

ΑΝΑΖΗΤΗΣΗ ΣΤΟ GOOGLE SCHOLAR ΜΕ ΑΠΟΚΛΕΙΣΜΟ ΑΥΤΟΑΝΑΦΟΡΩΝ

ΠΛΗΡΟΦΟΡΙΕΣ: Γιώργος Γιαννόπουλος, 210 772 1402, giann@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η ανάπτυξη μίας εφαρμογής η οποία θα επεκτείνει και θα βελτιώνει τη λειτουργικότητα του Google scholar με έμφαση στην ακριβέστερη αναζήτηση αναφορών για δημοσιεύσεις.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοικτή – θα καθοριστεί σε συνεννόηση με τον υποψήφιο

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Το Google scholar (<http://scholar.google.com/>) είναι μία εξειδικευμένη έκδοση του Google που αφορά στην αναζήτηση ακαδημαϊκών κειμένων όπως δημοσιεύσεις σε συνέδρια, περιοδικά κ.τ.λ.. Προσφέρει επιλογές όπως αναζήτηση ανά συγγραφέα, έκδοση, ημερομηνία και ανά γλώσσα. Στα αποτελέσματα αναζήτησης εμφανίζονται, πέρα από τον τίτλο και το snippet του άρθρου, πληροφορίες όπως τα ονόματα των συγγραφέων, το συνέδριο ή το περιοδικό στο οποίο δημοσιεύτηκε. Επίσης, δίνεται η δυνατότητα εύρεσης όλων των εκδοχών του άρθρου, σχετικών άρθρων και εύρεσης άρθρων από τα οποία γίνεται αναφορά σε αυτό το άρθρο.

Οι αναφορές (citations) σε ένα άρθρο, δηλαδή ο αριθμός των άρθρων τα οποία το συμπεριλαμβάνουν στη βιβλιογραφική αναφορά τους είναι ένας σημαντικός δείκτης για την αξιολόγηση του άρθρου αλλά και συνολικά του επιστημονικού έργου ενός συγγραφέα. Το πρόβλημα με τη λειτουργία του google scholar είναι ότι για κάθε άρθρο-αποτέλεσμα, εμφανίζει το σύνολο των αναφορών που βρίσκει. Μέσα σε αυτές, όμως, τις περισσότερες φορές υπάρχουν αναφορές του ίδιου συγγραφέα, από διαφορετικό άρθρο του, στο συγκεκριμένο άρθρο (αυτοαναφορές). Επιπλέον, ενδέχεται να υπάρχουν αναφορές από στενούς συνεργάτες του συγγραφέα, προς το συγκεκριμένο άρθρο του συγγραφέα. Οι αυτοαναφορές, λοιπόν, αποτελούν αρνητικό παράγοντα για την αξιολόγηση ενός άρθρου, αφού ένας συγγραφέας ενδέχεται σε κάθε άρθρο του να κάνει αναφορά σε προηγούμενα δικά του, με σκοπό να διογκώσει τον αριθμό αναφορών.

Στόχος της διπλωματικής είναι η δημιουργία ενός μηχανισμού αποκλεισμού/διαχωρισμού των αυτοαναφορών και παρουσίας των διαχωρισμένων αποτελεσμάτων στο χρήστη. Για την υλοποίηση ενός τέτοιου μηχανισμού ο υποψήφιος θα αντιμετωπίσει θέματα σχετικά με:

1. Ολοκλήρωση διαφορετικών πηγών αποτελεσμάτων. Τα αποτελέσματα που επιστρέφονται από το google scholar ενδέχεται να είναι ελλιπή ή σε μη επιθυμητή μορφή. Σε αυτήν την περίπτωση, θα πρέπει να αναζητείται πληροφορία και σε άλλες πηγές, π.χ. ACM Portal.
2. String matching. Τα ονόματα των συγγραφέων μπορεί να μην είναι πάντα στην ίδια μορφή, για παράδειγμα το μικρό όνομα μπορεί να δίνεται ολόκληρο, ή μόνο με το αρχικό γράμμα. Το σύστημα θα πρέπει να ξεπερνάει τέτοια προβλήματα, είτε με string matching τεχνικές, είτε λαμβάνοντας υπόψιν το context του ονόματος, δηλαδή πληροφορία γύρω από αυτό.

Επιμέρους στόχοι της εργασίας θα είναι θέματα αποδοτικότερης οργάνωσης και παρουσίας των αποτελεσμάτων όπως αυτόματη κατηγοριοποίηση τους ανάλογα με τους συγγραφείς, τις ερευνητικές περιοχές στις οποίες ανήκουν κ.α.

ΣΧΕΤΙΚΕΣ ΑΝΑΦΟΡΕΣ:

<http://www.dbnet.ece.ntua.gr/pubs/uploads/DIPL-2007-9.pdf>

<http://www.harzing.com/pop.htm>