



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: Κλιμακώσιμη Απάντηση Ερωτήσεων Εκφρασμένων σε Φυσική Γλώσσα σε Μεγάλες Γεωγραφικές Βάσεις Γνώσεων

Επιστημονικός Υπεύθυνος: Καθ. Μανόλης Κουμπάρκης

Φιλικός προς τον αναγνώστη τίτλος: GeoQA

Επιστημονική Περιοχή: Μαθηματικά και Επιστήμες Πληροφορίας

Φορέας Προέλευσης και Χώρα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Ελλάδα

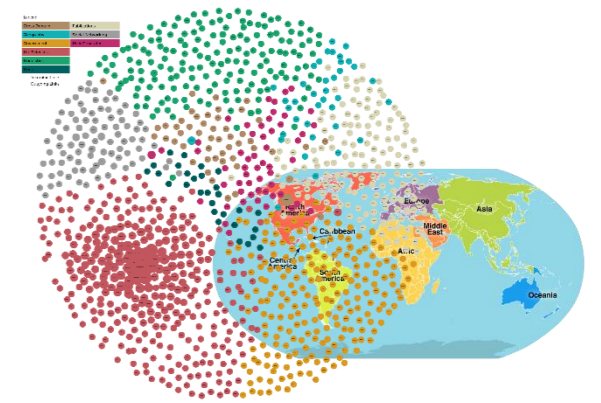
Φορέας Υποδοχής: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Συνεργαζόμενος Φορέας: Ινστιτούτο Πληροφορικής Max Planck

**Ιστοσελίδα προβολής του Έργου
(αν υφίσταται):** <http://geoqa.di.uoa.gr/>

Ποσό Χρηματοδότησης: 200.000 €

Διάρκεια Χρηματοδότησης: 36 μήνες



Σύνοψη Ερευνητικού Έργου

Ο στόχος του GeoQA είναι διττός:

1. Η μελέτη της επέκτασης γράφων γνώσης με γεωγραφική γνώση, όπως αυτή εμπεριέχεται σε σημαντικά γεωχωρικά σύνολα δεδομένων που διατίθενται στον Ιστό, και
2. Η ανάπτυξη αποτελεσματικών (ως προς την ακρίβεια και την ανάκληση) και αποδοτικών (ως προς τον χρόνο) τεχνικών και συστημάτων για την απάντηση πολύπλοκων ερωτήσεων πάνω σε γεωχωρικούς γράφους γνώσης.

Στο ερευνητικό έργο αυτό:

- Κατασκευάζουμε τον *γράφο γνώσης YAGO2geo*, επεκτείνοντας τον δημοφιλή γράφο γνώσης YAGO2 με γεωγραφικά δεδομένα από πλήθος αξιόπιστων συνόλων γεωχωρικών δεδομένων (π.χ. Global Administrative Areas)
- Αναπτύσσουμε αποδοτικές τεχνικές και, αντίστοιχα, *λογισμικό για τη συνεχή και αυτόματη ανανέωση του YAGO2geo* ακολουθώντας τις αντίστοιχες μεταβολές στα σύνολα δεδομένων.
- Δημιουργούμε ένα *πρότυπο σύνολο γεωγραφικών ερωτήσεων* (>1000 ερωτήσεις) σε φυσική γλώσσα με τις απαντήσεις τους.
- Αναπτύσσουμε την *μηχανή απάντησης ερωτήσεων GeoQA2*, η οποία θα είναι βασισμένη σε τεχνικές επεξεργασίας φυσικής γλώσσας καθώς και σε τεχνικές διανυσματικής αναπαράστασης των γράφων γνώσης. Η απαντήσεις θα είναι υπό την μορφή κειμένου αλλά και εικόνας για καλύτερη διαισθητική κατανόηση.

Πρωτοτυπία του Ερευνητικού Έργου

- Ο YAGO2geo θα είναι ο πιο πολυπληθής γράφος γνώσης ως προς τα γεωγραφικά δεδομένα για την Ευρωπαϊκή Ένωση και τις ΗΠΑ. Ενοποιεί δεδομένα από το YAGO2, τα γεωγραφικά δεδομένα των διοικητικών περιοχών όπως αυτά διατίθενται από επίσημες πηγές της Ελλάδας, της Μεγάλης Βρετανίας, την Δημοκρατία της Ιρλανδίας και τις ΗΠΑ, καθώς και από τα σύνολα δεδομένων Global Administrative Areas και OpenStreetMap.
- Πρώτη φορά μελετάται το πρόβλημα αποδοτικής ανανέωσης των γράφων γνώσης στην περιοχή των γεωχωρικών δεδομένων.
- Το πρότυπο σύνολο ερωταπαντήσεων που αναπτύσσουμε είναι το πρώτο το οποίο εστιάζει σε γεωχωρικά δεδομένα και θα αποτελέσει σημείο αναφοράς για μελλοντικές έρευνες στο πεδίο σε διεθνές επίπεδο.
- Η τεχνική διανυσματικής αναπαράστασης που αναπτύσσουμε κωδικοποιεί στον διανυσματικό χώρο για πρώτη φορά οντότητες, σχέσεις και λεκτικά τόσο πλούσια σε περιεχόμενο (π.χ. πολύγωνα) ενός γράφου γνώσης.
- Η μηχανή GeoQA2 θα είναι η πρώτη η μηχανή απάντησης ερωτήσεων σε διεθνές επίπεδο, η οποία θα απαντά πολύπλοκες ερωτήσεις γεωγραφικού ενδιαφέροντος. Δημοφιλείς μηχανές αναζήτησης, όπως είναι η Google, αδυνατούν να απαντήσουν πολύπλοκες ερωτήσεις της μορφής: «Ποιο είναι το μήκος του ποταμού που διασχίζει την Λάρισα, που βρίσκεται στην Ελλάδα;».
- Θα βελτιστοποιήσουμε την απόδοσή της GeoQA2 επιλύοντας το πρόβλημα βελτιστοποίησης πολλαπλών ερωτήσεων για την γλώσσα επερωτήσεων GeoSPARQL, για το οποίο δεν υπάρχει αντίστοιχη έρευνα στη βιβλιογραφία.

Αναμενόμενα αποτελέσματα & Αντίκτυπος του Ερευνητικού Έργου

Αναμενόμενα αποτελέσματα:

1. Ο μεγαλύτερος γράφος γνώσης για γεωχωρικά δεδομένα YAGO2geo.
2. Σύστημα αυτόματης ανανέωσης μεγάλων γράφων γνώσης.
3. Σύνολο ερωταπαντήσεων για γεωχωρικά δεδομένα σε φυσική γλώσσα.
4. Τεχνικές αναπαράστασης γράφων γνώσης με γεωχωρικά δεδομένα στον διανυσματικό χώρο.
5. Αποδοτικό και αποτελεσματικό σύστημα απάντησης ερωτήσεων γεωχωρικού ενδιαφέροντος εκφρασμένων σε φυσική γλώσσα.

Ο επιστημονικός αντίκτυπος:

- Ο YAGO2geo θα είναι ο πρώτος γράφος γνώσης γεωγραφικών δεδομένων τέτοιας έκτασης.
- Υπάρχει μόνο ένα πιλοτικό σύστημα απάντησης ερωτήσεων γεωχωρικού ενδιαφέροντος εκφρασμένων σε φυσική γλώσσα (GeoQA), το οποίο έχει αναπτυχθεί από την ερευνητική ομάδα του Καθ. Μ. Κουμπάρακη και αποτελεί σημείο έναρξης του έργου αυτού.
- Μόλις το σύνολο ερωταπαντήσεων και η μηχανή GeoQA2 αναπτυχθούν, θα αποτελέσουν σημείο αναφοράς διεθνώς σε αυτό το ερευνητικό πεδίο.

Ο οικονομικός αντίκτυπος:

- Θα χρηματοδοτηθεί η έρευνα μεταπτυχιακών και διδακτορικών φοιτητών.
- Θα επιτρέψει στην ομάδα του Καθ. Μ. Κουμπάρακη να αναπτύξει «έξυπνους βοηθούς» εξειδικευμένους στη γεωχωρική γνώση που θα μπορούν να αξιοποιηθούν εμπορικά από start-ups του Πανεπιστημίου Αθηνών.
- Ο YAGO2geo και η GeoQA2 μπορούν να αξιοποιηθούν από οποιαδήποτε Ευρωπαϊκή μηχανή αναζήτησης για την ανάπτυξη εξειδικευμένων εφαρμογών (π.χ. για τον τουρισμό).
- Η αξιοποίησή της GeoQA2 σε σύνολα δεδομένων από δορυφορικές εικόνες, τα οποία αποτελούν, ενδεχομένως, τις πιο σημαντικές ψηφιακές πηγές που είναι διαθέσιμες σήμερα στην ανθρωπότητα (π.χ., για την κλιματική αλλαγή).
- Συνεπώς, τα αποτελέσματα του ερευνητικού έργου θα έχουν πολύ σημαντικό οικονομικό αντίκτυπο στην Ελλάδα και στην Ευρώπη.

Η σημασία της χρηματοδότησης

Το παρόν έργο δίνει την ευκαιρία στον ΕΥ και την ομάδα του να αναπτύξουν καινοτόμα αποτελέσματα και εφαρμογές στη συγκεκριμένη ερευνητική περιοχή. Επίσης, δίνει τη δυνατότητα σε νέους ερευνητές να απασχοληθούν στη χώρα μας και να δυναμώσουν το βιογραφικό τους.



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