

ΔΗΜΗΤΡΙΟΣ Α. ΚΟΥΤΣΟΜΗΤΡΟΠΟΥΛΟΣ

**«Μέθοδοι και Τεχνικές Ανακάλυψης Γνώσης στο Σημαντικό Ιστό: Παραγωγική Απόκτηση Γνώσης από Οντολογικά Έγγραφα και η Τεχνική της Σημασιακής Προσαρμογής»**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΠΕΡΙΛΗΨΗ

Ο Σημαντικός Ιστός (Semantic Web) είναι ένας συνδυασμός τεχνολογιών και προτύπων με σκοπό να προσδοθεί στη διαδικτυακή πληροφορία αυστηρά καθορισμένη σημασιακή δομή και ερμηνεία. Στόχος είναι να μπορούν οι χρήστες του Παγκόσμιου Ιστού καθώς και αυτοματοποιημένοι πράκτορες να επεξεργάζονται, να διαχειρίζονται και να αξιοποιούν την κατάλληλα χαρακτηρισμένη πληροφορία με τρόπο ευφυή και αποδοτικό.

Ωστόσο, παρά τις τεχνικές που έχουν κατά καιρούς προταθεί, δεν υπάρχει ξεκάθαρη μέθοδος ώστε, αξιοποιώντας το φάσμα του Σημαντικού Ιστού, η διαδικτυακή πληροφορία να ανακτάται με τρόπο παραγωγικό, δηλαδή με βάση τα ήδη εκπεφρασμένα γεγονότα να συνάγεται νέα, άρρητη πληροφορία.

Για την αντιμετώπιση αυτής της κατάστασης, αρχικά εισάγεται και προσδιορίζεται το πρόβλημα της *Ανακάλυψης Γνώσης στο Σημαντικό Ιστό* (Semantic Web Knowledge Discovery, SWKD). Η Ανακάλυψη Γνώσης στο Σημαντικό Ιστό εκμεταλλεύεται το σημασιακό υπόβαθρο και τις αντίστοιχες σημασιακές περιγραφές των πληροφοριών, όπως αυτές είναι θεμελιωμένες σε μια λογική θεωρία (οντολογίες εκφρασμένες σε γλώσσα OWL). Βάσει αυτών και με τη χρήση των κατάλληλων μηχανισμών αυτοματοποιημένου συλλογισμού μπορεί να συμπεραθεί νέα, άδηλη γνώση, η οποία, μέχρι τότε, μόνο υπονοούνταν στα ήδη υπάρχοντα δεδομένα.

Για να απαντηθεί το ερώτημα αν και σε πιο βαθμό οι τεχνολογίες και η λογική θεωρία του Σημαντικού Ιστού συνεισφέρουν *αποδοτικά* και *εκφραστικά* στο πρόβλημα της SWKD καταρτίζεται μια πρότυπη *Μεθοδολογία Ανακάλυψης Γνώσης στο Σημαντικό Ιστό*, η οποία θεμελιώνεται σε πρόσφατα θεωρητικά αποτελέσματα, αλλά και στην ποιοτική και πειραματική συγκριτική αξιολόγηση διαδεδομένων μηχανισμών συμπερασμού (inference engines) που βασίζονται σε Λογικές Περιγραφής (Description Logics). Η αποδοτικότητα και η εκφραστικότητα της μεθόδου αυτής δείχνεται ότι εξαρτώνται από συγκεκριμένους θεωρητικούς, οργανωτικούς και τεχνικούς περιορισμούς.

Η πειραματική επαλήθευση της μεθοδολογίας επιτυγχάνεται με την κατασκευή και επίδειξη της *Διεπαφής Ανακάλυψης Γνώσης* (Knowledge Discovery Interface) μιας κατανεμημένης δηλαδή δικτυακής υπηρεσίας, η οποία έχει εφαρμοστεί με επιτυχία σε πειραματικά δεδομένα. Τα αποτελέσματα που προκύπτουν με τη χρήση της διεπαφής επαληθεύουν, μέχρι ορισμένο βαθμό, τις υποθέσεις που έχουν γίνει σχετικά κυρίως με την παράμετρο της *εκφραστικότητας* και δίνουν το έναυσμα για την αναζήτηση και εξέταση της υποστήριξης των νέων προτεινόμενων επεκτάσεων της λογικής θεωρίας του Σημαντικού Ιστού, δηλαδή της γλώσσας OWL 1.1.

Για την ενίσχυση της εκφραστικότητας της ανακάλυψης γνώσης στην περίπτωση συγκεκριμένων πεδίων γνώσης (knowledge domains) εισάγεται μια νέα τεχνική, αποκαλούμενη *Σημασιακή Προσαρμογή*. Η τεχνική αυτή εξελίσσει την *Προσαρμογή*

*Μεταδεδομένων Εφαρμογής (Metadata Application Profiling)* από μια επίπεδη συρραφή και συγχώνευση σχημάτων και πεδίων μεταδεδομένων, σε μία ουσιαστική επέκταση και σημασιακή αναγωγή και εμπλουτισμό του αντίστοιχου μοντέλου στο οποίο εφαρμόζεται. Έτσι, η σημασιακή προσαρμογή εξειδικεύει ένα οντολογικό μοντέλο ως προς μια συγκεκριμένη εφαρμογή, όχι απλά με την προσθήκη λεξιλογίου από ετερογενή σχήματα, αλλά μέσω της σημασιακής εμβάθυνσης (semantic intension) και εκλέπτυνσης (semantic refinement) του αρχικού μοντέλου. Η τεχνική αυτή και τα αποτελέσματά της επαληθεύονται πειραματικά με την εφαρμογή στο μοντέλο πληροφοριών πολιτιστικής κληρονομιάς CIDOC-CRM και δείχνεται ότι, με τη χρήση κατάλληλων μεθόδων, η γενική εφαρμοσιμότητα του μοντέλου μπορεί να διαφυλαχθεί.

Για να μπορεί όμως η Ανακάλυψη Γνώσης στο Σημαντικό Ιστό να δώσει ικανοποιητικά αποτελέσματα, απαιτούνται όσο το δυνατόν πληρέστερες και αυξημένες περιγραφές των δικτυακών πόρων. Παρόλο που πληροφορίες άμεσα συμβατές με τη λογική θεωρία του Σημαντικού Ιστού δεν είναι ευχερείς, υπάρχει πληθώρα δεδομένων οργανωμένων σε επίπεδα σχήματα μεταδεδομένων (flat metadata schemata). Διερευνάται επομένως αν η SWKD μπορεί να εφαρμοστεί *αποδοτικά και εκφραστικά* στην περίπτωση τέτοιων *ημιδομημένων μοντέλων γνώσης*, όπως για παράδειγμα στην περίπτωση του σχήματος μεταδεδομένων Dublin Core. Δείχνεται ότι το πρόβλημα αυτό ανάγεται μερικώς στην εφαρμογή της σημασιακής προσαρμογής στην περίπτωση τέτοιων μοντέλων, ενώ για τη διαφύλαξη της διαλειτουργικότητας και την επίλυση αμφισημιών που προκύπτουν εφαρμόζονται ανάλογες μέθοδοι και επιπλέον εξετάζεται η τεχνική της *παρονομασίας* (running) που εισάγει η OWL 1.1, βάσει της οποίας ο ορισμός ενός ονόματος μπορεί να έχει κυμαινόμενη σημασιακή ερμηνεία ανάλογα με τα συμφραζόμενα.

Συμπερασματικά, οι νέες μέθοδοι που προτείνονται μπορούν να βελτιώσουν το πρόβλημα της Ανακάλυψης Γνώσης στο Σημαντικό Ιστό ως προς την εκφραστικότητα, ενώ ταυτόχρονα η πολυπλοκότητα παραμένει η μικρότερη δυνατή. Επιτυγχάνουν επίσης την παραγωγή πιο εκφραστικών περιγραφών από υπάρχοντα μεταδεδομένα, προτείνοντας έτσι μια λύση στο πρόβλημα της *εκκίνησης* (bootstrapping) για το Σημαντικό Ιστό. Παράλληλα, μπορούν να χρησιμοποιηθούν ως βάση για την υλοποίηση πιο αποδοτικών τεχνικών κατανεμημένου και αυξητικού συλλογισμού.