ανδρόν	Η επιχείρ	Οι οδηγίες για	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Οι ασυμπτωματικοί μπορούν	24-1	Ανδρόν	Νέα μελέ	Μια διαφορε	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΨΕΥΔΟΕΠΙΣΤΗΜΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΨΕΥΔΟΕΠΙΣΤΗΜΗ
https://www.ellinikahoaxes.gr/2022/04/05/pfizer-adverse-reactions-due-to-covid-vaccine-long-list-hoax/	Καταρρίπτεται-ΣΟΚ!!!ΑΜΦΙΠ	14-9	Γιώργα	-	-	HOAXES	-

Εικόνα 5.5.10 Άρθρο με πολλούς ισχυρισμούς μετά την επεξεργασία των κατηγοριών του

1	Column1	Column2	Column3	Column100	Column101	Column102
2	Link	Title	Date	Sp	Category	Verdict NO DUPLICATES
23	https://www.ell	Άντε πάλι τα ίδια.	18-5-2013	HO	HOAXES	
24	https://www.ell	Κουρκουμίνη, μύθοι και αλήθειες	22-9-2017	ΥΓ	ΥΓΕΙΑ	
25	https://www.ell	Καταρρίπτεται - Σάρα Πέυλιν - Η πρώτη ατομ	22-1-2016	FA	HOAXES	FAKE NEWS
26	https://www.ell	Κρύβουν την αλήθεια για τις χειρόγραφες επι	6-7-2019	Ψ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΨΕΥΔΕΣ
27	https://www.ell	Ο Πούτιν ΔΕΝ διέταξε την καταστροφή των ει	8-3-2023	Π	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	
28	https://www.ell	Ο Bill Gates ΔΕΝ παρευρέθηκε στη σύνοδο τω	26-11-2022	Π	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	
29	https://www.ell	Εφαρμογή που βρίσκει σε δευτερόλεπτα κάθε	19-3-2017	FA	HOAXES	FAKE NEWS
30	https://www.ell	Καταρρίπτεται - ΠΙΛΟΤΟΣ ΕΛΛΗΝΙΚΟΥ F-16 ΚΑ	20-7-2015	FA	HOAXES	FAKE NEWS
31	https://www.ell	Ψευδής είδηση με υποτιθέμενο μήνυμα της α	25-4-2019	Ψ	HOAXES	ΨΕΥΔΕΣ
32	https://www.ell	Καταρρίπτεται - Φορηγάκι Σταματάει Αεροπ	8-11-2014	FA	FAKE NEWS	
33	https://www.ell	Όχι, βίντεο που κυκλοφορεί στα social media	5-8-2021	Π	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	
34	https://www.ell	Στο Τόκιο ανοίγει εστιατόριο ανθρωπίνης σά	8-12-2017	HO	HOAXES	
35	https://www.ell	Καταρρίπτεται-Θεία Δίκη; Μόλις είχαν ληστέ	21-12-2014	HO	HOAXES	
36	https://www.ell	Συνωμοσιολογικό βίντεο αναπαράγει αντιση	5-11-2022	Σ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
37	https://www.ell	Οι ασυμπτωματικοί μπορούν να μεταδώσουν	24-11-2020	Ψ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΨΕΥΔΟΕΠΙΣΤΗΜΗ

Προκειμένου να λυθεί αυτό το θέμα, έχουν πραγματοποιηθεί τα παρακάτω βήματα

1. Στο φύλλο fakeLinksDataUNIQUE πάμε στη δημιουργούμε μια καινούργια στήλη I (Category CLEANED), που θα περιέχει τη συνάρτηση SUBSTITUTE(SUBSTITUTE(fakeLinksDataUNIQUE!\$GX,"-",""),"-",""), όπου το X είναι η γραμμή μιας εγγραφής. Αυτή η συνάρτηση αντικαθιστά την παύλα με το κενό (Εικόνα 5.5.11).

Εικόνα 5.5.11 Επεξεργασία των κατηγοριών (1)

1	Column7	Column82	Column83
2	Link	Category CLEANED	Verdict CLEANED
3	https://www.ell	HOAXES, FAKE NEWS	HOAXES, FAKE NEWS
4	https://www.ell	ΥΓΕΙΑ, ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
5	https://www.ell	HOAXES	HOAXES
6	https://www.ell	FAKE NEWS	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
7	https://www.ell	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
8	https://www.ell	ΥΓΕΙΑ, ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
9	https://www.ell	ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ	
10	https://www.ell	HOAXES, FAKE NEWS, ΔΙΑΦΟΡΑ	
11	https://www.ell	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ, ΔΙΑΦΟΡΑ	
12	https://www.ell	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
13	https://www.ell	ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ, ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
14	https://www.ell	ΨΕΥΔΟΕΠΙΣΤΗΜΗ	ΨΕΥΔΟΕΠΙΣΤΗΜΗ
15	https://www.ell	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	
16	https://www.ell	HOAXES, ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	
17	https://www.ell	ΔΙΑΦΟΡΑ	
18	https://www.ell	HOAXES, FAKE NEWS	
19	https://www.ell	ΑΠΑΘΗ, LIKE FARMING	
20	https://www.ell	ΨΕΥΔΟΕΠΙΣΤΗΜΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣ

2. Στο φύλλο fakeLinksDataUNIQUE δημιουργούμε μια καινούργια στήλη J (Verdict CLEANED), που θα περιέχει τη συνάρτηση `SUBSTITUTE(SUBSTITUTE(fakeLinksDataUNIQUE!$HX,"-",""),"-","")`, όπου το **X** είναι η γραμμή μιας εγγραφής. Αυτή η συνάρτηση αντικαθιστά την παύλα με το κενό (Εικόνα 5.5.12).

Εικόνα 5.5.12 Επεξεργασία των κατηγοριών (2)

	A	B	C	D	E	F	G	H	I	J	
1	Col	Col	Col	Col	Col	Col	Column7	Column8	Col	Column83	Column9
2	Lin	Tit	Da	Au	Cl	Co	Category	Verdict	Cate	Verdict CLEANED	Category Verdi
3	htt	Nέ	28	Θο	-	-	HOAXES,	-	HOA		HOAXES,FAKE I
4	htt	Ψε	3-7	Στε	Η	Η	ΥΓΕΙΑ, ΠΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ	ΥΓΕΙ	ΠΑΡΑΠΛΗΡΟΦΟ	ΥΓΕΙΑ,ΠΑΡΑΠΛ
5	htt	Κα	14	Σπ	-	-	HOAXES	-	HOA		HOAXES
6	htt	Οχ	26	Στε	Η	Σύ	FAKE NEV	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	FAK	ΠΑΡΑΠΛΗΡΟΦΟ	FAKE NEWS,ΠΑ
7	htt	Η	12	Θο	-	-	ΣΥΝΩΜΟΣ	-	ΣΥΝ		ΣΥΝΩΜΟΣΙΟΛΟ
8	htt	Πα	9-1	Στε	Σύ	Η	ΥΓΕΙΑ, ΠΑ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΥΓΕΙ	ΠΑΡΑΠΛΗΡΟΦΟ	ΥΓΕΙΑ,ΠΑΡΑΠΛ
9	htt	Κα	22	Γιά	-	-	ΨΕΥΤΙΚΑ	-	ΨΕΥ		ΨΕΥΤΙΚΑ ΡΗΤΑ
10	htt	Θέ	8-6	Θά	-	-	HOAXES,	-	HOA		HOAXES,FAKE I
11	htt	Κα	14	Γιά	-	-	ΣΥΝΩΜΟΣ	-	ΣΥΝ		ΣΥΝΩΜΟΣΙΟΛΟ
12	htt	Η	27	Στε	Ο	Πρ	ΣΥΝΩΜΟΣ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝ	ΣΥΝΩΜΟΣΙΟΛΟΓ	ΣΥΝΩΜΟΣΙΟΛΟ
13	htt	Γα	19	Στε	Εκ	Τα	ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙ	ΛΕΙΠ	ΠΑΡΑΠΛΗΡΟΦΟ	ΛΕΙΠΕΙ ΘΕΜΑΤ
14	htt	Οχ	21	Στε	-	-	ΥΓΕΙΑ	ΨΕΥΔΟΕΠΙΣΤΗΜΗ	ΥΓΕΙ	ΨΕΥΔΟΕΠΙΣΤΗΜ	ΥΓΕΙΑ,ΨΕΥΔΟΕΙ
15	htt	Δι	14	Αν	-	-	ΠΑΡΑΠΛΗ	-	ΠΑΡ		ΠΑΡΑΠΛΗΡΟΦ
16	htt	Κα	22	Θο	-	-	HOAXES,	-	HOA		HOAXES,ΣΥΝΩ
17	htt	Να	9-1	Γιά	-	-	ΔΙΑΦΟΡΑ	-	ΔΙΑ		ΔΙΑΦΟΡΑ
18	htt	Πή	9-2	Θο	-	-	HOAXES,	-	HOA		HOAXES,FAKE I
19	htt	Ξα	24	Θο	-	-	ΑΠΑΤΗ, L	-	ΑΠΑ		ΑΠΑΤΗ,LIKE FA
20	htt	Πα	3-9	Αν	Στε	Η	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ	ΠΑΡ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣ	ΠΑΡΑΠΛΗΡΟΦ

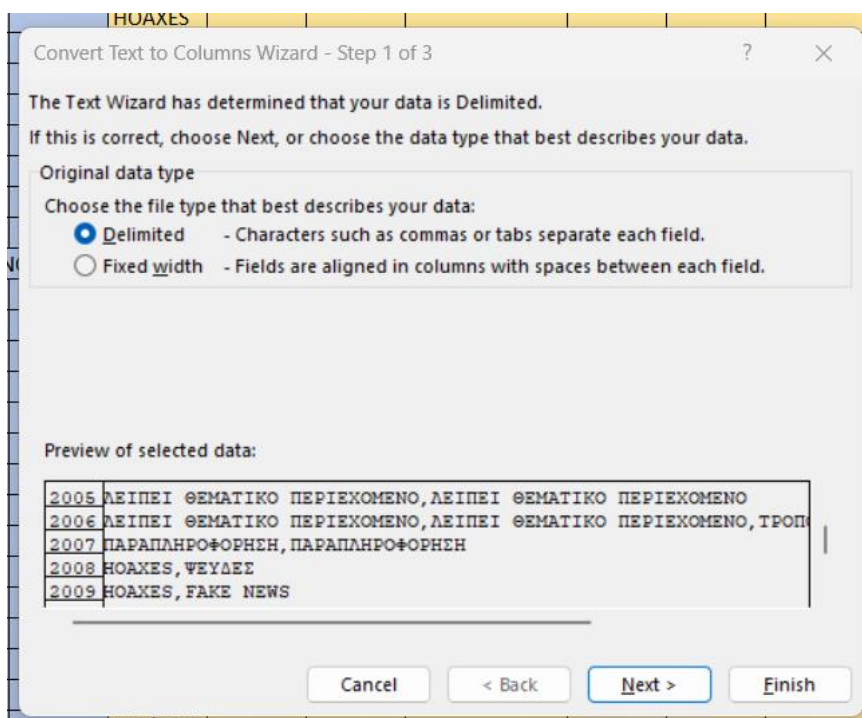
3. Ενώνουμε αυτές τις δυο στήλες (I και J) με τη βοήθεια της συνάρτησης `SUBSTITUTE(IFS(IX="",JX, JX="",IX, TRUE,CONCATENATE(fakeLinksDataUNIQUE!$IX,"",fakeLinksDataUNIQUE!$JX)),"",",")`, όπου το **X** είναι η γραμμή μιας εγγραφής, στη στήλη K (Category Verdict JOINED). Παρατηρούμε, ότι σε κάποιες στήλες υπάρχουν επαναλήψεις, πχ K4 (Εικόνα 5.5.13).

Εικόνα 5.5.13 Επεξεργασία των κατηγοριών (3)

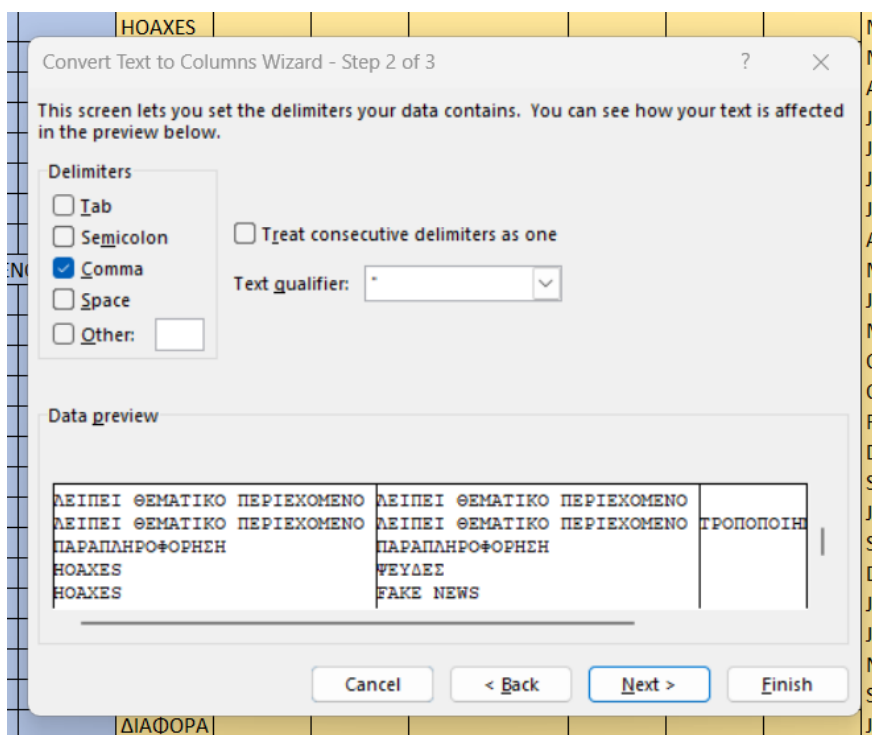
Column7	Column8	Column9
Category	Verdict	Category Verdict JOINED
HOAXES, FAKE NEWS	-	HOAXES,FAKE NEWS
ΥΓΕΙΑ, ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ	ΥΓΕΙΑ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ,ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ
HOAXES	-	HOAXES
FAKE NEWS	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	FAKE NEWS,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	-	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
ΥΓΕΙΑ, ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΥΓΕΙΑ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ
ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ	-	ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ
HOAXES, FAKE NEWS, ΔΙΑΦΟΡΑ	-	HOAXES,FAKE NEWS,ΔΙΑΦΟΡΑ
ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ, ΔΙΑΦΟΡΑ	-	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ,ΔΙΑΦΟΡΑ
ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ,ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ
ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ, ΛΕΙΠΕΙ	ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ,ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ,ΛΕΙΠ

- Επειδή οι επαναλήψεις είναι στην ίδια κελί, το Remove Duplicates δε θα βοηθήσει, η διαδικασία είναι πιο περιπλοκή. Πρώτα, στην επόμενη στήλη L (Category Verdict JOINED – COPIED VALUES) αντιγράφουμε όλες τις τιμές της στήλης K. Τώρα επιλέγουμε αυτή τη στήλη και πατάμε Data → Data Tools → Text to Columns → Delimited → Next → Delimiters: Comma → Next → Destination: =\$M:\$S → Finish (Δείτε τις Εικόνες 5.5.14 – 5.5.16 με τα βήματα.).

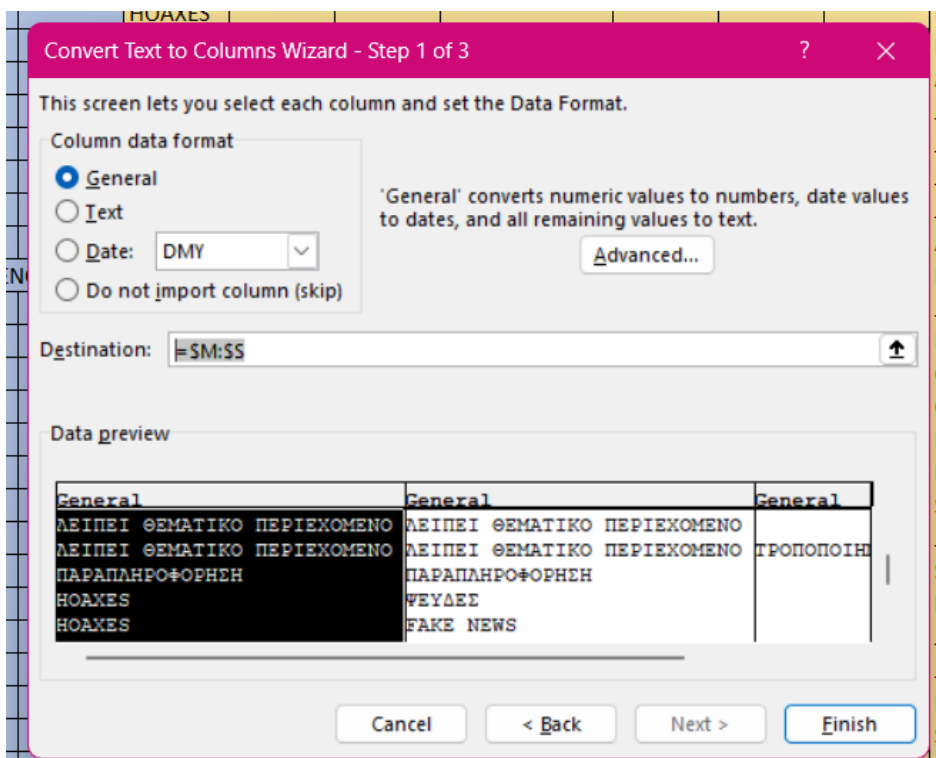
Εικόνα 5.5.14 Επεξεργασία των κατηγοριών (4)



Εικόνα 5.5.15 Επεξεργασία των κατηγοριών (5)



Εικόνα 5.5.16 Επεξεργασία των κατηγοριών (6)



5. Τέλος, για να διαγραφούν οι επαναλήψεις εφαρμόζεται η συνάρτηση `SUBSTITUTE(IF(COUNTIF($ZX:ZX,ZX)>1,"",ZX),0,"")` για όλες τις γραμμές των στηλών M:S, όπου το X είναι η γραμμή της εγγραφής και το Z είναι η στήλη της εγγραφής, που παίρνει τιμές από M μέχρι S (Εικόνα 5.5.17).

Εικόνα 5.5.17 Επεξεργασία των κατηγοριών (7)

	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Column107	Column10	Column10	Column11	Column11	Column11	Column11	Column10	Column10	Column10	Column103	Column10	Column10	Column10
2	Split by comma							Category	Verdict	NO DUPLICATES				
3	HOAXES	FAKE NEWS						HOAXES	FAKE NEW					
4	ΥΓΕΙΑ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ				ΥΓΕΙΑ	ΠΑΡΑΠΛΗ		ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜ			
5	HOAXES							HOAXES						
6	FAKE NEWS	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ						FAKE NEW	ΠΑΡΑΠΛΗ					
7	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ							ΣΥΝΩΜΟΣ						
8	ΥΓΕΙΑ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ					ΥΓΕΙΑ	ΠΑΡΑΠΛΗ					
9	ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ							ΨΕΥΤΙΚΑ						
10	HOAXES	FAKE NEW	ΔΙΑΦΟΡΑ					HOAXES	FAKE NEW	ΔΙΑΦΟΡΑ				
11	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΔΙΑΦΟΡΑ						ΣΥΝΩΜΟΣ	ΔΙΑΦΟΡΑ					
12	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ						ΣΥΝΩΜΟΣ						
13	ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜ	ΠΑΡΑΠΛΗ	ΠΑΡΑΠΛΗ	ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ				ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗ					
14	ΥΓΕΙΑ	ΨΕΥΔΟΕΠΙΣΤΗΜΗ						ΥΓΕΙΑ	ΨΕΥΔΟΕΠ					
15	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ							ΠΑΡΑΠΛΗ						
16	HOAXES	ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ						HOAXES	ΣΥΝΩΜΟΣ					
17	ΔΙΑΦΟΡΑ							ΔΙΑΦΟΡΑ						
18	HOAXES	FAKE NEWS						HOAXES	FAKE NEW					
19	ΑΠΑΘΗ	LIKE FARMING						ΑΠΑΘΗ	LIKE FARM					
20	ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ	ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ						ΠΑΡΑΠΛΗ	ΨΕΥΔΗΣ ΙΣ					
21	HOAXES	FAKE NEW	ΣΥΝΩΜΟΣ	ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ				HOAXES	FAKE NEW	ΣΥΝΩΜΟΣ	ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡ			

Προκειμένου να υπολογιστεί ο αριθμός άρθρων ανά κατηγορία, χρησιμοποιήθηκε η συνάρτηση COUNTIF (Εικόνα 5.5.18).

Εικόνα 5.5.18 Επεξεργασία των κατηγοριών (8)

T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
Column10	Column10	Column10	Column103	Column10	Column10	Column10	Month of publication					
Category	Verdict	NO DUPLICATES										
HOAXES	FAKE NEW						January-2017		HOAXES	=COUNTIF(T3:Z5027,AC3)		
ΥΓΕΙΑ	ΠΑΡΑΠΛΗ		ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜ				July-2020		ΠΑΡΑΠΛΗ	COUNTIF(range, criteria)		
HOAXES							May-2016		FAKE NEWS			
FAKE NEW	ΠΑΡΑΠΛΗ						March-2020		ΔΙΑΦΟΡΑ			
ΣΥΝΩΜΟΣ							April-2014		ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ			
ΥΓΕΙΑ	ΠΑΡΑΠΛΗ						January-2021		ΨΕΥΔΕΣ			
ΨΕΥΤΙΚΑ							January-2015		ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ			
HOAXES	FAKE NEW	ΔΙΑΦΟΡΑ					June-2017		ΥΓΕΙΑ			
ΣΥΝΩΜΟΣ	ΔΙΑΦΟΡΑ						January-2015		ΨΕΥΔΟΕΠΙΣΤΗΜΗ			
ΣΥΝΩΜΟΣ							April-2020		ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ			
ΛΕΙΠΕΙ ΘΕ	ΠΑΡΑΠΛΗ						May-2021		ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ			
ΥΓΕΙΑ	ΨΕΥΔΟΕΠ						July-2019		ΑΠΑΘΗ			
ΠΑΡΑΠΛΗ							May-2018		ΠΡΑΓΜΑΤΙΚΕΣ ΙΣΤΟΡΙΕΣ			
HOAXES	ΣΥΝΩΜΟΣ						October-2015		LIKE FARMING			
ΔΙΑΦΟΡΑ							October-2014		ΜΙΞΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ			
HOAXES	FAKE NEW						February-2017		ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ			

Προκειμένου να φιλτραριστούν τα δεδομένα ανά κατηγορία, χρησιμοποιήθηκε η συνάρτηση =IF(AND(ISBLANK(fakeLinksDataUNIQUE!AX)=FALSE,COUNTIF(fakeLinksDataUNIQUE!\$TX:\$ZX,F)>0),fakeLinksDataUNIQUE!AX,""), όπου το X είναι ο αριθμός της γραμμής μιας εγγραφής, το A είναι η στήλη του χαρακτηριστικού, που παίρνει τιμές από A μέχρι Z και το F είναι η κατηγορία που μας ενδιαφέρει (με κεφάλαια και σε εισαγωγικά, πχ «HOAXES» (Εικόνα 5.5.19).

Εικόνα 5.5.19 Επεξεργασία των κατηγοριών (9)

	A	B	C	D	E	F	T	U	V	W	X	Y	Z	AA	AB	AC
1	https://www.ellinikahoaxes.gr/20	Νέο κρούσμα από Αλβανό στον Ει 28-1 Θεοδ -	-	-	-	-	HOAXES	FAKE NEW								
2																
3	https://www.ellinikahoaxes.gr/20	Καταρρίπτεται - Γιατί οι Εβραίοι δ 14-5 Σπύι -	-	-	-	-	HOAXES									
4																
5																
6																
7																
8	https://www.ellinikahoaxes.gr/20	Θέλετε να μάθετε για τα εμβόλια; 8-6- Θάνι -	-	-	-	-	HOAXES	FAKE NEW ΔΙΑΦΟΡΑ								
9																
10																
11																
12																
13																
14	https://www.ellinikahoaxes.gr/20	Καταρρίπτεται - ΝΙΓΗΡΙΑΝΑ ΚΑΙ ΙΝ 22-1 Θεοδ -	-	-	-	-	HOAXES	ΣΥΝΩΜΟΣ								
15																
16	https://www.ellinikahoaxes.gr/20	Πήγε στην κηδεία του γιου της ή τ 9-2- Θεοδ -	-	-	-	-	HOAXES	FAKE NEW								
17																
18																
19	https://www.ellinikahoaxes.gr/20	Ο Αμερικανός πρόεδρος προειδοπ 31-1 Θεοδ -	-	-	-	-	HOAXES	FAKE NEW ΣΥΝΩΜΟΣ ΜΥΘΟΙ ΚΑ								
20	https://www.ellinikahoaxes.gr/20	ΟΧΙ, επιτήδειοι ΔΕΝ κλέβουν οχήμ 27-9 Στέλ -	-	-	-	-	HOAXES									
21	https://www.ellinikahoaxes.gr/20	«Οι διασώστες της Αλ Κάιντα έσω 21-1 Θεοδ -	-	-	-	-	HOAXES	FAKE NEW								
22																
23	https://www.ellinikahoaxes.gr/20	Καταρρίπτεται - ΣΑΛΟΣ- ΒΡΕΘΗΚΕ 9-6- Θεοδ -	-	-	-	-	HOAXES	FAKE NEW								
24	https://www.ellinikahoaxes.gr/20	Ανακοίνωση για το φάκελο ψεκας 20-1 Θεοδ -	-	-	-	-	HOAXES	ΣΥΝΩΜΟΣ								
25																
26																

5.5.3 ΕΞΑΓΩΓΗ ΛΕΞΕΩΝ – ΚΛΕΙΔΙΩΝ ΑΠΟ ΙΣΧΥΡΙΣΜΟΥΣ

Η επεξεργασία των ισχυρισμών για την περαιτέρω ανάλυση ήταν η πιο χρονοβόρα διαδικασία διότι

- Χρησιμοποιούνται όλες οι εγγραφές (το φύλλο fakeLinksDataRAW).
- Οι ισχυρισμοί αποτελούνται από πολλές λέξεις με μέσο όρο 9 λέξεις και ο μεγαλύτερος ισχυρισμός να αποτελείται από 94 λέξεις
- Υπάρχουν πολλά stop-words, πχ «με», «από», «αυτό» κ.α., που δημιουργούν θόρυβο στα αποτελέσματα.
- Επειδή στην ελληνική γλώσσα τα ρήματα, ουσιαστικά και επίθετα κλίνονται, για τα πιο έγκυρα αποτελέσματα πρέπει να γίνει stemming, που σε αυτή την εργασία αποτελούσε από τα τρία στάδια: η χειρωνακτική ανάλυση των λέξεων, η συγγραφή του

κώδικα στη Java για την αντικατάσταση των λέξεων και επεξεργασία των λέξεων μετά την αλλαγή τους με τη βοήθεια της Java.

Ο λόγος όλης της επεξεργασίας των ισχυρισμών είναι η περαιτέρω ανάλυση των τάσεων των ψευδών ειδήσεων βάσει τις λέξεις-κλειδιά (αγγλ. keywords), που θα εξαχθούν από τους ισχυρισμούς.

Πριν ξεκινήσει η επεξεργασία των ισχυρισμών, έχει δημιουργηθεί ένα φύλλο Stopwords που περιλαμβάνει τις λέξεις στα ελληνικά και τα αγγλικά που χρησιμοποιούνται πολύ συχνά και δε δίνουν καμία πληροφορία όσο αφορά τις τάσεις στις ψευδείς ειδήσεις (Εικόνα 5.5.20).

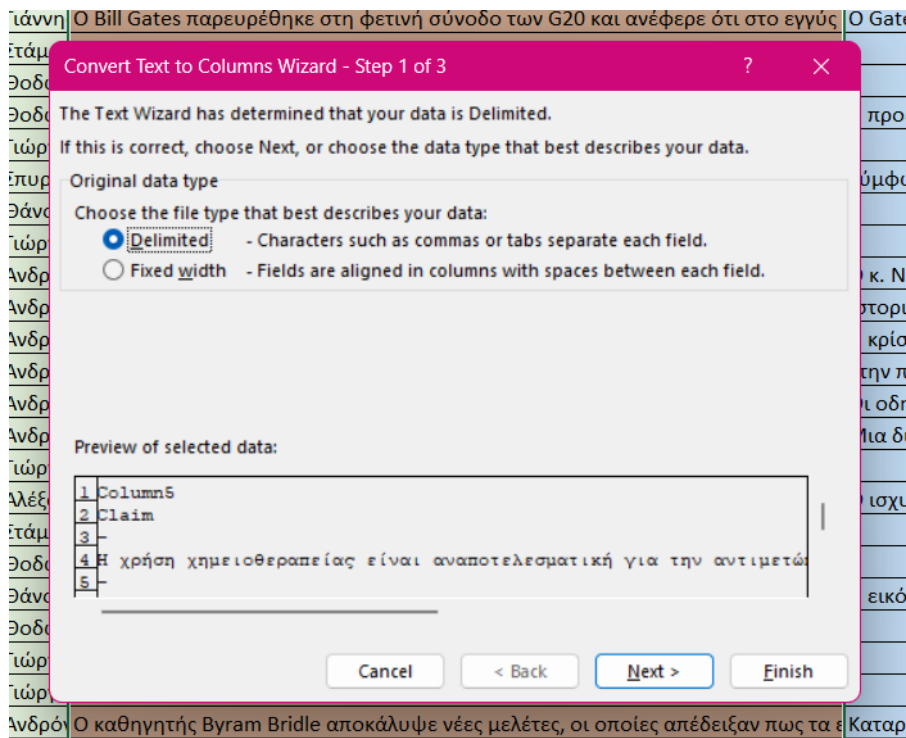
Εικόνα 5.5.20 Αρχείο keywordsDictionary.csv

	A
5	and
6	h
7	μία
8	n
9	o
10	oi
11	t
12	to
13	τα
14	to
15	ακόμη
16	αλλά
17	άλλα

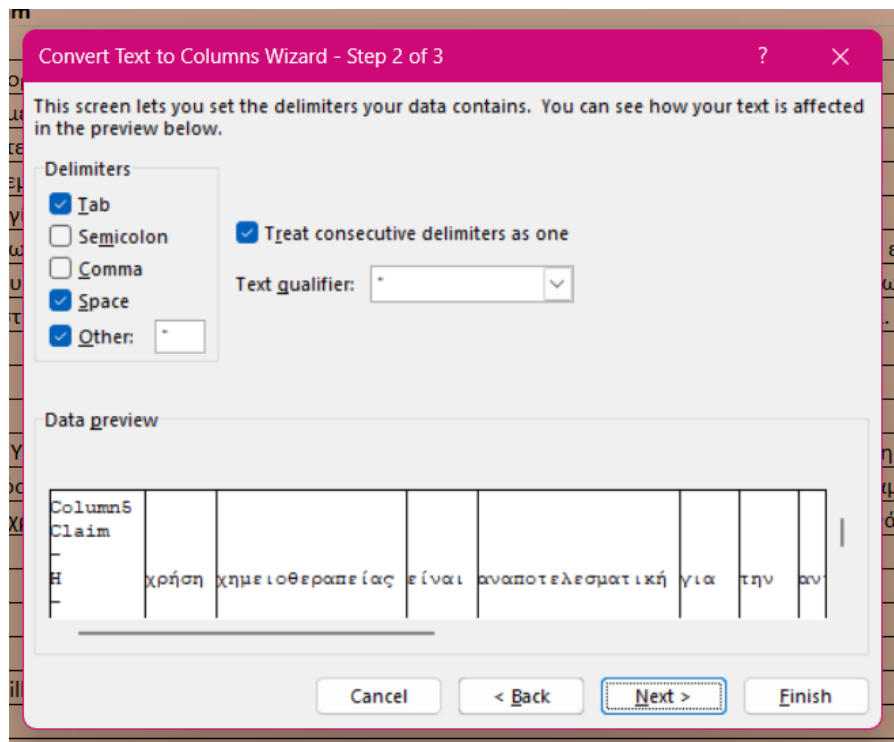
Τα βήματα επεξεργασίας των ισχυρισμών είναι τα εξής:

1. Όλοι οι ισχυρισμοί στο φύλλο επιλέγονται και το περιεχόμενο κάθε κελιού διασπάται λέξη-λέξη. Για να γίνει αυτό πατάμε Data → Data Tools → Text to Columns → Delimited → Next → Delimiters: Tab, Space, Other: “ → Treat consecutive delimiters as one: Ναι → Next → Destination: \$J\$1 → Finish (Δείτε τις Εικόνες 5.5.21 – 5.5.23 με τα βήματα.).

Εικόνα 5.5.21 Επεξεργασία των ισχυρισμών (1)



Εικόνα 5.5.22 Επεξεργασία των ισχυρισμών (2)



Εικόνα 5.5.24 Επεξεργασία των ισχυρισμών (4)

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
ξε είναι	αναποτελε	για	την	αντιμετώπι	καρκίνων	ενώ	είναι	επικίνδυν	τις	περισσότε	φορές									
είναι	θετική	στον	κορωνοϊό																	
ων της	αμερικανι	ρευματολογι	εταιρείας	οι	ασθενείς	αυτοάνοσι	νοσημάτων	συγκεκριμ	ασθενείς	με	ρευματολογ	προβλήμα	οι	οποίοι	χρησιμοπι	φάρμακα	όπως	η	κορτιζόνη	δε
/ σε	συνέντευξ	στο	graham	norton	show	δήλωσε	πως	στη	χόλιγουντ	διενεργού	τελετές	παιδρασα	και	κανιβαλοι	παιδιών					
ς τομ	χανκς	αποκαλύπ	την	παιδοφιλι	στο	χόλιγουντ														
γι του	δήμου	του	παρισιο	σαπίζου	και	διαρρέου	τα	τοξικά	προϊόντα	τους	στο	χώρα	οι	επιχειρη	ανακύκλω	δε	τα	δέχονται	αν	δε

4. Σε αυτά τα δεδομένα πρέπει να διαγραφούν τα stopwords, που είναι αποθηκευμένα σε ένα ξεχωριστό φύλλο. Προκειμένου να γίνει αυτό θα χρησιμοποιηθεί η συνάρτηση `IF(COUNTIF(Stopwords!A1:A1000,KeywordsRAW_SUBSTITUTED!AZ)>0,"",KeywordsRAW_SUBSTITUTED!AZ)`, όπου το `KeywordsRAW_SUBSTITUTED` είναι το όνομα του φύλλου, όπου βρίσκονται οι ισχυρισμοί σπασμένοι λέξη-λέξη μετά την αντικατάσταση, που έγινε στο βήμα 3, το `A` είναι η στήλη του κελιού και το `Z` είναι η γραμμή του κελιού (Εικόνα 5.5.25).

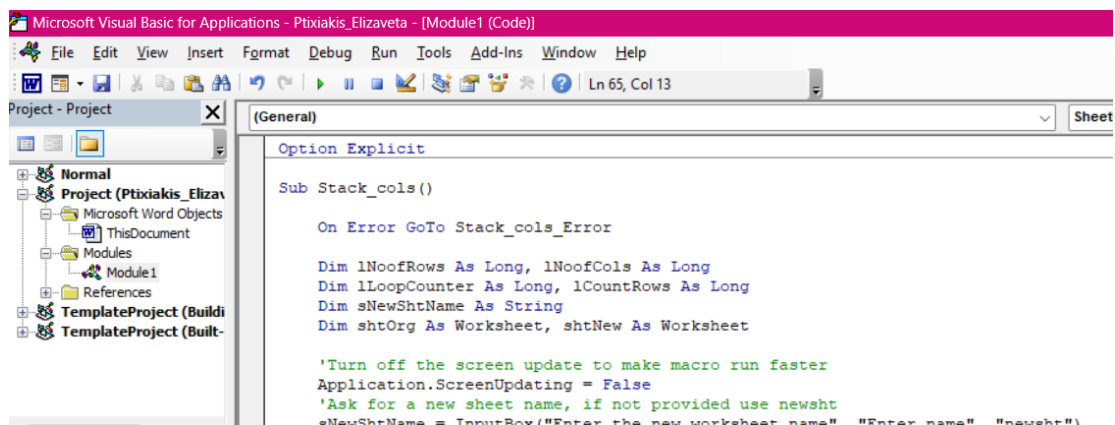
Εικόνα 5.5.25 Επεξεργασία των ισχυρισμών (5)

A	B	C	D	E	F	G	H	I	J
	χρήση	χημειοθεραπείας		αναποτελε			αντιμετώπιση	καρκίνων	
	άνγκελα	μέρκελ		θετική		κορωνοϊό			
		ανακοίνωση		αμερικανι	ρευματολογικη	εταιρειας		ασθενεις	αυτοανοσι
	μελ	γκίμπσον		συνέντευξ		graham	norton	show	
	γνωστός	ηθοποιός	τομ	χανκς	αποκαλύπτει	παρισιο	παιδοφιλία		χόλιγουντ
εκατοντάδες	ηλεκτρικά	αυτοκίνητα		δήμου		παρισιού	σαπίζου		διαρρέου

5. Οι τιμές των ληφθέντων δεδομένων αντιγράφονται σε ένα κενό φύλλο.
6. Διαγράφονται όλα τα κενά, όπως περιγράφηκε στην Ενότητα 5.5.1.1 πρώτα με το `Shift cells up` και μετά με το `Shift cells left`.

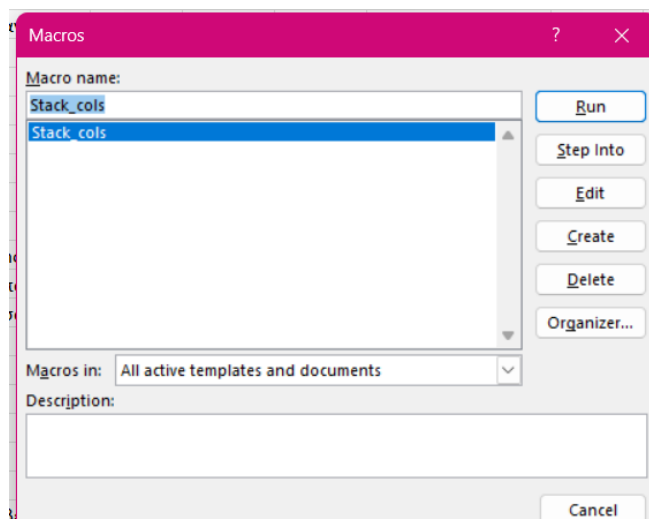
7. Με το πάτημα Alt + F11 ανοίγουμε το Microsoft Visual Basic. Επιλέγουμε το Insert → Model. Εισάγουμε του κώδικα, που προσπελάστηκε από τον ιστότοπο <https://nandeshwar.info/>¹⁰⁵ (Εικόνα 5.5.26).

Εικόνα 5.5.26 Επεξεργασία των ισχυρισμών (6)



8. Με το πάτημα Alt + F8 ανοίγουμε τα Macros. Επιλέγουμε το Stack_cols και πατάμε Run (Εικόνα 5.5.27).

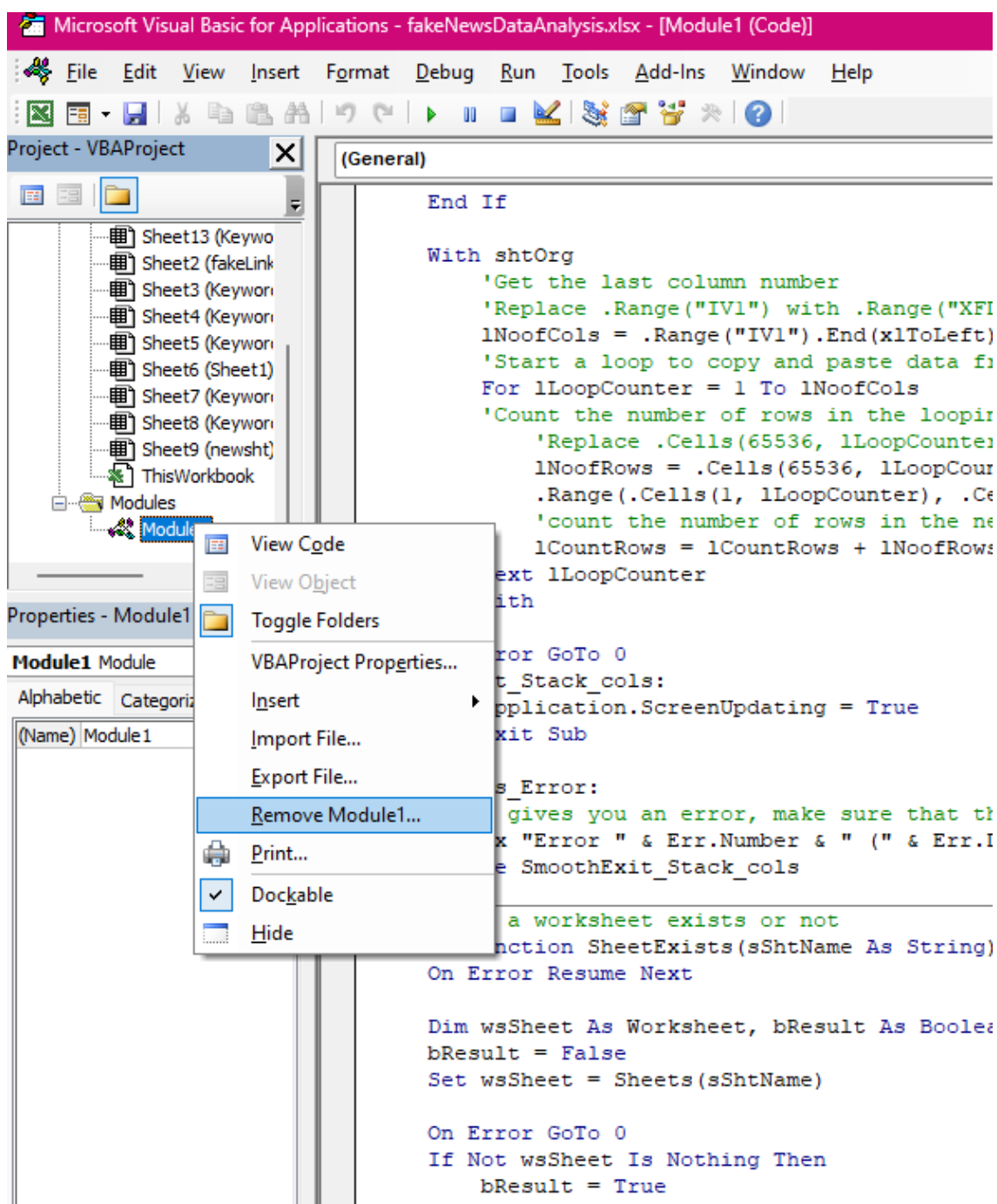
Εικόνα 5.5.27 Επεξεργασία των ισχυρισμών (7)



¹⁰⁵ <https://nandeshwar.info/useful-procedures/stack-columns-of-data-on-one-column/#:~:text=Press%20Alt%20%2B%20F8%20to%20bring,column%20with%20the%20column%20headers>. (προσπελάστηκε στις 26/01/2024)

9. Ανοίγουμε ξανά το παράθυρο Microsoft Visual Basic. Στο παράθυρο Project – VBAP Project κάνουμε το δεξί κλικ στο Module1 → Remove Module1 (Εικόνα 5.5.28). Το αποτέλεσμα της εκτέλεσης είναι διαθέσιμο στο φύλλο Keywords FINAL του αρχείου fakeLinksDataAnalysis.xlsx.

Εικόνα 5.5.28 Επεξεργασία των ισχυρισμών (8)



10. Το τελευταίο βήμα είναι το stemming και περιλάμβανε δυο στάδια προετοιμασίας.

- Χειρωνακτική ανάλυση:

Κάθε λέξη αναλύθηκε και μετά στην πρωτότυπη μορφή της κάθε λέξης, προστέθηκε η αρχική της μορφή χωρισμένη με ερωτηματικό, πχ «μονόπλευρα;μονόπλευρος». Κάποιες λέξεις που δεν φέρνουν καμία ουσία στα αποτελέσματα, διαγράφηκαν, πχ οι αριθμοί. Το πλήρες σύνολο των λέξεων-κλειδιών σε πρωτότυπη και αρχική μορφές είναι διαθέσιμο στο αρχείο keywordsDictionary.csv.

- Γραφή του κώδικα στη Java:

Για τον σκοπό αυτό δημιουργήθηκε η κλάση DictionaryReader. Στην κλάση αυτή υπάρχουν τρεις μέθοδοι:

1. Η μέθοδος chooseFile(), που επιλέγει το CSV-αρχείο μόνο με τις λέξεις-κλειδιά σε μια στήλη (Εικόνα 5.5.29).

Εικόνα 5.5.29 Μέθοδος chooseFile() της κλάσης DictionaryReader

```
public static String chooseFile(){
    FileDialog dialog = new FileDialog((Frame)null, "Select File to Open");
    dialog.setMode(FileDialog.LOAD);
    dialog.setVisible(true);
    String f = dialog.getFile();
    dialog.dispose();
    System.out.println(f + " was chosen.");
    return f;
}
```

2. Η μέθοδος retrieveDictionary() που ανακτά τα δεδομένα από το αρχείο keywordsDictionary.csv και τα αποθηκεύει σε ένα HashMap, όπου η πρώτη λέξη γίνεται το κλειδί και η δεύτερη λέξη γίνεται η τιμή (Εικόνα 5.5.30). Ο Πίνακας 5.5.1 αποδεικνύει την προσέγγιση ανάκτησης των δεδομένων και της αποθήκευσής τους στο HashMap dict.

Εικόνα 5.5.30 Μέθοδος retrieveDictionary() της κλάσης DictionaryReader

```

public static void retrieveDictionary(File f){
    try (BufferedReader in = new BufferedReader(
        new InputStreamReader(
            new FileInputStream(f), "UTF8")
        )) {
        String str;
        String strKeyword;
        String strValue;
        int counter = 0;

        while ((str = in.readLine()) != null) {
            counter +=1;
            String[] keywordValue= str.split(";", 2);
            strKeyword = keywordValue[0];
            strValue = keywordValue[1];
            dict.put(strKeyword, strValue);
            System.out.println(
                strKeyword + " - " + strValue + "counter = " + counter);
        }

    } catch (Exception e) {
    }
}

```

Πίνακας 5.5.1 Δεδομένα, αποθηκευμένα στο αρχείο keywordsDictionary.csv ως στοιχεία του HashMap dict

ΤΑ ΔΕΔΟΜΕΝΑ ΣΤΟ keywordsDictionary.csv	HashMap dict	
	ΚΛΕΙΔΙΑ	ΤΙΜΗ
απόθεσης;απόθεση	απόθεσης	αποθήκη
αποθήκες;αποθήκη	αποθήκες	αποθήκη
αποθηκευμένα;αποθηκευμένος	αποθηκευμένα	αποθηκευμένος
αποθήκη;αποθήκη	αποθήκη	αποθήκη

3. Η μέθοδος changeKeywords(), που μέσα στο βρόγχο διαβάζει κάθε γραμμή του αρχείου με τις λέξεις-κλειδιά, που πρέπει να αντικατασταθούν, και ελέγχει αν η λέξη αυτή υπάρχει ανάμεσα στα κλειδιά του HashMap dict. Αν υπάρχει ένα κλειδί X, που είναι το ίδιο με τη λέξη που ελέγχουμε, τότε η λέξη αυτή αντικαθίσταται με την τιμή του κλειδιού X (Εικόνα 5.5.31).

Εικόνα 5.5.31 Μέθοδος changeKeywords() της κλάσης DictionaryReader

```
public static void changeKeywords(File oldF, File newF) {
    try (BufferedReader in = new BufferedReader(
        new InputStreamReader(
            new FileInputStream(oldF), "UTF8")
        )) {
        String strOld;
        String strNew;
        int counter = 0;

        try (Writer textWriter = new BufferedWriter(
            new OutputStreamWriter(
                new FileOutputStream(newF, true),
                StandardCharsets.UTF_8)
            )) {
            System.setOut(new PrintStream(System.out, true, "UTF8"));

            while ((strOld = in.readLine()) != null) {

                if (counter < 1) {
                    strOld = strOld.substring(1);
                }

                counter +=1;

                if (dict.containsKey(strOld)) {
                    strNew = dict.get(strOld);
                    textWriter.append(strNew+"\n");
                }

                if (dict.containsKey(strOld)==false) {
                    System.out.println(strOld + " not in dictionary");
                }
            }

            textWriter.flush();
        }
    } catch (Exception e) {
        System.out.println(e.getMessage());
    }
}
```

Αφού έχουμε και το αρχείο keywordsDictionary.csv και τον κώδικα για την αντικατάσταση τιμών, αντιγράφουμε τα δεδομένα από το φύλλο keywordsFINAL σε ένα ξεχωριστό αρχείο CSV UTF-8, πχ υπό το όνομα keywordsAll.csv και εκτελούμε τον κώδικα της κλάσης DictionaryReader, όπου επιλέγουμε το αρχείο με τα δεδομένα, πχ keywordsAll.csv. Μετά την εκτέλεση, έχουμε ένα νέο αρχείο, όπου οι λέξεις-κλειδιά είναι σε αρχική μορφή.

Για την ανάλυση των ισχυρισμών ανά χρονιά ή ανά κατηγορία τα βήματα επεξεργασίας είναι τα ίδια, το μόνο που προηγείται είναι το φιλτράρισμα των άρθρων βάση τη χρονιά ή την κατηγορία, αντίστοιχα.

5.5.4 ΕΞΑΓΩΓΗ ΤΩΝ ΑΡΘΡΩΝ ΠΟΥ ΑΦΟΡΟΥΝ ΕΝΑ ΣΥΓΚΕΚΡΙΜΕΝΟ ΘΕΜΑ

Στα ελληνικά ΜΜΕ κατά την περίοδο 2013 – 2023 τα πιο δημοφιλή θέματα ήταν κορονοϊός, εμβολιασμός, πολιτική, η Ρωσία, η Ουκρανία, οι ΗΠΑ, η Τουρκία και η Ελλάδα. Κάποια από αυτά τα θέματα παραμένουν ακόμα επίκαιρα. Για να αναλυθούν τα άρθρα ανά κάθε από αυτά τα θέματα, αρχικά τα άρθρα αυτά πρέπει να εξαχθούν είτε χειρωνακτικά (πολύ χρονοβόρο και με μεγάλη πιθανότητα λαθών), είτε με τη βοήθεια του κώδικα σε μια γλώσσα προγραμματισμού. Στη συνέχεια, για να αναλυθούν οι ημερομηνίες δημοσίευσης και οι κατηγορίες αυτών των άρθρων, πρέπει να εκτελεστεί ο αντίστοιχος κώδικας.

Και για τα δυο στάδια αποφασίστηκε να συγγραφεί ο κώδικας στη Java. Για την εξαγωγή και τον υπολογισμό των δημοσιευμάτων ανά μήνα δημιουργήθηκε η κλάση ClaimDateRetrieval και για την εξαγωγή και τον υπολογισμό της συχνότητας των κατηγοριών δημιουργήθηκε η κλάση ClaimCategoryRetrieval.

Πριν εκτελεστεί ο κώδικας της κλάσης ClaimDateRetrieval και της κλάσης ClaimCategoryRetrieval από το φύλλο fakeLinksDataRAW του αρχείου fakeLinksDataAnalysis.xlsx σε ένα ξεχωριστό αρχείο claimDate.csv claimCat.csv, αντίστοιχα, αντιγράφηκαν οι ισχυρισμοί των άρθρων και οι ημερομηνίες («mmm-yyyy» του δημοσιεύματός τους (Εικόνα 5.5.32) ή οι ισχυρισμοί των άρθρων και οι κατηγορίες τους (Εικόνα 5.5.33), αντίστοιχα, με το διαχωριστικό (αγγλ. delimiter) «@@@».

Εικόνα 5.5.32 Αρχείο claimDate.csv με όλους τους ισχυρισμούς και τις ημερομηνίες δημοσίευσής τους

	A	B	C	D
1	-@@@January-2017			
2	Η χρήση χημειοθεραπείας είναι αναποτελεσματική για την αντιμετώπιση καρκίνων ενώ είναι επικίνδυνη 1			
3	-@@@May-2016			
4	Η Άνγκελα Μέρκελ είναι θετική στον κορωνοϊό@@@March-2020			
5	-@@@April-2014			
6	Σύμφωνα με ανακοίνωση της "Αμερικανικής Ρευματολογικής Εταιρείας" οι ασθενείς αυτοάνοσων νοσημέ			
7	-@@@January-2015			
8	-@@@June-2017			
9	-@@@January-2015			
10	Ο Μελ Γκίμπσον σε συνέντευξή στο Graham Norton Show δήλωσε πως στη Χόλιγουντ διενεργούνται τελετ			
11	ο γνωστός ηθοποιός Τομ Χανκς αποκαλύπτει την παιδοφιλία στο Χόλιγουντ. @@@April-2020			

Εικόνα 5.5.33 Αρχείο claimCat.csv με όλους τους ισχυρισμούς και οι κατηγορίες τους

	A	B	C	D	E	F	G	H	I
1	-@@@HOAXES;FAKE NEWS;								
2	Η χρήση χημειοθεραπείας είναι αναποτελεσματική για την αντιμετώπιση καρκίνων ενώ εί								
3	-@@@HOAXES;								
4	Η Άνγκελα Μέρκελ είναι θετική στον κορωνοϊό@@@FAKE NEWS;ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ;								
5	-@@@ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ;								
6	Σύμφωνα με ανακοίνωση της "Αμερικανικής Ρευματολογικής Εταιρείας" οι ασθενείς αυτο								
7	-@@@ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΕΓΜΑΤΑ;								
8	-@@@HOAXES;FAKE NEWS;ΔΙΑΦΟΡΑ;								
9	-@@@ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ;ΔΙΑΦΟΡΑ;								
10	Ο Μελ Γκίμπσον σε συνέντευξή στο Graham Norton Show δήλωσε πως στη Χόλιγουντ διενε								
11	ο γνωστός ηθοποιός Τομ Χανκς αποκαλύπτει την παιδοφιλία στο Χόλιγουντ. @@@ΣΥΝΩΜ								
12	Εκατοντάδες ηλεκτρικά αυτοκίνητα του δήμου του Παρισιού σαπίζουν και διαρρέουν τα τ								
13	-@@@ΥΓΕΙΑ;ΨΕΥΔΟΕΠΙΣΤΗΜΗ;								
14	-@@@ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ;								
15	-@@@HOAXES;ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ;								
16	-@@@ΔΙΑΦΟΡΑ;								
17	-@@@HOAXES;FAKE NEWS;								
18	-@@@ΑΠΑΤΗ;LIKE FARMING;								
19	Στη διαδήλωση κατά των μέτρων κορωνοϊού στο Βερολίνο συμμετείχαν 5.000.000 άνθρωπ								
20	Άρματα μάχης κατευθύνθηκαν προς το Βερολίνο λόγω της διαδήλωσης κατά της λήψης μέ								
21	-@@@HOAXES;FAKE NEWS;ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ;ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ;								
22	-@@@HOAXES;								
23	-@@@HOAXES;FAKE NEWS;								

Στην αρχή στο Java-κώδικα και στις δυο περιπτώσεις δημιουργούνται τα HashMaps, που συσχετίζονται με τα θέματα του ενδιαφέροντός μας. Η δημιουργία τους παρουσιάζεται στην Εικόνα 5.5.34 και η τα στοιχεία των HashMaps απεικονίζονται στον Πίνακα 5.5.2.

Εικόνα 5.5.34 Δημιουργία των HashMap των κατηγοριών Covid, Vaccination, Russia, Ukraine, Greece, Turkey, Politics, the USA

```
String covidKeywords [] = {"cov19", "covid", "κοροναϊ", "κορονοϊό",
    "κορονοϊ", "κορωνοϊό", "κορωνοϊό", "κορωνοϊ", "κορωνοϊ", "sars-cov",
    "2019-ncov", "mers-cov"};
String vacKeywords [] = {"εμβόλ", "εμβολ"};
String russiaKeywords [] = {"ρωσία", "ρωσικ", "ρώσ", "μόσχα", "russian",
    "moskva", "πούτιν"};
String ukraineKeywords [] = {"ουκραν", "ukrainskaya", "κίεβ", "ζελένσκα",
    "zelensk", "μαριούπολη", "χάρκοβο", "αζόφ"};
String greeceKeywords [] = {"ελλάδ", "ελλαδ", "έλλη", "ελλην", "ελλή",
    "αθή", "αθηναίων", "αθηνών", "greece", "πειραια", "πειραιά", "πειραιώς",
    "hellenic"};
String turkeyKeywords [] = {"turkish", "ερντογάν", "προύσα", "ταγίπ",
    "τουρκ", "τούρκ"};
String politicsKeywords[] = {"biden", "clinton", "macron", "mitsotakis",
    "obama", "trump", "αντιπολιτεύ", "αντιπολίτευ", "βουλευτ", "βουλή",
    "εκλογ", "ερντογάν", "κλίντον", "κογκρέσο", "κυβερνήσ", "κυβέρνησ",
    "κυβερνητ", "μακρόν", "μητισοτάκη", "μπαιντεν", "μπάιντεν", "ομπάμα",
    "πολιτικ", "πούτιν", "συριζα", "σύριζα", "ταγίπ", "τραμπ", "τσίπρα",
    "υπουργ"};
String usaKeywords [] = {"american", "biden", "clinton", "columbia",
    "florida", "georgia", "harvard", "obama", "stanford", "trump", "washington",
    "αμερικ", "αριζόνα", "ατλάντια", "ηπα", "καλιφόρνια", "καπιτώλιο", "κλίντον",
    "κογκρέσο", "μίσιγκαν", "μπαιντεν", "μπάιντεν", "ομπάμα", "ουάσιγκτον",
    "ουάσιγκτον", "ουϊσκόνσιν", "οχάιο", "τεξας", "τέξας", "τραμπ", "φλόριντα",
    "χόλιγουντ"};

//Adding new values to the dictionary "themeHashSet"
themeHashMap = MinorClass.addThemeDictValues(
    "Covid", covid, covidKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Vaccination", vaccination, vacKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Russia", russia, russiaKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Ukraine", ukraine, ukraineKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Greece", greece, greeceKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Turkey", turkey, turkeyKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "Politics", politics, politicsKeywords, themeHashMap);
themeHashMap = MinorClass.addThemeDictValues(
    "USA", usa, usaKeywords, themeHashMap);
```

Πίνακας 5.5.2 Στοιχεία του HashMap themeHashMap

themeHashMap <String, HashSet>	
ΚΛΕΙΔΙ (String)	TIMH (HashSet<String>)
"Covid"	("cov19", "covid", "κοροναϊ", "κορονοϊό", "κορονοϊ", "κορωνοϊό", "κορωνοϊό", "κορωνοϊ", "κορωνοϊ", "sars-cov", "2019-ncov", "mers-cov")

"Vaccination"	("εμβόλ", "εμβολ")
"Russia"	("ρωσία", "ρωσικ", "ρώσ", "μόσχα", "russian", "moskva", "πούτιν")
"Ukraine"	("ουκραν", "ukrainskaya", "κίεβ", "ζελένσκ", "zelensk", "μαριούπολη", "χάρκοβο", "αζόφ")
"Greece"	("ελλάδ", "ελλαδ", "έλλην", "ελλην", "ελλην", ".αθήν", "αθηναίων", "αθηνών", "greece", "πειραια", "πειραιά", "πειραιώς", "hellenic")
"Turkey"	("turkish", "ερντογάν", "προύσα", "ταγίπ", "τουρκ", "τούρκ")
"Politics"	("biden", "clinton", "macron", "mitsotakis", "obama", "trump", "αντιπολιτεύ", "αντιπολίτευ", "βουλευτ", "βουλή", "εκλογ", "ερντογάν", "κλίντον", "κογκρέσο", "κυβερνήσ", "κυβέρνησ", "κυβερνητ", "μακρόν", "μητσοτάκη", "μπαιντεν", "μπάιντεν", "ομπάμα", "πολιτικ", "πούτιν", "συριζα", "σύριζα", "ταγίπ", "τραμπ", "τσίπρα", "υπουργ")
"USA"	("american", "biden", "clinton", "columbia", "florida", "georgia", "harvard", "obama", "stanford", "trump", "washington", "αμερικ", "αριζόνα", "ατλάντα", "ηπα", "καλιφόρνια", "καπιτώλιο", "κλίντον", "κογκρέσο", "μίσιγκαν", "μπαιντεν", "μπάιντεν", "ομπάμα", "ουάσιγκτον", "ουάσιγκτον", "ουϊσκόνσιν", "οχάιο", "τεξας", "τέξας", "τραμπ", "φλόριντα", "χόλιγουντ")

Έπειτα επιλέγεται το θέμα και το αρχείο με ισχυρισμούς (claimDate.csv ή claimCat.csv, αντίστοιχα).

Προκειμένου να ληφθούν υπόψιν μόνο οι ισχυρισμοί ενός θέματος, πχ που αφοράνε τον εμβολιασμό, κάθε γραμμή του κώδικα διαβάζεται και επεξεργάζεται μέσα στη μέθοδο extractClaimDate() (για την επεξεργασία των ισχυρισμών και των ημερομηνιών) ή τη μέθοδο extractClaimCategories() (για την επεξεργασία των ισχυρισμών και των κατηγοριών) (Εικόνες 5.5.35 και 5.5.36, αντίστοιχα).

Εικόνα 5.5.35 Μέθοδος extractClaimDate()

```
public static void extractClaimDate(  
    String row, int i, String rowClaim, String rowDate,  
    HashSet <String> hs){  
    boolean continueCheck = true;  
  
    String[] splitLine = row.split("@@@",2);  
    i +=1;  
    rowClaim = splitLine[0].toLowerCase();  
    rowDate = splitLine[1];  
  
    if (rowDate.contains("\\")) {  
        rowDate = rowDate.replaceAll("\\", "");  
    }  
  
    System.out.println("\n");  
    System.out.println("Printing new line");  
  
    for (String word : hs){  
        if (continueCheck){  
            System.out.println("new line: "+ rowClaim);  
  
            if (rowClaim.contains(word)) {  
                int valueOld = dateHashMap.get(rowDate);  
                int valueNew = valueOld + 1;  
                System.out.println(rowClaim);  
                System.out.println("contains: "+ word);  
                System.out.println(  
                    "Old value of: " + rowDate + " = " + valueOld);  
                System.out.println(  
                    "New value of: " + rowDate + " = " + valueNew);  
                System.out.println("\n");  
                dateHashMap.replace(rowDate, valueNew);  
                continueCheck = false;  
            }  
        }  
    }  
}
```

Εικόνα 5.5.36 Μέθοδος extractClaimCategories()

```

public static void extractClaimCategories(
    String row, int i, String rowClaim, String rowCat, HashSet<String> hs) {
    String[] splitLine = row.split("@@@", 2);
    i += 1;
    rowClaim = splitLine[0].toLowerCase();
    rowCat = splitLine[1];

    for (String word : hs) {

        if (rowClaim.contains(word)) {
            System.out.println("\n");
            System.out.println(rowClaim);
            System.out.println("Line contains a word: " + word);
            String[] splitLineCats = rowCat.split(";");
            System.out.println(
                "Analyzing its categories: "
                + Arrays.toString(splitLineCats));

            for (String cat : splitLineCats) {

                if (catHashMap.keySet().contains(cat)) {
                    int valueOld = catHashMap.get(cat);
                    int valueNew = valueOld + 1;
                    catHashMap.replace(cat, valueNew);
                    System.out.println("keyset cointains: " + cat);
                    System.out.println(
                        "Old value of: " + cat + ": " + valueOld);
                    System.out.println(
                        "New value of: " + cat + ": " + valueNew);
                } else {
                    System.out.println(cat + " is not in the dictionary");
                }
            }

            break;
        }
    }
}

```

Για τον υπολογισμό της συχνότητας, πχ των άρθρων του θέματος «Vaccination» ανά μήνα ή ανά κατηγορία χρησιμοποιείται η μετρητής που αποθηκεύει δεδομένα σε ένα HashMap. Στον Πίνακα 5.5.3 απεικονίζεται η σχετική δομή του HashMap catHashMap και στον Πίνακα 5.5.4 η σχετική δομή του HashMap dateHashMap,

Πίνακας 5.5.3 Σχετική δομή του HashMap catHashMap

catHashMap <String, Integer>	
ΚΛΕΙΔΙ (String)	ΤΙΜΗ (Integer)
"May-2013"	0
"June-2013"	0
"July-2013"	0
"August-2013"	0
"September-2013"	0
"October-2013"	0
...	0
"May-2023"	0

Πίνακας 5.5.4 Σχετική δομή του HashMap dateHashMap

dateHashMap <String, Integer>	
ΚΛΕΙΔΙ (String)	ΤΙΜΗ (Integer)
"HOAXES"	0
"ΠΑΡΑΠΛΗΡΟΦΟΡΗΣΗ"	0
"FAKE NEWS"	0
"ΔΙΑΦΟΡΑ"	0
"ΣΥΝΩΜΟΣΙΟΛΟΓΙΑ"	0
"ΨΕΥΔΕΣ"	0
...	0
"ΕΙΚΟΝΑ ΑΠΟ ΠΡΟΓΡΑΜΜΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ"	0

5.6 ΑΠΟΤΕΛΕΣΜΑΤΑ

Η EllhnikaHoaxes.gr είναι η ελληνική υπηρεσία ελέγχου ψευδών ειδήσεων και όχι μια πηγή ψευδών ειδήσεων. Η ομάδα της πάντα ελέγχει μόνο τις πιο επίκαιρες ειδήσεις στο ελληνικό Διαδίκτυο, άρα είναι μια καλή πηγή πληροφόρησης για το τι είναι ευρέως δημοφιλές (αγγλ. viral) στα ελληνικά ΜΜΕ και τα μέσα κοινωνικής δικτύωσης.

Οι κατηγορίες που ανατίθενται από τους ελεγκτές της ομάδας της EllhnikaHoaxes.gr σχετίζονται με τον ισχυρισμό ή την είδηση που αναλύεται και όχι με το ίδιο το άρθρο.

Οι λέξεις-κλειδιά έχουν εξαχθεί από τους ύποπτους ισχυρισμούς που ελέγχονται από την EllhnikaHoaxes.gr.

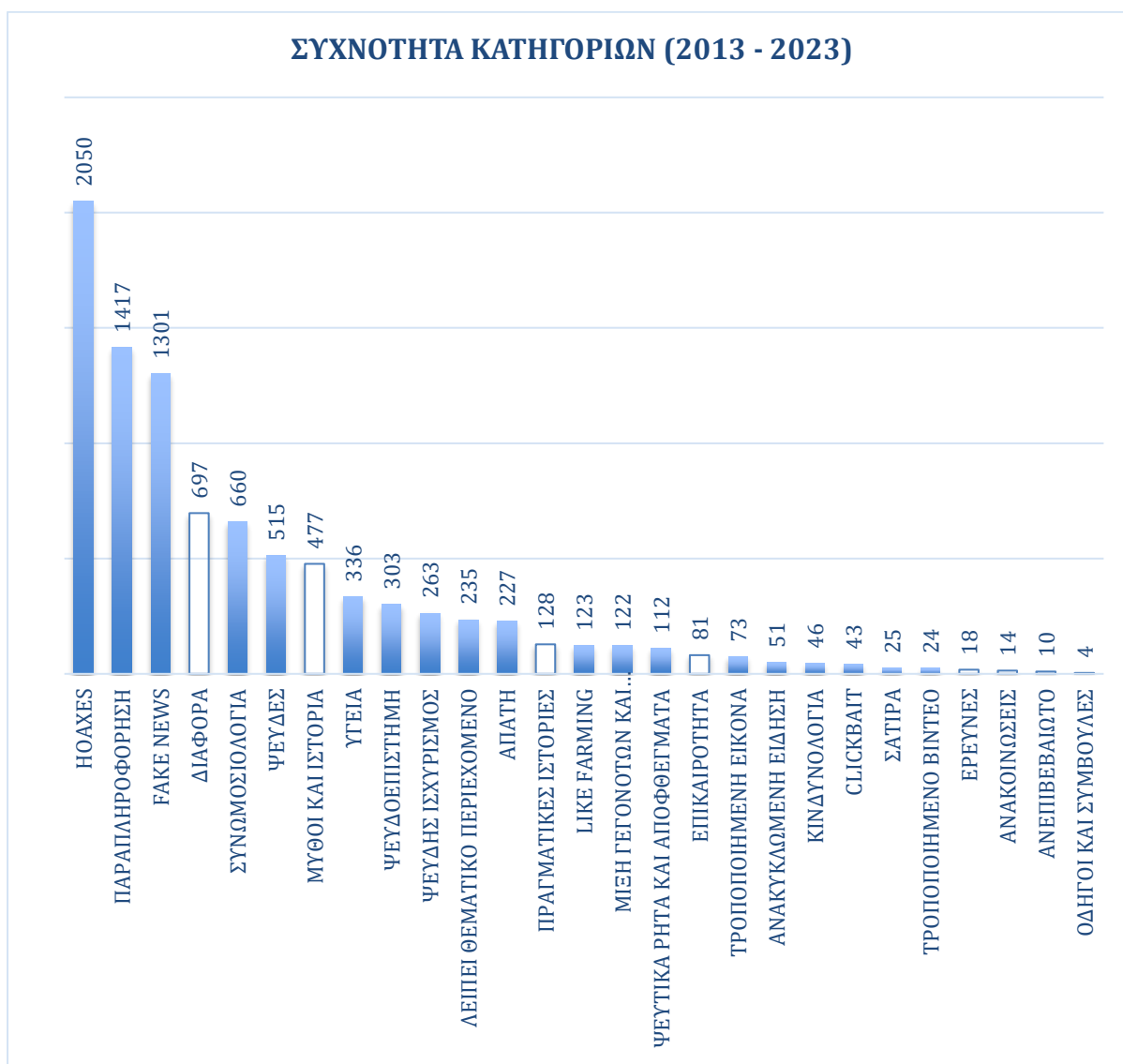
Τα θέματα αποτελούνται από τις λέξεις-κλειδιά που σχετίζονται μεταξύ τους, πχ οι λέξεις-κλειδιά «Ρωσία», «Μόσχα», «Κρεμλίνο», «Πούτιν» και «ρωσικός» - όλες σχετίζονται με τη Ρωσία.

5.6.1 ΟΙ ΚΑΤΗΓΟΡΙΕΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ ΔΙΑΔΙΚΤΥΟ

Βάση τα δεδομένα που ανακτήθηκαν από τον ιστότοπο EllhnikaHoaxes.gr και στη συνέχεια επεξεργάστηκαν με τη βοήθεια του MS Excel, η πιο συχνή κατηγορία των ψευδών ειδήσεων στην Ελλάδα κατά την περίοδο 2013 – 2023 ήταν η κατηγορία «Hoaxes» με 2050 άρθρα, οι κατηγορίες «Παραπληροφόρηση» και «Fake news» την ακολουθούν με 1417 και 1301 άρθρα, αντίστοιχα (Διάγραμμα 5.6.1).

Οι κατηγορίες «Διάφορα», «Μύθοι και ιστορίες», «Επικαιρότητα», «Έρευνες», «Ανακοινώσεις», «Ανεπιβεβαίωτο» και «Οδηγοί και συμβουλές» δε σχετίζονται απευθείας με ψευδείς ειδήσεις και απεικονίζονται με άσπρο χρώμα. Αν και αυτές οι κατηγορίες δε λαμβάνονται υπόψιν στα αποτελέσματα, είναι ωφέλιμο να διατηρηθούν γιατί μαζί τους μπορούμε να έχουμε μια πλήρη εικόνα των κατηγοριών που χρησιμοποιούνται από τους ελεγκτές της EllhnikaHoaxes.gr.

Διάγραμμα 5.6.1 Κατηγορίες των άρθρων στον EllhnikaHoaxes.gr, που δημοσιεύτηκαν το 2013 - 2023

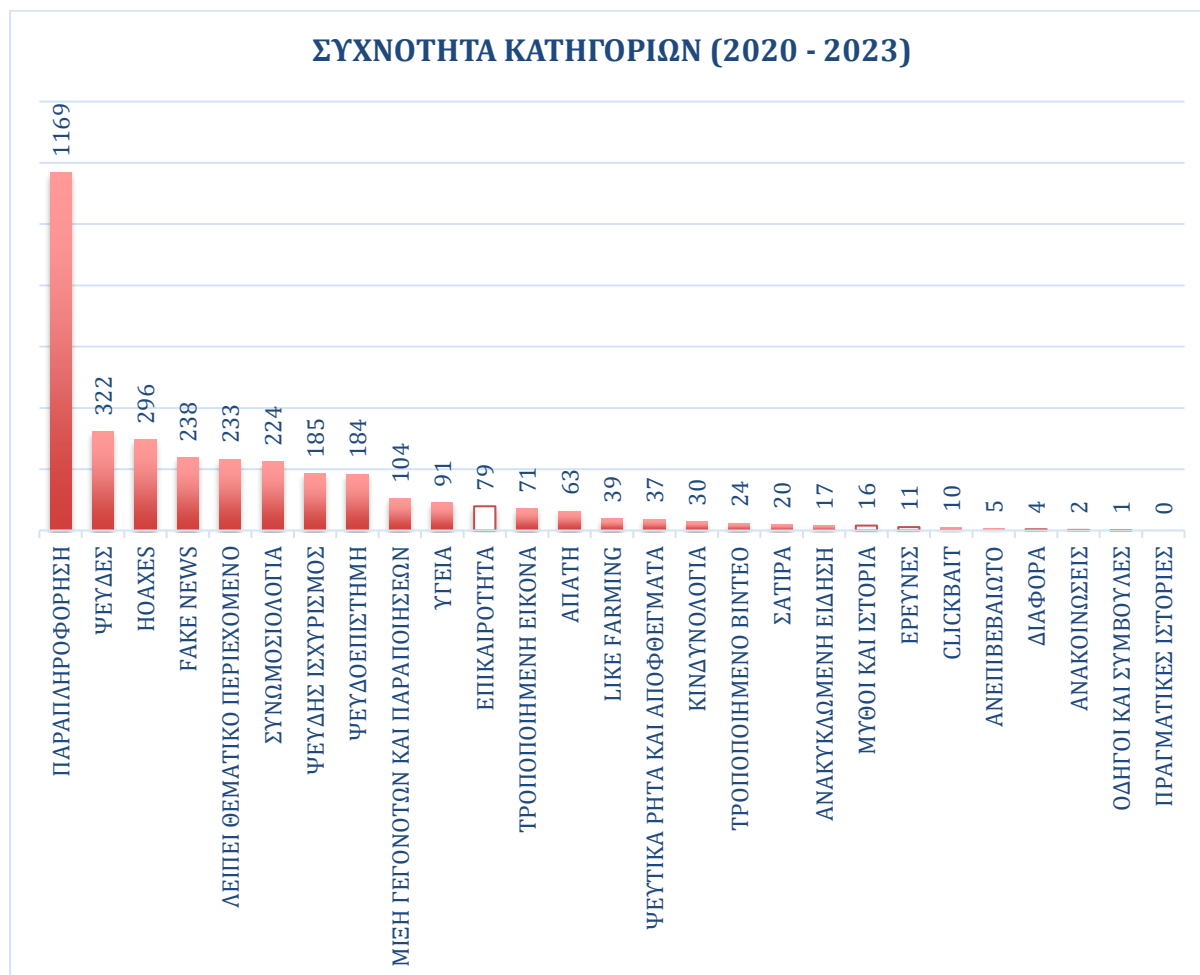


Το 2020 η δομή των άρθρων στον EllhnikaHoaxes.gr άλλαξε πολύ και σχεδόν όλα τα άρθρα ξεκίνησαν να είναι πλήρη, δηλαδή με τους ισχυρισμούς, τα συμπεράσματα και τις δυο μορφές κατηγοριοποίησης (Δείτε τον Πίνακα 5.4.1). Αυτή η πλέον πληρότητα πρόκειται να έχει αλλάξει τα στατιστικά δεδομένα. Για αυτό το λόγο, τα άρθρα που δημοσιεύτηκαν το 2020 και μετά αναλύθηκαν ξεχωριστά (Διάγραμμα 5.6.2).

Πράγματι, το 2020 – 2023 τα περισσότερα άρθρα είχαν την κατηγορία «Παραπληροφόρηση» (1169 άρθρα) και όχι την κατηγορία «Hoaxes». Αντιθέτως, οι άλλες κατηγορίες, όπως οι «Ψευδές», «Hoaxes» ή «Fake news» δεν αποτελούν ούτε ένα τρίτο των άρθρων που σχετίζονται με την κατηγορία «Παραπληροφόρηση». Οι κατηγορίες «Διάφορα» και «Μύθοι

και ιστορία» που ήταν τόσο συχνές, αν ληφθεί υπόψιν όλη η δεκαετία, την περίοδο 2020 – 2023 δε χρησιμοποιούταν σχεδόν καθόλου. Γενικώς, όλες οι κατηγορίες που δε σχετίζονται απευθείας με ψευδείς ειδήσεις ήταν σε χρήση πολύ ελάχιστα.

Διάγραμμα 5.6.2 Κατηγορίες των άρθρων στον Ellhnikahoaxes.gr, που δημοσιεύτηκαν το 2020 - 2023

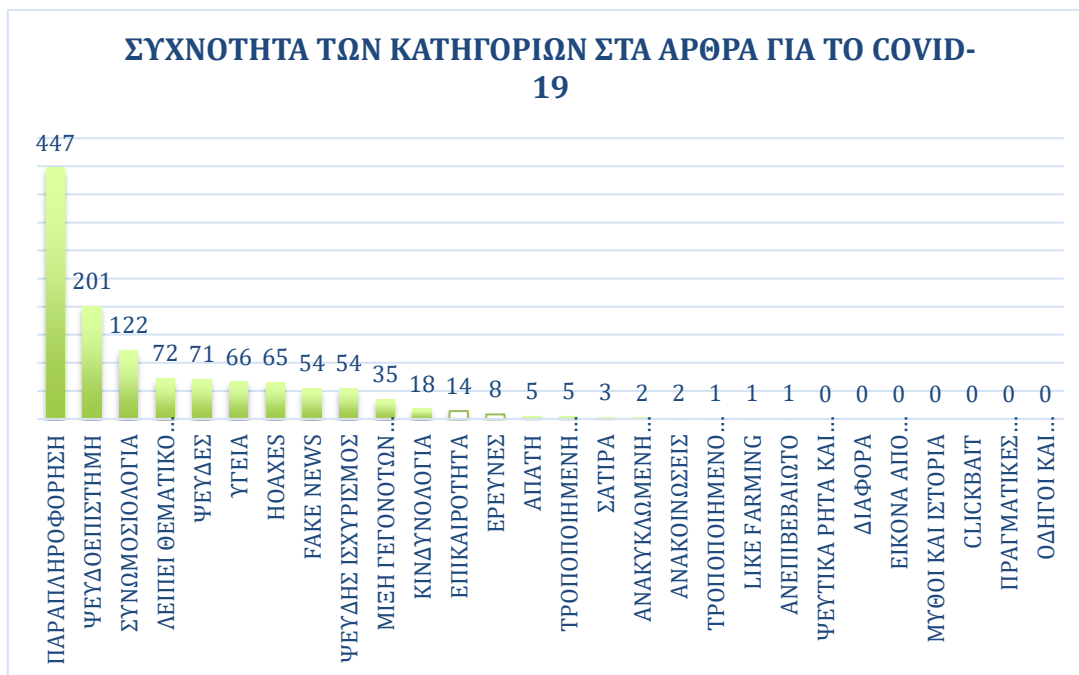


5.6.1.1 ΟΙ ΚΑΤΗΓΟΡΙΕΣ ΑΝΑ ΘΕΜΑ

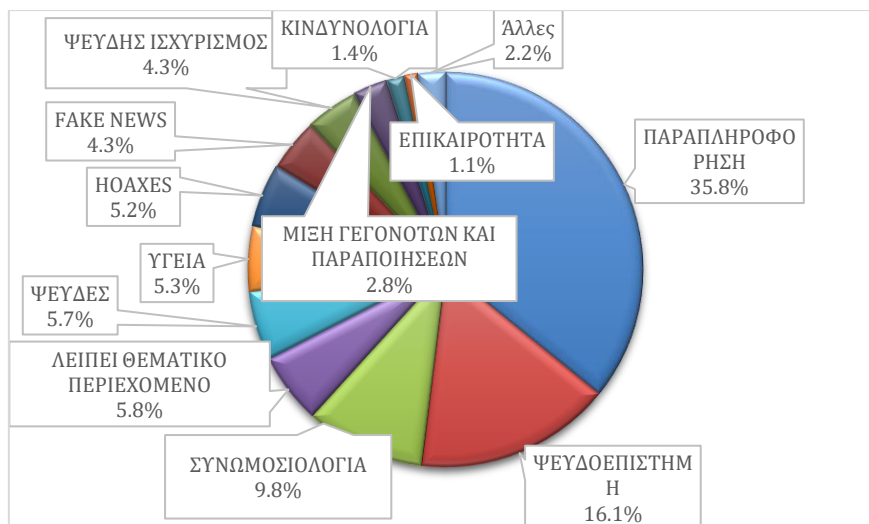
Ήταν πολύ ενδιαφέρον να αναλυθούν οι κατηγορίες ανά διαφορά θέματα.

Στα άρθρα που αφορούν το Covid-19 οι πιο συχνές κατηγορίες ήταν η «Παραπληροφόρηση» με μεγάλη διαφορά (το 35,8% όλων των άρθρων), η «Ψευδοεπιστήμη» και η «Συνωμοσιολογία» (Διαγράμματα 5.6.3 και 5.6.4).

Διάγραμμα 5.6.3 Κατηγορίες των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν το COVID-19 και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.4 Κατανομή των κατηγοριών των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν το COVID-19 και δημοσιεύτηκαν το 2013 - 2023

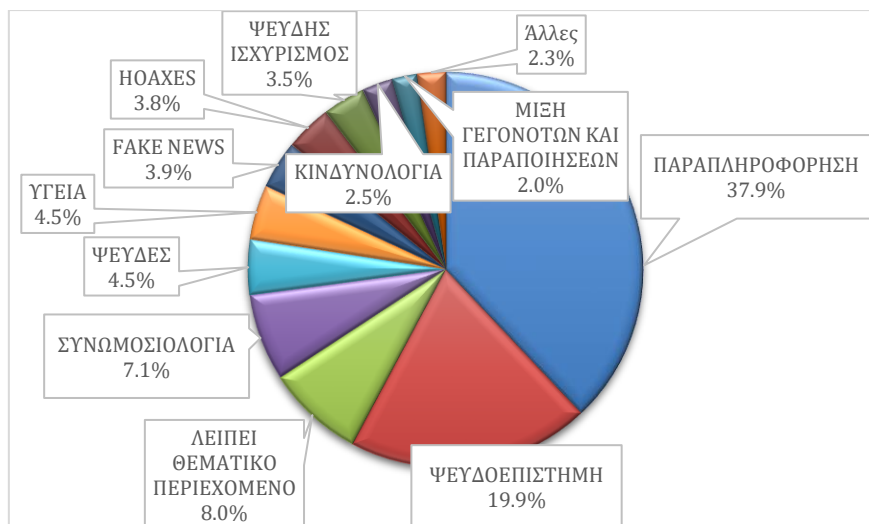


Στα άρθρα που αφορούν τον εμβολιασμό οι πιο συχνές κατηγορίες ήταν επίσης η «Παραπληροφόρηση» με μεγάλη διαφορά (το 37.9% όλων των άρθρων) και η «Ψευδοεπιστήμη» (το 19.9% όλων των άρθρων) (Διαγράμματα 5.6.5 και 5.6.6).

Διάγραμμα 5.6.5 Κατηγορίες των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν τον εμβολιασμό και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.6 Κατανομή των κατηγοριών των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν τον εμβολιασμό και δημοσιεύτηκαν το 2013 - 2023

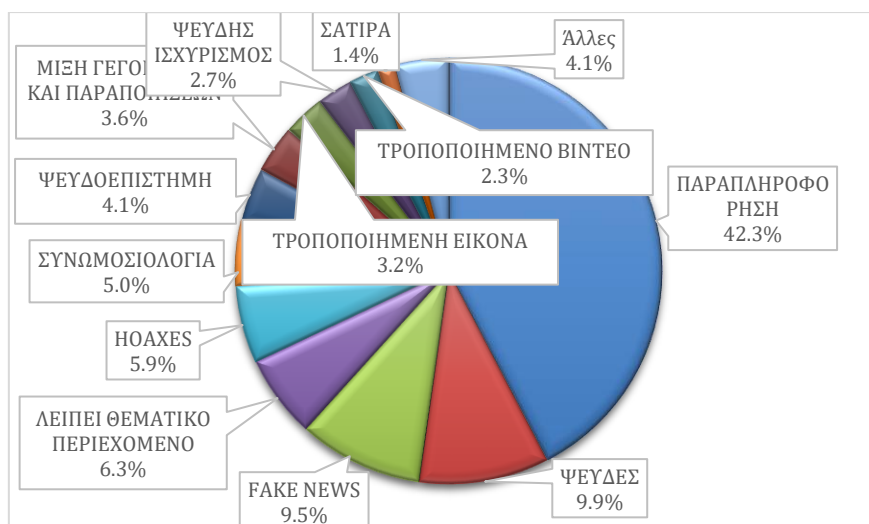


Στα άρθρα που αφορούν τη Ρωσία η πιο συχνή κατηγορία ήταν η «Παραπληροφόρηση» με μεγάλη διαφορά (το 42.3% όλων των άρθρων). Οι άλλες κατηγορίες σε σύγκριση με αυτήν δεν ήταν σχεδόν καθόλου δημοφιλείς (Διαγράμματα 5.6.7 και 5.6.8).

Διάγραμμα 5.6.7 Κατηγορίες των άρθρων στον Ellhnikahoaxes.gr, που αφορούν τη Ρωσία και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.8 Κατανομή των κατηγοριών των άρθρων στον Ellhnikahoaxes.gr, που αφορούν τη Ρωσία και δημοσιεύτηκαν το 2013 - 2023

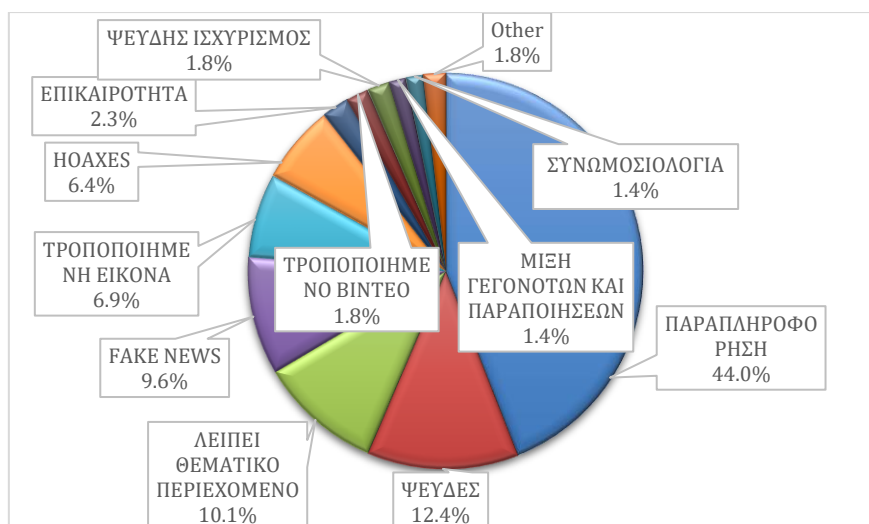


Στα άρθρα που αφορούν την Ουκρανία η πιο συχνή κατηγορία ήταν επίσης η «Παραπληροφόρηση» με μεγάλη διαφορά (το 44% όλων των άρθρων). Οι άλλες κατηγορίες σε σύγκριση με αυτήν δεν ήταν σχεδόν καθόλου δημοφιλείς, όπως και στα άρθρα που αφορούν τη Ρωσία (Διαγράμματα 5.6.9 και 5.6.10).

Διάγραμμα 5.6.9 Κατηγορίες των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Ουκρανία και δημοσιεύτηκαν το 2013 - 2023

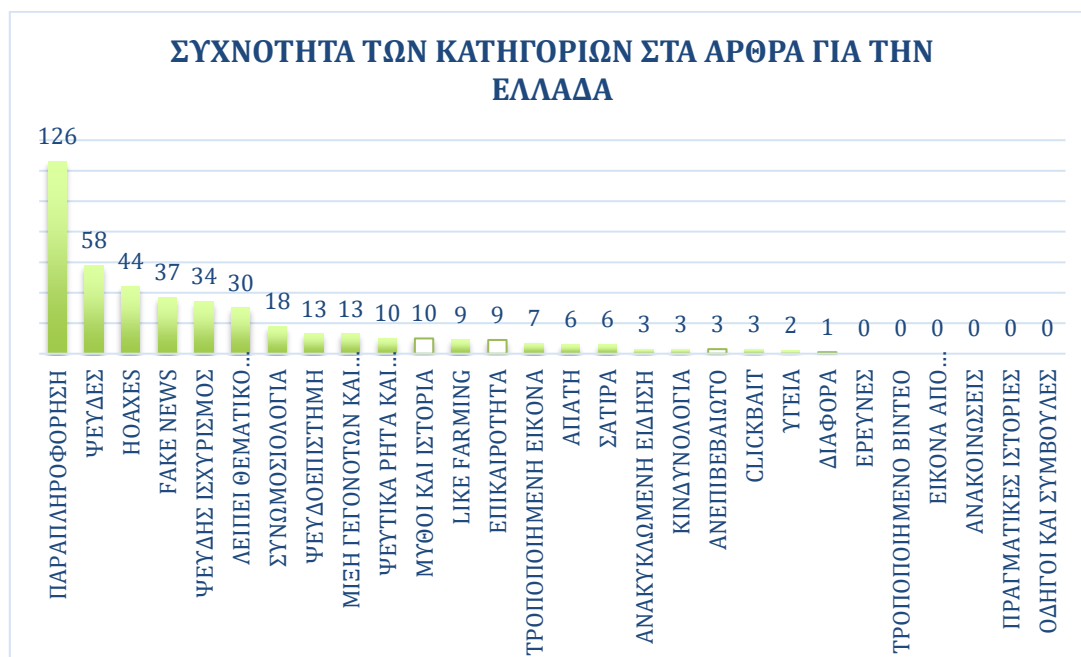


Διάγραμμα 5.6.10 Κατανομή των κατηγοριών των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Ουκρανία και δημοσιεύτηκαν το 2013 - 2023

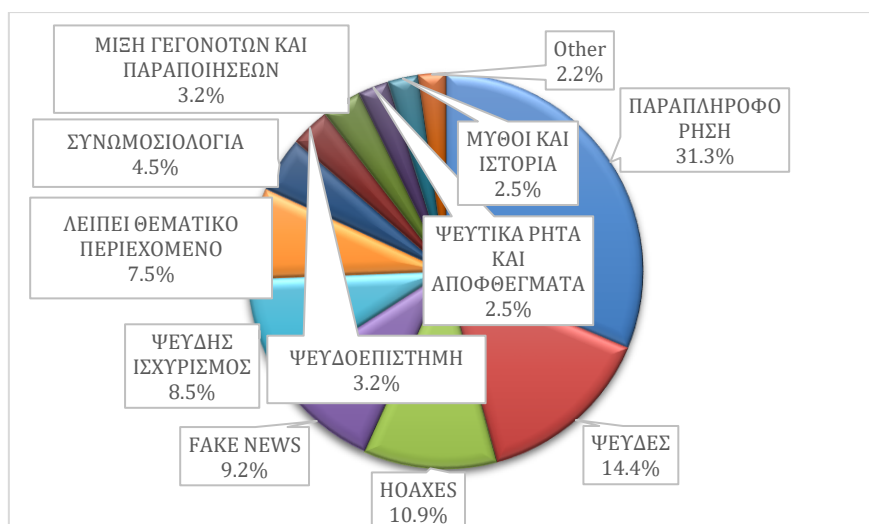


Στα άρθρα που αφορούν την Ελλάδα οι πιο συχνές κατηγορίες ήταν η «Παραπληροφόρηση» (το 31.3% όλων των άρθρων), η «Ψευδές» (το 14.4% όλων των άρθρων) και η «Hoaxes» (το 10.9% όλων των άρθρων) (Διαγράμματα 5.6.11 και 5.6.12).

Διάγραμμα 5.6.11 Κατηγορίες των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Ελλάδα και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.12 Κατανομή των κατηγοριών των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Ελλάδα και δημοσιεύτηκαν το 2013 - 2023

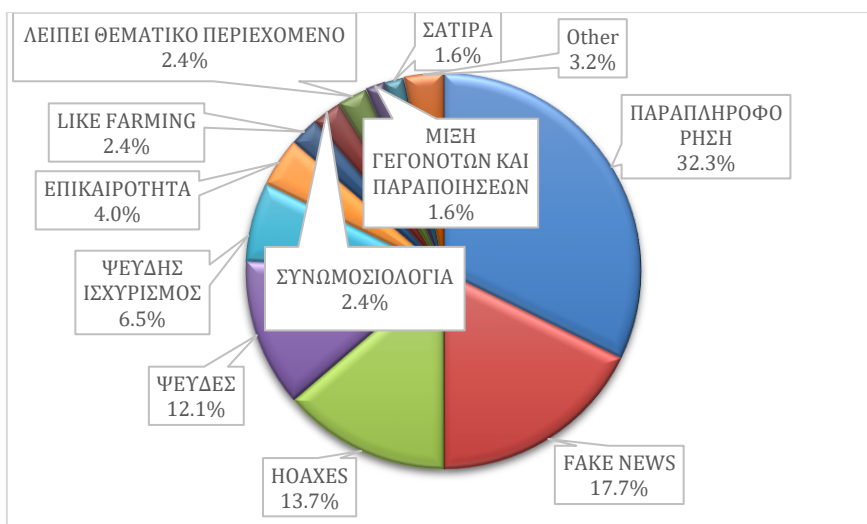


Στα άρθρα που αφορούν την Τουρκία οι πιο συχνές κατηγορίες ήταν η «Παραπληροφόρηση» (το 32.3% όλων των άρθρων), η «Fake news» (το 17.7% όλων των άρθρων), η «Hoaxes» (το 13.7% όλων των άρθρων) και η «Ψευδές» (το 12.1% όλων των άρθρων) (Διαγράμματα 5.6.13 και 5.6.14).

Διάγραμμα 5.6.13 Κατηγορίες των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Τουρκία και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.14 Κατανομή των κατηγοριών των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την Τουρκία και δημοσιεύτηκαν το 2013 - 2023

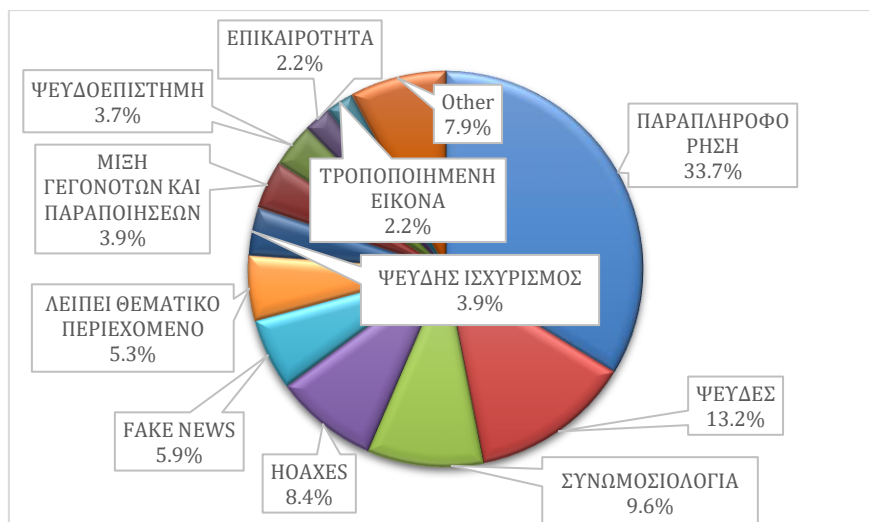


Στα άρθρα που αφορούν τις ΗΠΑ η πιο συχνή κατηγορία ήταν η «Παραπληροφόρηση» (το 33.7% όλων των άρθρων) με μεγάλη διαφορά (Διαγράμματα 5.6.15 και 5.6.16).

Διάγραμμα 5.6.15 Κατηγορίες των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν τις ΗΠΑ και δημοσιεύτηκαν το 2013 - 2023

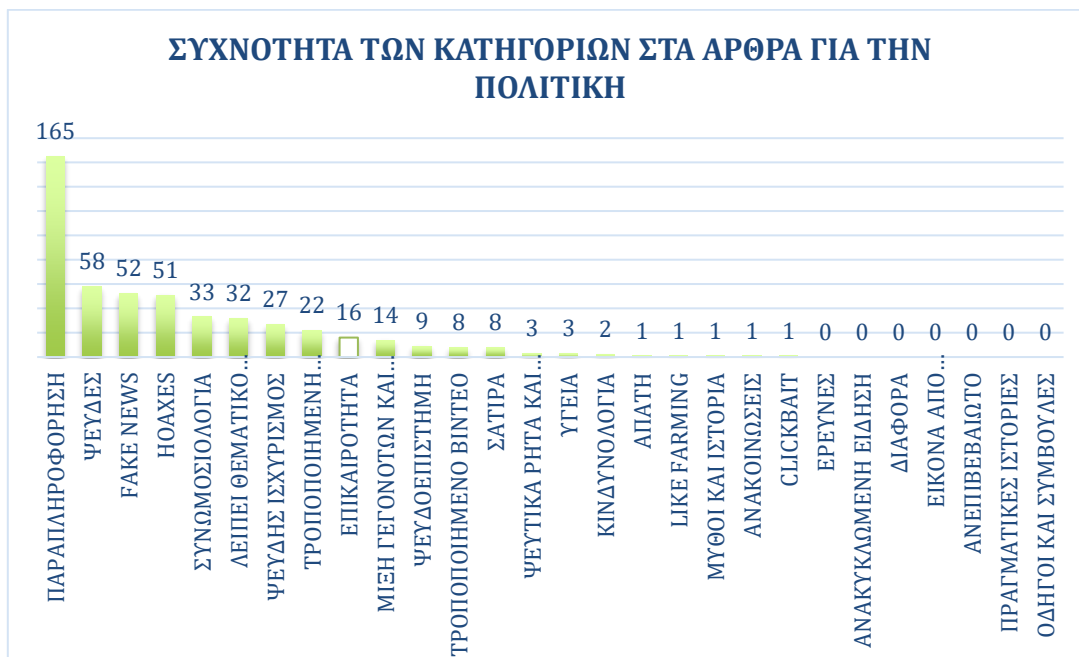


Διάγραμμα 5.6.16 Κατανομή των κατηγοριών των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν τις ΗΠΑ και δημοσιεύτηκαν το 2013 - 2023

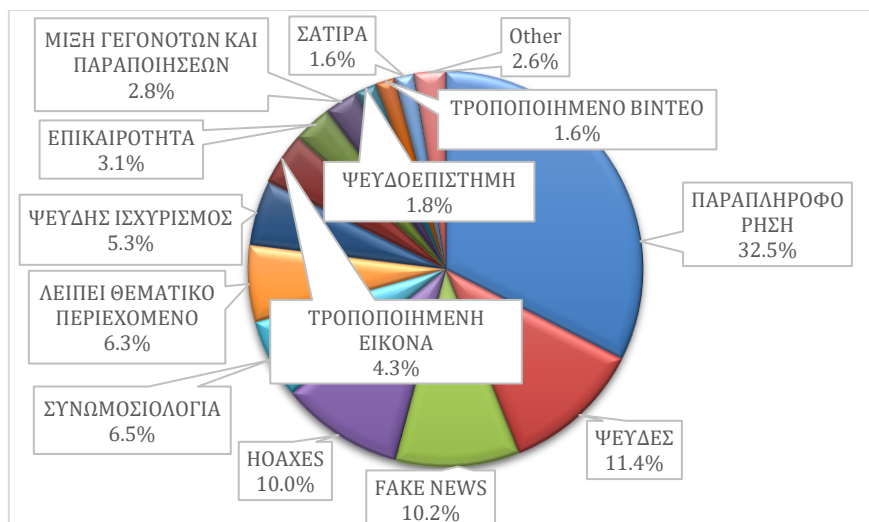


Τέλος, στα άρθρα που αφορούν την πολιτική η πιο συχνή κατηγορία ήταν η «Παραπληροφόρηση» (το 32.5% όλων των άρθρων). Οι κατηγορίες «Ψευδές», «Fake news» και «Hoaxes» υπάρχουν σε περίπου 10% των άρθρων η καθεμία (Διαγράμματα 5.6.17 και 5.6.18).

Διάγραμμα 5.6.17 Κατηγορίες των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την πολιτική και δημοσιεύτηκαν το 2013 - 2023



Διάγραμμα 5.6.18 Κατανομή των κατηγοριών των άρθρων στον ΕλληναHoaxes.gr, που αφορούν την πολιτική και δημοσιεύτηκαν το 2013 - 2023

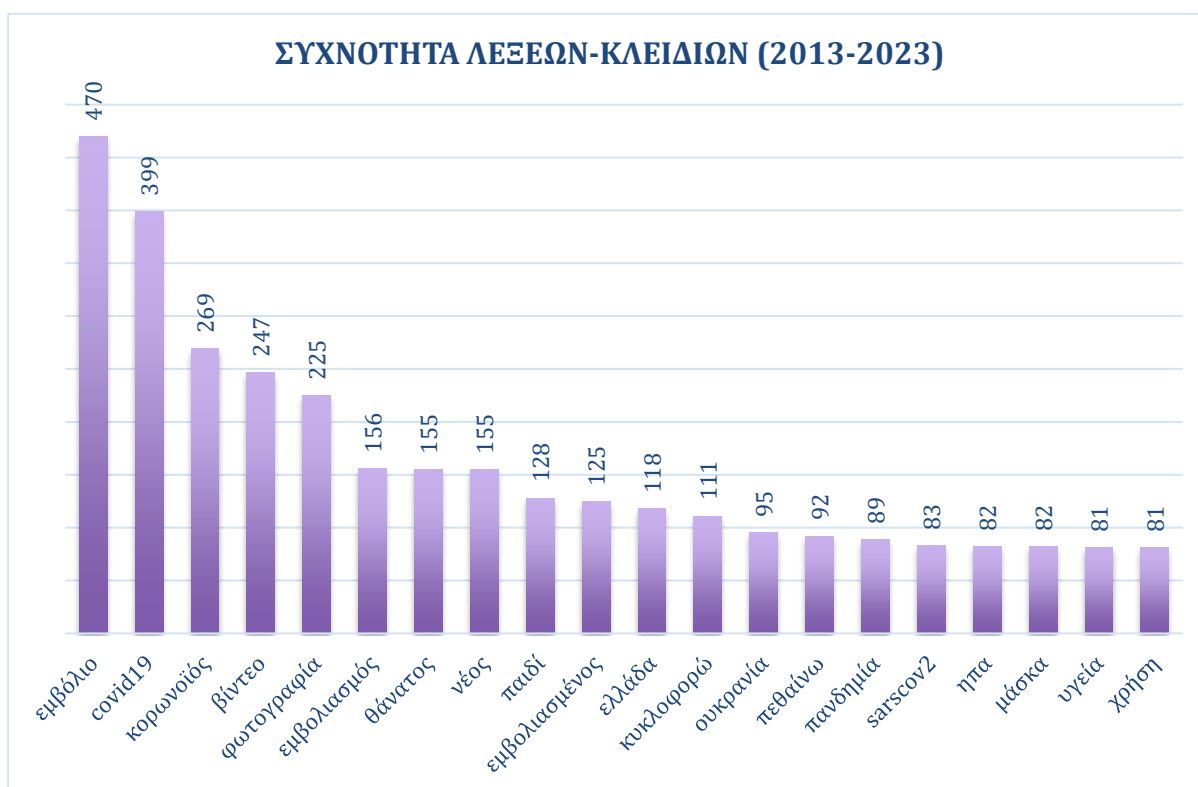


5.6.2 ΤΑ ΘΕΜΑΤΑ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ ΔΙΑΔΙΚΤΥΟ

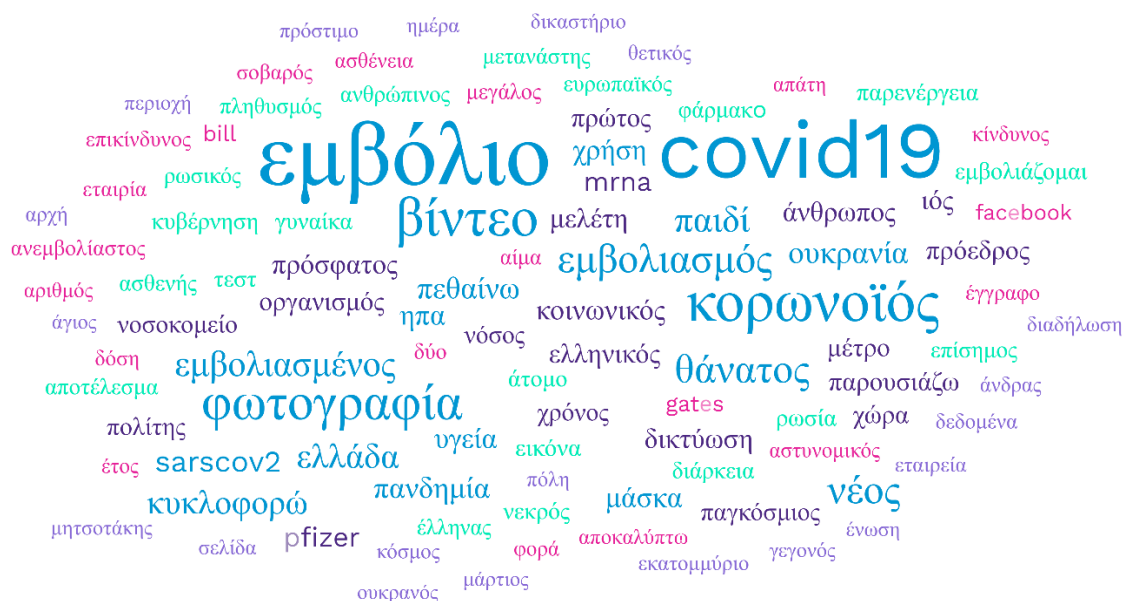
Παρότι η πανδημία του Covid-19 ξεκίνησε μόνο το 2020, αυτό το θέμα επηρέασε τον τομέα των ψευδών ειδήσεων πάρα πολύ με αποτέλεσμα στα θέματα ψευδών ειδήσεων τα θέματα του κορονοϊού και του εμβολιασμού να επικρατούν.

Βάση τα δεδομένα που συλλέχθηκαν από τον Ellhnikahoaxes.gr η πιο συχνές λέξεις-κλειδιά ήταν «εμβόλιο» (470 εμφανίσεις), «covid19» (399 εμφανίσεις), «κορωνοϊός» (269 εμφανίσεις), «βίντεο» (247 εμφανίσεις), «φωτογραφία» (225 εμφανίσεις), «εμβολιασμός» (156 εμφανίσεις), «θάνατος» (155 εμφανίσεις), «νέος» (155 εμφανίσεις), «παιδί» (128 εμφανίσεις), «εμβολιασμένος» (125 εμφανίσεις), «ελλάδα» (118 εμφανίσεις), «κυκλοφορώ» (111 εμφανίσεις), «ουκρανία» (95 εμφανίσεις), «πεθαίνω» (92 εμφανίσεις), «πανδημία» (89 εμφανίσεις), «sarscov2» (83 εμφανίσεις), «ηπα» (82 εμφανίσεις), «μάσκα» (82 εμφανίσεις), «υγεία» (81 εμφανίσεις), «χρήση» (81 εμφανίσεις) (Διάγραμμα 5.6.19 και Εικόνα 5.6.1).

Διάγραμμα 5.6.19 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της Ellhnikahoaxes.gr το 2013 - 2023

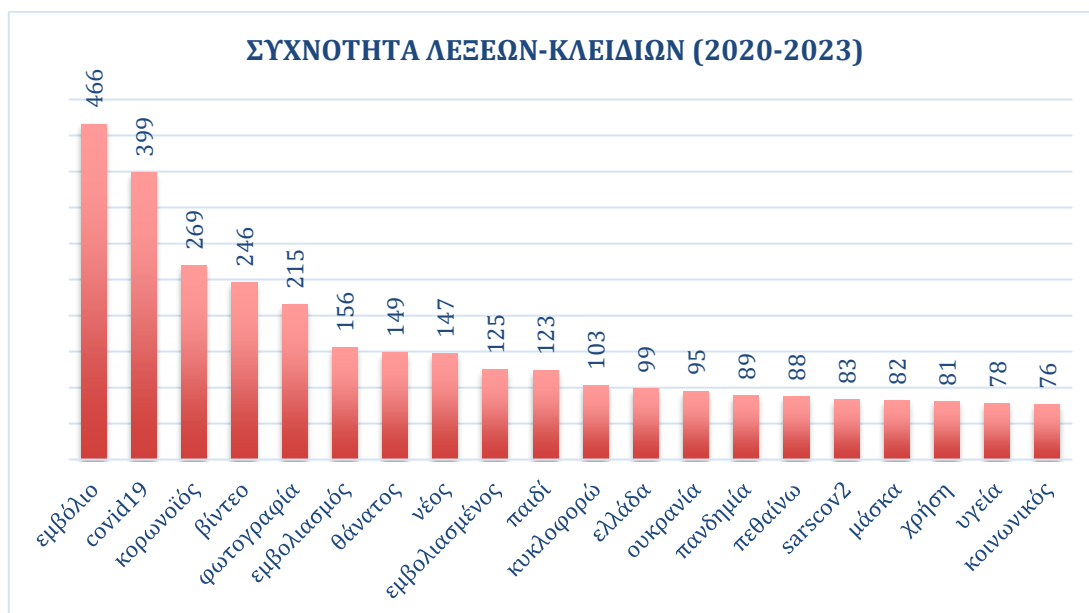


Εικόνα 5.6.1 Word-cloud των ισχυρισμών στον EllhnikaHoaxes.gr



Αν ληφθούν υπόψιν μόνο τα άρθρα που δημοσιεύτηκαν το 2020 – 2023, τα στατιστικά δεδομένα δεν αλλάζουν πολύ (Διάγραμμα 5.6.20). Ο λόγος είναι η εξαιρετικά διαφορετική δομή των άρθρων πριν το 2020, όπου οι ισχυρισμοί υπήρχαν μόνο σε περίπου 6% των άρθρων.

Διάγραμμα 5.6.20 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της EllhnikaHoaxes.gr το 2020 - 2023



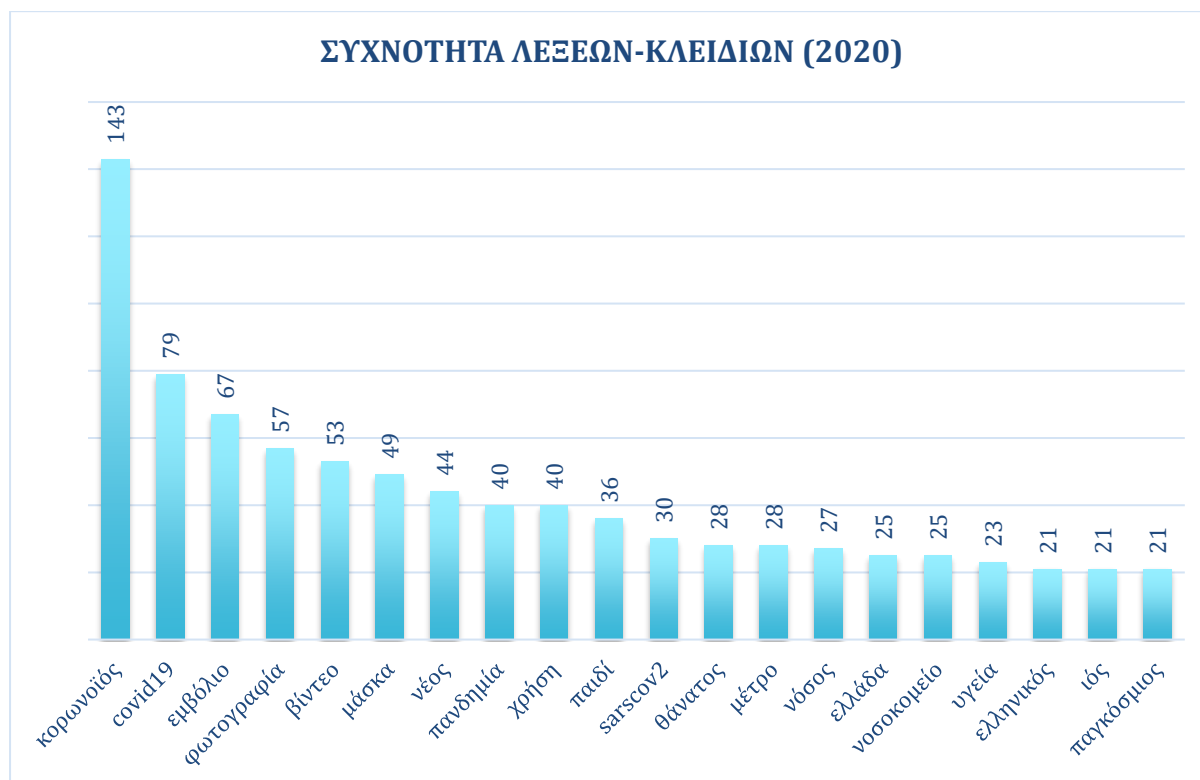
5.6.2.1 ΤΑ ΘΕΜΑΤΑ ΑΝΑ ΧΡΟΝΙΑ

Αναλύοντας τις λέξεις-κλειδιά της περιόδου 2020 – 2023 ανά χρονιά, παρατηρείται ότι το πιο δημοφιλές θέμα και τότε ήταν ο κορονοϊός και ο εμβολιασμός με πιο πολλές εμφανίσεις της λέξης-κλειδιού «εμβόλιο» το 2021 (277 εμφανίσεις) (Διαγράμματα 5.6.21 – 5.6.24).

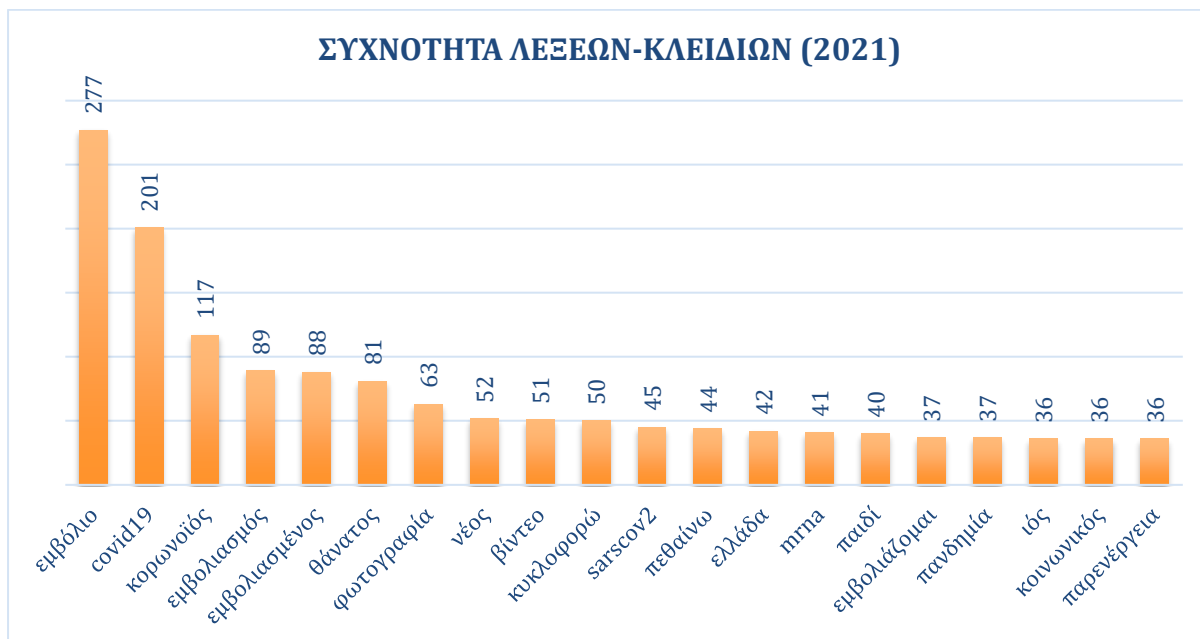
Στα στατιστικά δεδομένα του 2022 και του 2023 εμφανίζονται οι λέξεις-κλειδιά που αφορούν τη Ρωσία και την Ουκρανία, πχ «ουκρανία», «ρωσία», «ουκρανός», «ρωσικός».

Το 2023 υπάρχουν πολλές εμφανίσεις των λέξεων-κλειδιών που αφορούν την τραγωδία στην Τέμπη, πχ «τέμπη», «αμαξοστοιχία» και «δυστύχημα». Επίσης, το 2023 είναι η πρώτη χρονιά με την έξαρση της πανδημίας του κορονοϊού, όταν η πιο συχνή λέξη-κλειδί δε σχετίζεται με την πανδημία (η λέξη-κλειδί «βίντεο»).

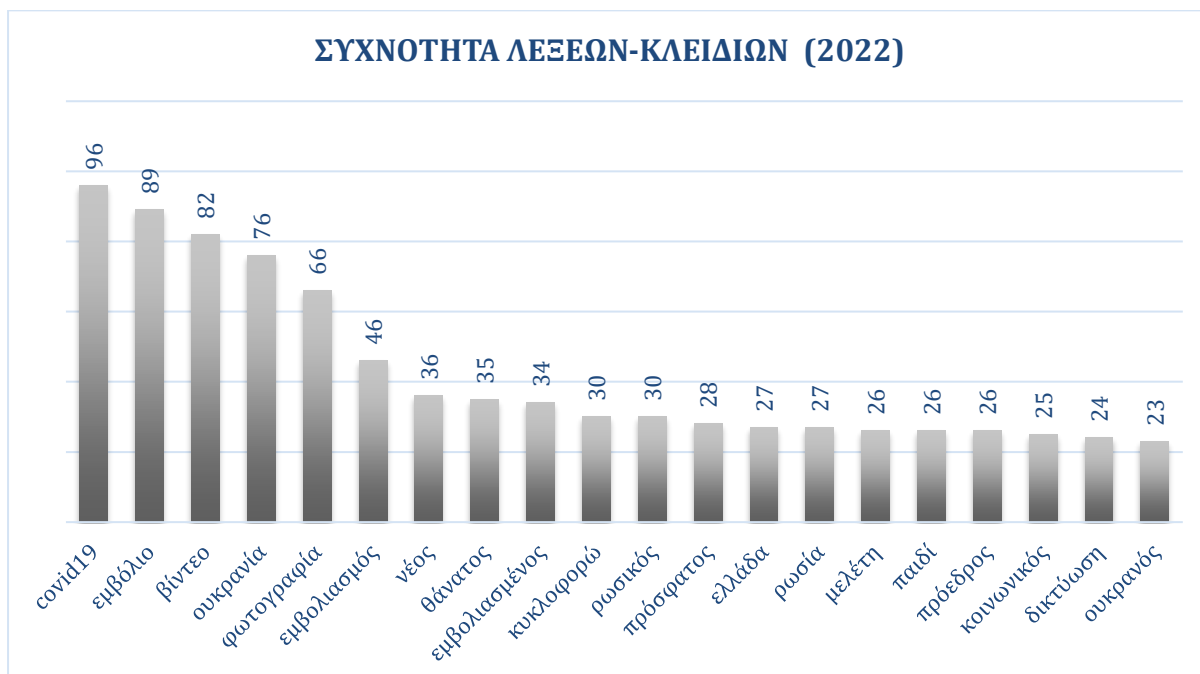
Διάγραμμα 5.6.21 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της Ellhnikahoaxes.gr το 2020



Διάγραμμα 5.6.22 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της Ellhnikahoaxes.gr το 2021



Διάγραμμα 5.6.23 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της Ellhnikahoaxes.gr το 2022



Διάγραμμα 5.6.24 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της Ellhnikahoaxes.gr το 2023



5.6.2.2 ΤΑ ΘΕΜΑΤΑ ΑΝΑ ΚΑΤΗΓΟΡΙΑ

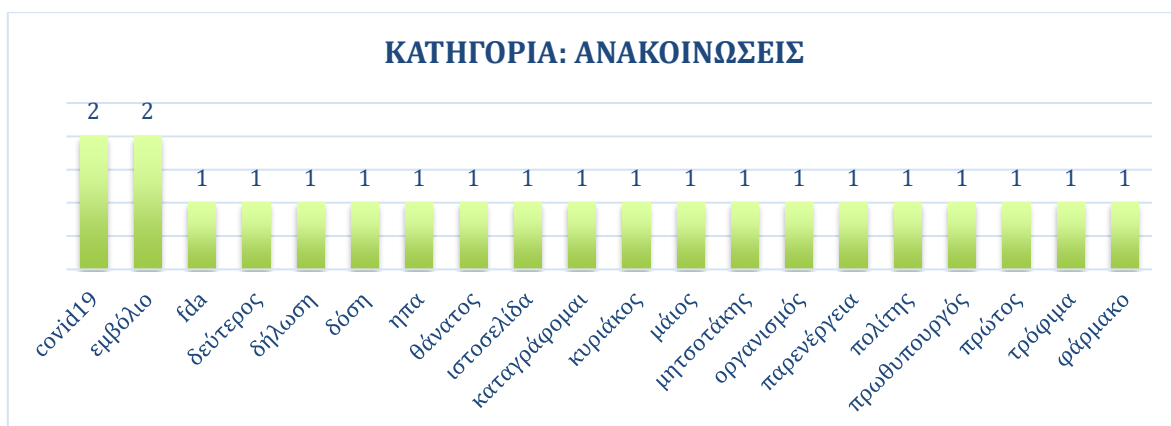
Αναλύοντας τα θέματα ανά κατηγορία, παρατηρείται ότι σ

- Στα άρθρα των κατηγοριών «Ανακοινώσεις», «Έρευνες», «Fake news», «Hoaxes», «Κινδυνολογία», «Λείπει θεματικό περιεχόμενο», «Μίξη γεγονότων και παραποιήσεων», «Παραπληροφόρηση», «Ψευδής ισχυρισμός», «Ψευδοεπιστήμη», «Συνωμοσιολογία» και «Υγεία» επικρατεί το θέμα του Covid-19 και του εμβολιασμού.
- Στα άρθρα της κατηγορίας «Απάτη» επικρατεί το θέμα των διαγωνισμών στα μέσα κοινωνικής δικτύωσης, πχ «σελίδα», «facebook», «διαγωνισμός», «κερδίζω», «πλατφόρμα», «χρήστης», «διαδικτυακός» και «χαρίζω».
- Στα άρθρα της κατηγορίας «Like farming» επικρατούν τα θέματα των φωτογραφιών και των μέσων κοινωνικής δικτύωσης. Επίσης, υπάρχουν πολλές εμφανίσεις των λέξεων-κλειδιών «γυναίκα» και «παιδί».

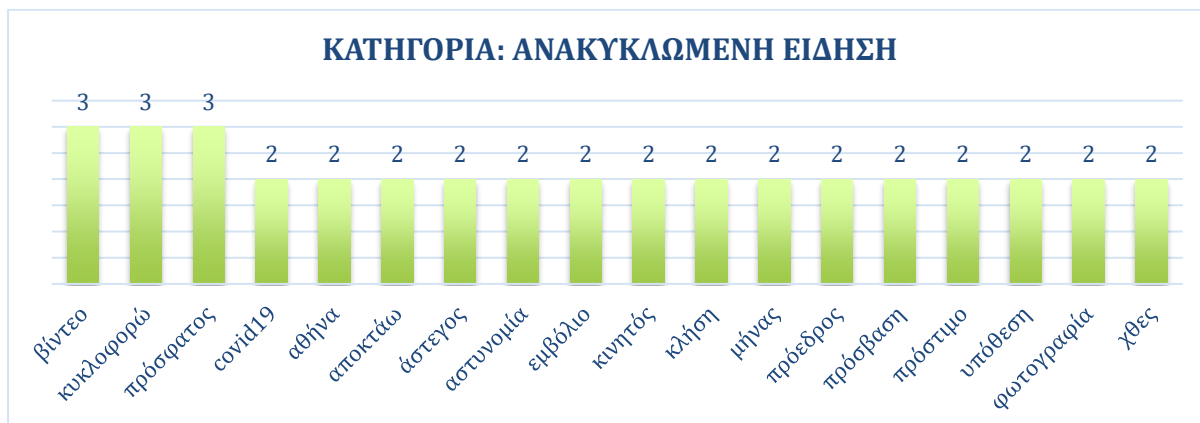
- Στα άρθρα της κατηγορίας «Ψευδές» επικρατούν τα θέματα του οπτικού και του ακουστικού υλικού, πχ «φωτογραφία» και «βίντεο», του Covid-19 και της Ουκρανίας.
- Στα άρθρα της κατηγορίας «Ψεύτικα ρητά και αποφθέγματα» επικρατούν τα θέματα της ιστορίας και των διάφορων ισχυρισμών, πχ «κείμενο», «σωκράτης», «φράση», «ιστορία», «τίτλος», «έλληνας» και «ποίηση».
- Στα άρθρα της κατηγορίας «Τροποποιημένη εικόνα» επικρατούν τα θέματα των εικόνων και των πολιτικών, πχ «φωτογραφία», «εικόνα», «πρόεδρος», «μητσοτάκης», «ουκρανία», «κυριάκος», «πρωθυπουργός», «ρεπορτάζ», «είδηση», «ζελένσκι» και «κυκλοφορώ».
- Στα άρθρα της κατηγορίας «Τροποποιημένο βίντεο» επικρατεί το θέμα των βίντεο και των πολιτικών, πχ «βίντεο», «πρόεδρος», «μπάιντεν», «ηπα», «ουκρανία» και «ομιλία».
- Στα άρθρα των κατηγοριών «Ανακυκλωμένη είδηση», «Ανεπιβεβαίωτο», «Clickbait», «Διάφορα», «Επικαιρότητα» και «Σάτιρα» είναι δύσκολο να επιλεγεί ένα επικρατούν θέμα, διότι λίγα από αυτά τα άρθρα περιλαμβάνουν ισχυρισμό.
- Τέλος, στα άρθρα της κατηγορίας «Μύθοι και ιστορία» το κύριο θέμα είναι η αρχαία ιστορία της Ελλάδας.

Τα στατιστικά δεδομένα κάθε κατηγορίας παρουσιάζονται στα Διαγράμματα 5.6.25 – 5.6.49.

Διάγραμμα 5.6.25 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανακοινώσεις» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.26 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανακυκλωμένη είδηση» της Ellhnikahoaxes.gr το 2013 - 2023



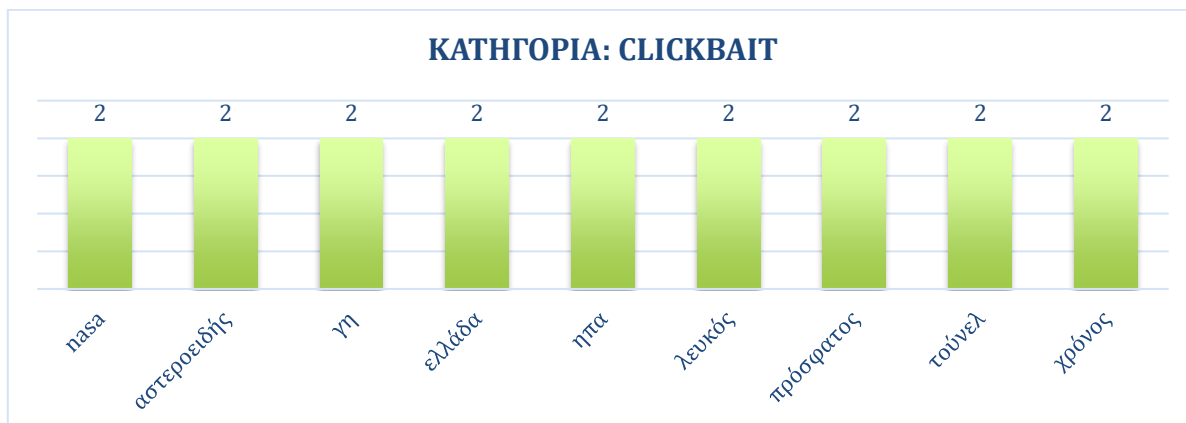
Διάγραμμα 5.6.27 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ανεπιβεβαίωτο» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.28 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Απάτη» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.29 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Clickbait» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.30 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Διάφορα» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.31 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Επικαιρότητα» της EllhnikaHoaxes.gr το 2013 - 2023



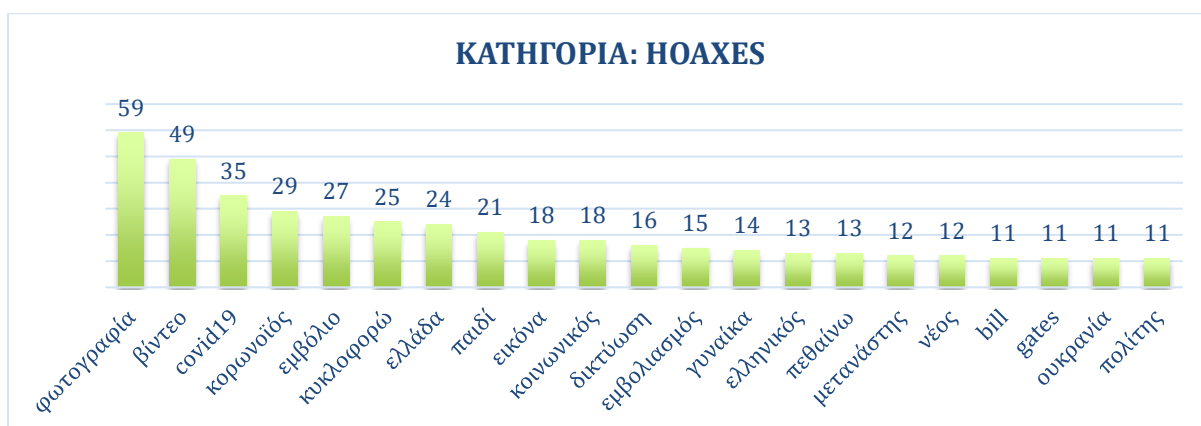
Διάγραμμα 5.6.32 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Έρευνες» της EllhnikaHoaxes.gr το 2013 - 2023



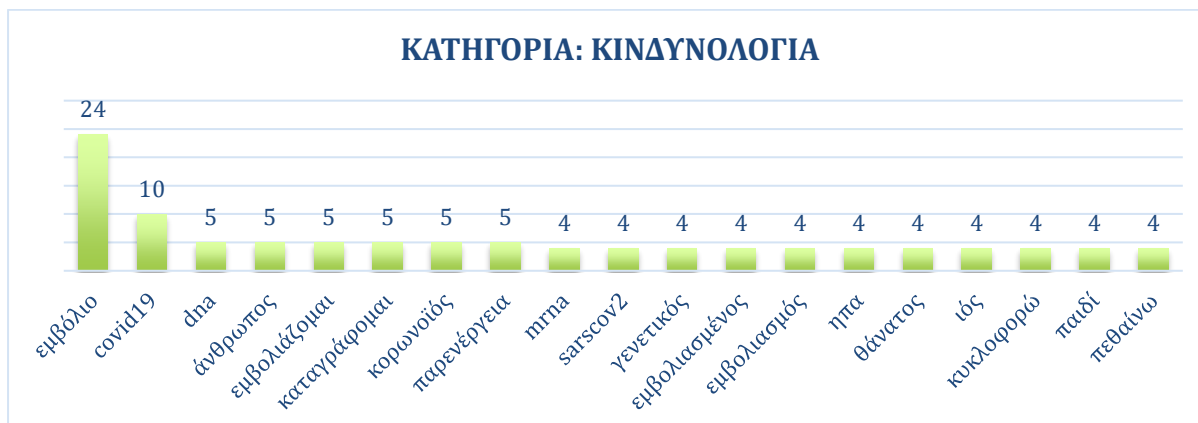
Διάγραμμα 5.6.33 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Fake news» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.34 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Hoaxes» της EllhnikaHoaxes.gr το 2013 - 2023



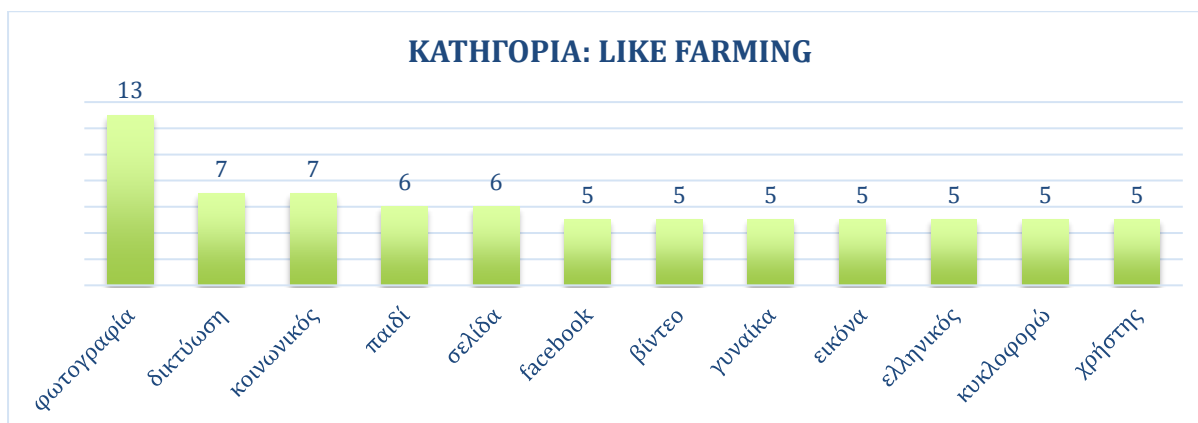
Διάγραμμα 5.6.35 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Κινδυνολογία» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.36 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Λείπει θεματικό περιεχόμενο» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.37 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Like farming» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.38 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Μίξη γεγονότων και παραποιήσεων» της ΕλληνικάHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.39 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Μύθοι και ιστορία» της ΕλληνικάHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.40 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Παραπληροφόρηση» της ΕλληνικάHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.41 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδές» της EllhnikaHoaxes.gr το 2013 - 2023



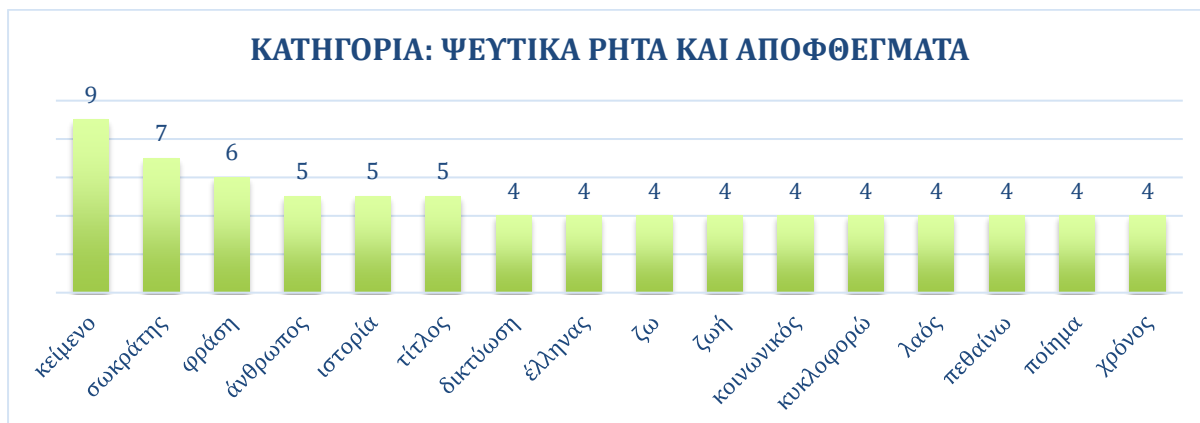
Διάγραμμα 5.6.42 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδείς ισχυρισμός» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.43 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψευδοεπιστήμη» της EllhnikaHoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.44 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Ψεύτικα ρητά και αποφθέγματα» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.45 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Σάτιρα» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.46 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Συνωμοσιολογία» της Ellhnikahoaxes.gr το 2013 - 2023



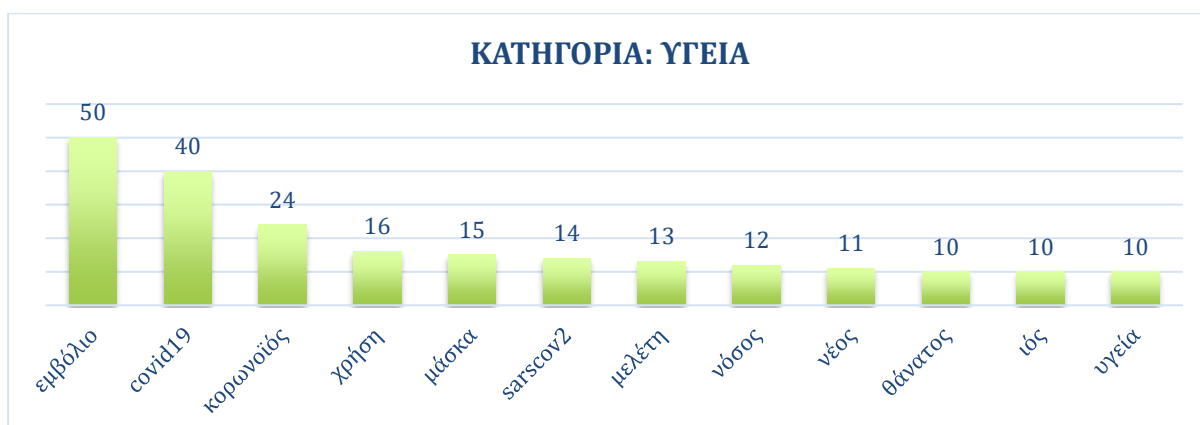
Διάγραμμα 5.6.47 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Τροποποιημένη εικόνα» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.48 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Τροποποιημένο βίντεο» της Ellhnikahoaxes.gr το 2013 - 2023



Διάγραμμα 5.6.49 Λέξεις-κλειδιά που αναφέρθηκαν περισσότερο στα άρθρα της κατηγορίας «Υγεία» της Ellhnikahoaxes.gr το 2013 - 2023



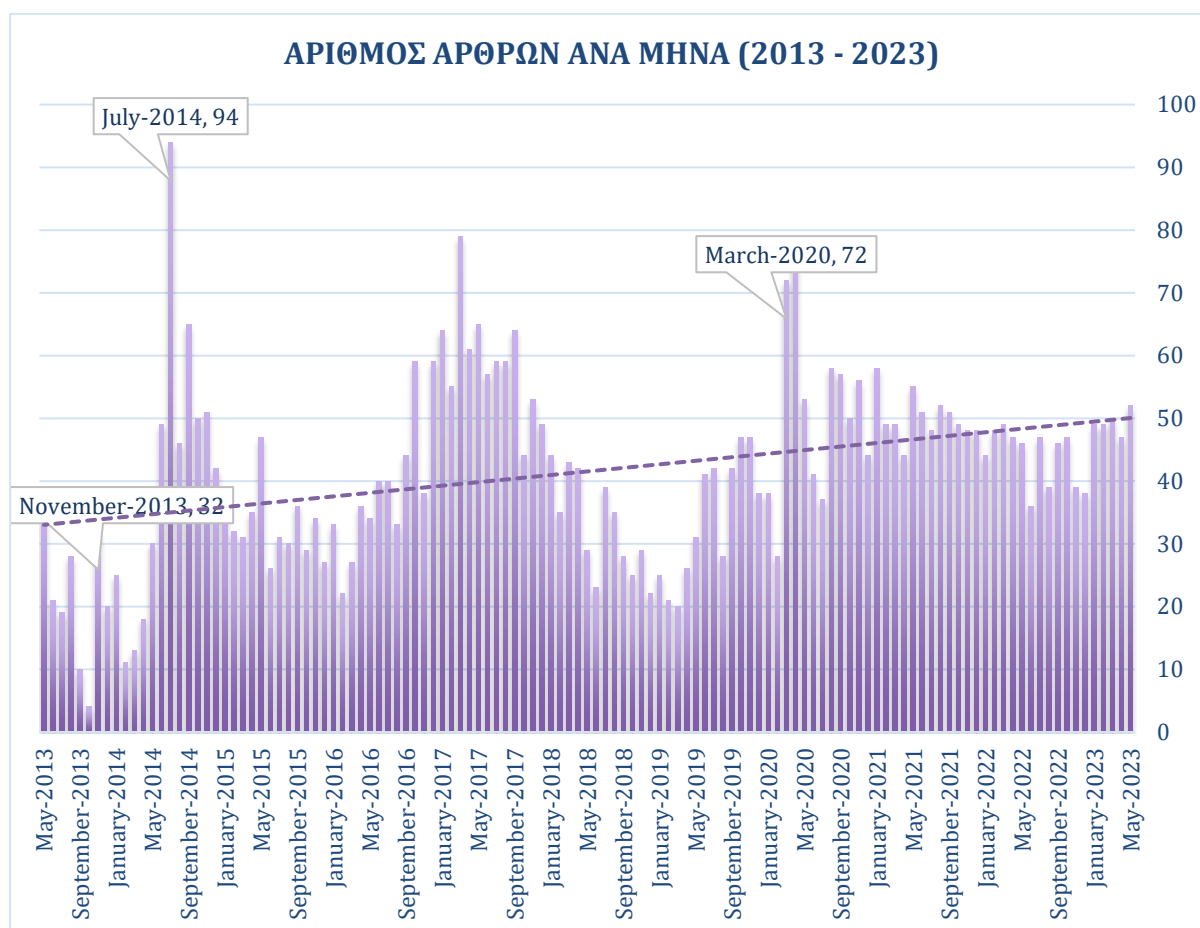
5.6.3 ΠΕΡΙΟΔΟΙ ΕΞΑΡΣΗΣ ΤΩΝ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΤΟ ΕΛΛΗΝΙΚΟ

ΔΙΑΔΙΚΤΥΟ

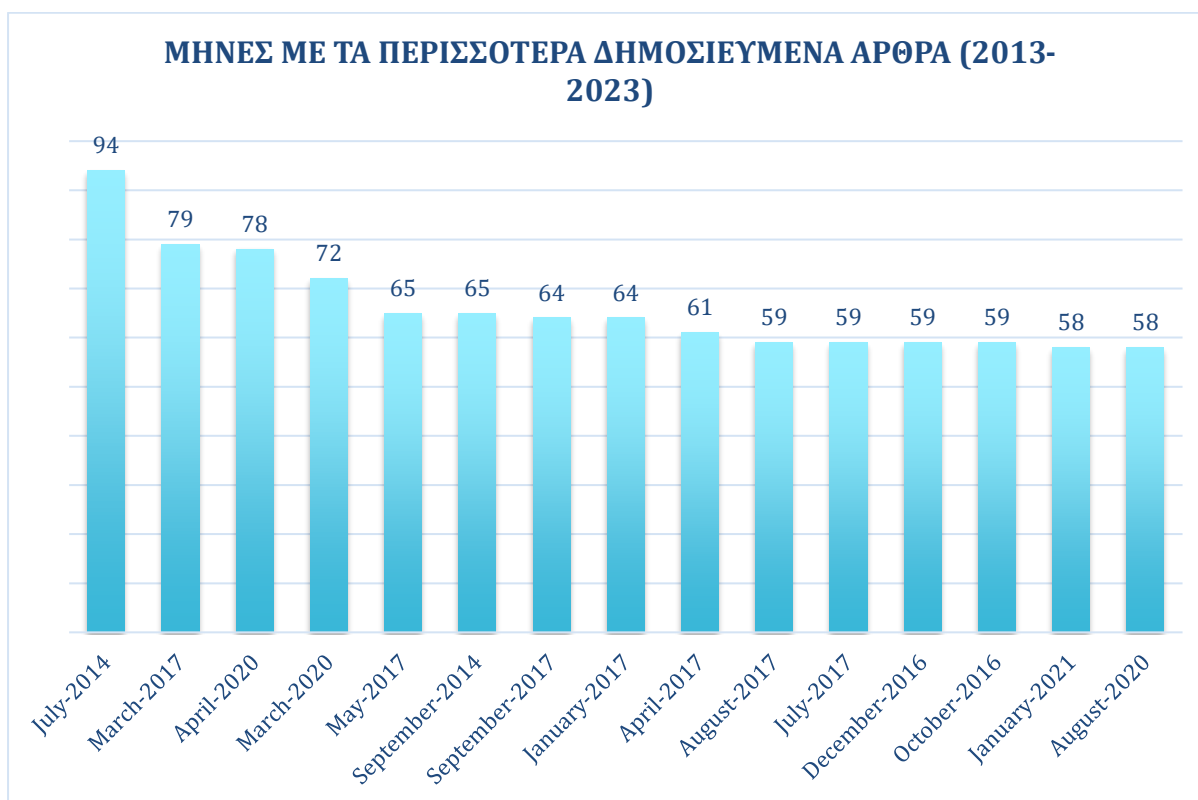
Η υπηρεσία Ellhnikahoaxes.gr ξεκίνησε να λειτουργεί το 2013 και από τότε ο αριθμός των άρθρων κάθε μήνα διπλασιάστηκε. Η χρονοσειρά των μηνιαίων δημοσιεύσεων ακολουθεί ανοδική τάση με ακανόνιστες συνιστώσες, δηλαδή απότομες αυξομειώσεις που οφείλονται σε απρόβλεπτα γεγονότα, το Νοέμβριο 2013, τον Ιούλιο-Αύγουστο 2014 και το Μάρτιο-Μάιο 2020. Δεν υπάρχει εποχικότητα (Διάγραμμα 5.6.50).

Ο μήνας με τις περισσότερες δημοσιεύσεις ήταν Ιούλιος 2014 (Διάγραμμα 5.6.51).

Διάγραμμα 5.6.50 Αριθμός των δημοσιευμένων άρθρων της Ellhnikahoaxes.gr ανά μήνα το 2013 - 2023



Διάγραμμα 5.6.51 Μήνες με τα περισσότερα δημοσιευμένα άρθρα στον ΕλληναHoaxes.gr το 2013 - 2023



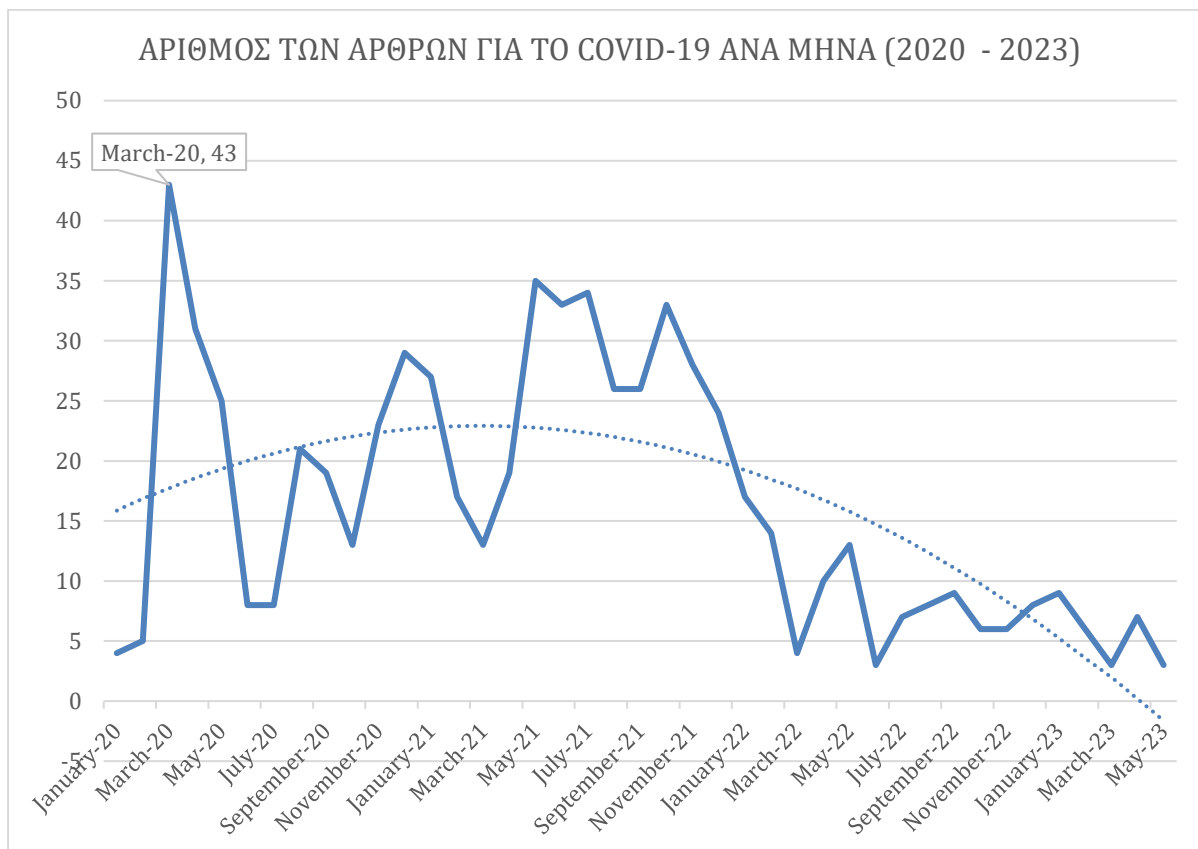
5.6.3.1 ΟΙ ΠΕΡΙΟΔΟΙ ΕΞΑΡΣΗΣ ΑΝΑ ΘΕΜΑ

Επειδή μέχρι το 2019 μόνο το 6% των άρθρων περιείχε ισχυρισμούς, τα στατιστικά δεδομένα πριν το 2020 δε λαμβάνονται υπόψιν.

Λαμβάνοντας υπόψιν αυτή τη λεπτομέρεια, τα πιο δημοφιλή θέματα στην Ελλάδα έχουν το παρακάτω ιστορικό:

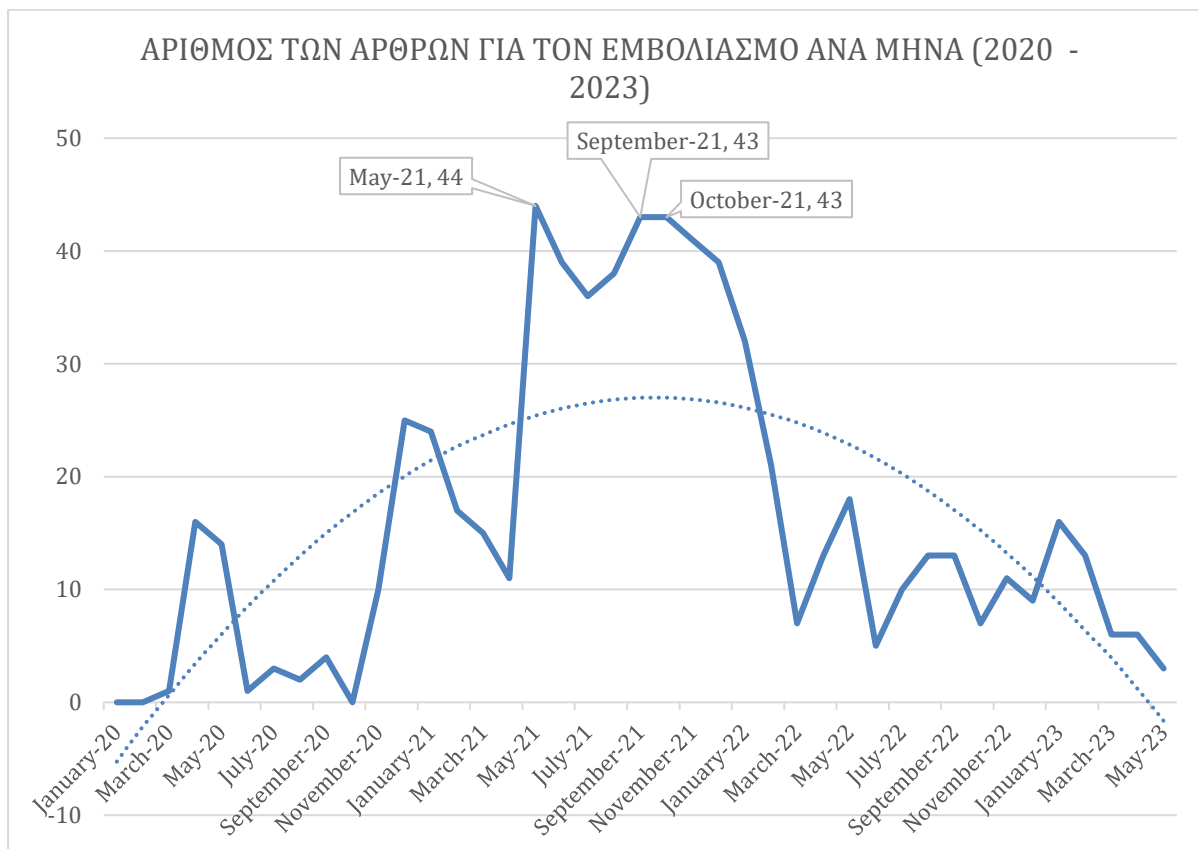
- Το θέμα του Covid-19 ήταν η πιο δημοφιλής το Μάρτιο 2020 (43 άρθρα). Το Νοέμβριο 2021 αυτό το θέμα ξεκίνησε να αναφέρεται λιγότερα, φτάνοντας το ελάχιστο το καλοκαίρι 2022 (Διάγραμμα 5.6.52).

Διάγραμμα 5.6.52 Αριθμός των άρθρων στον ΕλληναHoaxes.gr, που αφορούν το COVID-19, ανά μήνα από το 2020 έως 2023



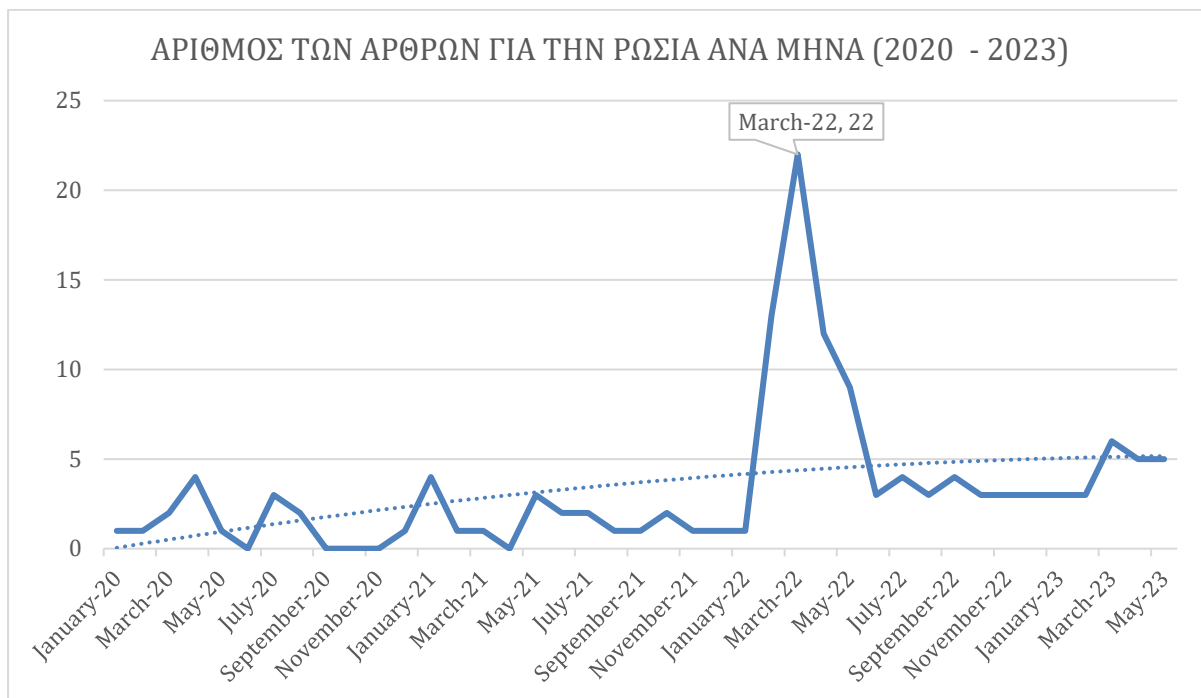
- Τον Απρίλιο 2020 ήταν η πρώτη αύξηση της δημοτικότητας του θέματος του εμβολιασμού, ωστόσο τον Οκτώβριο 2020 δεν αναφέρθηκε ούτε μια φορά. Η δεύτερη απότομη αύξηση ήταν το Δεκέμβριο 2020, αλλά μετά ξεκίνησε να μειώνεται έως τον Απρίλιο 2021. Το Μάιο 2021 το θέμα αυτό έφτασε το μέγιστο σημείο (44 άρθρα), παραμένοντας πολύ δημοφιλές έως το Δεκέμβριο 2021. Ο αριθμός των άρθρων πάνω σε αυτό το θέμα έπειτα μειώθηκε κατά πέντε φορές, παραμένοντας μετά στο επίπεδο των 10 άρθρων το μήνα (Διάγραμμα 5.6.53).

Διάγραμμα 5.6.53 Αριθμός των άρθρων στον ΕλληνικάΗοαxes.gr, που αφορούν τον εμβολιασμό, ανά μήνα από το 2020 έως 2023

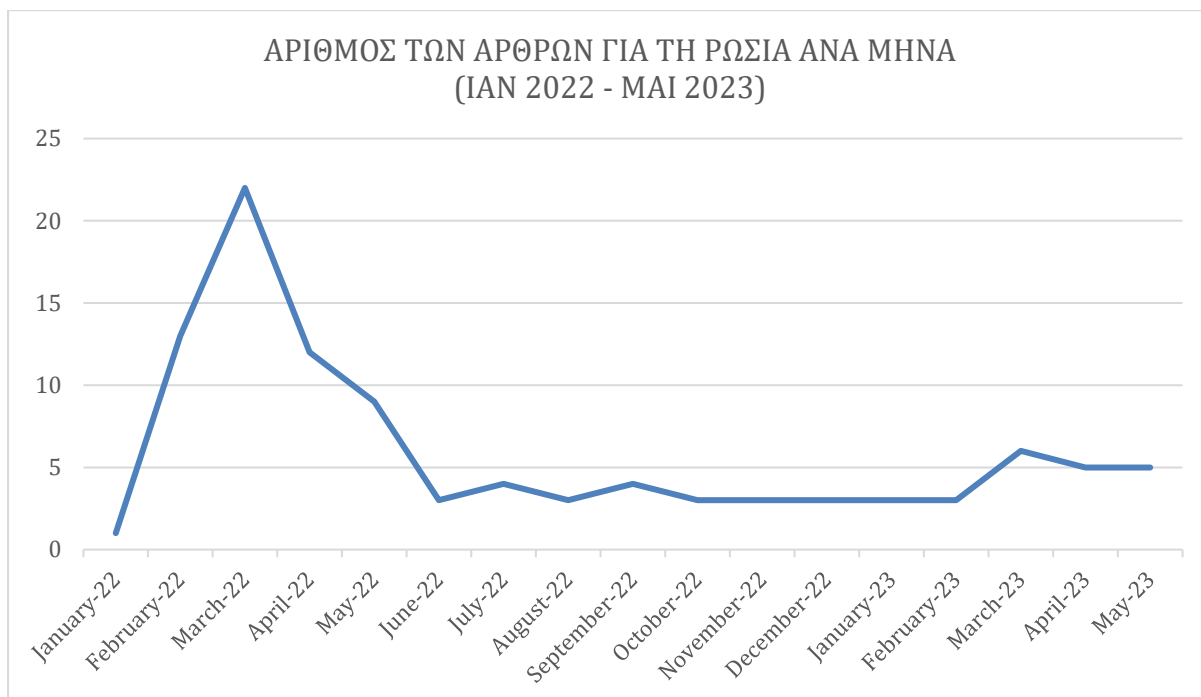


- Η δημοτικότητα του θέματος της Ρωσίας απότομα ξεκίνησε να αυξάνεται το Φεβρουάριο 2022, φτάνοντας το μέγιστο σημείο το Μάρτιο 2022 (22 άρθρα). Τον Απρίλιο 2022 η δημοτικότητα του θέματος άρχισε να μειώνεται, φτάνοντας το σημείο των τριών δημοσιεύσεων το μήνα τον Ιούνιο 2022. Από τότε ο αριθμός των άρθρων για τη Ρωσία δεν άλλαξε πολύ, παραμένοντας στο μέσο επίπεδο των τεσσάρων άρθρων κάθε μήνα (Διαγράμματα 5.6.54 και 5.6.55).

Διάγραμμα 5.6.54 Αριθμός των άρθρων στον ΕλληνικαΗοαχες.gr, που αφορούν τη Ρωσία, ανά μήνα από το 2020 έως 2023



Διάγραμμα 5.6.55 Αριθμός των άρθρων στον ΕλληνικαΗοαχες.gr, που αφορούν τη Ρωσία, ανά μήνα από τον Ιανουάριο 2022 έως το Μάιο 2023

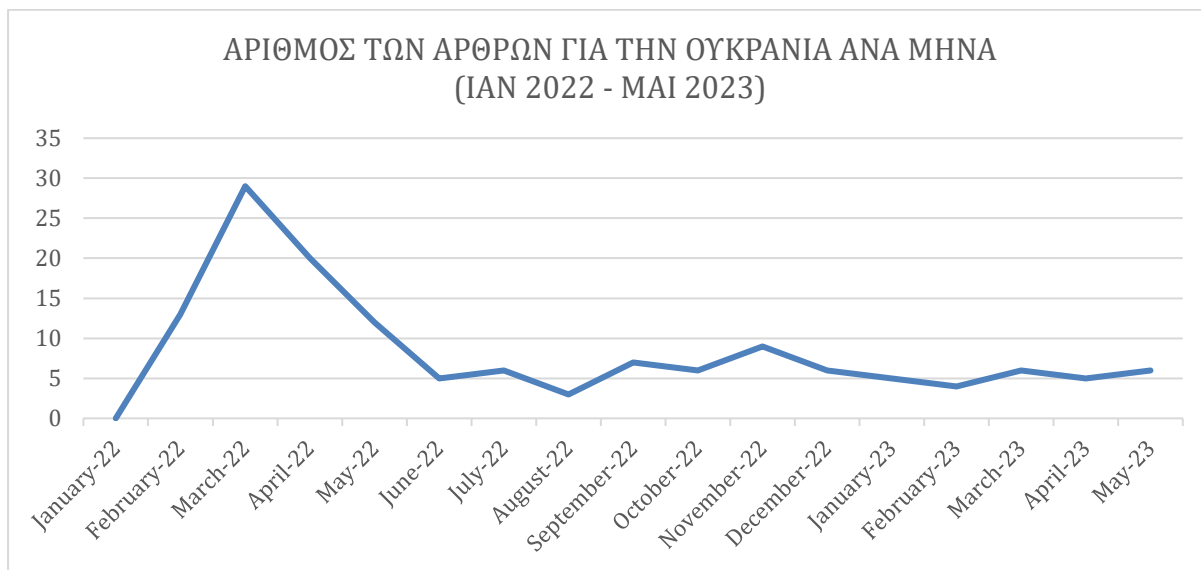


- Η δημοτικότητα των άρθρων για την Ουκρανία συσχετίζεται με τη δημοτικότητα των άρθρων για τη Ρωσία. Η δημοτικότητα αυτού του θέματος απότομα ξεκίνησε να αυξάνεται το Φεβρουάριο 2022, φτάνοντας το μέγιστο σημείο το Μάρτιο 2022 (29 άρθρα). Τον Απρίλιο 2022 η δημοτικότητα του θέματος άρχισε να μειώνεται, φτάνοντας το σημείο των πέντε δημοσιεύσεων το μήνα τον Ιούνιο 2022. Από τότε ο αριθμός των άρθρων για την Ουκρανία δεν άλλαξε πολύ, παραμένοντας στο μέσο επίπεδο των έξι άρθρων κάθε μήνα (Διαγράμματα 5.6.56 και 5.6.57).

Διάγραμμα 5.6.56 Αριθμός των άρθρων στον Ellhnikahoaxes.gr, που αφορούν την Ουκρανία, ανά μήνα από το 2020 έως 2023

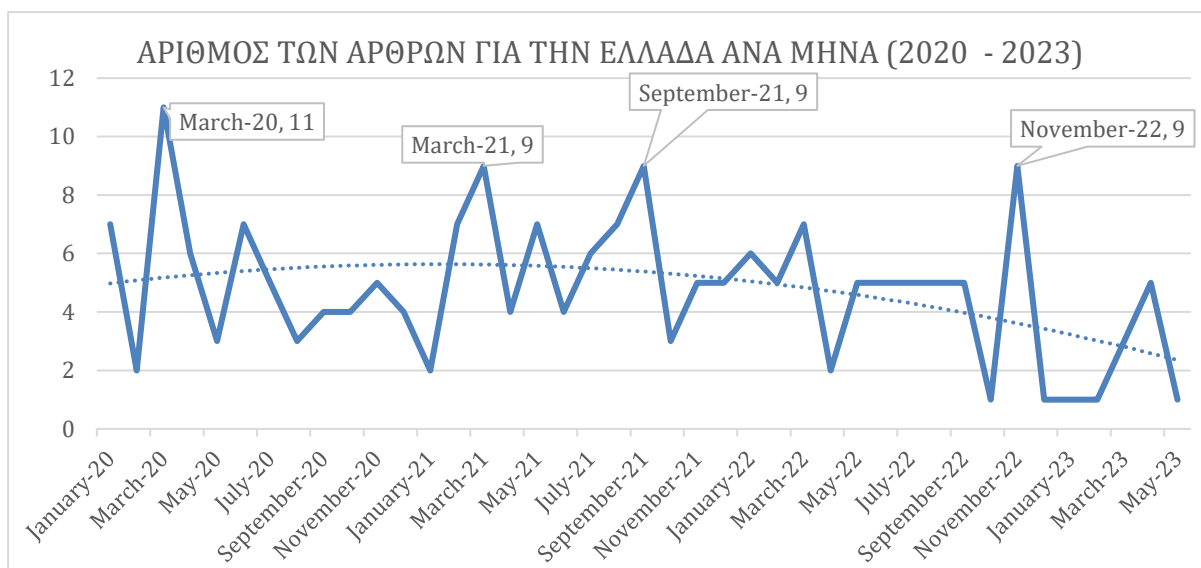


Διάγραμμα 5.6.57 Αριθμός των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν την Ουκρανία, ανά μήνα από τον Ιανουάριο 2022 έως το Μάιο 2023



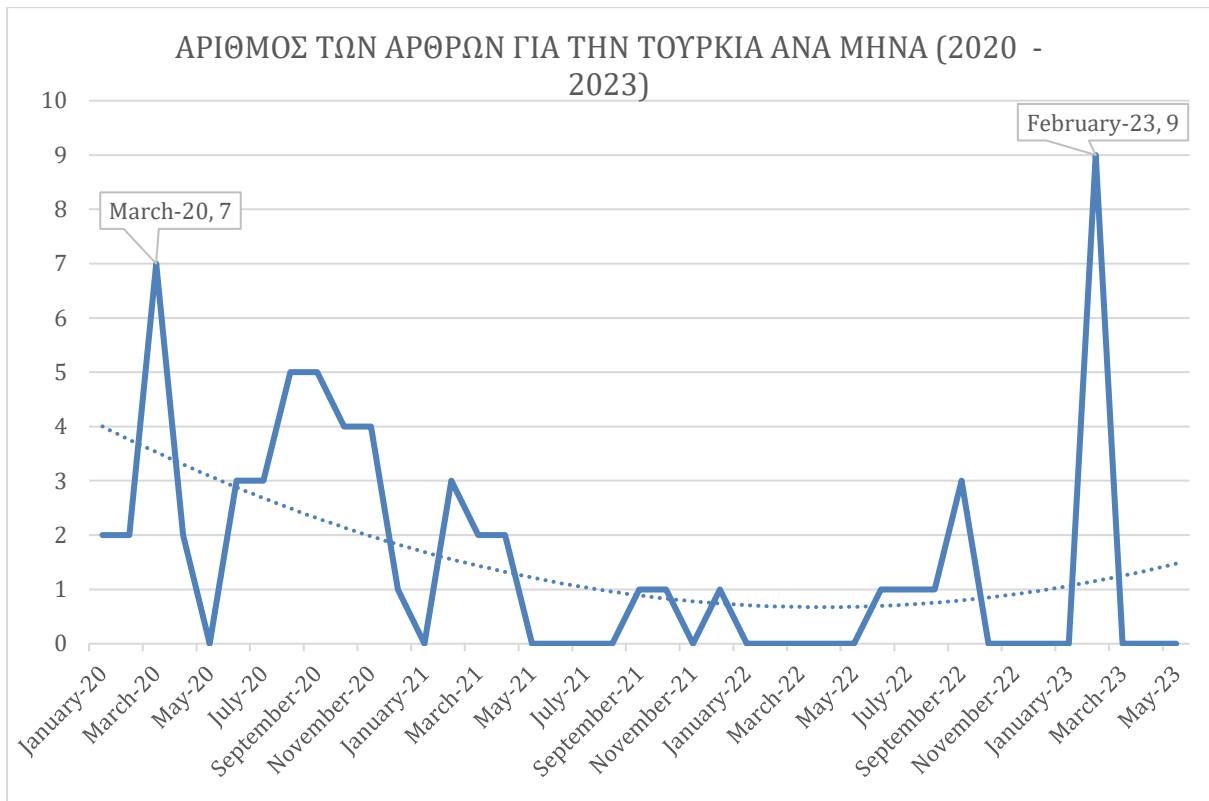
- Από το Νοέμβριο 2019 κάθε μήνα δημοσιεύονται έστω μερικά άρθρα για την Ελλάδα. Τα πιο πολλά άρθρα που αφορούν αυτό το θέμα δημοσιεύτηκαν το Μάρτιο 2020 (11 άρθρα) (Διάγραμμα 5.6.58).

Διάγραμμα 5.6.58 Αριθμός των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν την Ελλάδα, ανά μήνα από το 2020 έως 2023



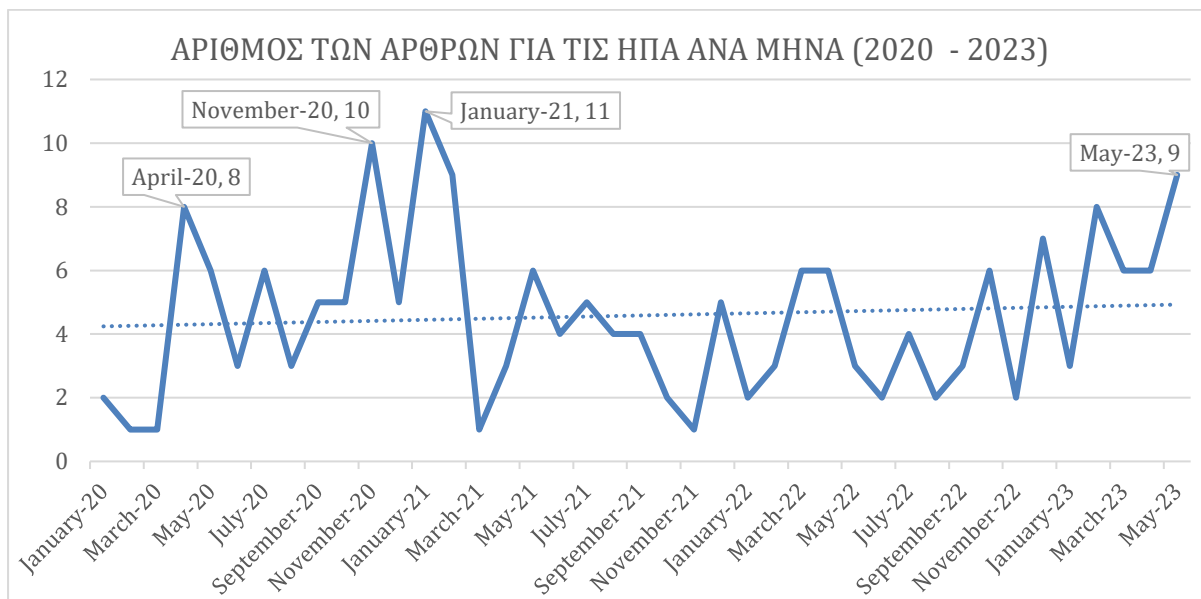
- Από το Νοέμβριο 2019 και μέχρι τον Απρίλιο 2021 ο μέσος αριθμός δημοσιεύσεων για την Τουρκία το μήνα ήταν τρία άρθρα το μήνα. Έπειτα υπήρχαν μη σημαντικές αυξομειώσεις μέχρι το Φεβρουάριο 2023, όταν ο αριθμός των δημοσιεύσεων ήταν ο μέγιστος (9 άρθρα) (Διάγραμμα 5.6.59).

Διάγραμμα 5.6.59 Αριθμός των άρθρων στον Ellhnikahoaxes.gr, που αφορούν την Τουρκία, ανά μήνα από το 2020 έως 2023



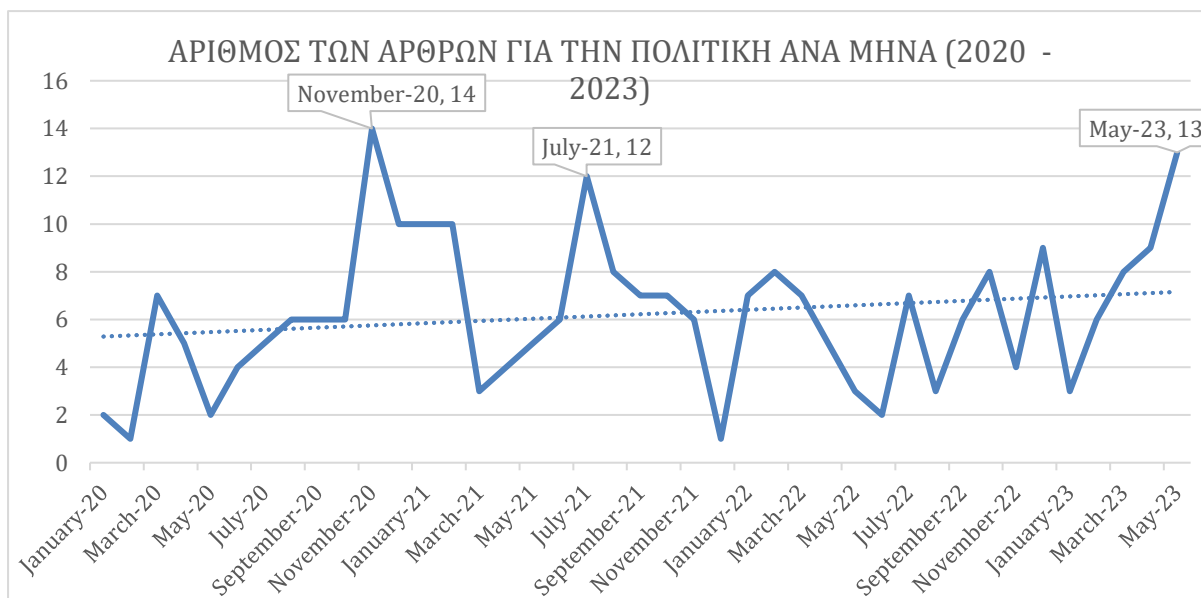
- Τα πιο πολλά άρθρα που αφορούν αυτό τις ΗΠΑ δημοσιεύτηκαν τον Απρίλιο 2020 (8 άρθρα), το Νοέμβριο 2020 (10 άρθρα), τον Ιανουάριο 2021 (11 άρθρα) και το Μάιο 2023 (9 άρθρα) (Διάγραμμα 5.6.60).

Διάγραμμα 5.6.60 Αριθμός των άρθρων στον **ΕλληνικάHoaxes.gr**, που αφορούν τις ΗΠΑ, ανά μήνα από το 2020 έως 2023

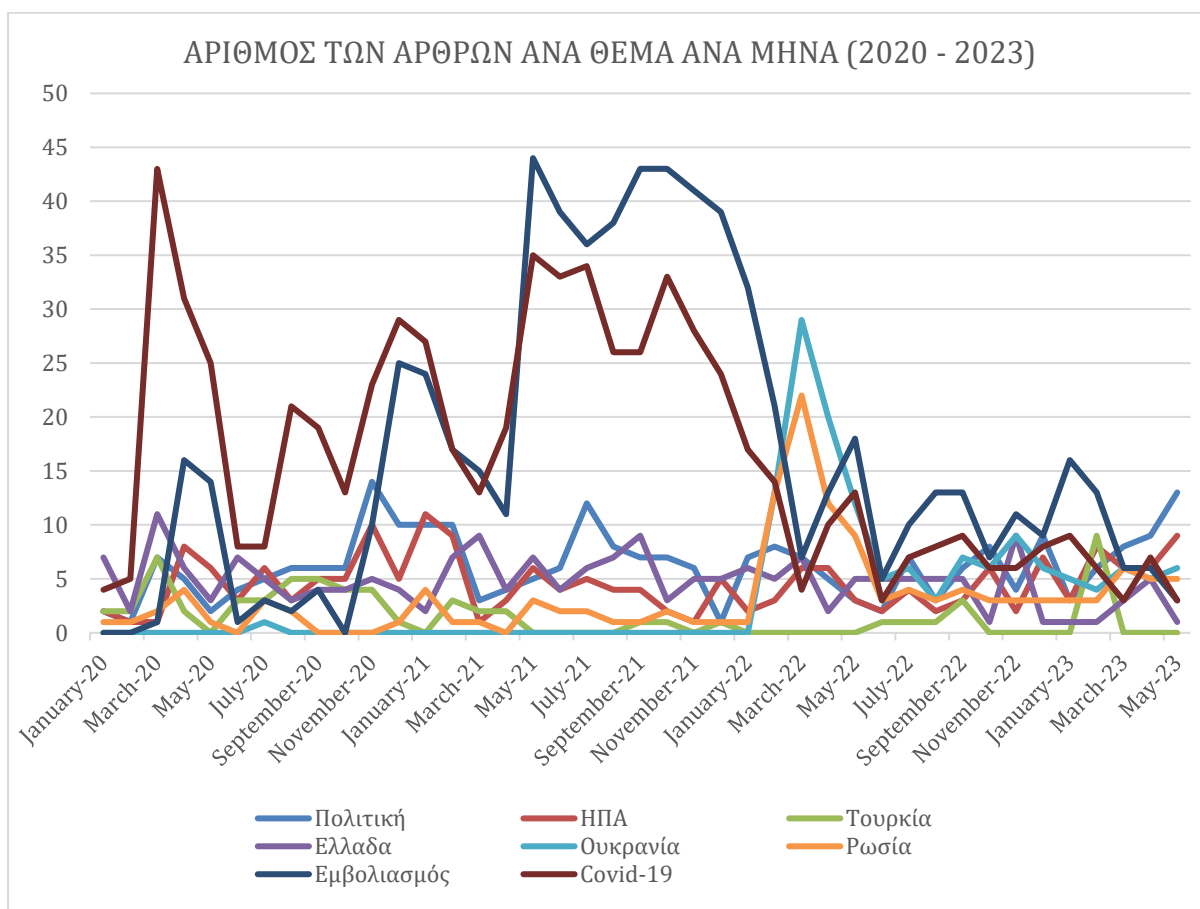


- Τα πιο πολλά άρθρα που αφορούν την πολιτική δημοσιεύτηκαν το Νοέμβριο 2020 (14 άρθρα), τον Ιούλιο 2021 (12 άρθρα) και το Μάιο 2023 (13 άρθρα) (Διάγραμμα 5.6.61).

Διάγραμμα 5.6.61 Αριθμός των άρθρων στον **ΕλληνικάHoaxes.gr**, που αφορούν την πολιτική, ανά μήνα από το 2020 έως 2023



Διάγραμμα 5.6.62 Αριθμός των άρθρων στον ΕλληναHoaxes.gr ανά θέμα ανά μήνα από το 2020 έως 2023



ΚΕΦΑΛΑΙΟ 6

ΕΠΙΛΟΓΟΣ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

6.1 ΕΠΙΛΟΓΟΣ

Το φαινόμενο των ψευδών ειδήσεων είναι παρόν σε όλες τις χώρες του κόσμου, αλλά επειδή η κουλτούρα και η νοοτροπία επηρεάζουν πολύ την αντίληψη των ανθρώπων, οι αναληθείς ειδήσεις από χώρα σε χώρα μπορούν να διαφέρουν πολύ. Επίσης, οι ψευδείς ειδήσεις δεν αφορούν μόνο τα παγκόσμια γεγονότα αλλά και εσωτερικά, με αποτέλεσμα να υπάρχουν πολλές παραπλανητικές ειδήσεις που αφορούν πράγματα που είναι κατανοητά μόνο από τους κατοίκους μιας συγκεκριμένης χώρας. Για το λόγο αυτό, είναι χρήσιμο να αναλύονται οι ψευδείς ειδήσεις σε κάθε χώρα ξεχωριστά. Ο σκοπός αυτής της εργασίας ήταν η συγγραφή του κώδικα για αυτοματοποιημένη ανάκτηση δεδομένων από τον ιστότοπο της ελληνικής υπηρεσίας ελέγχου γεγονότων EllhnikaHoaxes.gr, που επικεντρώνεται στις κυκλοφορούσες στο ελληνικό Διαδίκτυο ψευδείς ειδήσεις, που ελέγχθηκαν προς την εγκυρότητά τους, και η ανάλυση των ψευδών ειδήσεων στο ελληνικό Διαδίκτυο βάσει αυτά τα δεδομένα.

Μετά την ανάλυση, συμπεραίνεται ότι ο όγκος των ψευδών ειδήσεων που ελέγχονται από την υπηρεσία EllhnikaHoaxes.gr αυξήθηκε και σε σύγκριση με το 2013 ο αριθμός των ψευδών ειδήσεων διπλασιάστηκε.

Η πανδημία του Covid-19 δεν ήταν περίοδος με το μέγιστο αριθμό άρθρων το μήνα, αν και από τότε κάθε μήνα ελέγχονται έστω πενήντα ψευδείς ειδήσεις. Με άλλα λόγια, κατά εκείνη την περίοδο οι ψευδείς ειδήσεις για το Covid-19 ήταν το πιο κύριο θέμα αλλά οι ειδήσεις αυτές διαδιδόταν ομοιόμορφα χωρίς μεγάλη διακύμανση. Τα πιο πολλά άρθρα δημοσιεύτηκαν τον Ιούλιο 2014, ενώ τότε ο όρος των ψευδών ειδήσεων δεν ήταν κάτι ευρέως γνωστό.

Επειδή η διαφορά μεταξύ των κατηγοριών "Hoaxes", "Παραπληροφόρηση", "Fake news" και "Ψευδές" είναι πολύ λεπτή, αυτές οι κατηγορίες μπορούν να αναλυθούν ως μια γενικευμένη κατηγορία όλων των ψευδών ειδήσεων. Αν πάμε στα πιο ειδικευμένα θέματα, οι κατηγορίες "Συνωμοσιολογία", "Ψευδοεπιστήμη", "Υγεία" και "Ψευδής ισχυρισμός" είναι πολύ συχνές στις ψευδείς ειδήσεις στο ελληνικό Διαδίκτυο.

Στα άρθρα για τον κορονοϊό και τον εμβολιασμό επικρατούσαν οι παραπληροφορήσεις και οι ψευδείς επιστημονικοί ισχυρισμοί. Οι περισσότερες ύποπτες ειδήσεις που αφορούν τις χώρες όπως την Ελλάδα, την Τουρκία, τις ΗΠΑ, τη Ρωσία και την Ουκρανία ήταν

παραπληροφόρηση. Οι πιο πολλές ψευδείς ειδήσεις για τα πολιτικά θέματα ήταν επίσης παραπληροφόρηση.

Μολονότι η εποχή της πανδημίας δεν ήταν εποχή με τα περισσότερα άρθρα, φαίνεται πως σε σύγκριση με τις άλλες χρονιές, όταν οι ψευδείς ειδήσεις ήταν πολύ διαφοροποιημένες και αφορούσαν πολλά θέματα, το 2020-2023 σχεδόν όλες οι αναληθείς ειδήσεις ήταν για το Covid-19, με αποτέλεσμα τα θέματα του κορονοϊού και του εμβολιασμού να επικρατούν ανάμεσα σε όλα άλλα τα θέματα των ψευδών ειδήσεων των 2020-2023. Το 2020 και το 2021 το θέμα του κορονοϊού και του εμβολιασμού ήταν το πιο επίκαιρο. Στις αρχές του 2022 μετά την εισβολή της Ρωσίας στην Ουκρανία το Φλεβάρη 2022, το θέμα της σύγκρουσης αυτών των δυο χωρών ήταν πολύ δημοφιλές στις ψευδείς ειδήσεις στην Ελλάδα.

Ενδιαφέρον παρουσιάζουν τα θέματα κάποιων κατηγοριών ψευδών ειδήσεων. Σύμφωνα με τις ψευδείς ειδήσεις που έχουν αναλυθεί από την EllhnikaHoaxes.gr, οι απάτες στο ελληνικό Διαδίκτυο σχετίζονται με διάφορους διαγωνισμούς στο Ίντερνετ και τα μέσα κοινωνικής δικτύωσης. Οι ειδήσεις που ανήκουν στο "Like farming" επίσης σχετίζονται τα μέσα κοινωνικής δικτύωσης. Πολύ πιθανόν, οι "Like farming" ειδήσεις χρησιμοποιούν διάφορες φωτογραφίες, συχνά των παιδιών ή των γυναικών, στα μέσα κοινωνικής δικτύωσης για να αποκτήσουν όσο το δυνατό πιο πολλά likes. Οι ψευδείς ειδήσεις που περιέχουν τροποποιημένα βίντεο ή φωτογραφίες συνήθως αφορούν τα πολιτικά πρόσωπα.

Επειδή αυτή η πτυχιακή εργασία αποτελείται από δυο μέρη, ανάκτηση και ανάλυση των δεδομένων, αξίζει να σημειωθεί η χρησιμότητα του κώδικα στη Java που δεν είναι μόνο ένα εργαλείο για τις επόμενες ανακτήσεις των άρθρων από τον EllhnikaHoaxes.gr, αλλά και μια καλή βάση για την ανάπτυξη άλλων σχετιζόμενων προγραμμάτων.

6.2 ΙΔΕΕΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Ασφαλώς αυτή η εργασία δεν απαντάει μόνο σε ερωτήματα άλλα και παράγει νέες ιδέες για εργασία. Αξιοποιώντας αυτά τα δυο εργαλεία, τα δεδομένα και τον κώδικα, μερικά από τα θέματα που μπορούν να μελετηθούν είναι

- Ανάλυση των περιόδων με εξαιρετικά πολλά άρθρα, όπως Ιούλιος 2014, Μάρτιος 2017 και Απρίλιος-Μάιος 2020. Σε τι οφειλόταν ο αυξημένος αριθμός των ψευδών ειδήσεων που ελέγχθηκαν από την EllhnikaHoaxes.gr εκείνη την περίοδο;
- Στα περισσότερα άρθρα της EllhnikaHoaxes.gr υπάρχει είτε έστω ένας σύνδεσμος στην πρωτότυπη πηγή παραπληροφόρησης, είτε το αντιγραμμένο κείμενο της ψευδούς είδησης. Τα κείμενα των πρωτότυπων παραπλανητικών άρθρων η ισχυρισμών θα

μπορούσαν να ανακτηθούν και να αναλυθούν λεξιλογικά ή να αναλυθούν γενικά οι ιστότοποι που παράγουν ή/και διαδίδουν ψευδείς ειδήσεις.

- Ανάπτυξη μιας εφαρμογής που θα προσθέτει στη ΒΔ τα νέα άρθρα αυτόματα κάθε φορά που αυτά δημοσιεύονται.
- Ανάπτυξη μιας εφαρμογής, με τη βοήθεια της οποίας ένας απλός χρήστης θα μπορούσε να ελέγχει αν μια συγκεκριμένη είδηση είναι αληθής η ψευδής, βάση τα δεδομένα που ανακτώνται από τον EllhnikHoaxes.gr.

BIBΛΙΟΓΡΑΦΙΑ

7.1 ΞΕΝΟΓΛΩΣΣΗ

- Agudelo, G.E.R., Parra, O.J.S. and Velandia, J.B., 2018. Raising a model for fake news detection using machine learning in Python. In *Challenges and Opportunities in the Digital Era: 17th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2018, Kuwait City, Kuwait, October 30–November 1, 2018, Proceedings 17* (pp. 596-604). Springer International Publishing.
- Allcott, H., Gentzkow, M. and Yu, C., 2019. Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2), p.2053168019848554.
- Awan, M.J., Yasin, A., Nobanee, H., Ali, A.A., Shahzad, Z., Nabeel, M., Zain, A.M. and Shahzad, H.M.F., 2021. Fake news data exploration and analytics. *Electronics*, 10(19), p.2326.
- Berduygina, O.N., Vladimirova, T.N. and Chernyaeva, E.V., 2019. Trends in the spread of fake news in the mass media. *Media Watch*, 10(1), pp.122-132.
- Berger, J. and Milkman, K.L., 2012. What makes online content viral? *Journal of marketing research*, 49(2), pp.192-205.
- Bhoir, S.V., 2020, January. An efficient fake news detector. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-9). IEEE.
- Bounegru, L., Gray, J., Venturini, T. and Mauri, M., 2018. A field guide to 'Fake News' and other information disorders. *A Field Guide to "Fake News" and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods*, Public Data Lab, Amsterdam (2018).
- Braşoveanu, A.M. and Andonie, R., 2019. Semantic fake news detection: a machine learning perspective. In *Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part I 15* (pp. 656-667). Springer International Publishing.
- Ceron, W., de-Lima-Santos, M.F. and Quiles, M.G., 2021. Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content. *Online Social Networks and Media*, 21, p.100116.

- Collins, B., Hoang, D.T., Nguyen, N.T. and Hwang, D., 2021. Trends in combating fake news on social media—a survey. *Journal of Information and Telecommunication*, 5(2), pp.247-266.
- Cui, L., Wang, S. and Lee, D., 2019, August. Same: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 41-48).
- Das, D. and Clark, A.J., 2019, September. Satire vs fake news You can tell by the way they say it. In *2019 First International Conference on Transdisciplinary AI (TransAI)* (pp. 22-26). IEEE.
- Dey, A., Rafi, R.Z., Hasan, S. and Kundu, S., 2018. *Fake news pattern recognition using linguistic analysis* (Doctoral dissertation, BRAC University).
- Figueira, Á. And Oliveira, L., 2017. The current state of fake news: challenges and opportunities. *Procedia computer science*, 121, pp.817-825.
- Fourney, A., Racz, M.Z., Ranade, G., Mobius, M. and Horvitz, E., 2017, November. Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. In *CIKM* (Vol. 17, pp. 6-10).
- Go, S.G. and Lee, M.R., 2020. Analysis of fake news in the 2017 Korean presidential election. *Asian Journal for Public Opinion Research*, 8(2), pp.105-125.
- Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K.H., Bonk, C., Bouzaghrane, M.A., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J.B. and Falak, W., 2018, May. Fake news vs satire: A dataset and analysis. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 17-21).
- Gutiérrez-Coba, L.M., Coba-Gutiérrez, P. and Gómez-Díaz, J.A., 2020. Fake news about Covid-19: A comparative analysis of six iberoamerican countries. *Revista Latina de Comunicación Social*, (78), pp.237-264.
- Han, W. and Mehta, V., 2019, November. Fake news detection in social networks using machine learning and deep learning: Performance evaluation. In *2019 IEEE international conference on industrial internet (ICII)* (pp. 375-380). IEEE.
- Higdon, N., 2020. *The anatomy of fake news: A critical news literacy education*. University of California Press.

- Iosifidis, P. and Nicoli, N., 2020. The battle to end fake news: A qualitative content analysis of Facebook announcements on how it combats disinformation. *International Communication Gazette*, 82(1), pp.60-81.
- Iwendi, C., Mohan, S., Ibeke, E., Ahmadian, A. and Ciano, T., 2022. Covid-19 fake news sentiment analysis. *Computers and electrical engineering*, 101, p.107967.
- Jang, S.M., Geng, T., Li, J.Y.Q., Xia, R., Huang, C.T., Kim, H. and Tang, J., 2018. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior*, 84, pp.103-113.
- Jones, M.O., 2019. The gulf information war |propaganda, fake news, and fake trends: The weaponization of twitter bots in the gulf crisis. *International journal of communication*, 13, p.27.
- Kapusta, J., Benko, L. and Munk, M., 2020. Fake news identification based on sentiment and frequency analysis. In *Innovation in Information Systems and Technologies to Support Learning Research: Proceedings of EMENA-ISTL 2019 3* (pp. 400-409). Springer International Publishing.
- Klyuev, V., 2018, August. Fake news filtering: Semantic approaches. In *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 9-15). IEEE.
- Krittanawong, C., Narasimhan, B., Virk, H.U.H., Narasimhan, H., Hahn, J., Wang, Z. and Tang, W.W., 2020. Misinformation dissemination in twitter in the COVID-19 era. *The American journal of medicine*, 133(12), pp.1367-1369.
- Maliaroudakis, E., Boland, K., Dietze, S., Todorov, K., Tzitzikas, Y. and Fafalios, P., 2021, April. ClaimLinker: Linking text to a knowledge graph of fact-checked claims. In *Companion Proceedings of the Web Conference 2021* (pp. 669-672).
- Mayakul, T., Srisawat, W. and Kiattisin, S., 2019, December. Fake news and online disinformation. A perspectives of Thai government officials. In *2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)* (pp. 1-4). IEEE.
- Mukerji, N.S., 2018. A conceptual analysis of fake news.
- Nakov, P., Alam, F., Shaar, S., Martino, G.D.S. and Zhang, Y., 2021. A second pandemic? Analysis of fake news about COVID-19 vaccines in Qatar. *arXiv preprint arXiv:2109.11372*.

- Oshikawa, R., Qian, J. and Wang, W.Y., 2018. A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.
- Paul, C. and Matthews, M., 2016. The Russian “firehose of falsehood” propaganda model. *Rand Corporation*, 2(7), pp.1-10.
- Raponi, S., Khalifa, Z., Oligeri, G. and Di Pietro, R., 2022. Fake news propagation: a review of epidemic models, datasets, and insights. *ACM Transactions on the Web (TWEB)*, 16(3), pp.1-34.
- Riffkin, R., 2015. Americans' trust in media remains at historical low. *Gallup*, September 28.
- Righetti, N., 2021. Four years of fake news: A quantitative analysis of the scientific literature. *First Monday*
- Rubin, V.L., Chen, Y. and Conroy, N.K., 2015. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1), pp.1-4.
- Shao, C., Ciampaglia, G.L., Flammini, A. and Menczer, F., 2016, April. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web* (pp. 745-750).
- Shu, K., Bernard, H.R. and Liu, H., 2019. Studying fake news via network analysis: detection and mitigation. *Emerging research challenges and opportunities in computational social network analysis and mining*, pp.43-65.
- Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1), pp.22-36.
- Soll, J., 2016. The long and brutal history of fake news. *Politico Magazine*.
- Sousa-Silva, R., 2022. Fighting the fake: A forensic linguistic analysis to fake news detection. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 35(6), pp.2409-2433.
- Tandoc Jr, E.C., Lim, Z.W. and Ling, R., 2018. Defining “fake news” A typology of scholarly definitions. *Digital journalism*, 6(2), pp.137-153.
- Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., Zloch, M., Zapilko, B., Dietze, S. and Todorov, K., 2019. ClaimsKG: A knowledge graph of fact-checked claims. In *The Semantic Web–ISWC 2019: 18th International Semantic Web*

Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18 (pp. 309-324). Springer International Publishing.

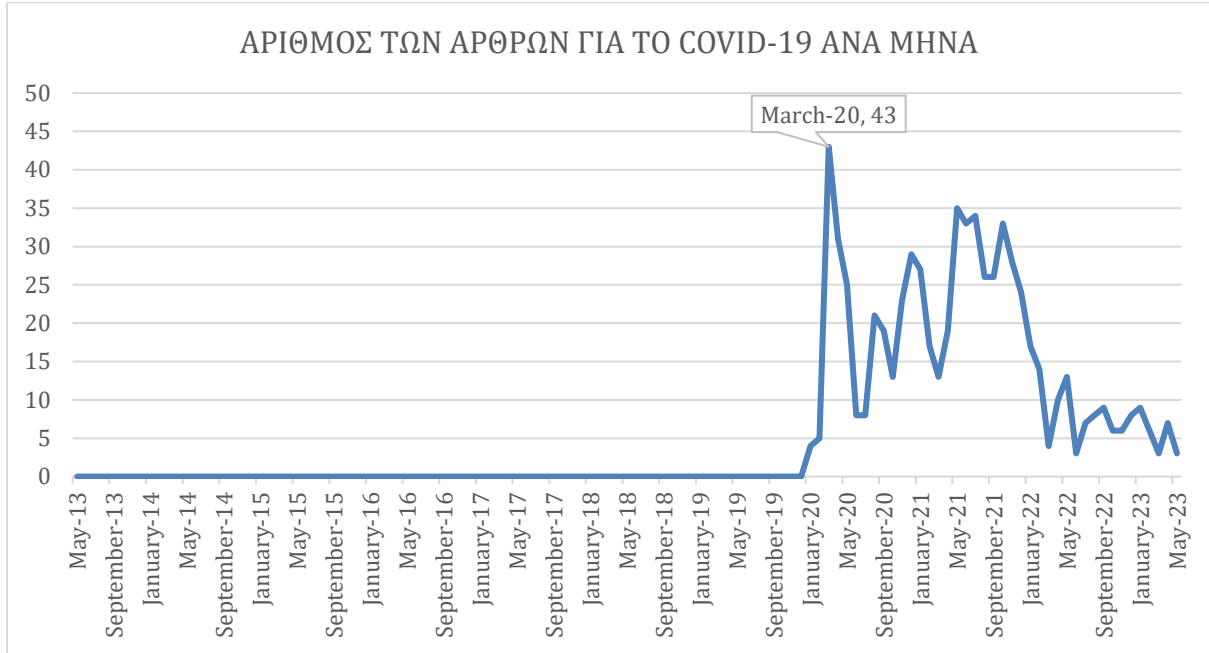
- Vosoughi, S., Roy, D. and Aral, S., 2018. The spread of true and false news online. *science*, 359(6380), pp.1146-1151.
- Wardle, C., 2017. Fake news. It's complicated. *firstdraftnews.org*. Retrieved April 22, 2017.
- Zhang, X. and Ghorbani, A.A., 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, 57(2), p.102025.
- Zhou, X. and Zafarani, R., 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5), pp.1-40.

7.2 ΗΛΕΚΤΡΟΝΙΚΕΣ ΠΗΓΕΣ

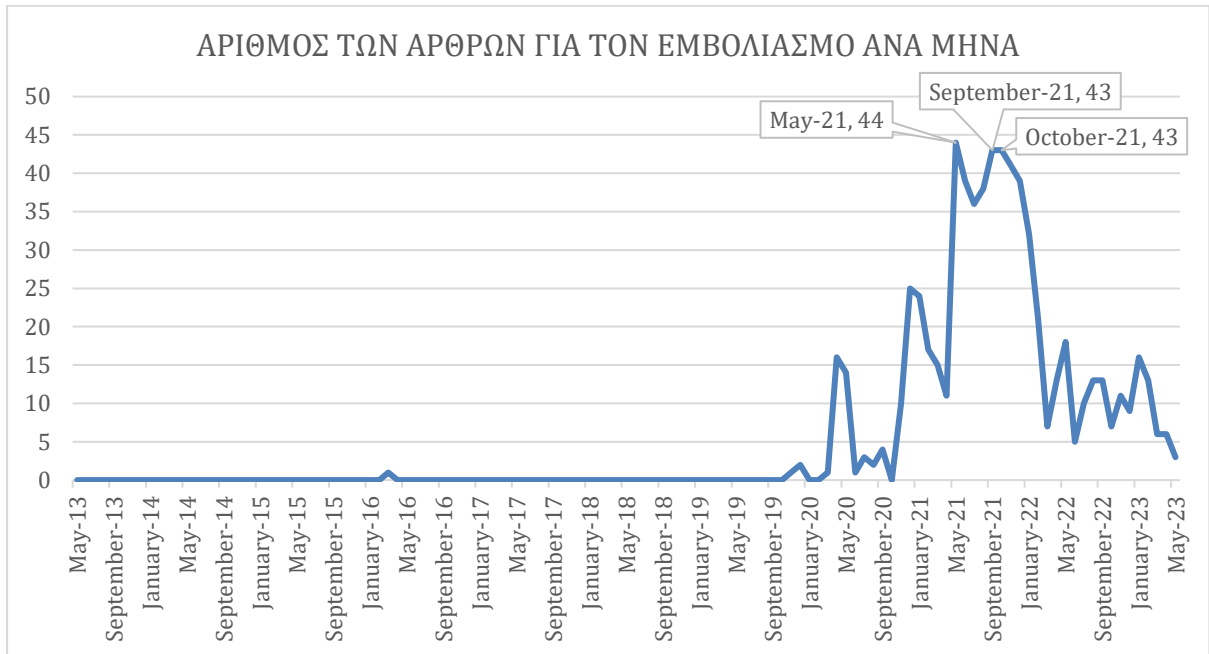
- <https://nandeshwar.info/>
- <https://wikipedia.org/>

ΠΑΡΑΡΤΗΜΑ Α

Διάγραμμα 8.1 Αριθμός των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν το Covid-19, ανά μήνα από το 2013 έως 2023



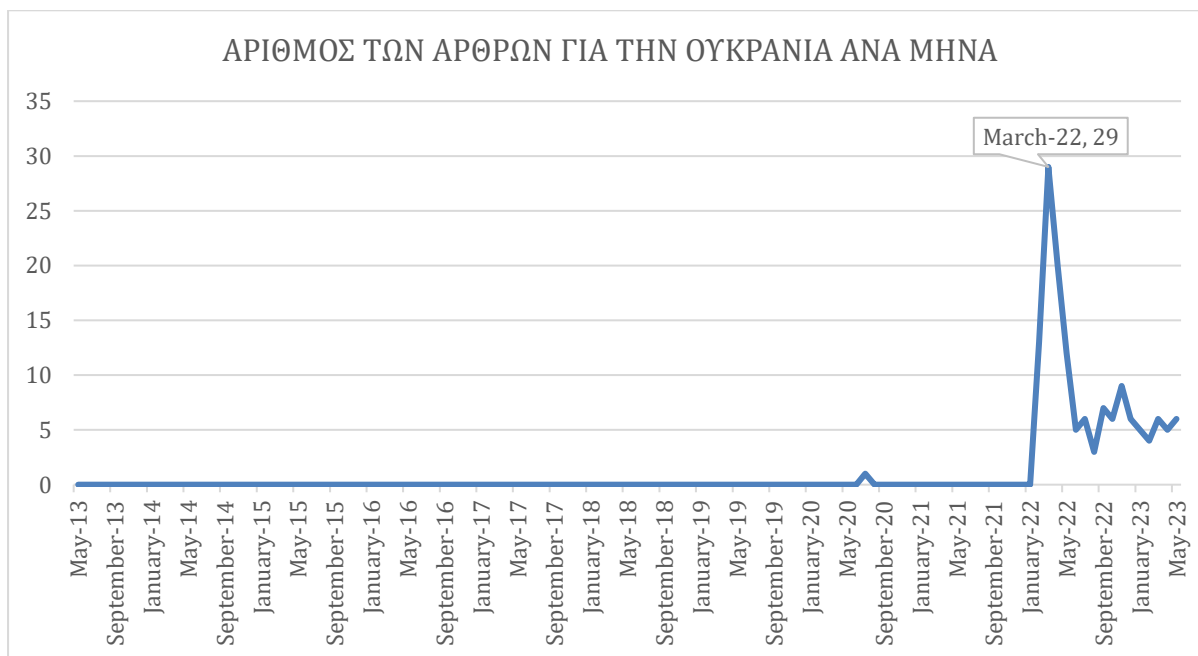
Διάγραμμα 8.2 Αριθμός των άρθρων στον ΕλληνικάHoaxes.gr, που αφορούν τον εμβολιασμό, ανά μήνα από το 2013 έως 2023



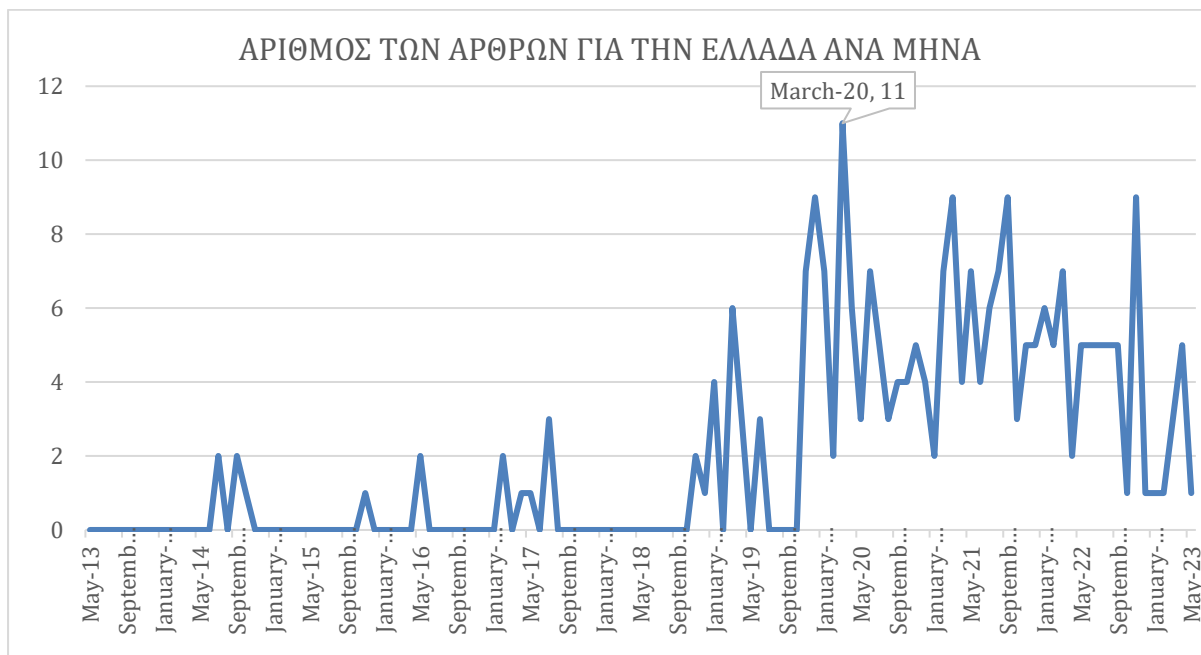
Διάγραμμα 8.3 Αριθμός των άρθρων στον ΕλληνικαΗοαχες.gr, που αφορούν τη Ρωσία, ανά μήνα από το 2013 έως 2023



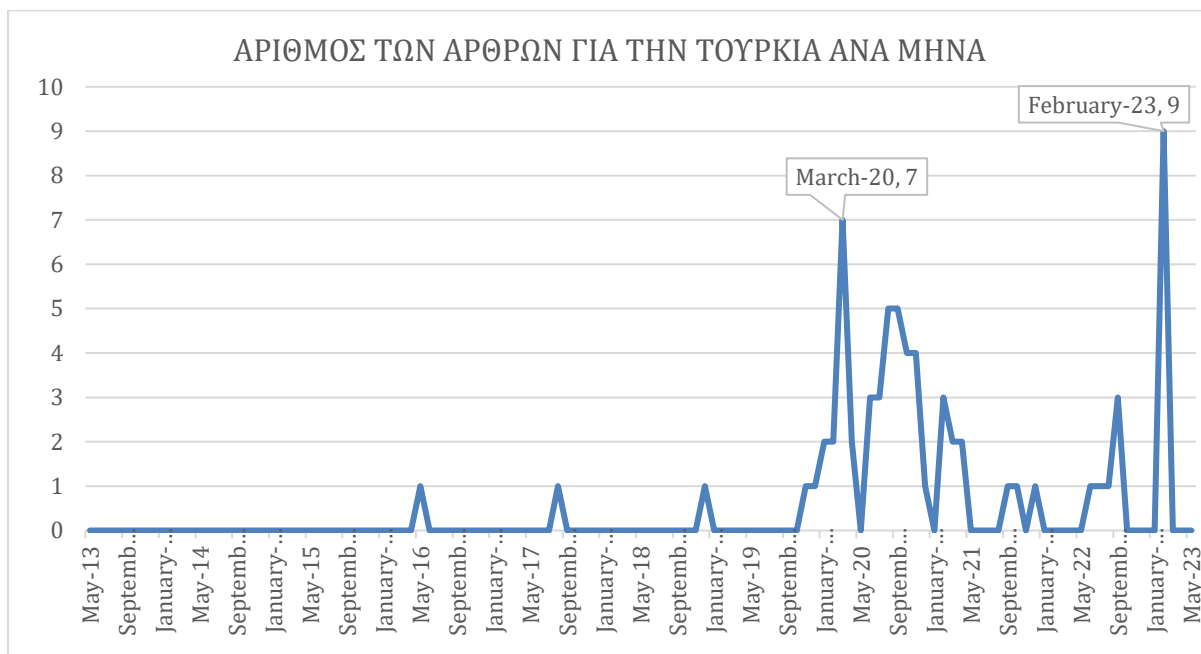
Διάγραμμα 8.4 Αριθμός των άρθρων στον ΕλληνικαΗοαχες.gr, που αφορούν την Ουκρανία, ανά μήνα από το 2013 έως 2023



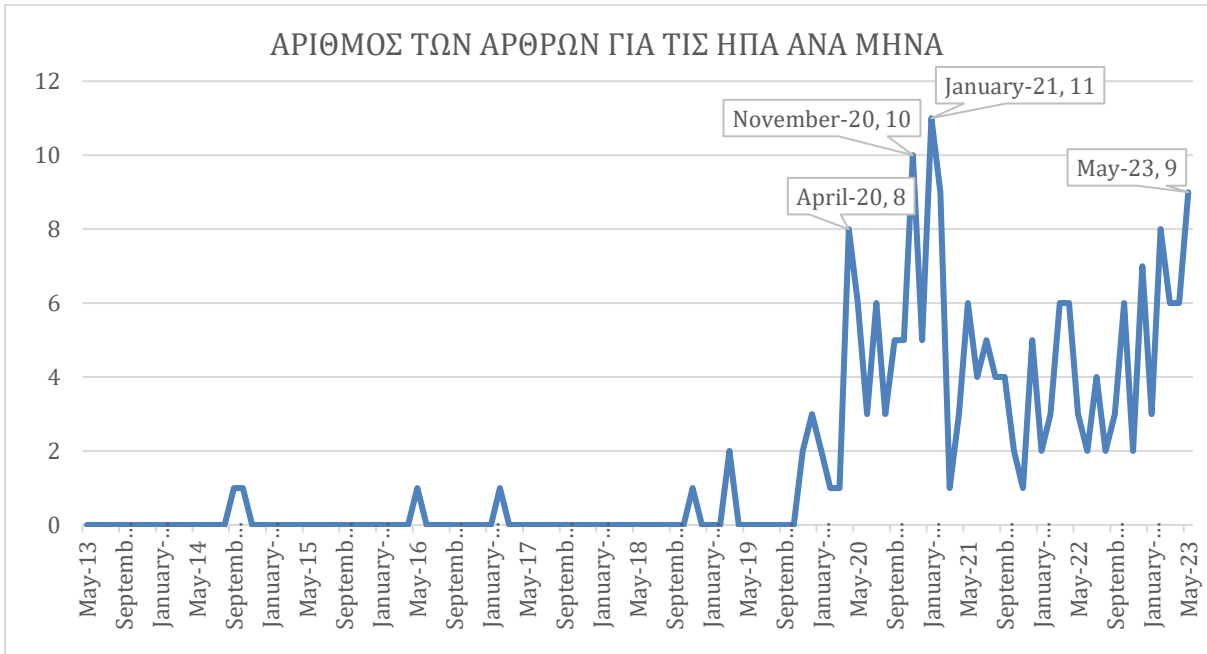
Διάγραμμα 8.5 Αριθμός των άρθρων στον ΕλληνικαHoaxes.gr, που αφορούν την Ελλάδα, ανά μήνα από το 2013 έως 2023



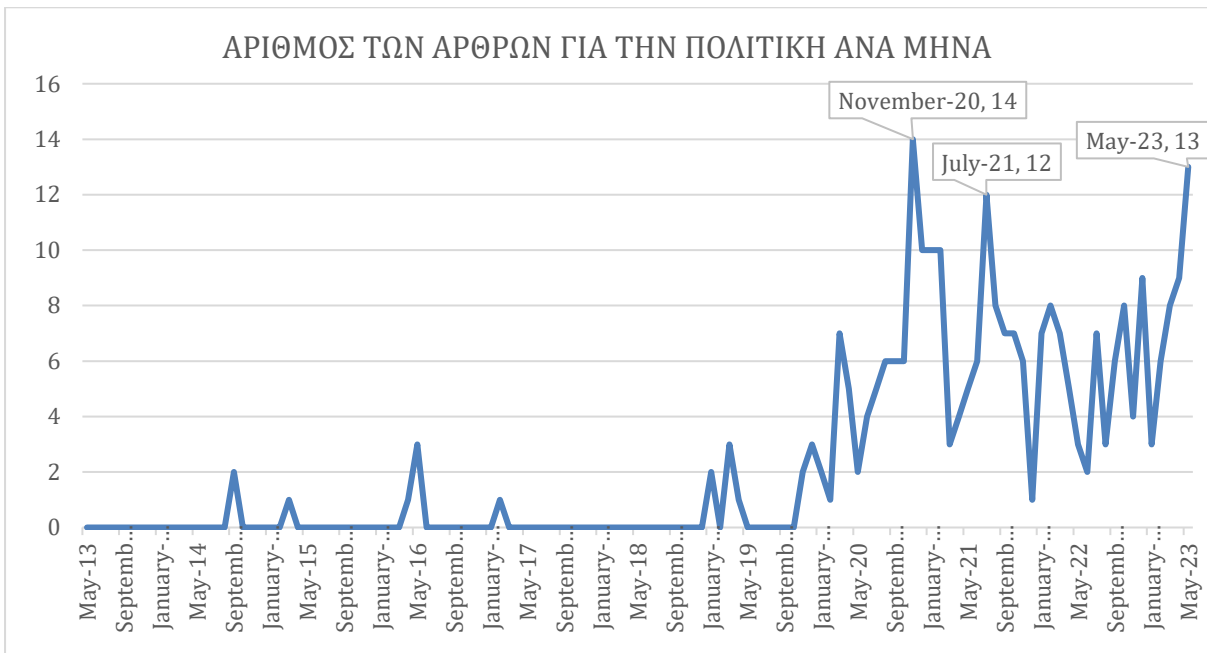
Διάγραμμα 8.6 Αριθμός των άρθρων στον ΕλληνικαHoaxes.gr, που αφορούν την Τουρκία, ανά μήνα από το 2013 έως 2023



Διάγραμμα 8.7 Αριθμός των άρθρων στον ΕλληνικάΗοαχες.gr, που αφορούν τις ΗΠΑ, ανά μήνα από το 2013 έως 2023



Διάγραμμα 8.8 Αριθμός των άρθρων στον ΕλληνικάΗοαχες.gr, που αφορούν την πολιτική, ανά μήνα από το 2013 έως 2023



ΠΑΡΑΡΤΗΜΑ Β

Εικόνα 9.1 Κλάση LinkRetrieval (1)

```
1 package gr.hmu.mst.fakenews;
2
3 import java.io.BufferedReader;
4 import java.io.BufferedWriter;
5 import java.io.File;
6 import java.io.FileInputStream;
7 import java.io.FileOutputStream;
8 import java.io.IOException;
9 import java.io.InputStreamReader;
10 import java.io.OutputStreamWriter;
11 import java.io.PrintStream;
12 import java.io.Writer;
13 import java.net.MalformedURLException;
14 import java.net.URL;
15 import java.net.URLConnection;
16 import java.nio.charset.StandardCharsets;
17 import java.util.HashSet;
18
19
20 /**
21  *
22  * @author zabal
23  */
24 public class LinkRetrieval {
25
26     private static final String HOAXES_LINKS_FILENAME = "fakeLinks.csv";
27     private static HashSet<String> oldRetrievedLinks = new HashSet<>();
28     private static HashSet<String> newRetrievedLinks = new HashSet<>();
29
30
31
32
33     public static void main(String[] args) throws MalformedURLException,
34         IOException, InterruptedException {
35
36         getOldLinks();
37
38         File linkFile = new File(HOAXES_LINKS_FILENAME);
39
40         //Writing a new file or extending the existing one
41
42         System.setOut(new PrintStream(System.out, true, "UTF8"));
43         int maxPage = getLastPageNum();
44         int numberOfNewLinks = addLinks(linkFile,maxPage+1);
45         System.out.println(
46             "\n# Number of new links added in the file: " + numberOfNewLinks);
47     }
48
49     /* Scans the HTML code of the webpage and returns the HTML code as a string
50     */
51     public static String getURLContent(String urlPath) throws
52         MalformedURLException, IOException {
53         URL fakeNewsPage = new URL(urlPath);
54         URLConnection fakeNewsCon = fakeNewsPage.openConnection();
```

Εικόνα 9.2 Κλάση LinkRetrieval (2)

```

55
56     String pageContent;
57
58     try (BufferedReader in = new BufferedReader(
59         new InputStreamReader(
60             fakeNewsCon.getInputStream(), "UTF-8")
61     )) {
62         pageContent = "";
63         String inputLine;
64
65         while ((inputLine = in.readLine()) != null) {
66             pageContent += (inputLine + "\n");
67         }
68     }
69
70     return pageContent;
71 }
72
73 /* Reads the content of the first page and retrieves the number of the last
74 page from the bottom of the page.
75 */
76 public static int getLastPageNum() throws IOException{
77     String firstPage = "https://www.ellinikahoaxes.gr/category/kathgories/";
78     String firstPageContent = getURLContent(firstPage);
79     int lastPageNumElementStart = firstPageContent.indexOf(
80         "\"page-numbers dots\"");
81     int lastPageNumElementEnd = firstPageContent.indexOf(
82         "</a>", lastPageNumElementStart);
83     String lastPageNumElement = firstPageContent.substring(
84         lastPageNumElementStart, lastPageNumElementEnd);
85     int lastPageNumStart = lastPageNumElement.indexOf(
86         "kathgories/page/");
87     int lastPageNumEnd = lastPageNumElement.indexOf(
88         "/" data-wpel-link="internal");
89     String lastPageNumStr = lastPageNumElement.substring(
90         lastPageNumStart+16, lastPageNumEnd);
91     System.out.println("Total number of pages: " + lastPageNumStr);
92     int lastPage = Integer.parseInt(lastPageNumStr);
93     return lastPage;
94 }
95
96 /* Reads the file with old links and adds these links
97 to the hashset "alreadyRetrievedLinks".
98 */
99 public static void getOldLinks() throws IOException {
100     System.out.println(
101         "# Reading the links in the text file '"
102         + HOAXES_LINKS_FILENAME + "'");
103     File file = new File(HOAXES_LINKS_FILENAME);
104
105     try (BufferedReader in = new BufferedReader(
106         new InputStreamReader(
107             new FileInputStream(file),
108             "UTF8")

```

Εικόνα 9.3 Κλάση LinkRetrieval (3)

```
109     )) {
110
111         String str;
112
113         while ((str = in.readLine()) != null) {
114
115             if (!str.trim().equals("")) {
116                 oldRetrievedLinks.add(str);
117             }
118         }
119     } catch (Exception e) {
120         System.out.println(e.getMessage());
121     }
122     System.out.println(
123         " - Total number of links in the file: "
124         + oldRetrievedLinks.size());
125 }
126
127
128 /* Retrieves links from the website
129 1. If a retrieved link has already been retrieved and saved in the file
130 with retrieved links during of the previous executions, the loop stops and
131 the execution is over.
132 2. If a retrieved link has not already been retrieved and saved in the file
133 with retrieved links during one of the previous executions, it is added
134 to the file with retrieved links and the executions contionues.
135 3. Oif a retrieved link has already been retrieved during this current
136 execution but it is not in the file with retrieved links (dublicate
137 publications), this link is just ignored and the execution continues.
138 */
139 public static int addLinks(File f, int maxP) throws
140     IOException, InterruptedException {
141     int n;
142
143     try (Writer textWriter = new BufferedWriter(
144         new OutputStreamWriter(
145             new FileOutputStream(f, true),
146             StandardCharsets.UTF_8)
147     )) {
148
149         boolean canStopReadingPages = false;
150         n = 0;
151         for (int pageNum = 1; pageNum < maxP; pageNum++) {
152             String pageAddress =
153                 "https://www.ellinikahoaxes.gr/category/kathgories/";
154             String pageNumStr = String.valueOf(pageNum);
155
156             if (pageNum >= 2) {
157                 pageAddress = pageAddress + "page/" + pageNumStr + "/";
158             }
159
160             System.out.println(
161                 "\n# Reading the contents of the page: " + pageAddress);
162             String fakeNewsContent = getURLContent(pageAddress);
```

Εικόνα 9.4 Κλάση LinkRetrieval (4)

```
163     int i1 = fakeNewsContent.indexOf(
164         "<div class=\"col-lg-9 col-md-9 col-sm-12 col-sm-12\"");
165     int i2 = fakeNewsContent.indexOf(
166         "<div class=\"pagination-list text-center\"");
167     String newsPart = fakeNewsContent.substring(i1, i2);
168     int divStartPos = newsPart.indexOf(
169         "<div class=\"blog-post-content\"");
170
171     while (divStartPos != -1) {
172         int divEndPos = newsPart.indexOf("</div", divStartPos);
173         String divContent = newsPart.substring(divStartPos, divEndPos);
174         System.out.println("divContent: " + divContent);
175         int aStartPos = divContent.indexOf("<a ");
176         int aEndPost = divContent.indexOf(
177             "\" data-wpel-link=\"internal\">", aStartPos);
178         String hrefValue = divContent.substring(
179             aStartPos + 9, aEndPost);
180
181         if (oldRetrievedLinks.contains(hrefValue)) {
182             System.out.println(
183                 " - The link '" + hrefValue
184                 + "' already exists in the list! No need to check the remaining...");
185             canStopReadingPages = true;
186             break;
187         } else {
188
189             if (newRetrievedLinks.contains(hrefValue) == false) {
190                 System.out.println(" - Writing the link: " + hrefValue);
191                 textWriter.append(hrefValue).append("\n");
192                 newRetrievedLinks.add(hrefValue);
193                 n++;
194             }
195         }
196
197         divStartPos = newsPart.indexOf(
198             "<div class=\"blog-post-content\"", divEndPos);
199     }
200
201     if (canStopReadingPages == true) {
202         break;
203     }
204
205     Thread.sleep(2000);
206 }
207 textWriter.flush();
208 }
209 return n;
210 }
211 }
212 }
```

Εικόνα 9.5 Κλάση PageDataRetrieval

```

1 package gr.hmu.mst.fakenews;
2
3 import java.io.BufferedReader;
4 import java.io.BufferedWriter;
5 import java.io.File;
6 import java.io.FileInputStream;
7 import java.io.FileOutputStream;
8 import java.io.IOException;
9 import java.io.InputStreamReader;
10 import java.io.OutputStreamWriter;
11 import java.io.PrintStream;
12 import java.io.Writer;
13 import java.net.MalformedURLException;
14 import java.net.URL;
15 import java.net.URLConnection;
16 import java.nio.charset.StandardCharsets;
17 import java.util.Arrays;
18 import java.util.HashMap;
19 import java.util.HashSet;
20
21 /**
22  *
23  * @author zabal
24  */
25 public class PageDataRetrieval {
26     private static final String HOAXES_LINKS_FILENAME = "fakeLinks.csv"; //only links
27     public static String allData = "fakeLinksData.csv"; //links data
28
29     private static HashSet<String> links = new HashSet<>();
30
31     static HashSet<String> accessedLinks = new HashSet<>();
32     public static HashSet<String> validCatVerdictValues = new HashSet<>();
33
34     public static HashMap<String, String> imageCategory = new HashMap<>();
35     public static HashMap<String, String> linkCategory = new HashMap<>();
36     public static HashMap<String, String> monthConvert= new HashMap<>();
37
38     private static HashSet<String> parasplhr_imageLinks = new HashSet<>();
39     private static HashSet<String> leipeiThemPeriex_imageLinks = new HashSet<>();
40     private static HashSet<String> tropopEikona_imageLinks = new HashSet<>();
41     private static HashSet<String> pseudes_imageLinks = new HashSet<>();
42     private static HashSet<String> anakyklEidhsh_imageLinks = new HashSet<>();
43     private static HashSet<String> synvmos_imageLinks = new HashSet<>();
44     private static HashSet<String> tropopVinteo_imageLinks = new HashSet<>();
45     private static HashSet<String> miksi_imageLinks = new HashSet<>();
46     private static HashSet<String> psevdospis_imageLinks = new HashSet<>();
47     private static HashSet<String> apath_imageLinks = new HashSet<>();
48     private static HashSet<String> satira_imageLinks = new HashSet<>();
49     private static HashSet<String> likeFarming_imageLinks = new HashSet<>();
50     private static HashSet<String> kindynolog_imageLinks = new HashSet<>();
51     private static HashSet<String> anepiv_imageLinks = new HashSet<>();
52     private static HashSet<String> clickbait_imageLinks = new HashSet<>();

```


Εικόνα 9.6 Κλάση PageDataRetrieval (2)

```

57 private static HashSet<String> psevdheTexyr_imageLinks = new HashSet<>();
58 private static HashSet<String> sikonaTexnNohm_imageLinks = new HashSet<>();
59
60 public static void main(String[] args) throws
61     IOException, InterruptedException {
62     System.setOut(new PrintStream(System.out, true, "UTF8"));
63     getOldLinks();
64
65     /*Adds values of all existing categories and verdicts (text values) of
66     fake news to the hashset "validCatVerdictValues"
67     */
68     String validCatVerdictValuesStr [] = {"ΠΑΡΑΠΛΗΡΟΦΩΡΗΗ",
69     "ΔΕΙΝΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ", "ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ", "ΨΕΥΔΕΣ",
70     "ΑΝΑΚΥΚΛΩΜΕΝΗ ΕΙΔΗΣΗ", "ΕΥΝΗΣΜΟΣΙΟΛΟΓΙΑ", "ΤΡΟΠΟΠΟΙΗΜΕΝΟ ΒΙΝΤΕΟ",
71     "ΜΙΣΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ", "ΨΕΥΔΟΕΠΙΣΤΗΜΗ", "ΑΠΑΤΗ",
72     "LIKE FARMING", "ΚΙΝΔΥΝΟΛΟΓΙΑ", "ΑΝΕΠΙΒΕΒΑΙΩΤΟ", "CLICKBAIT", "EATIPA",
73     "ΨΕΥΔΗ ΙΣΧΥΡΙΣΜΟΣ", "HOAXES", "ΨΕΥΤΙΚΑ ΠΗΓΑ ΚΑΙ ΑΠΟΘΕΤΜΑΤΑ",
74     "FAKE NEWS", "ΥΠΕΙΑ", "ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ", "ΔΙΑΦΟΡΑ"};
75     validCatVerdictValues.addAll(Arrays.asList(validCatVerdictValuesStr));
76
77
78     /* Adds values of all existing categories (image values)
79     to the hashmap "imageCategory"
80     */
81     String paraplhr_images[] = {
82     "https://media.ellinikahoaxes.gr/uploads/2020/08/paraplhrforhsh-full.png",
83     "https://media.ellinikahoaxes.gr/uploads/2019/05/Missinformation-1-1024x369.png",
84     "https://media.ellinikahoaxes.gr/uploads/2019/05/Missinformation-2-1024x369.png",
85     "https://media.ellinikahoaxes.gr/uploads/2019/01/12-1024x369.png",
86     "https://media.ellinikahoaxes.gr/uploads/2019/03/49727936_2513812705512635_1756180280151900160_n-1024x369.png",
87     "https://media.ellinikahoaxes.gr/uploads/2019/11/Parapliroforisi_2-300x108.png",
88     "https://media.ellinikahoaxes.gr/uploads/2019/11/Parapliroforisi_2-1-1024x369.png",
89     "https://media.ellinikahoaxes.gr/uploads/2019/01/12-300x108.png",
90     "https://media.ellinikahoaxes.gr/uploads/2019/03/49727936_2513812705512635_1756180280151900160_n-300x108.png",
91     "https://media.ellinikahoaxes.gr/uploads/2019/05/Missinformation-1-300x108.png",
92     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/49727936_2513812705512635_1756180280151900160_n-300x108.png",
93     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/12-300x108.png",
94     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/12-1024x369.png",
95     "https://media.ellinikahoaxes.gr/uploads/2019/06/Missinformation-1024x369.png",
96     "https://media.ellinikahoaxes.gr/uploads/2019/06/Missinformation-300x108.png",
97     "https://i2.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/12.png?fit=696%2C251&ssl=1",
98     "https://media.ellinikahoaxes.gr/uploads/2019/05/Missinformation-3-1024x369.png"
99     };
100     paraplhr_imageLinks.addAll(Arrays.asList(paraplhr_images));
101     imageCategory = MinorClass.addDictValues(
102         paraplhr_imageLinks,
103         imageCategory,
104         "ΠΑΡΑΠΛΗΡΟΦΩΡΗΗ");
105
106     String leipeiThemPeriex_images[] = {
107     "https://media.ellinikahoaxes.gr/uploads/2020/08/leipei-thematiko-periexomeno-full.png"
108     }

```

Εικόνα 9.7 Κλάση PageDataRetrieval (3)

```

109     };
110     leipeiThemPeriex_imageLinks.addAll(Arrays.asList(leipeiThemPeriex_images));
111     imageCategory = MinorClass.addDictValues(
112         leipeiThemPeriex_imageLinks,
113         imageCategory,
114         "ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ");
115
116
117     String tropopEikona_images[] = {
118         "https://media.ellinikahoaxes.gr/uploads/2020/08/tropopoihmenh-eikona-full.png"
119     };
120     tropopEikona_imageLinks.addAll(Arrays.asList(tropopEikona_images));
121     imageCategory = MinorClass.addDictValues(
122         tropopEikona_imageLinks,
123         imageCategory,
124         "ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ");
125
126
127     String pseudes_images[] = {
128         "https://media.ellinikahoaxes.gr/uploads/2019/05/False-2.png",
129         "https://media.ellinikahoaxes.gr/uploads/2019/01/50649971_401410660631160_5938137417407004672_n-1-300x108.png",
130         "https://media.ellinikahoaxes.gr/uploads/2019/01/50649971_401410660631160_5938137417407004672_n-1-1024x369.png",
131         "https://media.ellinikahoaxes.gr/uploads/2019/04/Hoaxes-New-Seals-Fake-1024x369.png",
132         "https://media.ellinikahoaxes.gr/uploads/2019/04/Hoaxes-New-Seals-Fake-300x108.png",
133         "https://media.ellinikahoaxes.gr/uploads/2019/06/False-1-300x108.png",
134         "https://media.ellinikahoaxes.gr/uploads/2019/06/False-1-1024x369.png",
135         "https://media.ellinikahoaxes.gr/uploads/2019/11/Pseudes-1024x369.png",
136         "https://media.ellinikahoaxes.gr/uploads/2019/10/Pseudes-300x108.png",
137         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50649971_401410660631160_5938137417407004672_n-1-300x108.png",
138         "https://media.ellinikahoaxes.gr/uploads/2019/06/False-300x108.png",
139         "https://media.ellinikahoaxes.gr/uploads/2019/05/False-2-1024x369.png",
140         "https://i2.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50649971_401410660631160_5938137417407004672_n-1.png?fit=696%2C251&ssl=1",
141         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50649971_401410660631160_5938137417407004672_n-1-1024x369.png",
142         "https://media.ellinikahoaxes.gr/uploads/2019/05/False-1024x369.png"
143     };
144     pseudes_imageLinks.addAll(Arrays.asList(pseudes_images));
145     imageCategory = MinorClass.addDictValues(
146         pseudes_imageLinks,
147         imageCategory,
148         "ΨΕΥΔΕΣ");
149
150
151     String anakyklEidshsh_images[] = {
152         "https://media.ellinikahoaxes.gr/uploads/2019/01/50614956_969046413293560_8288493841338597376_n-300x108.png",
153         "https://media.ellinikahoaxes.gr/uploads/2019/05/Recycled-News-2-1024x369.png",
154         "https://media.ellinikahoaxes.gr/uploads/2019/05/Recycled-News-1-300x108.png",
155         "https://i0.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50614956_969046413293560_8288493841338597376_n.png?fit=696%2C251&ssl=1",
156         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50614956_969046413293560_8288493841338597376_n-300x108.png",
157         "https://media.ellinikahoaxes.gr/uploads/2019/05/Recycled-News-1024x369.png",
158         "https://media.ellinikahoaxes.gr/uploads/2019/01/50614956_969046413293560_8288493841338597376_n-1024x369.png",
159         "https://i0.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/4.png?fit=696%2C251&ssl=1",
160         "https://media.ellinikahoaxes.gr/uploads/2019/05/Recycled-News-2-300x108.png"
161     };
162     anakyklEidshsh_imageLinks.addAll(Arrays.asList(anakyklEidshsh_images));

```

Εικόνα 9.8 Κλάση PageDataRetrieval (4)

```

163     imageCategory = MinorClass.addDictValues(
164         anakyklEidhsh_imageLinks,
165         imageCategory,
166         "ΑΝΑΚΥΚΛΩΜΕΝΗ ΕΙΔΗΣΗ");
167
168
169     String synwmos_images[] = {
170         "https://media.ellinikahoaxes.gr/uploads/2020/08/sinomosologia-full.png",
171         "https://media.ellinikahoaxes.gr/uploads/2019/10/Conspiracy-1024x369.png",
172         "https://media.ellinikahoaxes.gr/uploads/2019/01/7-1-300x108.png",
173         "https://media.ellinikahoaxes.gr/uploads/2019/10/Conspiracy-300x108.png",
174         "https://media.ellinikahoaxes.gr/uploads/2019/03/7-300x108.png",
175         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/7-1-300x108.png"
176     };
177     synwmos_imageLinks.addAll(Arrays.asList(synwmos_images));
178     imageCategory = MinorClass.addDictValues(
179         synwmos_imageLinks,
180         imageCategory,
181         "ΣΥΝΩΜΟΣΕΙΟΛΟΓΙΑ");
182
183
184     String tropopVinteo_images[] = {
185         "https://media.ellinikahoaxes.gr/uploads/2020/08/tropopoihmeno-video-full.png"
186     };
187     tropopVinteo_imageLinks.addAll(Arrays.asList(tropopVinteo_images));
188     imageCategory = MinorClass.addDictValues(
189         tropopVinteo_imageLinks,
190         imageCategory,
191         "ΤΡΟΠΟΠΟΙΗΜΕΝΟ ΒΙΝΤΕΟ");
192
193
194     String miksi_images[] = {
195         "https://media.ellinikahoaxes.gr/uploads/2019/05/Mix-True-False-1.png",
196         "https://media.ellinikahoaxes.gr/uploads/2019/05/Mix-True-False-1-300x108.png",
197         "https://media.ellinikahoaxes.gr/uploads/2019/05/Mix-True-False-1-1024x369.png",
198         "https://media.ellinikahoaxes.gr/uploads/2019/05/Mix-True-False-300x108.png",
199         "https://media.ellinikahoaxes.gr/uploads/2019/05/Mix-True-False-1024x369.png"
200     };
201     miksi_imageLinks.addAll(Arrays.asList(miksi_images));
202     imageCategory = MinorClass.addDictValues(miksi_imageLinks,
203         imageCategory,
204         "ΜΙΞΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ");
205
206
207     String pseypoepis_images[] = {
208         "https://media.ellinikahoaxes.gr/uploads/2020/08/pseypoepisthmh-full.png",
209         "https://media.ellinikahoaxes.gr/uploads/2019/02/3-1-300x108.png",
210         "https://media.ellinikahoaxes.gr/uploads/2019/02/3-1-1024x369.png",
211         "https://media.ellinikahoaxes.gr/uploads/2019/05/Pseudoscience-2-300x108.png",
212         "https://media.ellinikahoaxes.gr/uploads/2019/05/Pseudoscience-1-1024x369.png",
213         "https://media.ellinikahoaxes.gr/uploads/2019/06/Pseudoscience-1-1024x369.png",
214         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/02/3-1-300x108.png",
215         "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/02/3-300x108.png",
216         "https://media.ellinikahoaxes.gr/uploads/2019/05/Pseudoscience-3-300x108.png",

```

Εικόνα 9.9 Κλάση PageDataRetrieval (5)

```

217     "https://media.ellinikahoaxes.gr/uploads/2019/05/Pseudoscience-300x108.png",
218     "https://media.ellinikahoaxes.gr/uploads/2019/05/Pseudoscience-4-1024x369.png"
219 };
220 pseydoepis_imageLinks.addAll(Arrays.asList(pseydoepis_images));
221 imageCategory = MinorClass.addDicitValues(
222     pseydoepis_imageLinks,
223     imageCategory,
224     "ΨΕΥΔΟΕΠΙΣΤΗΜΗ");
225
226
227 String apath_images[] = {
228     "https://media.ellinikahoaxes.gr/uploads/2020/08/apati-full.png",
229     "https://media.ellinikahoaxes.gr/uploads/2019/08/Scam-300x108.png",
230     "https://media.ellinikahoaxes.gr/uploads/2019/11/Apati-1024x369.png",
231     "https://media.ellinikahoaxes.gr/uploads/2019/08/Scam-1024x369.png",
232     "https://media.ellinikahoaxes.gr/uploads/2019/02/απατη.png",
233     "https://media.ellinikahoaxes.gr/uploads/2019/02/απατη-300x108.png",
234     "https://media.ellinikahoaxes.gr/uploads/2019/08/Scam-300x108.png",
235     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/02/απατη-300x108.png"
236 };
237 apath_imageLinks.addAll(Arrays.asList(apat_images));
238 imageCategory = MinorClass.addDicitValues(
239     apath_imageLinks,
240     imageCategory,
241     "ΑΠΑΤΗ");
242
243
244 String satira_images[] = {
245     "https://media.ellinikahoaxes.gr/uploads/2020/08/satira-full.png",
246     "https://media.ellinikahoaxes.gr/uploads/2019/06/Εάτιπα.png",
247     "https://media.ellinikahoaxes.gr/uploads/2019/06/Εάτιπα-300x108.png"
248 };
249 satira_imageLinks.addAll(Arrays.asList(satira_images));
250 imageCategory = MinorClass.addDicitValues(
251     satira_imageLinks,
252     imageCategory,
253     "ΕΑΤΙΠΑ");
254
255
256 String likeFarming_images[] = {
257     "https://media.ellinikahoaxes.gr/uploads/2019/11/Like framing.png",
258     "https://media.ellinikahoaxes.gr/uploads/2019/05/Like-Farming-300x108.png",
259     "https://media.ellinikahoaxes.gr/uploads/2019/01/10-1-300x108.png",
260     "https://media.ellinikahoaxes.gr/uploads/2019/11/Like framing-300x108.png",
261     "https://media.ellinikahoaxes.gr/uploads/2019/05/Like-Farming-1024x369.png",
262     "https://media.ellinikahoaxes.gr/uploads/2019/11/Like framing-1024x369.png",
263     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/10-1-300x108.png"
264 };
265 likeFarming_imageLinks.addAll(Arrays.asList(likeFarming_images));
266 imageCategory = MinorClass.addDicitValues(
267     likeFarming_imageLinks,
268     imageCategory,
269     "LIKE FARMING");
270

```

Εικόνα 9.10 Κλάση PageDataRetrieval (6)

```
271
272 String kindynolog_images[] = {
273     "https://media.ellinikahoaxes.gr/uploads/2020/08/kindinologia-full.png",
274     "https://media.ellinikahoaxes.gr/uploads/2019/06/Fear-Mongering-300x108.png",
275     "https://media.ellinikahoaxes.gr/uploads/2019/06/Fear-Mongering-1024x369.png",
276     "https://media.ellinikahoaxes.gr/uploads/2019/05/Fear-Mongering-1-300x108.png",
277     "https://media.ellinikahoaxes.gr/uploads/2019/05/Fear-Mongering-1024x369.png",
278     "https://media.ellinikahoaxes.gr/uploads/2019/04/Fear-Mongering-1-300x108.png"
279 };
280 kindynolog_imageLinks.addAll(Arrays.asList(kindynolog_images));
281 imageCategory = MinorClass.addDictValues(
282     kindynolog_imageLinks,
283     imageCategory,
284     "KINDYNOROTIA");
285
286
287 String anepiv_images[] = {
288     "https://media.ellinikahoaxes.gr/uploads/2020/08/anepebebaioto-full.png",
289     "https://media.ellinikahoaxes.gr/uploads/2019/03/49748456_22461226524239483_1258659364596613120_n-1024x369.png"
290 };
291 anepiv_imageLinks.addAll(Arrays.asList(anepiv_images));
292 imageCategory = MinorClass.addDictValues(
293     anepiv_imageLinks,
294     imageCategory,
295     "ANENIBEBAIOTO");
296
297
298 String clickbait_images[] = {
299     "https://media.ellinikahoaxes.gr/uploads/2020/08/clickbait-full.png",
300     "https://media.ellinikahoaxes.gr/uploads/2019/01/49213291_964918363707769_4401650851953770496_n-1-300x108.png",
301     "https://media.ellinikahoaxes.gr/uploads/2019/04/Clickbait-300x108.png",
302     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/49213291_964918363707769_4401650851953770496_n-1-300x108.png",
303     "https://i0.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/49213291_964918363707769_4401650851953770496_n-1.png?fit=696&2C251&ssl=1",
304     "https://i2.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/49213291_964918363707769_4401650851953770496_n.png?fit=696&2C251&ssl=1",
305     "https://media.ellinikahoaxes.gr/uploads/2019/04/Clickbait-1024x369.png",
306     "https://media.ellinikahoaxes.gr/uploads/2019/03/Clickbait-1024x369.png",
307     "https://www.ellinikahoaxes.gr/wp-content/uploads/2019/01/49213291_964918363707769_4401650851953770496_n-1-1024x369.png"
308 };
309 clickbait_imageLinks.addAll(Arrays.asList(clickbait_images));
310 imageCategory = MinorClass.addDictValues(
311     clickbait_imageLinks,
312     imageCategory,
313     "CLICKBAIT");
314
315
316 String pseudsIsxyr_images[] = {
317     "https://media.ellinikahoaxes.gr/uploads/2019/01/50266413_631324187332633_7251492597812690944_n-300x108.png",
318     "https://media.ellinikahoaxes.gr/uploads/2019/01/50266413_631324187332633_7251492597812690944_n-1024x369.png",
319     "https://media.ellinikahoaxes.gr/uploads/2019/10/Pseudis_isxurismos-1-1024x369.png",
320     "https://media.ellinikahoaxes.gr/uploads/2019/06/False-Assertion-1024x369.png",
321     "https://media.ellinikahoaxes.gr/uploads/2019/06/False-Assertion-300x108.png",
322     "https://media.ellinikahoaxes.gr/uploads/2019/06/False-Assertion.png",
323     "https://media.ellinikahoaxes.gr/uploads/2019/04/False-Assertion-1-300x108.png",
324     "https://i0.wp.com/www.ellinikahoaxes.gr/wp-content/uploads/2019/01/50266413_631324187332633_7251492597812690944_n.png?fit=696&2C251&ssl=1",
```


Εικόνα 9.13 Κλάση PageDataRetrieval (9)

```

433 monthConvert.put("Οκτωβρίου", "10");
434 monthConvert.put("Νοεμβρίου", "11");
435 monthConvert.put("Δεκεμβρίου", "12");
436
437 System.setOut(new PrintStream(System.out, true, "UTF8"));
438
439 //Reads a file with already retrieved links data
440 File readFile = new File(allData);
441
442 /* If the file exists, all the links in the file are added to the
443 hashset "accessedLinks"
444 */
445 try (BufferedReader in = new BufferedReader(
446     new InputStreamReader(
447         new FileInputStream(readFile), "UTF8")
448 )) {
449     String str;
450
451     while ((str = in.readLine()) != null) {
452
453         if (!str.trim().equals("")) {
454             String[] strSplit;
455             strSplit = str.split("\t");
456             accessedLinks.add(strSplit[0].replace("\"", ""));
457         }
458     }
459 //If the file does not exist, it is created and a header is added
460 } catch (Exception e) {
461     System.out.println(e.getMessage());
462     File linkFile = new File(allData);
463     try (Writer textWriter = new BufferedWriter(
464         new OutputStreamWriter(
465             new FileOutputStream(linkFile, true),
466             StandardCharsets.UTF_8)
467     )) {
468         System.setOut(new PrintStream(System.out, true, "UTF8"));
469         String fieldName =
470             "Link\tTitle\tDate\tAuthor\tClaim\tConclusion\tCategory\tVerdict\n";
471         textWriter.append(fieldName);
472         textWriter.flush();
473     }
474 }
475
476 System.out.println(accessedLinks);
477
478 /* Runs the hashset of the links that were retrieved from the file
479 "FakeData.csv (function getOldLinks())
480 */
481 for (String l : links) {
482     System.out.println("\n-----\n");
483
484     //If the link had been already accessed we ignore it
485     if (accessedLinks.contains(l)) {
486         System.out.println(

```

Εικόνα 9.14 Κλάση PageDataRetrieval (10)

```

487         " - The link '" + l + "' already exists in the list!");
488     } else {
489         Link link = new Link (l);
490
491         if (Link.processNewLink(l)){
492             System.out.println("Link: " + link.getLink());
493             System.out.println("Title: " + link.getTitle());
494             System.out.println("Date: " + link.getDate());
495             System.out.println("Author: " + link.getAuthor());
496             System.out.println("Claims: ");
497
498             for (String cl : link.getClaimsAll()) {
499                 System.out.println(cl);
500             }
501
502             System.out.println("Conclusions: ");
503
504             for (String con : link.getConclusionsAll()) {
505                 System.out.println(con);
506             }
507
508             System.out.println("Categories: " + link.getCategory());
509             System.out.println("Verdicts: " + link.getVerdict());
510         } else {
511             System.out.println("Link: " + l);
512             System.out.println("does not contain any fakenews");
513         }
514
515         Thread.sleep(1000);
516     }
517 }
518
519
520 }
521
522 /* Scans the HTML code of the webpage and returns the HTML code as a string
523 */
524 public static String getURLContent(String urlPath) throws
525     MalformedURLException, IOException {
526     URL fakeNewsPage = new URL(urlPath);
527     URLConnection fakeNewsCon = fakeNewsPage.openConnection();
528
529     String pageContent;
530
531     try (BufferedReader in = new BufferedReader(
532         new InputStreamReader(
533             fakeNewsCon.getInputStream(), "UTF-8")
534     )) {
535         pageContent = "";
536         String inputLine;
537
538         while ((inputLine = in.readLine()) != null) {
539             pageContent += (inputLine + "\n");
540         }

```


Εικόνα 9.15 Κλάση PageDataRetrieval (11)

```
541     }
542
543     return pageContent;
544
545 }
546
547 /* Reads the file with old links and adds these links
548 to the hashset "alreadyRetrievedLinks".
549 */
550 public static void getOldLinks() throws IOException {
551     System.out.println(
552         "# Reading the links in the text file '"
553         + HOAXES_LINKS_FILENAME + "'");
554
555     File file = new File(HOAXES_LINKS_FILENAME);
556
557     try (BufferedReader in = new BufferedReader(
558         new InputStreamReader(
559             new FileInputStream(file), "UTF8")
560     )) {
561         String str;
562
563         while ((str = in.readLine()) != null) {
564
565             if (!str.trim().equals("")) {
566                 links.add(str);
567             }
568         }
569
570     } catch (Exception e) {
571         System.out.println(e.getMessage());
572     }
573
574     System.out.println(
575         " - Total number of links in the file: " + links.size());
576 }
577 }
```

Εικόνα 9.16 Κλάση Link (1)

```

1 package gr.hmu.mst.fakenews;
2
3 import static gr.hmu.mst.fakenews.PageDataRetrieval.imageCategory;
4 import static gr.hmu.mst.fakenews.PageDataRetrieval.linkCategory;
5 import static gr.hmu.mst.fakenews.PageDataRetrieval.monthConvert;
6 import static gr.hmu.mst.fakenews.PageDataRetrieval.validCatVerdictValues;
7 import java.io.BufferedWriter;
8 import java.io.File;
9 import java.io.FileOutputStream;
10 import java.io.IOException;
11 import java.io.OutputStreamWriter;
12 import java.io.PrintStream;
13 import java.io.Writer;
14 import java.nio.charset.StandardCharsets;
15 import java.util.ArrayList;
16 import java.util.Arrays;
17 import java.util.HashSet;
18
19 /**
20  *
21  * @author zabal
22  */
23 public class Link {
24
25     private static String link;
26     private static String recordedData;
27     private static String title;
28     private static String date;
29     private static String author;
30     private static HashSet<String> claimsAll;
31     private static HashSet<String> conclusionsAll;
32
33     private static String category;
34     private static String verdict;
35
36     public Link(String link) {
37         Link.link = link;
38     }
39
40     public String getLink() {
41         return link;
42     }
43     public static void setLink(String link) {
44         Link.link = link;
45     }
46
47     public String getTitle() {
48         return title;
49     }
50     public static void setTitle(String title) {
51         Link.title = title;
52     }
53

```

Εικόνα 9.17 Κλάση Link (2)

```

54 | public String getDate() {
55 |     return date;
56 | }
57 | public static void setDate(String date) {
58 |     Link.date = date;
59 | }
60 |
61 | public String getAuthor() {
62 |     return author;
63 | }
64 | public static void setAuthor (String author) {
65 |     Link.author = author;
66 | }
67 |
68 | public HashSet<String>getClaimsAll() {
69 |     return claimsAll;
70 | }
71 | public static void setClaimsAll(HashSet <String> claimsAll) {
72 |     Link.claimsAll = claimsAll;
73 | }
74 |
75 | public HashSet<String> getConclusionsAll() {
76 |     return conclusionsAll;
77 | }
78 | public static void setConclusionsAll(HashSet <String>conclusionsAll) {
79 |     Link.conclusionsAll = conclusionsAll;
80 | }
81 |
82 | public String getCategory() {
83 |     return category;
84 | }
85 | public static void setCategory(String category) {
86 |     Link.category = category;
87 | }
88 |
89 | public String getVerdict() {
90 |     return verdict;
91 | }
92 | public static void setVerdict(String verdict) {
93 |     Link.verdict = verdict;
94 | }
95 |
96 | public String getRecordedData() {
97 |     return recordedData;
98 | }
99 |
100 | public static void setRecordedData(String recordedData) {
101 |     Link.recordedData = recordedData;
102 | }
103 |
104 | /**
105 |  * @param args the command line arguments
106 |  * @throws java.io.IOException

```

Εικόνα 9.18 Κλάση Link (3)

```

107 L      */
108 □ public static void main(String[] args) throws IOException {
109     System.setOut(new PrintStream(System.out, true, "UTF8"));
110     processNewLink(link);
111 }
112 □ public static boolean processNewLink(String str) throws IOException {
113
114
115     ArrayList<String> categories = new ArrayList<>();
116     ArrayList<String> claims = new ArrayList<>();
117     ArrayList<String> conclusions = new ArrayList<>();
118     ArrayList<String> rawCategories = new ArrayList<>();
119     ArrayList<String> rawImageValues = new ArrayList<>();
120
121     String openedLinkContent = PageDataRetrieval.getURLContent(link);
122     System.out.println(" - Writing the link: "+link);
123
124     //Adds the link to the hashset after accessing it
125     PageDataRetrieval.accessedLinks.add(link);
126
127     //Extracts categories from H1 field
128     ArrayList<String> catVerdicts = new ArrayList<>();
129     int spanCategoryStart = openedLinkContent.indexOf(
130         "<span class=\"breadcrumb_last\"");
131     int spanCategoryEnd = openedLinkContent.indexOf(
132         "<h1", spanCategoryStart);
133     String spanCategory = openedLinkContent.substring(
134         spanCategoryStart, spanCategoryEnd);
135     String[] categoryList= spanCategory.split("<a href=\"");
136     rawCategories.addAll(Arrays.asList(categoryList));
137     rawCategories.remove(0);
138     String categoryString = "";
139
140     //Normalizes links in the hashset
141     for (String rawCat : rawCategories) {
142         int rawCatEnd = rawCat.indexOf("\" rel");
143         String categoryKey = rawCat.substring(0, rawCatEnd);
144         String categoryValue = linkCategory.get(categoryKey);
145
146         /* If the category exists in linkCategory dictionary,
147         adds the category to the "categories" hashset,
148         expands the string "categoryString",
149         and adds the category to the "catVerdicts" hashset
150         */
151         if (categoryValue != null){
152             categories.add(categoryValue);
153             categoryString = categoryString + "," + categoryValue;
154             catVerdicts.add(categoryValue);
155         }
156     }
157
158     if (!categories.isEmpty()) {

```

Εικόνα 9.19 Κλάση Link (4)

```
159     categoryString = categories.toString();
160     categoryString = categoryString.replace("[", "");
161     categoryString = categoryString.replace("]", "");
162 }
163
164 int divImageStart;
165 int divImageEnd;
166 String divImage;
167 int figImageStart;
168 int figImageEnd;
169 String figImage;
170 int imageStart;
171 int imageEnd;
172 String openedLinkImage;
173 String openedLinkImageValue;
174 String imageValueString;
175
176 divImageStart = openedLinkContent.indexOf(
177     "<div class=\"col-12 text-center stams\"");
178 divImageEnd = openedLinkContent.indexOf("</div",divImageStart);
179 divImage = openedLinkContent.substring(divImageStart, divImageEnd);
180 imageStart = divImage.indexOf("data-lazy-src");
181
182 //Tries to extract values from the images on the webpage (if exist)
183
184 while (imageStart != -1) {
185     imageEnd = divImage.indexOf("\" ",imageStart);
186     openedLinkImage = divImage.substring(imageStart+15, imageEnd);
187     openedLinkImageValue = imageCategory.get(openedLinkImage);
188
189     /* If this value exists in the "imageCategory" dictionary,
190     adds the value to the hashset "catVerdicts"
191     */
192     if (openedLinkImageValue != null) { //if
193         catVerdicts.add(openedLinkImageValue);
194         rawImageValues.add(openedLinkImageValue);
195     }
196
197     imageStart = divImage.indexOf("data-lazy-src", imageEnd);
198
199 }
200
201 imageValueString = rawImageValues.toString();
202 imageValueString = imageValueString.replace("[", "");
203 imageValueString = imageValueString.replace("]", "");
204
205 //If the previous if-clause had not been executed
206 if (imageValueString.isEmpty()) {
207
208     try {
209         divImageStart = openedLinkContent.indexOf(
210             "<div class=\"wp-block-columns is-layout-flex wp-container-6\"");
211         divImageEnd = openedLinkContent.indexOf("<p",divImageStart);
212         divImage = openedLinkContent.substring(
```

Εικόνα 9.20 Κλάση Link (5)

```
213         divImageStart, divImageEnd);
214     figImageStart = divImage.indexOf(
215         "<figure class=\"wp-block-image size-large is-resized\"");
216
217     while (figImageStart != -1) {
218         figImageEnd = divImage.indexOf("</figure", figImageStart);
219         figImage = divImage.substring(figImageStart, figImageEnd);
220         imageStart = figImage.indexOf("data-lazy-src");
221         imageEnd = figImage.indexOf("<noscript", imageStart);
222         openedLinkImage = figImage.substring(
223             imageStart+15, imageEnd-3).replace("\\", "");
224         openedLinkImageValue = imageCategory.get(openedLinkImage);
225
226         if (openedLinkImageValue != null) {
227             rawImageValues.add(openedLinkImageValue);
228             catVerdicts.add(openedLinkImageValue);
229         }
230
231         figImageStart = divImage.indexOf(
232             "<figure class=\"wp-block-image size-large is-resized\"",
233             figImageEnd);
234
235     }
236
237     imageValueString = rawImageValues.toString();
238     imageValueString = imageValueString.replace("[", "");
239     imageValueString = imageValueString.replace("]", "");
240
241     } catch (Exception e1){
242     }
243 }
244
245 int mainImageDivStart;
246 int mainImageDivEnd;
247 String mainDivImage;
248
249 //If the previous if-clause had not been executed
250 if (imageValueString.isEmpty()) {
251
252     try {
253         mainImageDivStart = openedLinkContent.indexOf(
254             "<div class=\"single-article post\"");
255         mainImageDivEnd = mainImageDivStart + 5000;
256         mainDivImage = openedLinkContent.substring(
257             mainImageDivStart, mainImageDivEnd);
258         int divClassStart = mainDivImage.indexOf(
259             "<div class=\"wp-block-image\"");
260         int figureClassStart= mainDivImage.indexOf(
261             "figure class=\"wp-block-image\"");
262
263         if
264         (
265             (mainDivImage.contains("figure class=\"wp-block-image\"")) &&
266             (mainDivImage.contains("<div class=\"wp-block-image\"")) &&
```

Εικόνα 9.21 Κλάση Link (6)

```

267         (divClassStart < figureClassStart)
268     ){
269         divImageStart = mainDivImage.indexOf(
270             "<div class=\"wp-block-image\"");
271     } else {
272         divImageStart = mainDivImage.indexOf(
273             "class=\"wp-block-image\"");
274     }
275
276     while (divImageStart != -1 ) {
277         divImageEnd = mainDivImage.indexOf(
278             "</figure\"", divImageStart);
279
280         if (divImageEnd == -1) {
281             break;
282         }
283         divImage = mainDivImage.substring(
284             divImageStart, divImageEnd);
285         imageStart = divImage.indexOf("data-lazy-src");
286         imageEnd = divImage.indexOf("<noscript\"", imageStart);
287         openedLinkImage = divImage.substring(
288             imageStart+15, imageEnd-3).replace("\"", "");
289         openedLinkImageValue = imageCategory.get(openedLinkImage
290
291         if (openedLinkImageValue != null) {
292             rawImageValues.add(openedLinkImageValue);
293             catVerdicts.add(openedLinkImageValue);
294         }
295
296         divImageStart = mainDivImage.indexOf(
297             "class=\"wp-block-image\"", divImageEnd);
298
299     }
300
301     imageValueString = rawImageValues.toString();
302     imageValueString = imageValueString.replace("[", "");
303     imageValueString = imageValueString.replace("]", "");
304 } catch (Exception e2) {
305     imageValueString = "";
306 }
307
308
309 Boolean noRecord = true;
310
311 if (catVerdicts.isEmpty() == false){
312
313     for (String cV : catVerdicts) {
314
315         if (validCatVerdictValues.contains(cV)) {
316             noRecord = false;
317             break;
318         }
319     }

```

Εικόνα 9.22 Κλάση Link (7)

```

320 }
321
322 if (catVerdicts.isEmpty()){
323     noRecord = false;
324 }
325
326 if (noRecord == false) {
327
328     //Retrieves the title
329     int titleStart = openedLinkContent.indexOf("<h1");
330     int titleEnd = openedLinkContent.indexOf("</h1>");
331     String openedLinkTitle = openedLinkContent.substring(
332         titleStart+4, titleEnd);
333
334     if (openedLinkTitle.contains("<strong>")){
335         openedLinkTitle = openedLinkTitle.replace(
336             "</strong>", "").replace("<strong>", "");
337     }
338
339     //Retrieves the date
340     int dateStart = openedLinkContent.indexOf(
341         "<div class=\"blog-date\"");
342     int dateEnd = openedLinkContent.indexOf("</div>",dateStart);
343     String openedLinkDate = openedLinkContent.substring(
344         dateStart+23, dateEnd);
345
346     //Splits the date in Greek into 3: date, month, and year
347     String[] dateName = openedLinkDate.split(" ");
348     dateName[1] = monthConvert.get(dateName[1]);
349     String dateNumber = dateName[0] + "-" + dateName[1] + "-"
350         + dateName[2];
351
352     //Retrieves the author
353     int authorStart = openedLinkContent.indexOf(
354         "rel=\"author\" data-wpel-link=\"internal\"");
355     int authorEnd = openedLinkContent.indexOf("</a>", authorStart);
356     String openedLinkAuthor = openedLinkContent.substring(
357         authorStart+39, authorEnd);
358
359     String openedLinkClaim = "";
360     int divClaimStart;
361     int divClaimEnd;
362     int pClaimStart;
363     int pClaimEnd;
364     String pClaim;
365     String divClaim;
366     int claimStart;
367     int claimEnd;
368
369     String openedLinkConclusion = "";
370     int pConclusionStart;
371     int pConclusionEnd;

```


Εικόνα 9.23 Κλάση Link (8)

```
372 String pConclusion;
373 int conclusionStart;
374
375 //Retrieves the claim/claims
376 divClaimStart = openedLinkContent.indexOf("<div class=\"Claim\"");
377
378 while (divClaimStart != -1) {
379     divClaimEnd = openedLinkContent.indexOf(
380         "</div", divClaimStart);
381     divClaim = openedLinkContent.substring(
382         divClaimStart, divClaimEnd);
383     claimStart = divClaim.indexOf("<p");
384     claimEnd = divClaim.indexOf("</p");
385     openedLinkClaim = divClaim.substring(
386         claimStart + 3, claimEnd);
387     //Adds all found claims to the hashset "claims"
388     claims.add(openedLinkClaim);
389     divClaimStart = openedLinkContent.indexOf(
390         "<div class=\"Claim\"", divClaimEnd);
391 }
392
393 int divConclusionStart;
394 int divConclusionEnd;
395 String divConclusion;
396 int conclusionEnd;
397
398 //Retrieves the conclusion/conclusions
399 divConclusionStart = openedLinkContent.indexOf("<div class=\"conclusion\"");
400
401 while (divConclusionStart != -1) {
402     divConclusionEnd = openedLinkContent.indexOf(
403         "</div", divConclusionStart);
404     divConclusion = openedLinkContent.substring(
405         divConclusionStart, divConclusionEnd);
406     conclusionStart = divConclusion.indexOf("<p");
407     conclusionEnd = divConclusion.indexOf("</p");
408     openedLinkConclusion = divConclusion.substring(
409         conclusionStart + 3, conclusionEnd);
410     //Adds all found conclusions to the hashset "conclusions"
411     conclusions.add(openedLinkConclusion);
412     divConclusionStart = openedLinkContent.indexOf(
413         "<div class=\"conclusion\"", divConclusionEnd);
414 }
415
416 /* If claims had not been retrieved earlier,
417 The webpage has a different structure, need to try again
418 */
419 if (openedLinkClaim.isEmpty() ) {
420
421     try {
422         pClaimStart = openedLinkContent.indexOf(
423             "<p class=\"has-background has-luminous-vivid-orange-background-color\"");
424         pConclusionStart = openedLinkContent.indexOf(
```

Εικόνα 9.24 Κλάση Link (9)

```
425         "<p class=\"has-background has-vivid-green-cyan-background-color\"");
426
427     while (pConclusionStart != -1) {
428
429         //Tries to retrieve a claim
430         pClaimEnd = openedLinkContent.indexOf(
431             "</p", pClaimStart);
432         pClaim = openedLinkContent.substring(
433             pClaimStart, pClaimEnd);
434         claimStart = pClaim.indexOf("<br");
435         openedLinkClaim = pClaim.substring(
436             claimStart+4).replace("<br>", "");
437         claims.add(openedLinkClaim);
438         pClaimStart = openedLinkContent.indexOf(
439             "<p class=\"has-background has-luminous-vivid-orange-background-color\"",
440             pClaimEnd);
441
442         //Tries to retrieve a conclusion
443         pConclusionEnd = openedLinkContent.indexOf(
444             "</p", pConclusionStart);
445         pConclusion = openedLinkContent.substring(
446             pConclusionStart, pConclusionEnd);
447         conclusionStart = pConclusion.indexOf("<br");
448         openedLinkConclusion = pConclusion.substring(
449             conclusionStart + 4).replace("<br>", "");
450         conclusions.add(openedLinkConclusion);
451         pConclusionStart = openedLinkContent.indexOf(
452             "< p class=\"has-background has-vivid-green-cyan-background-color\"",
453             pConclusionEnd);
454     }
455 } catch (Exception e) {
456 }
457 }
458
459 //Adds retrieved data to the file
460 if (openedLinkClaim.isEmpty()){
461     openedLinkClaim = "-";
462     claims.add(openedLinkClaim);
463 }
464
465 if (openedLinkConclusion.isEmpty()){
466     openedLinkConclusion = "-";
467     conclusions.add(openedLinkConclusion);
468 }
469
470 if (imageValueString.isEmpty()){
471     imageValueString = "-";
472 }
473
474 if (categoryString.isEmpty()) {
475     categoryString = "-";
476 }
477 }
```

Εικόνα 9.25 Κλάση Link (10)

```

478      /* Writes as many records as were retrieved from this webpage
479      (Number of claims and conclusions is always equal)
480      */
481      int numOfRecords = claims.size();
482
483      setRecordedData("");
484
485      File linkFile = new File(PageDataRetrieval.allData);
486
487      try (Writer textWriter = new BufferedWriter(
488          new OutputStreamWriter(
489              new FileOutputStream(
490                  linkFile, true), StandardCharsets.UTF_8)
491          )) {
492          System.setOut(new PrintStream(System.out, true, "UTF8"));
493          addNewLine (numOfRecords, claims, conclusions,
494                      openedLinkTitle, date[Close], openedLinkAuthor,
495                      categoryString, imageValueString, textWriter);
496          textWriter.flush();
497      }
498      return true;
499  } else {
500      return false;
501  }
502  }
503
504  private static void addNewLine (int numOfRecords,
505      ArrayList<String> claims, ArrayList<String> conclusions,
506      String openedLinkTitle, String dateNumber, String openedLinkAuthor,
507      String categoryString, String imageValueString, Writer textWriter)
508      throws IOException {
509      String recordData = "";
510      claimsAll = new HashSet<>();
511      conclusionsAll = new HashSet<>();
512
513      for (int i = 0; i < numOfRecords; i++ ) {
514
515          openedLinkTitle = openedLinkTitle.
516              replace("\n", " ").replace("\r", " ").
517              replace("&#8220;", "<")&#8221;".replace("&#8221;", ">").
518              replace("&#8211;", "-").replace("&#8216;", "<").
519              replace("&#8217;", ">").replace("&#8230;", "...").
520              replace("&#038;", "&");
521          openedLinkAuthor = openedLinkAuthor.
522              replace("\n", " ").replace("\r", " ");
523          String claimsStr = claims.get(i).
524              replace("\n", " ").replace("\r", " ").
525              replace("&#8220;", "<")&#8221;".replace("&#8221;", ">").
526              replace("&#8211;", "-").replace("&#8216;", "<").
527              replace("&#8217;", ">").replace("&#8230;", "...").
528              replace("&#038;", "&");
529          String conclusionsStr = conclusions.get(i).

```

Εικόνα 9.26 Κλάση Link (11)

```

530         replace("\n", " ").replace("\r", " ").
531         replace("&#8220;", "<")&#8221;,">").
532         replace("&#8211;", "-").replace("&#8216;", "`").
533         replace("&#8217;", "\'").replace("&#8230;", "...").
534         replace("&#038;", "&");
535     categoryString = categoryString.
536         replace("\n", " ").replace("\r", " ");
537     imageUrlString = imageUrlString.
538         replace("\n", " ").replace("\r", " ");
539     recordData = link + "\t"
540         + openedLinkTitle + "\t"
541         + dateNumber + "\t"
542         + openedLinkAuthor + "\t"
543         + claimsStr + "\t"
544         + conclusionsStr + "\t"
545         + categoryString + "\t"
546         + imageUrlString;
547     textWriter.append(recordData).append("\n");
548     claimsAll.add(claimsStr);
549     conclusionsAll.add(conclusionsStr);
550
551 }
552
553 setRecordedData(
554     link + "\t"
555         + openedLinkTitle + "\t"
556         + dateNumber + "\t"
557         + openedLinkAuthor + "\t"
558         + claimsAll
559         + conclusionsAll
560         + categoryString + "\t"
561         + imageUrlString);
562 setLink(link);
563 setTitle(openedLinkTitle);
564 setAuthor(openedLinkAuthor);
565 setDate(dateNumber);
566 setClaimsAll(claimsAll);
567 setConclusionsAll(conclusionsAll);
568 setCategory(categoryString);
569 setVerdict(imageUrlString);
570
571 }
572

```

Εικόνα 9.27 Κλάση MinorClass (1)

```
1 package gr.hmu.mst.fakenews;
2
3 import java.awt.FileDialog;
4 import java.awt.Frame;
5 import java.io.IOException;
6 import java.io.Writer;
7 import java.util.Arrays;
8 import java.util.HashMap;
9 import java.util.HashSet;
10
11 /**
12  *
13  * @author zabal
14  */
15 public class MinorClass {
16
17     public static void main(String[] args) {
18     }
19
20     public static String chooseFile(){
21         FileDialog dialog = new FileDialog((Frame)null, "Select File to Open");
22         dialog.setMode(FileDialog.LOAD);
23         dialog.setVisible(true);
24         String f = dialog.getFile();
25         dialog.dispose();
26         System.out.println(f + " was chosen.");
27         return f;
28     }
29
30     public static void addThemes(HashSet hs,String [] str){
31         hs.addAll(Arrays.asList(str));
32     }
33
34     public static HashMap addThemeDictValues (String str, HashSet hs,String [] strK, HashMap hm){
35         addThemes(hs,strK);
36         hm.put(str,hs);
37         return hm;
38     }
39
40     public static void addNewRecord(
41         Writer tWriter, HashMap<String,Integer> dict) throws IOException{
42         try (tWriter) {
43             int keysetFrequency;
44             String catFreq;
45             String keysetFrequencyStr;
46             System.out.println("Includes the keyset:");
47
48             for (String i: dict.keySet()) {
```

Εικόνα 9.28 Κλάση MinorClass (2)

```
49         keysetFrequency = dict.get(i);
50         keysetFrequencyStr = String.valueOf(keysetFrequency);
51         System.out.println(
52             "category = " + i + " value = " + keysetFrequency);
53         catFreq = i + ";" + keysetFrequencyStr;
54         tWriter.append(catFreq);
55         tWriter.append("\n");
56
57     }
58
59     tWriter.flush();
60
61 }
62 }
63
64 public static HashMap addDictValues(HashSet <String> hs,HashMap hm, Object obj){
65     for (String k : hs) {
66         hm.put(k, obj);
67     }
68     return hm;
69 }
70 }
```

Εικόνα 9.29 Κλάση DictionaryReader (1)

```

1 package gr.hmu.mst.takenews;
2
3 import java.awt.FileDialog;
4 import java.awt.Frame;
5 import java.io.BufferedReader;
6 import java.io.BufferedWriter;
7 import java.io.File;
8 import java.io.FileInputStream;
9 import java.io.FileNotFoundException;
10 import java.io.FileOutputStream;
11 import java.io.InputStreamReader;
12 import java.io.OutputStreamWriter;
13 import java.io.PrintStream;
14 import java.io.UnsupportedEncodingException;
15 import java.io.Writer;
16 import java.nio.charset.StandardCharsets;
17 import java.util.HashMap;
18
19
20 /**
21  *
22  * @author zabal
23  */
24
25 public class DictionaryReader {
26
27     private static final String KEYWORDS_DICTIONARY = "keywordsDictionary.csv";
28     private static String changedKeywords;
29     private static HashMap<String, String> dict = new HashMap<>();
30
31     /**
32      * @param args the command line arguments
33      * @throws java.io.FileNotFoundException
34      * @throws java.io.UnsupportedEncodingException
35      */
36     public static void main(String[] args) throws
37         FileNotFoundException, UnsupportedEncodingException {
38         System.setOut(new PrintStream(System.out, true, "UTF8"));
39         String unchangedKeywords = chooseFile();
40
41         changedKeywords = unchangedKeywords.substring(
42             0, unchangedKeywords.length()-4)
43             + "Changed.csv";
44
45         File readFile = new File(KEYWORDS_DICTIONARY);
46
47         retrieveDictionary(readFile);
48
49         File oldFile = new File(unchangedKeywords);
50         File newFile = new File(changedKeywords);
51
52         changeKeywords(oldFile, newFile);
53     }
54

```

Εικόνα 9.30 Κλάση DictionaryReader (2)

```

55
56
57 public static String chooseFile(){
58     FileDialog dialog = new FileDialog((Frame)null, "Select File to Open");
59     dialog.setMode(FileDialog.LOAD);
60     dialog.setVisible(true);
61     String f = dialog.getFile();
62     dialog.dispose();
63     System.out.println(f + " was chosen.");
64     return f;
65 }
66
67 public static void retrieveDictionary(File f){
68     try (BufferedReader in = new BufferedReader(
69         new InputStreamReader(
70             new FileInputStream(f), "UTF8")
71     )) {
72         String str;
73         String strKeyword;
74         String strValue;
75         int counter = 0;
76
77         while ((str = in.readLine()) != null) {
78             counter +=1;
79             String[] keywordValue= str.split(";", 2);
80             strKeyword = keywordValue[0];
81             strValue = keywordValue[1];
82             dict.put(strKeyword, strValue);
83             System.out.println(
84                 strKeyword + " - " + strValue + "counter = " + counter);
85         }
86
87     } catch (Exception e) {
88     }
89 }
90
91
92 public static void changeKeywords(File oldF, File newF) {
93     try (BufferedReader in = new BufferedReader(
94         new InputStreamReader(
95             new FileInputStream(oldF), "UTF8")
96     )) {
97         String strOld;
98         String strNew;
99         int counter = 0;
100
101         try (Writer textWriter = new BufferedWriter(
102             new OutputStreamWriter(
103                 new FileOutputStream(newF, true),
104                 StandardCharsets.UTF_8)
105         )) {
106             System.setOut(new PrintStream(System.out, true, "UTF8"));
107
108             while ((strOld = in.readLine()) != null) {

```


Εικόνα 9.31 Κλάση DictionaryReader (3)

```

109
110         if (counter < 1) {
111             strOld = strOld.substring(1);
112         }
113
114         counter +=1;
115
116         if (dict.containsKey(strOld)){
117             strNew = dict.get(strOld);
118             textWriter.append(strNew+"\n");
119         }
120
121         if (dict.containsKey(strOld)==false){
122             System.out.println(strOld + " not in dictionary");
123         }
124     }
125
126     textWriter.flush();
127
128     }
129 } catch (Exception e) {
130     System.out.println(e.getMessage());
131 }
132 }
133 }

```

Εικόνα 9.32 Κλάση SubjectSelectionMenu

```

1  package gr.hmu.mst.fakenews;
2
3  import javax.swing.JFrame;
4  import javax.swing.JOptionPane;
5
6
7  /**
8   *
9   * @author zabal
10  */
11  public class SubjectSelectionMenu {
12
13      public static void main(String[] args){
14
15      }
16
17      public static String showMenu(){
18          JFrame frame = new JFrame();
19          frame.setAlwaysOnTop(true);
20
21          Object [] cats = {"Covid", "Vaccination", "Russia", "Ukraine", "Greece",
22                          "Turkey", "USA", "Politics"};
23
24          Object selectionObject = JOptionPane.showInputDialog(
25              frame, "Theme", "Choose a theme:", JOptionPane.PLAIN_MESSAGE,
26              null, cats, cats[0]);
27          frame.dispose();
28          String selectionString = selectionObject.toString();
29          return selectionString;
30      }
31  }

```

Εικόνα 9.33 Κλάση ClaimDateRetrieval (1)

```

1 package gr.hmu.mst.fakenews;
2
3
4 import java.io.BufferedReader;
5 import java.io.BufferedWriter;
6 import java.io.File;
7 import java.io.FileInputStream;
8 import java.io.FileNotFoundException;
9 import java.io.FileOutputStream;
10 import java.io.InputStreamReader;
11 import java.io.OutputStreamWriter;
12 import java.io.PrintStream;
13 import java.io.UnsupportedEncodingException;
14 import java.io.Writer;
15 import java.nio.charset.StandardCharsets;
16 import java.util.Arrays;
17 import java.util.HashMap;
18 import java.util.HashSet;
19
20
21 /**
22  *
23  * @author zabal
24  */
25 public class ClaimDateRetrieval {
26
27     private static HashMap<String, Integer> dateHashMap= new HashMap<> ();
28     private static HashSet<String> vaccination = new HashSet<> ();
29     private static HashSet<String> covid = new HashSet<> ();
30     private static HashSet<String> ukraine = new HashSet<> ();
31     private static HashSet<String> russia = new HashSet<> ();
32     private static HashSet<String> greece = new HashSet<> ();
33     private static HashMap<String,HashSet> themeHashMap = new HashMap<> ();
34     private static HashSet<String> dates = new HashSet<> ();
35     private static HashSet<String> turkey = new HashSet<> ();
36     private static HashSet<String> politics = new HashSet<> ();
37     private static HashSet<String> usa = new HashSet<> ();
38     /**
39      * @param args the command line arguments
40      * @throws java.io.FileNotFoundException
41      * @throws java.io.UnsupportedEncodingException
42      */
43     public static void main(String[] args) throws
44         FileNotFoundException, UnsupportedEncodingException {
45         System.setOut(new PrintStream(System.out, true, "UTF8"));
46
47         String newFileName = MinorClass.chooseFile();
48
49         //Creates a hashset of month-years
50         String [] datesStr = {"May-2013", "June-2013", "July-2013", "August-2013",
51             "September-2013", "October-2013", "November-2013", "December-2013",

```

Εικόνα 9.34 Κλάση ClaimDateRetrieval (2)

```

52     "January-2014", "February-2014", "March-2014", "April-2014", "May-2014",
53     "June-2014", "July-2014", "August-2014", "September-2014", "October-2014",
54     "November-2014", "December-2014", "January-2015", "February-2015",
55     "March-2015", "April-2015", "May-2015", "June-2015", "July-2015",
56     "August-2015", "September-2015", "October-2015", "November-2015",
57     "December-2015", "January-2016", "February-2016", "March-2016",
58     "April-2016", "May-2016", "June-2016", "July-2016", "August-2016",
59     "September-2016", "October-2016", "November-2016", "December-2016",
60     "January-2017", "February-2017", "March-2017", "April-2017", "May-2017",
61     "June-2017", "July-2017", "August-2017", "September-2017", "October-2017",
62     "November-2017", "December-2017", "January-2018", "February-2018",
63     "March-2018", "April-2018", "May-2018", "June-2018", "July-2018",
64     "August-2018", "September-2018", "October-2018", "November-2018",
65     "December-2018", "January-2019", "February-2019", "March-2019",
66     "April-2019", "May-2019", "June-2019", "July-2019", "August-2019",
67     "September-2019", "October-2019", "November-2019", "December-2019",
68     "January-2020", "February-2020", "March-2020", "April-2020", "May-2020",
69     "June-2020", "July-2020", "August-2020", "September-2020", "October-2020",
70     "November-2020", "December-2020", "January-2021", "February-2021",
71     "March-2021", "April-2021", "May-2021", "June-2021", "July-2021",
72     "August-2021", "September-2021", "October-2021", "November-2021",
73     "December-2021", "January-2022", "February-2022", "March-2022",
74     "April-2022", "May-2022", "June-2022", "July-2022", "August-2022",
75     "September-2022", "October-2022", "November-2022", "December-2022",
76     "January-2023", "February-2023", "March-2023", "April-2023", "May-2023"};
77     dates.addAll(Arrays.asList(datesStr));
78
79     /* Adds keys and values to the dictionary consisting of dates and
80     their frequency
81     */
82     dateHashMap = MinorClass.addDictValues(dates, dateHashMap, 0);
83
84     //Creates strings of subjects of our interest
85     String covidKeywords [] = {"cov19", "covid", "κοροναϊ", "κορονοϊό",
86     "κορονοϊ", "κορωνοϊό", "κορωνοϊό", "κορωνοϊ", "κορωνοι", "sars-cov",
87     "2019-ncov", "mers-cov"};
88     String vacKeywords [] = {"εμβόλ", "εμβολ"};
89     String russiaKeywords [] = {"ρωσία", "ρωσικ", "ρώσ", "μόσχα", "russian",
90     "moskva", "πούτιν"};
91     String ukraineKeywords [] = {"ουκραν", "ukrainskaya", "κίεβ", "ζελένσκ",
92     "zelensk", "μαριούπολη", "χάρκοβο", "αζόφ"};
93     String greeceKeywords [] = {"ελλάδ", "ελλαδ", "έλληνη", "ελλην", "ελλην",
94     "αθήνη", "αθηναίων", "αθηνών", "greece", "πειραια", "πειραιά", "πειραιώς",
95     "hellenic"};
96     String turkeyKeywords [] = {"turkish", "ερντογάν", "προύσα", "ταγίπ",
97     "τουρκ", "τούρκ"};
98     String politicsKeywords[] = {"biden", "clinton", "macron", "mitsotakis",
99     "obama", "trump", "αντιπολιτεύ", "αντιπολίτευ", "βουλευτι", "βουλή",
100     "εκλογ", "ερντογάν", "κλίντον", "κογκρέσο", "κυβερνήσ", "κυβέρνησ",
101     "κυβερνητι", "μακρόν", "μητσοτάκη", "μπαιντεν", "μπάιντεν", "ομπάμα",
102     "πολιτικ", "πούτιν", "συριζα", "σύριζα", "ταγίπ", "τραμπ", "τισίπρα",

```

Εικόνα 9.35 Κλάση ClaimDateRetrieval (3)

```

103     "υπουργ");
104     String usaKeywords [] = {"american", "biden", "clinton", "columbia",
105     "florida", "georgia", "harvard", "obama", "stanford", "trump", "washington",
106     "αμερικ", "αριζόνα", "ατλάντα", "ηπα", "καλιφόρνια", "καπιτώλιο", "κλίντον",
107     "κογκρέσο", "μίσιγκον", "μπαιντεν", "μπάιντεν", "ομπάμα", "ουάσιγκτον",
108     "ουάσιγκτον", "ουίσκόνσιν", "οχάιο", "τεξας", "τέξας", "τραμπ", "φλόριντα",
109     "χόλιγουντ"};
110
111     //Adds new values to the dictionary "themeHashSet"
112     themeHashMap = MinorClass.addThemeDictValues(
113     "Covid", covid, covidKeywords, themeHashMap);
114     themeHashMap = MinorClass.addThemeDictValues(
115     "Vaccination", vaccination, vacKeywords, themeHashMap);
116     themeHashMap = MinorClass.addThemeDictValues(
117     "Russia", russia, russiaKeywords, themeHashMap);
118     themeHashMap = MinorClass.addThemeDictValues(
119     "Ukraine", ukraine, ukraineKeywords, themeHashMap);
120     themeHashMap = MinorClass.addThemeDictValues(
121     "Greece", greece, greeceKeywords, themeHashMap);
122     themeHashMap = MinorClass.addThemeDictValues(
123     "Turkey", turkey, turkeyKeywords, themeHashMap);
124     themeHashMap = MinorClass.addThemeDictValues(
125     "Politics", politics, politicsKeywords, themeHashMap);
126     themeHashMap = MinorClass.addThemeDictValues(
127     "USA", usa, usaKeywords, themeHashMap);
128
129     try {
130         String inputValue = SubjectSelectionMenu.showMenu();
131         HashSet<String> theme = themeHashMap.get(inputValue);
132         File readFile = new File(newFileName);
133         File writeFile = new File("monthFrequency_" + inputValue
134         + "_Claims.csv");
135
136         try (BufferedReader in = new BufferedReader(
137         new InputStreamReader(
138         new FileInputStream(readFile), "UTF8")
139         )) {
140             String line;
141             String lineClaim = "";
142             String lineDate = "";
143             int counter = 0;
144
145             try (Writer textWriter =
146             new BufferedWriter(
147             new OutputStreamWriter(
148             new FileOutputStream(writeFile, false),
149             StandardCharsets.UTF_8)
150             )) {
151                 System.setOut(new PrintStream(System.out, true, "UTF8"));
152

```

Εικόνα 9.36 Κλάση ClaimDateRetrieval (4)

```

153         while ((line = in.readLine()) != null) {
154             extractClaimDate(line, counter, lineClaim, lineDate, theme)
155         }
156
157         MinorClass.addNewRecord(textWriter, dateHashMap);
158
159     }
160     } catch (Exception e1) {
161     }
162
163     } catch (Exception e2) {
164         System.out.println("You haven't selected a subject!");
165     }
166 }
167
168 //Retrieves dates from the old file
169 public static void extractClaimDate(
170     String row, int i, String rowClaim, String rowDate,
171     HashSet <String> hs) {
172     boolean continueCheck = true;
173
174     String[] splitLine = row.split("@@", 2);
175     i +=1;
176     rowClaim = splitLine[0].toLowerCase();
177     rowDate = splitLine[1];
178
179     if (rowDate.contains("\n")) {
180         rowDate = rowDate.replaceAll("\n", "");
181     }
182
183     System.out.println("\n");
184     System.out.println("Printing new line");
185
186     for (String word : hs) {
187
188         if (continueCheck) {
189             System.out.println("new line: " + rowClaim);
190
191             if (rowClaim.contains(word)) {
192                 int valueOld = dateHashMap.get(rowDate);
193                 int valueNew = valueOld + 1;
194                 System.out.println(rowClaim);
195                 System.out.println("contains: " + word);
196                 System.out.println(
197                     "Old value of: " + rowDate + " = " + valueOld);
198                 System.out.println(
199                     "New value of: " + rowDate + " = " + valueNew);
200                 System.out.println("\n");
201                 dateHashMap.replace(rowDate, valueNew);
202                 continueCheck = false;
203             }
204         }
205     }
206 }
207

```

Εικόνα 9.37 Κλάση ClaimCategoryRetrieval (1)

```

1 package gr.hmu.mst.fakenews;
2
3 import java.io.BufferedReader;
4 import java.io.BufferedWriter;
5 import java.io.File;
6 import java.io.FileInputStream;
7 import java.io.FileNotFoundException;
8 import java.io.FileOutputStream;
9 import java.io.InputStreamReader;
10 import java.io.OutputStreamWriter;
11 import java.io.PrintStream;
12 import java.io.UnsupportedEncodingException;
13 import java.io.Writer;
14 import java.nio.charset.StandardCharsets;
15 import java.util.Arrays;
16 import java.util.HashMap;
17 import java.util.HashSet;
18
19 /**
20  *
21  * @author zabal
22  */
23 public class ClaimCategoryRetrieval {
24
25     /**
26      * @param args the command line arguments
27      */
28     private static HashMap<String, Integer> catHashMap = new HashMap<>();
29     private static HashSet<String> vaccination = new HashSet<>();
30     private static HashSet<String> covid = new HashSet<>();
31     private static HashSet<String> ukraine = new HashSet<>();
32     private static HashSet<String> russia = new HashSet<>();
33     private static HashSet<String> greece = new HashSet<>();
34     private static HashMap<String, HashSet> themeHashMap = new HashMap<>();
35     private static HashSet<String> keys = new HashSet<>();
36     private static HashSet<String> turkey = new HashSet<>();
37     private static HashSet<String> politics = new HashSet<>();
38     private static HashSet<String> usa = new HashSet<>();
39
40     /**
41      * @param args the command line arguments
42      * @throws java.io.FileNotFoundException
43      * @throws java.io.UnsupportedEncodingException
44      */
45     public static void main(String[] args) throws
46         FileNotFoundException, UnsupportedEncodingException {
47         System.setOut(new PrintStream(System.out, true, "UTF8"));
48         String newFileName = MinorClass.chooseFile();
49
50         //Creates an array of categories
51         String[] keysStr = {"HOAXES", "ΠΑΡΑΠΛΗΡΟΦΟΡΕΗ", "FAKE NEWS", "ΔΙΑΦΟΡΑ",
52             "ΣΥΝΚΟΜΟΙΟΛΟΓΙΑ", "ΨΕΥΔΕΣ", "ΜΥΘΟΙ ΚΑΙ ΙΣΤΟΡΙΑ", "ΥΓΕΙΑ",
53             "ΨΕΥΔΟΕΠΙΣΤΗΜΗ", "ΨΕΥΔΗΣ ΙΣΧΥΡΙΣΜΟΣ", "ΛΕΙΠΕΙ ΘΕΜΑΤΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ",
54             "ΑΠΑΘ", "ΠΡΑΓΜΑΤΙΚΕΣ ΙΣΤΟΡΙΕΣ", "LIKE FARMING",

```

Εικόνα 9.38 Κλάση ClaimCategoryRetrieval (2)

```

55     "ΜΙΣΗ ΓΕΓΟΝΟΤΩΝ ΚΑΙ ΠΑΡΑΠΟΙΗΣΕΩΝ", "ΨΕΥΤΙΚΑ ΡΗΤΑ ΚΑΙ ΑΠΟΦΘΙΓΜΑΤΑ",
56     "ΕΠΙΚΑΙΡΟΤΗΤΑ", "ΤΡΟΠΟΠΟΙΗΜΕΝΗ ΕΙΚΟΝΑ", "ΑΝΑΚΥΚΛΩΜΕΝΗ ΕΙΔΗΣΗ",
57     "ΚΙΝΔΥΝΟΛΟΓΙΑ", "CLICKBAIT", "ΣΑΤΙΡΑ", "ΤΡΟΠΟΠΟΙΗΜΕΝΟ ΒΙΝΤΕΟ",
58     "ΕΡΕΥΝΕΣ", "ΑΝΑΚΟΙΝΩΣΕΙΣ", "ΑΝΕΠΙΒΕΒΑΙΩΤΟ", "ΟΔΗΓΟΙ ΚΑΙ ΣΥΜΒΟΥΛΕΣ",
59     "ΕΙΚΟΝΑ ΑΠΟ ΠΡΟΓΡΑΜΜΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ");
60     keys.addAll(Arrays.asList(keysStr));
61
62     /* Adds keys and values to the dictionary consisting of categories and
63     their frequency
64     */
65     catHashMap = MinorClass.addDictValues(keys, catHashMap, 0);
66
67     //Creates strings of subjects of our interest
68     String covidKeywords [] = {"cov19", "covid", "κοροναϊ", "κορονοϊό",
69     "κορονοϊ", "κορωναιό", "κορωνοϊό", "κορωνοϊ", "κορωνοι", "sars-cov",
70     "2019-ncov", "mers-cov"};
71     String vacKeywords [] = {"εμβόλ", "εμβολ"};
72     String russiaKeywords [] = {"ρωσία", "ρωσικ", "ρώσ", "μόσχα", "russian",
73     "moskva", "πούτιν"};
74     String ukraineKeywords [] = {"ουκραν", "ukrainskaya", "κίεβ", "ζελένσκ",
75     "zelensk", "μαριούπολη", "χάρκοβο", "αζόφ"};
76     String greeceKeywords [] = {"ελλάδ", "ελλαδ", "έλλην", "ελλην", "ελλην",
77     "αθήν", "αθηναίων", "αθηνών", "greece", "πειραια", "πειραιά", "πειραιώς",
78     "hellenic"};
79     String turkeyKeywords [] = {"turkish", "ερντογάν", "προύσα", "ταγίπ",
80     "τουρκ", "τούρκ"};
81     String politicsKeywords[] = {"biden", "clinton", "macron", "mitsotakis",
82     "obama", "trump", "αντιπολιτεύ", "αντιπολίτευ", "βουλευτι", "βουλή",
83     "εκλογ", "ερντογάν", "κλίντον", "κογκρέσο", "κυβερνήσ", "κυβέρνησ",
84     "κυβερνητι", "μακρόν", "μητσοτάκη", "μπαινιεν", "μπάινιεν", "ομπάμα",
85     "πολιτικ", "πούτιν", "συριζα", "σύριζα", "ταγίπ", "τραμπ", "τσιπρά",
86     "υπουργ"};
87     String usaKeywords [] = {"american", "biden", "clinton", "columbia",
88     "florida", "georgia", "harvard", "obama", "stanford", "trump", "washington",
89     "αμερικ", "αριζόνα", "ατλάντα", "ηπα", "καλιφόρνια", "καπιτώλιο", "κλίντον",
90     "κογκρέσο", "μίσιγκαν", "μπαινιεν", "μπάινιεν", "ομπάμα", "ουάσιγκτον",
91     "ουάσινγκτον", "ουϊσκόνσιν", "οχάιο", "τεξας", "τέξας", "τραμπ", "φλόριντα",
92     "χόλιγουντ"};
93
94     //Adds new values to the dictionary "themeHashSet"
95     themeHashMap = MinorClass.addThemeDictValues(
96     "Covid", covid, covidKeywords, themeHashMap);
97     themeHashMap = MinorClass.addThemeDictValues(
98     "Vaccination", vaccination, vacKeywords, themeHashMap);
99     themeHashMap = MinorClass.addThemeDictValues(
100     "Russia", russia, russiaKeywords, themeHashMap);
101     themeHashMap = MinorClass.addThemeDictValues(
102     "Ukraine", ukraine, ukraineKeywords, themeHashMap);
103     themeHashMap = MinorClass.addThemeDictValues(
104     "Greece", greece, greeceKeywords, themeHashMap);
105     themeHashMap = MinorClass.addThemeDictValues(
106     "Turkey", turkey, turkeyKeywords, themeHashMap);
107     themeHashMap = MinorClass.addThemeDictValues(
108     "Politics", politics, politicsKeywords, themeHashMap);

```

Εικόνα 9.39 Κλάση ClaimCategoryRetrieval (3)

```

109 themeHashMap = MinorClass.addThemeDictValues(
110     "USA", usa, usaKeywords, themeHashMap);
111
112 try {
113     String inputValue = SubjectSelectionMenu.showMenu();
114     HashSet<String> theme = themeHashMap.get(inputValue);
115     File readFile = new File(newFileName);
116     File writeFile = new File("categoryFrequency_" + inputValue + "_Claims.cs
117
118     try ( BufferedReader in = new BufferedReader(
119         new InputStreamReader(
120             new FileInputStream(readFile), "UTF8")
121     )) {
122         String line;
123         String lineClaim = "";
124         String lineCats = "";
125         int counter = 0;
126
127         try (Writer textWriter = new BufferedWriter(
128             new OutputStreamWriter(
129                 new FileOutputStream(writeFile, false),
130                 StandardCharsets.UTF_8)
131         )) {
132             System.setOut(new PrintStream(System.out, true, "UTF8"));
133
134             while ((line = in.readLine()) != null) {
135                 extractClaimCategories(
136                     line, counter, lineClaim, lineCats, theme);
137             }
138
139             MinorClass.addNewRecord(textWriter, catHashMap);
140
141         }
142     } catch (Exception e1) {
143     }
144
145 } catch (Exception e2) {
146     System.out.println("You haven't selected a subject!");
147 }
148 }
149
150 //Retrieves categories from the old file
151 public static void extractClaimCategories(
152     String row, int i, String rowClaim, String rowCat, HashSet<String> hs) {
153     String[] splitLine = row.split("@@@", 2);
154     i += 1;
155     rowClaim = splitLine[0].toLowerCase();
156     rowCat = splitLine[1];
157
158     for (String word : hs) {
159
160         if (rowClaim.contains(word)) {
161             System.out.println("\n");
162             System.out.println(rowClaim);

```


Εικόνα 9.40 Κλάση ClaimCategoryRetrieval (4)

```
163 System.out.println("Line contains a word: " + word);
164 String[] splitLineCats = rowCat.split(";");
165 System.out.println(
166     "Analyzing its categories: "
167     + Arrays.toString(splitLineCats));
168
169 for (String cat : splitLineCats) {
170
171     if (catHashMap.keySet().contains(cat)) {
172         int valueOld = catHashMap.get(cat);
173         int valueNew = valueOld + 1;
174         catHashMap.replace(cat, valueNew);
175         System.out.println("keyset cointains: " + cat);
176         System.out.println(
177             "Old value of: " + cat + ": " + valueOld);
178         System.out.println(
179             "New value of: " + cat + ": " + valueNew);
180     } else {
181         System.out.println(cat + " is not in the dictionary");
182     }
183 }
184
185 break;
186
187 }
188 }
189 }
190 }
```