Τμήμα Μηχανικών Η/Υ & Πληροφορικής - Σεμινάρια 2021-2022

ΔΙΑΛΕΞΗ

" Natural Language Processing: Focus on detecting toxic language in online conversations "

Ιωάννης Παυλόπουλος

Λέκτορας

Department of Computer and System Sciences

Πανεπιστήμιο της Στοκχόλμης

Θα μεταδοθεί διαδικτυακά μέσω MS Teams

Link MsTeams

Περίληψη – Abstract

User generated posts play a central role in social media and online discussion fora. News portals and blogs often also allow their readers to comment in order to get feedback, engage their readers, and build customer loyalty. User posts, however, and more generally User Generated Content (UGC) can be abusive (e.g., bullying, profanity, hate speech). Social media are increasingly under pressure to combat abusive content. News portals also suffer from abusive UGC, which damage their reputation and make them liable to fines, e.g., when hosting posts encouraging illegal actions. They often employ moderators, who are frequently overwhelmed by the volume of posts. Systems that detect abusive language are used to promote healthy conversations online and protect minority voices. Apart from a growing volume of press articles concerning toxicity online, there is increased research interest on detecting abusive and other unwelcome comments labeled 'toxic' by moderators, both for English and other languages. Despite this growing interest for the field of toxicity detection, three weakness hinder its further progress. First, information about the author is often being disregarded by systems. Second, systems focus on classifying a whole post as toxic or not, when human moderators could be assisted more if spans of that post (that made the system classify it as toxic) were automatically highlighted. Third and foremost, both toxicity systems and datasets still completely disregard the conversational context. With this talk I will discuss recent work addressing these three weaknesses.

Ο Ιωάννης Παυλόπουλος είναι λέκτορας στο Πανεπιστήμιο της Στοκχόλμης (Department of Computer and System Sciences), επισκέπτης καθηγητής στο Οικονομικό Πανεπιστήμιο Αθηνών (Τμήμα Πληροφορικής) και επισκέπτης ερευνητής στο Πανεπιστήμιο Ritsumeikan (Kinugasa, 2021) και στο Πανεπιστήμιο της Βενετίας (Ca' Foscari, 2022). Η έρευνά του εστιάζεται στη Μηχανική Μάθηση για την Επεξεργασία Φυσικής Γλώσσας (ΕΦΓ) και εφαρμογές της ΕΦΓ για την Υγεία, την Εκπαίδευση και τις Ανθρωπιστικές Επιστήμες. Κάποιες από τις εφαρμογές αυτές αφορούν εντοπισμό και μετρίαση ανεπιθύμητου λόγου, ανάλυση συναισθήματος, αυτόματη παραγωγή κειμένου από ιατρικές εικόνες, αυτόματη παραγωγή βαθμών εξετάσεων, και αυτόματη διόρθωση γραμματικών λαθών.

Πέμπτη 30/09/2021 – 12:30 – 13:30

Η διάλεξη θα γίνει διαδικτυακά