
**Πρόγραμμα Μεταπτυχιακών Σπουδών:**

**«Διαχείριση Πληροφοριών σε Βιβλιοθήκες, Αρχεία, Μουσεία»**

**ΤΜΗΜΑ ΑΡΧΕΙΟΝΟΜΙΑΣ, ΒΙΒΛΙΟΘΗΚΟΝΟΜΙΑΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΗΣΗΣ**

**ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**DEPARTMENT OF ARCHIVAL, LIBRARY AND INFORMATION STUDIES**

**SCHOOL OF MANAGEMENT, ECONOMICS AND SOCIAL SCIENCES**

**Διπλωματική Εργασία**

# «Μελέτη και σύγκριση αλγορίθμων, μοντέλων και συστημάτων εντοπισμού ψευδών ειδήσεων»

# «Study and comparison of fake news detection algorithms, models and systems»

**Αικατερίνη-Σαπφώ Κωλέτση (ΑΜ: 206682008)**

Επιβλέπων: Δημήτριος Κουής

Dimitrios Kouis

Digitally signed by Dimitrios Kouis
Date: 2022.12.10 17:28:40 +02'00'

Αθήνα, Δεκέμβριος 2022

# Σκοπός

- Μελέτη κύριων χαρακτηριστικών των αλγορίθμων, μοντέλων & συστημάτων για τον εντοπισμό των ψευδών ειδήσεων
- Ομοιότητες & διαφοροποιήσεις τους, ως προς τα χαρακτηριστικά και τους τρόπους λειτουργίας
- Οδηγός εξοικείωσης των επιστημόνων της πληροφόρησης με το πεδίο ανάπτυξης αλγορίθμων & συστημάτων εντοπισμού ψευδών ειδήσεων

# Εισαγωγή στις Ψευδείς ειδήσεις - Fake news

**Εικόνα 1. Σάτιρα**


**Εικόνα 2. Προπαγάνδα**

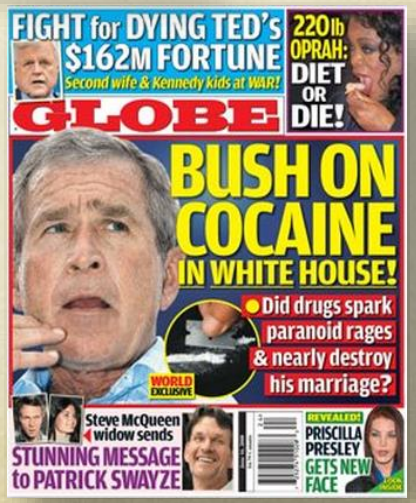

**Εικόνα 3. Παραπληροφόρηση**


**Εικόνα 4. Κακή πληροφόρηση**

# Εισαγωγή στις Ψευδείς ειδήσεις - Fake news


**Εικόνα 5. Ισχυρισμός (Hoax)**


**Εικόνα 6. Φήμη**




**Εικόνα 7. «Κίτρινος Τύπος/Συναισθηματισμός»**

# Εισαγωγή στις Ψευδείς ειδήσεις - Fake news


**Εικόνα 8. Θεωρίες συνωμοσίας**


**Εικόνα 9. Clickbait**


**Εικόνα 10. μεροληπτική προπαγάνδα**

# Ψευδείς ειδήσεις & επιστήμη της πληροφόρησης

**IFLA 2017- Πρωτοβουλία για αναβάθμιση δεξιοτήτων επιστημόνων πληροφόρησης & αντιμετώπιση του φαινομένου**

# Μεθοδολογία

## 1ο στάδιο

**Εντοπισμός πηγών από Scopus & Web of Science (έτη 2019-2021) με βάση 4 ερωτήματα**

**Εξαγωγή δεδομένων σε csv & συστηματική καταχώρησή τους σε Excel**

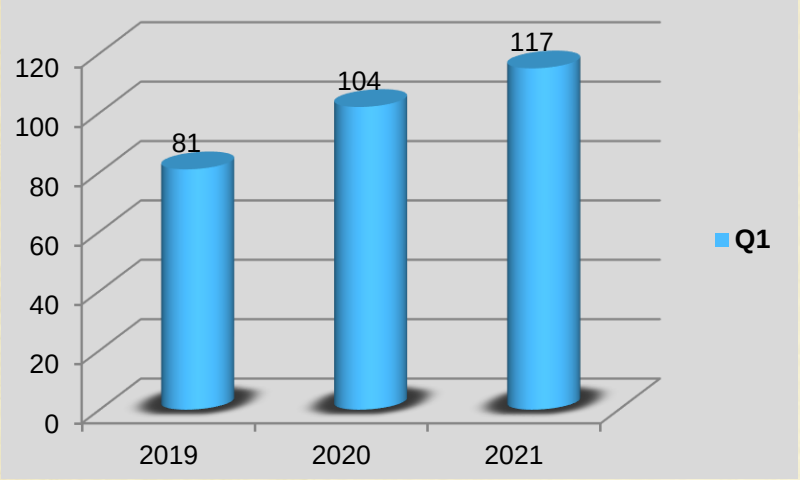
## 2ο στάδιο

**Ενοποίηση αποτελεσμάτων από Scopus & WoS και στα 4 ερωτήματα**

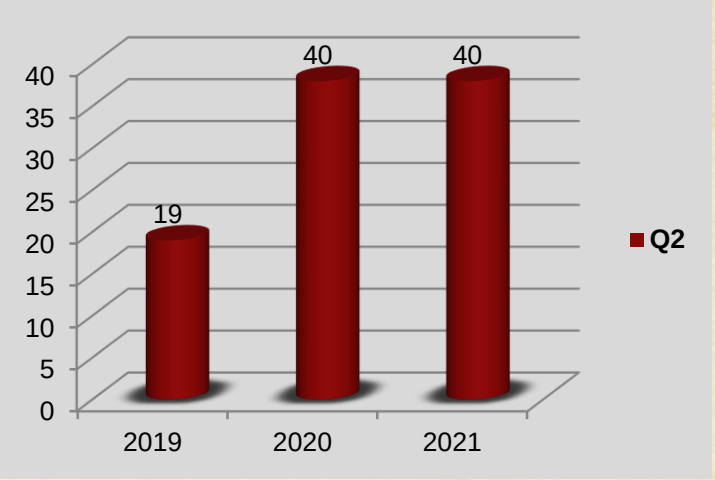
## 3ο στάδιο

**Εντοπισμός, διαγραφή διπλοεγγραφών & συγχώνευση εγγραφών σε νέο Excel**

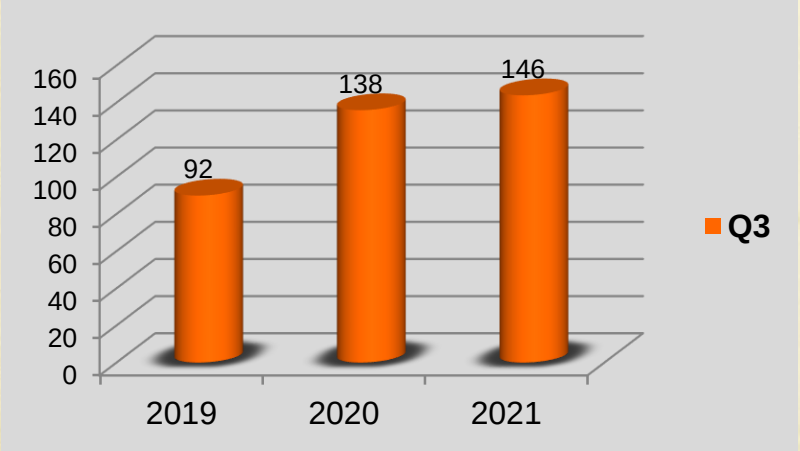
# Μεθοδολογία



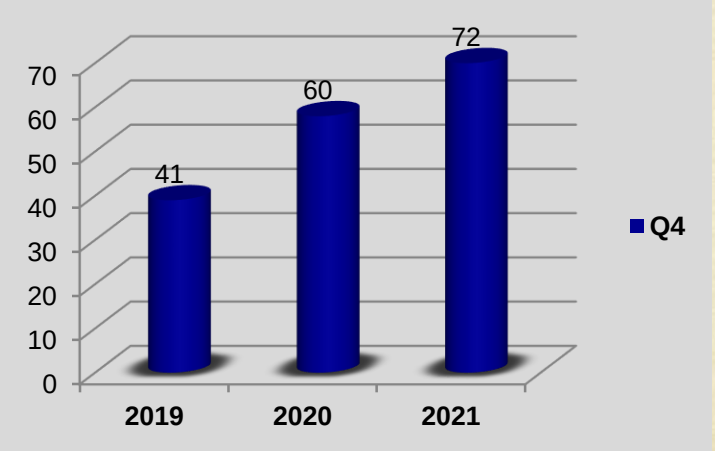

Query 1: fake news + algorithms = 302 πηγές




Query 2: disinformation + algorithms= 99 πηγές




Query 3: fake news + machine learning = 376 πηγές




Query 4: misinformation + algorithms = 173 πηγές

# Μεθοδολογία

## 4ο στάδιο

Από τις συνολικά **970** πηγές μελετήθηκαν διεξοδικά οι **60** πιο συναφείς, πιο πλήρεις & πιο συχνά αναφερόμενες πηγές

## 5ο στάδιο

Κατά τη μελέτη των **60** πηγών έγινε χρήση της μεθόδου **«snowball»:** *«επιτρέπει την συλλογή δημοσιεύσεων σε ένα συγκεκριμένο θέμα έρευνας και την σύνδεσή τους με το δίκτυο των αναφορών» (Dobrovolskyi & Keberle, 2019)*

## 6ο στάδιο

Διεξοδική ανάλυση του περιεχομένου από 60 πηγές της βιβλιογραφικής επισκόπησης σε Scopus & Web of Science & από 50 και πάνω πηγές με τη μέθοδο «snowball» => πάνω από 110 πηγές στις οποίες εντοπίστηκαν επιμέρους κοινές θεματικές κατηγορίες

# Αποτελέσματα

## 9 κατηγορίες χαρακτηριστικών των αλγορίθμων, μοντέλων & συστημάτων εντοπισμού ψευδών ειδήσεων

1. Έλεγχος γεγονότων που βασίζεται σε χαρακτηριστικά «γνώσης»

2. Γλωσσολογικά χαρακτηριστικά

3. Στυλομετρικά χαρακτηριστικά: *αξιολόγηση γραφής συγγραφέα-έμφαση σε αυτούς που αποπροσανατολίζουν το κοινό και υιοθέτηση εξελιγμένων τεχνικών NLP*

4. Χαρακτηριστικά προσεγγίσεων μηχανικής μάθησης

5. Χαρακτηριστικά θεματικών προσεγγίσεων

6. Συναισθηματικά χαρακτηριστικά

7. Υβριδικά χαρακτηριστικά & μοντέλα

8. Οπτικά χαρακτηριστικά

9. Χαρακτηριστικά κοινωνικού περιεχομένου

## Αποτελέσματα

### 1. Έλεγχος γεγονότων που βασίζονται σε χαρακτηριστικά «γνώσης»

**1.1 Έλεγχος από επαγγελματίες ειδικούς – Manual fact checking**

**1.2.Έλεγχος βάση πληθοπορισμού - Crowdsourcing**

**1.3. Αυτοματοποιημένος έλεγχος - Automatic fact-checking**

a. Fact-checking

b. Fact extraction (εξαγωγή γνώσης από τον ιστό)

c. Ιστοσελίδες αυτοματοποιημένου ελέγχου

## Αποτελέσματα

**3. Στυλομετρικά χαρακτηριστικά: *αξιολόγηση γραφής συγγραφέα-έμφαση σε αυτούς που αποπροσανατολίζουν το κοινό και υιοθέτηση εξελιγμένων τεχνικών NLP***

**3.1 Συντακτικά χαρακτηριστικά**

**3.2 Σημασιολογικά χαρακτηριστικά**

**3.3 Χαρακτηριστικά λεξικών (LIWC, Bias lexicon, MPQA Subjectivity lexicon): καταμέτρηση χαρ/κων υποκειμενικότητας (δηλώσεων υπεκφυγής, κατηγορημάτων) & μεροληψίας**

**3.4 Χαρακτηριστικά ιστοτόπου ειδήσεων**

**3.5 Χαρακτηριστικά αναπαραστάσεων (μη παρατηρήσιμα)**

**Μοντέλα : Word2vec, Doc2vec, Link2vec, Glove, Fast Text**

**RST(Θεωρία Ρητορικής Δομής)** https://www.sfu.ca/rst/01intro/intro.html

# 4. Χαρακτηριστικά προσεγγίσεων μηχανικής μάθησης

4.1 Αλγόριθμοι επιβλεπόμενης μάθησης

4.2 Αλγόριθμοι πολλαπλής μάθησης

4.3 Αλγόριθμοι μη επιβλεπόμενης μάθησης

4.4 Αλγόριθμοι Μεταευριστικοί

4.5 Αλγόριθμοι βαθιάς μάθησης

# 4.4 Αλγόριθμοι Μεταευριστικοί

Grey Wolf

Whale Optimization

## Συμπεράσματα

Το φαινόμενο των ψευδών ειδήσεων δεν είναι τωρινό αλλά έχει ιστορικές ρίζες από τις απαρχές της ανθρωπότητας

Η διπλωματική εργασία εστιάστηκε σε βασικά χαρακτηριστικά των αλγορίθμων που εντοπίστηκαν έπειτα από τη βιβλιογραφική επισκόπηση και επιλογή των σχετικών πηγών.

Δεν μελετήθηκαν διεξοδικά τα email spam και τα bots καθώς αποτελούν υποπεριπτώσεις των ψευδών ειδήσεων και δεν αποτελούν πρωτογενείς επιρροές ως προς τη διάδοσή τους

Η πλειονότητα των πηγών ταυτίζει την παραπληροφόρηση και τη διασπορά των ψευδών ειδήσεων με την **προπαγάνδα** (πολιτική, κρατική ή θρησκευτική) αλλά και με τα **κοινωνικά μέσα δικτύωσης** (Facebook, Twitter)

**Χαρακτηριστικά αλγορίθμων**

Στα πρώτα στάδια της έρευνας ήταν δυσδιάκριτες οι διαφορές και η κατηγοριοποίησή τους π.χ. λεξιλόγιο, σύνταξη, σημασιολογία, θεματικά, υβριδικά, λόγω της μεγάλης ποικιλομορφίας τους.

Εξαίρεση αποτελούν τα συναισθηματικά, τα ψυχολογικά καθώς και τα χαρακτηριστικά κοινωνικού περιεχομένου που ήταν ευδιάκριτα από την αρχή

## Συμπεράσματα

# ΚΥΡΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΕΡΙΕΧΟΜΕΝΟΥ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ

**Στίξη !!! ;;;**

**Μικρό ποσοστό μοναδικών λέξεων**

**Λέξεις που εγείρουν κυρίως αρνητικά συναισθήματα**


**Απλοϊκή γραφή (υπολογίζεται με εργαλεία όπως Flech reading Ease)**

**Λέξεις και τρόποι διατύπωσης που ασκούν μεγάλη επιρροή**

**Μεγάλη συχνότητα επανάληψης λέξεων-κλειδιών, π.χ. Covid-19 (υπολογίζεται με εργαλεία όπως Rake Library)**

**Επίθετα, αντωνυμίες, λέξεις μεροληψίας, χρήση υπερθετικών βαθμών & τροπικών επιρρημάτων (υπολογίζονται με εργαλεία όπως LIWC, NELA)**

**"Democrats are sadly the true enemy of our America!"**

**Μεγάλη συχνότητα επανάληψης μίας λέξης-unigrams, δύο λέξεων-bigrams (υπολογίζεται με εργαλεία όπως το Tf-Idf και αναπαρίσταται με μεθόδους όπως το Bag of words)**

**Αργκό ή νεολογισμούς**

**Χρήση ρημάτων σε τρίτο ενικό και τρίτο πληθυντικό και κυρίως βοηθητικών**

## Συμπεράσματα

# ΚΥΡΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΨΕΥΔΟΥΣ ΚΟΙΝΩΝΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ ή ΑΝΑΞΙΟΠΙΣΤΙΑΣ ΧΡΗΣΤΗ

- Όταν είμαι χρήστης που με ακολουθεί μεγάλος αριθμός άλλων χρηστών (my followers) ή έχω μεγάλο δίκτυο φίλων

- Όταν είμαι ένας από τους ακόλουθους (followees) ενός χρήστη με μεγάλο αριθμό ακολούθων ή με μεγάλο δίκτυο φίλων

+ Πιστοποιημένο προφίλ χρήστη

- Μεγάλος αριθμός (ανα)δημοσιεύσεων τόσο ειδήσεων όσο και σχολίων ενός χρήστη

- Μεγάλος αριθμός hashtag

- Μεγάλος αριθμός προτιμήσεων χρήστη (likes)

+ Ύπαρξη URL

Ευχαριστώ πολύ!