ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασία

Διπλωματικη Εργασια

της

ΤΣΑΜΟΥ ΟΛΥΜΠΙΑΣ

Επιβλέποντες: Ευάγγελος Δερματάς

Πάτρα, Μάιος 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ Υ-ΠΟΛΟΓΙΣΤΩΝ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασία

Διπλωματικη Εργασια

της

ΤΣΑΜΟΥ ΟΛΥΜΠΙΑΣ

Επιβλέποντες: Ευάγγελος Δερματάς

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 19η Μαΐου 2022.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ευάγγελος Δερματάς	Θεόδωρος Αντωνακόπουλος	****
Καθηγητής	Καθηγητής	Καθηγητής

Περίληψη

Περίληψη

Λέξεις Κλειδιά

Παράλληλος Προγραμματισμός, Συνελικτικά Νευρωνικά Δίκτυα, OpenACC, Αναγνώριση Εικόνας

Abstract

Abstract

Keywords

Parallel Programming, Convolutional Neural Networks, OpenACC,Image Classification

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. ..... για την επίβλεψη αυτής της διπλωματικής εργασίας [ ] ευχαριστώ ιδιαίτερα τον Δρ. ..... για την καθοδήγησή του και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Εισαγωγή.[3]

1.1 Ορισμός του Προβλήματος

Εφαρμογή τεχνικών βελτιστοποίησης σε πρόβλημα ταξινόμησης εικόνων κάνοντας χρήση ενός Συνελικτικού Νευρωνικού Δικτύου [1]

1.2 Στόχοι της Διπλωματικής Εργασίας

Γίνεται έρευνα σχετικά με τις διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων και τις μεθόδους εκπαίδευσης, ωστόσο, μια άλλη κρίσιμη πτυχή των νευρωνικών δικτύων είναι, δεδομένου ενός εκπαιδευμένου δικτύου, η γρήγορη και ακριβής ταξινόμηση εικόνων. Σε αυτό το project, δίνετε ένα εκπαιδευμένο νευρωνικό δίκτυο που ταξινομεί εικόνες 32x32 RGB σε 10 κατηγορίες. Οι εικόνες ανήκουν στο dataset CIFAR-10. Μας δίνετε ο αλγόριθμος εμπρόσθιας διάδοσης του νευρωνικού δικτύου και τα τελικά βάρη του δικτύου. Σκοπός είναι η βελτίωση της ταχύτητας της εμπρόσθιας διάδοσης ώστε να γίνει ταξινόμηση με πιο γρήγορο ρυθμό.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Κατηγοριοποίηση Εικόνων με χρήση Συνελικτικών Νευρωνικών Δικτύων classifying images using a Convolutional Neural Network

2.1 Αναγνώριση Εικόνων

2.1.1 Πώς ένας υπολογιστής αναγνωρίζει εικόνες·

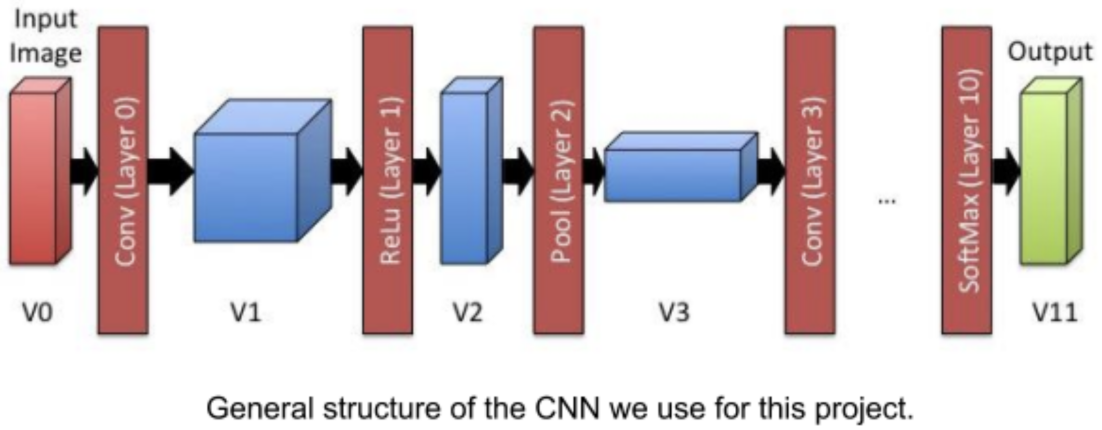
Η ταξινόμηση εικόνων περιγράφει ένα πρόβλημα στο οποίο σε ένα υπολογιστή δίνεται μία εικόνα και πρέπει να καταλάβει τι απεικονίζει (από ένα σύνολο πιθανών κατηγοριών) Σήμερα, τα Συνελικτικά Νευρωνικά Δίκτυα (CNNs) αποτελούν μια πολύ καλή προσέγγιση αυτού το προβλήματος. Γενικά, τα νευρωνικά δίκτυα υποθέτουν πως υπάρχει κάποια συνάρτηση από την είσοδο (π.χ. εικόνες) σε μία έξοδο (π.χ. ένα σύνολο κατηγοριών εικόνων). Ενώ οι κλασσικοί αλγόριθμοι προσπαθούν να κωδικοποιήσουν κάποια πληροφορία του πραγματικού κόσμου στη συνάρτηση τους, τα CNN μαθαίνουν την συνάρτηση δυναμικά από ένα σύνολο ταξινομημένων εικόνων (labelled images)—αυτή η διαδικασία ονομάζετε εκπαίδευση. Μόλις καταλήξει σε μια σταθερή συνάρτηση (δηλαδή σε μια προσέγγιση αυτής), μπορεί να εφαρμόσει τη συνάρτηση σε εικόνες που δεν έχει ξαναδεί.

2.1.2 Τι μπορεί να κάνει ένα νευρωνικό δίκτυο·

Ένα νευρωνικό δίκτυο αποτελείται από πολλαπλά επίπεδα. Κάθε επίπεδο λαμβάνει έναν πολυδιάστατο πίνακα αριθμών ως είσοδο και παράγει έναν άλλο πολυδιάστατο πίνακα αριθμών ως έξοδο (ο οποίος στη συνέχεια γίνεται η είσοδος του επόμενου επιπέδου). Κατά την ταξινόμηση εικόνων, η είσοδος του πρώτου επιπέδου είναι η εικόνα εισόδου (π.χ. 32x32x3 αριθμοί για εικόνες 32x32 pixel με 3 κανάλια χρώματος), ενώ η έξοδος του τελευταίου επιπέδου αποτελείται ένα σύνολο πιθανοτήτων των διαφόρων κατηγοριών (π.χ., 1x1x10 αριθμοί αν υπάρχουν 10 κατηγορίες).

Αρχιτεκτονική CNN 2.1

Κάθε επίπεδο έχει ένα σύνολο από βάρη που σχετίζονται με αυτό — αυτά τα βάρη είναι που "μαθαίνει" το νευρωνικό όταν του δοθούν δεδομένα εκπαίδευσης. Ανάλογα με το επίπεδο,



Σχήμα 2.1: Δομή Συνελικτικού Νευρωνικού Δικτύου

τα βάρη έχουν διαφορετικές ερμηνείες, αλλά δεν είναι αντικείμενο μελέτης του συγκεκριμένου project, φτάνει να γνωρίζουμε ότι κάθε επίπεδο λαμβάνει μία είσοδο, εκτελεί κάποια διεργασία σε αυτή, που εξαρτάται από τα βάρη και παράγει μια έξοδο. Αυτό το βήμα ονομάζεται forward pass: παίρνουμε μία είσοδο και την προωθούμε στο δίκτυο, παράγοντας το επιθυμητό αποτέλεσμα ως έξοδο. Το forward pass είναι το μόνο που χρειάζεται για την ταξινόμηση εικόνων σε ένα ήδη εκπαιδευμένο CNN.

Στην πράξη, ένα νευρωνικό δίκτυο αποτελεί μια πολύ απλή μηχανή αναγνώρισης προτύπων (με εξαιρετικά περιορισμένη χωρητικότητα), αλλά μπορεί να είναι αρκετά παράξενο αυτό που καταλήγει να αναγνωρίσει. Για παράδειγμα, κάποιος μπορεί να εκπαιδεύσει ένα νευρωνικό δίκτυο να αναγνωρίζει τη διαφορά μεταξύ "σκύλων" και "λύκων", και να δουλέψει καλά κοιτώντας το χιόνι και το δάσος στο φόντο των φωτογραφιών με τους λύκους.

Κεφάλαιο 3

Περιγραφή θέματος

Στο κεφάλαιο αυτό αρχικά γίνεται μια περιγραφή των συστημάτων ομότιμων κόμβων που είναι βασισμένα σε σχήματα (schema-based peer-to-peer systems). Στη συνέχεια περιγράφονται τρία βασικά συστήματα που ανήκουν σε αυτή την κατηγορία, καθώς και ένα σύστημα για τη διαχείρηση RDF σχημάτων, και τέλος αναλύεται ο στόχος της παρούσας εργασίας.

3.1 Σχετικές εργασίες

Οι βάσεις δεδομένων εισήγαγαν ένα τρόπο αποθήκευσης και ανάκτησης των δεδομένων που βασιζόταν στο σχήμα [2]. Τα πρώτα συστήματα ομότιμων κόμβων που περιγράψαμε στην Υποενότητα 2.1.2 έδιναν μεγάλη σημασία στην αρχιτεκτονική του συστήματος και την δρομολόγηση των ερωτήσεων και λιγότερη στον τρόπο αναπαράστασης και τις δυνατότητες αναζήτησης. Η αναζήτηση σε αυτά τα συστήματα ομότιμων κόμβων γίνεται με βάση προκαθορισμένα χαρακτηριστικά - δείκτες, ή με προσπάθεια αντιστοίχισης μιας λέξης κλειδί.

Η ανάγκη λοιπόν για πιο εκφραστικές λειτουργίες οδήγησε στα συστήματα ομότιμων κόμβων τα οποία είναι βασισμένα σε σχήματα (schema based peer-to-peer systems). Πρόκειται για ομότιμες υποδομές διαχείρισης δεδομένων που όμως διατηρούν όλα τα χαρακτηριστικά των συστημάτων ομότιμων κόμβων. .....

Κεφάλαιο 4

Ανάλυση και σχεδίαση

Στο κεφάλαιο αυτό παρουσιάζεται η μελέτη που έγινε για την υλοποίηση του συστήματος. Αρχικά περιγράφεται η αρχιτεκτονική του συστήματος και γίνεται ο διαχωρισμός του στα επιμέρους υποσυστήματα, ενώ στη συνέχεια περιγράφονται οι εφαρμογές του συστήματος.

4.1 Ανάλυση - περιγραφή αρχιτεκτονικής

Στην ενότητα αυτή παρουσιάζεται η ανάλυση του συστήματος και ο χωρισμός του σε υποσυστήματα όσον αφορά την αρχιτεκτονική.

4.1.1 Διαχωρισμός υποσυστημάτων

Το σύστημα αποτελείται από τους απλούς κόμβους και ένα κόμβο διαχειριστή. Στο σημείο αυτό αναλύουμε το σύστημα ενός απλού κόμβου, το οποίο αποτελείται από τα εξής υποσυστήματα:

- Υποσύστημα δημιουργίας σχήματος.
- Υποσύστημα ενσωμάτωσης δεδομένων στο σχήμα.
- Υποσύστημα επικοινωνίας κόμβου.

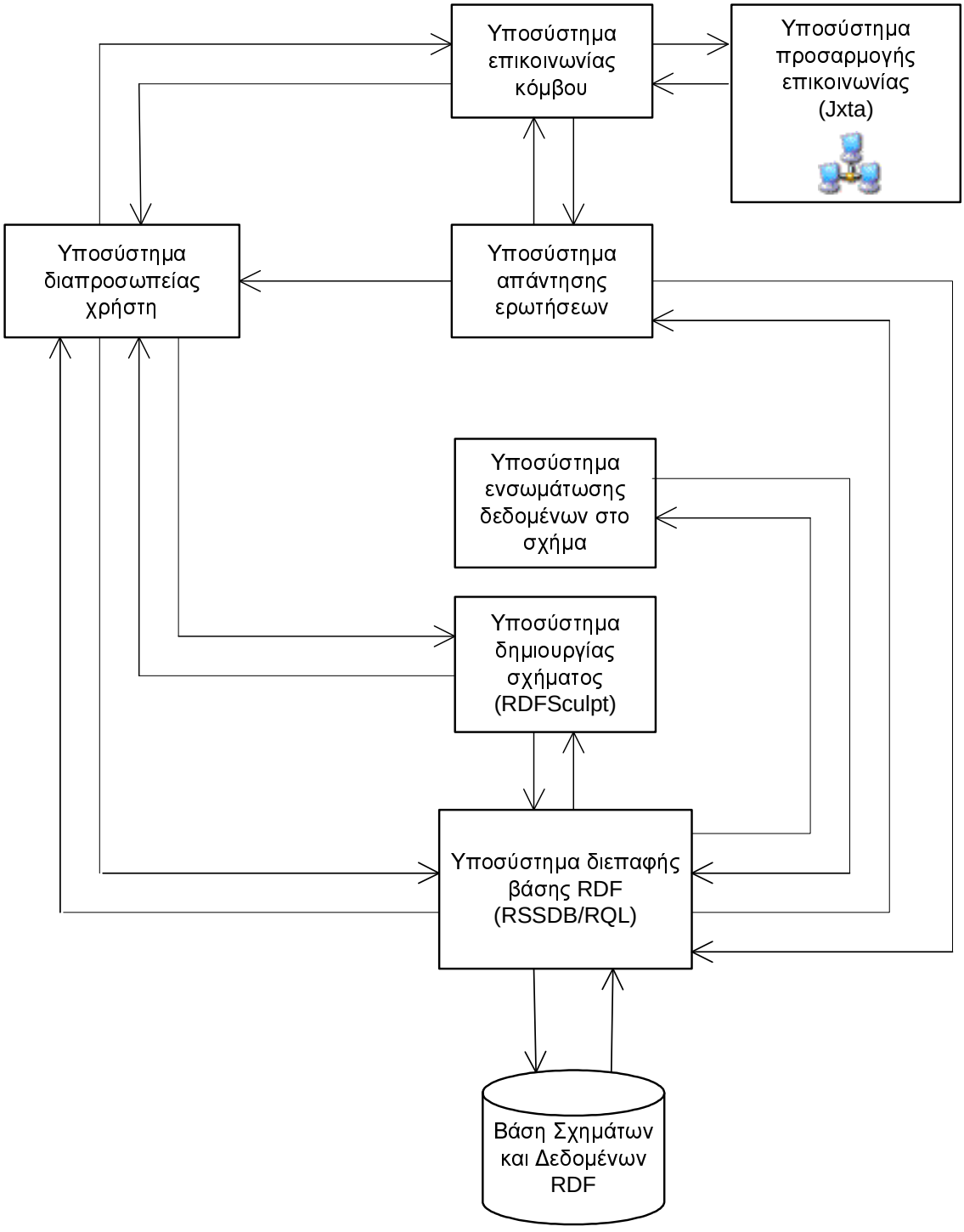
Το Σχήμα 4.1 απεικονίζει .....

4.1.2 Περιγραφή υποσυστημάτων

Παρακάτω δίνεται λεπτομερής περιγραφή για καθένα από τα συστήματα που αναφέραμε. Η περιγραφή αυτή γίνεται με βάση τα διαγράμματα ροής δεδομένων.

Υποσύστημα δημιουργίας σχήματος

Το υποσύστημα αυτό .....



Σχήμα 4.1: Αρχιτεκτονική Απλού Κόμβου

Κεφάλαιο 5

Υλοποίηση

Στο κεφάλαιο αυτό περιγράφεται η υλοποίηση του συστήματος, με βάση τη μελέτη που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Αρχικά παρουσιάζεται η πλατφόρμα και τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στη συνέχεια δίνονται οι λεπτομέρειες υλοποίησης για τους βασικούς αλγορίθμους του συστήματος καθώς και η δομή του κώδικα.

5.1 Λεπτομέρειες υλοποίησης

Στην ενότητα αυτή παρουσιάζονται οι βασικοί αλγόριθμοι που αναπτύχθηκαν καθώς και λεπτομέρειες σχετικά με την υλοποίηση της επικοινωνίας των κόμβων.

5.1.1 Αλγόριθμοι

Αλγόριθμος εισαγωγής δεδομένων

Όταν ένας κόμβος εισέρχεται για πρώτη φορά στο σύστημα, αρχικά δημιουργεί το σχήμα που θέλει χρησιμοποιώντας το RDFSculpt. Στη συνέχεια.....
Κατασκευή του διανύσματος groupedMapping.
Περιέχει ομαδοποιημένα τα στοιχεία του mapping που ανήκουν στην ίδια κλάση.
Το διάνυσμα groupedMapping έχει τη μορφή:
[[[[Κλάση1,Κυριολεκτικό1],Χαρακτηριστικό1],[[Κλάση1,Κυριολεκτικό2]
,Χαρακτηριστικό2],...],[[[Κλάση2,Κυριολεκτικό3],Χαρακτηριστικό3], [[Κλάση2,
Κυριολεκτικό4],Χαρακτηριστικό4],...]]

```
Για κάθε εγγραφή
Δημιούργησε αντίγραφο του groupedMapping, που ονομάζεται imapping
΄Οσο το imapping έχει στοιχεία
  Πάρε το πρώτο στοιχείο του διανύσματος έστω classMapping
     Βάλε την κλάση που ανήκει στο πρώτο στοιχείο, στο διάνυσμα
     classesToWrite
     ΄Οσο το διάνυσμα classesToWrite έχει στοιχεία
         Πάρε το στοιχείο-κλάση που βρίσκεται στην αρχή του διανύσματος
```

Παράδειγμα

Έστω ότι ο κόμβος έχει επιλέξει να συμμετέχει στο σύστημα με το RDF σχήμα που φαίνεται στο Σχήμα. Για τις ανάγκες του παραδείγματος θεωρούμε ότι η όψη αυτή περιέχει μόνο μία εγγραφή.

.....

5.2 Περιγραφή κλάσεων

Στην ενότητα αυτή δίνεται μια σύντομη περιγραφή των κλάσεων, των πεδίων και των μεθόδων που τις απαρτίζουν.

5.2.1 public class FirstUi

Η κλάση αυτή κατασκευάζει την οθόνη εισαγωγής του χρήστη στο σύστημα.

Πεδία

- `private GridBagLayout blayout`
  Το layout για όλα τα Panel.
- `private GridBagConstraints con`
  Τα constraints για το layout.
- `private Icon arrowR`
  Εικονίδιο για το κουμπί Next.

Μέθοδοι

- `public FirstUi()`
  Ο κατασκευαστής της κλάσης ο οποίος καλεί την createEntryFrame().
- `private void createEntryFrame()`
  Μέθοδος που κατασκευάζει το ενφραμε.

Κεφάλαιο 6

Έλεγχος

Στο κεφάλαιο αυτό γίνεται ο έλεγχος καλής λειτουργίας του συτσήματος.

6.1 Μεθοδολογία Ελέγχου

Ο έλεγχος του συστήματος αυτού πραγματοποιήθηκε με τη χρήση ενός σεναρίου λειτουργίας. Σύμφωνα με το σενάριο αυτό θεωρούμε ότι στο σύστημα υπάρχουν τρεις κόμβοι (peer1,peer2,peer3). Θεωρούμε επίσης ότι οι κόμβοι peer2 και peer3 έχουν ήδη σχήμα και δεδομένα.

Επίσης η τοπολογία του συστήματος έχει ως εξής: ο peer2 είναι γείτονας του peer1 και ο peer3 γείτονας του peer2.

Αρχικά λοιπόν θα δημιουργήσουμε σχήμα για τον κόμβο peer1 και στη συνέχεια θα εισάγουμε σε αυτό δεδομένα εξετάζοντας έτσι την καλή λειτουργία του υποσυστήματος δημιουργίας σχήματος και του υποσυστήματος εισαγωγής δεδομένων. Στη συνέχεια από τον κόμβο αυτό στέλνουμε ερωτήσεις στους υπόλοιπους για τον έλεγχο του υποσυστήματος απάντησης ερωτήσεων και επικοινωνίας κόμβων.

6.2 Αναλυτική παρουσίαση ελέγχου

Στην ενότητα αυτή παρουσιάζουμε αναλυτικά τον έλεγχο του συστήματος σύμφωνα με το σενάριο που περιγράφηκε στην προηγούμενη ενότητα.

Κεφάλαιο 7

Παράδειγμα Πίνακα

7.1 Συμπεράσματα

Τα συστήματα ομότιμων κόμβων, προκειμένου να υποστηρίζουν πιο εκφραστικές λειτουργίες αναπαράστασης και αναζήτησης δεδομένων, εξελίχθηκαν στα συστήματα ομότιμων κόμβων τα οποία βασίζονται στις τεχνολογίες του Σημασιολογικού Ιστού για την αναπαράσταση των δεδομένων μέσω σχημάτων που τα περιγράφουν (Schema-based peer-to-peer systems).

Συμπερασματικά το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής είναι ένα πλήρες σύστημα ομότιμων κόμβων βασισμένο σε σχήματα, το οποίο καθιστά δυνατή την αναζήτηση της πληροφορίας με ένα διαφορετικό τρόπο απ΄ ότι τα προϋπάρχοντα συστήματα.

7.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα:

- Ενσωμάτωση διαδικασίας επιλογής σχήματος με βάση το οποίο ο κόμβος θα συμμετέχει στο σύστημα. Έτσι όπως έχει σχεδιαστεί το σύστημα, κάθε κόμβος έχει τη δυνατότητα να δημιουργήσει πολλά σχήματα και να αποθηκεύσει δεδομένα σε περισσότερα από ένα. Ως σχήμα του κόμβου (με βάση το οποίο απαντάει τις ερωτήσεις), θεωρείται το τελευταίο στο οποίο αποθήκευσε δεδομένα. Η δυνατότητα επιλογής θα του παρείχε περισσότερη ευελιξία.

- Δυνατότητα αντιστοίχισης δεδομένων τα οποία να μην είναι αποθηκευμένα σε βάση δεδομένων αλλά σε αρχεία. Η αποδέσμευση από τη βάση δεδομένων θα έκανε το σύστημα πιο εύκολο στην εγκατάσταση και τη χρήση.

- Αξιολόγηση του συστήματος ως προς τη συμπεριφορά του αν συμμετέχει σε αυτό μεγάλος αριθμός κόμβων (scalability testing) και αν χρησιμοποιηθεί ένα πολύ μεγάλο καθολικό σχήμα. Η αξιολόγηση αυτή αφορά την ταχύτητα με την οποία ένας κόμβος παίρνει απαντήσεις σε μια ερώτηση καθώς και την ποιότητα των απαντήσεων.

Πίνακας 7.1: Πίνακας αλήθειας της λογικής συνάρτησης F

A	**B**	**C**	**F**
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

Κεφάλαιο 8

Παράδειγμα Μαθηματικών Σχέσεων – Εκφράσεων

8.1 Συμπεράσματα

Τα συστήματα ομότιμων κόμβων, προκειμένου να υποστηρίζουν πιο εκφραστικές λειτουργίες αναπαράστασης και αναζήτησης δεδομένων, εξελίχθηκαν στα συστήματα ομότιμων κόμβων τα οποία βασίζονται στις τεχνολογίες του Σημασιολογικού Ιστού για την αναπαράσταση των δεδομένων μέσω σχημάτων που τα περιγράφουν (Schema-based peer-to-peer systems).

Στα συστήματα αυτά κάθε $y = \int_0^1 f(x)dx$ $y = \int_0^1 f(x)dx$ κόμβος χρησιμοποιεί ένα σχήμα για την $\sum_{i=0}^{100} a_i$ αναπαράσταση των δεδομένων του. Όμως σε ένα σύστημα ομότιμων κόμβων, κάθε κόμβος έχει διαφορετικές απαιτήσεις αναπαράστασης δεδομένων. Επομένως πρέπει να υπάρχει ευελιξία στην επιλογή $\frac{1}{1+x^2}$ σχήματος. Τα συστήματα που έχουν προταθεί μέχρι τώρα και παρέχουν αυτή την ευελιξία, για να είναι δυνατή η αναζήτηση πληροφορίας, απαιτούν την ύπαρξη κανόνων αντιστοίχισης μεταξύ των σχημάτων με βάση τους οποίους να μετασχηματίζονται οι ερωτήσεις. Όμως δεν υποστηρίζεται ακόμα αυτόματη δημιουργία και δυναμική ανανέωση των κανόνων, που είναι απαραίτητα για τα συστήματα ομότιμων κόμβων.

$$y = \int_0^1 f(x)dx \tag{8.1}$$

Η συνεισφορά της (8.1) παρούσας διπλωματικής εργασίας έχει δύο σκέλη. Το πρώτο αφορά τη δημιουργία ενός πλήρους συστήματος ομότιμων κόμβων βασισμένο σε σχήματα RDF το οποίο παρέχει: (α) την υποδομή για την επικοινωνία των κόμβων,(β) μηχανισμό δημιουργίας σχήματος, (γ) μηχανισμό ενσωμάτωσης σχεσιακών δεδομένων στο σχήμα με τη χρήση αντιστοιχίσεων που δημιουργεί ο χρήστης με τη βοήθεια ειδικής διαπροσωπείας, (δ) ευέλικτη διαπροσωπεία χρήστη για τη διατύπωση ερωτημάτων και (ε) μηχανισμό απάντησης και επεξεργασίας ερωτήσεων.

Το δεύτερο σκέλος αφορά το γεγονός ότι το συγκεκριμένο σύστημα προσφέρει μια σχετική ευελιξία ως προς την επιλογή του σχήματος από τον κάθε κόμβο, ενώ ταυτόχρονα δίνει τη

δυνατότητα μετασχηματισμού ερωτήσεων χωρίς τη χρήση κανόνων αντιστοίχισης. Συγκεκριμένα, τα σχήματα των κόμβων αποτελούν υποσύνολα−όψεις(views) ενός βασικού σχήματος που ονομάζεται καθολικό σχήμα. Εκμεταλλευόμενοι λοιπόν το γεγονός ότι τα σχήματα αυτά είναι συμβατά μεταξύ τους, έχουμε τη δυνατότητα ελέγχου της ικανοποιησιμότητας μιας ερώτησης και μετατροπής της όπου χρειάζεται, χρησιμοποιώντας τόσο το σχήμα του κόμβου όσο και το καθολικό σχήμα.

Συμπερασματικά το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής είναι ένα πλήρες σύστημα ομότιμων κόμβων βασισμένο σε σχήματα, το οποίο καθιστά δυνατή την αναζήτηση της πληροφορίας με ένα διαφορετικό τρόπο απ΄ ότι τα προϋπάρχοντα συστήματα.

8.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα:

- Ενσωμάτωση διαδικασίας επιλογής σχήματος με βάση το οποίο ο κόμβος θα συμμετέχει στο σύστημα. Έτσι όπως έχει σχεδιαστεί το σύστημα, κάθε κόμβος έχει τη δυνατότητα να δημιουργήσει πολλά σχήματα και να αποθηκεύσει δεδομένα σε περισσότερα από ένα. Ως σχήμα του κόμβου (με βάση το οποίο απαντάει τις ερωτήσεις), θεωρείται το τελευταίο στο οποίο αποθήκευσε δεδομένα. Η δυνατότητα επιλογής θα του παρείχε περισσότερη ευελιξία.

- Δυνατότητα αντιστοίχισης δεδομένων τα οποία να μην είναι αποθηκευμένα σε βάση δεδομένων αλλά σε αρχεία. Η αποδέσμευση από τη βάση δεδομένων θα έκανε το σύστημα πιο εύκολο στην εγκατάσταση και τη χρήση.

- Αξιολόγηση του συστήματος ως προς τη συμπεριφορά του αν συμμετέχει σε αυτό μεγάλος αριθμός κόμβων (scalability testing) και αν χρησιμοποιηθεί ένα πολύ μεγάλο καθολικό σχήμα. Η αξιολόγηση αυτή αφορά την ταχύτητα με την οποία ένας κόμβος παίρνει απαντήσεις σε μια ερώτηση καθώς και την ποιότητα των απαντήσεων.

Κεφάλαιο 9

Επίλογος

9.1 Συμπεράσματα

Τα συστήματα ομότιμων κόμβων, προκειμένου να υποστηρίζουν πιο εκφραστικές λειτουργίες αναπαράστασης και αναζήτησης δεδομένων, εξελίχθηκαν στα συστήματα ομότιμων κόμβων τα οποία βασίζονται στις τεχνολογίες του Σημασιολογικού Ιστού για την αναπαράσταση των δεδομένων μέσω σχημάτων που τα περιγράφουν (Schema-based peer-to-peer systems).

Στα συστήματα αυτά κάθε κόμβος χρησιμοποιεί ένα σχήμα για την αναπαράσταση των δεδομένων του. Όμως σε ένα σύστημα ομότιμων κόμβων, κάθε κόμβος έχει διαφορετικές απαιτήσεις αναπαράστασης δεδομένων. Επομένως πρέπει να υπάρχει ευελιξία στην επιλογή σχήματος. Τα συστήματα που έχουν προταθεί μέχρι τώρα και παρέχουν αυτή την ευελιξία, για να είναι δυνατή η αναζήτηση πληροφορίας, απαιτούν την ύπαρξη κανόνων αντιστοίχισης μεταξύ των σχημάτων με βάση τους οποίους να μετασχηματίζονται οι ερωτήσεις. Όμως δεν υποστηρίζεται ακόμα αυτόματη δημιουργία και δυναμική ανανέωση των κανόνων, που είναι απαραίτητα για τα συστήματα ομότιμων κόμβων.

Η συνεισφορά της παρούσας διπλωματικής εργασίας έχει δύο σκέλη. Το πρώτο αφορά τη δημιουργία ενός πλήρους συστήματος ομότιμων κόμβων βασισμένο σε σχήματα RDF το οποίο παρέχει: (α) την υποδομή για την επικοινωνία των κόμβων,(β) μηχανισμό δημιουργίας σχήματος, (γ) μηχανισμό ενσωμάτωσης σχεσιακών δεδομένων στο σχήμα με τη χρήση αντιστοιχίσεων που δημιουργεί ο χρήστης με τη βοήθεια ειδικής διαπροσωπείας, (δ) ευέλικτη διαπροσωπεία χρήστη για τη διατύπωση ερωτημάτων και (ε) μηχανισμό απάντησης και επεξεργασίας ερωτήσεων.

Το δεύτερο σκέλος αφορά το γεγονός ότι το συγκεκριμένο σύστημα προσφέρει μια σχετική ευελιξία ως προς την επιλογή του σχήματος από τον κάθε κόμβο, ενώ ταυτόχρονα δίνει τη δυνατότητα μετασχηματισμού ερωτήσεων χωρίς τη χρήση κανόνων αντιστοίχισης. Συγκεκριμένα, τα σχήματα των κόμβων αποτελούν υποσύνολα−όψεις(views) ενός βασικού σχήματος που ονομάζεται καθολικό σχήμα. Εκμεταλλευόμενοι λοιπόν το γεγονός ότι τα σχήματα αυτά είναι συμβατά μεταξύ τους, έχουμε τη δυνατότητα ελέγχου της ικανοποιησιμότητας μιας ερώτησης και μετατροπής της όπου χρειάζεται, χρησιμοποιώντας τόσο το σχήμα του κόμβου όσο και το καθολικό σχήμα.

Συμπερασματικά το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής είναι ένα πλήρες σύστημα ομότιμων κόμβων βασισμένο σε σχήματα, το οποίο καθιστά δυνατή την αναζήτηση της πληροφορίας με ένα διαφορετικό τρόπο απ΄ ότι τα προϋπάρχοντα συστήματα.

9.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα:

- Ενσωμάτωση διαδικασίας επιλογής σχήματος με βάση το οποίο ο κόμβος θα συμμετέχει στο σύστημα. Έτσι όπως έχει σχεδιαστεί το σύστημα, κάθε κόμβος έχει τη δυνατότητα να δημιουργήσει πολλά σχήματα και να αποθηκεύσει δεδομένα σε περισσότερα από ένα. Ως σχήμα του κόμβου (με βάση το οποίο απαντάει τις ερωτήσεις), θεωρείται το τελευταίο στο οποίο αποθήκευσε δεδομένα. Η δυνατότητα επιλογής θα του παρείχε περισσότερη ευελιξία.

- Δυνατότητα αντιστοίχισης δεδομένων τα οποία να μην είναι αποθηκευμένα σε βάση δεδομένων αλλά σε αρχεία. Η αποδέσμευση από τη βάση δεδομένων θα έκανε το σύστημα πιο εύκολο στην εγκατάσταση και τη χρήση.

- Αξιολόγηση του συστήματος ως προς τη συμπεριφορά του αν συμμετέχει σε αυτό μεγάλος αριθμός κόμβων (scalability testing) και αν χρησιμοποιηθεί ένα πολύ μεγάλο καθολικό σχήμα. Η αξιολόγηση αυτή αφορά την ταχύτητα με την οποία ένας κόμβος παίρνει απαντήσεις σε μια ερώτηση καθώς και την ποιότητα των απαντήσεων.

Παράρτημα Α΄

Παράδειγμα Παραρτήματος

Α΄.1 Πρώτη ενότητα

Τα συστήματα ομότιμων κόμβων, προκειμένου να υποστηρίζουν πιο εκφραστικές λειτουργίες αναπαράστασης και αναζήτησης δεδομένων, εξελίχθηκαν στα συστήματα ομότιμων κόμβων τα οποία βασίζονται στις τεχνολογίες του Σημασιολογικού Ιστού για την αναπαράσταση των δεδομένων μέσω σχημάτων που τα περιγράφουν (Schema-based peer-to-peer systems).

Συμπερασματικά το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής είναι ένα πλήρες σύστημα ομότιμων κόμβων βασισμένο σε σχήματα, το οποίο καθιστά δυνατή την αναζήτηση της πληροφορίας με ένα διαφορετικό τρόπο απ΄ ότι τα προϋπάρχοντα συστήματα.

Α΄.2 Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε στα πλαίσια αυτής της διπλωματικής εργασίας θα μπορούσε να βελτιωθεί και να επεκταθεί περαιτέρω, τουλάχιστον ως προς τρεις κατευθύνσεις. Συγκεκριμένα, αναφέρονται τα ακόλουθα:

- Ενσωμάτωση διαδικασίας επιλογής σχήματος με βάση το οποίο ο κόμβος θα συμμετέχει στο σύστημα. Έτσι όπως έχει σχεδιαστεί το σύστημα, κάθε κόμβος έχει τη δυνατότητα να δημιουργήσει πολλά σχήματα και να αποθηκεύσει δεδομένα σε περισσότερα από ένα. Ως σχήμα του κόμβου (με βάση το οποίο απαντάει τις ερωτήσεις), θεωρείται το τελευταίο στο οποίο αποθήκευσε δεδομένα. Η δυνατότητα επιλογής θα του παρείχε περισσότερη ευελιξία.

- Δυνατότητα αντιστοίχισης δεδομένων τα οποία να μην είναι αποθηκευμένα σε βάση δεδομένων αλλά σε αρχεία. Η αποδέσμευση από τη βάση δεδομένων θα έκανε το σύστημα πιο εύκολο στην εγκατάσταση και τη χρήση.

- Αξιολόγηση του συστήματος ως προς τη συμπεριφορά του αν συμμετέχει σε αυτό μεγάλος αριθμός κόμβων (scalability testing) και αν χρησιμοποιηθεί ένα πολύ μεγάλο καθολικό σχήμα. Η αξιολόγηση αυτή αφορά την ταχύτητα με την οποία ένας κόμβος παίρνει απαντήσεις σε μια ερώτηση καθώς και την ποιότητα των απαντήσεων.

Πίνακας Α΄.1: Πίνακας αλήθειας της λογικής συνάρτησης F

A	**B**	**C**	**F**
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

Παράρτημα Β΄

Απόδειξη της σχέσης (8.1)

Στο κεφάλαιο αυτό παρουσιάζεται η μελέτη που έγινε για την υλοποίηση του συστήματος. Αρχικά περιγράφεται η αρχιτεκτονική του συστήματος και γίνεται ο διαχωρισμός του στα επιμέρους υποσυστήματα, ενώ στη συνέχεια περιγράφονται οι εφαρμογές του συστήματος.

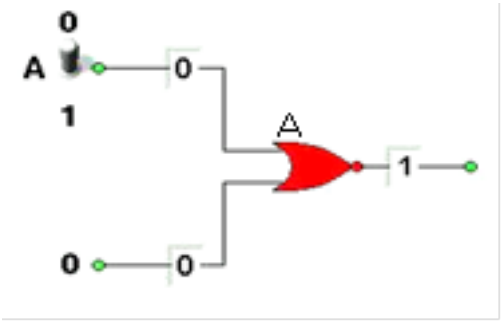
Β΄.1 Ανάλυση - περιγραφή αρχιτεκτονικής

Στην ενότητα αυτή παρουσιάζεται η ανάλυση του συστήματος και ο χωρισμός του σε υποσυστήματα όσον αφορά την αρχιτεκτονική.

Β΄.1.1 Διαχωρισμός υποσυστημάτων

Το σύστημα αποτελείται από τους απλούς κόμβους και ένα κόμβο διαχειριστή. Στο σημείο αυτό αναλύουμε το σύστημα ενός απλού κόμβου, το οποίο αποτελείται από τα εξής υποσυστήματα:

- Υποσύστημα δημιουργίας σχήματος.
- Υποσύστημα ενσωμάτωσης δεδομένων στο σχήμα.
- Υποσύστημα επικοινωνίας κόμβου.



Σχήμα Β΄.1: Προσομοίωση Πύλης NOR

Το Σχήμα Β΄.1 απεικονίζει .....

Β΄.1.2 Περιγραφή υποσυστημάτων

Παρακάτω δίνεται λεπτομερής περιγραφή για καθένα από τα συστήματα που αναφέραμε. Η περιγραφή αυτή γίνεται με βάση τα διαγράμματα ροής δεδομένων.

Υποσύστημα δημιουργίας σχήματος

Το υποσύστημα αυτό .....

Βιβλιογραφία

[1] UC Berkeley. *CS 61C Project*. https://inst.eecs.berkeley.edu/ cs61c/sp19/projects/proj4/, 2019.

[2] Wolfgang Nejdl, Wolf Siberski και Michael Sintek. Design issues and challenges for rdf- and schema-based peer-to-peer systems. *SIGMOD Rec.*, 32(3):41–46, 2003.

[3] Ζωή Καούδη. Πρότυπο Σύστημα Αποθήκευσης και Διαχείρισης σχημάτων RDFS. KDBS Lab, NTU Athens, 2004.

Συντομογραφίες - Αρκτικόλεξα - - Ακρωνύμια

βλπ	βλέπε
κ.λπ.	και λοιπά
κ.ο.κ	και ούτω καθεξής
ΤΕΙ	Τεχνολογικό Εκπαιδευτικό Ίδρυμα
BPF	Band Pass Filter

Απόδοση ξενόγλωσσων όρων

Απόδοση	Ξενόγλωσσος όρος
αδερφός	sibling
αμεταβλητότητα	idempotency
ανάκτηση πληροφορίας	information retrieval
αντιμεταθετικότητα	commutativity
απόγονος	descedant
απορρόφηση	absorption
βάση δεδομένων	database
γνώρισμα	attribute
διαπροσωπεία	interface
διαφορά	difference
δικτυακός κατάλογος	portal catalog
δικτυωτή δομή	lattice
δομικές επερωτήσεις	structural queries
δομικές σχέσεις	structural relationships
δομικό σχήμα	schema
εγκυρότητα	validity
ένωση	union

Ευρετήριο όρων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ - ΠΟΛΙΤΕΞΝΙΚΗ ΣΞΟΛΗ - ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΞΑΝΙΚΩΝ ΚΑΙ ΤΕΞΝΟΛΟΓΙΑΣ ΨΠΟΛΟΗΙΣΤΩΝ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασίας

Ολυμπία Ι. Τσάμου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΤΡΑ,
Μάιος
2022

Θέση barcode

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ - ΠΟΛΙΤΕΞΝΙΚΗ ΣΞΟΛΗ - ΤΜΗΜΑ ΜΗΞΑΝΙΚΩΝ ΨΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασίας

Ολυμπία Ι. Τσάμου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΤΡΑ,
Μάιος
2022

Θέση barcode

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασίας

Ολυμπία Ι. Τσάμου

ΠΑΤΡΑ

Μάιος 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναζήτηση και Εξόρυξη Πληροφορίας σε Μεγάλες Βάσεις Αδόμητων Δεδομένων με Μεθόδους Παράλληλης Επεξεργασίας

Ολυμπία Ι. Τσάμου

ΠΑΤΡΑ

Μάιος 2022