

Μελέτη Περίπτωσης: Εξόρυξη και ανάλυση ανοιχτών δεδομένων Twitter

Η παρούσα μελέτη πραγματοποιήθηκε από τις φοιτήτριες Τσαχειρίδου Μαρία και Τσελιγκάκη Δήμητρα υπό την επίβλεψη του καθηγητή Ταμπούρη Ευθύμιου, ως μέρος του μαθήματος Τεχνολογίες Ιστού και Ανάλυση Δεδομένων του προγράμματος μεταπτυχιακών σπουδών του τμήματος Εφαρμοσμένης Πληροφορικής του Πανεπιστημίου Μακεδονίας και ως σκοπό έχει να φέρει τους αναγνώστες πιο κοντά στην κατανόηση αλλά και στη χρήση ανοιχτών διασυνδεδεμένων δεδομένων (linkedopendata), τα οποία αναμφισβήτητα τα τελευταία χρόνια έχει ενταθεί σε ολοένα και περισσότερους κλάδους. Στα πλαίσια της εργασίας, χρησιμοποιήσαμε την υπηρεσία Influence Tracker και δεδομένα του Twitter τα οποία εξορύξαμε από αυτήν.

Η υπηρεσία η οποία ξεκίνησε ως μέρος μιας PhD έρευνας προτείνει μια οντολογία για τη σημασιοποίηση των λογαριασμών Twitter και των πληροφοριών που διαδίδονται καθώς και των αντίστοιχων οντοτήτων ως LinkedData, όπου οι χρήστες μπορούν να αναζητήσουν οποιοδήποτε λογαριασμό Twitter επιθυμούν, ανακαλύπτοντας μετρήσεις απευθείας από τις πληροφορίες που παρέχονται από το Twitter. Παράλληλα όμως, η υπηρεσία μας παρέχει την δυνατότητα να θέσουμε στοχευμένα ερωτήματα στο SparqlEndpoint ώστε να αντλήσουμε συγκεκριμένα δεδομένα τα οποία στη συνέχεια μπορούμε να τα χρησιμοποιήσουμε σε γραφικές απεικονίσεις στατιστικές αναλύσεις κλπ.

Από τη βάση αυτή, εμείς επιλέξαμε να αντλήσουμε δεδομένα για λογαριασμούς Twitter της παγκόσμιας και ελληνικής πολιτικής σκηνής, δεδομένου ότι η χρήση του twitter είναι ιδιαίτερα διαδεδομένη και προσφιλής στα πολιτικά πρόσωπα. Έτσι, οι λογαριασμοί τους οποίους μελετήσαμε μέσω της υπηρεσίας InfluenceTracker και για τους οποίους χρησιμοποιήσαμε δεδομένα σχετίζονται με πολιτικούς τόσο της Ελλάδας, όσο και του εξωτερικού χωρίς να σημαίνει όμως ότι το παρόν άρθρο υπηρετεί πολιτικές σκοπιμότητες, εφόσον η παρουσίαση αυτή γίνεται καθαρά για εκπαιδευτικούς σκοπούς. Αφού έγινε εξόρυξη των δεδομένων χρησιμοποιώντας αντίστοιχα sparql ερωτήματα, αναλύσαμε και οπτικοποιήσαμε τα δεδομένα χρησιμοποιώντας τα λογισμικά tableau και exploratory.

Χαρακτηριστικά παραδείγματα αποτελούν το ερώτημα που θέσαμε σχετικά τα accounts που έχουν χρησιμοποιήσει το hashtag #Greece στα tweets τους, με τα δεδομένα του οποίου κατασκευάστηκε ένα γράφημα με μπάρες. Τα άτομα-χρήστες αυτά ανήκουν στο πολιτικό χώρο. Παράλληλα βλέπουμε τον αριθμό των followers που έχουν προκειμένου να κατανοήσουμε και σχηματικά την «δύναμη» των προσώπων – οργανισμών στα κοινωνικά δίκτυα με βάση τον αριθμό αυτό. Τα ονόματα των προσώπων-οργανισμών δίνονται σε μορφή ονόματος account ώστε να μπορούν εύκολα να αναζητηθούν στο Twitter για περαιτέρω πληροφορίες. Ακόμη από τα δεδομένα που αντλήσαμε όταν αναζητήσαμε τους 100 λογαριασμούς twitter οι οποίοι ακολουθούν τον Donald Trump και αντίστοιχα 100 λογαριασμούς που ακολουθούν τον Barack Obama πήραμε δύο γραφήματα τύπου bubble graph σε καθένα από τα οποία οι followers ομαδοποιούνται σε 3 σύνολα βάσει του ποσοστού επιρροής που κατέχουν.

Συνολικά για την παρούσα έρευνα, τρέξαμε 8 διαφορετικά ερωτήματα τύπου sparql από τα οποία οπτικοποιήσαμε τα 6. Τα 4 πρώτα ερωτήματα αφορούσαν πολιτικούς της Ελλάδας, ενώ τα 4 επόμενα ήταν σχετικά με τους πολιτικούς Donald Trump και Barack Obama. Τα αποτελέσματα τα οποία λάβαμε είναι ενδεικτικά των δυνατοτήτων των ανοιχτών δεδομένων. Οι αξιοποιήσεις των ανοιχτών δεδομένων είναι πολύ περισσότερες, όμως ένα από τα ζητήματα που θα πρέπει να αντιμετωπιστεί πρώτα, είναι αποδοχή της πρόκλησης απελευθέρωσης και ελεύθερης χρήσης των δεδομένων από κρατικούς οργανισμούς, δημόσιους φορείς και ιδιωτικές επιχειρήσεις.