

5 Μαρτίου 2017

Τεχνολογία εντοπισμού κακόβουλων σχολίων από τη Google

/ [Επιστήμες, Τέχνες & Πολιτισμός](#)





REUTERS/NEIL HALL

Η συγκεκριμένη τεχνολογία αποκαλείται Perspective και αυτό που κάνει είναι να εξετάζει σχόλια και να τα βαθμολογεί ανάλογα με το πόσο παρόμοια είναι με σχόλια που έχουν χαρακτηριστεί «τοξικά» από ανθρώπους, ή που τους έχουν κάνει να εγκαταλείψουν μια συζήτηση.

Η Google και η θυγατρική Jigsaw λάνσαραν την Πέμπτη μια νέα τεχνολογία που θα βοηθά ειδησεογραφικούς οργανισμούς και online πλατφόρμες να εντοπίζουν κακόβουλα και προσβλητικά σχόλια στις ιστοσελίδες τους.

Η συγκεκριμένη τεχνολογία αποκαλείται Perspective και αυτό που κάνει είναι να εξετάζει σχόλια και να τα βαθμολογεί ανάλογα με το πόσο παρόμοια είναι με σχόλια που έχουν χαρακτηριστεί «τοξικά» από ανθρώπους, ή που τους έχουν κάνει να εγκαταλείψουν μια συζήτηση.

Όπως αναφέρει το Reuters, έχει δοκιμαστεί στους New York Times και οι εταιρείες ελπίζουν να την επεκτείνουν και σε άλλα μέσα (The Guardian, The Economist κ.α.).

«Οι ειδησεογραφικοί οργανισμοί θέλουν να ενθαρρύνουν την αλληλεπίδραση και τη συζήτηση γύρω από το περιεχόμενό τους, αλλά διαπιστώνουν πως ο εντοπισμός ανάμεσα σε εκατομμύρια σχόλια αυτών που είναι “trolling” ή προσβλητικά απαιτεί πολύ χρόνο, εργασία και χρήμα.

Ως αποτέλεσμα, πολλά sites έχουν κλείσει εντελώς τα σχόλια» έγραψε ο Τζάρεντ Κόεν, πρόεδρος της Jigsaw, που υπάγεται στην Alphabet, σε blogpost. «Αλλά μας λένε πως αυτή δεν είναι η λύση που θέλουν. Πιστεύουμε πως η τεχνολογία μπορεί να βοηθήσει».

Η Perspective εξέτασε εκατοντάδες χιλιάδες σχόλια που είχαν επισημανθεί ως προσβλητικά από ανθρώπους εξεταστές, ώστε να «εκπαιδευτεί» έτσι στο να αναγνωρίζει την προσβλητική γλώσσα.

Ο Σι Τζέι Άνταμς, product manager της εταιρείας, είπε πως η εταιρεία ήταν ανοιχτή στο να παρουσιάσει την τεχνολογία σε όλες τις πλατφόρμες, περιλαμβανομένων μεγαλύτερων όπως το Facebook και το Twitter, όπου το «τρολάρισμα» μπορεί να εξελιχθεί σε σημαντικό πρόβλημα.

Κατά τον Κόεν, η συγκεκριμένη τεχνολογία θα μπορούσε στο μέλλον να επεκταθεί για την ταυτοποίηση των προσωπικών επιθέσεων ή των off-topic σχολίων.

Όσον αφορά στο τι γίνεται με τα σχόλια που υποδεικνύονται, δεν θα αποφασίζει το Perspective, αλλά θα προωθούνται στους moderators ή θα αναπτυχθούν εργαλεία που θα «βοηθούν τους σχολιαστές να αντιληφθούν τη σημασία αυτών που γράφουν».

Όπως σημείωσε ο Κόεν, ένα σημαντικό μέρος των προσβλητικών σχολίων προέρχεται από ανθρώπους που «είχαν απλά άσχημη ημέρα». Ωστόσο, όπως επεσήμανε, η τεχνολογία Perspective είναι ακόμα σε πρώιμο στάδιο, με πολλά περιθώρια βελτίωσης.

Πηγή: naftemporiki.gr