

"Δημιουργία εφαρμογής παρουσίασης  
ενοποιημένων προφίλ χρηστών από  
διαφορετικά κοινωνικά δίκτυα  
(Twitter, Facebook) "

# Σύγχρονες Διαδικτυακές Εφαρμογές

Project 4ου Εξαμήνου  
Ακαδημαϊκό έτος 2020-2021

Γιάννης Αξαόπουλος – Βασίλης Παπαδάκης

---

# Περιγραφή προβλήματος

Η παρακολούθηση του ίδιου χρήστη σε διαφορετικά κοινωνικά δίκτυα αποτελεί μια δύσκολη διαδικασία με πολλαπλές αναζητήσεις, καθώς σε κάθε κοινωνικό δίκτυο αναρτάται διαφορετικό είδος υλικού. Για παράδειγμα το Facebook κατά κύριο λόγο περιέχει οπτικοακουστικό υλικό, σε αντίθεση με το twitter που συγκεντρώνει κατά πλειοψηφία άρθρα. Συνεπώς για να μπορέσει ένας χρήστης να έχει μια ολοκληρωμένη άποψη για ένα προφίλ θα πρέπει να περιηγηθεί στο κάθε ένα ξεχωριστά κοινωνικό του δίκτυο.

## Προσέγγιση προβλήματος

Η συλλογή δεδομένων που πραγματοποιήθηκε από το κοινωνικό δίκτυο Facebook, έγινε μέσω facebook scrapper<sup>1</sup>. Αντίστοιχα για την εφαρμογή του Twitter χρησιμοποιήθηκε το tweepy, το οποίο είναι ένα πακέτο σε γλώσσα Python που επικοινωνεί με το Twitter API. Για το Twitter συγκεκριμένα πήραμε άδεια από το API<sup>2</sup>, μας δόθηκαν credentials και access tokens (διαβάζονται από το αρχείο *twitter\_credentials.json*) και είχαμε περιθώριο αναζήτησης δεδομένων με βάση το χρόνο.

Στον τομέα του backend η γλώσσα που χρησιμοποιήθηκε ήταν η Python, ενώ το σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων ήταν η MySQL. Η Sqlite παρόλο που ήταν ενσωματωμένη στην γλώσσα, απορρίφθηκε λόγω του ότι η MySQL είναι πιο διαδεδομένη και έχει μεγαλύτερη επεξεργαστική ισχύ. Η ανάπτυξη που πραγματοποιήθηκε στο frontend και στο backend είναι τελείως ανεξάρτητη και μπορεί να λειτουργήσει το καθένα αυτόνομα. Μαζί με τα αρχεία του backend και του frontend στον φάκελο υπήρχε ένα αρχείο csv, που περιείχε τα usernames που είχε το κάθε προφίλ για Twitter και Facebook ταυτόχρονα. Συνεπώς αυτό το csv αρχείο

---

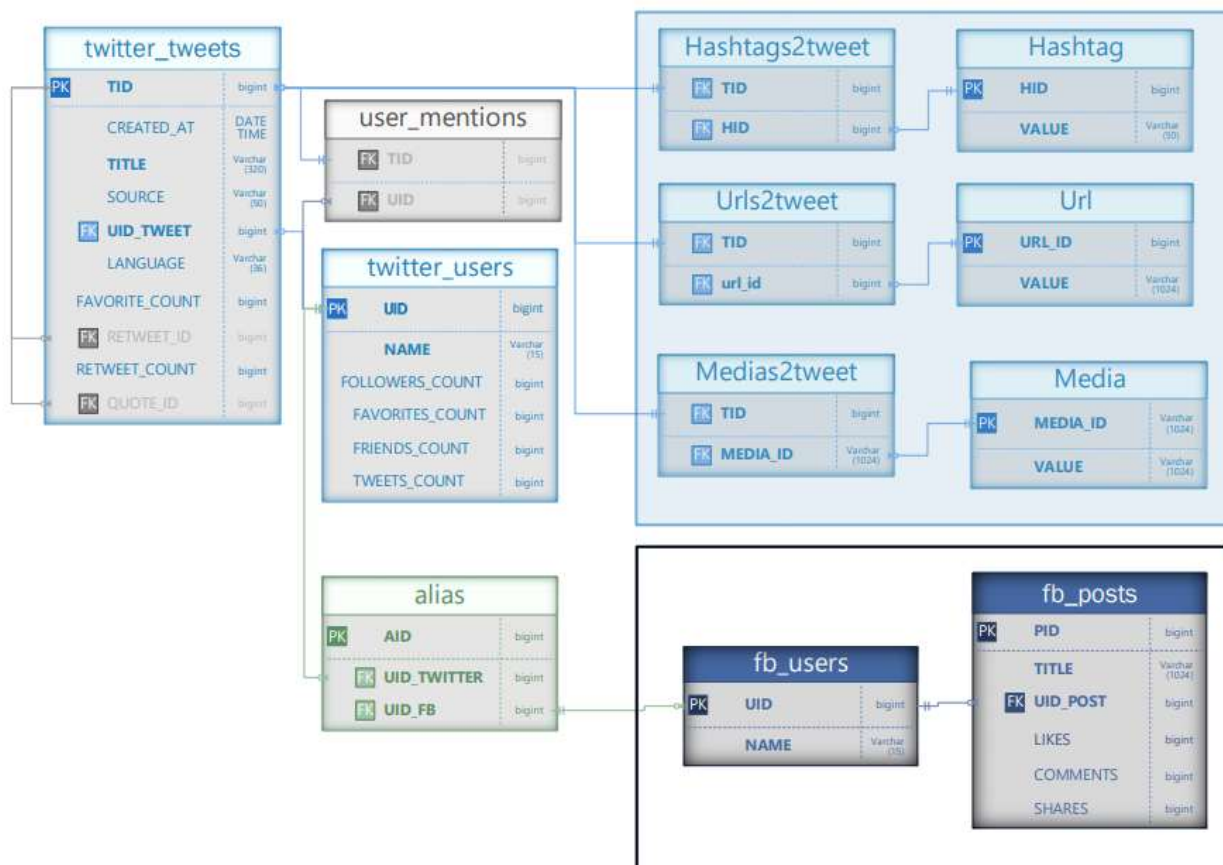
<sup>1</sup> [https://github.com/kevinzg/facebook-scraper/tree/master/facebook\\_scraper](https://github.com/kevinzg/facebook-scraper/tree/master/facebook_scraper)

<sup>2</sup> <https://developer.twitter.com/en>

μας καθορίζει από ποια προφίλ θα “τραβήξουμε” δεδομένα και είναι απαραίτητο για να λειτουργήσει το backend.

Η σύνδεση της βάσης δεδομένων MySQL με τον local host πραγματοποιήθηκε αφότου εισήγαμε username και password.

Η κωδικοποίηση που χρησιμοποιήσαμε για να είναι τα αρχεία μας ορατά με την γλώσσα που χρησιμοποιήσαμε ήταν η utf8. Παρακάτω εμφανίζεται το διάγραμμα σχεδίασης για την βάση δεδομένων (από τον συγκεκριμένο πίνακα δεν υλοποιήθηκαν οι πίνακες user\_mentions, Hashtags2tweet, Hashtag, Urls2tweet, Url, Media και Medias2tweet, καθώς και οι αυτοαναφορές RETWEET\_ID και QUOTE\_ID στον πίνακα twitter\_tweets).

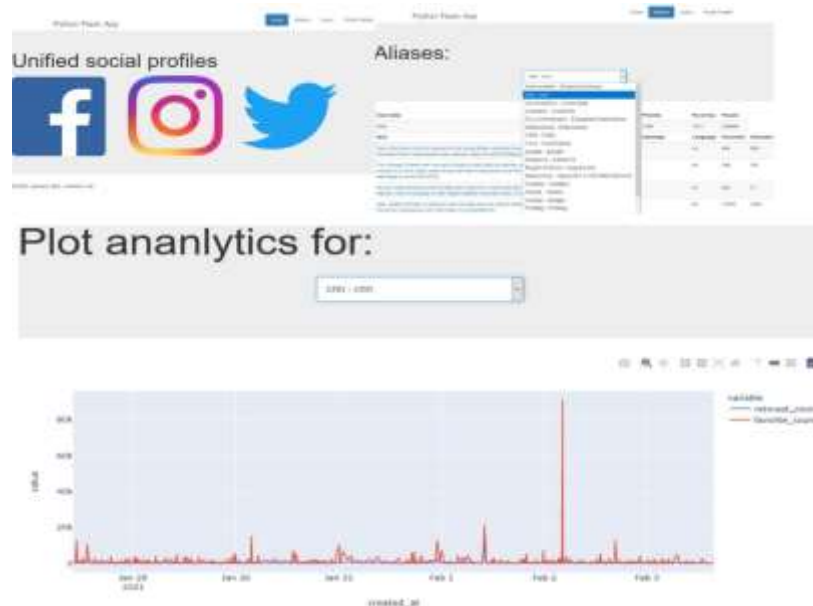


Το διάγραμμα αποτελείται από 3 διαφορετικά μέρη. Τα κεφάλαια Facebook, Twitter και η σύνδεση μεταξύ τους, που είναι ο πίνακας alias. Σε μελλοντική επέκταση της βάσης μπορούν να προστεθούν και άλλα δίκτυα.

Η βάση ονομάστηκε socialmedia\_db και συλλέγει πληροφορίες και από τα δύο κοινωνικά δίκτυα για 59 ισάριθμα κοινωνικά προφίλ. Τα δεδομένα που επιλέξαμε να εξάγουμε από το Facebook ήταν για τους χρήστες το : UserID, Name ενώ για τα posts: Postid, Userid\_post, Title, Likes, Comments, Shares, Time

Αντίστοιχα για το Twitter για τους χρήστες διαλέξαμε: UserID, Name, Favorites\_Count, Followers\_Count, Friends\_Count, Tweets\_Count ενώ για τα tweets είχαμε: TwitterID, UserID\_Tweet, Title, Retweet\_Count, Favorite\_Count, Created\_at, Language, Source και Hashtags.

Το κομμάτι του frontend είναι και αυτό στην γλώσσα Python με χρήση του Flask, χρησιμοποιήθηκε επίσης για την παρουσίαση στην ιστοσελίδα η βιβλιοθήκη Bootstrap και για τα γραφήματα το plot.ly (στην Python). Το frontend αποτελείται από τρεις βασικές σελίδες: την αρχική σελίδα, την σελίδα alias (που παρουσιάζει τα δεδομένα όπως είναι συγκεντρωμένα στην βάση) και την σελίδα των διαγραμμάτων-analytics.



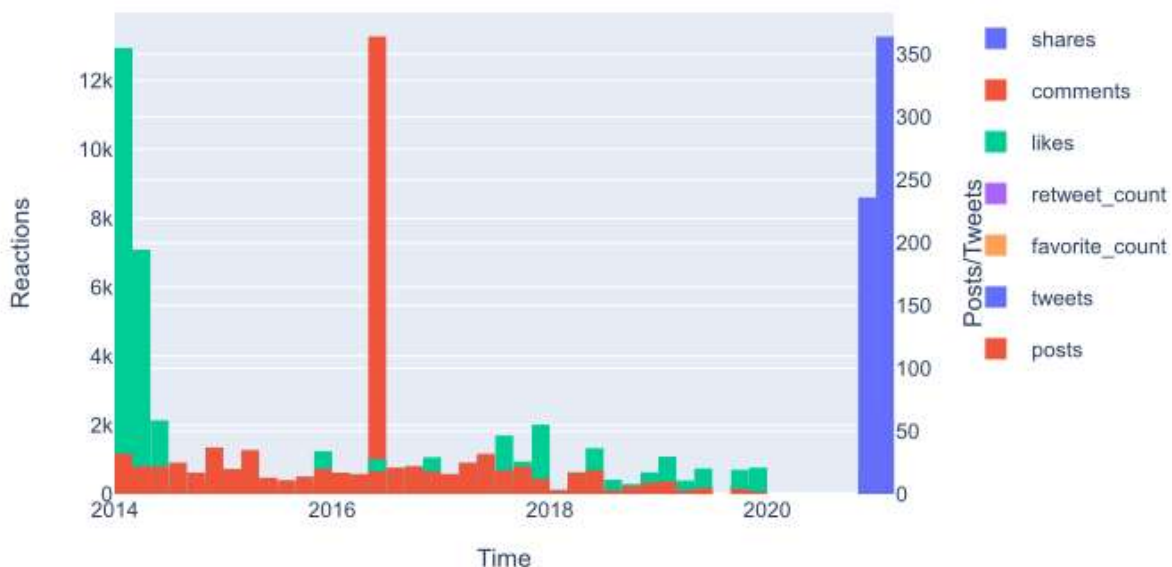
## Παρουσίαση αποτελεσμάτων

Όπως παρατηρούμε στον παρακάτω πίνακα της βάσης δεδομένων μας, τα δεδομένα που πήραμε ήταν για 53 προφίλ έκαστος σε Twitter και

Facebook. Τα posts του Facebook ήταν 25.312, ενώ τα tweets στο σύνολο ήταν 31.645. Να σημειωθεί ότι το όριο αναζήτησης για το κάθε προφίλ στο Facebook ήταν 150 σελίδες δημοσίευσης. Στο twitter τα items ήταν 600 για κάθε προφίλ. Ο χρόνος που απαιτήθηκε για συλλογή δεδομένων σε κάθε χρήστη ανά 600 post και 600 tweets ήταν 3 λεπτά.

Table	Rows	Type	Collation	Size
alias	44	InnoDB	utf8_general_mysql500_ci	48.0 KiB
fb_posts	25,312	InnoDB	utf8_general_mysql500_ci	14.0 MiB
fb_users	53	InnoDB	utf8_general_mysql500_ci	16.0 KiB
twitter_tweets	31,645	InnoDB	utf8_general_mysql500_ci	16.0 MiB
twitter_users	53	InnoDB	utf8_general_mysql500_ci	16.0 KiB
<b>5 tables</b>	<b>57,107</b>	<b>InnoDB</b>	<b>utf8_general_mysql500_ci</b>	<b>30.1 MiB</b>

Στη συνέχεια βλέπουμε ένα διάγραμμα που παρουσιάζονται με χρονική σειρά για τον τελευταίο μήνα posts, tweets, favorites\_count, retweets\_count, likes, comments, shares από ένα προφίλ για διάρκεια 6 ετών.



Τέλος βλέπουμε ένα πίνακα με τα πιο συχνά hashtags από το αντίστοιχο προφίλ. Τα hashtag είναι συνολικά από το Facebook και το Twitter για όλα τα δεδομένα που υπάρχουν στην βάση.



παρουσιάζεται στο διάγραμμα της βάσης στην ενότητα «Προσέγγιση προβλήματος».

Επίσης στο κομμάτι του frontend, περισσότερα στατιστικά μπορούν να παρουσιαστούν από την ανάλυση των δεδομένων που έχουν ήδη συλλεχθεί.