# **Εργασία**: Ανίχνευση αντιστοιχίας («στάσης») του κειμένου σε άρθρα /ιστοσελίδες σε σχέση με τον τίτλο τους.

## Εισαγωγή

Στην παρούσα εργασία γίνεται μια προσπάθεια δημιουργίας εργαλείου με σκοπό την ανίχνευση της «στάσης» που τηρεί ένα άρθρο σε σχέση με τον τίτλο του. Βασίζεται σε διαγωνισμό που έτρεξε στα μέσα του 2017 και είχε ως αποτέλεσμα τη δημιουργία αντίστοιχων αλγορίθμων μηχανικής μάθησης. Στο πλαίσιο της εργασίας, χρησιμοποιούμε τον αλγόριθμο της τρίτης κατά σειρά νικήτριας ομάδας ως πιο εύκολο να ενσωματωθεί στο περιβάλλον προγραμματισμού που χρησιμοποιούμε. Η ακρίβεια στην πρόβλεψη για τις τρεις πρώτες ομάδες κυμαίνεται από 81.72% έως 82.02%, κατά συνέπεια και η λύση που χρησιμοποιούμε μπορεί να θεωρηθεί state of art.

## Επισκόπηση

Η αρχική σκέψη για την εργασία ήταν η δημιουργία εργαλείου για την ανίχνευση ψευδών ειδήσεων. Η σχετική έρευνα, έδειξε ότι δεν υπάρχουν κατάλληλοι αλγόριθμοι διαθέσιμοι για αυτό το σκοπό. Για παράδειγμα, η πρόταση:

*“Στα χρόνια της κρίσης αυξήθηκε η ανεργία”*

Είναι δυνατό να ελεγχθεί αυτόματα σε σχέση με την αλήθεια με κατάλληλο ερευνητικό έργο, αλλά η πρόταση:

*«Οι Έλληνες στρατιωτικοί όντως πέρασαν τα σύνορα και για αυτό το λόγο συνελήφθησαν»*

Δεν μπορεί να ελεγχθεί με τα μέσα που διαθέτουμε, και πρέπει να περιμένουμε η έρευνα να φτάσει την τεχνητή νοημοσύνη στο επίπεδο της ανθρώπινης.

Επί πλέον, η τιτλοφόρηση (εφ’ εξής labeling /annotation) άρθρων με σκοπό τη χρήση τους για την εκμάθηση αλγορίθμων τεχνητής νοημοσύνης, ενώ σε πρώτη ματιά φαίνεται δελεαστική προσέγγιση, στην πράξη αποδεικνύεται πολύ δύσκολη, καθώς δεν υπάρχουν σύνολα δεδομένων ανοιχτά προς χρήση (αν υπάρχουν καθόλου – οπότε πρέπει να δημιουργηθούν), ότι υπάρχει είναι προστατευμένο υλικό και σίγουρα χρειάζεται ανθρώπινη παρέμβαση με χρήση κατάλληλων εργαλείων.

## Κίνητρο

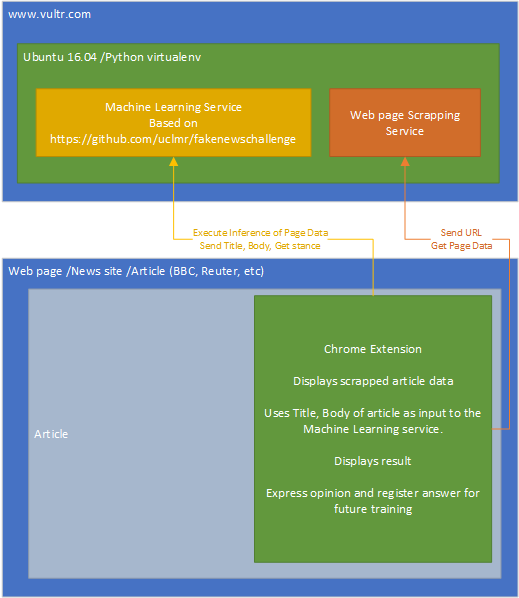
Ο έλεγχος της «στάσης» ενός άρθρου σε σχέση με τον τίτλο του είναι πιθανό να αποτελεί το πρώτο βήμα για τον έλεγχο των ψευδών ειδήσεων.

* Μπορεί να οδηγήσει στην υποστήριξη ανθρώπων /ελεγκτών ψευδών ειδήσεων δίνοντας τη δυνατότητα για γρήγορο έλεγχο της πληροφορίας που καλούνται να αξιολογήσουν.
* Είναι δυνατό επίσης να κατασκευαστεί ένα εργαλείο για τον έλεγχο ψευδών ειδήσεων βασισμένο στο εργαλείο ανίχνευσης «στάσης». Θα μπορούσε να χρησιμοποιήσει το αποτέλεσμα ανίχνευσης στάσης για το ίδιο θέμα ανάμεσα σε πολλούς οργανισμούς, να δώσει βαρύτητα ανάλογα με την αξιολόγηση του εκάστοτε οργανισμού και εν τέλει να συμπεράνει την αλήθεια /ψέμα για ένα συγκεκριμένο θέμα.

Θεωρώντας ότι η παραπάνω ακολουθία είναι αξιόλογη και μπορεί να οδηγήσει σε αποτέλεσμα η εργασία βασίστηκε στην ανίχνευση στάσης.

## Περιγραφή αρχιτεκτονικής

Στο παρακάτω διάγραμμα φαίνεται η αρχιτεκτονική του συστήματος ανίχνευσης στάσης. Χρησιμοποιούμε δύο υπηρεσίες, μία για την αυτόματη ανίχνευση των δεδομένων του άρθρου (ιστοσελίδας), για παράδειγμα του τίτλου και του κειμένου (γραμμένη σε javascript /node.js) και μία για την ανίχνευση της στάσης του άρθρου με είσοδο τον τίτλο και το κείμενο (γραμμένη σε python). Η δεύτερη υπηρεσία βασίζεται στην εργασία της τρίτης νικήτριας ομάδας στο διαγωνισμό που περιγράψαμε, παραλλαγμένη ώστε να δέχεται είσοδο από τα εργαλεία που χρησιμοποιούμε.

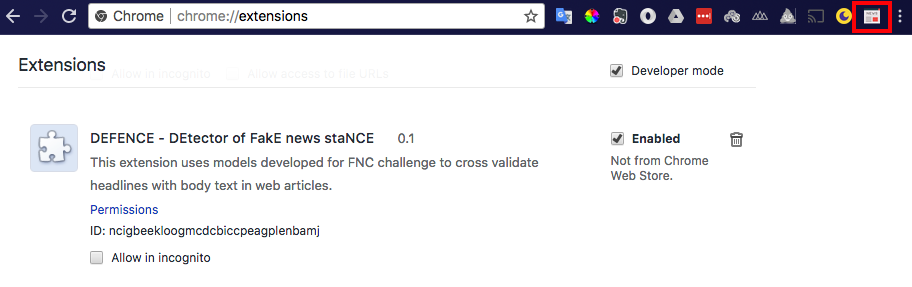


Σχήμα 1

Οι λειτουργίες των υπηρεσιών ενοποιούνται /ενσωματώνονται σε κατάλληλη γραφική διασύνδεση χρήστη με τη μορφή επέκτασης για τον browser Google Chrome.

## Εγκατάσταση

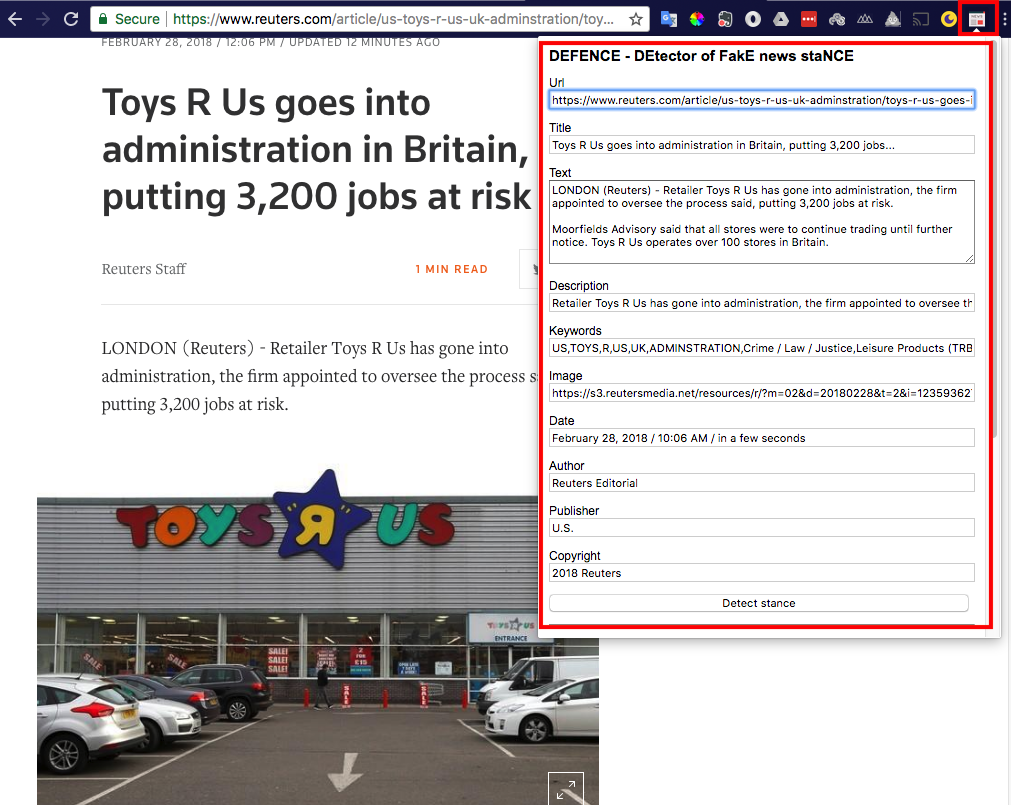
Το μόνο που απαιτείται είναι η εγκατάσταση της επέκτασης για τον browser Google Chrome. Η επέκταση μπορεί να αποθηκευτεί στον τοπικό υπολογιστή του χρήστη από τη θέση <https://github.com/antalakas/fakenewschallenge/blob/master/javascript/defence.crx>. Στη συνέχεια ανοίγουμε τη σελίδα των επεκτάσεων του Chrome και με drag n drop ρίχνουμε το αρχείο επέκτασης defence.crx.



Εικόνα 1

## Περιγραφή λειτουργίας

Φορτώνουμε κάποια ιστοσελίδα.



Εικόνα

Πατάμε στο εικονίδιο της επέκτασης και αυτόματα η σελίδα αναλύεται από την υπηρεσία scrapping, ενώ τα αποτελέσματα εμφανίζονται στα κατάλληλα πεδία. Πατάμε το κουμπί “Detect stance” και ενεργοποιείται η υπηρεσία μηχανικής μάθησης που εκτελεί τον αλγόριθμο και επιστρέφει το αποτέλεσμα.

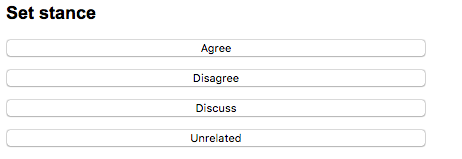


Εικόνα

Τα δυνατά αποτελέσματα είναι **agree**, **disagree**, **discuss** και **unrelated**.

## Συμπέρασμα /Πιθανές επεκτάσεις.

Μία πιθανή επέκταση είναι η συλλογή πληροφοριών για σελίδες που ο χρήστης έχει άποψη και θέλει να εκτελέσει annotation. Για το σκοπό αυτό η υπηρεσία μηχανικής μάθησης έχει επεκταθεί ώστε να δέχεται και να αποθηκεύει την άποψη του χειριστή. Μελλοντικά τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για μετεκπαίδευση αλγορίθμου.



Εικόνα

Επίσης μπορούμε να διορθώσουμε περιπτώσεις που ο scrapper δε λειτουργεί (παράδειγμα στους NYT).

Στο πλαίσιο της εργασίας η συνολική λύση φαίνεται να λειτουργεί αξιόπιστα.

/Ανδρέας Νταλάκας

### Πηγαίος Κώδικας

<https://github.com/antalakas/fakenewschallenge>

### Αναφορές

* Fake News Challenge. (2017, June). Fake News Challenge. Retrieved From <http://www.fakenewschallenge.org/>
* Fake News Challenge. (2017, June). Fake News Challenge Source Code. Retrieved From
* <https://github.com/FakeNewsChallenge>
* Talos Group at CISCO. (2017, June). Model Source Code for Fake News Challenge. Retrieved From <https://github.com/Cisco-Talos/fnc-1/blob/master/deep_learning_model/clf.py>

### Talos Group at CISCO. (2017, June). Talos Targets Disinformation with Fake News Challenge Victory. Retrieved From <https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html>

* [Andreas Hanselowski](mailto:hanselowski@aiphes.tu-darmstadt.de), [Avinesh PVS](mailto:avinesh@aiphes.tu-darmstadt.de), [Benjamin Schiller](mailto:schiller.benny@googlemail.com) and Felix Caspelherr. (June 2017). Athene System Source Code for Fake News Challenge. Retrieved From <https://github.com/hanselowski/athene_system>
* UCL Machine Reading Group. (June 2017). Source Code for Fake News Challenge. Retrieved From <https://github.com/uclmr/fakenewschallenge>
* William Ferreira, Andreas Vlachos. (Sep 26, 2016). Emergent: a novel data-set for stance classification. Retrieved From <https://arxiv.org/abs/1606.05464>