



ΕΛΙΔΕΚ.
Ελληνικό Ίδρυμα Έρευνας & Καινοτομίας

Περιγραφή Χρηματοδοτούμενου Ερευνητικού Έργου
1η Προκήρυξη Ερευνητικών Έργων ΕΛ.ΙΔ.Ε.Κ. για την
ενίσχυση των Μελών ΔΕΠ και Ερευνητών/τριών και την
προμήθεια ερευνητικού εξοπλισμού μεγάλης αξίας

Τίτλος Ερευνητικού Έργου:

Προηγμένες Τεχνικές Μάθησης Πολλαπλών Ετικετών

Επιστημονικός/ή Υπεύθυνος/η:

Γρηγόριος Τσουμάκας

Φιλικός προς τον αναγνώστη τίτλος:

AMULET

Επιστημονική Περιοχή:

Μαθηματικά και Επιστήμες Πληροφορίας

Φορέας Προέλευσης και Χώρα:

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Ελλάδα

Φορέας Υποδοχής:

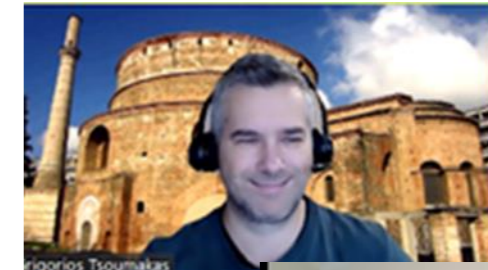
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Συνεργαζόμενος Φορέας:

Κανέννας

Ιστοσελίδα προβολής του Έργου

(αν υφίσταται): <https://amulet.csd.auth.gr>



Ποσό Χρηματοδότησης: 158,270.00 €

Διάρκεια Χρηματοδότησης: 36 Μήνες

Σύνοψη Ερευνητικού Έργου

Τα δεδομένα πολλαπλών ετικετών αποτελούνται από παραδείγματα τα οποία σχετίζονται με έναν πίνακα δυαδικών τιμών ο οποίος παίζει τον ρόλο της ετικέτας. Την τελευταία δεκαετία, το αντικείμενο της μάθησης από δεδομένα πολλαπλών ετικετών έχει γνωρίσει μία μεγάλη πρόοδο, πράγμα που γίνεται φανερό από τον αυξημένο αριθμό σχετικών εργασιών σε μερικά από τα σημαντικότερα συνέδρια του επιστημονικού πεδίου της Μηχανικής Μάθησης όπως τα KDD, AAAI και ICML. Παρόλη αυτή την έρευνα στο τρέχον αντικείμενο, υπάρχουν διάφορα προβλήματα που προκύπτουν με την εφαρμογή των τεχνικών αυτών σε πρακτικές εφαρμογές. Στόχος του AMULET είναι η δημιουργία καινούργιων τεχνικών πολλαπλών ετικετών που θα γεφυρώσουν την απόσταση θεωρίας και πράξης για την αντιμετώπιση δύο περίπλοκων προβλημάτων της τελευταίας κατηγορίας: i) την εννοιολογική εξέλιξη, ii) την επεξήγηση των αποτελεσμάτων των τεχνικών. Πιο συγκεκριμένα, για το πρώτο πρόβλημα που καλείται να μελετηθεί από το AMULET, σκοπεύουμε να αντιμετωπίσουμε την εννοιολογική εξέλιξη του χώρου των ετικετών σε περίπλοκες Βιοϊατρικές ιεραρχίες. Οι αλλαγές αυτές μπορεί να είναι απλές αλλαγές όπως ή εμφάνιση μίας νέας ετικέτας, αλλά και περίπλοκες όπως είναι η διαίρεση ή ένωση των ετικετών καθώς και η αλλαγή στο νόημα κάποιων ήδη υπαρχόντων ετικετών. Για το δεύτερο πρόβλημα έχουμε σκοπό στην δημιουργία τεχνικών για την επεξήγηση των αποτελεσμάτων τέτοιων αλγορίθμων με έμφαση στα δεδομένα κειμένου. Οι τεχνικές αυτές πρέπει να είναι σε θέση ένα εξηγούν τα αποτελέσματα βάσει κατανοητών αναπαραστάσεων όπως είναι οι λέξεις/φράσεις κλειδιά.

Πρωτοτυπία του Ερευνητικού Έργου

Οι υπάρχουσες εφαρμογές για Κατηγοριοποίηση/Ταξινόμηση δεδομένων πολλαπλών ετικετών σε ροές δεδομένων δεν λαμβάνουν υπόψιν τους τις αλλαγές που συχνά συμβαίνουν σε πραγματικά δεδομένα όπως στην περίπτωση της MeSH (Medical Subject Headings). Στο AMULET, θα μελετήσουμε τεχνικές οι οποίες αντιμετωπίζουν ρητά το πρόβλημα εμφάνισης νέων ετικετών καθώς και τις περισσότερο περίπλοκες αλλαγές που μπορεί να συμβούν στις ετικέτες, ενώ ταυτόχρονα θα μπορούν να εντοπίζουν και να προσαρμόζονται στην εννοιολογική εξέλιξη των δεδομένων. Βασική πρωτοτυπία του συγκεκριμένου ερευνητικού έργου είναι πως σκοπεύουμε να χρησιμοποιήσουμε σαφή πληροφορία σε συνδυασμό με τις συνήθεις στατιστικές τεχνικές για τον εντοπισμό της εννοιολογικής εξέλιξης. Αυτό θα γίνει με την ανάλυση των αλλαγών στην περίπτωση της MeSH που ανακοινώνει κάθε χρόνο η National Library of Medicine. Επιπλέον, στα πλαίσια του AMULET στοχεύουμε στο να επεξηγήσουμε της αποφάσεις των αλγορίθμων κατηγοριοποίησης πολλαπλών ετικετών πάνω στο συγκεκριμένο πρόβλημα του εντοπισμού ρητορικής μίσους στα σχόλια της πλατφόρμας του YOUTUBE, κάτι που από όσο γνωρίζουμε, δεν έχει διερευνηθεί ακόμη.

Αναμενόμενα αποτελέσματα & Αντίκτυπος του Ερευνητικού Έργου

Το αναμενόμενο αντίκτυπο των δράσεων που έχουν σχεδιαστεί στα πλαίσια του AMULET να λάβουν χώρα εντοπίζεται σε τρεις τομείς: την επιστήμη, την οικονομία και την κοινωνία.

Στον πρώτο τομέα, το AMULET στοχεύει στο να επιλύσει δύο βασικά ερευνητικά προβλήματα στην περιοχή της κατηγοριοποίησης δεδομένων πολλαπλών ετικετών, κάτι που αναμένεται να ξεκινήσει νέες ερευνητικές δουλειές από άλλους επιστήμονες. Η ροή δεδομένων πολλαπλών ετικετών είναι ένα πολύ ενδιαφέρον αντικείμενο, ειδικά για τις εφαρμογές του Διαδικτύου των Πραγμάτων, όπου μεγάλος όγκος τέτοιων δεδομένων παράγεται συνεχώς από τους διάφορους αισθητήρες. Η εννοιολογική εξέλιξη είναι ένα υπαρκτό φαινόμενο στις πραγματικές εφαρμογές και η επεξήγηση των αποτελεσμάτων από αλγορίθμους πολλαπλών ετικετών είναι απαραίτητη σε διάφορα πεδία εφαρμογής (π.χ. τραπεζικές συναλλαγές, υγεία).

Στον τομέα της οικονομίας, σκοπεύουμε να κάνουμε την μάθηση από δεδομένα πολλαπλών ετικετών περισσότερο προσβάσιμη στην βιομηχανία. Όσον αφορά την βιομηχανία εκδόσεων ακαδημαϊκών εντύπων, η αντιμετώπιση της εννοιολογικής εξέλιξης θα αυξήσει την ακρίβεια των τεχνικών αυτόματης κατηγοριοποίησης των εγγραφών, μειώνοντας το χειρωνακτικό φόρτο από τη πλευρά των ανθρώπων. Αντίστοιχα οφέλη αναμένεται να παρατηρηθούν και σε άλλες βιομηχανίες με μεγάλο αριθμό δεδομένων (π.χ. ενημέρωση και ειδήσεις). Σε εφαρμογές από το πεδίο της οικονομίας ή της υγειονομικές αποφάσεις, η καλύτερη κατανόηση των προβλέψεων θα οδηγήσει σε μικρότερο αριθμό από λανθασμένες αποφάσεις.

Στον τρίτο τομέα, στοχεύουμε στην βελτίωση των τεχνικών αυτόματης κατηγοριοποίησης αντιμετωπίζοντας το πρόβλημα της εννοιολογικής εξέλιξης των δεδομένων. Αυτές οι βελτιώσεις βοηθούν στην καταπολέμηση του προβλήματος της υπερβολικής διαθέσιμης πληροφορίας που αντιμετωπίζουν οι ερευνητές. Επιπλέον η ρητορική μίσους στο διαδίκτυο έχει πολλές αρνητικές επιπτώσεις στα άτομα τα οποία την υφίστανται. Στα πλαίσια του AMULET, θα σχεδιαστούν τεχνικές που θα βοηθήσουν στην καταπολέμηση του φαινομένου αυτού.

Η σημασία της χρηματοδότησης

Η χρηματοδότηση από το φορέα του ΕΛ.Ι.Δ.ΕΚ. παίζει πολύ σημαντικό ρόλο στην πραγματοποίηση αυτής της έρευνας. Δεν είναι μυστικό πως η εκάστοτε πηγή χρηματοδότησης αποτελεί το «καύσιμο» που «κινεί» μια ερευνητική ομάδα. Συγκεκριμένα, βοηθάει τα μέλη της ομάδας να μπορούν να αφιερώσουν τον απαραίτητο χρόνο στην τριβή τους με το αντικείμενο της προδιαγεγραμμένης έρευνας. Σε περίπτωση που δεν υπήρχε αυτή, τα μέλη της ομάδας θα έπρεπε να αναζητήσουν κάποιον διαφορετικό τρόπο ικανοποίησης/κάλυψης των βιοτικών τους αναγκών, γεγονός που δεν θα τους επέτρεπε να μπορούν να είναι ενήμεροι με τις νέες τάσεις/εξελίξεις και τους μηχανισμούς μάθησης που εμφανίζονται συνεχώς στο εξεταζόμενο επιστημονικό πεδίο, καθώς και να συμβάλλουν ουσιαστικά σε ένα τόσο ανταγωνιστικό περιβάλλον. Τέλος η χρηματοδότηση είναι απαραίτητη για την απόκτηση των υπολογιστικών πόρων που απαιτούνται για τα πειράματα καθώς και για την παρακολούθηση αξιόπαινων συνεδρίων στο εσωτερικό αλλά και το εξωτερικό.



ΕΛΙΔΕΚ.
Ελληνικό Ίδρυμα Έρευνας & Καινοτομίας

ΕΠΙΚΟΙΝΩΝΙΑ

Λ. Συγγρού 185 & Σάρδεων 2
ΤΚ. 17121, Νέα Σμύρνη, Ελλάδα
210 64 12 410, 420
communication@elidek.gr
www.elidek.gr