Ονοματεπώνυμο (Ελληνικά): Ευάγγελος Συμεών Μπιστολάκης

Ονοματεπώνυμο (Αγγλικά): Evangelos Symeon Bistolakis

Όνομα πατρός: Σπυρίδων

Τίτλος της εργασίας στην ελληνική γλώσσα: Πολύγλωσση Αφαιρετική Περίληψη Κειμένου

Τίτλος της εργασίας στην αγγλική γλώσσα: Multilingual Abstractive Text Summarization

Επιβλέπων Καθηγητής: : Γρηγόριος Τσουμάκας

Λέξεις κλειδιά (Ελληνικά)

1. Αφαιρετική Περίληψη 2. Πολύγλωσση Περίληψη 3. Βαθιά Μάθηση

Λέξεις κλειδιά (Αγγλικά)

1. Abstractive Summarization 2. Multilingual Summarization 3. Deep Learning

Περίληψη στην ελληνική γλώσσα (έως 500 λέξεις)

Η επεξεργασία φυσικής γλώσσας έχει αναπτυχθεί πολύ στην σύγχρονη εποχή των μεγάλων δεδομένων. Ένας από τους μεγαλύτερους στόχους της επιστημονικής κοινότητας είναι η  
δημιουργία αφαιρετικών περιλήψεων από τις μηχανές. Σε αυτό τον στόχο συμπεριλαμβάνεται και η δημιουργία μοντέλων τα οποία θα είναι ικανά να παράγουν περιλήψεις σε πολλές  
γλώσσες. Η παρούσα πτυχιακή χρησιμοποιεί πολύγλωσσα μοντέλα, όπως το mT5, για την  
παραγωγή αφαιρετικών περιλήψεων με χρήση συναφών γλωσσών κατά την εκπαίδευση. Ειδικότερα, φαίνεται ότι μοντέλα σαν αυτά, μπορούν να γενικεύσουν την ικανότητα που έχουν  
στις περιλήψεις, σε πολλές γλώσσες, ακόμη και εάν έχουν λίγα δείγματα στην εκπαίδευση.  
Η χρήση συναφών γλωσσών στην εκπαίδευση δείχνει να βοηθά τις επιμέρους γλώσσες με  
καλύτερης ποιότητας περιλήψεις σε σύγκριση με μοντέλα που έχουν εκπαιδευτεί σε μια  
γλώσσα. Επίσης, αυτή η μέθοδος φαίνεται να είναι μια καλή εναλλακτική όταν δεν υπάρχει μεγάλο πλήθος δεδομένων από μια γλώσσα, αφού η ικανότητα περίληψης φαίνεται να  
μεταφέρεται πιο εύκολα από μια γλώσσα σε άλλη όταν αυτές είναι συγγενικές. Η συνολική  
έρευνα δείχνει ότι η χρήση συναφών γλωσσών για την εκπαίδευση πολύγλωσσων μοντέλων  
είναι πιο αποδοτική στον ίδιο ή και σε λιγότερο χρόνο εκπαίδευσης, με τις περιλήψεις να  
είναι καλύτερες από την εκπαίδευση μοντέλων σε κάθε γλώσσα ξεχωριστά. Προκύπτει ότι  
είναι προτιμότερο να χρησιμοποιούνται πολύγλωσσα μοντέλα από την επιστημονική κοινότητα αφού είναι πιο αποδοτικά και πιο εύχρηστα ταυτόχρονα.

Περίληψη στην αγγλική γλώσσα (έως 500 λέξεις)

Natural language processing has grown a lot in the modern age of Big Data. One of the major goals of the scientific community is the creation of abstract summaries from machines.  
This goal also includes the creation of models that will be capable of producing summaries  
in multiple languages. This thesis uses multilingual models such as mT5 to generate abstract  
summaries using related languages during fine-tuning. In particular, it seems that models  
like these can generalize their ability to summarize in many languages, even if they have few  
samples in fine-tuning. The use of related languages seems to help individual languages with  
improved quality in summaries compared to models witch are trained with one language  
data. Also, this method seems to be a good alternative when there is not enough amount  
of data from a language, since the ability to summarize seems to be transferred more easily  
from one language to another when these are related. Overall research shows that using  
related languages to train multilingual models is more efficient, because in less training time  
the summaries are improved compared to models which where trained in each language  
separately. As a conclusion, it is preferable by the scientific community to use multilingual  
models since they are more efficient and useful as well.