

**«Αξιοποίηση Οντολογιών και Χρήση Τεχνολογιών του Σημαντικού Ιστού για
Αναπαράσταση, Διαχείριση και Ανακάλυψη Γνώσης μέσω του Διαδικτύου»**

Διπλωματική Εργασία για το Μεταπτυχιακό Δίπλωμα Ειδίκευσης (ΜΔΕ)

Δημήτριος Κουτσομητρόπουλος

ΠΕΡΙΛΗΨΗ

Η έρευνα αναφορικά με την εξέλιξη του Παγκόσμιου Ιστού (WWW) κινείται τώρα προς πιο ευφυείς και αυτοματοποιημένους τρόπους ανακάλυψης και εξαγωγής πληροφορίας. Ο *Σημαντικός Ιστός* (Semantic Web) είναι μία επέκταση του σημερινού Ιστού, όπου στην πληροφορία δίνεται σαφώς προσδιορισμένη σημασία, δίνοντας έτσι τη δυνατότητα στις μηχανές να μπορούν πλέον να επεξεργάζονται καλύτερα και να «κατανοούν» τα δεδομένα, τα οποία σήμερα περιορίζονται απλώς να παρουσιάζουν. Για να λειτουργήσει ο Σημαντικός Ιστός, οι υπολογιστές θα πρέπει να έχουν πρόσβαση σε οργανωμένες συλλογές πληροφοριών που καλούνται *οντολογίες* (ontologies). Οι οντολογίες παρέχουν μια μέθοδο αναπαράστασης της γνώσης στο Σημαντικό Ιστό και μπορούν επομένως να αξιοποιηθούν από τα υπολογιστικά συστήματα για τη διεξαγωγή *αυτοματοποιημένου συλλογισμού* (automated reasoning).

Για την περιγραφή και αναπαράσταση των οντολογιών του Σημαντικού Ιστού σε γλώσσες αναγνώσιμες από τη μηχανή, έχουν προταθεί και βρίσκονται υπό εξέλιξη διάφορες πρωτοβουλίες, με πιο σημαντικές τη γλώσσα *DAML+OIL* (DARPA) και την πιο πρόσφατη *OWL* (W3C). Οι γλώσσες αυτές, μεταξύ άλλων, συγκρίνονται και αξιολογούνται συστηματικά, βάσει ενός διεξοδικού πλαισίου αξιολόγησης. Εξάλλου δίνονται μοντέλα και προσεγγίσεις που βασίζονται σε οντολογίες και αντιμετωπίζουν το ζήτημα της διαχείρισης της γνώσης, σε όλο τον κύκλο ζωής της.

Το θεωρητικό υπόβαθρο που εγγράφει την εκφραστική και συλλογιστική ισχύ των οντολογιών, παρέχεται από τις *Λογικές Περιγραφής* (Description Logics). Οι Λογικές Περιγραφής αποτελούν ένα καλώς ορισμένο αποφασίσιμο υποσύνολο της Λογικής Πρώτης Τάξης και εξηγείται πώς οι τυπικές ιδιότητές τους καθιστούν δυνατή την αναπαράσταση και ανακάλυψη γνώσης στο Σημαντικό Ιστό. Επιπλέον, για την OWL παρατίθενται πλέον πρόσφατα αποτελέσματα αναγωγής σε εκφραστικές Λογικές Περιγραφής, καθώς και αποτελέσματα πολυπλοκότητας του συλλογισμού.

Για την ανακάλυψη άρρηκτης πληροφορίας ενδείκνυται επομένως να αξιοποιηθούν συστήματα βασισμένα σε Λογικές Περιγραφής. Τα συστήματα αυτά δείχνεται ότι

πλεονεκτούν, τη δεδομένη στιγμή, έναντι εναλλακτικών προσεγγίσεων, όπως τα *προγράμματα Απόδειξης Θεωρημάτων* (Theorem Provers) και τα *Βασισμένα σε Κανόνες συστήματα* (Rule-Based systems). Έτσι παρουσιάζονται και αξιολογούνται, κριτικά και πειραματικά, τρεις κινήτρες συμπερασμού Λογικών Περιγραφής.

Το σύστημα RACER κρίθηκε ως το καταλληλότερο για να χρησιμοποιηθεί ως πυρήνας συλλογισμού από την πιλοτική *διεπαφή ανακάλυψης γνώσης*, η οποία αναπτύχθηκε για να επιδείξει τη δυνατότητα συμπερασμού νέων γεγονότων από OWL έγγραφα κατανεμημένα στον Παγκόσμιο Ιστό. Η διεπαφή ανακάλυψης γνώσης υλοποιεί μια εκφραστική μεθοδολογία διεξαγωγής συμπερασμών και υποστηρίζει τη διαισθητική και δηλωτική σύνθεση και υποβολή ευφώνων ερωτημάτων από την πλευρά του χρήστη.

Χρησιμοποιώντας τη διεπαφή, διεξάγεται μια σειρά πειραματικών συμπερασμών πάνω σε πολιτισμική γνώση, όπως αυτή αναπαρίσταται από την οντολογία CIDOC-CRM. Η οντολογία CIDOC-CRM ανήχθη σε OWL μορφή και εμπλουτίστηκε, με κατανεμημένο τρόπο, με συγκεκριμένα στιγμιότυπα και εκφραστικές δομές, γεγονός που κατέστη δυνατό ύστερα από αυτήν την αναγωγή.