



ΕΛΙΔΕΚ.
Ελληνικό Ίδρυμα Έρευνας & Καινοτομίας

Περιγραφή Χρηματοδοτούμενου Ερευνητικού Έργου
2η Προκήρυξη Ερευνητικών Έργων ΕΛ.ΙΔ.Ε.Κ. για την
ενίσχυση Μεταδιδακτορικών Ερευνητών/τριών



Τίτλος Ερευνητικού Έργου: iQARuS – Αποδοτική αποτίμηση ερωτήσεων πάνω από διασυνδεδεμένα δεδομένα ροής με χρήση συνόψεων

Επιστημονικός/ή Υπεύθυνος/η: Χαρίδημος Κονδυλάκης

Φιλικός προς τον αναγνώστη τίτλος: iQARuS

Επιστημονική Περιοχή: Μαθηματικά και Επιστήμες της Πληροφορίας

Φορέας Προέλευσης και Χώρα: ΙΔΡΥΜΑ
ΤΕΧΝΟΛΟΓΙΑΣ & ΕΡΕΥΝΑΣ (ΙΤΕ), ΕΛΛΑΔΑ

Φορέας Υποδοχής: ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ & ΕΡΕΥΝΑΣ (ΙΤΕ)
Συνεργαζόμενος Φορέας: -

Ιστοσελίδα προβολής του Έργου:
<http://www.ics.forth.gr/~kondylak/iQARuS>



Ποσό Χρηματοδότησης: 130.000€

Διάρκεια Χρηματοδότησης: 3 έτη

Σύνοψη

Μαζικές ροές δεδομένων πλέον φτάνουν συνεχώς, από το Ίντερνετ των πραγμάτων Devices (IoT), τους αισθητήρες και τα μέσα κοινωνικής δικτύωσης, και απαιτούν γρήγορη επεξεργασία και αναζήτηση πληροφορίας καθώς και την άμεση συσχέτιση της με περαιτέρω γνώση προκειμένου να υποστηριχθούν ροές εργασιών ανάλυσης. Σε αυτή την κατεύθυνση, πλατφόρμες επεξεργασίας ροών δεδομένων RDF αποτελούν πολύτιμα εργαλεία, επιτρέποντας την γρήγορη απάντηση ερωτήσεων. Ωστόσο, μέχρι στιγμής, οι τεχνολογίες αιχμής στο πεδίο αυτό παρέχουν είτε μηχανές που βασίζονται σε ένα κεντρικό υπολογιστικό κόμβο είτε κατακεντρωμένες μηχανές που ωστόσο προσφέρουν περιορισμένες δυνατότητες συλλογιστικής και συνδυασμού με προγενέστερη πληροφορία. Από την άλλη πλευρά, τεχνικές δημιουργίας συνόψεων έχουν ήδη αποδείξει την αξία τους για δεικτοδότηση, απάντηση επερωτήσεων, συλλογιστική, επιλογή της κατάλληλης πηγής δεδομένων, οπτικοποίησης και ανακάλυψης του σχήματος. Ωστόσο, μέχρι σήμερα τέτοιες τεχνικές δεν έχουν χρησιμοποιηθεί για RDF ροές δεδομένων πράγμα που παραμένει μια εντελώς ανεξερεύνητη περιοχή. Το αντικείμενο λοιπόν της τρέχουσας πρότασης, με το ακρωνύμιο iQARuS, είναι η αποτελεσματική και αποδοτική απάντηση επερωτήσεων πάνω από ροές δεδομένων RDF χρησιμοποιώντας συνόψεις. Οι συνόψεις που θα δημιουργούνται θα είναι μικρότερες από τους αρχικούς γράφους δεδομένων, μειώνοντας έτσι δραστικά το χώρο των δεδομένων και επιτρέποντας την αποτελεσματική και αποδοτική αποτίμηση επερωτήσεων και συλλογισμούς πάνω σε αυτά. Ωστόσο εξαιτίας αυτής ακριβώς της μείωσης του χώρου των δεδομένων υπάρχει η περίπτωση να μην είναι δυνατή η πλήρης αποτίμηση των επερωτήσεων από τις συνόψεις. Σε αυτές τις περιπτώσεις σκοπεύουμε να διερευνήσουμε προσεγγιστικές απαντήσεις σε ερωτήματα και στη συνέχεια να προσφέρουμε τελεστές εξερεύνησης που θα επιτρέπουν την επέκταση των συνόψεων, για την πλήρη αποτίμηση των επερωτήσεων. Επιπλέον, αυξητικοί αλγόριθμοι θα επιτρέψουν την σταδιακή ενημέρωση των συνόψεων αποφεύγοντας το κόστος του επαναυπολογισμού τους από την αρχή. Η λύση που θα αναπτυχθεί θα καλύπτει και θα συνδυάζει τόσο δεδομένα πάνω από παράθυρα ροών δεδομένων όσο και προϋπάρχουσα γνώση. Επιπλέον η λύση θα αξιολογηθεί εκτενώς, χρησιμοποιώντας τόσο καθιερωμένα σύνολα αξιολόγησης ροών δεδομένων RDF, όσο και ένα νέο που θα δημιουργηθεί κατά την διάρκεια του έργου.

Πρωτοτυπία του Ερευνητικού Έργου

Μέχρι στιγμής, οι προσεγγίσεις αιχμής για την επεξεργασία ροών δεδομένων RDF έχουν επικεντρωθεί σε συνεχείς γλώσσες επερωτήσεων και έχουν παράσχει είτε κεντριοποιημένες μηχανές που δεν μπορούν να αντιμετωπίσουν ροές δεδομένων RDF μεγάλου όγκου, είτε καταναεμημένες δομές που προσφέρουν περιορισμένες δυνατότητες συλλογιστικής.

Από την άλλη, τεχνικές δημιουργίας συνόψεων, έχουν ήδη αποδείξει την αξία τους για την ευρετηρίαση δεδομένων, την απάντηση ερωτημάτων, την εκτίμηση του μεγέθους των αποτελεσμάτων, την επιλογή πηγής, την οπτικοποίηση και την ανακάλυψη του σχήματος μιας πηγής. Ωστόσο, από όσο γνωρίζουμε, δεν έχουν ακόμη αξιοποιηθεί για ροές δεδομένων, πράγμα που παραμένει μια εντελώς ανεξερεύνητη περιοχή.

Οι κύριες πρωτότυπες συνιστώσες του έργου iQARuS είναι οι εξής:

- 1. Θα αναπτυχθούν νέες τεχνικές σύνοψης για δεδομένα ροών, αξιοποιώντας τεχνικές περιλήψεων σημασιολογικών γράφων, διερευνώντας προσεγγίσεις τόσο για την δημιουργία αντιπροσωπευτικών κόμβων, όσο και για την επιλογή υφιστάμενων κόμβων.*
- 2. Οι συνόψεις που θα δημιουργηθούν θα χρησιμοποιηθούν για την απάντηση ερωτημάτων, μειώνοντας δραστικά το μέγεθος των δεδομένων προς επεξεργασία, επιτρέποντας την αποτελεσματική απάντηση ερωτημάτων και τη συλλογιστική πάνω από αυτά. Λόγω των ταχέως μεταβαλλόμενων πληροφοριών θα αξιοποιηθούν τόσο οι κατά προσέγγιση όσο και οι ακριβείς απαντήσεις για την αποτίμηση επερωτήσεων, πάνω από τις υφιστάμενες περιλήψεις.*
- 3. Επιπλέον θα δημιουργηθούν αλγόριθμοι σταδιακής ενημέρωσης για τις περιλήψεις, διασφαλίζοντας ότι οι περιλήψεις που δημιουργούνται έχουν πάντα τις απαραίτητες πληροφορίες που ενδιαφέρουν.*
- 4. Τέλος θα δημιουργηθούν αποτελεσματικοί αλγόριθμοι για την απάντηση ερωτημάτων που απαιτούν το συνδυασμό των δυναμικών δεδομένων ροής με ιστορικά στατικά δεδομένα. Αυτό θα οδηγήσει σε μια ενοποιημένη πλατφόρμα που θα επιτρέψει τη συλλογιστική πάνω στα αποτελέσματα των επερωτήσεων, επιτρέποντας περαιτέρω εργασίες ανάλυσης.*

Από όσα γνωρίζουμε, το πεδίο της επεξεργασίας ροών δεδομένων χρησιμοποιώντας περιλήψεις δεν έχει ακόμη εξερευνηθεί και έχει τη δυνατότητα να προσφέρει απτά αποτελέσματα.

Αναμενόμενα αποτελέσματα & Αντίκτυπος του Ερευνητικού Έργου

Η αποτελεσματική και αποδοτική επεξεργασία ροών δεδομένων έχει υψηλό επιστημονικό και κοινωνικό αντίκτυπο. Έχει τη δυνατότητα να επιτρέψει την ταχεία ανίχνευση καταστροφικών γεγονότων, πρόγνωσης καιρού, για χρήση στην εξόρυξη γνώμης, ανίχνευση παράτυπων συμβάντων για την υγεία, εκδηλώσεις μολυσματικών ασθενειών, ανίχνευση απάτης κ.α. που βασίζονται στην επεξεργασία ταχέως μεταβαλλόμενων ροών δεδομένων.

Καθώς η πλατφόρμα που θα αναπτυχθεί θα είναι ανεξάρτητη από το πεδίο εφαρμογής, θα μπορεί να επωφεληθεί ποικίλες επιστημονικές και κοινωνικές προσπάθειες που απαιτούν ταχεία επεξεργασία μεγάλου όγκου ροών δεδομένων. Ήδη έχουν επιλεγεί σενάρια αξιολόγησης που σχετίζονται με έξυπνες εφαρμογές πόλεων, ηλεκτρονικό εμπόριο και κοινωνικά δίκτυα και εντός του έργου μας, σκοπεύουμε επίσης να αναζητήσουμε ενδιαφέροντα σενάρια εκμετάλλευσης για ροές δεδομένων ροής δεδομένων υγείας. Εν κατακλείδι, η πρότασή μας έχει τη δυνατότητα για πολύ υψηλό επιστημονικό και κοινωνικό αντίκτυπο.

Η σημασία της χρηματοδότησης

Ήδη πολλές μελέτες έχουν αναλύσει τον αντίκτυπο της χρηματοδότησης της έρευνας και έχουν αναδείξει τον μεγάλο αντίκτυπο στην παραγωγικότητα και την απόδοση των χρηματοδοτούμενων ερευνητών όσον αφορά την ποσότητα και την ποιότητα των δημοσιεύσεών τους και στις πιθανές επιστημονικές συνεργασίες.

Η έρευνα που προτείνεται σε αυτό το έργο δεν θα είχε προχωρήσει διαφορετικά, καθώς οι προτεινόμενες ιδέες βρίσκονται στο στάδιο της σύλληψης και απαιτείται περαιτέρω έρευνα μέχρι να ωριμάσουν και να δημιουργήσουν πραγματικά οφέλη.



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