

Αναφορά αξιολόγησης για συστήματα επεξεργασίας και εξόρυξης κειμένου χρησιμοποιώντας την Ανάλυση SWOT

Δημήτρης Ρουσίδης, Πάνος Μπαλατσούκας, Εμμανουήλ Γαρουφάλλου

Περίληψη

Επί σειρά ετών έχει παρατηρηθεί η συνεχόμενα εκθετική αύξηση των ακαδημαϊκών άρθρων, γεγονός που έχει προκαλέσει αντιστοίχως μια διακαή ανάγκη για αποτελεσματικά εργαλεία λογισμικού επεξεργασίας κειμένου. Στην προσπάθεια ανταπόκρισης στην ανάγκη αυτή μια πληθώρα από εργαλεία εξόρυξης κειμένου δημιουργήθηκαν, κάποια γενικής χρήσεως και άλλα πιο εξειδικευμένα για συγκεκριμένους κλάδους, κυρίως για τις βιολογικές και ιατρικές επιστήμες. Από τα υπολογιστικά αυτά εργαλεία απαιτείται να επεξεργάζονται, να αναλύουν, να ερμηνεύουν, να ενσωματώνουν και να παράγουν πληροφορίες διαχειριζόμενα μεγάλες αποθήκες δεδομένων. Τα εργαλεία πρέπει να είναι λειτουργικά και να ανταποκρίνονται στις απαιτήσεις του επιστημονικού αλλά και του πραγματικού κόσμου. Σκοπός αυτού του άρθρου είναι η αναφορά αξιολόγησης (benchmark evaluation) ενός επιλεγμένου συνόλου από τα πιο επαγγελματικά, λειτουργικά, ανοικτού κώδικα (open source) εργαλεία λογισμικού (GATE, KNIME, Rapid Miner, κ.τ.λ.) τα οποία μέσω της επεξεργασίας κειμένου παράγουν πληροφορίες χρησιμοποιώντας τεχνικές εξόρυξης δεδομένων.

Στόχος αυτής της έρευνας είναι να προτείνει ένα κοινό πλαίσιο αξιολόγησης το οποίο να προσδιορίζει, να βοηθάει στη λήψη αποφάσεων και να αποτιμά τα παραπάνω συστήματα στο χώρο των ακαδημαϊκών βιβλιοθηκών. Για το σκοπό αυτής της έρευνας εφαρμόστηκε η τεχνική αξιολόγησης και ανάλυσης SWOT (Πλεονεκτήματα, Αδυναμίες, Ευκαιρίες και Απειλές) για να προσδιοριστούν ενδογενείς και εξωγενείς παράγοντες (όπως αποτελεσματικότητα αλγορίθμων, βάθος σημασιολογικής ανάλυσης, ποιότητα δεδομένων, επεκτασιμότητα, διαλειτουργικότητα με άλλα συστήματα, κ.α.) ως κριτήρια για μια «δοκιμή πρόσκρουσης» μεταξύ των σημαντικότερων εργαλείων.

Η ανάλυση αυτή θα χρησιμοποιηθεί ως εφαλτήριο και οδηγός κατανόησης και λήψης αποφάσεων για τη δημιουργία ενός πρότυπου εργαλείου εξόρυξης δεδομένων το οποίο θα εντοπίζει σημασιολογικές ομοιότητες μεταξύ ακαδημαϊκών συγγραμμάτων. Ταυτόχρονα θα αναγνωρίζει τις κύριες μεθοδολογίες έρευνας που χρησιμοποιήθηκαν και κύριος σκοπός του εργαλείου θα είναι η διευκόλυνση της ακαδημαϊκής μελέτης και έρευνας και η αύξηση της ικανότητας άντλησης πληροφοριών και συσχετισμών από τα ακαδημαϊκά περιεχόμενα.

Λέξεις Κλειδιά

Αναφορά αξιολόγησης; ανάλυση SWOT; εξόρυξη κειμένου; εξόρυξη γνώσης; συστήματα επεξεργασίας κειμένου