MAI4CAREU

Επεξεργασία Φυσικής Γλώσσας

Μελέτες περιπτώσεων και εφαρμογές

Εντοπισμός διαδικτυακής ρητορικής μίσους

Demetris Paschalides

Department of Computer Science

University of Cyprus

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Τι είναι η ρητορική μίσους;

- Η ρητορική μίσους δεν έχει επίσημο ορισμό, αλλά υπάρχει συναίνεση:

  Η ρητορική μίσους είναι η ακραία αντιπάθεια ενός ατόμου ή μιας ομάδας ανθρώπων λόγω της φυλής, της εθνικότητας, της θρησκείας, του φύλου ή του προσανατολισμού φύλου τους.

- Διαφορετικοί νόμοι που ισχύουν για διαφορετικές χώρες και οργανισμούς.
- Μικρή διαφορά μεταξύ ρητορικής μίσους και ελευθερίας του λόγου.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Νομοθεσίες και πολιτικές ρητορικής μίσους

- Η ρητορική μίσους προστατεύεται από τις διατάξεις περί ελευθερίας του λόγου της Πρώτης Τροπολογίας.
- Νόμοι που απαγορεύουν τη ρητορική μίσους καθώς προωθεί τη βία ή την κοινωνική αναταραχή.
- Το Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων υιοθέτησε έναν ανοικτό ορισμό της ρητορικής μίσους που καλύπτει όλες τις μορφές έκφρασης μίσους.
- Οι άνθρωποι που καταδικάζονται για χρήση ρητορικής μίσους μπορεί συχνά να αντιμετωπίσουν μεγάλα πρόστιμα, ακόμη και φυλάκιση.
- Αυτοί οι νόμοι επεκτείνονται στο διαδίκτυο και τα μέσα κοινωνικής δικτύωσης, οδηγώντας πολλά OSN να δημιουργήσουν τις δικές τους διατάξεις κατά της ρητορικής μίσους.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ρητορική μίσους στα OSN

- Τα διαδικτυακά κοινωνικά δίκτυα (OSN), όπως το Twitter και το Facebook, έχουν φέρει επανάσταση στην επικοινωνία.
- Μοιραστείτε ελεύθερα σκέψεις, απόψεις και εμπειρίες πραγματικής ζωής.
- **Δυστυχώς, ορισμένοι άνθρωποι εκμεταλλεύονται αυτήν την υποδομή**:
  - Αξιοποίηση του ανοικτού χαρακτήρα των OSN, των ιδιοτήτων της ανωνυμίας και της κινητικότητας
  - Διαδώστε τη ρητορική μίσους και οργανώστε δραστηριότητες μίσους.

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ρητορική μίσους στα OSN

- Σημαντική αυξηση του hate-speech προς τις μεταναστευτικές κοινότητες
- Άνοδος του hate-speech και hate-crimes με βάση τις θρησκευτικές πεποιθήσεις, την εθνικότητα ή το φύλο
  - 80% των ερωτηθέντων έχουν αντιμετωπίσει ρητορική μίσους στο διαδίκτυο και το 40% αισθάνθηκε επίθεση ή απειλή.
- Αύξηση της ρητορικής μίσους και των εγκλημάτων μίσους απο την εκλογή του Trump.
- Επείγουσα ανάπτυξη countermeasures.
- Φάσμα διεθνών πρωτοβουλιών που δρομολογήθηκαν για την αντιμετώπιση του προβλήματος (**MANDOLA**).

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus

Online Social Networks and Hate-speech

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ορισμός της ρητορικής μίσους

- Τα ερευνητικά έργα στην ανίχνευση ρητορικής μίσους υιοθετούν τον δικό τους ορισμό ρητορικής μίσους.
- ***Silva et al.*** 2016 - Η ρητορική μίσους είναι κάθε αδίκημα που υποκινείται, εν όλω ή εν μέρει, από την προκατάληψη του δράστη εναντίον μιας πτυχής μιας ομάδας ανθρώπων.
- ***Gitari et al.*** 2015 - Η ρητορική μίσους ορίζεται σε τρία μέρη:
  a. Απευθύνεται σε μια ομάδα ανθρώπων και όχι σε ένα μόνο άτομο (εθνικότητα, θρησκεία, φύλο, κοινωνικό προσανατολισμό, αναπηρία και τάξη).
  b. Μπορεί να περιέχει μερικά από τα χαρακτηριστικά γνωρίσματα της επικίνδυνης ομιλίας (συγκρίνοντας ομάδα ανθρώπων με έντομα).
  c. Ο επικίνδυνος λόγος συχνά ενθαρρύνει το κοινό να εγκρίνει ή να διαπράξει βίαιες πράξεις στην ομάδα-στόχο (διακρίσεις, λεηλασίες, ταραχές, ξυλοδαρμός, βίαιη έξωση και δολοφονία).

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ορισμός της ρητορικής μίσους

- Υιοθετούμε τον ορισμό από ***Davidson et al.*** 2017:

> ***Η γλώσσα που χρησιμοποιείται για να εκφράσει μίσος προς μια στοχευμένη ομάδα ή προορίζεται να είναι υποτιμητική, να ταπεινώσει ή να προσβάλει τα μέλη της ομάδας.***

- Δεν περιλαμβάνει όλες τις περιπτώσεις προσβλητικής γλώσσας.
- Οι άνθρωποι χρησιμοποιούν εξαιρετικά προσβλητικούς όρους σε ορισμένες ομάδες, αλλά με ποιοτικά διαφορετικό τρόπο.
- Για παράδειγμα, στην καθημερινή γλώσσα και στο διαδίκτυο, σε στίχους τραγουδιών π.χ. ραπ τραγούδια, ή από εφήβους ενώ παίζουν βιντεοπαιχνίδια.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of ...ion
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MANDOLA Big-Data Processing System

Υπάρχουν δύο σημαντικές δυσκολίες στην αντιμετώπιση της ρητορικής μίσους στο διαδίκτυο::

1. **Έλλειψη αξιόπιστων δεδομένων**
2. **Κακή επίγνωση**

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Quiz Time

MAI4CAREU

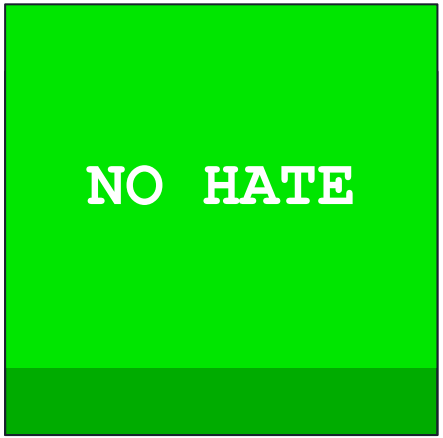
Είναι Hate-speech ή Όχι?

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

#California is full of white trash who moved from #Oklahoma

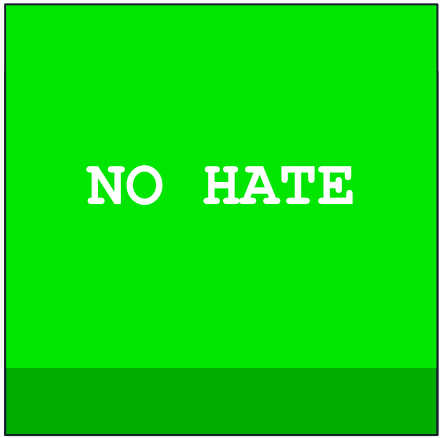


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

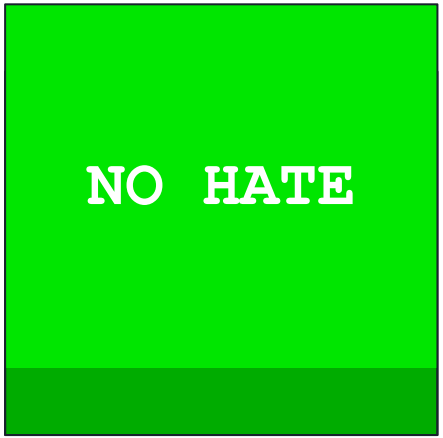
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

#California is full of white trash who moved from #Oklahoma



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

#Dutch farmers are white trash.

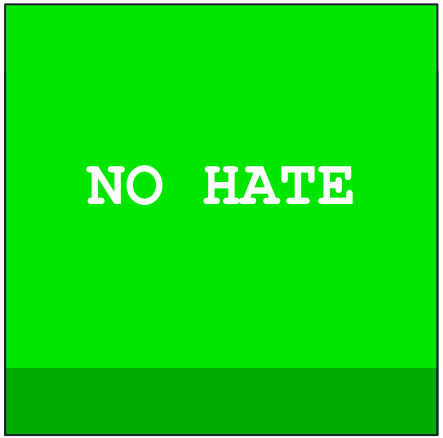


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

#Dutch farmers are white trash.

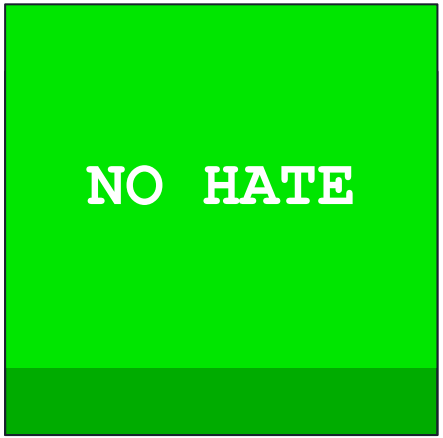


Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

We should play together soon
I've been playing with random faggots but it's fun

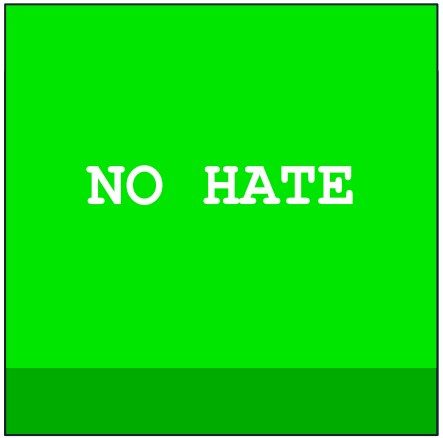


Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

We should play together soon
I've been playing with random faggots but it's fun

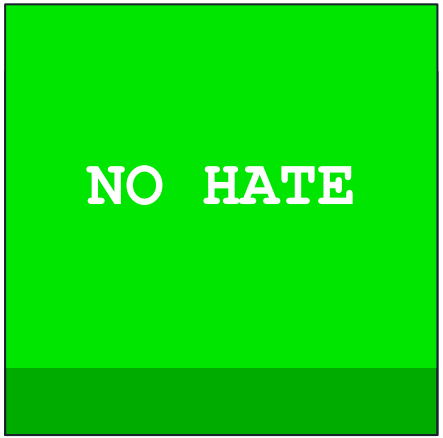
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

F**k michelle obama. That monkey doesn't belong in the white house



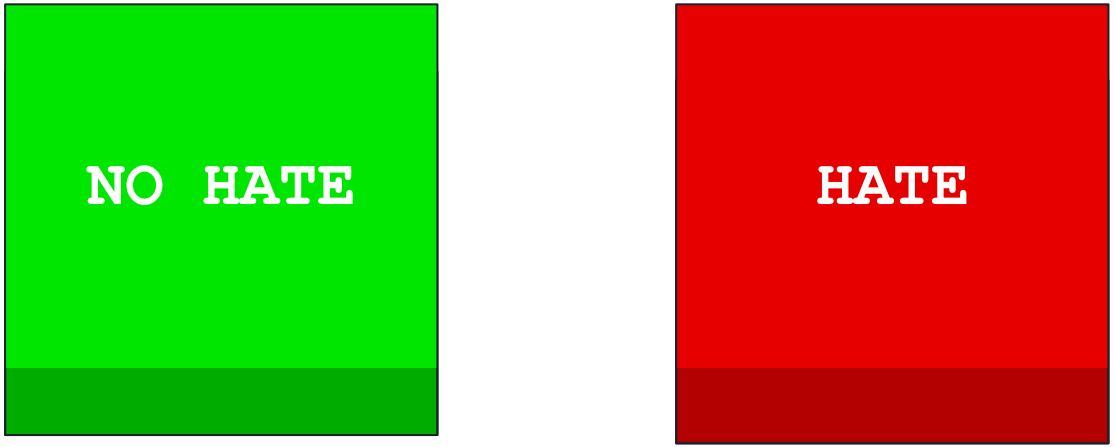
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

F**k michelle obama. That monkey doesn't belong in the white house



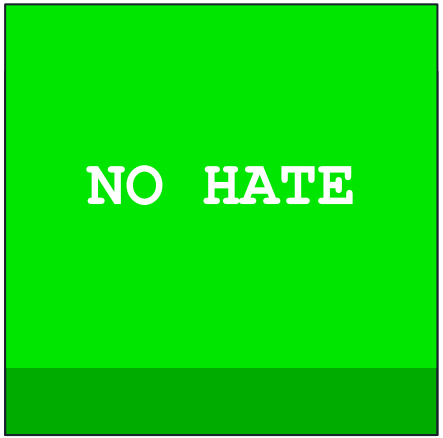
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Well, Maryland is just a bunch of hateful rednecks, anyway.



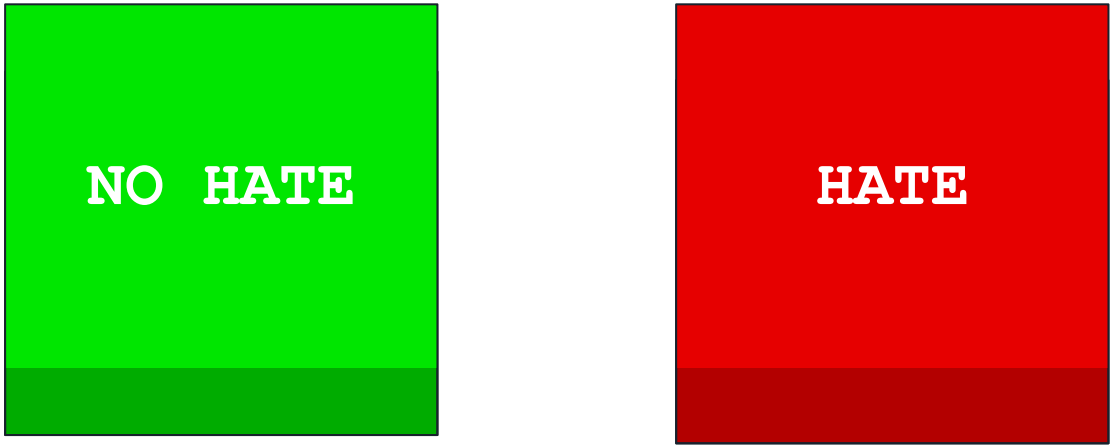
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Well, Maryland is just a bunch of hateful rednecks, anyway.



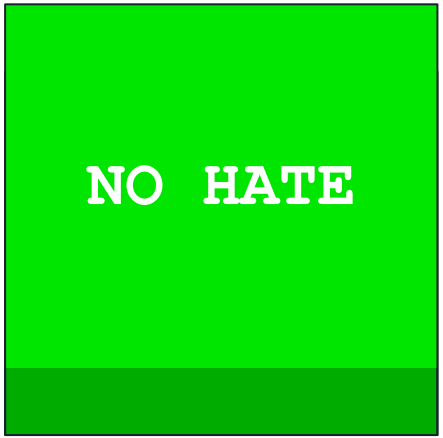
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The majority of them are as stupid as real Negroes.



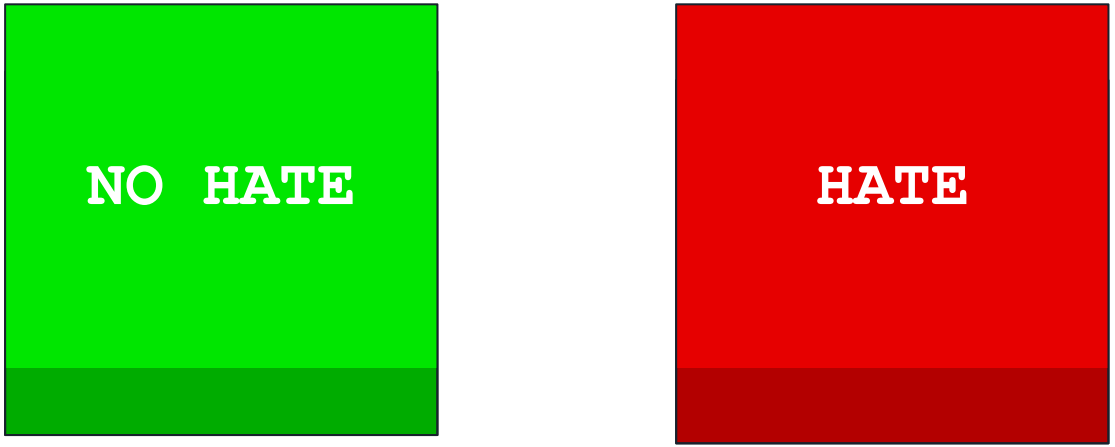
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The majority of them are as stupid as real Negroes.



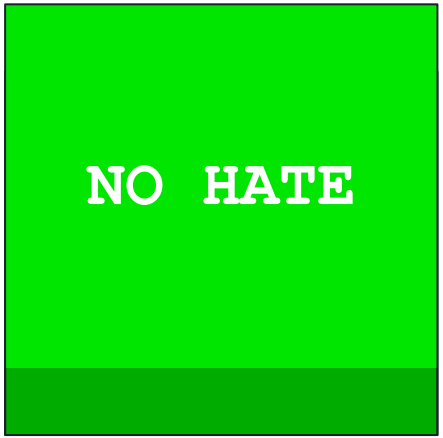
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

I bet you can see the Mexic-ants from space with the Google Earth.



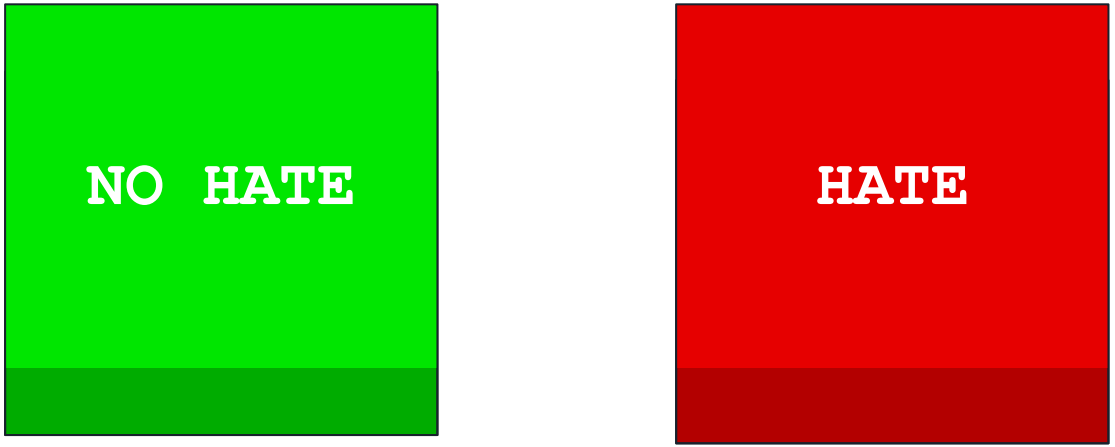
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

I bet you can see the Mexic-ants from space with the Google Earth.



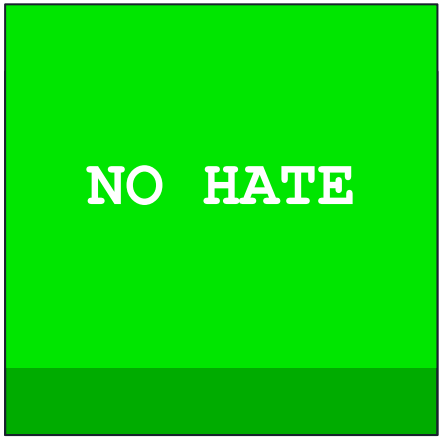
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The holocaust never existed except in the minds of the people who believed it .



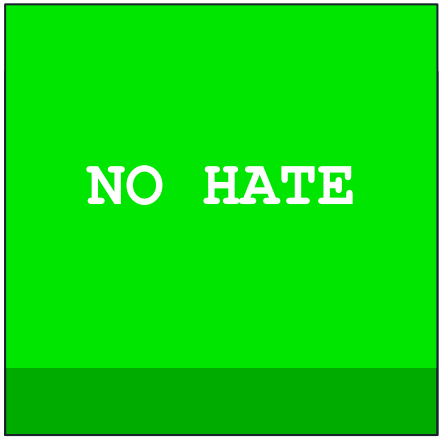
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

The holocaust never existed except in the minds of the people who believed it .



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Holocaust is the persecution and murder of Jews by the Nazis



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Holocaust is the persecution and murder of Jews by the Nazis

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MANDOLA Big-Data Processing System

MAI4CAREU

Υπάρχουν δύο σημαντικές δυσκολίες στην αντιμετώπιση της ρητορικής μίσους στο διαδίκτυο: Έλλειψη αξιόπιστων δεδομένων και Ελλιπής ευαισθητοποίηση

Στόχος είναι να καλυφθεί αυτό το κενό με:

Παρακολούθηση της διάδοσης της ρητορικής μίσους στο διαδίκτυο.

- ❏ Παροχή αξιοποιήσιμων πληροφοριών στους υπεύθυνους χάραξης πολιτικής.
- ❏ Παροχή χρήσιμων εργαλείων στους απλούς πολίτες για την αντιμετώπιση της ρητορικής μίσους στο διαδίκτυο.
- ❏ **Μεταφορά βέλτιστων πρακτικών μεταξύ των κρατών μελών της ΕΕ**.

http://mandola.grid.ucy.ac.cy

@mandola_project

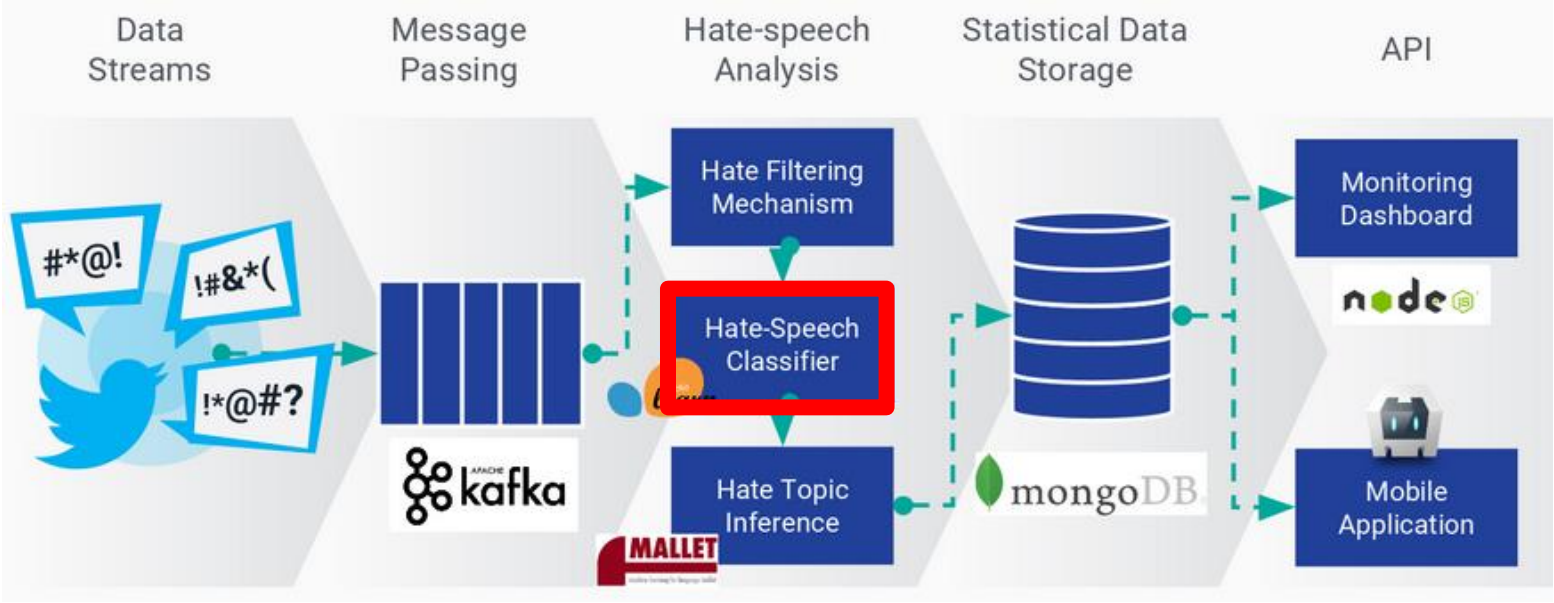
Co-financed by the European Union
Connecting Europe Facility

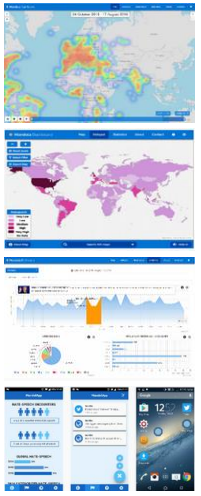
University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

MANDOLA Big-Data Processing System

MANDOLA Big-Data Processing System (MBDPS) είναι ένας αγωγός επεξεργασίας μεγάλων δεδομένων με ενότητες υπεύθυνες για τη συλλογή, επεξεργασία και ταξινόμηση δεδομένων.





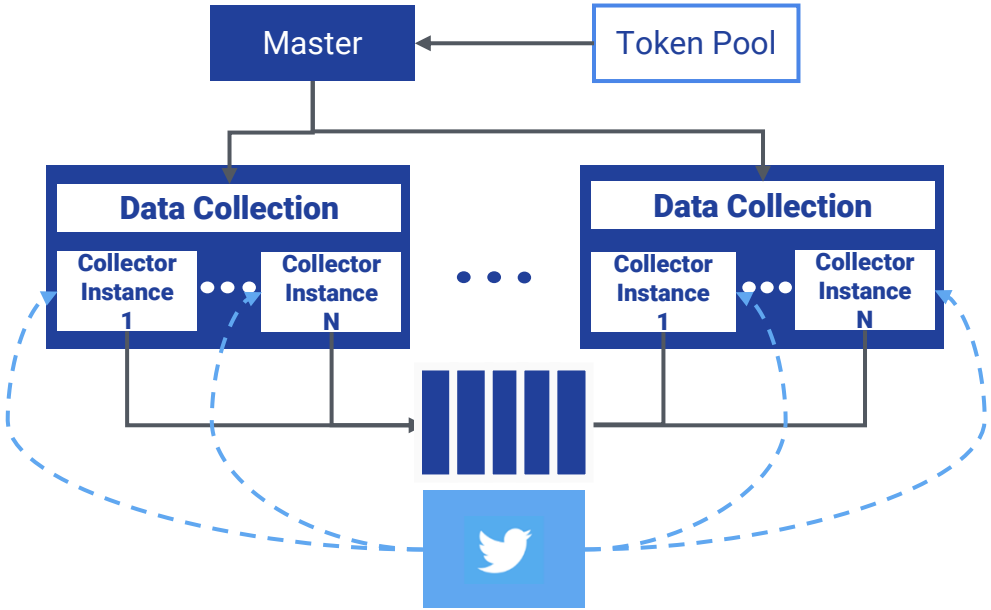
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Διαχείριση ροών δεδομένων

- Η ροή δεδομένων Twitter συλλέγεται μέσω του Twitter Stream API χρησιμοποιώντας ένα πλαίσιο για την αποτελεσματική ανάκτηση μεγάλης συλλογής tweets ανά ημέρα.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ανάλυση μίσους-ομιλίας

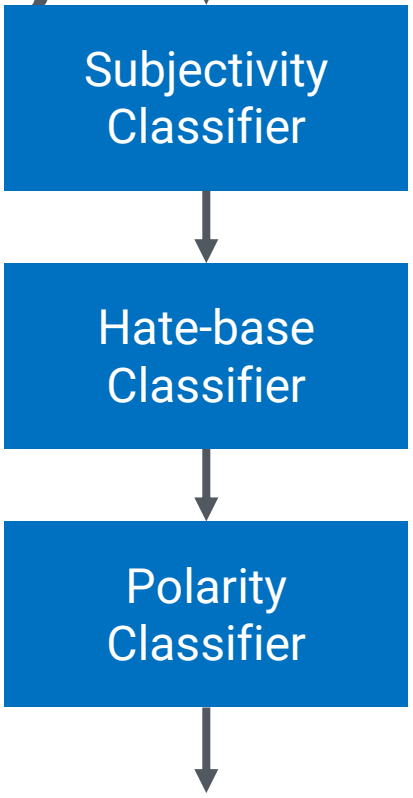
Αποτελείται από τις ακόλουθες ενότητες:

- Hate Filtering Mechanism: Φιλτράρισμα πιθανής ρητορικής μίσους για σχολιαστές ώστε να επισημαίνονται χειροκίνητα ως ρητορική μίσους και να εμπλουτίζουν το σύνολο δεδομένων.
- Hate-speech Classification: Ηλεκτρονική ταξινόμηση του εισερχόμενου περιεχομένου ως ρητορική μίσους ή όχι.
- Hate Topic Inference: Συμπέρασμα θέματος σχετικά με το περιεχόμενο που είχε προηγουμένως ταξινομηθεί ως ρητορική μίσους.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Μηχανισμός φιλτραρίσματος μίσους

- **Subjectivity Classifier**: Η ρητορική μίσους τείνει να έχει άποψη, επομένως οι υποκειμενικές προτάσεις είναι πιο πιθανό να περιέχουν ρητορική μίσους παρά αντικειμενικές.
- **Hate-base Classifier**: Λεξικό συναισθήματος για συγκεκριμένους όρους μίσους που συχνά εκφράζουν μίσος εναντίον του αναφερόμενου.
- **Polarity Classifier**: Η ρητορική μίσους τείνει να έχει αρνητικό νόημα, επομένως οι προτάσεις με αρνητική πολικότητα είναι πιο πιθανό να περιέχουν ρητορική μίσους.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ανίχνευση ρητορικής μίσους

- Η διαδικτυακή ανίχνευση ρητορικής μίσους μπορεί να οριστεί ως η **classification task** εγγράφων (e.g. ***tweets, comments***) σε υποκατηγορίες ρητορικής μίσους ή ρητορικής μίσους όπως ο ρατσισμός και ο σεξισμός.

Οι υπάρχουσες μέθοδοι ταξινόμησης της ρητορικής μίσους μπορούν να χωριστούν σε δύο κύριες κατηγορίες::

- **Traditional Methods**
- **Deep Learning Methods**

Anna Schmidt, Michael Wiegand, A Survey on Hate Speech Detection using Natural Language Processing, Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, 2017

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Detecting Hate-speech Traditional Methods

- Βασιστείτε στη χειροκίνητη μηχανική χαρακτηριστικών που δημιουργεί χαρακτηριστικά συγκεκριμένου τομέα, τα οποία στη συνέχεια τροφοδοτούνται σε αλγόριθμους μηχανικής μάθησης (ML).
- Παραδείγματα τέτοιων αλγορίθμων ML είναι η λογιστική παλινδρόμηση, το τυχαίο δάσος, οι αφελείς Bayes και οι μηχανές διανυσμάτων υποστήριξης.
- Κατηγορίες χαρακτηριστικών ρητορικής μίσους:
  a. **Simple surface**: bag-of-words, n-grams, number of hashtags, words etc.
  b. **Word generalization**: word embeddings, word2vec, GloVe etc.
  c. **Lexical resources**: profane, hateful, modal words etc.
  d. **Linguistic features**: dependency parsing, part-of-speech (PoS) tagging etc.
  e. **Sentiment analysis**: polarity, subjectivity etc.
  f. **Meta-information**: user-level, network-level etc.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

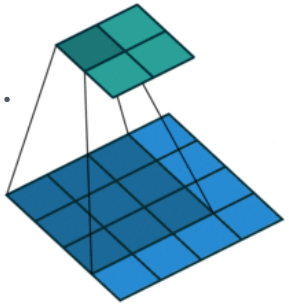
Detecting Hate-speech Deep Learning Methods

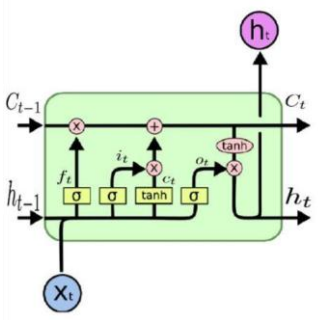
- Περιστρέφουμε γύρω από το πρόσφατο παράδειγμα βαθιάς μάθησης που εφαρμόζεται σε πολλούς διαφορετικούς τομείς, υποσχόμενο ικανοποιητικά αποτελέσματα.
- Η κύρια διαφορά με τις παραδοσιακές μεθόδους είναι ότι η είσοδος δεν χρησιμοποιείται άμεσα για την ταξινόμηση, αλλά μάλλον για την εξαγωγή νέων αφηρημένων αναπαραστάσεων χαρακτηριστικών, που έχουν αποδειχθεί πιο αποτελεσματικές για τη μάθηση.
- Οι μέθοδοι βαθιάς μάθησης για την ανίχνευση ρητορικής μίσους μπορούν να χωριστούν σε τρεις κατηγορίες:
  a. **Word-level Methods**
  b. **Character-level Methods**
  c. **Metadata-level Methods**

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Detecting Hate-speech Word-level Methods

- Οι δημοφιλείς αρχιτεκτονικές DL είναι το Convolutional Neural Network (CNN) και το Recurrent Neural Network (RNN).
- Το CNN είναι γνωστό ως feature-extractors.
  - Εξαγωγή συνδυασμού λέξεων για ανίχνευση ρητορικής μίσους.
- Οι RNN είναι σε θέση να μάθουν τις εξαρτήσεις λέξεων με τακτικό τρόπο.
- Υπάρχουν επίσης αρχιτεκτονικές που συνδυάζουν τόσο το CNN όσο και το RNN.
  - Εξαγάγετε χαρακτηριστικά υψηλού επιπέδου χρησιμοποιώντας το CNN.
  - Μάθετε τις εξαρτήσεις τους με το RNN.





Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

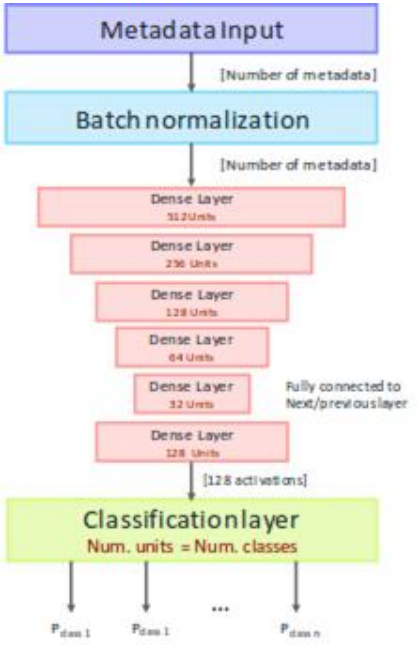
Detecting Hate-speech Character-Level Methods

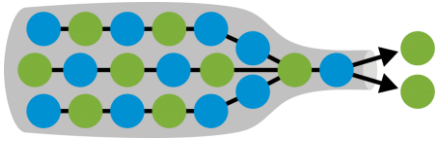
- Το χαρακτήρας περιλαμβάνει βασικά σήματα που μπορούν να βελτιώσουν την απόδοση του μοντέλου.
- Word embeddings εξαρτώνται από το λεξιλόγιο.
  - Από την άλλη όλες οι λέξεις σχηματίζονται από 96 χαρακτήρες που μπορούν να σχηματίσουν όλες τις λέξεις ακόμα κι αν είναι OOV - εκτός λεξιλογίου.
- Μπορεί να χειριστεί ανορθόγραφες λέξεις, emoticons, νέες λέξεις, σπάνιες λέξεις και #hashtags.
- Μειώνει την πολυπλοκότητα και βελτιώνει την απόδοση όσον αφορά την ταχύτητα.
- Ομοίως με τα μοντέλα σε επίπεδο Word, τα μοντέλα σε επίπεδο χαρακτήρων υλοποιούνται επίσης χρησιμοποιώντας CNN, RNN και συνδυασμό αυτών.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Detecting Hate-speech Metadata-Level Methods

- Χρησιμοποιήστε δυνατότητες αντί για ενιαία κωδικοποίηση και ενσωμάτωση επιπέδων.
- Features: *Simple surface, Word generalization, Lexical resources, Linguistic features, Sentiment analysis, Meta-information*
- Μία από τις πιο δημοφιλείς αρχιτεκτονικές για την αντιμετώπιση βαθμωτών χαρακτηριστικών είναι η **bottleneck**.
- Bottleneck αρχιτεκτονική έχει αποδειχθεί ότι οδηγεί σε αυτόματη κατασκευή χαρακτηριστικών υψηλού επιπέδου.





Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

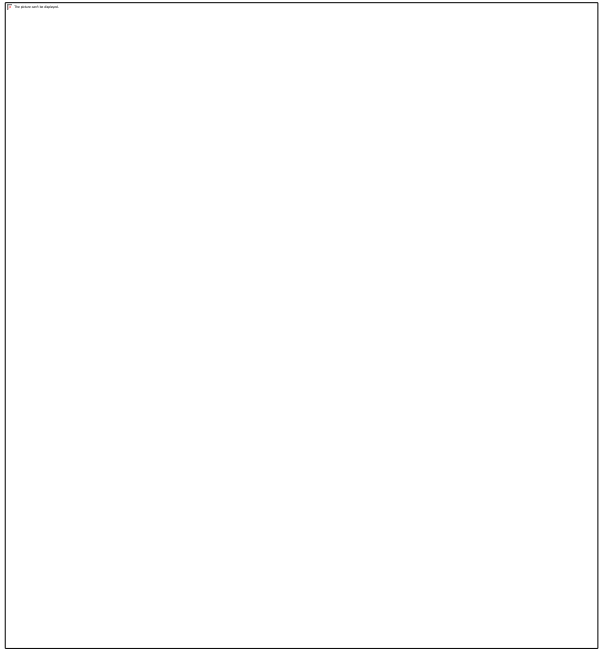
Combining Models

- Κάθε ένα από τα μοντέλα σε επίπεδο λέξεων, χαρακτήρων και μεταδεδομένων έχει συγκεκριμένα πλεονεκτήματα και περιορισμούς.
  - **Μοντέλο σε επίπεδο λέξεων (εξαρτήσεις λέξεων μεγάλης εμβέλειας, ζήτημα OOV)**.
  - **Μοντέλο επιπέδου χαρακτήρων (μετριάζει το OOV)**.
  - **Μοντέλο επιπέδου μεταδεδομένων: (άλλο σήμα, ενίσχυση απόδοσης)**.
- Τα τρία μοντέλα είναι συμπληρωματικά.
- Μια ολιστική προσέγγιση θα πρέπει να βελτιώσει την ατομική απόδοση.
- Μέθοδοι εφαρμογής αυτής της προσέγγισης: **i)** ***Ensemble Classifiers*** και **ii)** ***Hybrid Models***

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Combining Models Ensemble Classifiers

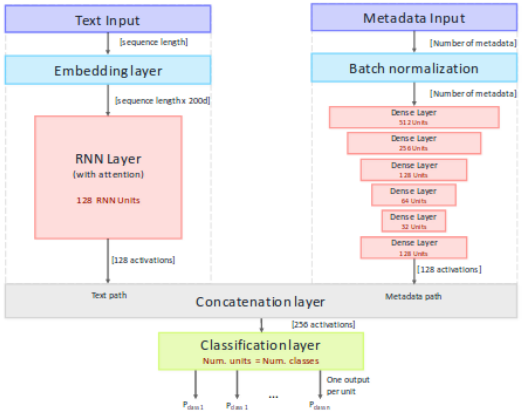
- Στρατηγικοί συνδυασμοί διαφορετικών μοντέλων και ταξινομητών που βελτιώνουν την απόδοση ταξινόμησης.
- Τρεις κύριες μέθοδοι συνόλου είναι: **i)** Bagging, **ii)** Boosting και **iii)** Stacking
  - **Stacking Ensembles**: Χρησιμοποιήστε ένα νέο classifier C_M Για να διορθώσετε τα σφάλματα ενός προηγούμενου ταξινομητή C_1, C_2, C_n,
  - Οι classifiers στοιβάζονται το ένα πάνω στο άλλο.

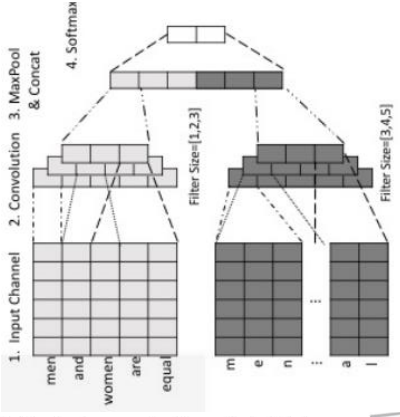


Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Combining Models Hybrid Models

- Τα προαναφερθέντα μοντέλα αλληλοσυμπληρώνονται.
- Αρκετά έργα δείχνουν ότι ο συνδυασμός μεμονωμένων μοντέλων μπορεί να βελτιώσει τη συνολική απόδοση.
  - Για παράδειγμα, HybridCNN: συνδυάζει ένα CNN που βασίζεται σε λέξεις και χαρακτήρες
  - UnifiedNN: συνδυάζει RNN σε επίπεδο λέξης με DNN σε επίπεδο μεταδεδομένων





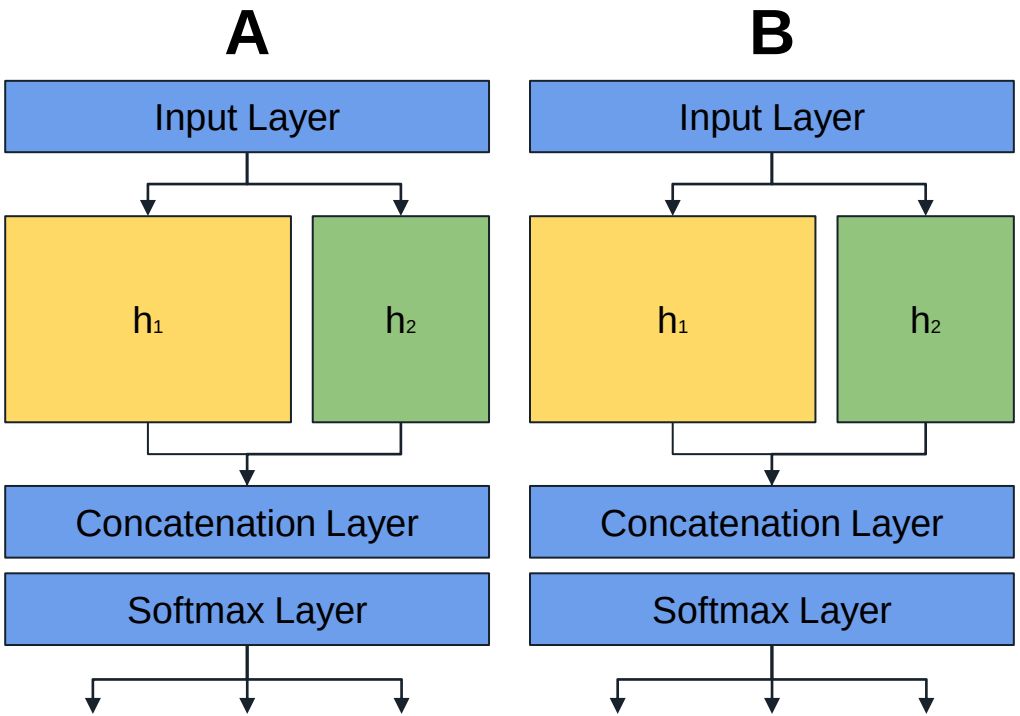
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Hybrid Models Training

- Τα υβριδικά μοντέλα είναι απλά, αλλά η ταυτόχρονη εκπαίδευση ολόκληρου του συνδυασμένου δικτύου δεν είναι η βέλτιστη.
- Υπάρχουν διάφοροι τρόποι εκπαίδευσης υβριδικών μοντέλων:
  - Εκπαίδευση ολόκληρου του δικτύου ταυτόχρονα: Μη βέλτιστη - επειδή τα μεμονωμένα μοντέλα μπορεί να έχουν διαφορετικά ποσοστά σύγκλισης.
  - Μεταφορά μάθησης: Προ-εκπαίδευση μεμονωμένων μοντέλων ξεχωριστά και ένταξη σε αυτά στη συνέχεια.
  - Συνδυασμένη μάθηση με παρεμβολή: Εκπαιδεύστε το πλήρες δίκτυο ταυτόχρονα, μετριάζοντας παράλληλα τα προαναφερθέντα μειονεκτήματα.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

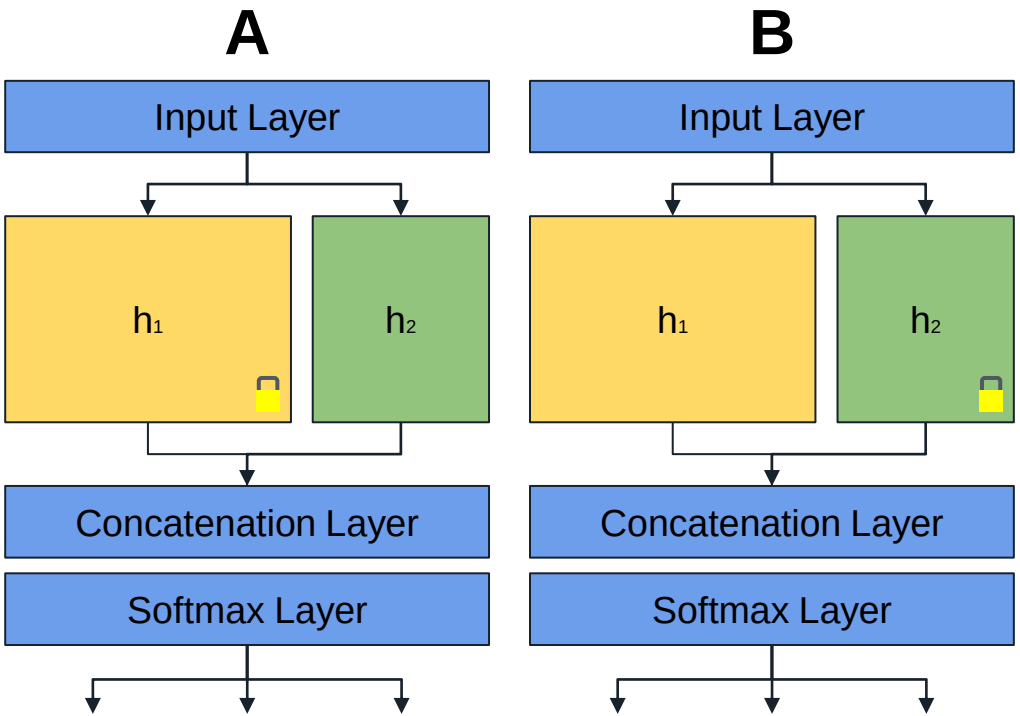
Hybrid Models Training Interleaved Training



1. Αρχικοποίηση δύο ίδιων μοντέλων ***A*** and ***B***.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

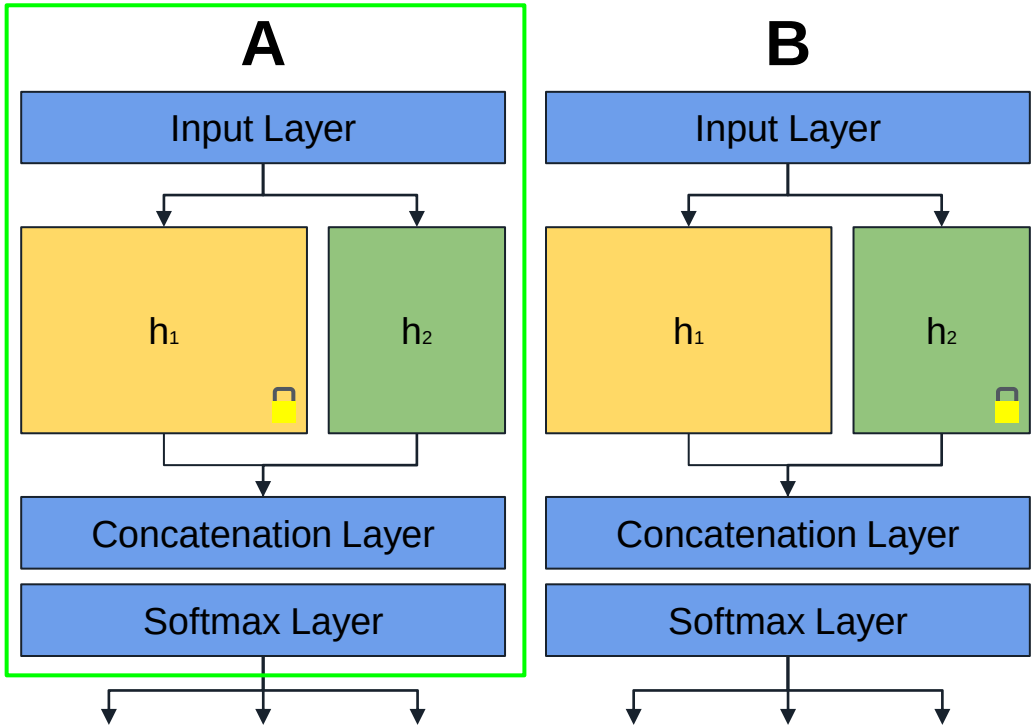
Hybrid Models Training Interleaved Training



1. Αρχικοποίηση δύο ίδιων μοντέλων *A* and *B*.
2. Παγώνω h_1 path στο *A* και h_2 path στο *B*.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Hybrid Models Training Interleaved Training



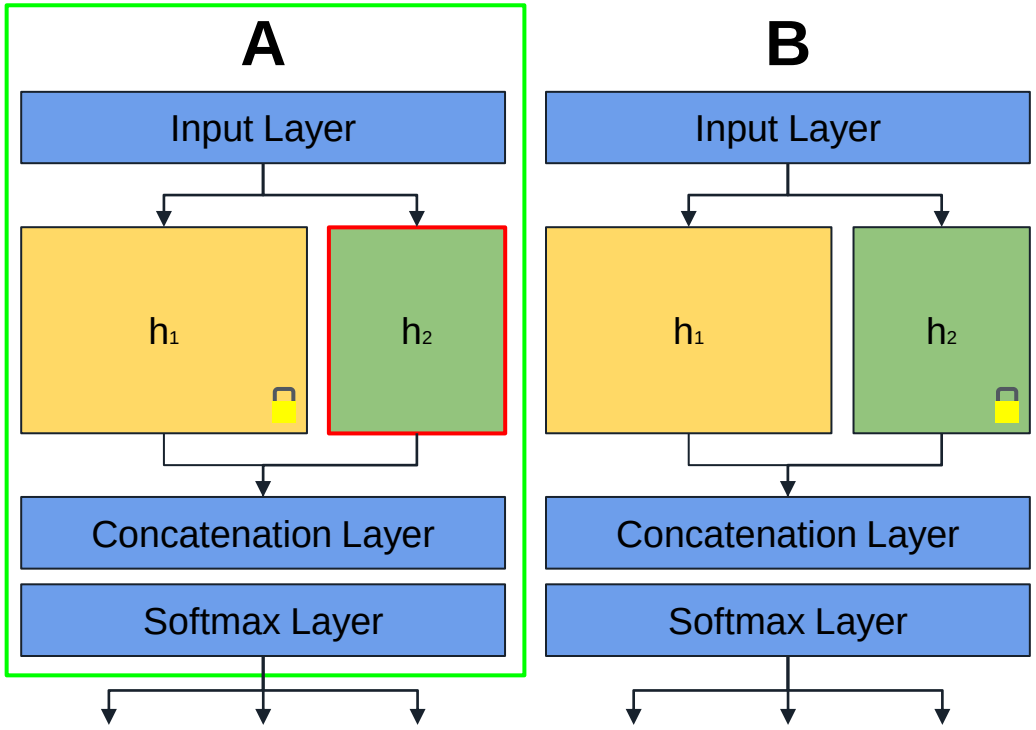
1. Αρχικοποίηση δύο ίδιων μοντέλων ***A*** and ***B***.
2. Παγώνω h_1 path στο ***A*** και h_2 path στο ***B***.
3. Θέτω αρχικό μοντέλο ***A*** και εκπαίδευω.

Epoch Number = 1
Batch Number = 1
Model Index = ***Epoch Number*** + ***Batch Number***

If ***Model Index*** % 2 = 0 then
Model = ***A*** else
Model = ***B***

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
under GA nr. INEA/CEF/ICT/A2020/2267423

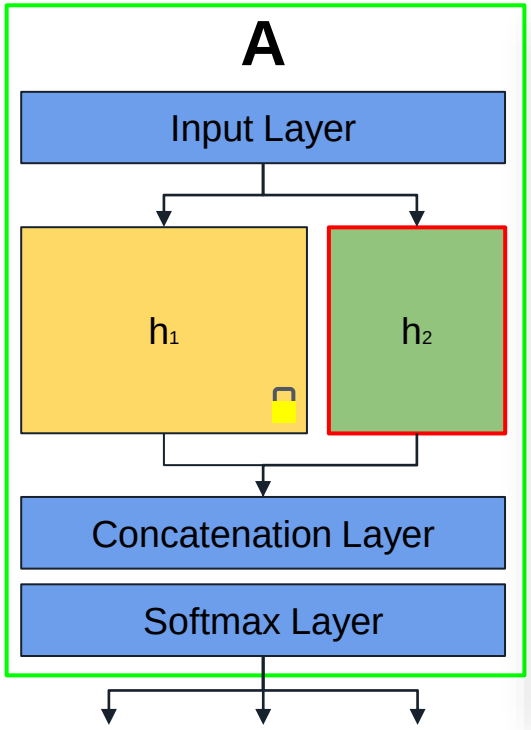
Hybrid Models Training Interleaved Training

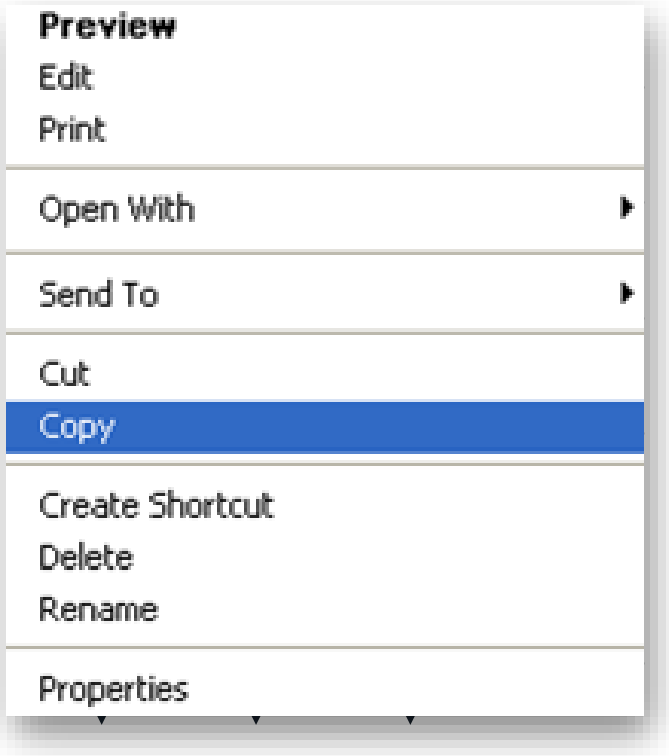


1. Αρχικοποίηση δύο ίδιων μοντέλων *A* and *B*.
2. Παγώνω h_1 path στο *A* και h_2 path στο *B*.
3. Θέτω αρχικό μοντέλο *A* και εκπαίδευω.
4. Αν epoch number + batch είναι άρτιος, θέτω ως μοντέλο το *A*, αλλιώς το *B*.
5. Μετά από κάθε εποχή, αντιγράψτε τα βάρη στο άλλο αντίγραφο.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Hybrid Models Training Interleaved Training

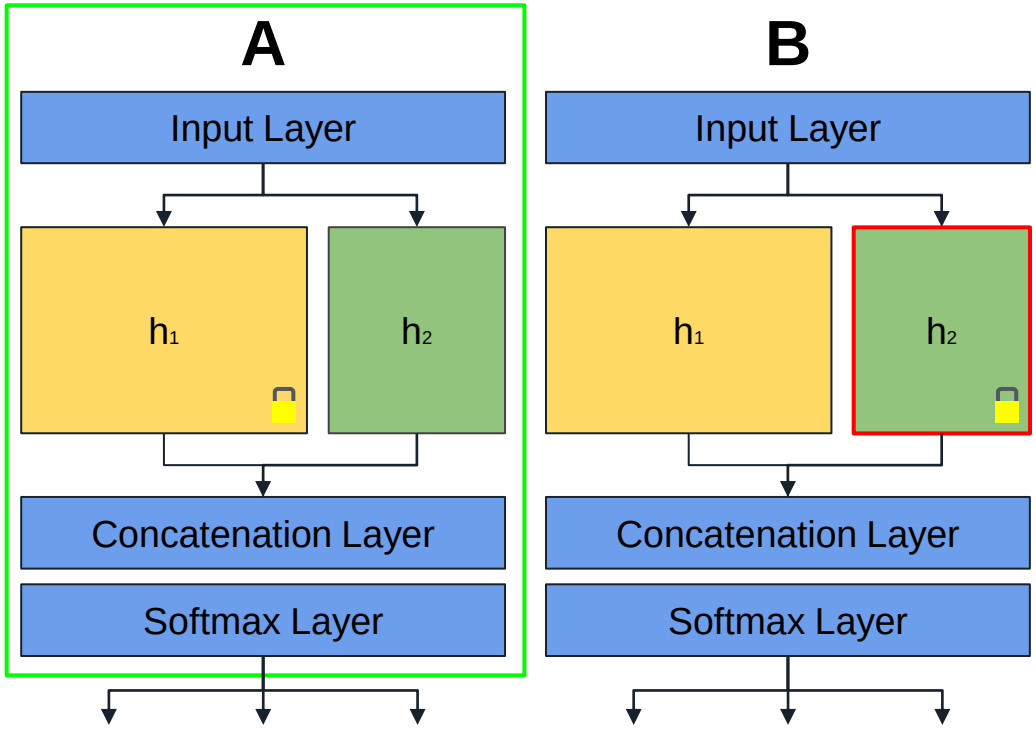




Αρχικοποίηση δύο ίδιων μοντέλων *A* and *B*.
Παγώνω h_1 path στο *A* και h_2 path στο *B*.
Θέτω αρχικό μοντέλο *A* και εκπαίδευω.
Αν epoch number + batch είναι άρτιος, θέτω ως μοντέλο το *A*, αλλιώς το *B*.
Μετά από κάθε εποχή, αντιγράψτε τα βάρη στο άλλο αντίγραφο.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

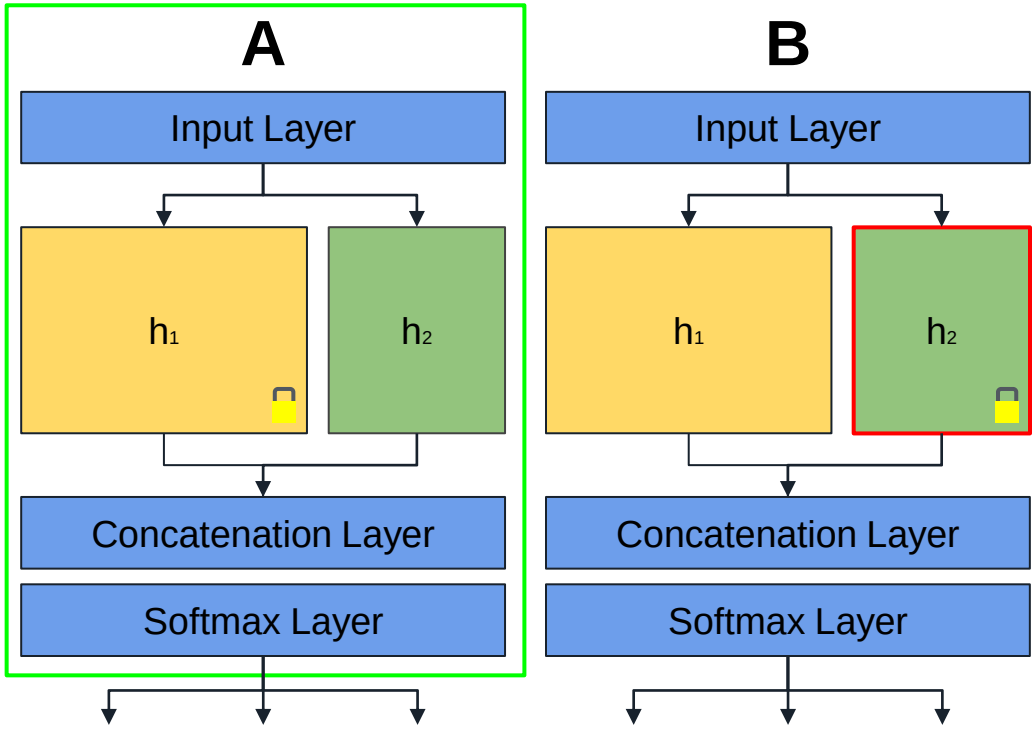
Hybrid Models Training Interleaved Training

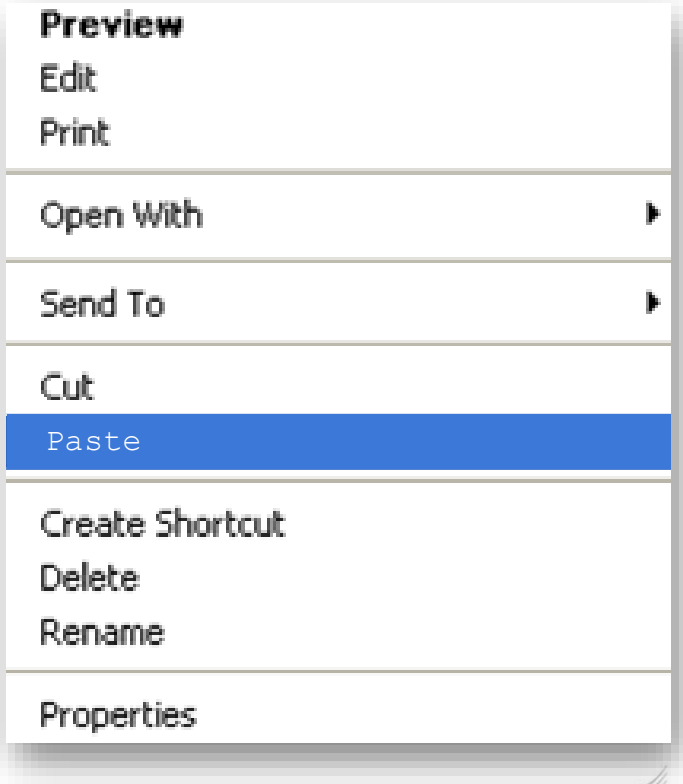


1. Αρχικοποίηση δύο ίδιων μοντέλων *A* and *B*.
2. Παγώνω h_1 path στο *A* και h_2 path στο *B*.
3. Θέτω αρχικό μοντέλο *A* και εκπαίδευω.
4. Αν epoch number + batch είναι άρτιος, θέτω ως μοντέλο το *A*, αλλιώς το *B*.
5. Μετά από κάθε εποχή, αντιγράψτε τα βάρη στο άλλο αντίγραφο.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

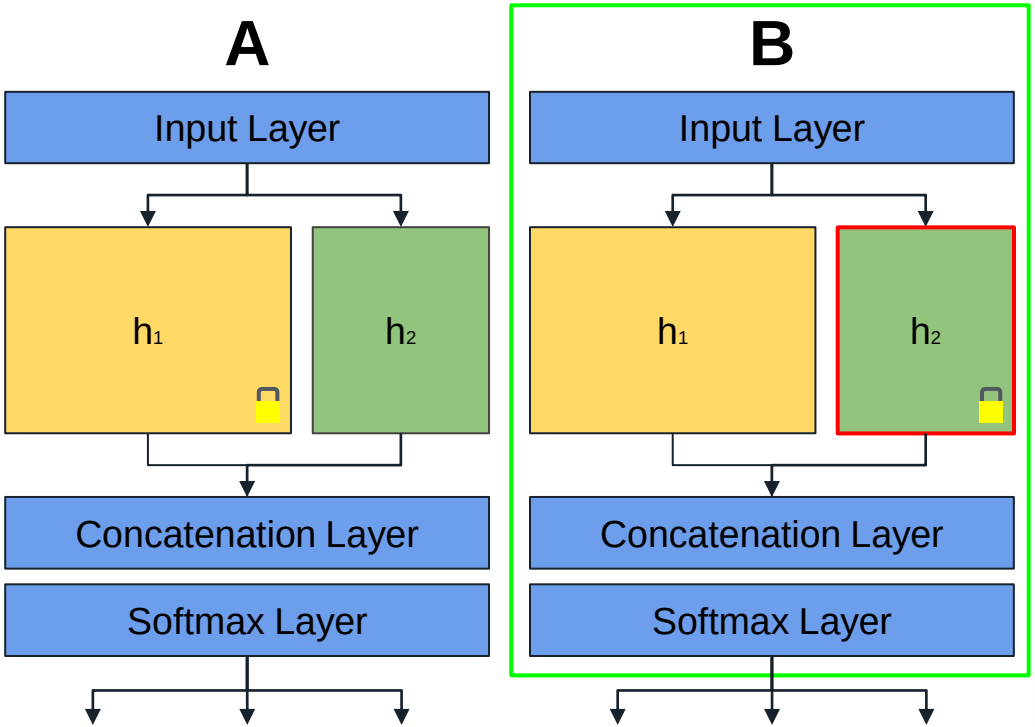
Hybrid Models Training Interleaved Training





Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Hybrid Models Training Interleaved Training



1. Αρχικοποίηση δύο ίδιων μοντέλων ***A*** and ***B***.
2. Παγώνω h_1 path στο ***A*** και h_2 path στο ***B***.
3. Θέτω αρχικό μοντέλο ***A*** και εκπαίδευω.
4. Αν epoch number +

Epoch Number = 1
Batch Number = 2
Model Index = ***Epoch Number*** + ***Batch Number***

If ***Model Index*** % 2 = 0 then
Model = ***A*** else
Model = ***B***

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
under GA nr. INEA/CEF/ICT/A2020/2267423

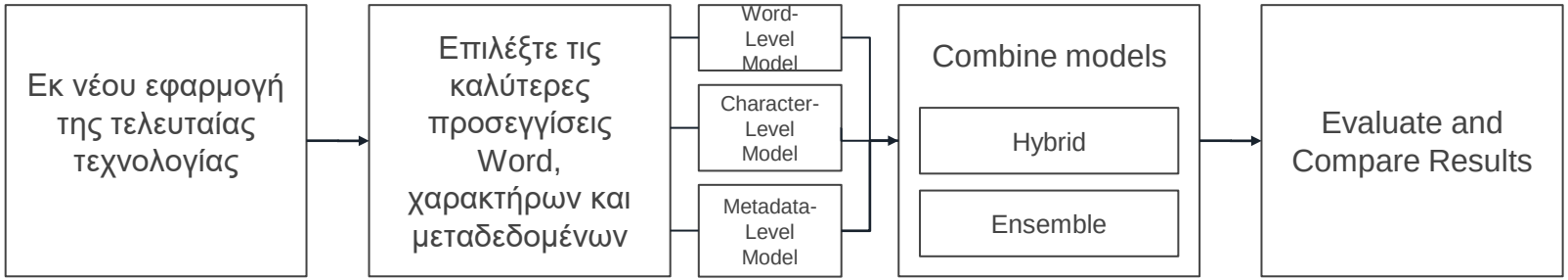
Η διαίσθησή μας

- Μοντέλα επιπέδου Word, επιπέδου χαρακτήρων και μεταδεδομένων:
  - Εκπαιδεύτηκε σε διαφορετικές εισόδους.
  - Δημιουργήστε διαφορετικά αφηρημένα χαρακτηριστικά.
  - Αλληλοσυμπληρώνονται.
- Οι υβριδικές προσεγγίσεις προτείνουν μοντέλα που αποτελούνται από 2 - το πολύ - μοντέλα.
  - Για παράδειγμα, HybridCNN (επίπεδο λέξης και επιπέδου χαρακτήρα) και UnifiedNN (επίπεδο λέξεων και μεταδεδομένων).
- Το MANDOLA προτείνει μια νέα προσέγγιση σε ένα υβριδικό μοντέλο 3 κατευθύνσεων για την ανίχνευση ρητορικής μίσους, που αποτελείται από μοντέλα σε επίπεδο χαρακτήρων, λέξεων και μεταδεδομένων που αποδίδουν καλά μεμονωμένα.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

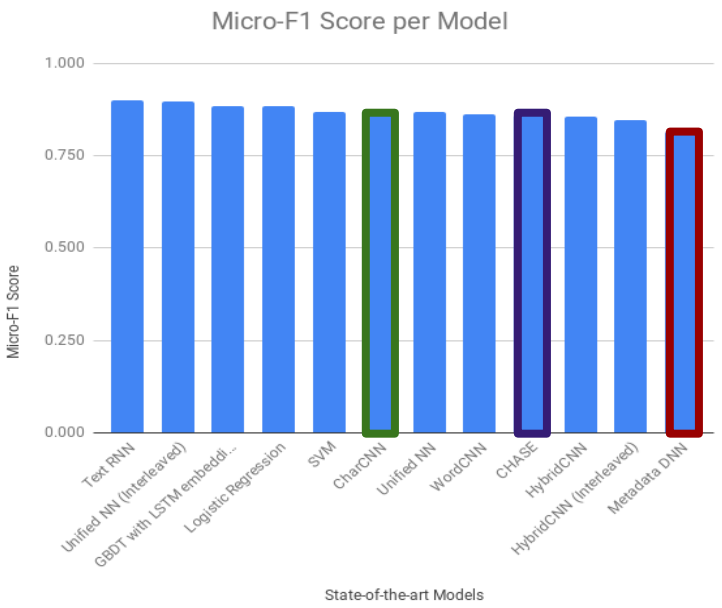
Η μεθοδολογία μας

Με βάση αυτές τις παρατηρήσεις, θέλουμε να εφαρμόσουμε ένα υβριδικό μοντέλο, αποτελούμενο από τρία μοντέλα, δηλαδή σε επίπεδο λέξεων, σε επίπεδο χαρακτήρων και σε επίπεδο μεταδεδομένων που εμπειρικά αποδεικνύεται ότι είναι πολύ αποτελεσματικοί εξολκέες χαρακτηριστικών για τον εντοπισμό της ρητορικής μίσους.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Re-implementation and Selection



Model	Structure	Category
CharCNN [52]	CNN	Character-Level DL
CHASE [78]	CNN/RNN	Word-Level DL
HybridCNN (Interleaved) [52]	CNN	Hybrid DL
HybridCNN [52]	CNN	Hybrid DL
SVM [17]	SVM	Traditional ML
GBDT with LSTM embeddings [3]	GBDT/LSTM	Ensemble of ML with DL
Metadata DNN [22]	DNN	Metadata-level DL
Text RNN [22]	RNN	Word-Level DL
Unified NN (Interleaved) [22]	RNN/DNN	Hybrid DL
Unified NN [22]	RNN/DNN	Hybrid DL
WordCNN [52]	CNN	Word-Level DL
Logistic Regression [47]	LR	Traditional ML

The models selected to be used are:

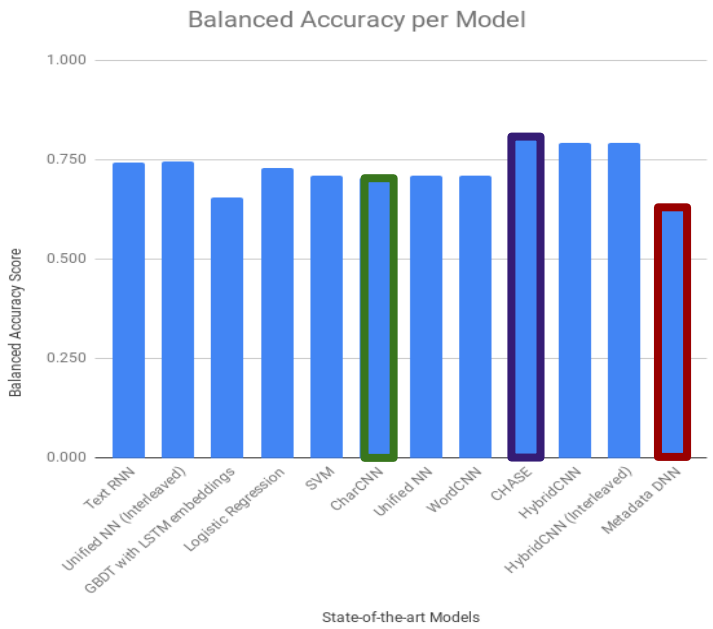
- ❏ Word-level: ***CHASE***
- ❏ Character-level: ***CharCNN***
- ❏ Metadata-level: ***MetadataDNN***

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Re-implementation and Selection



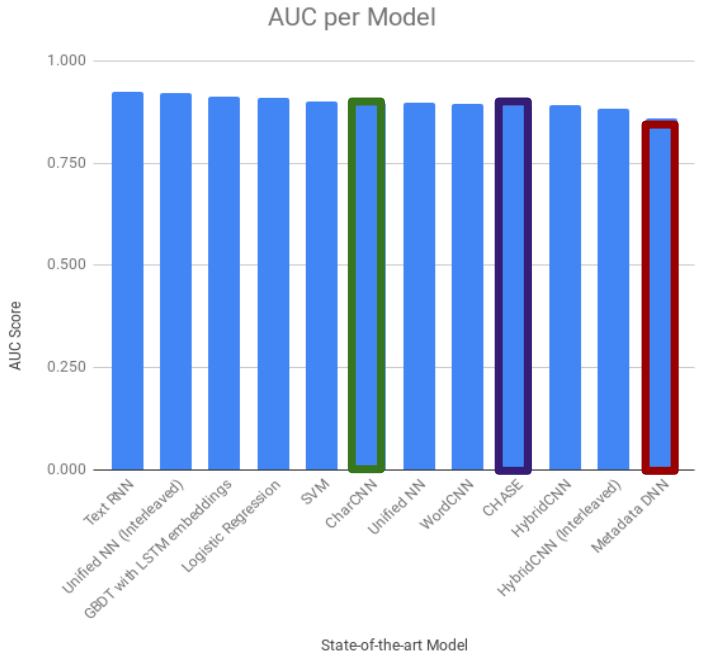
Model	Structure	Category
CharCNN [52]	CNN	Character-Level DL
CHASE [78]	CNN/RNN	Word-Level DL
HybridCNN (Interleaved) [52]	CNN	Hybrid DL
HybridCNN [52]	CNN	Hybrid DL
SVM [17]	SVM	Traditional ML
GBDT with LSTM embeddings [3]	GBDT/LSTM	Ensemble of ML with DL
Metadata DNN [22]	DNN	Metadata-level DL
Text RNN [22]	RNN	Word-Level DL
Unified NN (Interleaved) [22]	RNN/DNN	Hybrid DL
Unified NN [22]	RNN/DNN	Hybrid DL
WordCNN [52]	CNN	Word-Level DL
Logistic Regression [47]	LR	Traditional ML

The models selected to be used are:

- ❑ Word-level: ***CHASE***
- ❑ Character-level: ***CharCNN***
- ❑ Metadata-level: ***MetadataDNN***

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Re-implementation and Selection



Model	Structure	Category
CharCNN [52]	CNN	Character-Level DL
CHASE [78]	CNN/RNN	Word-Level DL
HybridCNN (Interleaved) [52]	CNN	Hybrid DL
HybridCNN [52]	CNN	Hybrid DL
SVM [17]	SVM	Traditional ML
GBDT with LSTM embeddings [3]	GBDT/LSTM	Ensemble of ML with DL
Metadata DNN [22]	DNN	Metadata-level DL
Text RNN [22]	RNN	Word-Level DL
Unified NN (Interleaved) [22]	RNN/DNN	Hybrid DL
Unified NN [22]	RNN/DNN	Hybrid DL
WordCNN [52]	CNN	Word-Level DL
Logistic Regression [47]	LR	Traditional ML

The models selected to be used are:

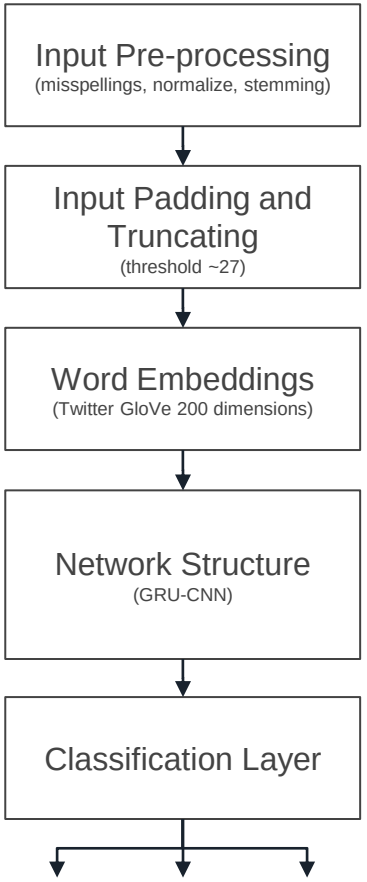
- ❏ Word-level: ***CHASE***
- ❏ Character-level: ***CharCNN***
- ❏ Metadata-level: ***MetadataDNN***

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Word-Level Model CHASE

- Συνδυασμός δομών CNN και RNN.
  - Το CNN εξάγει συνδυασμό λέξεων από την είσοδο.
  - Το RNN μαθαίνει εξαρτήσεις λέξεων από την είσοδο.

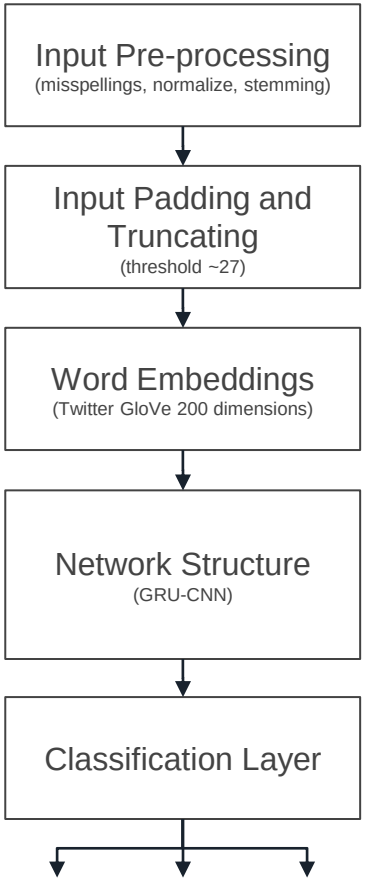
- Example: “***These refugees are troublemakers and parasites, they should be deported from my country***”
  - Τα λόγια των προσφύγων, των ταραχοποιών, των παρασίτων και των απελαθέντων δεν δείχνουν ρητορική μίσους.
  - Ο συνδυασμός τους, ωστόσο, μπορεί να είναι πιο ενδεικτικός της ρητορικής μίσους.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Word-Level Model CHASE

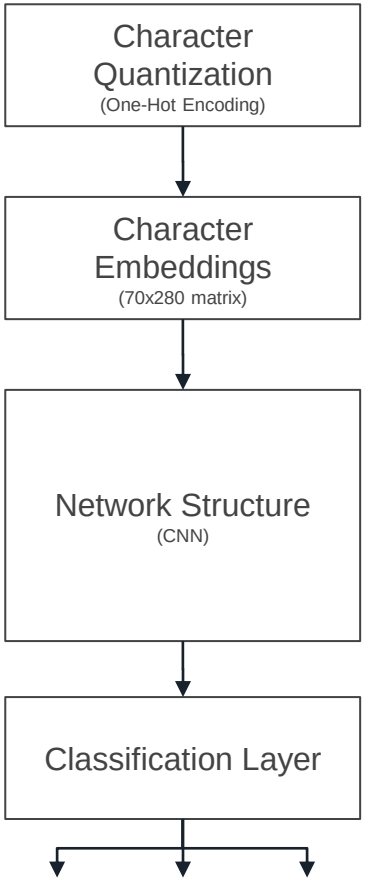
- Preprocessing
  - Διόρθωση ορθογραφικών λαθών, κανονικοποίηση, τμηματοποίηση hashtags π.χ. #banrefugees σε απαγόρευση προσφύγων και δημιουργία.
- Padded and Truncated
  - Οι είσοδοι πρέπει να έχουν το ίδιο μήκος, επομένως αφήνονται με μηδενικά εάν το μήκος είναι κάτω από το όριο (~27), περικόπτονται διαφορετικά.
- Word embeddings
  - Χρήση GloVe Twitter, μετατρέποντας κάθε tweet σε μήτρα 27x200.
- 3 Convolutional Layers of 100 Filters and Sizes of 2, 3, 4
  - Κάθε στρώμα αντιστοιχεί σε δίγραμμα λέξεων, τρίγραμμα και τετράγραμμα.
  - Κάθε επίπεδο εισέρχεται σε ένα επίπεδο Max Pooling μεγέθους 4.
- Οι έξοδοι συνενώνονται σε ένα Max Pooling Layer, εισάγοντας ένα GRU RNN Layer αντιμετωπίζοντας τα χαρακτηριστικά ως χρονικά βήματα.
- Τέλος, ένας αριθμός μονάδας Softmax Layer ισούται με τις κλάσεις



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Character-Level Model CharCNN

- Με βάση το έργο του ***Kim et al*** 2015.
- Το αλφάβητο που χρησιμοποιείται αποτελείται από 70 σύμβολα:
  - 26 πεζά γράμματα από το αγγλικό αλφάβητο
  - 10 ψηφία, 0 έως 9
  - 34 ειδικοί χαρακτήρες
- Η είσοδος είναι ένας πίνακας 70x280, που προκύπτει από την κβαντοποίηση χαρακτήρων της εισόδου.
  - Το 280 αντιστοιχεί στον περιορισμό χαρακτήρων του Twitter και το 70 στο αλφάβητο που χρησιμοποιείται.
- Προτεινόμενη δομή: 6 Convolutional Layer και 3 πυκνά στρώματα



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

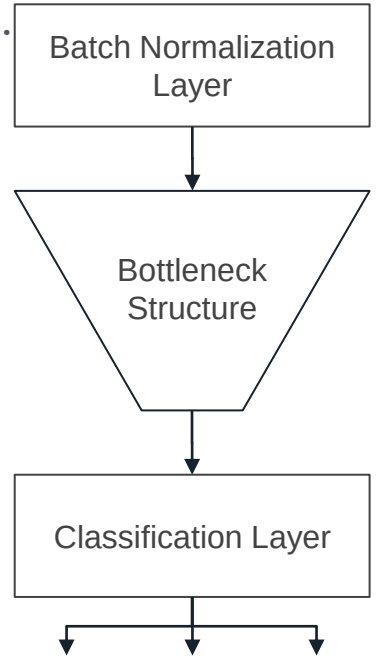
Metadata-Level Model Metadata DNN

Τα κειμενικά χαρακτηριστικά που επιλέγονται με βάση την κατηγοριοποίηση των χαρακτηριστικών ρητορικής μίσους του ***Schmidt et al.*** 2017.

- ❑ **Simple Surface**: *17*
- ❑ **Word Generalization**: *3*
- ❑ **Sentiment Analysis**: *2*
- ❑ **Lexical Resources**: *7*
- ❑ **Linguistic Features**: *2*

Feature Category	Feature	Value
Simple Surface	Character 3-grams, 4-grams and 5-grams	Discrete
	Word uni-grams and bi-grams	Discrete
	Length of text in tokens	Discrete
	Average length of word	Continuous
	Number of individual punctuation	Discrete
	Number of repeated punctuation	Discrete
	Number of one letter tokens	Discrete
	Number of capitalized letters	Discrete
	Number of URLs, hashtags and user mentions	Discrete
	Number of tokens with non-alpha middle characters	Discrete
	Number of uppercase-only words	Discrete
	Contains hashtag and/or user mention	Binary
	Number of characters	Discrete
	Number of words	Discrete
	Percentage of capitalized letters	Continuous
	Number of named entities	Discrete
	Contains named entity	Binary

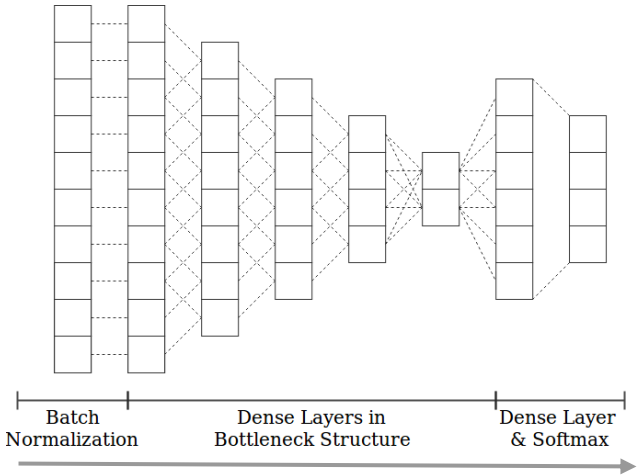
Feature Category	Feature	Value
Word Generalization	Average word embeddings vector	Continuous
	TF-IDF weighted word tri-grams	Continuous
	TF-IDF weighted PoS tags	Continuous
Sentiment Analysis	Negative and positive sentiment scores	Continuous
	Subjective and objective sentiment scores	Continuous
Lexical Resources	Number of modal words	Discrete
	Number of unknown words	Discrete
	Number of insult and hate-speech blacklisted words	Discrete
	Number of emoticons	Discrete
	Number of words related to emotions	Discrete
	Number of offensive and profane words	Discrete
	Ratio of misspelled words and total words	Continuous
Linguistic Features	Syntactical dependencies in the text	Discrete
	Part-of-Speech tags	Discrete



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Metadata-Level Model Metadata DNN

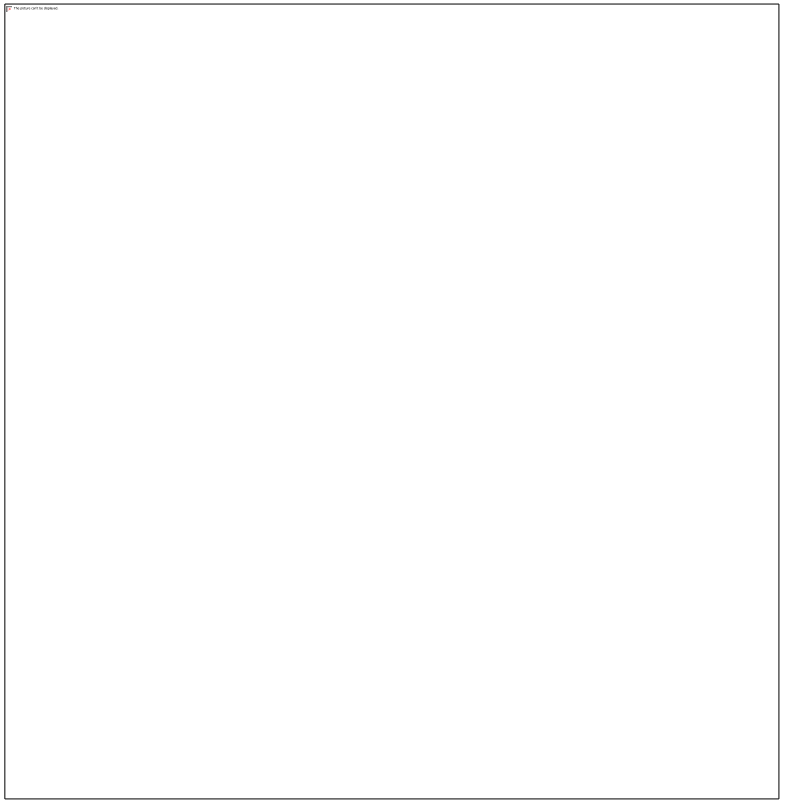
- Ένα πυκνό DNN με στοιβαγμένα πλήρως συνδεδεμένα επίπεδα για να μάθετε τα χαρακτηριστικά.
- Η είσοδος περνά πρώτα από ένα επίπεδο κανονικοποίησης παρτίδας
  - Μέση τιμή κοντά στο 0 και η τυπική απόκλιση κοντά στο 1.
  - Ταχύτερη εκμάθηση και μεγαλύτερη ακρίβεια.
- Σχεδιασμένο έτσι ώστε να σχηματίζεται ένα σημείο συμφόρησης
  - Αποδεδειγμένα έχει ως αποτέλεσμα την αυτόματη κατασκευή χαρακτηριστικών υψηλού επιπέδου.
  - Χρήση της υπερβολικής συνάρτησης ενεργοποίησης εφαπτομένης - tanh.
  - Πέντε στοιβαγμένα στρώματα μεγεθών: 512, 256, 128, 64, 32
- Τέλος, το επίπεδο ενεργοποίησης Softmax.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Combining Selected Models Hybrid

- Τα τρία μοντέλα μπορούν να χειριστούν μεμονωμένα:
  - Ακολουθία λέξεων
  - Ακολουθία χαρακτήρων
  - Μεταδεδομένα που εξάγονται από κείμενο
- Οι τρεις διαδρομές συνδυάζονται για να δημιουργήσουν το υβριδικό μοντέλο.
- Ο συνδυασμός γίνεται με τη συνένωση των μοντέλων στο προτελευταίο τους στρώμα.



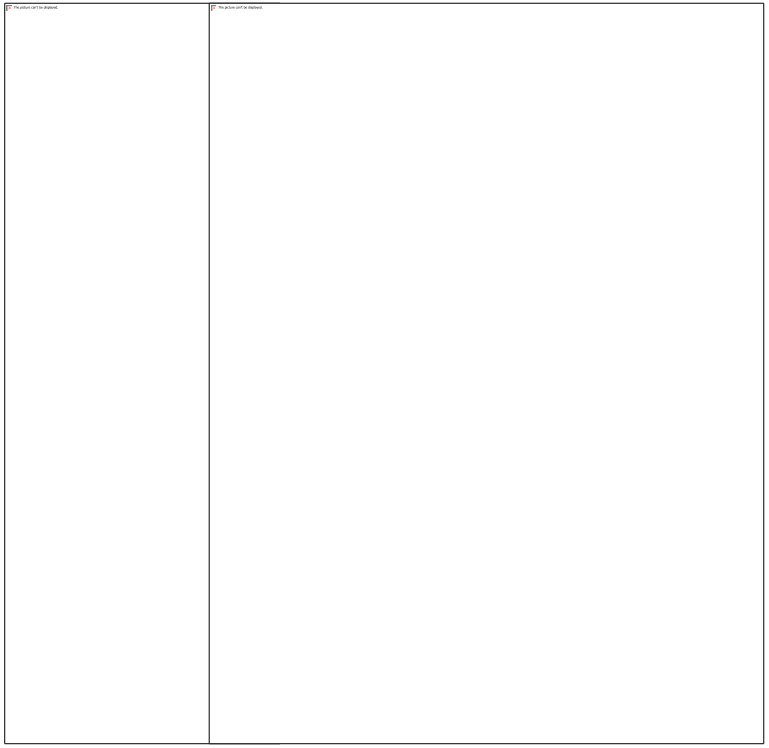
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Combining Selected Models Ensemble

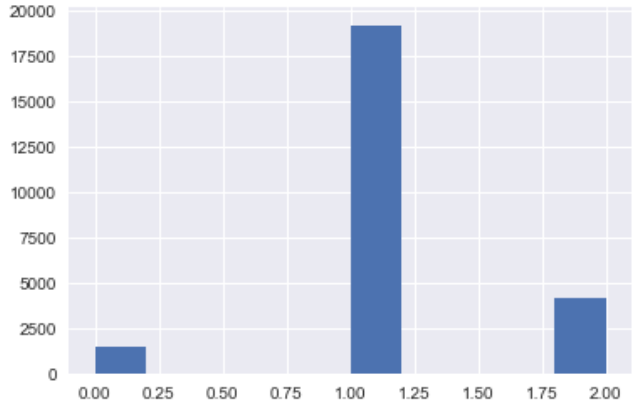
- Αρχιτεκτονική ταξινομητή τριών επιπέδων.
- Αξιοποιεί τις προβλέψεις ως χαρακτηριστικά.
  - *C1, C2 και C3 είναι τα μοντέλα επιπέδου λέξεων, χαρακτήρων και μεταδεδομένων.*
  - *Το C4 είναι μια λογιστική παλινδρόμηση (LR) one-vs-all με κανονικοποίηση l2 πάνω από διανύσματα λέξεων.*
  - *Το C5 είναι ένα LR one-vs-all με l2 regularization πάνω από τα χαρακτηριστικά.*
  - *Το C6 είναι ένα σύνολο 2ου στρώματος LR, εκπαιδευμένο στις προβλέψεις από τα C1, C2 και C3.*
  - *CM είναι ο κύριος ταξινομητής του συνόλου, ένα SVM με l2 κανονικοποίηση.*



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Data Overview

- Λίγα σύνολα δεδομένων, όλα με μη ισορροπημένη φύση.
- Για τους σκοπούς της παρούσας διπλωματικής εργασίας, ***Davidson et al.*** 2017 χρησιμοποιήθηκε σύνολο δεδομένων.
  - Τα δεδομένα έχουν τρεις τύπους: μίσος, προσβλητικό και μη μίσος.
- Τα δεδομένα σχολιάστηκαν από ειδικούς (3+ άτομα) με συμφωνία intercoder 92%.
- Σημαντική προκατάληψη προς τις τάξεις που δεν μισούν
  - *Hate* **6%,**
  - *Offensive* **77%**
  - *Non-hate* **17%**



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Evaluation Metrics

- Τυποποιημένες μετρήσεις αξιολόγησης ταξινόμησης ακρίβειας, ανάκλησης και F1.
- **Χρησιμοποιήθηκαν προσαρμογές μικρομέσου όρου των μετρήσεων λόγω της μη ισορροπημένης φύσης των δεδομένων**.
- Οι υπάρχουσες μελέτες σχετικά με την ανίχνευση ρητορικής μίσους ανέφεραν κυρίως τα αποτελέσματά τους χρησιμοποιώντας micro-Precision, micro-Recall και micro-F1 - Mask πραγματική απόδοση μειονοτικών τάξεων.
- **Ισορροπημένη ακρίβεια - προσαρμογή της ακρίβειας για μη ισορροπημένα δεδομένα.**
- **Περιοχή κάτω από την καμπύλη (AUC)**

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Setup

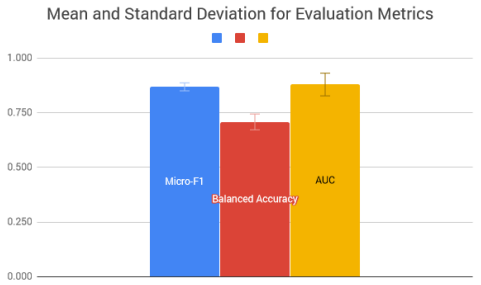
- Τα μοντέλα υλοποιήθηκαν σε Python Keras με backend Tensorflow, καθώς και σε Scikit-Learn.
- **Jupyter Notebooks https://jupyter.org/ χρησιμοποιήθηκαν για κάθε προσέγγιση για λόγους αναγνωσιμότητας και χρηστικότητας**.
- VM Ubuntu 16.4 με 16 VCPU και 32GB RAM
- **Google's Colab** VM με 13GB RAM και ένα Tesla K80 GPU.
- Για την εκπαίδευση μοντέλων:
  - Για κάθε μοντέλο οι εποχές εκπαίδευσης καθορίστηκαν σε 100 με μίνι-παρτίδες 128.
  - Εισήχθη μηχανισμός πρόωρης διακοπής προκειμένου να σταματήσει η υπερφόρτωση των μοντέλων, σταματώντας την εκπαίδευση εάν η απώλεια επικύρωσης δεν μειωθεί για 10 συνεχόμενες εποχές.
  - Όλα τα πειράματα εκτελέστηκαν σε 10πλάσια στρωματοποιημένη διασταυρούμενη επικύρωση.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Overall Results

- Κείμενο RNN επιτυγχάνει το καλύτερο F1 του 0,898 και AUC του 0,924
- Το CHASE επιτυγχάνει την καλύτερη βαθμολογία ακρίβειας 0,802
- Τα παρεμβαλλόμενα μοντέλα είχαν χειρότερες επιδόσεις από αυτά που εκπαιδεύτηκαν στο σύνολό τους. Αυτό οφείλεται σε:
  - Ακραία ανισορροπία του συνόλου δεδομένων.
  - Ποικιλομορφία υποδειγμάτων και ποσοστών σύγκλισης.
- Οι μέθοδοι συνόλου τοποθετούνται κοντά στην κορυφή με ακρίβεια 0,890 F1, 0,765 και 0,917 AUC.
- Αξιολόγηση ανά τάξη λόγω μικρών διαφορών στην αξιολόγηση.

Model	F1 Score	Accuracy	AUC
CharCNN [52]	0.866	0.703	0.900
CHASE [78]	0.858	**0.802**	0.893
HybridCNN (Interleaved) [52]	0.846	0.791	0.884
HybridCNN [52]	0.856	0.791	0.892
SVM [17]	0.869	0.709	0.902
GBDT with LSTM embeddings [3]	0.884	0.655	0.913
MANDOLA Hybrid Character-Metadata	0.866	0.719	0.900
MANDOLA Hybrid (Interleaved) Character-Metadata	0.882	0.717	0.912
MANDOLA Hybrid Word-Character	0.883	0.703	0.912
MANDOLA Hybrid (Interleaved) Word-Character	0.837	0.617	0.878
MANDOLA Hybrid Word-Metadata	0.871	0.731	0.903
MANDOLA Hybrid (Interleaved) Word-Metadata	0.837	0.628	0.878
MANDOLA Hybrid Word-Character-Metadata	0.879	0.739	0.909
MANDOLA Hybrid (Interleaved) Word-Character-Metadata	0.858	0.742	0.821
MANDOLA Three-layer Stacked Ensemble	0.890	0.769	0.917
MANDOLA Two-layer Stacked Ensemble	0.889	0.765	0.917
Metadata DNN [22]	0.814	0.632	0.861
Text RNN [22]	**0.898**	0.743	**0.924**
Unified NN (Interleaved) [22]	0.897	0.745	0.923
Unified NN [22]	0.867	0.709	0.900
WordCNN [52]	0.862	0.709	0.897
Logistic Regression [47]	0.882	0.730	0.911



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
of Action
Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Per-Class Results

- Παρά το γεγονός ότι τα συνολικά αποτελέσματα είναι υψηλά, τα αποτελέσματα ανά τάξη δείχνουν ότι οι βαθμολογίες ρητορικής μίσους είναι κακές.
- Αυτό οφείλεται στην ακραία ανισορροπία του συνόλου δεδομένων.
  - Οι συνολικές μετρήσεις αξιολόγησης αποκρύπτουν την απόδοση της μειονοτικής τάξης.
- Οι προσεγγίσεις μας για το σύνολο αποδίδουν καλύτερα στην ταξινόμηση ομιλίας μίσους με βαθμολογία F1 0,445 (3 επιπέδων) και 0,436 (2 επιπέδων), ενώ το CHASE ακολουθεί με 0,403.

F1 Scores			
Model	**Hate**	**Offensive**	**Neither**
CharCNN [52]	0.323	0.920	0.815
CHASE [78]	0.403	0.910	0.867
HybridCNN (Interleaved) [52]	0.393	0.907	0.833
HybridCNN [52]	0.400	0.910	0.853
SVM [17]	0.345	0.922	0.824
GBDT with LSTM embeddings [3]	0.310	0.930	0.810
MANDOLA Hybrid Character-Metadata	0.356	0.920	0.822
MANDOLA Hybrid (Interleaved) Character-Metadata	0.362	0.930	0.836
MANDOLA Hybrid Word-Character	0.344	0.932	0.830
MANDOLA Hybrid (Interleaved) Word-Character	0.275	0.903	0.699
MANDOLA Hybrid Word-Metadata	0.364	0.928	0.836
MANDOLA Hybrid (Interleaved) Word-Metadata	0.268	0.902	0.710
MANDOLA Hybrid Word-Character-Metadata	0.365	0.921	0.826
MANDOLA Hybrid (Interleaved) Word-Character-Metadata	0.380	0.910	0.820
MANDOLA Three-layer Stacked Ensemble	**0.445**	0.934	0.856
MANDOLA Two-layer Stacked Ensemble	0.436	0.934	0.853
Metadata DNN [22]	0.272	0.888	0.687
Text RNN [22]	0.401	**0.939**	0.878
Unified NN (Interleaved) [22]	0.397	0.938	**0.882**
Unified NN [22]	0.333	0.921	0.835
WordCNN [52]	0.369	0.916	0.838
Logistic Regression [47]	0.387	0.930	0.827

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Experiments and Evaluations Per-Class Results

- Τα μοντέλα σε επίπεδο λέξεων που εκπαιδεύονται με παρεμβολές, επιτυγχάνουν χειρότερα αποτελέσματα από τα παραδοσιακά εκπαιδευμένα.
- Αυτό οφείλεται στη μεγαλύτερη διάσταση του χώρου χαρακτηριστικών (μήκος λεξιλογίου).
- Τα αποτελέσματα δικαιολογούν επίσης ότι οι προσεγγίσεις μας επιτυγχάνουν τα καλύτερα αποτελέσματα τόσο στην ακρίβεια όσο και στην ανάκληση.

Precision Scores			
Model	**Hate**	**Offensive**	**Neither**
MANDOLA Hybrid Character-Metadata	0.330	0.936	0.788
MANDOLA Hybrid (Interleaved) Character-Metadata	0.394	0.936	0.796
MANDOLA Hybrid Word-Character	0.395	0.932	0.805
MANDOLA Hybrid (Interleaved) Word-Character	0.303	0.889	0.440
MANDOLA Hybrid Word-Metadata	0.385	0.928	0.795
MANDOLA Hybrid (Interleaved) Word-Metadata	0.360	0.902	0.670
MANDOLA Hybrid Word-Character-Metadata	0.343	0.939	0.786
MANDOLA Hybrid (Interleaved) Word-Character-Metadata	0.300	0.940	0.800
MANDOLA Three-layer Stacked Ensemble	**0.413**	**0.953**	0.808
MANDOLA Two-layer Stacked Ensemble	0.407	0.952	**0.809**

Recall Scores			
Model	**Hate**	**Offensive**	**Neither**
MANDOLA Hybrid Character-Metadata	0.394	0.902	0.858
MANDOLA Hybrid (Interleaved) Character-Metadata	0.350	0.924	0.876
MANDOLA Hybrid Word-Character	0.318	0.928	0.861
MANDOLA Hybrid (Interleaved) Word-Character	0.264	0.916	0.673
MANDOLA Hybrid Word-Metadata	0.359	0.916	0.883
MANDOLA Hybrid (Interleaved) Word-Metadata	0.226	0.900	0.760
MANDOLA Hybrid Word-Character-Metadata	0.395	0.904	0.871
MANDOLA Hybrid (Interleaved) Word-Character-Metadata	**0.510**	0.890	0.830
MANDOLA Three-layer Stacked Ensemble	0.483	0.916	**0.910**
MANDOLA Two-layer Stacked Ensemble	0.472	**0.952**	0.809

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Περίληψη

- Τα τελευταία χρόνια, μια δυσοίωνη εικόνα σχετικά με τη ρητορική μίσους στο διαδίκτυο έχει αρχίσει να υλοποιείται εντός των OSN (ανωνυμία, κινητικότητα).
- Οι προσπάθειες αντιμετώπισης της ρητορικής μίσους στο διαδίκτυο βασίζονται στην αυτοματοποιημένη σημασιολογική ανάλυση
- Κρίσιμο έργο: Ο εντοπισμός και η ταξινόμηση της ρητορικής μίσους με βάση διακριτά γλωσσικά χαρακτηριστικά.
- Παρουσιάσαμε το Σύστημα Επεξεργασίας Μεγάλων Δεδομένων MANDOLA, με έμφαση στην ενότητα ταξινόμησης ρητορικής μίσους.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Συμπεράσματα

Αυτή η διατριβή έχει δύο σημαντικές συνεισφορές:

1. Η μεγαλύτερη σύγκριση μοντέλων τελευταίας τεχνολογίας για την ανίχνευση ρητορικής μίσους στο διαδίκτυο, προκειμένου να βρεθούν τα καλύτερα μοντέλα σε επίπεδο λέξεων, χαρακτήρων και μεταδεδομένων.
    - ❏ Δώδεκα έργα τελευταίας τεχνολογίας υλοποιήθηκαν εκ νέου και αξιολογήθηκαν ανά τάξη.
    - ❏ Αυτό είχε ως αποτέλεσμα την έκθεση των μεροληπτικών βαθμολογιών αξιολόγησης λόγω της ανισορροπίας των συνόλων δεδομένων.
2. Προτεινόμενες συνδυαστικές προσεγγίσεις των συμπληρωματικών μοντέλων σε επίπεδο χαρακτήρων, λέξεων και μεταδεδομένων.
    - ❏ Τα αποτελέσματα δείχνουν ότι ο ταξινομητής 3 επιπέδων stacked ensemble επιτυγχάνει τα καλύτερα αποτελέσματα ανά κατηγορία με βαθμολογία F1 0,445 με την επόμενη υπερσύγχρονη να είναι 0,403.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Thank You

MAI4CAREU

MANDOLA

This work has been produced with the financial support of the Rights, Equality and Citizenship (REC) Programme of the European Union under Grant Agreement:

JUST/2014/RRAC/AG/HATE/6652