

Περίληψη

Στις επιστήμες της Πληροφορικής και της Πληροφόρησης, η χρήση Σχεσιακών Βάσεων Δεδομένων (Relational Databases) αποτελεί για πολλά χρόνια έναν από τους βασικούς τρόπους οργάνωσης και αποθήκευσης μεγάλων όγκων δεδομένων, με αποτέλεσμα να έχει υπάρξει έντονη ερευνητική δραστηριότητα, πολλών χρόνων, στην ανάπτυξη και υλοποίηση αλγορίθμων που προσφέρουν βελτιστοποίηση στο τρόπο επεξεργασίας των δεδομένων που βρίσκονται αποθηκευμένα σε αυτές, πάνω στους οποίους στηρίζονται όλα τα σύγχρονα Συστήματα Διαχείρισης Βάσεων Δεδομένων (Relational Database Managements Systems - RDBMS).

Η SQL (Structured Query language) αποτελεί τη βασική γλώσσα ερωτημάτων των σχεσιακών βάσεων δεδομένων, η χρήση της οποίας, όμως, προαπαιτεί από το χρήστη τη γνώση του σχήματος της βάσης δεδομένων στην οποία απευθύνει τα ερωτήματά του.

Η αλματώδης ανάπτυξη του Παγκόσμιου Ιστού (World Wide Web) στις μέρες μας, δημιουργεί μια μεγάλη πηγή γνώσης για τον άνθρωπο αφού το πλήθος των αποθηκευμένων δεδομένων που παρέχονται σε αυτόν, αλλά και το πλήθος της πληροφορίας που καθημερινά διακινείται, είναι τεράστια, όμως η έλλειψη σημασιολογίας στα δεδομένα που διακινούνται κάνει τους υπολογιστές ανίσχυρους να επεξεργαστούν τα δεδομένα χωρίς την ανθρώπινη διαμεσολάβηση.

Για την οργάνωση των δεδομένων με τρόπο ώστε να ορίζεται και η σημασιολογία τους, στις μέρες μας έχει επικρατήσει η RDF (Resource Description Framework), που αποτελεί τρία κομμάτια (triples) πληροφορίας που συνδυάζονται για να παράγουν μια ελάχιστα πλήρως ορισμένη γνώση (knowledge) και επιτρέπουν τόσο στον άνθρωπο να καταλαβαίνει την πληροφορία που αυτά εκφράζουν όσο και στη μηχανή να μπορεί να κατανοεί τη σημασιολογία των δεδομένων που περιγράφουν.

Η SPARQL είναι μια σημασιολογική γλώσσα ερωτημάτων που μας παρέχει τη δυνατότητα αναζήτησης της πληροφορίας που βρίσκεται μέσα στα αρχεία RDF χωρίς να γνωρίζουμε τη δομή των δεδομένων που περιέχονται σε αυτά.

Για την υλοποίηση συστημάτων εκτέλεσης ερωτημάτων SPARQL μπορούν να χρησιμοποιηθούν δύο διαφορετικές προσεγγίσεις. Η πρώτη απαιτεί την ανάπτυξη αλγορίθμων που προγραμματιστικά θα υλοποιούν την άλγεβρα SPARQL ώστε να γίνεται η επεξεργασία της πληροφορίας μέσα από τα αρχεία RDF. Το πλήθος των δεδομένων όμως που διακινούνται στις μέρες μας στον Παγκόσμιο Ιστό είναι τόσο μεγάλο, που η ανάπτυξη προγραμμάτων που θα επεξεργάζονται τα δεδομένα μέσα από τα αρχεία κειμένου RDF, θα απαιτούσε υπολογιστικά συστήματα με τεράστιους υπολογιστικούς πόρους.

Μια δεύτερη προσέγγιση είναι η μεταφορά της πληροφορίας που περιέχεται σε ένα αρχείο RDF, σε μια σχεσιακή βάση δεδομένων ενός προεπιλεγμένου γνωστού σχήματος και η υλοποίηση αλγορίθμου αντιστοίχισης των ερωτημάτων SPARQL στα αντίστοιχα ερωτήματα SQL. Η δεύτερη προσέγγιση, βασίζεται στη μετατροπή του ερωτήματος SPARQL στο αντίστοιχο ερώτημα SQL και τη χρήση της SQL στην επεξεργασία των δεδομένων, αφού η SQL έχει ήδη μεγάλη ερευνητική εξέλιξη και βελτιστοποίηση διότι αποτελεί για πολλά χρόνια την καθιερωμένη γλώσσα επερωτήσεων των σχεσιακών βάσεων δεδομένων.

Στην παρούσα Διπλωματική εργασία, χρησιμοποιείται η δεύτερη προσέγγιση στην υλοποίηση ενός συστήματος εκτέλεσης ερωτημάτων SPARQL. Αρχικά προτείνεται ένα σχήμα σχεσιακής βάσης δεδομένων (RDBMS), μέσα στο οποίο, μπορεί να γίνει αποθήκευση των δεδομένων και της πληροφορίας που περιέχεται σε οποιοδήποτε αρχείο RDF και στη συνέχεια παρουσιάζεται αλγόριθμος, που σχεδιάστηκε και υλοποιήθηκε, μετατροπής ερωτημάτων SPARQL στα αντίστοιχα ερωτήματα SQL (SPARQL to SQL Mapping)

Λέξεις Κλειδιά: Σημασιολογικός Ιστός, RDF, SPARQL, SPARQL to SQL Mapping