

## ΠΕΡΙΛΗΨΗ

Οι γενικές μηχανές αναζήτησης, όπως η Google και η Yahoo!, παρέχουν ένα εύκολο μηχανισμό για τους χρήστες τους για να βρίσκουν πληροφορίες στο Διαδίκτυο. Πέραν των φανερών πλεονκτημάτων τους όμως, έχουν ένα σημαντικό αριθμό περιορισμών, επειδή δεν μπορούν να προσεγγίσουν και να αναλύσουν ένα σημαντικό μέρος της πληροφορίας που είναι διαθέσιμη.

Τα Κατανεμημένα Συστήματα Αναζήτησης Πληροφοριών, κάνοντας χρήση αλγορίθμων συγχώνευσης συλλογών, παρέχουν μία λύση στο παραπάνω πρόβλημα, επιτρέποντας στους χρήστες τους να υποβάλλουν ερωτήματα συγχρόνως σε πολλαπλές πηγές πληροφόρησης, παρέχοντας μία πολύ μεγαλύτερη κάλυψη της διαθέσιμης πληροφορίας.

Αυτή η διατριβή ασχολείται με δύο από τα βασικά προβλήματα που αφορούν στον σχεδιασμό και στην υλοποίηση αποτελεσματικών και αποδοτικών Κατανεμημένων Συστημάτων Αναζήτησης Πληροφοριών: την *επιλογή πηγών* και την *σύνθεση αποτελεσμάτων*. Το πρώτο πρόβλημα ασχολείται με την ικανότητα του συστήματος να επιλέγει τις πιο κατάλληλες πηγές πληροφόρησης για να μεταβιβάσει το ερώτημα του χρήστη και το δεύτερο αποβλέπει στο να παράξει την καλύτερη δυνατή τελική λίστα κειμένων μέσω της σύνθεσης των επιμέρους ανακτημένων κειμένων από τις επιλεγμένες πηγές.

Οι νέοι αλγόριθμοι που παρουσιάζονται σε αυτή τη διατριβή έχουν σχεδιαστεί ώστε να λειτουργούν αποτελεσματικά σε περιβάλλοντα όπου οι πηγές δεν παρέχουν καμία συνεργασία, με αποτέλεσμα να είναι εφαρμόσιμη στην μεγαλύτερο δυνατό σύνολο περιβαλλόντων και συνθηκών. Ο αλγόριθμος επιλογής πηγών που προάγεται παρέχει έναν καινοτόμο τρόπο μοντελοποίησης των πηγών ως περιοχές σε ένα χώρο που παράγεται από τα κείμενα τα οποία περιέχουν. Διατυπώνει ένα πλήρες θεωρητικό πλαίσιο επίλυσης του προβλήματος της επιλογής πηγών, ενώ παράλληλα αποτελεσματικά συλλαμβάνει πειραματικές παρατηρήσεις και γενικά αποδεκτές αντιλήψεις στον τομέα της Ανάκτησης Πληροφοριών. Εκτεταμένα πειράματα επιδεικνύουν ότι είναι ικανός να διασφαλίσει μία απόδοση που είναι τουλάχιστον τόσο καλή όσο άλλες μεθοδολογίες αιχμής και συχνότερα καλύτερη.

Οι νέοι αλγόριθμοι σύνθεσης αποτελεσμάτων που παρουσιάζονται είναι βασισμένη στην υπόθεση ότι οι μηχανές αναζήτησης επιστρέφουν μονάχα κατατάξεις κειμένων, χωρίς σκορ σχετικότητας, ένα σενάριο που είναι συνήθης

πρακτική σε σύγχρονα συστήματα ανάκτησης πληροφοριών. Και οι δύο επιλύουν το πρόβλημα της έλλειψης πληροφόρησης πολύ αποτελεσματικά, επιδεικνύοντας σημαντικά οφέλη στην απόδοση συγκρίσει με άλλους αλγορίθμους αιχμής. Επιπροσθέτως, ο δεύτερος αλγόριθμος ενοποιεί τις δύο γενικές κατευθύνσεις από τις οποίες έχει προσεγγιστεί το πρόβλημα στην έρευνα, συνδυάζοντας τα πλεονεκτήματά τους, ενώ συγχρόνως ελαχιστοποιώντας τα μειονεκτήματά τους.