Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Μελέτη επιδόσεως του μηχανισμού διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού

Γεωργία Μαργαρίτη

Μεταπτυχιακή Εργασία

Ηράκλειο, Φεβρουάριος 1997

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημων
Τμήμα Επιστήμης Υπολογιστών

Μελέτη επιδόσεως του μηχανισμού διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού

Εργασία που υποβλήθηκε από την
ΓΕΩΡΓΙΑ ΜΑΡΓΑΡΙΤΗ
ως μερική εκπλήρωση των απαιτήσεων
για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Γεωργία Μαργαρίτη
Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής, Επόπτης

Δημήτρης Σερπάνος, Επίκουρος Καθηγητής, Μέλος

Γεώργιος Δ. Σταμούλης, Επίκουρος Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, Φεβρουάριος 1997

Μελέτη επιδόσεως του μηχανισμό διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού

Γεωργία Μαργαρίτη

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

ΠΕΡΙΛΗΨΗ

Η κρυφή μνήμη είναι ταχεία μνήμη ενταμιευτών, μικρή σε μέγεθος, η οποία χρησιμοποιείται στα υπολογιστικά συστήματα για προσωρινή αποθήκευση δεδομένων της κύριας μνήμης. Με δεδομένο ότι η πρόσβαση στο δίσκο είναι 40-200 χιλιάδες φορές πιο αργή σε σύγκριση με την πρόσβαση στην κρυφή μνήμη, είναι φανερό ότι η διαχείριση της κρυφής μνήμης πρέπει να στοχεύει στον περιορισμό των πράξεων ανάγνωσης/γραφής από/στο δίσκο, ώστε να ελαττώνεται ο χρόνος απόκρισης του συστήματος.

Παρά το γεγονός ότι τα λειτουργικά συστήματα των σύγχρονων υπολογιστικών συστημάτων διαθέτουν ολοκληρωμένους μηχανισμούς διαχείρισης κρυφής μνήμης, τα συστήματα διαχείρισης βάσεων δεδομένων διαθέτουν τους δικούς τους μηχανισμούς προκειμένου να εκμεταλλευτούν τα ιδιαίτερα χαρακτηριστικά προσπελάσεων δεδομένων που πραγματοποιείται σ'αυτά.

Προκειμένου να μελετηθεί η επίδοση των μηχανισμών κρυφής μνήμης στα συστήματα διαχείρισης βάσεων δεδομένων, χρησιμοποιούνται μέθοδοι μετρήσεων. Οι μετρήσεις αυτές περιλαμβάνουν υλοποίηση και χρήση προσομοιωτικών μοντέλων, χρήση προκαθορισμένων δοκιμασιών επιδόσεων (benchmarks) ή εκτέλεση πειραμάτων και συγκέντρωση μετρήσεων στο πραγματικό σύστημα.

Το Σύστημα Σημασιολογικού Ευρετηριασμού (ΣΣΕ), που αναπτύχθηκε στο Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας, διαχειρίζεται βάσεις γνώσεις στις οποίες η γνώση παριστάνεται με τη μορφή σημασιολογικού δικτύου. Το ΣΣΕ διαθέτει δικό του μηχανισμό διαχείρισης κρυφής μνήμης.

Αντικείμενο της εργασίας αυτής είναι η μελέτη της συμπεριφοράς του μηχανισμού διαχείρισης κρυφής μνήμης του ΣΣΕ. Για τη μελέτη αυτή υιοθετήθηκε η προσέγγιση της εκτέλεσης πειραμάτων στο πραγματικό σύστημα. Οι μετρήσεις στο ΣΣΕ πραγματοποιήθηκαν για τρεις διαφορετικές βάσεις δεδομένων, που επιλέχθηκαν με βάση ιδιαίτερα χαρακτηριστικά τους, ενώ χρησιμοποιήθηκαν σενάρια τυπικής χρήσης του συστήματος.

Τα μεγέθη των επιμέρους μηχανισμών που απαρτίζουν την κρυφή μνήμη του συστήματος αποτέλεσαν τις παραμέτρους των μετρήσεων. Τα μεγέθη που καταγράφονταν από τις μετρήσεις είναι η συνολική κατανάλωση μνήμης από το σύστημα καθώς επίσης και τα ποσοστά επιτυχών προσπελάσεών της.

Το αποτέλεσμα της εργασίας είναι η εξαγωγή συμπερασμάτων για τη συμπεριφορά του μηχανισμού διαχείρισης κρυφής μνήμης καθώς επίσης και η εξαγωγή καμπύλων επίδοσης των επιμέρους μηχανισμών που την απαρτίζουν συναρτήσει της χωρητικότητάς τους, κατά την τυπική χρήση του συστήματος. Οι καμπύλες αυτές μπορούν να χρησιμοποιηθούν για τον προδιορισμό των παραμέτρων μεγέθους του μηχανισμού με στόχο τον περιορισμό της συνολικής κατανάλωσης μνήμης από το σύστημα, χωρίς να θίγεται η επίδοσή του όσον αφορά το χρόνο απόκρισης.

Επόπτης : Πάνος Κωνσταντόπουλος
Αναπληρωτής Καθηγητής Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Performance metrics for the cache management mechanism of the Semantic Index System

Georgia Margariti

Master of Science Thesis

Department of Computer Science
University of Crete

ABSTRACT

Cache memory is small in size, fast buffer memory that is used in computer systems for the temporary storage of data of the main memory. Given that disk access is 40-200 thousand times slower than memory access, it is obvious that cache mamagement must aim at diminishing read/write operations from/to the disk, so that the system's response time is restrained.

Despite the fact that the operating systems of modern computer systems have complete cache management mechanisms, database management systems have their own mechanisms in order to take advantage of their special data access characteristics.

Metrics methods are used to study the performance of cache management mechanisms in database management systems. These metrics can involve simulator models, benchmarks or experimentation and metrics' aquisistion on the real system.

The Semantic Index System, that was developed in the Institute of Computer Science, FORTH, manages knowledge bases in which knowledge is modeled in the form of a semantic network. The Semantic Index System deployes its own cache management mechanism.

The study of the cache management mechanism's behaviour in the Semantic Index System is the goal of this work. The method of experimenting and collecting metris on the real system was adopted for this study. The experiments on the Sematic Index System were conducted for three different databases, chosen based on their special characteristics, and typical system operation scenarios were used.

The parameters of the experiments have been the sizes of the separate mechanisms that cache memory consists of. Total memory consumption and hit rates have been the outcome of the experiments.

Interesting conclusions have been drawn concerning the cache management mechanism, as a result of the abovementioned work, as well as performance curves of the separate mechanisms that make up the cache memory given their sizes, during the system's typical usage. These curves can be used for determining the size parameters of the mechanism with the aim of restraining the system's total memory consumption, without allowing its performance deteriorate.

Supervisor : Panos Constantopoulos
Associate Professor of Computer Science
University of Crete

Ευχαριστίες

Αισθάνομαι την ανάγκη να ευχαριστήσω όλους όσους μου συμπαραστάθηκαν και με βοήθησαν κατά τη διάρκεια αυτής της εργασίας.

Ιδιαίτερα θα ήθελα να ευχαριστήσω τον επόπτη καθηγητή μου κ. Πάνο Κωνσταντόπουλο που μου εμπιστεύθηκε αυτή την εργασία. Τον ευχαριστώ για τις συμβουλές και τις υποδείξεις του οποτεδήποτε το χρειάστηκα.

Επίσης ευχαριστώ θερμά τους γονείς μου Τάκη και Χρυσούλα και τις αδελφές μου Μαρία και Τζοβάνα για το κουράγιο, την υπομονή και την αγάπη τους.

Χρωστώ ένα μεγάλο ευχαριστώ στο Γιάννη Δασκαλάκη για την υπομονή και την κατανόηση που έδειχνε στις ώρες δουλειάς που απαιτούσε η εργασία αυτή, καθώς επίσης για την ηθική συμπαράσταση που μου προσέφερε.

Για τις συμβουλές τους, τη συμπαράσταση και το ξενύχτι, θα ήθελα ιδιαίτερα να ευχαριστήσω την Αθηνά Τραψιώτη, το Γιάννη Τζίτζικα, το Μάνο Θεοδωράκη, τη Χριστίνα Γκριτζάπη, το Δημήτρη Αγγελάκη, την Ανθή Γιώρτσου και τη Σούλα Αναστασιάδη.

Θα ήθελα να ευχαριστήσω ακόμη το Γιώργο Γεωργιαννάκη καθώς επίσης και όλα τα μέλη (νυν και πρώην) της Ομάδας Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας για την πολύτιμη συνεργασία τους.

Τέλος, θα ήθελα να ευχαριστήσω τα μέλη της εξεταστικής επιτροπής κ. Δημήτρη Σερπάνο και κ. Γεώργιο Δ. Σταμούλη για τα σχόλιά τους στην τελική διαμόρφωση του κειμένου καθώς επίσης και το Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας για την οικονομική και υλικοτεχνική υποστήριξη που μου παρείχε κατά τη διάρκεια των μεταπτυχιακών μου σπουδών.

Στους γονείς μου Τάκη και Χρυσούλα

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

1 Εισαγωγή **1**

1.1 Το πρόβλημα 1

1.2 Η προσέγγιση που υλοποιήθηκε 2

1.3 Υλοποίηση 3

1.4 Οργάνωση της εργασίας 3

2 Βιβλιογραφική εργασία **5**

2.1 Διαχείριση κρυφής μνήμης 5

2.1.1 Στοιχεία διαχείρισης κρυφής μνήμης 6

2.1.2 Διαχείριση κρυφής μνήμης σε συστήματα βάσεων δεδομένων 8

2.2 Αλγόριθμοι αντικατάστασης 9

2.2.1 Γνωστοί αλγόριθμοι 9

2.2.2 Εφαρμογή σε γνωστά συστήματα διαχείρισης οντοκεντρικών βάσεων δεδομένων 11

2.3 Αρχιτεκτονικές πελάτη-εξυπηρετητή 12

2.3.1 Το μοντέλο πελάτη-εξυπηρετητή για συστήματα διαχείρισης οντοκεντρικών βάσεων δεδομένων 12

2.3.2 Διαχείριση κρυφής μνήμης στον πελάτη 13

2.4 Διατήρηση της συνέπειας 14

2.4.1 Γενικά 14

2.4.2 Αλγόριθμοι διατήρησης της συνέπειας 15
2.4.3 Αλγόριθμοι διατήρησης της συνέπειας σε γνωστά συστήματα 22
2.5 Μετρήσεις 23
2.6 Προανάκληση 25
2.7 Συμπεράσματα 26

3 Διαχείριση κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού 29
3.1 Τύποι οντοτήτων στην SIS-Telos 29
3.1.1 Η γλώσσα SIS-Telos και το Σύστημα Σημασιολογικού Ευρετηριασμού 29
3.1.2 Τύποι οντοτήτων στην SIS-Telos 31
3.1.3 Υλοποίηση των οντοτήτων στη γλώσσα SIS-Telos 32
3.2 Διαχείριση οντοτήτων στο Σύστημα Σημασιολογικού Ευρετηριασμού . . . 36
3.2.1 Κατάλογος συστήματος 36
3.2.2 Μηχανισμός διαχείρισης οντοτήτων 39
3.2.3 Κατάλογος λογικών ονομάτων 40
3.3 Μηχανισμός κρυφής μνήμης για το σύστημα διαχείρισης οντοτήτων 44
3.3.1 Επιμέρους μηχανισμοί κρυφής μνήμης 45

4 Ανάλυση της διαχείρισης κρυφής μνήμης 47
4.1 Στοιχεία που χαρακτηρίζουν ένα σύστημα διαχείρισης κρυφής μνήμης . . 47
4.2 Σχεδίαση κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού . . 48
4.3 Ανάλυση των παραμέτρων 49

5 Σχεδίαση πειραμάτων 53
5.1 Εισαγωγή 53
5.2 Παράμετροι μέτρησης 54
5.2.1 Τιμές των παραμέτρων (μεταβλητών) των πειραμάτων 56
5.2.2 Επιλογή συνδυασμών μετρήσεων για τις διαφορετικές τιμές των παραμέτρων 58
5.3 Μεγέθη που καταγράφονται από τα πειράματα 58
5.4 Βάσεις μετρήσεων 61
5.5 Σενάρια χρήσης 63

6 Τυπική συμπεριφορά συστήματος : βάσεις ΚΛΕΙΩ και ΑΑΤ **67**

6.1 Εισαγωγή 67

6.2 ΚΛΕΙΩ 68

6.2.1 Αποτελέσματα πειραμάτων 68

6.2.2 Συμπεράσματα 87

6.3 ΑΑΤ 89

6.3.1 Αποτελέσματα πειραμάτων 90

6.3.2 Συμπεράσματα 106

7 Ακραία συμπεριφορά συστήματος : βάση ULAN **109**

7.1 Εισαγωγή 109

7.2 Αποτελέσματα μετρήσεων 110

7.2.1 Αποτελέσματα 110

7.2.2 Συμπεράσματα 126

7.3 Διεξαγωγή πρόσθετων μετρήσεων επίδοσης 127

7.3.1 Αποτελέσματα 129

7.3.2 Συμπεράσματα 141

8 Συμπεράσματα **143**

8.1 Εισαγωγή 143

8.2 Κατανάλωση μνήμης από το σύστημα 144

8.3 Χρόνος απόκρισης συστήματος 155

9 Επίλογος **157**

9.1 Συμπεράσματα 157

9.2 Μελλοντική εργασία 158

Βιβλιογραφία **160**

A Σενάριο τυπικής χρήσης του ΚΛΕΙΩ **167**

B Σενάριο τυπικής χρήσης του ΑΑΤ **173**

C Σενάριο τυπικής χρήσης του ULAN **179**

Κεφάλαιο 1

Εισαγωγή

1.1 Το πρόβλημα

Το Σύστημα Σημασιολογικού Ευρετηριασμού είναι μια εφαρμογή, που αναπτύχθηκε στο Ινστιτούτο Πληροφορικής στο Ιδρυμα Τεχνολογίας και Ερευνας, με αντικείμενο την περιγραφή και τεκμηρίωση βάσεων δεδομένων. Χρησιμοποιεί τη γλώσσα παράστασης γνώσης SIS-Telos για την περιγραφή των περιεχομένων της βάσης και διαθέτει για τη διαχείρισή τους μηχανισμούς αποθήκευσης αντικειμένων, ερωτήσεων και εισαγωγής δεδομένων.

Η κρυφή μνήμη, εξάλλου, είναι το ενδιάμεσο σημείο προσωρινής αποθήκευσης δεδομένων μεταξύ δίσκου και επεξεργαστή. Αν και όλα τα λειτουργικά συστήματα διαθέτουν μηχανισμούς για την εύκολη και αποδοτική διαχείριση της κρυφής μνήμης, τα συστήματα διαχείρισης βάσεων δεδομένων, κι ανάμεσα σ'αυτά και το Σύστημα Σημασιολογικού Ευρετηριασμού, διαθέτουν δικούς τους μηχανισμούς για το σκοπό αυτό.

Ο μηχανισμός διαχείρισης κρυφής μνήμης στο ΣΣΕ υλοποιήθηκε στα πλαίσια της κατασκευής του γενικότερου μηχανισμού αποθήκευσης οντοτήτων για τη γλώσσα παράστασης γνώσης SIS-Telos, η οποία αποτέλεσε το αντικείμενο της μεταπτυχιακής εργασίας του Γ. Γεωργιαννάκη [47].

Η ανάγκη για τη διεξαγωγή μετρήσεων επίδοσης για το μηχανισμό διαχείρισης κρυφής μνήμης του ΣΣΕ δημιουργήθηκε όταν το σύστημα άρχισε να χρησιμοποιείται για την περιγραφή βάσεων δεδομένων πολύ μεγάλου μεγέθους - πολύ μεγάλου σε σχέση με τη βάση του συστήματος ΚΛΕΙΩ, η οποία αποτελούσε το αντιπροσωπευτικότερο δείγμα των προηγούμενων εφαρμογών. Για τις νέες εφαρμογές, ο χρόνος απόκρισης του συστήματος

είναι στοιχείο εξαιρετικής σημασίας και, καθώς ο χρόνος που δαπανάται για τη μεταφορά δεδομένων από και προς τη βάση αποτελεί το συντριπτικά μεγαλύτερο ποσοστό του ενώ η σωστή διαχείριση της κρυφής μνήμης μπορεί να προκαλέσει μεγάλη ελάττωσή του, αποφασίστηκε η εκτενής μελέτη της επίδοσης του μηχανισμού.

1.2 Η προσέγγιση που υλοποιήθηκε

Η μέθοδος που επιλέχθηκε για τη μελέτη του μηχανισμού διαχείρισης κρυφής μνήμης του ΣΣΕ είναι αυτή της εκτέλεσης μετρήσεων στο πραγματικό σύστημα με τη βοήθεια προκαθορισμένων σεναρίων χρήσης του. Οι μετρήσεις πραγματοποιήθηκαν για διαφορετικά μεγέθη των επιμέρους μηχανισμών κρυφής μνήμης ενώ υπολογίζονταν σ'αυτές τόσο οι επιμέρους όσο και η συνολική κατανάλωση μνήμης από το σύστημα, καθώς επίσης και τα επιμέρους και συνολικά ποσοστά επιτυχών προσπελάσεων στην κρυφή μνήμη.

Οι μέθοδοι που παρουσιάζονται στη βιβλιογραφία σχετικά με την πραγματοποίηση μετρήσεων σε συστήματα διαχείρισης βάσεων δεδομένων συνοψίζονται σε :

- Μετρήσεις σε πραγματικό σύστημα, όμοια με τη δική μας προσέγγιση.
- Μετρήσεις σε προσομοιωτικά μοντέλα, για τη μελέτη νέων μεθόδων που δεν έχουν ακόμη υλοποιηθεί.
- Μετρήσεις με βάση πρότυπα ερωτήσεων (benchmarks) σε βάσεις δεδομένων με προκαθορισμένο περιεχόμενο.

Ο λόγος για τον οποίο επιλέχθηκε, στην περίπτωση του ΣΣΕ, η πρώτη από τις τρεις προσεγγίσεις είναι διότι τόσο τα πρότυπα ερωτήσεων όσο και τα προσομοιωτικά μοντέλα δεν ικανοποιούν συγκεκριμένες προϋποθέσεις. Σε ό,τι αφορά τα προσομοιωτικά μοντέλα, αυτά χρησιμοποιούνται, όπως προαναφέρθηκε, για τη συγκριτική μελέτη πρωτότυπων μεθόδων, πράγμα που δεν ισχύει στη δική μας περίπτωση. Τα πρότυπα ερωτήσεων, εξάλλου, κατασκευάστηκαν με κύριο προσανατολισμό συστήματα βάσεων δεδομένων σε εφαρμογές CAD/CAM, γι'αυτό και δεν αποτελούν ικανοποιητικό μέτρο σύγκρισης για το δικό μας σύστημα.

Οι μετρήσεις που έγιναν για το σύστημα διαχείρισης κρυφής μνήμης του ΣΣΕ οδήγησαν - σε μία περίπτωση βάσης δεδομένων - στην εξαγωγή καμπυλών επίδοσης του μηχανισμού

για διαφορετικές τιμές των μεγεθών των επιμέρους κρυφών μνημών. Με τη βοήθεια τέτοιου είδους καμπυλών επίδοσης για διαφορετικές κατηγορίες - ως προς το συνολικό μέγεθος και την αναλογία σχήματος και δεδομένων - βάσεων δεδομένων, είναι δυνατό να επιλέγονται τα μεγέθη των κρυφών μνημών ώστε να εξασφαλίζεται για το σύστημα κάποια επιθυμητή επίδοση, χωρίς να ξεπερνιέται ένα προκαθορισμένο όριο συνολικής κατανάλωσης μνήμης.

1.3 Υλοποίηση

Οι μετρήσεις στο Σύστημα Σημασιολογικού Ευρετηριασμού είχαν στόχο τη μελέτη τόσο του γραφικού συστήματος διατύπωσης ερωτήσεων που αυτό διαθέτει, όσο και του συστήματος Διαλογικών Δελτίων Ενημέρωσης, το οποίο είναι μια εφαρμογή του ΣΣΕ για εισαγωγή δεδομένων στη βάση.

Οι βάσεις δεδομένων για τις οποίες πραγματοποιήθηκαν τα πειράματα επίδοσης είναι η βάση του συστήματος ΚΛΕΙΩ, το οποίο σχεδιάστηκε και κατασκευάστηκε για τεκμηρίωση πολιτιστικών αγαθών, ο θησαυρός όρων τέχνης και αρχιτεκτονικής (Art & Architecture Thesaurus - AAT) καθώς επίσης και η βάση ULAN που περιέχει πληροφορία για καλλιτέχνες και καταχωρήσεις σχετικά μ'αυτούς και το έργο τους.

1.4 Οργάνωση της εργασίας

Στο δεύτερο κεφάλαιο παρουσιάζεται μια αναλυτική βιβλιογραφική εργασία σχετικά με τη διαχείριση κρυφής μνήμης σε συστήματα διαχείρισης βάσεων δεδομένων. Παρουσιάζονται τα κύρια στοιχεία που καθορίζουν το σχεδιασμό ενός μηχανισμού διαχείρισης κρυφής μνήμης, όπως ο αλγόριθμος αντικατάστασης σελίδων, ο αλγόριθμος διατήρησης της συνέπειας σε συστήματα που διαθέτουν περισσότερες από μία κρυφές μνήμες καθώς επίσης και το ζήτημα της προανάκλησης σελίδων στην κρυφή μνήμη. Γίνεται αναφορά, τέλος, στο θέμα της διεξαγωγής μετρήσεων προκειμένου να μελετηθεί η συμπεριφορά του μηχανισμού διαχείρισης κρυφής μνήμης σε συστήματα βάσεων δεδομένων.

Στο τρίτο κεφάλαιο περιγράφεται ο μηχανισμός αναπαράστασης και αποθήκευσης οντοτήτων στο Σύστημα Σημασιολογικού Ευρετηριασμού. Επίσης περιγράφεται αναλυτικά ο μηχανισμός διαχείρισης κρυφής μνήμης για το σύστημα.

Στο τέταρτο κεφάλαιο αναλύεται ο μηχανισμός διαχείρισης κρυφής μνήμης του Συστήματος Σημασιολογικού Ευρετηριασμού. Στόχος της ανάλυσης αυτής είναι να εντοπιστούν τα στοιχεία εκείνα στο σχεδιασμό του μηχανισμού διαχείρισης κρυφής μνήμης τα οποία επιδέχονται οποιαδήποτε τροποποίηση, ώστε να διαφανεί σε ποια κατεύθυνση θα μπορούσε να κινηθεί μια αλλαγή του σχεδιασμού με σκοπό τη βελτίωση της επίδοσής του.

Στο πέμπτο κεφάλαιο περιγράφεται η πειραματική σχεδίαση. Παρατίθενται όλες οι παράμετροι στο σχεδιασμό των πειραμάτων, τα οποία υλοποιήθηκαν για τη μελέτη της επίδοσης του μηχανισμού διαχείρισης κρυφής μνήμης.

Στο έκτο κεφάλαιο περιγράφονται τα αποτελέσματα των πειραμάτων για τις βάσεις ΚΛΕΙΩ και ΑΑΤ. Τα αποτελέσματα αυτά χαρακτηρίζονται τυπικά σε σχέση με αυτά της βάσης του ULAN, η οποία παρουσιάζει μεγάλες ιδιαιτερότητες, τόσο ως προω το μέγεθος όσο και ως προς την αναλογία σχήματος-δεδομένων που τη χαρακτηρίζει.

Στο έβδομο κεφάλαιο περιγράφονται τα πειραματικά αποτελέσματα για τη βάση του ULAN. Εκεί παρατίθενται και τα αποτελέσματα για μια επιπλέον σειρά πειραμάτων που πραγματοποιήθηκε για τη βάση αυτή, με σκοπό τον εντοπισμό των ”ορίων” της συμπεριφοράς της.

Στο όγδοο κεφάλαιο παρατίθενται τα συμπεράσματα από τη διεξοδική μελέτη της συμπεριφοράς της βάσης του ULAN. Πιο συγκεκριμένα, τα συμπεράσματα συνοψίζονται σε μια πρώτη συνδυαστική μελέτη της επίδοσης του συστήματος με παραμέτρους τόσο τη συνολική κατανάλωση μνήμης, όσο και τα ποσοστά επιτυχημένω προσπελάσεων.

Στο ένατο κεφάλαιο υπάρχει ο επίλογος και αναφορές σε μελλοντικές κατευθύνσεις στη διεξαγωγή πειραμάτων για το μηχανισμό διαχείρισης κρυφής μνήμης.

Κεφάλαιο 2

Βιβλιογραφική εργασία

Το κεφάλαιο αυτό ασχολείται με τη βιβλιογραφική ανασκόπηση που έχει γίνει με στόχο την υλοποίηση πειραμάτων για τη μελέτη της επίδοσης του μηχανισμού διαχείρισης της κρυφής μνήμης του Συστήματος Σημασιολογικού Ευρετηριασμού. Επίσης, μελετούνται εδώ εναλλακτικοί μηχανισμοί σχεδίασης για συστήματα διαχείρισης κρυφής μνήμης, κύρια για αρχιτεκτονικές πελάτη-εξυπηρετητή, οι οποίες αποτελούν τη μελλοντική κατεύθυνση του συστήματός μας.

Ο μηχανισμός διαχείρισης της κρυφής μνήμης που τώρα υπάρχει και ο οποίος δημιουργήθηκε σαν τμήμα του συνολικού μηχανισμού διαχείρισης και αποθήκευσης οντοτήτων για την Telos, απαιτεί βελτιώσεις για την καλύτερη επίδοσή του και την καλύτερη επίδοση ολόκληρου του συστήματος SIS και των εφαρμογών του [47].

2.1 Διαχείριση κρυφής μνήμης

Η κρυφή μνήμη είναι μικρή σε μέγεθος, ταχεία μνήμη ενταμιευτών, η οποία χρησιμοποιείται στα υπολογιστικά συστήματα για προσωρινή αποθήκευση δεδομένων της κύριας μνήμης, τα οποία (πιστεύεται ότι) πρόκειται να χρησιμοποιηθούν. Ουσιαστικά, η κρυφή μνήμη είναι η επαφή ανάμεσα στην κύρια μνήμη και το δίσκο [42]. Με δεδομένο ότι η πρόσβαση στο δίσκο είναι 4-10 φορές πιο αργή σε σύγκριση με την πρόσβαση στην κρυφή μνήμη, είναι φανερό ότι η διαχείριση της κρυφής μνήμης πρέπει να στοχεύει στον περιορισμό των πράξεων ανάγνωσης/γραφής από/στο δίσκο. Η επιτυχία της χρήσης της αποδίδεται στο φαινόμενο που παρατηρείται ότι οι διευθύνσεις της μνήμης οι οποίες χρησιμοποιούνται από ένα πρόγραμμα ανήκουν σε ‘‘ομάδες’’ που

είναι πιθανό να μην αλλάξουν για μεγάλα χρονικά διαστήματα. Το φαινόμενο αυτό, το οποίο είναι γενικά γνωστό ως ‘‘ιδιότητα τοπικότητας’’, έχει δύο διαστάσεις :

- **χρονική τοπικότητα**, σύμφωνα με την οποία η πληροφορία που τώρα χρησιμοποιείται είναι πολύ πιθανό να χρειαστεί στο εγγύς μέλλον και
- **χωρική τοπικότητα**, σύμφωνα με την οποία οι διευθύνσεις αναφοράς του προγράμματος στο εγγύς μέλλον είναι πιθανό να βρίσκονται κοντά στις τρέχουσες διευθύνσεις αναφοράς.

2.1.1 Στοιχεία διαχείρισης κρυφής μνήμης

Τα στοιχεία που χαρακτηρίζουν ένα σύστημα διαχείρισης κρυφής μνήμης είναι ποικίλα. Τα σημαντικότερα από αυτά αναφέρονται στη συνέχεια [42] :

- **Αλγόριθμος ανάκλησης**

  Ο αλγόριθμος αυτός χρησιμοποιείται για να αποφασιστεί πότε μεταφέρονται σελίδες στην κρυφή μνήμη. Υπάρχουν διάφορες εναλλακτικές λύσεις. Οι σελίδες ανακαλούνται μόλις ζητηθούν (fetch on demand) ή προανακαλούνται (prefetch). Οι αλγόριθμοι προανάκλησης προσπαθούν να ”μαντέψουν” ποια πληροφορία θα χρειαστεί σύντομα ώστε να τη μεταφέρουν νωρίτερα στην κρυφή μνήμη.

- **Αλγόριθμος κατανομής της μνήμης [42]**

  Ο αλγόριθμος κατανομής της κρυφής μνήμης μοιράζει το διαθέσιμο χώρο της ανάμεσα στις σύνδρομες (concurrent) διεργασίες στη βάση.

  Οι αλγόριθμοι κατανομής διακρίνονται σε :

  -- **τοπικούς**, οι οποίοι παραχωρούν τμήματα της κρυφής μνήμης σε κάθε διεργασία, χωρίς να λαμβάνουν υπόψη τις αναφορές των υπόλοιπων διεργασιών. Οι τοπικοί αλγόριθμοι κατανομής χωρίζονται περαιτέρω σε **στατικούς**, στους οποίους το τμήμα της κρυφής μνήμης που ανήκει σε μια διεργασία παραμένει σταθερό σε όλη της τη διάρκεια και σε **δυναμικούς** στους οποίους το τμήμα της κρυφής μνήμης που αποδίδεται σε κάθε διεργασία μπορεί να μεγαλώνει και να μικραίνει ανάλογα με τη συμπεριφορά της διεργασίας ως προς τις αναφορές (αλγόριθμοι working-set [17, 18], page-fault-frequency [13]).

-- **καθολικούς**, οι οποίοι παραχωρούν τμήματα της κρυφής μνήμης στις διεργασίες λαμβάνοντας υπόψη τη συμπεριφορά αναφορών όλων των σύνδρομων διεργασιών.

-- αλγορίθμους οι οποίοι διαχωρίζουν τη μνήμη σε τμήματα, σε καθένα από τα οποία υπάρχει ένα μόνο είδος σελίδων.

- **Αλγόριθμος αναζήτησης στην κρυφή μνήμη [20]**

  Οποτεδήποτε μια διεργασία πραγματοποιήσει μια λογική αναφορά σε σελίδα, γίνεται αναζήτηση στην κρυφή μνήμη για τη σελίδα αυτή. Με δεδομένο ότι αυτό είναι ένα πολύ συχνό γεγονός, ο αλγόριθμος αναζήτησης πρέπει να είναι πολύ αποδοτικός σε ταχύτητα.

  Οι γνωστότεροι αλγόριθμοι αυτής της κατηγορίας είναι οι ακόλουθοι :

  -- **απευθείας αναζήτηση**, κατά την οποία εξετάζονται σειριακά όλες οι σελίδες της κρυφής μνήμης.

  -- **πίνακας μετάφρασης**, για την προσπέλαση του οποίου χρησιμοποιείται ο αριθμός της σελίδας. Ο πίνακας μετάφρασης μπορεί να είναι **αταξινόμητος** ή **ταξινομημένος**.

  -- **πίνακας κατακερματισμού**, για την προσπέλαση του οποίου χρησιμοποιείται αλγεβρικός μετασχηματισμός του αριθμού της σελίδας.

- **Αλγόριθμος αντικατάστασης**

  Στην αποτυχία (σφάλμα) μιας λογικής αναφοράς, εάν η κρυφή μνήμη είναι γεμάτη, απαιτείται επιλογή μιας σελίδας για αντικατάστασή της από αυτήν που εισάγεται. Αυτό αποτελεί αντικείμενο του αλγορίθμου αντικατάστασης [42].

  Βασικός στόχος κάθε αλγορίθμου αντικατάστασης είναι η ελαχιστοποίηση του ρυθμού σφαλμάτων στην κρυφή μνήμη, για δεδομένο μέγεθός της και δεδομένη κατανομή της.

  Οι διάφοροι αλγόριθμοι αντικατάστασης κατατάσσονται με βάση το κατά πόσο λαμβάνουν υπόψη τα ακόλουθα κριτήρια για την επιλογή της σελίδας που θα αντικατασταθεί (της σελίδας-‘‘θύματος’’) :

-- χρόνος από την πρώτη ή την τελευταία αναφορά στη σελίδα (π.χ. αλγόριθμος FIFO)

-- όλες ή οι πιο πρόσφατες αναφορές στη σελίδα (π.χ. αλγόριθμος LFU)

-- συνδυασμός των παραπάνω (π.χ. αλγόριθμοι LRU, CLOCK, GCLOCK, DG-CLOCK, LRD)

- **Αλγόριθμος ενημέρωσης των δεδομένων**

  Η πράξη της γραφής στα δεδομένα της κρυφής μνήμης δημιουργεί την απαίτηση ενημέρωσης του δίσκου για το γεγονός [42]. Η ενημέρωση αυτή ενδέχεται να πραγματοποιηθεί με δύο τρόπους :

  -- αμέσως μετά την ενημέρωση των στοιχείων της κρυφής μνήμης ή

  -- μετά την ολοκλήρωση της διεργασίας που προκάλεσε αλλαγή των δεδομένων

- **Αλγόριθμος διατήρησης της συνέπειας ανάμεσα σε πολλαπλές κρυφές μνήμες**

  Σε υπολογιστικό σύστημα το οποίο διαθέτει πολλές κρυφές μνήμες που αποθηκεύουν σελίδες ίδιου είδους, υπάρχει το πρόβλημα της εξασφάλισης ότι όλα τα αντίγραφα ενός τμήματος πληροφορίας είναι τα ίδια [42]. Για το σκοπό αυτό, είναι απαραίτητο να ενημερώνονται για την τροποποίηση δεδομένων όλες οι κρυφές μνήμες που διαθέτουν αντίγραφά τους. Οι μέθοδοι που εφαρμόζονται είναι :

  -- μετάδοση των αλλαγών σε όλες τις κρυφές μνήμες που διαθέτουν την ίδια πληροφορία

  -- ακύρωση, στις υπόλοιπες κρυφές μνήμες, του τμήματος πληροφορίας που υπέστη την αλλαγή

  -- απαγόρευση ύπαρξης, σε περισσότερες από μια κρυφές μνήμες, δεδομένων των οποίων επιτρέπεται η τροποποίηση

2.1.2 Διαχείριση κρυφής μνήμης σε συστήματα βάσεων δεδομένων

Τα συστήματα διαχείρισης βάσεων δεδομένων παρέχουν τους δικούς τους μηχανισμούς για τη χρησιμοποίηση της κρυφής μνήμης, παρά το γεγονός ότι τα λειτουργικά συστήματα διαθέτουν μηχανισμούς για διαχείριση κρυφής μνήμης αρχείων [20]. Αυτό

συμβαίνει εξαιτίας των διαφορετικών μοντέλων πρόσβασης στη μνήμη από μια διαδικασία διαχείρισης της βάσης, σε σχέση με τα σειριακά (συνήθως) μοντέλα πρόσβασης που παρουσιάζουν τα προγράμματα.

Σε οντοκεντρικές βάσεις δεδομένων, εξάλλου [27], η πρόσβαση στη βάση γίνεται ακολουθώντας δείκτες σε γράφο αντικειμένων, γεγονός που καθιστά ακόμη δυσκολότερη την υπαγωγή σε κάποιο μοντέλο πρόσβασης.

Μια ακόμη διαφορά που παρουσιάζεται σε βάσεις δεδομένων - οντοκεντρικές ή όχι - είναι το γεγονός ότι η διαχείρισή τους πραγματοποιείται μέσω δοσοληψιών στο τέλος των οποίων είναι απαραίτητη η ενημέρωση της βάσης για τυχόν τροποποιήσεις. Αυτό εμφανίζεται σε αντιδιαστολή με τα προγράμματα των οποίων οι τροποποιήσεις σε δεδομένα είναι άμεσα ορατές στο δίσκο.

Με βάση όλα τα παραπάνω, είναι φανερή η ανάγκη ξεχωριστής μελέτης του μηχανισμού διαχείρισης της κρυφής μνήμης για βάσεις δεδομένων, με τρόπο ώστε να μη μένουν ανεκμετάλλευτα (ή να εξυπηρετούνται αποδοτικότερα) τα ιδιαίτερα χαρακτηριστικά τους.

2.2 Αλγόριθμοι αντικατάστασης

2.2.1 Γνωστοί αλγόριθμοι

Ο αλγόριθμος αντικατάστασης σελίδων στην κρυφή μνήμη είναι ένα από τα σημαντικότερα ζητήματα για την επίτευξη ικανοποιητικών επιδόσεων. Για το λόγο αυτό, οι αλγόριθμοι που περιγράφονται στη βιβλιογραφία είναι πολυάριθμοι και αφορούν ποικίλες περιπτώσεις προγραμμάτων.

Οι γνωστότεροι από τους αλγορίθμους αυτούς είναι οι ακόλουθοι [20] :

- **RANDOM** : η σελίδα η οποία πρόκειται να αντικατασταθεί επιλέγεται τυχαία ανάμεσα σε αυτές που βρίσκονται στην κρυφή μνήμη. Η επίδοση του αλγορίθμου αυτού χρησιμοποιείται στις μετρήσεις ως τιμή της χειρότερης δυνατής επίδοσης.
- **FIFO** : η σελίδα η οποία πρόκειται να αντικατασταθεί είναι αυτή που πρώτη είχε εισαχθεί στην κρυφή μνήμη από αυτές που βρίσκονται εκεί. Ο αλγόριθμος αυτός είναι ιδανικός για σειριακή ακολουθία αναφορών.
- **LFU** : η σελίδα που αντικαθίσταται είναι αυτή με τη μικρότερη συχνότητα αναφορών.

- **LRU** : η σελίδα που αντικαθίσταται είναι αυτή που χρησιμοποιήθηκε λιγότερο πρόσφατα από αυτές που βρίσκονται στην κρυφή μνήμη. Με βάση τον ορισμό της χρησιμοποίησης μιας σελίδας, ο αλγόριθμος εμφανίζεται σε δύο εκδόσεις :

  -- Αν χρησιμοποίηση σελίδας θεωρείται η αναφορά σε αυτήν, τότε ο αλγόριθμος αντικαθιστά τη σελίδα η οποία πρώτη έγινε FIXed στην κρυφή μνήμη με βάση τον αλγόριθμο FIX-UNFIX [20].

  -- Αν χρησιμοποίηση σελίδας θεωρείται η ολοκληρωμένη χρησιμοποίησή της - η οποία περιλαμβάνει τις φάσεις FIX και UNFIX - τότε ο αλγόριθμος αντικαθιστά τη σελίδα η οποία πρώτη έγινε UNFIXed από τη μνήμη.

- **CLOCK (SECOND CHANCE)** : για την υλοποίηση του αλγορίθμου αυτού, με κάθε σελίδα που βρίσκεται στην κρυφή μνήμη σχετίζεται ένα bit χρήσης, ενώ υπάρχει κι ένας δείκτης της επόμενης θέσης στην κρυφή μνήμη της οποίας η σελίδα θα αντικατασταθεί. Οταν η σελίδα χρησιμοποιείται, το bit χρήσης της γίνεται 1. Οταν χρειαστεί η αντικατάσταση μιας σελίδας, εξετάζεται το bit χρήσης της σελίδας που βρίσκεται στη θέση που δείχνει ο δείκτης αντικατάστασης. Αν το bit χρήσης είναι 0, η σελίδα αντικαθίσταται. Διαφορετικά, το bit χρήσης τίθεται στην τιμή 0 και ο δείκτης προχωρά στην επόμενη θέση στον πίνακα των σελίδων της κρυφής μνήμης, ο οποίος θεωρούμε ότι έχει κυκλική οργάνωση. Με τον τρόπο αυτό, κάθε σελίδα έχει μια ‘‘δεύτερη ευκαιρία’’ παραμονής στην κρυφή μνήμη.

- **GCLOCK** : παραλλαγή του προηγούμενου αλγορίθμου, στην οποία το bit χρήσης αντικαθίσταται από έναν μετρητή αναφορών που αρχικοποιείται στην τιμή 1, αυξάνεται κάθε φορά που γίνεται αναφορά στη σελίδα και μειώνεται κάθε φορά που τον προσπερνά ο δείκτης αντικατάστασης. Η αντικατάσταση μιας σελίδας πραγματοποιείται και πάλι όταν ο μετρητής αναφορών της μειωθεί στην τιμή 0.

- **LRD** : η σελίδα η οποία πρόκειται να αντικατασταθεί, είναι αυτή με τη μικρότερη ‘‘πυκνότητα αναφορών’’ στο σύνολο των αναφορών στις σελίδες της κρυφής μνήμης.

- **Αλλοι αλγόριθμοι** :

  Εκτός από τους παραπάνω, στη βιβλιογραφία παρουσιάζονται αλγόριθμοι οι οποίοι επιλέγουν σελίδες για αντικατάσταση με βάση κάποιο (γνωστό) μοντέλο προσπελάσεων κάθε διεργασίας [10, 13, 31, 39]. Η μέθοδος αυτή ενδείκνυται -

και παρουσιάζει πολύ καλά αποτελέσματα - σε συστήματα στα οποία η εκτέλεση διεργασιών επαναλαμβάνεται - συνεχώς ή σε τακτά χρονικά διαστήματα (π.χ. τραπεζικές δοσοληψίες).

Οι παραπάνω αλγόριθμοι μελετήθηκαν και εφαρμόζονται κύρια στην επεξεργασία δοσοληψιών (transaction processing) όπου ανάλογα με το μοντέλο προσπελάσεων (access pattern) κάθε κατηγορίας προγραμμάτων, επιλέγεται ο κατάλληλος για αυτά αλγόριθμος αντικατάστασης.

2.2.2 Εφαρμογή σε γνωστά συστήματα διαχείρισης οντοκεντρικών βάσεων δεδομένων

Παρά το μεγάλο πλήθος αλγορίθμων αντικατάστασης που περιγράφονται στη βιβλιογραφία, τόσο αυτών που περιγράφονται παραπάνω όσο και διάφορων παραλλαγών τους, η εφαρμογή τους στα συστήματα διαχείρισης οντοκεντρικών βάσεων δεδομένων είναι περιορισμένη.

Η διάκριση, στο σημείο αυτό, ανάμεσα σε σχεσιακές και οντοκεντρικές βάσεις δεδομένων οφείλεται στο γεγονός ότι, εξαιτίας συγκεκριμένων πράξεων που είναι δυνατό να πραγματοποιηθούν σε σχεσιακές βάσεις δεδομένων - π.χ η πράξη της φυσικής σύνδεσης (natural join) - υπάρχει η δυνατότητα προσδιορισμού συγκεκριμένων μοντέλων προσπέλασης και εκμετάλλευσής τους για την επιλογή του καταλληλότερου αλγορίθμου αντικατάστασης σελίδων στην κρυφή μνήμη. Αντίθετα, οι αναφορές στις οντοκεντρικές βάσεις δεδομένων είναι τυχαίες και δεν παρουσιάζουν μοντέλο προσπέλασης.

Επιπλέον, τα σύγχρονα υπολογιστικά συστήματα διαθέτουν κατά κανόνα ικανοποιητικό αποθηκευτικό χώρο στη μνήμη, γεγονός που σημαίνει ότι η χωρητικότητα της κρυφής μνήμης είναι αρκετά σημαντική.

Για το λόγο αυτό, συστήματα όπως τα O_2 [37, 38], EXODUS [7], Objectstore [36], ORION [32], δε διαθέτουν άλλον αλγόριθμο αντικατάστασης στην κρυφή μνήμη από τον LRU, ενώ οι προσπάθειες βελτίωσης της επίδοσης των συστημάτων προσανατολίζεται στον έλεγχο άλλων παραγόντων καθυστέρησης.

2.3 Αρχιτεκτονικές πελάτη-εξυπηρετητή

Ενα σύστημα πελάτη-εξυπηρετητή είναι μια κατανεμημένη αρχιτεκτονική λογισμικού που εκτελείται σε ένα δίκτυο από υπολογιστές-πελάτες και κοινούς υπολογιστές-εξυπηρετητές [23]. Οι διεργασίες-πελάτες εκτελούνται στους υπολογιστές-πελάτες, όπου παρέχεται διαλογική επικοινωνία με το χρήστη. Οι διεργασίες-εξυπηρετητές είναι κοινοί πόροι που παρέχουν υπηρεσίες κατά παραγγελία των πελατών.

Ο διαχωρισμός των ευθυνών που υποστηρίζεται στο μοντέλο πελάτη-εξυπηρετητή ικανοποιεί τις ανάγκες των χρηστών ενώ, παράλληλα, παρέχει υποστήριξη για τη διαχείριση, το συντονισμό και τον έλεγχο των κοινών πόρων του συστήματος. Ετσι, το μοντέλο πελάτη-εξυπηρετητή αποτελεί έναν αποτελεσματικό τρόπο δόμησης κατανεμημένων συστημάτων.

2.3.1 Το μοντέλο πελάτη-εξυπηρετητή για συστήματα διαχείρισης οντοκεντρικών βάσεων δεδομένων

Η συγχώνευση του μοντέλου πελάτη-εξυπηρετητή και των συστημάτων διαχείρισης οντοκεντρικών βάσεων δεδομένων έχει δημιουργήσει μια νέα τάξη συστημάτων βάσεων δεδομένων. Τα συστήματα αυτά χρησιμοποιούν το μοντέλο πελάτη-εξυπηρετητή για την παροχή αμεσότητας στο χρήστη [1] και την υποστήριξη πολύπλοκων κοινών δεδομένων σε ένα κατανεμημένο περιβάλλον.

Οι αρχιτεκτονικές πελάτη-εξυπηρετητή για συστήματα διαχείρισης βάσεων δεδομένων κατηγοριοποιούνται με βάση τη μονάδα επικοινωνίας ανάμεσα στον πελάτη και τον εξυπηρετητή [23]. Ετσι, οι πελάτες είναι δυνατό να στέλνουν αιτήσεις για δεδομένα στον εξυπηρετητή είτε ως ερωτήσεις, οποότε αναφερόμαστε σε συστήματα μεταφοράς ερωτήσεων (query-shipping systems), είτε ως αιτήσεις για συγκεκριμένα αντικείμενα πληροφορίας, οπότε αναφερόμαστε σε συστήματα μεταφοράς πληροφορίας (data-shipping systems).

Στην πρώτη από τις παραπάνω κατηγορίες, ο πελάτης στέλνει την ερώτηση στον εξυπηρετητή, ο οποίος την εκτελεί και στέλνει τα αποτελέσματα στον πελάτη. Στη δεύτερη από αυτές, η ερώτηση εκτελείται στον πελάτη, ο οποίος ζητά από τον

[1] ο οποίος δεν αναγκάζεται να εξοικειωθεί με το πολύπλοκο, ενδεχομένως, λογισμικό του εξυπηρετητή

εξυπηρετητή τα δεδομένα που απαιτούνται για την ολοκλήρωση αυτής της διεργασίας. Τα συστήματα μεταφοράς δεδομένων διαχωρίζονται περαιτέρω σε συστήματα μεταφοράς σελίδων (page-server systems) στα οποία η επικοινωνία πραγματοποιείται χρησιμοποιώντας φυσικές μονάδες αντικειμένων (π.χ. σελίδες ή τμήματα) και συστήματα μεταφοράς αντικειμένων (object-server systems) στα οποία μεταφέρονται λογικές μονάδες αντικειμένων (π.χ. πλειάδες αντικειμένων).

Με δεδομένη τη συνεχή βελτίωση των υπολογιστικών συστημάτων που λειτουργούν σαν πελάτες, οι αρχιτεκτονικές μεταφοράς δεδομένων προτιμούνται από αυτές μεταφοράς ερωτήσεων [2]. Για το λόγο αυτό, η πλειοψηφία υπολογιστικής ισχύος και μνήμης του συστήματος μεταφέρεται στον υπολογιστή-πελάτη. Παρά το γεγονός ότι ο υπολογιστής-εξυπηρετητής είναι σχεδόν σίγουρο ότι διαθέτει περισσότερη μνήμη και υπολογιστική ισχύ από οποιονδήποτε μεμονωμένο υπολογιστή, η συνολική μνήμη και υπολογιστική ισχύς των υπολογιστών-πελατών του συστήματος θα είναι μεγαλύτερη από αυτήν του εξυπηρετητή. Επομένως, μεταφέροντας λειτουργικότητα από τον εξυπηρετητή προς τους πελάτες μπορεί να βελτιωθεί σημαντικά η επίδοση του συστήματος, καθώς ο πρώτος αποτελεί συχνά σημείο συσσώρευσης.

2.3.2 Διαχείριση κρυφής μνήμης στον πελάτη

Στα πλαίσια της μεταφοράς λειτουργικότητας από τους εξυπηρετητές στους πελάτες, μελετάται εδώ η χρησιμοποίηση της κρυφής μνήμης των πελατών για την προσωρινή αποθήκευση δεδομένων απαραίτητων για την εκτέλεση ερωτήσεων από τον υπολογιστή-πελάτη.

Η αποθήκευση δεδομένων στην κρυφή μνήμη των πελατών, έχει τα ακόλουθα αποτελέσματα :

- **δυναμική δημιουργία και καταστροφή αντιγράφων** σελίδων από τους πελάτες, χωρίς την ανάγκη αναδιοργάνωσης του φυσικού σχεδιασμού της βάσης δεδομένων, όπως συμβαίνει σε συστήματα όπου υποστηρίζεται στατική δημιουργία αντιγράφων.
- **ιδιοκτησία δευτέρου επιπέδου** κατά την οποία τα αντίγραφα των σελίδων που αποθηκεύονται στους πελάτες είναι αποδεκτό να έχουν υποστεί κάποιου βαθμού

[2] το συμπέρασμα αυτό, στο οποίο καταλήγουν πειραματικά αποτελέσματα μελετών, έρχεται σε σύγκρουση με το βιομηχανικό πρότυπο που αφορά το ίδιο ζήτημα, σύμφωνα με το οποίο τα συστήματα μεταφοράς ερωτήσεων αποτελούν απαίτηση

τροποποίηση σε σχέση με τις πραγματικές σελίδες, οι οποίες αποθηκεύονται στους εξυπηρετητές. Ετσι, οι σελίδες στους πελάτες μπορούν να καταστραφούν οποιαδήποτε στιγμή χωρίς να προκαλέσουν την απώλεια δεδομένων. Με τον τρόπο αυτό, εξασφαλίζεται η αξιοπιστία του συστήματος [1].

Τα παραπάνω στοιχεία, σε συνδυασμό με την εκμετάλλευση της υπολογιστικής ισχύος των υπολογιστών-πελατών, καθιστούν φανερό ότι η χρησιμοποίηση της κρυφής μνήμης των υπολογιστών-πελατών μπορεί να αποδειχθεί πολύ σημαντική για την αύξηση της υπολογιστικής ισχύος των συστημάτων.

Ιδιαίτερη σημασία έχει πάντα η διατήρηση της συνέπειας των δεδομένων, στοιχείο πάντοτε σημαντικό σε συστήματα διαχείρισης κρυφής μνήμης στα οποία υπάρχουν αντίγραφα δεδομένων. Το στοιχείο αυτό έχει αποτελέσει αντικείμενο εκτεταμένης μελέτης, η οποία παρουσιάζεται σε επόμενο κεφάλαιο.

2.4 Διατήρηση της συνέπειας

2.4.1 Γενικά

Το ζήτημα της διατήρησης της συνέπειας τίθεται από τη στιγμή που οι κρυφές μνήμες στις οποίες αποθηκεύονται δεδομένα της ίδιας βάσης γίνουν περισσότερες από μία. Στην περίπτωση αυτή, ο σχεδιασμός του συστήματος διαχείρισης κρυφής μνήμης πρέπει να είναι τέτοιος ώστε να εξασφαλίζεται η ταύτιση αντιγράφων του ίδιου τμήματος πληροφορίας, όταν αυτά πρόκειται να χρησιμοποιηθούν [23].

Οπως είναι γνωστό από τη θεωρία βάσεων δεδομένων, όταν σε μια βάση πραγματοποιούνται πολλές ταυτόχρονες αλλαγές, η ενημέρωση των δεδομένων πρέπει να γίνει όπως αν αυτές είχαν πραγματοποιηθεί σειριακά. Η αρχή αυτή ονομάζεται αρχή σειριακότητας (serializability) και τηρείται σε όλα τα συστήματα διαχείρισης βάσεων δεδομένων για να διατηρηθεί η συνέπεια της πληροφορίας.

Η προέκταση της έννοιας της σειριακότητας σε συστήματα στα οποία επιτρέπεται η ύπαρξη αντιγράφων πληροφορίας, ονομάζεται σειριακότητα ενός αντιγράφου (one-copy serializability). Η αθέατη για τους χρήστες επανάληψη πληροφορίας, δημιουργεί την απαίτηση η εκτέλεση μιας σειράς διεργασιών στη βάση να πρέπει να εμφανίζεται σαν σειριακή εκτέλεση των ίδιων διεργασιών σε βάση χωρίς αντίγραφα.

Για την επίτευξη αυτού του στόχου, είναι απαραίτητο, η τροποποίηση ενός από τα αντίγραφα, να αντικατοπτρίζεται σε όλα τα υπόλοιπα.

Στη συνέχεια, παρουσιάζονται σε μεγαλύτερη λεπτομέρεια συγκεκριμένοι αλγόριθμοι διατήρησης της συνέπειας.

2.4.2 Αλγόριθμοι διατήρησης της συνέπειας

Οι αλγόριθμοι διατήρησης της συνέπειας της κρυφής μνήμης που εμφανίζονται στη βιβλιογραφία κατηγοριοποιούνται με βάση διάφορα κριτήρια [23]. Λαμβάνοντας υπόψη, ωστόσο, ότι κύριο αντικείμενό τους είναι η εξασφάλιση μη ύπαρξης άκυρης πληροφορίας στη βάση - άκυρη θεωρείται η πληροφορία που είναι παλαιότερη από την τελευταία ενημέρωση της ίδιας πληροφορίας από ολοκληρωμένη διεργασία - προκύπτει ότι είναι λογικό να τους διαχωρίσουμε με βάση την τακτική που ακολουθούν στον τομέα αυτό. Ετσι, οι αλγόριθμοι διατήρησης της συνέπειας της κρυφής μνήμης διακρίνονται σε :

- **Πρωτόκολλα ανίχνευσης** , που επιτρέπουν την ύπαρξη άκυρης (stale) πληροφορίας σε κρυφή(ές) μνήμη(ες). Επομένως, οι διεργασίες οφείλουν να ελέγχουν την εγκυρότητα των δεδομένων που χρησιμοποιούν ώστε να τους επιτραπεί να τερματίσουν, ενώ ο εξυπηρετητής (server) είναι υπεύθυνος για τη διατήρηση πληροφορίας που θα επιτρέπει την πραγματοποίηση του ελέγχου.

- **Πρωτόκολλα αποφυγής** , που δεν επιτρέπουν την ύπαρξη άκυρης πληροφορίας. Τα πρωτόκολλα αυτά βασίζονται στην αρχή ROWA (read one/write all), σύμφωνα με την οποία η ανάγνωση πληροφορίας γίνεται ελεύθερα, αλλά για την ενημέρωση πληροφορίας πρέπει να ενημερωθούν όλα τα αντίστοιχα αντίγραφα για να επιτραπεί στη διεργασία να τερματίσει.

- **Οιονεί-αντίγραφα** : σε συστήματα που υποστηρίζεται η έννοια των Οιονεί-αντιγράφων αποθηκεύεται στην περιοχή του χρήστη πληροφορία που αυτός επιλέγει. Τα οιονεί-αντίγραφα αποτελούν τμήματα πληροφορίας, τα οποία σχετίζονται με την (οπωσδήποτε έγκυρη) πληροφορία που διαθέτει ο εξυπηρετητής, με συγκεκριμένες συνθήκες που καθορίζουν το χρονικό περιθώριο που δίνεται για την ενημέρωση του αντιγράφου, όταν τροποποιηθούν τα δεδομένα του εξυπηρετητή. Στα συστήματα αυτά, η τροποποίηση δεδομένων είναι τελείως κεντρικοποιημένη.

Είναι φανερό ότι η τεχνική που βασίζεται στην ανίχνευση είναι απλούστερη από αυτή που βασίζεται στην αποφυγή. Κι αυτό, γιατί στα πρωτόκολλα ανίχνευσης οι διεργασίες-πελάτες επικοινωνούν με τη διεργασία-εξυπηρετητή μόνο όταν αυτές το θέλουν [3] και όχι οποτεδήποτε πραγματοποιηθεί μια ενημέρωση.

Εξάλλου, με τη χρήση πρωτοκόλλων ανίχνευσης αποφεύγεται η απαίτηση χρησιμοποίησης κατανεμημένων αλγορίθμων για την επίλυση προβλημάτων που προκύπτουν σε ανάλογες περιπτώσεις. Εάν, για παράδειγμα, πρέπει να πραγματοποιηθεί η ανίχνευση αδιεξόδου, αυτή μπορεί να πραγματοποιηθεί στον πελάτη, ο οποίος διαθέτει κάθε στιγμή όλη την απαραίτητη πληροφορία.

Βασικό μειονέκτημα της χρησιμοποίησης αλγορίθμων ανίχνευσης, ωστόσο, είναι η μεγάλη εξάρτηση από τη διεργασία-εξυπηρετητή για την πραγματοποίηση οποιασδήποτε πράξης στα δεδομένα. Κι αυτό, γιατί στα πρωτόκολλα αποφυγής, οι ενημερώσεις σχετικά με την τροποποίηση δεδομένων πραγματοποιείται απευθείας ανάμεσα στους υπολογιστές-πελάτες, γεγονός που συμβάλλει στην ευελιξία του συστήματος.

Σε γενικές γραμμές, η χρήση του ενός από τα δύο πρωτόκολλα δεν αποκλείει το άλλο. Οι συνδυασμοί πρωτοκόλλων δίνουν συχνά τη χρυσή τομή που απαιτείται για την αποδοτική λειτουργία του συστήματος.

Detection-based πρωτόκολλα

Εξαιτίας της κεντρικοποιημένης μορφής τους, τα πρωτόκολλα που βασίζονται στην ανίχνευση αποτελούν επεκτάσεις των παραδοσιακών αλγορίθμων κλειδώματος σε δύο φάσεις ή ‘‘αισιόδοξου’’ ελέγχου συνδρομικότητας.

Οι περιπτώσεις πρωτοκόλλων αυτής της κατηγορίας διαφοροποιούνται ως προς τις ακόλουθες παραμέτρους :

- **Εναρξη πράξης συνέπειας** :

  Ο έλεγχος της εγκυρότητας της πληροφορίας που προσπελαύνεται μπορεί να πραγματοποιηθεί σε διαφορετικά χρονικά σημεία της εκτέλεσης μιας διεργασίας (ή και να μην πραγματοποιηθεί). Με δεδομένο ότι η εγκυρότητα μπορεί να ελεγχθεί οποιαδήποτε στιγμή μέχρι την ολοκλήρωση της διεργασίας, οι στρατηγικές που μπορούν να ακολουθηθούν είναι οι ακόλουθες :

[3] κυρίως για την πραγματοποίηση των απαραίτητων ελέγχων στα δεδομένα τους ώστε να τους επιτραπεί να τερματίσουν

-- **Σύγχρονες**, στις οποίες ο έλεγχος εγκυρότητας της πληροφορίας πρέπει να ολοκληρωθεί πριν την προσπέλαση της πληροφορίας. Οταν το τμήμα πληροφορίας κατοχυρωθεί σαν έγκυρο, η μέθοδος θεωρεί ότι παραμένει έτσι μέχρι την - ομαλή ή μη - ολοκλήρωση της διεργασίας. Αυτό επιτυγχάνεται με τη βοήθεια του εξυπηρετητή, ο οποίος δεν επιτρέπει την ολόκλήρωση πριν από αυτήν άλλης διεργασίας που επίσης χρησιμοποιεί την πληροφορία.

-- **Ασύγχρονες**, στις οποίες ο έλεγχος εγκυρότητας ξεκινά μαζί με την προσπέλαση της πληροφορίας, χωρίς ωστόσο αυτή να περιμένει το αποτέλεσμά του. Αν ο έλεγχος εγκυρότητας επιστρέψει αρνητική τιμή, η διεργασία οφείλει να τερματίσει μη ομαλά.

-- **Καθυστερημένου ελέγχου**, στις οποίες ο έλεγχος εγκυρότητας πραγματοποιείται όταν η διεργασία μπει στη φάση τερματισμού της, για το σύνολο των δεδομένων που έχει προσπελάσει.

Από τα παραπάνω, είναι φανερό ότι η ‘‘αισιοδοξία’’ των στρατηγικών ασύγχρονου και καθυστερημένου ελέγχου είναι δυνατό να οδηγήσουν πολλές φορές σε μη ομαλό τερματισμό των διαδικασιών. Οταν, όμως, δεν παρατηρείται εκτεταμένη χρήση κοινής πληροφορίας, οι τεχνικές αυτές προσφέρουν το πλεονέκτημα σημαντικής οικονομίας χρόνου.

- **Μηνύματα κοινοποίησης αλλαγών** :

  Κατά την πράξη τροποποίησης δεδομένων από μια διεργασία, στην περίπτωση χρησιμοποίησης των δεδομένων από τις υπόλοιπες, ειδοποιούνται αυτές που διαθέτουν κοινά τμήματα πληροφορίας. Η ειδοποίηση αυτή πραγματοποιείται ώστε να αποφεύγονται κατά το δυνατό οι μη ομαλοί τερματισμοί διεργασιών και στέλνεται ασύγχρονα.

  Για να αποφευχθεί η περίπτωση κατά την οποία, η διεργασία που κοινοποιεί τις ενημερώσεις, τερματίσει μη ομαλά παρασύροντας σε λανθασμένες προσβάσεις τις διεργασίες που έχουν λάβει την ειδοποίησή της, οι κοινοποιήσεις δεν πραγματοποιούνται παρά μόνο κατά τον ομαλό τερματισμό της διεργασίας.

- **Υλοποίηση ενημερώσεων** :

  Η λήψη μηνύματος ενημέρωσης από μια διεργασία που διαχειρίζεται κοινά δεδομένα,

σημαίνει ότι πρέπει να γίνει μια κίνηση προκειμένου να διατηρηθεί η συνέπεια. Η κίνηση αυτή μπορεί να είναι :

-- **Μετάδοση** των δεδομένων στη νέα τους μορφή σε όλες τις διεργασίες που τα χρησιμοποιούν.

-- **Ακύρωση** των δεδομένων που έχουν τροποποιηθεί από όλες τις υπόλοιπες διεργασίες που τα χρησιμοποιούν.

-- **Συνδυασμός** των δύο παραπάνω με εφαρμογή ευρηματικού αλγορίθμου.

Η επιλογή ης κατάλληλης για το σύστημα στρατηγικής σε καθέναν από τους παραπάνω τομείς, οδηγεί στη διαμόρφωση μιας ποικιλίας αλγορίθμων, που θα παρουσιαστούν παρακάτω.

Avoidance-based πρωτόκολλα

Τα πρωτόκολλα αποφυγής βασίζονται στο γεγονός ότι η πληροφορία που αποθηκεύεται σε μια κρυφή μνήμη είναι πάντα έγκυρη. Για να επιτευχθεί αυτός ο στόχος, είναι απαραίτητο για κάθε διεργασία-πελάτη να επεμβαίνει στα περιεχόμενα της κρυφής μνήμης άλλης διεργασίας-πελάτη, κατά την ενημέρωση ενός δεδομένου. Η απαίτηση αυτή επιβάλλει τη διεύρυνση των δυνατοτήτων των διεργασιών-πελατών με πρόσθετες λειτουργίες, οι οποίες αυξάνουν την πολυπλοκότητα του λογισμικού τους.
Στην πλευρά της διεργασίας-εξυπηρετητή, από την άλλη μεριά, τα πρωτόκολλα αποφυγής απαιτούν την ύπαρξη πλήρους πληροφορίας σχετικά με τις θέσεις κάθε σελίδας στο σύστημα, γεγονός που δημιουργεί έναν πρόσθετο φόρτο.
Παρά τις πρόσθετες ανάγκες που δημιουργούν, ωστόσο, οι αλγόριθμοι αποφυγής είναι πιο διαδεδομένοι από αυτούς της ανίχνευσης, καθώς εγγυώνται την εγκυρότητα των δεδομένων που χρησιμοποιούνται, οδηγώντας έτσι σε λιγότερους μη ομαλούς τερματισμούς διεργασιών.
Η κατηγοριοποίηση αυτής της ομάδας αλγορίθμων πραγματοποιείται με βάση τα ακόλουθα κριτήρια :

- **Δήλωση πρόθεσης γραφής** :

  Οπως και στην περίπτωση των αλγορίθμων ανίχνευσης, το πρώτο κριτήριο κατηγοριοποίησης αφορά στη στιγμή κατά την οποία αρχίζει η πράξη συνέπειας σε

σχέση με την πρόσβαση στα δεδομένα. Για την ανάγνωση δεδομένων, όταν αυτά δε βρίσκονται στην κρυφή μνήμη - αφού όταν βρίσκονται είναι σίγουρο πως είναι έγκυρα - οι αντίστοιχες σελίδες ζητούνται από τον εξυπηρετητή. Οταν αυτές δοθούν στον πελάτη, δίνεται έμμεσα και η άδεια ανάγνωσής τους. Για την τροποποίηση δεδομένων, ζητείται άδεια από τον εξυπηρετητή είτε εκ μέρους μιας διεργασίας, είτε εκ μέρους του υπολογιστή-πελάτη για όλες τις διεργασίες που πρόκειται να εκτελεστούν εκεί.

Σε κάθε περίπτωση, η άδεια τροποποίησης δεδομένων μπορεί να ζητηθεί :

-- **Σύγχρονα** οπότε η διεργασία περιμένει την άδεια του εξυπηρετητή πριν την τροποποίηση.

-- **Ασύγχρονα** οπότε η διεργασία δεν περιμένει την άδεια από τον εξυπηρετητή.

-- Στη φάση ολοκλήρωσης της διεργασίας.

Εδώ πρέπει να σημειωθεί ότι σε κάθε περίπτωση τροποποίησης δεδομένων, τα μέτρα για τη διατήρηση της συνέπειας στους υπόλοιπους πελάτες πρέπει να υλοποιηθούν πριν την ολοκλήρωση της διεργασίας, για να θεωρηθεί αυτή επιτυχής.

- **Διάρκεια πρόθεσης γραφής** :

  Ανάλογα με το εάν η άδεια γραφής ζητείται εκ μέρους της διεργασίας ή του υπολογιστή-πελάτη, καθορίζεται η διάρκεια της πρόθεσης γραφής, καθώς όταν η άδεια γραφής δοθεί, διαρκεί όσο το δεδομένο βρίσκεται στην κρυφή μνήμη.

- **Πράξη αντιμετώπισης σύγκρουσης** :

  Με την άφιξη μιας πράξης συνέπειας από κάποια διεργασία, υπάρχουν δύο δυνατότητες για τη διεργασία-παραλήπτη σε σχέση με τις πράξεις της που είναι σε σύγκρουση με την πράξη συνέπειας :

  -- Να θέσει την πράξη συνέπειας σε αναμονή (**wait-policy**) έως ότου ολοκληρωθεί η τρέχουσα διεργασία ή

  -- Να εγκαταλείψει την εκτέλεσή της (**preempt-policy**) εφόσον υπάρχει σύγκρουση ανάμεσα σε αυτήν και την πράξη συνέπειας.

Στην πλειοψηφία των γνωστών αλγορίθμων, ακολουθείται η πρώτη από τις παραπάνω μεθόδους, καθώς αυτή εξασφαλίζει την ολοκλήρωση μιας διεργασίας, άσχετα από τα εισερχόμενα μηνύματα ενημερώσεων.

- **Υλοποίηση ενημερώσεων** :

  Οπως και στην προηγούμενη κατηγορία πρωτοκόλλων, η ενημέρωση για την τροποποίηση δεδομένων μπορεί να πραγματοποιηθεί είτε ακυρώνοντας τα κοινά δεδομένα στις κρυφές μνήμες των υπόλοιπων πελατών, είτε μεταδίδοντας τις τροποποιήσεις σε όλες τις κρυφές μνήμες, είτε επιλέγοντας δυναμικά ένα από τα παραπάνω, ανάλογα με την περίπτωση.

Γνωστοί αλγόριθμοι

Στη συνέχεια παρουσιάζονται ορισμένοι γνωστοί αλγόριθμοι διατήρησης της συνέπειας που ανήκουν στις παραπάνω κατηγορίες. Οι ομάδες στις οποίες ταξινομούνται οι αλγόριθμοι αυτοί είναι οι ακόλουθες :

- **Server-based two-phase locking (S2PL)** :

  Οι αλγόριθμοι αυτής της κατηγορίας είναι σύγχρονοι αλγόριθμοι ανίχνευσης οι οποίοι δε χρησιμοποιούν μηνύματα ενημέρωσης.

  Για την υλοποίηση ενός τέτοιου αλγορίθμου, θεωρείται ότι κάθε κομμάτι πληροφορίας έχει ένα έγκυρο αντίγραφο, το οποίο θεωρείται τοποθετημένο στην κρυφή μνήμη του υπολογιστή-εξυπηρετητή. Πριν από την ανάγνωση ενός δεδομένου, η διεργασία είναι υποχρεωμένη να συγκρίνει τα δεδομένα της με αυτά του εξυπηρετητή για να σιγουρευτεί για την εγκυρότητά τους. Ομοια, μετά την ενημέρωση ενός αντικειμένου, η νέα του τιμή πρέπει να εγκατασταθεί στην κρυφή μνήμη του εξυπηρετητή.

  Σε αυτήν την κατηγορία, ανήκουν αλγόριθμοι όπως οι ακόλουθοι [23, 26] :

  -- **Basic two-phase locking (B2PL)** : στον αλγόριθμοι αυτό, τα δεδομένα δεν επιτρέπεται να παραμένουν στις κρυφές μνήμες των πελατών μετά την ολοκλήρωση μιας διεργασίας. Για την ανάγνωση δεδομένου, η διεργασία ζητά κλείδωμα ανάγνωσης από τον εξυπηρετητή, το οποίο μπορεί αργότερα να

αναβαθμιστεί σε κλείδωμα γραφής. Ολα τα κλειδώματα διατηρούνται μέχρι την ολοκλήρωση της διεργασίας.

-- **Caching two-phase locking (C2PL)** : στον αλγόριθμο αυτό, τα δεδομένα επιτρέπεται να διατηρούνται στις κρυφές μνήμες των πελατών και μετά την ολοκλήρωση μιας διεργασίας. Οι υπόλοιπες λειτουργίες του πραγματοποιούνται όπως παραπάνω, ενώ τα κλειδώματα διατηρούνται για όσο χρόνο το δεδομένο βρίσκεται στην κρυφή μνήμη.

- **Optimistic two-phase locking (O2PL)** :

  Οι αλγόριθμοι αυτής της κατηγορίας ανήκουν στα πρωτόκολλα αποφυγής ενώ χαρακτηρίζονται σαν ‘‘αισιόδοξοι’’ γιατί αναβάλλουν όλες τις πράξεις συνέπειας για το τέλος της διεργασίας. Επιπλέον, οι αλγόριθμοι αυτοί υποστηρίζουν διατήρηση δεδομένων στην κρυφή μνήμη μετά την ολοκλήρωση μιας διεργασίας. Τέλος, κάθε πελάτης - σε συστήματα όπου εφαρμόζεται αυτό το πρωτόκολλο - διαθέτει έναν πλήρη διαχειριστή κλειδαριών (lock manager).

  Οι αλγόριθμοι αυτής της κατηγορίας διακρίνονται με βάση τη μέθοδο που ακολουθούν για την κοινοποίηση των τροποποιήσεων των δεδομένων στις οποίες προβαίνουν. Οι κατηγορίες που υπάρχουν χρησιμοποιούν μια από τις ακόλουθες μεθόδους - δίνοντας και το ανάλογο όνομα στους αντίστοιχους αλγορίθμους :

  -- **Ακύρωση** των δεδομένων σε απομακρυσμένες κρυφές μνήμες (O2PL-I).

  -- **Ενημέρωση** των αντίστοιχων δεδομένων (O2PL-P).

  -- **Δυναμική επιλογή** ανάμεσα σε ακύρωση ή ενημέρωση των δεδομένων με βάση ευρετικό αλγόριθμο (O2PL-D).

- **Callback locking (CBL)** :

  Οι αλγόριθμοι αυτής της κατηγορίας ανήκουν και πάλι στους αλγόριθμους αποφυγής, αλλά χρησιμοποιούν σύγχρονες μεθόδους για τις πράξεις συνεπειας [24, 25]. Η ιδιαιτερότητα των αλγορίθμων αυτής της κατηγορίας είναι ότι, για να δοθεί σε μια διεργασία δικαίωμα τροποποίησης δεδομένων, πρέπει αυτά τα δεδομένα να ανακληθούν προηγουμένως από όλους τους υπόλοιπους πελάτες.

  Οι ομάδες αλγορίθμων που εμφανίζονται εδώ είναι :

-- **Callback-Read (CBL-R)**, στην οποία τα δεδομένα παραμένουν στην κρυφή μνήμη μόνο στη διάρκεια μιας διεργασίας και

-- **Callback-All (CBL-A)**, στην οποία τα δεδομένα μπορούν να παραμείνουν στην κρυφή μνήμη και μετά την ολοκλήρωση μιας διεργασίας.

- **Τεχνικές διαχείρισης συνολικής μνήμης** :

Στα πλαίσια της εκμετάλλευσης των πόρων που διαθέτουν οι υπολογιστές- πελάτες [25], αυτά τα συστήματα χαρακτηρίζονται από ιεραρχία μνήμης 4 επιπέδων :

1. Τοπική μνήμη πελάτη
2. Μνήμη εξυπηρετητή
3. Μνήμη απομακρυσμένου πελάτη
4. Δίσκος

Με άλλα λόγια, στα συστήματα αυτά, όλη η παραπάνω μνήμη θεωρείται προσπελάσιμη από όλους - με ορισμένους πάντοτε κανόνες. Οι τεχνικές που χρησιμοποιούνται είναι οι ακόλουθες :

-- **Προώθηση (forwarding)** : Η αίτηση σελίδας προς τον εξυπηρετητή, προωθείται από αυτόν σε πελάτη που είναι γνωστό ότι τη διαθέτει

-- **Hate hints** : για τη διατήρηση μεγαλύτερου ποσοστού της βάσης στη συνολική μνήμη του συστήματος - όπως αυτή ορίστηκε παραπάνω - κατά την αποθήκευση σελίδας στην κρυφή μνήμη υπολογιστή-πελάτη, αυτή σημειώνεται ώστε να είναι η πρώτη που θα απομακρυνθεί από την κρυφή μνήμη του εξυπηρετητή, εφόσον βρίσκεται εκεί

-- **Dropped pages** : μοναδικά αντίγραφα σελίδων που πρόκειται να διαγραφούν από την κρυφή μνήμη πελάτη-υπολογιστή, τοποθετούνται στην κρυφή μνήμη του εξυπηρετητή.

2.4.3 Αλγόριθμοι διατήρησης της συνέπειας σε γνωστά συστήματα

Η διατήρηση της συνέπειας των δεδομένων που αποθηκεύονται στις κρυφές μνήμες των υπολογιστών είναι ένα πολύ σημαντικό ζήτημα, γεγονός που δικαιολογεί την εκτεταμένη μελέτη που εμφανίζεται στον αντίστοιχο τομέα [23, 24, 25, 26]. Η ποικιλία των

αλγορίθμων που εμφανίζονται στη βιβλιογραφία, ωστόσο, σε καμμία περίπτωση δεν αντικατοπτρίζει την κατάσταση που υπάρχει στα γνωστά εμπορικά συστήματα, όπως τα ORION [32], EXODUS [7], Objectstore [36] και O2 [37, 38]. Στα συστήματα αυτά, η διατήρηση της συνέπειας πραγματοποιείται με συνδυασμό σύγχρονου κλειδώματος των δεδομένων - σε συνεννόηση με τον εξυπηρετητή - και ακύρωσης των δεδομένων που τροποποιούνται από άλλη διεργασία-πελάτη.

Η απλότητα των αλγορίθμων που χρησιμοποιούνται εξηγείται από το γεγονός ότι, σε ένα ολόκληρο σύστημα, ένας αλγόριθμος που παρουσιάζει ‘‘καλή’’ επίδοση είναι ικανοποιητικός, εφόσον συνδυαστεί με ‘‘καλής’’ επίδοσης αλγορίθμους και στα υπόλοιπα μέρη του. Με άλλα λόγια, στα συστήματα αυτά η επιλογή αλγορίθμου διατήρησης της συνέπειας δεν αποτελεί τη μοναδική απασχόληση - αντίθετα με αυτό που συμβαίνει σε πειραματικές εφαρμογές στις οποίες αναφέρονται τα [23, 24, 25, 26]. Σε εμπορικά συστήματα διαχείρισης βάσεων δεδομένων, η έρευνα στρέφεται περισσότερο σε αποδοτικές μεθόδους διαχείρισης των δεδομένων από τις ερωτήσεις στο σύστημα. Ικανοποιητικές επιδόσεις μπορούν, επομένως, να αναζητηθούν εκεί.

2.5 Μετρήσεις

Προκειμένου για τη συγκριτική αξιολόγηση των αλγορίθμων των σχετικών με τη διαχείριση της κρυφής μνήμης ενός συστήματος, είναι απαραίτητη η πραγματοποίηση μετρήσεων. Οι μετρήσεις αυτές πραγματοποιούνται, κατά κανόνα, όχι στο πραγματικό σύστημα [34], αλλά σε προσομοιωτικά μοντέλα του [23, 24, 25, 26].

Πιο συγκεκριμένα, οι εργασίες του Michael J. Franklin και των συνεργατών του αφορούν μελέτες οι οποίες πραγματοποιήθηκαν σε μοντέλα προσομοίωσης. Κι αυτό, γιατί οι μελέτες τους έχουν ως αντικείμενο αλγορίθμους διατήρησης της συνέπειας ανάμεσα σε πολυάριθμες κρυφές μνήμες του ίδιου συστήματος, οι οποίοι δεν είναι υλοποιημένοι σε κανένα πραγματικό σύστημα.

Οι παράμετροι του συστήματος προσομοίωσης, με δεδομένο ότι οι αλγόριθμοι αφορούν αρχιτεκτονικές πελάτη-εξυπηρετητή, είναι μεταξύ άλλων :

- ο αριθμός των πελατών στο σύστημα,
- κάποια χαρακτηριστικά υλοποίησης στους πελάτες,

- χαρακτηριστικά υλοποίησης στον εξυπηρετητή,
- χαρακτηριστικά στοιχεία των δίσκων του συστήματος και της βάσης δεδομένων,
- στοιχεία υλοποίησης του συστήματος και του δικτύου επικοινωνίας και
- ο αλγόριθμος διατήρησης της συνέπειας των ίδιων δεδομένων που διατηρούνται σε περισσότερες από μία κρυφές μνήμες.

Για την πραγματοποίηση των μετρήσεων στο προσομοιωτικό μοντέλο χρησιμοποιούνται ίχνη από προηγούμενες χρήσεις της βάσης δεδομένων. Τα ίχνη αυτά δεν είναι άλλο από ακολουθίες αριθμών σελίδων τις οποίες προσπελαύνουν οι πελάτες του συστήματος.

Τα μεγέθη που μετρούνται είναι ο χρόνος απόκρισης καθώς επίσης και ο ρυθμός εκτέλεσης των ερωτήσεων που θέτουν οι πελάτες. Τα τελικά συμπεράσματα προκύπτουν από τη σύγκριση των παραπάνω μεγεθών για διαφορετικές τιμές των παραμέτρων στο μοντέλο προσομοίωσης.

Με σκοπό την αξιολόγηση της επίδοσης των τεχνικών που επιλέγονται στο σχεδιασμό ενός συστήματος, και πάλι, κατασκευάζονται κοινά αποδεκτά πρότυπα ερωτήσεων σε βάσεις δεδομένων (benchmarks), που χρησιμοποιούνται σα μέτρα σύγκρισης [6, 8, 48].

Κύριο χαρακτηριστικό της τεχνικής αυτής είναι ότι τόσο το περιεχόμενο της βάσης δεδομένων που θα χρησιμοποιηθεί για την εκτέλεση των ερωτήσεων, όσο και η εκτέλεση των ερωτήσεων είναι πολύ αυστηρά προκαθορισμένες. Στόχος της χρήσης της τεχνικής αυτής είναι να εξεταστεί η συμπεριφορά διαφορετικών συστημάτων διαχείρισης βάσεων δεδομένων, όταν αυτά χρησιμοποιούν την ίδια ακριβώς βάση και εκτελούν σε αυτήν τις ίδιες ακριβώς ερωτήσεις τηρώντας συγκεκριμένη ακολουθία.

Το πρότυπο ερωτήσεων OO1, για παράδειγμα, προϋποθέτει βάση δεδομένων η οποία αποτελείται από :

- **Τμήματα** (parts) με γνωρίσματα : id, type, x, y, build, συγκεκριμένου τύπου και
- **Συνδέσεις** (connections) μεταξύ τμημάτων με γνωρίσματα : from, to, type, length συγκεκριμένου τύπου.

Οι ερωτήσεις στη βάση που προβλέπονται από το πρότυπο είναι οι ακόλουθες :

- **Αναζήτηση** 1000 τυχαίων τμημάτων και ανάκλησή τους από τη βάση. Για κάθε τμήμα, εκτέλεση μιας κενής διαδικασίας με ορίσματα τα περία του x, y και type.

- **Διάσχιση** : Αναζήτηση όλων των τμημάτων που συνδέονται με ένα τυχαία επιλεγμένο τμήμα, αναδρομικά. Για κάθε τμήμα, εκτέλεση της διαδικασίας του προηγούμενου βήματος.

- **Εισαγωγή** 100 τμημάτων και 3 συσχετίσεων για κάθε τμήμα με άλλα τυχαία επιλεγμένα τμήματα. Ενημέρωση της βάσης για τις εισαγωγές.

Για καθεμιά από τις παραπάνω ενέργειες, το πρότυπο προβλέπει υπολογισμό μέσου χρόνου εκτέλεσής τους και σύγκριση των αποτελεσμάτων με τους αντίστοιχους χρόνους για γνωστές βάσεις δεδομένων.

Σε ό,τι αφορά μετρήσεις που πραγματοποιούνται σε πραγματικά συστήματα, οι οποίες είναι αυτές που μας ενδιαφέρουν για το Σύστημα Σημασιολογικού Ευρετηριασμού, αυτές χρειάζονται τα ακόλουθα στοιχεία :

- Προσδιορισμό των χαρακτηριστικών εκείνων του συστήματος διαχείρισης κρυφής μνήμης που μπορούν να τροποποιηθούν, αλλάζοντας έτσι τη συμπεριφορά του συστήματος,

- προσδιορισμό των στοιχείων εκείνων που θεωρείται ότι αποτελούν μετρήσιμα μεγέθη και χαρακτηριστικά της επίδοσης του συστήματος και

- οργάνωση πειραμάτων που διεξάγονται για διαφορετικές τιμές των παραμέτρων σχεδίασης του συστήματος.

Ο λόγος για τον οποίο δεν είναι εφαρμόσιμες στο Σύστημα Σημασιολογικού Ευρετηριασμού οι τεχνικές προτύπων ερωτήσεων είναι διότι αυτές είναι προσανατολισμένες σε εμπορικές εφαρμογές συγκεκριμένου τύπου (CAD/CAM) χαρακτηριστικά των οποίων χρησιμοποιήθηκαν για την επιλογή, τόσο των περιεχομένων της βάσης όσο και των ερωτήσεων σ'αυτήν.

2.6 Προανάκληση

Η μέθοδος της προανάκλησης [5, 34] αναφέρεται στην εισαγωγή στην κρυφή μνήμη, εκτός από τη σελίδα που είναι κάθε φορά απαραίτητη, κάποιων ακόμη σελίδων που βρίσκονται πριν ή μετά από αυτήν στο δίσκο. Σκοπός μιας τέτοιας διαδικασίας είναι ο περιορισμός των προσπελάσεων στο δίσκο, οι οποίες είναι πράξεις μεγάλου κόστους για

το σύστημα. Απαραίτητη προϋπόθεση, βέβαια, για τον περιορισμό των προσπελάσεων στο δίσκο, είναι να γίνονται αναφορές στο μέλλον στις σελίδες που εισάγονται στην κρυφή μνήμη. Σε αντίθετη περίπτωση, οι επιπλέον εισαγωγές, όχι μόνο δε βελτιώνουν, αλλά χειροτερεύουν την επίδοση του συστήματος, καθώς είναι πολύ πιθανό να προκαλέσουν απομάκρυνση από την κρυφή μνήμη χρήσιμων, ενδεχομένως, σελίδων. Το ζήτημα της προανάκλησης είναι, επομένως, ένα πολύ σημαντικό ζήτημα για τη βελτίωση της επίδοσης ενός συστήματος, εφόσον χρησιμοποιηθεί σωστά.

Η προανάκληση είναι συνηθισμένη μέθοδος σε συστήματα διαχείρισης δοσοληψιών [34]. Τα συστήματα διαχείρισης βάσεων δεδομένων, ωστόσο, έχουν την ιδιαιτερότητα ότι η εισαγωγή μιας σελίδας στην κρυφή μνήμη του υπολογιστή, προκαλείται από την αναφορά σε ένα και μόνο αντικείμενο που περιέχεται σε αυτήν. Για το λόγο αυτό, είναι μάλλον αναποτελεσματικό να εισάγουμε στην κρυφή μνήμη μία ακόμη σελίδα, στην οποία πιθανότατα περιέχονται αντικείμενα άσχετα προς το αντικείμενο-στόχο. Από την άλλη μεριά, αντικείμενα που σχετίζονται σημασιολογικά με το αντικείμενο-στόχο, μπορούν να τοποθετηθούν στην ίδια με αυτό σελίδα κι έτσι να εισαχθούν μαζί του στην κρυφή μνήμη. Η διαδικασία της ομαδοποίησης (clustering) αντικειμένων μιας βάσης δεδομένων, με βάση ένα επιλεγμένο κριτήριο, είναι πολύ σημαντική. Γι'αυτό και αποτελεί έναν ολόκληρο τομέα μελέτης στα συστήματα διαχείρισης βάσεων δεδομένων.

Γεγονός είναι, πάντως, ότι η μελετημένη ομαδοποίηση των αντικειμένων στις σελίδες του δίσκου μπορεί να αποτελέσει σημαντικότατο παράγοντα βελτίωσης της επίδοσης του συστήματος.

2.7 Συμπεράσματα

Η παραπάνω μελέτη της βιβλιογραφίας σε σχέση με το ζήτημα της διαχείρισης της κρυφής μνήμης σε συστήματα διαχείρισης βάσεων δεδομένων οδήγησε στα ακόλουθα συμπεράσματα :

- Τα συστήματα διαχείρισης κρυφής μνήμης παρουσιάζουν μεγάλες διαφοροποιήσεις ανάμεσα σε σχεσιακές και οντοκεντρικές βάσεις δεδομένων. Κι αυτό, εξαιτίας της διαφορετικής φύσης των ερωτήσεων που μπορούν να τεθούν σε καθεμιά από τις παραπάνω περιπτώσεις, η οποία δημιουργεί τα ανάλογα μοντέλα πρόσβασης.

- Το παραπάνω έχει σαν αποτέλεσμα να είναι μικρή η δυνατότητα εφαρμογής αλγορίθμων αντικατάστασης σε οντοκεντρικές βάσεις δεδομένων - κι επομένως και στην Telos.

- Σε συστήματα τα οποία είναι σχεδιασμένα σε αρχιτεκτονική client-server, η χρησιμοποίηση των υπολογιστικών πόρων στους υπολογιστές-πελάτες είναι όλο και πιο διμοφιλής. Αυτό οφείλεται στο γεγονός ότι η μεταφορά αρμοδιοτήτων στους πελάτες δίνει μεγαλύτερη άνεση στον εξυπηρετητή να υποστηρίξει ακόμη περισσότερες διεργασίες. Για την εφαρμογή μιας τέτοιας πολιτικής, βέβαια, πρέπει να υπάρχουν ορισμένες προϋποθέσεις, όπως ταχύτητα των γραμμών στα διαθέσιμα δίκτυα υπολογιστών, ώστε να μην επιδεινώνεται η επίδοση του συστήματος από την ανάγκη συχνής επικοινωνίας των υπολογιστών-πελατών μεταξύ τους και με τον υπολογιστή-εξυπηρετητή.

- Αλγόριθμοι διατήρησης της συνέπειας υπάρχουν σε μεγάλη ποικιλία. Η επιλογή του κατάλληλου για κάθε σύστημα βασίζεται στα ιδιαίτερα χαρακτηριστικά του συστήματος (π.χ. ταχύτητα δικτύου, υπολογιστική ισχύ υπολογιστών κλπ.) και πραγματοποιείται με βάση τις επιδόσεις του σε μετρήσεις.

Κεφάλαιο 3

Διαχείριση κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού

3.1 Τύποι οντοτήτων στην SIS-Telos

3.1.1 Η γλώσσα SIS-Telos και το Σύστημα Σημασιολογικού Ευρετηριασμού

Η Telos είναι μια γλώσσα παράστασης γνώσης, που παριστάνει τη γνώση με τη μορφή σημασιολογικού δικτύου. Η Telos προτάθηκε από τον κ. Μυλόπουλο [33] και μια υλοποίησή της (SIS-Telos) έχει γίνει από την ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του ΙΠ-ΙΤΕ [22].

Κάθε αντικείμενο ή έννοια του πραγματικού κόσμου αποτελεί στη βάση γνώσης μια ξεχωριστή οντότητα (*Object*). Ολες οι οντότητες που μπορούν να υπάρξουν σε μια βάση δεδομένων ταξινομούνται στην κλάση του συστήματος *Object*, έχουν ένα μοναδικό όνομα που παράγεται από το σύστημα ενώ μπορούν να διαθέτουν λογικό όνομα, δοσμένο από το χρήστη του συστήματος (μηχανισμός ονοματοδοσίας).

Η κλάση *Object* περιλαμβάνει τέσσερις υποκλάσεις : *Individual*, *Attribute*, *Class* και *Token*. Στην κλάση *Individual* ταξινομούνται οι οντότητες, οι κλάσεις οντοτήτων, οι κλάσεις από κλάσεις οντοτήτων κ.ο.κ. Η κλάση αυτή χρησιμοποιείται για την αναπαράσταση των διακριτών αντικειμένων του πραγματικού κόσμου. Η κλάση *Attribute* χρησιμοποιείται για την αναπαράσταση των γνωρισμάτων και των σχέσεων μεταξύ αντικειμένων του

πραγματικού κόσμου. Στην κλάση αυτή ταξινομούνται τα γνωρίσματα, οι κλάσεις γνωρισμάτων κ.ο.κ. Τόσο οι οντότητες όσο και τα γνωρίσματα αποτελούν οντότητες στη γλώσσα SIS-Telos και τυγχάνουν ομοιόμορφης διαχείρισης. Η κλάση *Class* χρησιμοποιείται για την ταξινόμηση όλων των κλάσεων οντοτήτων και κλάσεων γνωρισμάτων, καθώς και κλάσεων από κλάσεις. Οι ατομικές οντότητες και τα ατομικά γνωρίσματα που αποτελούν ουσιαστικά τον πληθυσμό της βάσης δεδομένων ταξινομούνται στην κλάση του συστήματος *Token*.

Η SIS-Telos χρησιμοποιεί βασικούς μηχανισμούς δόμησης που χρησιμοποιούνται σε βάσεις παράστασης γνώσης και σημασιολογικά μοντέλα δεδομένων, παρέχοντας τη δυνατότητα της περιγραφής της πληροφορίας στο επίπεδο ακρίβειας που απαιτείται. Οι μηχανισμοί αυτοί είναι οι ακόλουθοι :

- **Μηχανισμός Ονοματοδοσίας αντικειμένων**
  Κάθε οντότητα του πραγματικού κόσμου παριστάνεται από μια ξεχωριστή οντότητα, *Object*. Η οντότητα αυτή έχει ένα μοναδικό όνομα που παράγεται από το σύστημα (SYSID). Επίσης σε κάθε οντότητα μπορεί να αποδωθεί και ένα *λογικό* όνομα από τον χρήστη του συστήματος.

- **Μηχανισμός Ταξινόμησης**
  Κάθε οντότητα του σημασιολογικού δικτύου της SIS-Telos ταξινομείται σε μία ή περισσότερες κλάσεις οντοτήτων και κληρονομεί τα χαρακτηριστικά τους. Η κλάση είναι κι αυτή οντότητα (άρα μπορεί να ταξινομηθεί κι αυτή). Η SIS-Telos απαιτεί κάθε οντότητα να είναι περίπτωση μιας τουλάχιστον κλάσης. Με τον τρόπο αυτό δημιουργείται μια ανω μη φραγμένη ιεραρχία.

- **Μηχανισμός Γενίκευσης-Εξειδίκευσης.**
  Με το μηχανισμό αυτό ορίζεται μια σχέση υποσυνόλου μεταξύ κλάσεων ίδιας στάθμης. Μία κλάση μπορεί να είναι εξειδίκευση (υποσύνολο-υποκλάση) μιας ή περισσοτέρων κλάσεων των οποίων κληρονομεί τα γνωρίσματα. Επίσης μια κλάση μπορεί να είναι γενίκευση (υπερσύνολο-υπερκλάση) μιας ή περισσότερων κλάσεων. Τέλος, μια κλάση που είναι γενίκευση ή εξειδίκευση μιας κλάσης μπορεί επίσης να έχει δικά της γνωρίσματα.

- **Μηχανισμός Απόδοσης Γνωρίσματος**

  Σε κάθε οντότητα που παριστάνεται στο σύστημα μπορεί να αποδοθούν γνωρίσματα. Στα γνωρίσματα αυτά μπορούν να δοθούν παραπάνω από μια τιμές (πλειότιμα) ή να μη δοθεί καμία (προαιρετικά). Κάθε γνώρισμα μπορεί να ταξινομηθεί, να αποτελεί γενίκευση ή εξειδίκευση κάποιου άλλου γνωρίσματος και να του απο δοθουν γνωρίσματα (ιδιότητες που προκύπτουν από την όμοια μεταχείριση οντοτήτων και γνωρισμάτων).

Η γλώσσα SIS-Telos αποτελεί τον πυρήνα ανάπτυξης του Συστήματος Σημασιολογικού Ευρετηριασμού. Το σύστημα αυτό χρησιμοποιείται για την αποθήκευση και τη διερεύνηση δεδομένων πολλαπλώς συνδεδεμένων και διαθέτει τους ακόλουθους μηχανισμούς διαχείρισης :

- **Μηχανισμός αποθήκευσης αντικειμένων**

  Ο μηχανισμός αυτός υποστηρίζει δοσοληψίες και ταυτόχρονη πρόσβαση από πολλούς χρήστες.

- **Μηχανισμός ερωτήσεων**

  Το ΣΣΕ παρέχει ένα μηχανισμό ερωτήσεων με τη βοήθεια του οποίου είναι δυνατή η πλοήγηση στην πληροφορία της βάσης, καθώς επίσης και η διατύπωση ερωτήσεων με βάση συγκεκριμένα κριτήρια.

- **Μηχανισμός εισαγωγής δεδομένων**

  Η εισαγωγή δεδομένων στο ΣΣΕ μπορεί να πραγματοποιηθεί είτε χρησιμοποιώντας το συντακτικό αναλυτή της SIS-Telos, είτε με τη βοήθεια των Διαλογικών Δελτίων Εισαγωγής Δεδομένων, που το σύστημα διαθέτει.

Το ΣΣΕ παρέχει γραφικό περιβάλλον για την επικοινωνία του χρήστη με το σύστημα. Στο γραφικό περιβάλλον αυτό παριστάνεται το σημασιολογικό μοντέλο του ΣΣΕ, καθώς επίσης και τα δελτία εισαγωγής δεδομένων, τα οποία αποτελούν αντικείμενο μελέτης των πειραμάτων που θα περιγραφούν στο 5ο κεφάλαιο.

3.1.2 Τύποι οντοτήτων στην SIS-Telos

Κάθε κλάση οντοτήτων αποτελεί, σύμφωνα με την παραπάνω περιγραφή, οντότητα στο οντοκεντρικό μοντέλο παράστασης γνώσης. Μπορεί, επομένως, να αποτελεί περίπτωση

μιας άλλης γενικότερης κλάσης οντοτήτων. Δημιουργώντας νέες κλάσεις οντοτήτων που περιέχουν υπάρχουσες κλάσεις οντοτήτων μπορεί να δημιουργηθεί μια μη φραγμένη ιεραρχία ταξινόμησης.Το πρώτο επίπεδο αυτής της ιεραρχίας αποτελούν οι ατομικές οντότητες (tokens), το επόμενο επίπεδο απλές κλάσεις οντοτήτων που ομαδοποιούν ατομικές οντότητες (simple classes), το επόμενο επίπεδο οι μετακλάσεις που ομαδοποιούν κλάσεις οντοτήτων (M1 classes) κ.ο.κ.

Οι ανεξάρτητες οντότητες της SIS-Telos ανάλογα με το επίπεδο ταξινόμησης στο οποίο βρίσκονται διακρίνονται σε κλάσεις οντοτήτων (classes) όταν βρίσκονται σε οποιοδήποτε επίπεδο ταξινόμησης εκτός από το πρώτο και ατομικές οντότητες (tokens) όταν ταξινομούνται στο πρώτο επίπεδο της παραπάνω ιεραρχίας. Καθώς τα γνωρίσματα (ή σχέσεις - Attribute) αποτελούν και αυτά οντότητες, διακρίνονται σε κλάσεις γνωρισμάτων (link classes) και ατομικά γνωρίσματα (link tokens).

Το εννοιολογικό σχήμα μιας βάσης που έχει περιγραφεί σε SIS-Telos έχει μορφή σημασιολογικού δικτύου, όπου οι κόμβοι του δικτύου αντιστοιχούν στις κλάσεις ανεξάρτητων οντοτήτων και τις ατομικές ανεξάρτητες οντότητες, ενώ οι σύνδεσμοι μεταξύ κόμβων αντιστοιχούν στις σχέσεις ταξινόμησης και γενίκευσης-εξειδίκευσης καθώς επίσης και στο μηχανισμό σύνθεσης, δηλαδή τις κλάσεις γνωρισμάτων και τα ατομικά γνωρίσματα.

3.1.3 Υλοποίηση των οντοτήτων στη γλώσσα SIS-Telos

Από τον ορισμό των μηχανισμών δόμησης που υποστηρίζονται από την SIS-Telos, προκύπτει ότι κάθε εφαρμογή ενός μηχανισμού περιλαμβάνει ακριβώς δύο οντότητες. Βέβαια, οι σχέσεις που δημιουργούνται με τους μηχανισμούς δόμησης κληρονομούνται μέσω των ιεραρχιών ταξινόμησης, γενίκευσης και ιεραρχίας σύνθεσης κλάσεων από περισσότερες από μια οντότητες, αλλά οι άμεσοι σύνδεσμοι στους μηχανισμούς δόμησης περιλαμβάνουν πάντα δύο μόνο οντότητες.

Ετσι, για την περιγραφή μιας οντότητας αρκεί να παρατεθούν στη δομή που την παριστάνει μόνο τα γνωρίσματα και οι οντότητες με τις οποίες συνδέεται άμεσα. Ολες οι υπόλοιπες συνδέσεις μπορούν να υπολογιστούν ως μεταβατικό κλειστό περίβλημα των συνδέσμων με αυτή την οντότητα για κάθε μηχανισμό δόμησης.

Επίσης, χρειάζεται η αποθήκευση στη δομή που παριστά μια οντότητα του αναγνωριστικού συστήματος, του μοναδικού αριθμού που παράγεται από το σύστημα

και αποτελεί την ταυτότητα της οντότητας, και η περιγραφή του τύπου της, δηλαδή το επίπεδο ταξινόμησης στο οποίο βρίσκεται και αν αποτελεί ανεξάρτητη οντότητα ή οντότητα-γνώρισμα.

Στη συνέχεια περιγράφεται η υλοποίηση της δομής παράστασης και αποθήκευσης για καθεμιά από τις τέσσερις βασικές κατηγορίες οντοτήτων της SIS-Telos.

- Κλάσεις κόμβων (Class Nodes)

  Εκτός από το αναγνωριστικό συστήματος και την περιγραφή του τύπου, η δομή παράστασης και αποοθήκευσης για κάθε τέτοια οντότητα περιλαμβάνει ένα διάνυσμα από σύνολα αναγνωριστικών συστήματος που αναφέρονται σε όλες τις οντότητες με τις οποίες η οντότητα αυτή συνδέεται άμεσα μέσω των μηχανισμών δόμησης. Τα σύνολα αναγνωριστικών συστήματος διακρίνονται σε:

  -- instance of : στο σύνολο αυτό περιλαμβάνονται όλες οι οντότητες των οποίων περίπτωση αποτελεί η συγκεκριμένη οντότητα.

  -- instantiated : στο σύνολο αυτό περιλαμβάνονται όλες οι οντότητες που αποτελούν περίπτωση της συκγεκριμένης οντότητας.

  -- isA : στο σύνολο αυτό περιλαμβάνονται όλες οι άμεσες υπερκλάσεις της συγκεκριμένης οντότητας.

  -- subclasses : στο σύνολο αυτό περιλαμβάνονται όλες οι άμεσες υποκλάσεις της συγκεκριμένης οντότητας.

  -- links : στο σύνολο αυτό περιλαμβάνονται όλοι οι σύνδεσμοι (γνωρίσματα) προς άλλες οντότητες που έχουν τη συγκεκριμένη οντότητα για πεδίο ορισμού.

  -- linked by : στο σύνολο αυτό περιλαμβάνονται όλοι οι σύνδεσμοι για τους οποίους η τρέχουσα οντότητα αποτελεί πεδίο τιμών.

- Κλάσεις συνδέσμων (Link Classes)

  Η δομή παράστασης για τις οντότητες αυτές είναι είναι αντίστοιχη με τη δομή παράστασης για τις κλάσεις κόμβων, καθώς επιτρέπεται ο ορισμός γνωρισμάτων και για τις κλάσεις γνωρισμάτων. Η δομή παράστασης περιλαμβάνει δύο επιπλέον πεδία τα οποία πρέπει απαραίτητα να παίρνουν τιμές για να δημιουργηθεί η συγκεκριμένη οντότητα :

-- from-node : περιλαμβάνει το αναγνωριστικό συστήματος που περιγράφει την κλάση κόμβων που αποτελεί το πεδίο ορισμού για τη συγκεκριμένη κλάση συνδέσμων.

-- to-node : περιλαμβάνει το αναγνωριστικό συστήματος που περιγράφει την κλάση κόμβων που αποτελεί το πεδίο τιμών για τη συγκεκριμένη κλάση συνδέσμων.

- Απλοί κόμβοι (Token nodes)

  Η δομή παράστασης για τις οντότητες αυτές είναι η απλούστερη δυνατή, καθώς τέτοιες οντότητες συνδέονται άμεσα μόνο με γνωρίσματα και τις κλάσεις των οποίων αποτελούν περιπτώσεις. Τα πεδία επομένως που περιλαμβάνονται στη δομή παράστασης και αποθήκευσής τους είναι :

  links, linked by και instance of, με την ερμηνεία που δόθηκε σ'αυτά παραπάνω.

- Απλοί σύνδεσμοι (Link tokens)

  Η δομή παράστασης και αποθήκευσης είναι και σ'αυτήν την περίπτωση απλούστερη από την αντίστοιχη των κλάσεων συνδέσμων. Τα πεδία που περιέχει είναι :

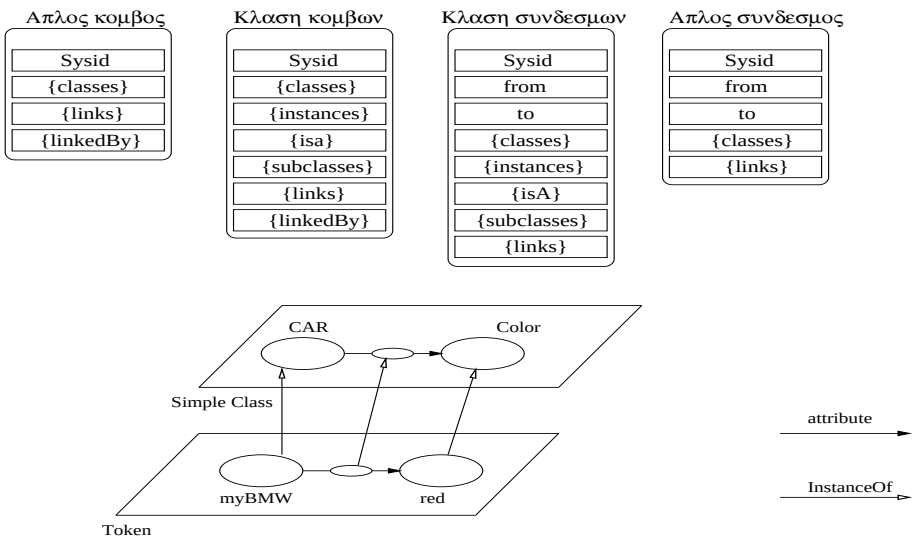
  instance of, links, linked by, from node, και to node, με την ερμηνεία που αυτά έχουν και παραπάνω.

  Παράδειγμα των παραπάνω δομών και της αντιστοίχισής τους σε αντικείμενα βάσης υλοποιημένης σε SIS-Telos παρουσιάζεται στο σχήμα 3.1.

- Μεγάλες οντότητες

  Για όλες τις παραπάνω δομές παράστασης και αποθήκευσης τα πεδία αναγνωριστικών συστήματος που χρησιμοποιούνται για την καταγραφή όλων των οντοτήτων που σχετίζονται με μια οντότητα μέσω κάποιου από τους διαθέσιμους μηχανισμούς ταξινόμησης, έχουν σταθερό, προκαθορισμένο μέγεθος. Το μέγεθος αυτό έχει επιλεγεί ώστε να είναι το καταλληλότερο για τις περισσότερες από τις οντότητες που ορίζονται από το χρήστη και φυσικά για λόγους οικονομίας χώρου και επίδοσης του συστήματος.

  Πολύ συχνά, κάποιες από τις οντότητες που ορίζει ο χρήστης έχουν περισσότερες άμεσες συνδέσεις με άλλες οντότητες απ'όσες μπορούν να



Σχήμα 3.1: Αντιστοίχιση των δομών αναπαράστασης και αποθήκευσης σε οντότητες και συσχετίσεις σε παράδειγμα εφαρμογής

αποθηκευτούν στα συγκεκριμένου μεγέθους σύνολα αναγνωριστικών συστήματος που χρησιμοποιούνται για κάθε οντότητα.

Για το λόγο αυτό έχει δημιουργηθεί μια ακόμη δομή παράστασης και αποθήκευσης συνόλων αναγνωριστικών συστήματος. Η δομή αυτή χρησιμοποιείται για να δημιουργούμε πιθανές επεκτάσεις στα σύνολα αναγνωριστικών συστήματος για κάθε δομή παράστασης οντοτήτων, όταν τα σύνολα αυτά υπερχειλίζουν. Η δομή αυτή οργανώνεται σε απλά συνδεδεμένες λίστες με κόμβους που περιέχουν σταθερό πλήθος αναγνωριστικών συστήματος. Κάθε περίπτωση της δομής αυτής, αποτελώντας κόμβο της συνδεδεμένης λίστας, εκτός από ένα σταθερό πίνακα αναγνωριστικών συστήματος, περιέχει ένα δείκτη στον επόμενο κόμβο της λίστας από επεκτάσεις για τη συγκεκριμένη οντότητα, καθώς και το αναγνωριστικό συστήματος για την οντότητα, για λόγους ελέγχου της συνέπειας των δεδομένων στη βάση.

3.2 Διαχείριση οντοτήτων στο Σύστημα Σημασιολογικού Ευρετηριασμού

Καθένας από τους τύπους οντοτήτων που χρησιμοποιούνται από την SIS-Telos για την παράσταση πληροφορίας απαιτεί, όπως είδαμε, διαφορετική δομή παράστασης και αποθήκευσης. Για το λόγο αυτό, οντότητες του ίδιου μόνο τύπου ομαδοποιούνται σε τμήματα (blocks) που έχουν μέγεθος ίδιο με αυτό του ελάχιστου χώρου που μπορεί να διαβαστεί/γραφτεί από/στο δίσκο, της σελίδας δίσκου (disk page). Επιτυγχάνεται με τον τρόπο αυτό ευκολότερη διαχείριση των τμημάτων που αποθηκεύουν οντότητες κατά τη δημιουργία και τη διαγραφή τους, ενώ ταυτόχρονα αποφεύγονται φαινόμενα κατακερματισμού.

Κάθε οντότητα στην SIS-Telos, εξάλλου, εκτός από το μοναδικό αναγνωριστικό συστήματος το οποίο την χαρακτηρίζει, διαθέτει κι ένα λογικό όνομα. Τα λογικά ονόματα των οντοτήτων ορίζονται από τους χρήστες και, όταν αυτό δε συμβεί, δημιουργούνται από το σύστημα με κατάλληλη μέθοδο. Σε αντίθεση με το αναγνωριστικό συστήματος, το οποίο αποτελεί τον τρόπο πρόσβασης του συστήματος στις οντότητες της βάσης, το λογικό όνομα αποτελεί τον τρόπο πρόσβασης των χρηστών στις οντότητες της βάσης, μέσω κάποιας εφαρμογής της SIS-Telos.

Για τη διαχείριση των δομών που απαιτούνται για τη δημιουργία της αντιστοίχισης ανάμεσα σε λογικά ονόματα και αναγνωριστικά συστήματος, υπάρχει στο SIS(Semantic Index System - Σύστημα Σημασιολογικού Ευρετηριασμού) ένας κατάλογος λογικών ονομάτων.

Τελος, κάθε σύστημα διαχείρισης βάσεων δεδομένων διατηρεί έναν εσωτερικό κατάλογο, ο οποίος περιέχει περιγραφές για τα δεδομένα που αποθηκεύονται, και τη θέση στην οποία βρίσκονται στη μνήμη και το δίσκο. Η δομή αυτή ονομάζεται κατάλογος συστήματος.

3.2.1 Κατάλογος συστήματος

Οπως σε όλα τα συστήματα διαχείρισης βάσεων δεδομένων, έτσι και στο SIS, η πληροφορία τόσο της σταθερής κατάστασης μιας οντότητας στο δίσκο, όσο και της ενδιάμεσης κατάστασης της οντότητας στη μνήμη διατηρείται σε μια δομή που ονομάζεται κατάλογος συστήματος.

Οι εφαρμογές που χρησιμοποιούν το Σύστημα Σημασιολογικού Ευρετηριασμού,

πραγματοποιούν όλες τις προσπελάσεις τους στα δεδομένα της βάσης μέσω του καταλόγου συστήματος.

Ο κατάλογος συστήματος είναι, επομένως, επιφορτισμένος με την εκτέλεση των παρακάτω λειτουργιών :

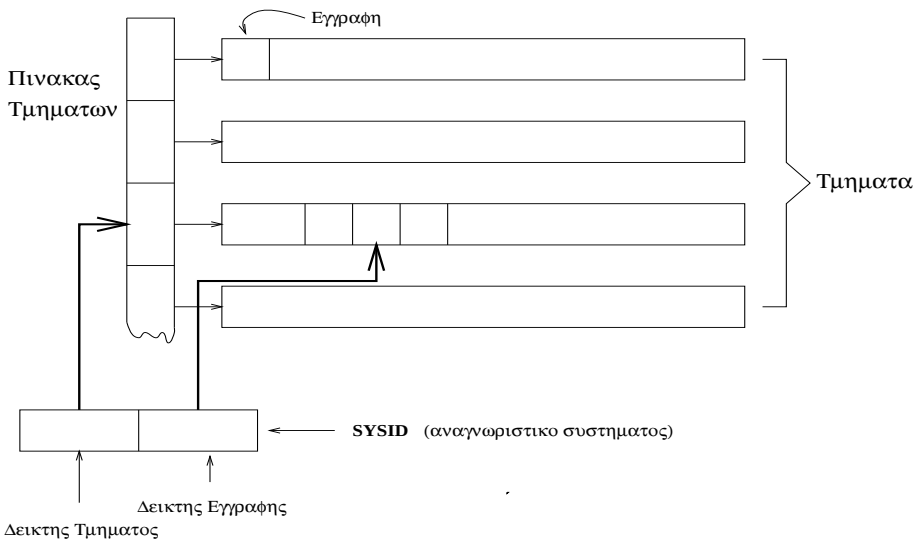
- δημιουργία νέων οντοτήτων
- διαγραφή οντοτήτων από τη βάση
- προσπέλαση στις οντότητες της βάσης
- ενημέρωση της βάσης με τις νέες οντότητες ή οντότητες που έχουν μεταβληθεί.

Ο κατάλογος συστήματος είναι ένας πίνακας από εγγραφές, μία για κάθε οντότητα της βάσης. Εξαιτίας του πολύ μεγάλου μεγέθους στο οποίο μπορεί να φτάσει, ο κατάλογος συστήματος οργανώνεται εικονικά σαν πίνακας από τμήματα με βαθμό εμμεσότητας 1. Κάθε τμήμα του πίνακα αυτού μπορεί να περιέχει ένα συγκεκριμένο πλήθος από εγγραφές.

Η προσπέλαση στις εγγραφές του καταλόγου συστήματος πραγματοποιείται μέσω του μοναδικού αναγνωριστικού συστήματος για την κάθε οντότητα. Για το σκοπό αυτό, τα αναγνωριστικά συστήματος που υλοποιούνται όπως προαναφέρθηκε με τη μορφή ακεραίων αριθμών, θεωρούνται ως διανύσματα της μορφής <δείκτης τμήματος, δείκτης εγγραφής>. Το πεδίο ”δείκτης τμήματος” που αποτελείται από τα δυαδικά ψηφία υψηλής τάξης του αναγνωριστικού συστήματος, καθορίζει το τμήμα του καταλόγου συστήματος στο οποίο βρίσκεται η εγγραφή για μια οντότητα. Το πεδίο ”δείκτης εγγραφής” που αποτελείται από τα δυαδικά ψηφία χαμηλής τάξης του αναγνωριστικού συστήματος, καθορίζει τη συγκεκριμένη εγγραφή για μια οντότητα στο τμήμα του υποκαταλόγου συστήματος που την περιέχει. Ο τρόπος δεικτοδότησης του καταλόγου συστήματος και η οργάνωσή του ως πίνακα από τμήματα φαίνεται στο σχήμα 3.2.

Η πληροφορία που αποθηκεύεται σε μια εγγραφή του καταλόγου συστήματος για μια οντότητα αφορά, όπως αναφέρθηκε και παραπάνω, τη σταθερή κατάσταση της οντότητας όπως αυτή αποθηκεύεται στο δίσκο και την ενδιάμεση κατάσταση της οντότητας όταν αυτή φορτωθεί στη μνήμη. Η σταθερή κατάσταση της οντότητας περιλαμβάνει τις εξής πληροφορίες :

- Θέση στο δίσκο όπου αποθηκεύεται η οντότητα.



Σχήμα 3.2: Οργάνωση και δεικτοδότηση του καταλόγου συστήματος

Ο *Πίνακας Τμημάτων* που φαίνεται στο παραπάνω σχήμα είναι *εικονικός* δηλαδή παρουσιάζεται μονάχα για να δείξει τον τρόπο που γίνεται η δεικτοδότηση του καταλόγου συστήματος (αναπαραγωγή από [47]).

- Θέση στο δίσκο όπου αποθηκεύονται οι δομές επέκτασης για την οντότητα, αν το πλήθος των συνδέσεων της οντότητας αυτής με άλλες οντότητες υπερβαίνει το μέγεθος του συγκεκριμένου τύπου που χρησιμοποιείται για την παράσταση της οντότητας αυτής.

- Το συνολικό μέγεθος των δομών επέκτασης για την οντότητα αυτή στο δίσκο, αν η παράσταση της οντότητας περιλαμβάνει τέτοιες δομές.

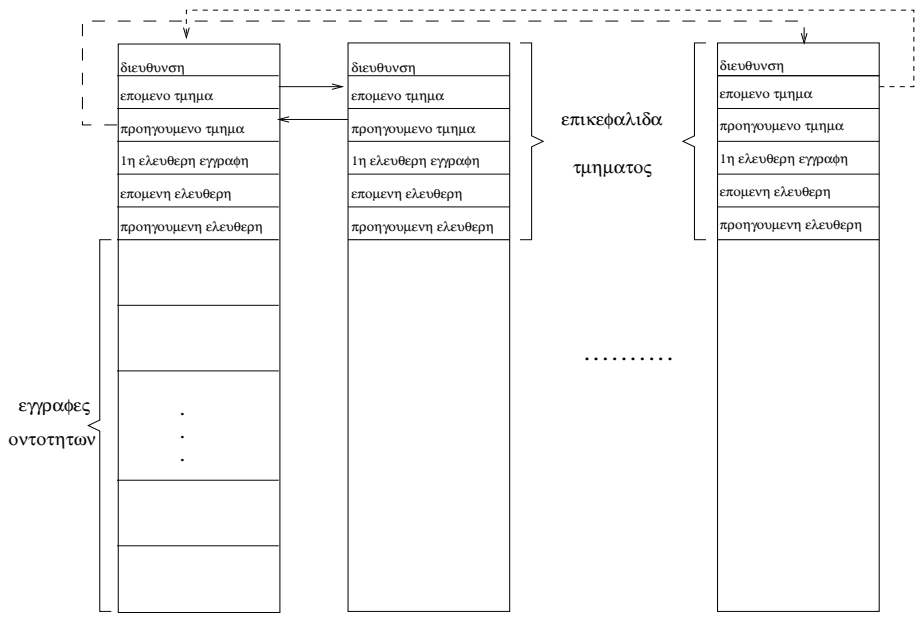
- Τον τύπο της οντότητας, ο οποίος περιγράφει το επίπεδο ταξινόμησης και τις κλάσεις του συστήματος στις οποίες ταξινομείται η οντότητα αυτή.

Για τις οντότητες που βρίσκονται στη μνήμη η εγγραφή του καταλόγου συστήματος περιλαμβάνει τις εξής πληροφορίες :

- Τη θέση στη μνήμη όπου βρίσκεται η οντότητα.

- Την κατάσταση της οντότητας. Μια οντότητα μπορεί να έχει μόλις δημιουργηθεί, διαγραφεί, να έχει υποστεί αλλαγές, ή τίποτα από τα προηγούμενα.

- Τη χρησιμότητας της οντότητας. Μια οντότητα μπορεί να παραμένει μονίμως στη



Σχήμα 3.3: Σύνδεση και περιγραφή των τμημάτων οντοτήτων για μια από τις δομές αποθήκευσης (αναπαραγωγή από [47]).

μνήμη ή να μη χρειάζεται πλέον, οπότε μπορεί να απελευθερωθεί ο χώρος της μνήμης που καταλαμβάνει.

3.2.2 Μηχανισμός διαχείρισης οντοτήτων

Οι οντότητες οργανώνονται σε τμήματα οντοτήτων και τα τμήματα οντοτήτων αποτελούν τη μονάδα μεταφοράς από το δίσκο στη μνήμη και αντίστροφα. Τα τμήματα οντοτήτων έχουν σταθερό μέγεθος (ίσο με το μέγεθος της σελίδας δίσκου) και περιέχουν πάντοτε οντότητες ενός μονάχα τύπου. Ο τύπος αυτός αντιστοιχεί στις δομές αναπαράστασης οντοτήτων που περιγράφηκαν παραπάνω, δηλαδή τις δομές class node, link class, token node, link token καθώς επίσης και τις δομές επέκτασης για τις διάφορες οντότητες. Τα τμήματα οντοτήτων που περιέχουν οντότητες του ίδιου τύπου οργανώνονται σε διπλά συνδεδεμένες λίστες οντοτήτων. Στο σύστημα διαχείρισης οντοτήτων για τη γλώσσα SIS-Telos υπάρχουν πέντε διαφορετικές λίστες τμημάτων οντοτήτων που αντιστοιχούν στις πέντε διαφορετικές δομές αποθήκευσης οντοτήτων. Η δομή των τμημάτων οντοτήτων και η λίστα που δημιουργείται από τα τμήματα οντοτήτων για μια δομή αποθήκευσης παρουσιάζονται στο σχήμα 3.3.

Για την προσπέλαση στις οντότητες, χρησιμοποιούνται μηχανισμοί κρυφής μνήμης, οι οποίοι περιέχουν τα τμήματα οντοτήτων που βρίσκονται στη μνήμη. Οι μηχανισμοί αυτοί περιγράφονται στο τμήμα 3.3.

Οταν απαιτείται πρόσβαση σε μια οντότητα για την οποία το τμήμα εγγραφών που την περιέχει βρίσκεται ήδη στη μνήμη, η δομή αναπαράστασης για την οντότητα αντιγράφεται από το τμήμα οντοτήτων που την περιέχει σε μια νέα θέση στη μνήμη, με τη βοήθεια του καταλόγου συστήματος. Αν και με την υλοποίηση αυτή δημιουργούνται αντίγραφα των οντοτήτων στη μνήμη και δυσκολεύει η διαχείρισή τους, η επιλογή επιβλήθηκε εξαιτίας προβλημάτων συμβατότητας σε διαφορετικά περιβάλλοντα της έκδοσης της γλώσσας C++ που χρησιμοποιήθηκε.

3.2.3 Κατάλογος λογικών ονομάτων

Οπως αναφέρθηκε και παραπάνω, το σύστημα διαχείρισης οντοτήτων για οποιαδήποτε εφαρμογή χρησιμοποιεί βάση δεδομένων κατασκευασμένη με την SIS-Telos χρησιμοποιεί τα αναγνωριστικά συστήματος για να προσπελάζει τις οντότητες της βάσης. Καθώς, όμως, ο χρήστης αντιλαμβάνεται και χρησιμοποιεί τις οντότητες μέσω των λογικών τους ονομάτων, δημιουργείται η ανάγκη ύπαρξης δύο μηχανισμών :

- ενός μηχανισμού μετάφρασης των λογικών ονομάτων των διαφόρων οντοτήτων στα αναγνωριστικά συστήματος που τους αντιστοιχούν και
- ενός μηχανισμού μετάφρασης των αναγνωριστικών συστήματος για τις διάφορες οντότητες στα λογικά ονόματα που τους αντιστοιχούν.

Υποκατάλογος αναγνωριστικών συστήματος

Ο κατάλογος αυτός χρησιμοποιείται από το Σύστημα Σημασιολογικού Ευρετηριασμού προκειμένου να ανακληθεί το λογικό όνομα μιας οντότητας για την οποία είναι γνωστό το αναγνωριστικό συστήματος που την προσδιορίζει.

Ο καθορισμός του λογικού ονόματος για κάποια οντότητα της οποίας το αναγνωριστικό συστήματος είναι γνωστό είναι μια πολύ συχνή πράξη που επιτελείται πριν από οποιαδήποτε παρουσίαση της οντότητας στο χρήστη.

Καθώς οι προσπελάσεις στον υποκατάλογο αναγνωριστικών συστήματος γίνονται μέσω των αναγνωριστικών συστήματος που προσδιορίζουν τα αντίστοιχα λογικά ονόματα για

τις οντότητες, ο υποκατάλογος αναγνωριστικών συστήματος μπορεί να οργανωθεί ανάλογα με τον κατάλογο συστήματος. Ετσι, ο υποκατάλογος αναγνωριστικών συστήματος είναι ένας πίνακας από εγγραφές για κάθε οντότητα της βάσης. Κάθε εγγραφή σχετίζει ένα αναγνωριστικό συστήματος κάποιας οντότητας με το λογικό όνομα της οντότητας.

Οπως και στην περίπτωση του καταλόγου συστήματος, ο πίνακας αυτός οργανώνεται σαν πίνακας από τμήματα με βαθμό εμμεσότητας 1. Ετσι, τα αναγνωριστικά συστήματος αποτελούν και πάλι τους δείκτες μέσω των οποίων γίνεται η προσπέλαση στον υποκατάλογο αναγνωριστικών συστήματος καθώς θεωρούνται ως διανύσματα της μορφής <δείκτης τμήματος, δείκτης εγγραφής>.

Υποκατάλογος λογικών ονομάτων

Ο κατάλογος αυτός χρησιμοποιείται από το Σύστημα Σημασιολογικού Ευρετηριασμού για τον εντοπισμό του αναγνωριστικού συστήματος μιας οντότητας η οποία προσδιορίζεται από το χρήστη μέσω του λογικού της ονόματος. Για το σχεδιασμό και την υλοποίηση του μηχανισμού αναζήτησης κατά τη μετάφραση λογικών ονομάτων στα αντίστοιχα αναγνωριστικά συστήματος, επιλέχθηκε ο μηχανισμός γραμμικού κατακερματισμού.

- **Γραμμικός κατακερματισμός**

  Ο μηχανισμός γραμμικού κατακερματισμού αναπτύχθηκε προκειμένου να χρησιμοποιηθεί για οργάνωση εξωτερικών αρχείων στα οποία χρησιμοποιούνται κλειδιά αναζήτησης.

  Ας θεωρήσουμε έναν κατάλογο κατακερματισμού (hash table) που αποτελείται από N ομάδες εγγραφών (buckets). Η συνάρτηση κατακερματισμού η οποία υπολογίζει σε ποια ομάδα βρίσκεται ένα κλειδί αναζήτησης K έχει τη μορφή :

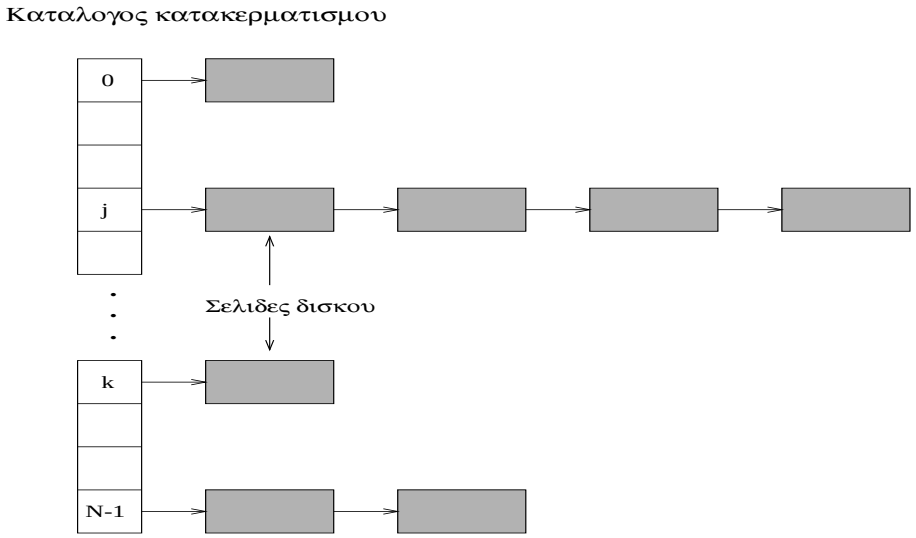
  $$h(K) = g(K) \bmod N$$

  Η συνάρτηση g(K), που ονομάζεται συνάρτηση κατακερματισμού, είναι μια συνάρτηση που μετασχηματίζει το κλειδί K σε μια διεύθυνση που βρίσκεται στο διάστημα [0, M] όπου το M είναι αρκετά μεγάλο.

Ο κατάλογος κατακερματισμού για τα λογικά ονόματα οργανώνεται με τη μορφή πίνακα από τμήματα. Κάθε τμήμα του καταλόγου αυτού αντιστοιχεί σε συγκεκριμένο πλήθος από ομάδες εγγραφών, ενώ κάθε ομάδα εγγραφών αντιστοιχεί σε μία ακριβώς θέση του καταλόγου κατακερματισμού. Το μέγεθος των τμημάτων είναι ίδιο με το μέγεθος της μονάδας μεταφοράς πληροφορίας μεταξύ δίσκου και μνήμης, της σελίδας δίσκου.

Η προσπέλαση στις εγγραφές οντοτήτων γίνεται με τρόπο αντίστοιχο με την προσπέλαση στις πλειάδες του υποκαταλόγου αναγνωριστικών συστήματος. Τα πιο σημαντικά δυαδικά ψηφία της τιμής h(K) επιλέγουν το τμήμα στο οποίο βρίσκεται η ζητούμενη ομάδα εγγραφών, ενώ τα λιγότερο σημαντικά ψηφία επιλέγουν την ομάδα εγγραφών στο τμήμα.

Κάθε ομάδα εγγραφών μπορεί να περιέχει μέχρι ένα καθορισμένο πλήθος από εγγραφές. Επειδή το πλήθος των εγγραφών που αντιστοιχούν σε μια συγκεκριμένη θέση του καταλόγου κατακερματισμού μπορεί να είναι μεγαλύτερο από το πλήθος εγγραφών που μπορεί να περιέχει μια ομάδα εγγραφών, οι ομάδες εγγραφών μπορούν να συνδέονται δημιουργώντας λίστες από ομάδες εγγραφών. Τότε η συγκεκριμένη θέση του καταλόγου κατακερματισμού περιέχει την πρωτεύουσα ομάδα εγγραφών (primary bucket), η οποία αποτελεί και την αρχή της λίστας εγγραφών για τη θέση αυτή. Τις υπόλοιπες ομάδες εγγραφών οι οποίες ονομάζονται ομάδες εγγραφών υπερχείλισης (overflow buckets) διαχειρίζεται ο μηχανισμός βοηθητικών ομάδων εγγραφών.



Σχήμα 3.4: Δομή του καταλόγου κατακερματισμού για την κρυφή μνήμη (αναπαραγωγή από [47]).

Κάθε εγγραφή περιλαμβάνει δύο πεδία : την τιμή g(K) για κάποιο συγκεκριμένο λογικό όνομα K και το αναγνωριστικό συστήματος για την οντότητα η οποία χαρακτηρίζεται από το λογικό όνομα K.

Οι εγγραφές μιας ομάδας ταξινομούνται ως προς την τιμή του πεδίου που περιέχει το αποτέλεσμα της συνάρτησης g(K). Ετσι, η εύρεση μιας συγκεκριμένης εγγραφής μέσα σε μια ομάδα γίνεται μέσω δυαδικής αναζήτησης στον πίνακα εγγραφών της ομάδας. Η δομή του καταλόγου κατακερματισμού παρουσιάζεται στο σχήμα 3.4.

Οταν σε περισσότερες από μία οντότητες αντιστοιχεί η ίδια τιμή της συνάρτησης μετασχηματισμού πρέπει να υπάρχει κάποιος τρόπος ώστε να αποθηκεύονται διαφορετικές εγγραφές για κάθε μία από τις οντότητες αυτές. Στην περίπτωση αυτή, δημιουργείται μια εγγραφή η οποία περιέχει την τιμή της συνάρτησης μετασχηματισμού στο ένα πεδίο της και έναν δείκτη σε βοηθητική ομάδα εγγραφών στο άλλο.

- **Μηχανισμός βοηθητικών εγγραφών ομάδων**

  Ο μηχανισμός διαχείρισης βοηθητικών εγγραφών αναλαμβάνει τη διάθεση τέτοιων ομάδων όποτε αυτές χρειάζονται, καθώς επίσης και κάθε προσπέλαση σ'αυτές. Ο μηχανισμός περιλαμβάνει έναν κατάλογο βοηθητικών ομάδων εγγραφών, ο οποίος οργανώνεται και λειτουργεί όπως ο κατάλογος συστήματος.

3.3 Μηχανισμός κρυφής μνήμης για το σύστημα διαχείρισης οντοτήτων

Ο μηχανισμός διαχείρισης κρυφής μνήμης για το σύστημα διαχείρισης οντοτήτων στη γλώσσα SIS-Telos δεν αποτελεί ένα ενιαίο σύστημα, αλλά διαχωρίζεται σε υπομηχανισμούς, κάθε ένας από τους οποίους αναλαμβάνει τη διαχείριση κρυφής μνήμης για μία από τις δομές αποθήκευσης και παράστασης που περιγράφηκαν παραπάνω. Η χρήση επιμέρους μηχανισμών προσφέρει τη δυνατότητα να ρυθμίζεται το μεγέθος της κρυφής μνήμης και οι αλγόριθμοι αναζήτησης για κάθε μηχανισμό ξεχωριστά. Με τον τρόπο αυτό το σύστημα γίνεται πιο ευέλικτο καθώς είναι δυνατό να ρυθμιστεί ξεχωριστά η συμπεριφορά για κάθε μηχανισμό διαχείρισης κρυφής μνήμης.

Οι επιμέρους αυτοί μηχανισμοί έχουν ανάλογη δομή και το μόνο που αλλάζει σε αυτούς είναι η δομή των περιεχομένων των τμημάτων για κάθε ενταμιευτή και ο τρόπος προσπέλασης για τους ενταμιευτές.

Οι κατάλογοι της κρυφής μνήμης που αποθηκεύουν εγγραφές του καταλόγου συστήματος και των υποκαταλόγων λογικών ονομάτων και αναγνωριστικών συστήματος οργανώνονται με τη μορφή καταλόγων κατακερματισμού (hash table). Το κλειδί αναζήτησης ενός αντικειμένου για καθέναν από τους καταλόγους κατακερματισμού είναι η λογική διεύθυνση της σελίδας που το περιέχει, με άλλα λόγια το μοναδικό αναγνωριστικό συστήματος που του αντιστοιχεί. Εξαιρέσεις στο μηχανισμό αυτό αποτελούν ο υποκατάλογος λογικών ονομάτων, στον οποίο η αναζήτηση μιας σελίδας πραγματοποιείται με βάση το μετασχηματισμό του λογικού ονόματος της οντότητας της οποίας αναζητάμε το αναγνωριστικό συστήματος και ο υποκατάλογος βοηθητικών εγγραφών στον οποίο κλειδί αναζήτησης αποτελεί η διεύθυνση βοηθητικής ομάδας εγγραφών.

Η τιμή της συνάρτησης κατακερματισμού για τους καταλόγους αυτούς υπολογίζεται από τα N λιγότερο σημαντικά ψηφία της λογικής διεύθυνσης του κάθε τμήματος. Ετσι, αν K είναι η λογική διεύθυνση κάποιου τμήματος και N είναι το μέγεθος του καταλόγου κατακερματισμού, η συνάρτηση κατακερματισμού ορίζεται ως εξής :

$$h(K) = K \bmod N$$

3.3.1 Επιμέρους μηχανισμοί κρυφής μνήμης

Κατάλογος συστήματος

Οι σελίδες του καταλόγου κατακερματισμού για την κρυφή μνήμη του καταλόγου συστήματος φιλοξενούν τμήματα εγγραφών με τη μορφή που παρουσιάστηκαν παραπάνω.

Το μέγεθος του καταλόγου κατακερματισμού είναι N = 128 και κάθε θέση του καταλόγου μπορεί να περιέχει μέχρι 8 τμήματα εγγραφών. Τα τμήματα εγγραφών, εξάλλου, περιέχουν δείκτες στη μνήμη για τις οντότητες που αντιγράφονται εκεί όπως αναφέρθηκε παραπάνω, προκειμένου να χρησιμοποιηθούν.

Ο αλγόριθμος αντικατάστασης που χρησιμοποιείται είναι μια παραλλαγή του αλγορίθμου LRU (Least Recently Used) μέσω της οποίας αντικαθίσταται κάθε φορά η σελίδα η οποία έχει μείνει αχρησιμοποίητη για τον περισσότερο χρόνο, με την προϋπόθεση να μην έχει τροποποιηθεί το περιεχόμενό της.

Στο σημείο αυτό, πρέπει να αναφερθεί ότι ο κατάλογος συστήματος διαιρείται σε δύο μέρη : το τμήμα του καταλόγου που αφορά οντότητες με δοσμένο από το χρήστη λογικό όνομα και το τμήμα που αφορά οντότητες χωρίς δοσμένο όνομα, οι οποίες ονοματοδοτούνται με κατάλληλο μηχανισμό από το σύστημα.

Τμήματα οντοτήτων

Καθεμιά από τις δομές αναπαράστασης για τους διαφορετικούς τύπους οντοτήτων στο εννοιολογικό μοντέλο της SIS-Telos αποθηκεύεται σε χωριστά τμήματα οντοτήτων. Ετσι, για καθέναν από τους τέσσερις διαφορετικούς τύπους οντοτήτων καθώς επίσης και για τις δομές επέκτασης οντοτήτων, αντιστοιχεί και ένας διαφορετικός μηχανισμός κρυφής μνήμης.

Σε αντίθεση με τους μηχανισμούς για τις υπόλοιπες δομές αποθήκευσης, ο μηχανισμός κρυφής μνήμης για τα τμήματα οντοτήτων και τις δομές επέκτασής τους χρησιμοποιεί έναν μονοδιάστατο πίνακα εγγραφών, μεγέθους N2 = 25, η πρόσβαση στον οποίο πραγματοποιείται όπως η πρόσβαση στα τμήματα των καταλόγων κατακερματισμού.

Ο αλγόριθμος αντικατάστασης που χρησιμοποιείται είναι η ίδια παραλλαγή του LRU

που περιγράφηκε παραπάνω.

Κατάλογος λογικών ονομάτων

Στον κατάλογο λογικών ονομάτων χρησιμοποιούνται 3 διαφορετικοί μηχανισμοί κρυφής μνήμης που αντιστοιχούν στις δομές :

- υποκαταλόγου αναγνωριστικών συστήματος,
- υποκαταλόγου λογικών ονομάτων και
- υποκαταλόγου βοηθητικών εγγραφών.

Τα μεγέθη των κρυφών μνημών για καθέναν από τους τρεις υποκαταλόγους είναι N = 128 και κάθε θέση του καταλόγου κατακερματισμού μπορεί να περιέχει έως 8 εγγραφές, με εξαίρεση τον υποκατάλογο λογικών ονομάτων, στην κρυφή μνήμη του οποίου κάθε θέση μπορεί να περιέχει έως 32 εγγραφές.
Ο αλγόριθμος αντικατάστασης που χρησιμοποιείται για τους μηχανισμούς αυτούς είναι ο LRU, με την παράλλαγη που ισχύει και για τον κατάλογο συστήματος.

Κεφάλαιο 4

Ανάλυση της διαχείρισης κρυφής μνήμης

4.1 Στοιχεία που χαρακτηρίζουν ένα σύστημα διαχείρισης κρυφής μνήμης

Η κρυφή μνήμη ενός συστήματος αποτελεί το σημείο της μνήμης στο οποίο μπορεί να αναζητήσει κάθε εφαρμογή την πληροφορία την απαραίτητη για τη διεκπεραίωσή της. Ο λόγος μεταφοράς της πληροφορίας από το δίσκο στη μνήμη προκειμένου να χρησιμοποιηθεί εντοπίζεται στο μεγάλο κόστος που συνεπάγεται η επικοινωνία με το δίσκο, τόσο για ανάγνωση, όσο και για εγγραφή δεδομένων. Αντίθετα, το κόστος επικοινωνίας μιας εφαρμογής με τη μνήμη είναι 40-200 χιλιάδες φορές μικρότερο. Ο λόγος για τον οποίο η κρυφή μνήμη πρέπει να διαθέτει μηχανισμούς, όχι μόνο για την αποθήκευση, αλλά και για την αποδοτική οργάνωση της πληροφορίας, εξάλλου, είναι ότι πληροφορία που περιέχεται σ'αυτήν είναι δυνατό να ξαναχρησιμοποιηθεί από την ίδια εφαρμογή. Είναι, επομένως, απαραίτητο να παρέχεται από την κρυφή μνήμη μια στοιχειώδης έστω οργάνωση των περιεχομένων της, ώστε πληροφορία που αποθηκεύεται να είναι δυνατό να ξαναβρεθεί για να ξαναχρησιμοποιηθεί. Η οργάνωση πρέπει να είναι τέτοια που να επιτρέπει τον κατά το δυνατό γρηγορότερο εντοπισμό της περιεχόμενης πληροφορίας.

Για την επίτευξη των παραπάνω, κάθε σύστημα διαχείρισης κρυφής μνήμης χαρακτηρίζεται από κάποιες παραμέτρους, οι οποίες μπορεί να είναι :

- Η δομή-οργάνωση της κρυφής μνήμης. Η παράμετρος αυτή έχει άμεση σχέση με την αποδοτική αναζήτηση στην κρυφή μνήμη. Ετσι, μια κρυφή μνήμη μπορεί να έχει τη μορφή μονοδιάστατου ή δι-διάστατου πίνακα, λίστας κ.ο.κ.

- Το πλήθος των δομών αποθήκευσης. Η παράμετρος αυτή, που είναι πιο συνηθισμένη στους μηχανισμούς κρυφής μνήμης για συστήματα διαχείρισης βάσεων δεδομένων, βοηθά στην ομαδοποίηση πληροφορίας διαφορετικού τύπου.

- Το μέγεθος της δομής αποθήκευσης στην κρυφή μνήμη. Το μέγεθος αυτό πρέπει να πληροί τις εξής προϋποθέσεις : να είναι αρκετό ώστε να περιλαμβάνει κατά το δυνατό το μεγαλύτερο ποσοστό της πληροφορίας που απαιτείται για την εκτέλεση μιας εφαρμογής αλλά και να μην είναι τόσο μεγάλο που η αναζήτηση σε αυτό να είναι υπολογιστικά δαπανηρή. Το μέγεθος περιορίζεται, εξάλλου, και από τους πόρους σε μνήμη που διαθέτει ο υπολογιστής στον οποίο εκτελείται η εφαρμογή.

- Ο αλγόριθμος αντικατάστασης. Ο αλγόριθμος αυτός προσδιορίζει την πληροφορία που πρέπει να αντικατασταθεί στην κρυφή μνήμη, όταν αυτή είναι πλήρης, προκειμένου να εισαχθεί νέα πληροφορία.

- Η μονάδα αποθήκευσης πληροφορίας, η οποία είναι κατά κανόνα ίδια με τη μονάδα αποθήκευσης της πληροφορίας στο δίσκο, τη σελίδα δίσκου.

- Ο τρόπος εισαγωγής πληροφορίας στη μνήμη. Οι διαθέσιμοι τρόποι είναι να εισάγεται κάθε φορά η απαιτούμενη πληροφορία μόνο ή να εισάγεται μαζί με πληροφορία η οποία πιστεύεται ότι θα χρησιμοποιηθεί από την εφαρμογή στο άμεσο μέλλον, ώστε να αποφευχθεί μελλοντική νέα πρόσβαση στο δίσκο για την εισαγωγή της. Η δεύτερη αυτή μέθοδος ονομάζεται προανάκληση (prefetching).

Ο καθορισμός των παραπάνω παραμέτρων προσδιορίζει πλήρως σε κάθε σύστημα το μηχανισμό διαχείρισης κρυφής μνήμης.

4.2 Σχεδίαση κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού

Τα λειτουργικά συστήματα των υπολογιστικών συστημάτων διαθέτουν μηχανισμούς διαχείρισης της κρυφής μνήμης. Παρ'όλα αυτά, τα συστήματα διαχείρισης βάσεων

δεδομένων διαθέτουν κατά κανόνα δικούς τους μηχανισμούς εξαιτίας της ιδιαιτερότητας που παρουσιάζουν τόσο στην οργάνωση, όσο και στη χρήση της πληροφορίας που διαχειρίζονται.

Ακολουθώντας, λοιπόν, το γενικό κανόνα, το Σύστημα Σημασιολογικού Ευρετηριασμού παρέχει, όπως περιγράφηκε στο προηγούμενο κεφάλαιο, δικούς τους μηχανισμούς για τη διαχείριση της κρυφής μνήμης του συστήματος. Τα χαρακτηριστικά των μηχανισμών αυτών, με βάση την παραπάνω κατηγοριοποίηση είναι συνοπτικά τα ακόλουθα :

- Υπαρξη επιμέρους μηχανισμών κρυφής μνήμης για καθεμιά από τις δομές αποθήκευσης του συστήματος.
- Δομή καταλόγου κατακερματισμού για την κρυφή μνήμη του καταλόγου συστήματος, του υποκαταλόγου λογικών ονομάτων, του υποκαταλόγου αναγνωριστικών συστήματος και του υποκαταλόγου βοηθητικών εγγραφών. Δομή μονοδιάστατου πίνακα-ουράς (με οργάνωση FIFO) για τις κρυφές μνήμες των διαφορετικών τύπων οντοτήτων και των δομών επέκτασής τους.
- Ιδιο μέγεθος για όλους τους καταλόγους κατακερματισμού και ίδιο μέγεθος για όλους τους πίνακες-ουρές.
- Αλγόριθμο αντικατάστασης LRU (Least Recently Used) για όλους τους καταλόγους κατακερματισμού και FIFO (First-In First-Out) για όλους τους πίνακες-ουρές.
- Μονάδα αποθήκευσης η οποία έχει μέγεθος ίδιο με αυτό της σελίδας δίσκου για λόγους συμβατότητας κατά τη μεταφορά δεδομένων.
- Προανάκληση δεδομένων, η οποία πραγματοποιείται μόνο με την εισαγωγή στην κρυφή μνήμη των δεδομένων σε σελίδες, όταν από καθεμιά από αυτές απαιτείται μόνο μια εγγραφή.

4.3 Ανάλυση των παραμέτρων

Οι παραπάνω παράμετροι σχεδίασης του συστήματος διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού διαδραματίζουν πολύ σημαντικό ρόλο στην επίδοση του συστήματος, με δεδομένο ότι καθορίζουν τη συμπεριφορά του συστήματος κρυφής μνήμης. Οι παράμετροι αυτές διαφοροποιούνται με βάση το κατά πόσον μπορούν

ή όχι να αλλάξουν ”τιμές” σε δεδομένη υλοποίηση, δίνοντας έτσι διαφορετικές επιδόσεις (όχι απαραίτητα καλύτερες).

Εξετάζοντας, λοιπόν, τις παραμέτρους αυτές διαπιστώνουμε τα ακόλουθα :

- Η ύπαρξη διαφορετικών μηχανισμών κρυφής μνήμης για καθεμιά από τις δομές αποθήκευσης του συστήματος είναι ένα στοιχείο το οποίο δεν ενδείκνυται να αλλάξει. Κι αυτό, γιατί η ομαδοποίηση της πληροφορίας που απαιτείται για τη διαχείριση μιας βάσης δεδομένων με βάση την εννοιολογία ή τη λειτουργικότητά της συμβάλλει στην αποδοτικότερη αναζήτηση στην κρυφή μνήμη, σε λιγότερα κάθε φορά δεδομένα. Σε διαφορετική περίπτωση, η αναζήτηση μιας εγγραφής του καταλόγου συστήματος π.χ. θα ενείχε προσπέλαση σε εγγραφές του καταλόγου λογικών ονομάτων και του καταλόγου διαχείρισης οντοτήτων. Οι διαφορετικοί μηχανισμοί κρυφής μνήμης, εξάλλου, προσφέρουν και περισσότερη ευελιξία στην επιλογή της οργάνωσης κάθε κρυφής μνήμης, ανάλογα με τη λειτουργία της και τη συχνότητα των προσβάσεων σ’αυτή.

- Η μονάδα μεταφοράς πληροφορίας ανάμεσα στο δίσκο και τη μνήμη είναι στοιχείο το οποίο επίσης δεν αλλάζει, αφού ενδεχόμενη αλλαγή του θα δημιουργούσε προβλήματα κατακερματισμού στη μνήμη εξαιτίας του διαφορετικού μεγέθους των δομών αποθήκευσης και αναπαράστασης για τις διαφορετικές μονάδες πληροφορίας που περιέχονται στη βάση καθώς επίσης και μεγάλο υπολογιστικό κόστος για την εξαγωγή και την εισαγωγή πληροφορίας από/προς τις σελίδες.

- Η επιλογή οργάνωσης των καταλόγων της κρυφής μνήμης με τη μορφή πινάκων κατακερματισμού έγινε έπειτα από μελέτη ([47]) ως πιο συμφέρουσα κι αποδοτική. Κύριο επιχείρημα που συνηγορεί υπέρ αυτής της επιλογής είναι η έλλειψη αναγκαιότητας διατήρησης ξεχωριστών δομών σχετικά με την οργάνωση και μόνο της πληροφορίας.

- Ο αλγόριθμος αντικατάστασης σελίδων στην κρυφή μνήμη ο οποίος αποδείχθηκε με τη διεξαγωγή μετρήσεων επίδοσης στα πλαίσια της μεταπτυχιακής εργασίας του Γ. Γεωργιαννάκη ότι είναι πολύ ικανοποιητικός, δεν κρίνεται σκόπιμο να τροποποιηθεί. Κι αυτό, γιατί σε αντίθεση με άλλα συστήματα, σε συστήματα διαχείρισης βάσεων δεδομένων δεν υπάρχει πρότυπο αναφορών σε σελίδες, το οποίο

να καθιστά αποδοτικότερη τη χρησιμοποίηση ενός αλγορίθμου αντικατάστασης που εκμεταλλεύεται την τοπικότητα των προσπελάσεων, σε σχέση με τον LRU.

- Τα μεγέθη των κρυφών μνημών, τέλος, καθώς επίσης και το στοιχείο της ομοιομορφίας τους για όλες τις κρυφές μνήμες του Συστήματος Σημασιολογικού Ευρετηριασμού με την ίδια δομή, είναι ένα χαρακτηριστικό που μπορεί και σε ορισμένες περιπτώσεις επιβάλλεται να αλλάξει. Το σχετικό μέγεθος του εννοιολογικού σχήματος μιας βάσης δεδομένων προς το μέγεθος των δεδομένων του είναι ένας από τους λόγους που μπορούν να οδηγήσουν στην επιλογή τροποποίησης των μεγεθών. Το γεγονός, εξάλλου, ότι κάθε βάση παρουσιάζει πληρότητα σε κάποιες από τις κρυφές μνήμες του συστήματος, σε αντίθεση με άλλες εξαιτίας της ίδιας της εννοιολογίας της συνηγορεί κι αυτό στην ίδια κατεύθυνση μελέτης.

Από την παραπάνω ανάλυση, προκύπτει ότι κύρια παράμετρος για τη μελέτη του συστήματος διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού αποτελούν τα μεγέθη των κρυφών μνημών. Τα υπόλοιπα στοιχεία, είτε επιβάλλεται είτε είναι προτιμότερο να μην αλλάξουν, αφού ενδεχόμενη αλλαγή τους θα είχε μικρό ή κανένα θετικό αποτέλεσμα.

Κεφάλαιο 5

Σχεδίαση πειραμάτων

5.1 Εισαγωγή

Το Σύστημα Σημασιολογικού Ευρετηριασμού παρέχει, όπως φάνηκε και στο 3ο κεφάλαιο, πολλές δυνατότητες και ευκολίες στο χρήστη που επιθυμεί ανάκτηση ή εισαγωγή πληροφορίας σε μια βάση γνώσης, χωρίς να διαθέτει ειδικές γνώσεις βάσεων δεδομένων και ερωτηματικών γλωσσών. Για να επιτύχει πλήρως το σκοπό του ένα τέτοιο σύστημα, επιβάλλεται, εκτός από λειτουργικό, να είναι και γρήγορο. Η ταχύτητα ενός συστήματος προσδιορίζει την ικανότητά του να διεκπεραιώσει ένα έργο σε ωφέλιμο χρόνο και αποτελεί σημαντικό παράγοντα για την ευχαρίστηση του χρήστη. Κύριος παράγοντας καθυστέρησης σε συστήματα διαχείρισης βάσεων δεδομένων είναι ο χρόνος αναζήτησης και ανάσυρσης της πληροφορίας από τη βάση. Τα συστήματα διαχείρισης βάσεων δεδομένων έχουν γενικά το χαρακτηριστικό της ανάγκης διαχείρισης μεγάλου όγκου πληροφορίας, γεγονός που τα καθιστά ”αργά” σε σχέση με συστήματα όπου η πληροφορία την οποία διαχειρίζονται είτε είναι ολιγοπληθέστερη είτε υπάρχει η δυνατότητα οργάνωσής της στην κρυφή μνήμη με βάση τον τρόπο χρησιμοποίησής της. Προκειμένου να εξακριβωθεί η επίδοση του συστήματος σε ό,τι έχει σχέση με την επίδοση της κρυφής του μνήμης, υπάρχουν στη βιβλιογραφία διάφορες μέθοδοι. Κυριότερη και περισσότερο διαδεδομένη από αυτές είναι η χρησιμοποίηση τυπικής δοκιμασίας επιδόσεων (benchmark), η οποία στηρίζεται στην κατασκευή μιας οντοκεντρικής βάσης δεδομένων με συγκεκριμένο περιεχόμενο και τον υπολογισμό της επίδοσης του συστήματος διαχείρισης της βάσης για συγκεκριμένες λειτουργίες. Βασικό μειονέκτημα της μεθόδου αυτής είναι ότι είναι προσανατολισμένη σε σχεδιαστικές

βάσεις δεδομένων (CAD/CAM) οι οποίες έχουν οργάνωση και δομή τέτοια ώστε να εξυπηρετούνται συγκεκριμένες λειτουργίες αναζήτησης κι επεξεργασίας, οι οποίες ωστόσο δε βρίσκουν εφαρμογή σε όλες τις οντοκεντρικές βάσεις δεδομένων. Οι συνηθέστερες λειτουργίες σε σχεδιαστικές βάσεις δεδομένων είναι η διάσχιση του γράφου των οντοτήτων κατά μήκος μονοπατιών καθώς επίσης και η εκτέλεση τμημάτων κώδικα για αντικείμενα που εντοπίζονται.

Το ΣΣΕ, από την άλλη μεριά, χρησιμοποιείται , κατά κύριο λόγο, σε πολιτισμικές εφαρμογές. Η ιδιαιτερότητα που αυτό το στοιχείο προσφέρει είναι ότι βασικό χαρακτηριστικό της πολιτισμικής πληροφορίας είναι η ύπαρξη πολλαπλών συνδέσεων μεταξύ των αντικειμένων που την απαρτίζουν και η αναζήτηση σ'αυτήν με τυχαίο, κατά κανόνα, τρόπο.

Κρίθηκε απαραίτητο, για το λόγο αυτό, να επιλεχθεί μια μέθοδος για τη μέτρηση της επίδοσης του συστήματος διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού, η οποία να προσφέρει την επιθυμητή πληροφορία, χωρίς τη χρήση των τυπικών δοκιμασιών επιδόσεων.

Η μέθοδος αυτή, είναι η μέτρηση χαρακτηριστικών τιμών συμπεριφοράς του συστήματος διαχείρισης κρυφής μνήμης χρησιμοποιώντας συγκεκριμένα σενάρια εκτέλεσης στο Σύστημα Σημασιολογικού Ευρετηριασμού για διαφορετικές βάσεις δεδομένων Κάθε μέτρηση εκτελείται για διαφορετικές τιμές κάποιων χαρακτηριστικών του συστήματος διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού, τα οποία αποτελούν τις παραμέτρους των μετρήσεων. Τα σενάρια εκτελούνται για διαφορετικές τιμές των παραμέτρων του πειράματος και συγκεντρώνονται μετρήσεις, οι οποίες χρησιμεύουν στην εξαγωγή συμπερασμάτων για τη συνολική συμπεριφορά του συστήματος,

Στη συνέχεια περιγράφονται οι παράμετροι του προβλήματος, τα μεγέθη που μετρούνται στα πειράματα, οι βάσεις δεδομένων που χρησιμοποιούνται καθώς επίσης και τα σενάρια χρήσης του συστήματος που επιλέχθηκαν.

5.2 Παράμετροι μέτρησης

Για τη διεξαγωγή των πειραμάτων χρειάστηκε αρχικά να καθοριστούν τα χαρακτηριστικά του συστήματος τα οποία θα αποτελούσαν σταθερές των μετρήσεων.

Σταθερές στις μετρήσεις επιλέγονται να είναι στοιχεία του συστήματος διαχείρισης κρυφής μνήμης τα οποία μπορούν να τροποποιηθούν σε ό,τι εξαρτάται από την υλοποίησή του συστήματος, αλλά επιλέγεται να διατηρούν αμετάβλητες τις τιμές τους σε όλα τα πειράματα.

Στις μετρήσεις επίδοσης του Συστήματος Σημασιολογικού Ευρετηριασμού σταθερές του συστήματος διαχείρισης κρυφής μνήμης, όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, αποφασίστηκε να αποτελέσουν :

- Το μέγεθος της μονάδας μεταφοράς πληροφορίας από το δίσκο στη μνήμη και αντίστροφα.
- Το πλήθος των κρυφών μνημών. Υπάρχει μία για κάθε δομή αναπαράστασης και αποθήκευσης της γλώσσας παράστασης γνώσης Telos, δηλαδή συνολικά δέκα κρυφές μνήμες για το σύστημα.
- Ο αλγόριθμος αντικατάστασης σελίδων στη κρυφή μνήμη όταν αυτή έχει γεμίσει και απαιτείται η μεταφορά μιας ακόμη σελίδας δεδομένων από το δίσκο στη μνήμη. Ο αλγόριθμος αυτός είναι ένας τροποποιημένος αλγόριθμος LRU (Least Recently Used) στον οποίο αντικαθίστανται στην κρυφή μνήμη μόνο σελίδες που δεν έχουν υποστεί καμμιά τροποποίηση για τις κρυφές μνήμες του καταλόγου συστήματος και των υποκαταλόγων λογικών ονομάτων, αναγνωριστικών συστήματος και βοηθητικών ομάδων εγγραφών. Για τις κρυφές μνήμες των δομών αναπαράστασης για τις τέσσερις διαφορετικές δομές αναπαράστασης οντοτήτων καθώς και για τις δομές επέκτασής τους χρησιμοποιείται αλγόριθμος FIFO (First-In First-Out) με την τροποποίηση που ισχύει και για τις υπόλοιπες κρυφές μνήμες.
- Η δομή των μηχανισμών κρυφής μνήμης, η οποία παραμένει δομή καταλόγου κατακερματισμού για τις κρυφές μνήμες του καταλόγου συστήματος, των υποκαταλόγων λογικών ονομάτων, αναγνωριστικών συστήματος και βοηθητικών ομάδων εγγραφών, ενώ για τις κρυφές μνήμες των εγγραφών για τις τέσσερις δομές αποθήκευσης της Telos καθώς επίσης και για τη εγγραφές δομών επέκτασης των οντοτήτων παραμένει η δομή του μονοδιάστατου πίνακα-ουράς.

Το χαρακτηριστικό του συστήματος διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού το οποίο αποφασίστηκε ότι θα αποτελέσει τη

μεταβλητή παράμετρο των μετρήσεων είναι το μέγιστο επιτρεπτό μέγεθος των κρυφών μνημών. Η επιλογή αυτή στηρίχθηκε κυρίως στο γεγονός ότι αποτέλεσμα των μετρήσεων αυτών επιθυμούσαμε να είναι μια πρόταση διαμόρφωσης του συστήματος σε διαφορετικές περιπτώσεις βάσεων δεδομένων με σκοπό τη βελτίωση της επίδοσής του σε ό,τι αφορά την ταχύτητά του. Σε αντίθεση με τις παραμέτρους που προαναφέρθηκαν, το μέγιστο επιτρεπτό μέγεθος μιας κρυφής μνήμης είναι ένα στοιχείο που εύκολα τροποποιείται για το σύστημα διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού, χωρίς να απαιτείται νέα υλοποίησή του.

5.2.1 Τιμές των παραμέτρων (μεταβλητών) των πειραμάτων

Στο 3ο κεφάλαιο έγινε αναφορά στο μηχανισμό διαχείρισης κρυφής μνήμης γαι το Σύστημα Σημασιολογικού Ευρετηριασμού. Εκεί αναφέρθηκε ότι το μέγεθος του πίνακα για όλες τις κρυφές μνήμες που έχουν δομή πίνακα κατακερματισμού εκτός από τον υποκατάλογο λογικών ονομάτων είναι $2^7 * 8$ θέσεις. Για τον πίνακα κατακερματισμού που αντιστοιχεί στην κρυφή μνήμη για τον υποκατάλογο λογικών ονομάτων, το μέγεθος του πίνακα ανεβαίνει σε $2^7 * 32$ θέσεις. Το μέγεθος των πινάκων για τις οντότητες αναπαράστασης και τις επεκτάσεις τους, εξάλλου, είναι 50 θέσεις.

Σε κάθε περίπτωση ορίζουμε μία μεταβλητή, η τιμή της οποίας θα καθορίζει το μέγεθος της αντίστοιχης κρυφής μνήμης :

- Πίνακας κατακερματισμού. Το μέγεθος της κρυφής μνήμης τίθεται $2^{Cache_Bits} * 8$ ή $2^{Cache_Bits} * 32$ στην περίπτωση του υποκαταλόγου λογικών ονομάτων.

- Μονοδιάστατος πίνακας. Το μέγεθος της κρυφής μνήμης τίθεται ίσο προς Max_Queue_Size.

Οι τιμές των παραπάνω μεταβλητών που αντιστοιχούν στην εν χρήσει έκδοση του ΣΣΕ είναι $Cache_Bits = 7$ και $Max_Queue_Size = 50$.

Για τη διεξαγωγή των πειραμάτων, επιλέγονται 3 τιμές για κάθε μεταβλητή έτσι ώστε να

ορίζουν μια ευρεία, για πρακτικούς σκοπούς, περιοχή.

$Cache_Bits$:

3	7	12
Small Cache	Medium Cache	Big Cache

Max_Queue_Size:

25	50	500
Small Queue	Medium Queue	Big Queue

Με δεδομένο ότι οι μέχρι τώρα τιμές των παραμέτρων για τις κρυφές μνήμες ήταν $Cache_Bits = 7$ γι'αυτές με δομή καταλόγου κατακερματισμού και $Max_Queue_Size = 50$ γι'αυτές με δομή μονοδιάστατου πίνακα-ουράς, η αιτία επιλογής των άλλων δύο τιμών για τις μετρήσεις ήταν η ανάγκη να ελεγχθεί η συμπεριφορά του συστήματος, τόσο όταν περιορίσουμε τη μέγιστη επιτρεπτή κατανάλωση μνήμης, όσο και όταν της επιτρέψουμε πολύ μεγάλες τιμές. Εκτιμώντας την επίδοση του συστήματος και στις δύο αυτές κατευθύνσεις σε σχέση με τις αρχικές προδιαγραφές, στοχεύουμε στο να διαπιστώσουμε σε ποια κατεύθυνση πρέπει να κινηθούν οι επιλογές μας στο μέλλον. Με άλλα λόγια, αν η επίδοση του συστήματος δεν επιδεινώνεται για κάποια βάση δεδομένων με τη μείωση του μέγιστου επιτρεπτού μεγέθους των κρυφών μνημών και με δεδομένο ότι η κατανάλωση μνήμης, εκτός από ευκολία, αποτελεί και κατανάλωση πόρων για το σύστημα, προκύπτει ότι συμφέρει η μείωση των μεγεθών των κρυφών μνημών για τη χρησιμοποίηση του Συστήματος Σημασιολογικού Ευρετηριασμού για τη συγκεκριμένη βάση.

Εκτός από μετρήσεις για τις παραπάνω τιμές των μεταβλητών παραμέτρων του προβλήματος, εκτελέστηκαν μετρήσεις και για ακόμη χαμηλότερες τους. Αφορμή για τις περαιτέρω μετρήσεις αποτέλεσαν συγκεκριμένα πειραματικά αποτελέσματα, οπότε οι επιπλέον αυτές μετρήσεις θα περιγραφούν μαζί με αυτά.

5.2.2 Επιλογή συνδυασμών μετρήσεων για τις διαφορετικές τιμές των παραμέτρων

Με δεδομένο ότι οι τιμές των παραπάνω παραμέτρων τροποποιούνται και οι δύο για τη διεξαγωγή μιας μέτρησης οι μετρήσεις πραγματοποιήθηκαν με συνδυασμούς τιμών των παραμέτρων. Οι συνδυασμοί παραμέτρων για τις οποίες έγιναν μετρήσεις είναι οι ακόλουθοι :

	Small Cache	Medium Cache	Big Cache
Small Queue	X	X	-
Medium Queue	X	X	X
Big Queue	-	X	X

Αποφασίστηκε να μη διεξαχθούν μετρήσεις για τους συνδυασμούς SmallCache-BigQueue και BigCache-SmallQueue γιατί θεωρήθηκε ότι η χρησιμοποίηση αυτών των συνδυασμών τιμών θα είχε ως αποτέλεσμα την παράλληλη ύπαρξη στους διαφορετικούς μηχανισμούς κρυφής μνήμης πληροφορίας που δεν αντιστοιχεί σε ίδιες οντότητες της βάσης, γεγονός που δε συμβάλλει στη λειτουργικότητα του συστήματος.

5.3 Μεγέθη που καταγράφονται από τα πειράματα

Προκειμένου να προσδιοριστεί η συμπεριφορά του μηχανισμού διαχείρισης κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού ήταν αναγκαίο να επιλεχθούν τα χαρακτηριστικά της που θα αποτελούσαν το κριτήριο για την αξιολόγηση.

- Hit Rate.

  Κύριο στοιχείο αξιολόγησης μηχανισμών κρυφής μνήμης σε όλη την έκταση της βιβλιογραφίας αποτελεί το ποσοστό επιτυχών προσβάσεων στην κρυφή μνήμη στο σύνολο των προσβάσεων που πραγματοποιούνται σ'αυτήν (hit rate). Μια πρόσβαση στην κρυφή μνήμη ενός συστήματος θεωρείται επιτυχής όταν, η πληροφορία για την οποία πραγματοποιείται η πρόσβαση στην κρυφή μνήμη βρίσκεται ήδη εκεί. Αποτυχία, αντίθετα, ονομάζεται η περίπτωση να μη βρίσκεται στην κρυφή μνήμη η

πληροφορία που χρειάζεται, γεγονός που προκαλεί την ανάγνωση της πληροφορίας από το μαγνητικό μέσο αποθήκευσης.

Καθώς, λοιπόν, το ποσοστό επιτυχών προσβάσεων στην κρυφή μνήμη καθορίζει και το ποσοστό των προσβάσεων που προκαλούν πρόσβαση στο μαγνητικό μέσο, κι εφόσον ο χρόνος απόκρισης ενός συστήματος αποτελεί κριτήριο αξιολόγησης της συμπεριφοράς του, αποφασίστηκε το ποσοστό επιτυχών προσβάσεων στην κρυφή μνήμη να αποτελέσει ένα από τα μεγέθη τα οποία θα υπολογίζουν οι μετρήσεις.

- Ποσοστό πληρότητας κρυφής μνήμης.
  Δεύτερο μέγεθος το οποίο καταγράφουν οι μετρήσεις είναι το τρέχον κάθε φορά μέγεθος όλων των επιμέρους μηχανισμών κρυφής μνήμης κι επομένως και το ποσοστό πληρότητας που επιτυγχάνουν. Η καταγραφή των τρεχόντων μεγεθών για τις επιμέρους κρυφές μνήμες είναι απαραίτητη προκειμένου να κριθεί ποιες από αυτές έχουν μεγαλύτερη ανάγκη χώρου στη μνήμη και παρ'όλα αυτά καλή επίδοση - η οποία προσδιορίζεται μέσω του ποσοστού επιτυχίας που αυτές εμφανίζουν - και ποιες δεν συμπληρώνουν καν το χώρο που τους διατίθεται. Οι μηχανισμοί που παρουσιάζουν κάποια από τις παραπάνω συμπεριφορές μπορούν να διαθέσουν τμήμα της μνήμης που τους διατίθεται για να καλυφθούν οι ανάγκες των απαιτητικότερων μηχανισμών.

- Οντότητες και επεκτάσεις οντοτήτων στη μνήμη.
  Εκτός από τους μηχανισμούς κρυφής μνήμης για τις διαφορετικές δομές αποθήκευσης κι αναπαράστασης, αναφέρθηκε στο 3ο κεφάλαιο ότι υπάρχει στο σύστημα άλλη μια πηγή κατανάλωσης μνήμης στο σύστημα. Αυτή προέρχεται από το γεγονός ότι, προκειμένου να χρησιμοποιηθεί μια δομή αναπαράστασης οντότητας ή επέκτασης οντότητας αυτή αντιγράφεται από την εγγραφή της κρυφής μνήμης που την περιέχει σε νέο σημείο της μνήμης, εξαιτίας προβλημάτων ακεραιότητας δεικτών σε άλλη περίπτωση (pointer swizzling). Η κατανάλωση μνήμης που προκαλείται από την αντιγραφή των δομών αυτών στη μνήμη είναι κι αυτή ένα στοιχείο που καταγράφεται από τις μετρήσεις. Κύριος στόχος αυτής της καταγραφής είναι να διαπιστωθεί η επιβάρυνση που προκαλεί στη σύστημα η μέθοδος αυτή, ώστε σε περίπτωση μεγάλης επιβάρυνσης να ελεγχθούν εναλλακτικές μέθοδοι αντιμετώπισης του προβλήματος.

- Σύνολα προσωρινής αποθήκευσης δεδομένων.

Μια τελευταία πηγή κατανάλωσης μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού είναι ο μηχανισμός διαχείρισης συνόλων ([28]). Προκειμένου να υποστηριχθούν μέθοδοι διαχείρισης συλλογών από αντικείμενα, κλάσεις και συσχετίσεις, όπως αυτά περιγράφονται στην Telos, όπως επίσης και για τους μηχανισμούς εκτέλεσης ερωτήσεων και σημασιολογικού ελέγχου (ο οποίος πραγματοποιείται κατά την εισαγωγή δεδομένων) υπάρχει ένας μηχανισμός με τη βοήθεια του οποίου υλοποιούνται ”σύνολα”. Αυτά δεν είναι τίποτε άλλο από ταξινομημένα σύνολα που περιέχουν αναγνωριστικά συστήματος οντοτήτων ή λογικά ονόματα και μπορούν να είναι σταθερού μεγέθους ή δυναμικά. Ο μηχανισμός διαχείρισης συνόλων για το Σύστημα Σημασιολογικού Ευρετηριασμού χρησιμεύει τόσο στην αποθήκευση της ”απάντησης” σε ένα ερώτημα (query) προς τη βάση, όσο και στην αποθήκευση ενδιάμεσων αποτελεσμάτων προκειμένου να υπολογιστεί μια τέτοια απάντηση.

Η αποθήκευση των συνόλων αυτών γίνεται στη μνήμη, γι’αυτό και το μέγεθός τους αποτελεί το τέταρτο μέγεθος που καταγράφεται από τις μετρήσεις. Από τις μετρήσεις επίσης καταγράφεται και το πλήθος των συνόλων που χρησιμοποιούνται.

Πιο συγκεκριμένα, σε ό,τι αφορά το μηχανισμό διαχείρισης συνόλων, τα μεγέθη που καταμετρούν τα πειράματα είναι η μέγιστη κι η ελάχιστη κατανάλωση μνήμης από το μηχανισμό στη διάρκεια των δοσοληψιών, ώστε να ελεγχθεί εκτός από την επιβάρυνση που προκαλεί ο μηχανισμός λόγω κατανάλωσης μνήμης και η ορθότητα του, καθώς η μνήμη επιβάλλεται να αποδεσμεύεται στο τέλος κάθε δοσοληψίας.

- Χρόνοι προσπέλασης.
  Τελευταίο στοιχείο καταμέτρησης από το σύστημα είναι οι χρόνοι πρόσβασης στη μνήμη σε περίπτωση επιτυχίας και σε περίπτωση αποτυχίας. Η μετρήσεις αυτές πραγματοποιήθηκαν για να δοθεί ένα πιο χειροπιαστό μέτρο της συμπεριφοράς του συστήματος, όπως αυτήν την αντιλαμβάνεται ο χρήστης.

  Πρέπει, ωστόσο, να αναφερθεί ότι καθώς οι μετρήσεις επίδοσης για το μηχανισμό διαχείρισης κρυφής μνήμης στα δύο συστήματα δεν πραγματοποιήθηκαν σε απομονωμένο υπολογιστή, οι χρόνοι που μετρήθηκαν είναι μόνο ενδεικτικοί και δεν μπορούν να αποτελέσουν στοιχείο χαρακτηρισμού της συμπεριφοράς των συστημάτων.

Βασικός στόχος των παραπάνω μετρήσεων είναι ο προσδιορισμός της συνολικής κατανάλωσης μνήμης στο σύστημα, καθώς επίσης και η συμπεριφορά του συστήματος σε ό,τι αφορά το μέσο χρόνο απόκρισης, στοιχείο που κύρια καθορίζεται από το ποσοστό των προσβάσεων στη μνήμη που είναι αποτυχημένες.

5.4 Βάσεις μετρήσεων

Προκειμένου να πραγματοποιηθούν οι μετρήσεις επίδοσης του μηχανισμού διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού έπρεπε να επιλεχθούν κάποιες βάσεις δεδομένων, για τις οποίες θα χρησιμοποιούσαμε το σύστημα, θέτοντάς τους ερωτήσεις κι ενημερώνοντάς τες.
Το κριτήριο επιλογής καθεμιάς από τις τρεις βάσεις που τελικά χρησιμοποιήθηκαν ήταν διαφορετικό. Βασική απαίτηση ήταν η παράσταση της βάσης μέσω της Telos.
Οι βάσεις δεδομένων που αποτέλεσαν την πλατφόρμα εκτέλεσης των μετρήσεων είναι οι ακόλουθες :

- **ΚΛΕΙΩ**

  Η βάση δεδομένων του συστήματος ΚΛΕΙΩ είναι μια βάση που κατασκευάστηκε για την παράσταση και την λεπτομερή τεκμηρίωση πολιτιστικών αντικειμένων. Κύριο χαρακτηριστικό της είναι ο πλούτος και η πολυπλοκότητα του σχήματός της. Τα δεδομένα της βάσης του ΚΛΕΙΩ, στη συγκεκριμένη εφαρμογή του που χρησιμοποιήθηκε για τα πειράματα, δεν είναι πολυάριθμα. Στην πραγματικότητα, το σχήμα και τα δεδομένα στη βάση αυτή παρουσιάζονται με σχετική αριθμητική ισοδυναμία. Πιο συγκεκριμένα, η βάση αυτή περιλαμβάνει 4500 περίπου ατομικές οντότητες, οι οποίες ταξινομούνται σε 720 κλάσεις και 10000 περίπου συσχετίσεις ανάμεσα σε ατομικές οντότητες που ταξινομούνται σε 310 κλάσεις συσχετίσεων. Οι ατομικές οντότητες και συσχετίσεις αποτελούν τα δεδομένα, ενώ οι κλάσεις οντοτήτων και συσχετίσεων αποτελούν το σχήμα της βάσης. Η αναλογία σχήματος και δεδομένων, επομένως, είναι 1 προς 15, ποσοστό αρκετά υψηλό, όπως θα φανεί και παρακάτω.

  Αυτό είναι και το κύριο χαρακτηριστικό που οδήγησε στην επιλογή της βάσης αυτής στη διεξαγωγή των μετρήσεων. Είναι μια βάση με πολύ πλούσιο σχήμα και όχι πάρα πολλά δεδομένα. Επιπλέον, η βάση του ΚΛΕΙΩ είναι μια βάση

καθαρά οντοκεντρική, με την έννοια ότι σχεδιάστηκε εξ αρχής χρησιμοποιώντας την εννοιολογία των οντοκεντρικών βάσεων δεδομένων, σε αντίθεση με άλλες βάσεις οι οποίες σχεδιάζονται σε σχεσιακό μοντέλο δεδομένων και στη συνέχεια μεταφέρονται στο οντοκεντρικό, πραγματοποιώντας μόνο τις απαραίτητες προσαρμογές. Ο εξ ολοκλήρου σχεδιασμός της βάσης του ΚΛΕΙΩ σε οντοκεντρικό μοντέλο δεδομένων, εξάλλου, είναι κι η αιτία του μεγέθους του σχήματός του.

- **AAT**

  Ο θησαυρός όρων AAT (Art & $Architecture\ Thesaurus$) είναι προϊόν του Art History Information Program (AHIP) του Ιδρύματος Getty και περιλαμβάνει ένα πρότυπο λεξικό όρων τέχνης και αρχιτεκτονικής. Η μεταφορά του στο SIS έγινε στα πλαίσια της συνεργασίας του ΙΠ/ΙΤΕ με το μουσείο Getty για τη δημιουργία ενός πιλοτικού συστήματος διαχείρισης λεξιλογίων, βασισμένου στο SIS.

  Η βάση του AAT προέρχεται από μεταφορά αντίστοιχης σχεσιακής βάσης δεδομένων σε Telos, χωρίς επανασχεδιασμό της. Για το λόγο αυτό, το σχήμα της βάσης είναι σχετικά περιορισμένο, κυρίως σε ό,τι αφορά τα επίπεδα στην ιεραρχία ταξινόμησης που περιλαμβάνει.

  Αντίθετα, η βάση του AAT συγκεντρώνει ένα μεγάλο όγκο δεδομένων, γεγονός που συνετέλεσε στην επιλογή της σε συνδυασμό με το ότι η υλοποίησή της σε Telos αποτελεί "μετάφραση" μιας σχεσιακής βάσης δεδομένων σε οντοκεντρικό σχήμα. Συγκεκριμένα, στη βάση περιλαμβάνονται 350000 ατομικές οντότητες που ταξινομούνται σε 4000 κλάσεις οντοτήτων και 1000000 συσχετίσεις ανάμεσα σε ατομικές οντότητες που ταξινομούνται σε 127 μόλις κλάσεις συσχετίσεων. Η αναλογία σχήματος - δεδομένων διαμορφώνεται εδώ σε 1 προς 270, αναλογία πολύ μικρότερη από την αντίστοιχη της βάσης δεδομένων του συστήματος ΚΛΕΙΩ.

- **ULAN**

  Η βάση του ULAN προέρχεται κι αυτή από το Μουσείο Μοντέρνας Τέχνης Getty και περιλαμβάνει πληροφορίες σχετικά με άτομα - καλλιτέχνες κι όχι μόνο - των οποίων τα προσωπικά στοιχεία ενδιαφέρουν το μουσείο, καθώς επίσης και πληροφορία σχετική με καταχωρήσεις που αφορούν τα άτομα αυτά και προέρχονται από διάφορες πηγές

Οπως και για τη βάση του AAT, η υλοποίησή της προέρχεται από μεταφορά σχεσιακής βάσης δεδομένων σε Telos, χωρίς εκ νέου σχεδιασμό της. Χαρακτηριστικό στοιχείο της βάσης αυτής είναι το πολύ μικρό (και πάλι) μέγεθος του σχήματός της, το οποίο περιορίζει την παράστασή του σε ένα μόνο επίπεδο ταξινόμησης και το μεγάλο πλήθος των δεδομένων της, τα οποία ξεπερνούν αριθμητικά τα 2,5 εκατομμύρια. Επίσης, χαρακτηριστικό της βάσης του ULAN είναι το πολύ μεγάλο πλήθος συνδέσμων (συσχετίσεων) που διαθέτει. Στη βάση αυτή, περιλαμβάνονται 850000 ατομικές οντότητες που ταξινομούνται σε 148 κλάσεις οντοτήτων και 2500000 συσχετίσεις ανάμεσα σε ατομικές οντότητες που ταξινομούνται σε 454 κλάσεις συσχετίσεων. Η αναλογία σχήματος - δεδομένων μειώνεται εδώ σε 1 προς 5000, μικρότερη κατά μία τάξη μεγέθους σε σχέση με την αντίστοιχη αναλογία στη βάση του AAT, η οποία αποτελεί στην ουσία μικρογραφία της βάσης ULAN.

Μια τέτοια βάση, βοηθά στη διερεύνηση της σχέσης των μεγεθών των επιμέρους κρυφών μνημών, σε περιπτώσεις δυσαναλογίας στην κατανομή των τύπων οντοτήτων.

5.5 Σενάρια χρήσης

Για τη διεξαγωγή των πειραμάτων ήταν αναγκαία η επιλογή σεναρίων εκτέλεσης τα οποία χρησιμοποιήθηκαν σε όλες τις μετρήσεις. Λέγοντας σενάρια εκτέλεσης εννοούμε συγκεκριμένα βήματα χρήσης του Συστήματος Σημασιολογικού Ευρετηριασμού, τα οποία εκτελούνται πάντα με την ίδια σειρά.

Τα σενάρια αυτά έπρεπε να αποτελούν τυπικά σενάρια χρήσης για καθεμιά από τις βάσεις. Τυπικό σενάριο χρήσης μιας βάσης δεδομένων θεωρείται αυτό που αποτελείται από βήματα τα οποία κατά κανόνα ακολουθούνται από τους πραγματικούς χρήστες των βάσεων στο χώρο εργασίας για τον οποίο αυτές κατ'αρχήν υλοποιήθηκαν.

Τα σενάρια θεωρήθηκε απαραίτητο να διαθέτουν κατά το δυνατό περισσότερες από τις δυνατότητες που προσφέρει το Σύστημα Σημασιολογικού Ευρετηριασμού για ανάκτηση και εισαγωγή πληροφορίας από/σε μια βάση δεδομένων.

Για να συνδυαστούν και οι δύο παραπάνω στόχοι, κατασκευάστηκαν σενάρια εκτέλεσης τα οποία αποτελούνται από τις ίδιες κατηγορίες βημάτων, οι οποίες όμως εφαρμόστηκαν με διαφορετικό τρόπο σε κάθε βάση ώστε να προσεγγίζουν τη συνηθισμένη εφαρμογή της.

Τα γενικά βήματα από τα οποία αποτελούνται τα σενάρια εκτέλεσης για τη διεξαγωγή των μετρήσεων είναι τα ακόλουθα :

1. Ερώτηση στη βάση που έχει ως αποτέλεσμα ένα σύνολο ατομικών οντοτήτων (tokens) της βάσης. Εκτέλεση ίδιας ερώτησης για όλα τα αντικείμενα της βάσης που περιλαμβάνονται στο αποτέλεσμα της πρώτης ερώτησης. Η διαφοροποίηση στο βήμα αυτό για κάθε βάση βρίσκεται στο είδος των ερωτήσεων που τίθενται.

   Στόχος της ύπαρξης αυτού του βήματος στο σενάριο εκτέλεσης είναι να διαπιστωθεί η συμπεριφορά του συστήματος σε αναδρομική εκτέλεση ερωτήσεων και να βγουν συμπεράσματα σχετικά με την αποδοτικότητα του μηχανισμού ομαδοποίησης των δομών αποθήκευσης και αναπαράστασης σε σελίδες που χρησιμοποιείται.

2. Εκτέλεση μιας ερώτησης με πολύ μεγάλο - αριθμητικά - αποτέλεσμα και μιας πολύ μικρότερης σε αποτέλεσμα ερώτησης σχετικά με κάποιο από τα στοιχεία που αποτελούν την απάντηση στην πρώτη ερώτηση.

   Στόχος του βήματος είναι να μελετηθεί η συμπεριφορά του συστήματος σε ερωτήσεις με μεγάλο αριθμητικά αποτέλεσμα, το οποίο επιστρέφεται στο σύνολό του στο χρήστη (και όχι τμηματικά). Η μετέπειτα εκτέλεση της ”μικρής” ερώτησης στοχεύει στο να εντοπιστεί το κατά πόσον η μεγάλη επιβάρυνση που επιφέρει στο σύστημα η εκτέλεση της προηγούμενης ερώτησης επιφέρει και το πλεονέκτημα εισαγωγής στην κρυφή μνήμη αρκετών σελίδων ώστε μετέπειτα αναφορές να είναι πάντα επιτυχημένες.

3. Χρησιμοποίηση των Δελτίων Εισαγωγής (Entry Forms) για εισαγωγή νέων δεδομένων στη βάση (εμπλουτισμός των δεδομένων). Τα Δελτία Εισαγωγής δεδομένων αρχικοποιούν και χρησιμοποιούν ένα δικό τους μηχανισμό διαχείρισης κρυφής μνήμης, ο οποίος είναι ανεξάρτητος από αυτόν του υπόλοιπου Συστήματος Σημασιολογικού Ευρετηριασμού, αλλά καθόλου διαφορετικός από αυτόν.

   Πριν την εισαγωγή των δεδομένων εκτελείται μια ερώτηση, η οποία ξαναεκτελείται μετά την ενημέρωση της βάσης. Στη συνέχεια, σβήνονται από τη βάση τα δεδομένα που είχαν εισαχθεί μέσω των παραπάνω βημάτων.

   Στην περίπτωση της εισαγωγής δεδομένων, τα νέα δεδομένα γράφονται από την κρυφή μνήμη των Δελτίων Εισαγωγής στη βάση ενώ οι σελίδες που τα περιέχουν και

βρίσκονται ενδεχόμενα στη μνήμη του Συστήματος Σημασιολογικού Ευρετηριασμού σημειώνονται ως άκυρες. Με τη νέα εκτέλεση της ίδιας ερώτησης, οι σελίδες που έχουν σημειωθεί ως άκυρες πρέπει να ξαναδιαβαστούν από το δίσκο.

Στόχος εδώ είναι να εξεταστεί η συμπεριφορά του συστήματος κατά την εισαγωγή δεδομένων καθώς επίσης και όταν στο σύστημα υπάρξει η ακύρωση των περιεχομένων της κρυφής μνήμης.

4. Χρησιμοποίηση των Δελτίων Εισαγωγής για τροποποίηση του σχήματος της βάσης και αναίρεση της τροποποίησης στη συνέχεια.

   Στόχος της εκτέλεσης αυτού του βήματος είναι η εξέταση της συμπεριφοράς του συστήματος κατά την τροποποίηση του σχήματος, η οποία επιφέρει πολυάριθμους σημασιολογικούς ελέγχους.

5. Χρησιμοποίηση των Δελτίων Εισαγωγής για εισαγωγή νέων δεδομένων στη βάση και σβήσιμό τους στη συνέχεια. Στο βήμα αυτό δεν πραγματοποιείται και ακύρωση των περιεχομένων της κρυφής μνήμης του Συστήματος Σημασιολογικού Ευρετηριασμού.

Τα ακριβή βήματα που περιλαμβάνονται στα σενάρια τυπικής χρήσης του συστήματος για καθεμιά από τις τρεις βάσεις δεδομένων που χρησιμοποιήθηκαν στα πειράματα, καθώς επίσης και ο αριθμός των δοσοληψιών στις οποίες το κάθε βήμα αντιστοιχεί παρουσιάζονται στο παράρτημα Α.

Τα αποτελέσματα των μετρήσεων καταγράφονται ανά δοσοληψία σε όλη την ακολουθία των δοσοληψιών που διεξάγονται κατά την εκτέλεση του σεναρίου χρήσης.

Στόχος των παραπάνω πειραμάτων είναι να εξεταστεί η συμπεριφορά του συστήματος σε όλες τις δυνατές καταστάσεις χρήσης του, ώστε να εντοπιστούν ενδεχόμενα προβλήματα ή στοιχεία του σχεδιασμού του που επιδέχονται βελτίωση. Στόχο αποτελεί, επίσης, ο εντοπισμός σημείων στη σχεδίαση με τα οποία πραγματοποιείται κατανάλωση πόρων, η οποία δεν προσφέρει τίποτα στη βελτίωση της επίδοσης του συστήματος, ενώ σε μερικές περιπτώσεις ενδέχεται και να το επιβαρύνει.

Κεφάλαιο 6

Τυπική συμπεριφορά συστήματος : βάσεις ΚΛΕΙΩ και ΑΑΤ

6.1 Εισαγωγή

Στο κεφάλαιο αυτό περιγράφονται τα αποτελέσματα που συγκεντρώθηκαν από πειράματα σε δύο από τις τρεις βάσεις δεδομένων που χρησιμοποιήθηκαν. Οι δύο αυτές βάσεις είναι η βάση του συστήματος πολιτιστικής τεκμηρίωσης ΚΛΕΙΩ και ο θησαυρός όρων τέχνης και αρχιτεκτονικής ΑΑΤ.

Τα αποτελέσματα από την εκτέλεση των πειραμάτων σε αυτές τις βάσεις περιγράφονται πρώτα γιατί οι βάσεις του ΚΛΕΙΩ και του ΑΑΤ θεωρούνται τυπικότερα παραδείγματα εφαρμογής και χρήσης του Συστήματος Σημασιολογικού Ευρετηριασμού.

Η βάση του συστήματος ΚΛΕΙΩ είναι μια βάση η οποία σχεδιάστηκε εξ αρχής σε μορφή σημασιολογικού δικτύου, χρησιμοποιώντας επομένως αρχές οντοκεντρικής σχεδίασης κι αυτός είναι ο λόγος που την καθιστά τυπικό παράδειγμα.

Η βάση ΑΑΤ, από την άλλη μεριά, αποτελεί μεταφορά σε Telos ενός σχεσιακού θησαυρού όρων τέχνης και αρχιτεκτονικής και το μέγεθός της είναι σαφώς μεγαλύτερο σε σχέση με αυτό της βάσης του συστήματος ΚΛΕΙΩ. Αποτελεί εξάλλου μια μικρογραφία της βάσης ULAN - με μεγαλύτερη ωστόσο αναλογία σχήματος-δεδομένων - η οποία είναι η τρίτη βάση που χρησιμοποιήθηκε στη διεξαγωγή των πειραμάτων, κι αποτελεί έτσι ένα ενδιάμεσο σημείο μεταξύ των δύο βάσεων. Για το λόγο αυτό επελέγη ως τυπικό παράδειγμα συμπεριφοράς για το Σύστημα Σημασιολογικού Ευρετηριασμού.

6.2 ΚΛΕΙΩ

Η βάση δεδομένων του συστήματος πολιτισμικής τεκμηρίωσης ΚΛΕΙΩ είναι, όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, η μόνη από τις τρεις βάσεις δεδομένων που χρησιμοποιήθηκαν στα πειράματα η οποία σχεδιάστηκε εξ αρχής στη γλώσσα παράστασης γνώσης Telos και υλοποιήθηκε στο Σύστημα Σημασιολογικού Ευρετηριασμού. Το γεγονός αυτό έχει σαν αποτέλεσμα μεγάλο αριθμό στοιχείων στο τμήμα της βάσης δεδομένων που ονομάζεται σχήμα.
Το μέγεθος της βάσης είναι σχετικά μικρό, συγκρινόμενο με τα μεγέθη των άλλων δύο βάσεων που χρησιμοποιήθηκαν, αυτής του AAT καθώς επίσης και αυτής του ULAN.

6.2.1 Αποτελέσματα πειραμάτων

Οι μετρήσεις που πραγματοποιήθηκαν για τη βάση του συστήματος πολιτισμικής τεκμηρίωσης ΚΛΕΙΩ ακολουθούν τις γενικές αρχές που διέπουν το σχεδιασμό όλων των πειραμάτων, οι οποίες παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Το σενάριο τυπικής χρήσης του συστήματος που χρησιμοποιήθηκε για τη διεξαγωγή των μετρήσεων δημιουργήθηκε με βάση τη χρήση του συστήματος στο Μουσείο Μπενάκη, στο οποίο είναι εγκατεστημένο. Τα μεγέθη που καταγράφηκαν κατά τις μετρήσεις και των οποίων οι τιμές παρουσιάζονται παρακάτω είναι αυτά που επιλέχθηκαν για όλα τα πειράματα.
Τα αποτελέσματα από τις μετρήσεις είναι τα ακόλουθα :

- Τα ποσοστά επιτυχίας κατά την πρόσβαση σε καθεμιά από τις κρυφές μνήμες.
- Τα μεγέθη στα οποία φτάνει καθεμιά από αυτές.
- Τα μεγέθη της μνήμης που καταναλώνονται εξαιτίας της αποθήκευσης σε αυτήν των οντοτήτων που χρησιμοποιούνται για την εκτέλεση ερωτήσεων και των επεκτάσεών τους.
- Τα μεγέθη των συνόλων που χρησιμοποιούνται για τον υπολογισμό και την αποθήκευση ενδιάμεσων αποτελεσμάτων κατά την εκτέλεση ερωτήσεων προς τη βάση.

Μηχανισμοί κρυφής μνήμης - Αποτελέσματα

Τα αποτελέσματα αυτά, όπως προαναφέρθηκε, διακρίνονται σε μεγέθη των μνημών κατά την εκτέλεση των σεναρίων χρήσης του συστήματος και ποσοστά επιτυχίας κατά την πρόσβαση στις κρυφές αυτές μνήμες. Τα αποτελέσματα διαχωρίζονται ανάλογα με το εάν προέρχονται από τη χρήση του Συστήματος Διατύπωσης Ερωτήσεων ή των Δελτίων Εισαγωγής Δεδομένων στο Σύστημα Σημασιολογικού Ευρετηριασμού.

Πριν την παρουσίαση των αποτελεσμάτων των μετρήσεων, είναι απαραίτητο να παρουσιαστούν οι χωρητικότητες των επιμέρους κρυφών μνημών, όπως αυτές διαμορφώθηκαν για τη διεξαγωγή των πειραμάτων. Για τις κρυφές μνήμες με δομή μονοδιάστατου πίνακα, δεν υπάρχει διαφοροποίηση των χωρητικοτήτων τους μεταξύ τους στο ίδιο πείραμα, καθώς τόσο οι παράμετροι μεγέθους, όσο και το μέγεθος της μονάδας μεταφοράς από και προς το δίσκο είναι η ίδια για όλους τους μηχανισμούς.

Χωρητικότητες κρυφών μνημών με δομή μονοδιάστατου πίνακα

Πείραμα	Χωρητικότητα
Small Queue	102400 bytes
Medium Queue	204800 bytes
Big Queue	2048000 bytes

Δεν ισχύει το ίδιο, όμως, για τις κρυφές μνήμες με δομή καταλόγου κατακερματισμού. Σ'αυτές υπάρχει διαφοροποίηση, όπως περιγράφηκε στο τρίτο κεφάλαιο, τόσο σε ό,τι αφορά τη μονάδα μεταφοράς από και προς το δίσκο, όσο και ως προς τις παραμέτρους μεγέθους των κρυφών μνημών. Αναλυτικότερα, οι χωρητικότητες στα διαφορετικά πειράματα έχουν ως εξής :

Χωρητικότητες κρυφών μνημών με δομή πίνακα κατακερματισμού

	Κρυφή Μνήμη				
Πείραμα	Hash	Symbol Table	Overflow	Named Sysid	Unnamed Sysid
Small Cache	1.06 Mb	450 Kb	263 Kb	463 Kb	463 Kb
Medium Cache	17 Mb	7.1 Mb	4.2 Mb	7.4 Mb	7.4 Mb
Big Cache	540 Mb	227 Mb	134 Mb	237 Mb	237 Mb

Οι ίδιες χωρητικότητες για τους επιμέρους μηχανισμούς κρυφής μνήμης ισχύουν για τα αντίστοιχα πειράματα και στις άλλες δύο βάσεις δεδομένων που χρησιμοποιήθηκαν. Για καθέναν από τους διαφορετικούς μηχανισμούς κρυφής μνήμης, τα αποτελέσματα των μετρήσεων παρουσιάζονται με τη μορφή πινάκων, για τους οποίους πρέπει να σημειωθούν τα εξής :

- Υπάρχει ένας πίνακας συνοπτικής παρουσίασης αποτελεσμάτων για καθένα από τα δύο συστήματα που χρησιμοποιήθηκαν στις μετρήσεις και για καθέναν από τους επιμέτους μηχανισμούς κρυφής μνήμης.

- Η πληροφορία που περιλαμβάνεται σε καθέναν από τους πίνακες είναι η ακόλουθη :

  -- Το μέσο ποσοστό επιτυχημένων προσπελάσεων στην κρυφή μνήμη.

  -- Η μέγιστη τιμή που φτάνει το μέγεθος της κρυφής μνήμης αυτής.

  -- Τα όρια τιμών στα οποία κυμαίνεται το ποσοστό επιτυχών προσπελάσεων.

  -- Ο τρόπος μεταβολής του μεγέθους της κρυφής μνήμης στη διάρκεια των δοσοληψιών καθενός από τα δύο συστήματα.

- Η παραπάνω πληροφορία περιγράφεται για καθένα από τα διαφορετικά πειράματα που διεξάγονται.

  Οπως αναφέρθηκε και στο πέμπτο κεφάλαιο, τα πειράματα διεξήχθησαν για διάφορους συνδυασμούς τιμών των παραμέτρων μεγέθους των μηχανισμών κρυφής μνήμης. Αποδείχθηκε πειραματικά, ωστόσο, ότι η συμπεριφορά κάθε μηχανισμού εξαρτάται αποκλειστικά και μόνο από τη χωρητικότητά του και όχι από τις χωρητικότητες άλλων μηχανισμών. Αυτό συμβαίνει διότι οι σελίδες που αποθηκεύονται σε καθέναν από τους επιμέρους μηχανισμούς είναι διαφορετικές,

ενώ το πλήθος κάθε κατηγορίας σελίδων που απαιτούνται για την εκτέλεση μιας ερώτησης εξαρτάται από την ερώτηση και μόνο. Για το λόγο αυτό, τα αποτελέσματα δεν παρουσιάζονται για το σύνολο των πειραμάτων, αλλά για καθεμιά από τις κατηγορίες χωρητικοτήτων για τις οποίες μελετήθηκε κάθε μηχανισμός.

- Ο τρόπος μεταβολής του μεγέθους της κρυφής μνήμης στη διάρκεια των δοσοληψιών κάθε συστήματος περιγράφεται με τη βοήθεια μαθηματικών συμβόλων με τις ακόλουθες ερμηνείες :

  $\nearrow$: Η τιμή του μεγέθους αυξάνεται συνεχώς.

  $\nearrow, \rightarrow$: Η τιμή του μεγέθους σταθεροποιείται μετά τις πρώτες μισές δοσοληψίες, κατά τις οποίες συνεχώς αυξάνεται.

  $\nearrow, \rightarrow, \rightarrow$: Η τιμή του μεγέθους πολύ γρήγορα αυξάνεται στο μέγιστό της, στο οποίο παραμένει σταθερά.

  $\rightarrow$: Η τιμή του μεγέθους παραμένει σταθερά στο μέγιστό της σε όλες τις δοσοληψίες.

Class Nodes Queue

Ο μηχανισμός αυτός διαχειρίζεται και αποθηκεύει στη μνήμη τις δομές του τύπου οντοτήτων που αναπαριστούν τις κλάσεις αντικειμένων (κόμβων) στο σχήμα της βάσης. Για τα διαφορετικά πειράματα που πραγματοποιήθηκαν, τα αποτελέσματα των μετρήσεων παρουσιάζονται στη συνέχεια σε πινακοποιημένη μορφή.

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	102 Kb	0% - 100%	$\nearrow, \rightarrow$
Medium Queue	90%	140 Kb	0% - 100%	$\nearrow$
Big Queue	90%	140 Kb	0% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	102 Kb	80% - 100%	↗, →
Medium Queue	100%	140 Kb	100% - 100%	↗
Big Queue	100%	140 Kb	100% - 100%	↗

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της και ανάγνωσής τους εκ νέου από το δίσκο (cache invalidation) μειώνεται στην τιμή των 40 Kb. Η τιμή του 0% στα ποσοστά επιτυχών προσπελάσεων της κρυφής μνήμης, εξάλλου, για το σύστημα διατύπωσης ερωτήσεων εμφανίζεται σε ολιγάριθμες περιπτώσεις και προκύπτει από την αναζήτηση μίας και μόνο σελίδας στη διάρκεια μιας δοσοληψίας, η οποία δε βρίσκεται στη μνήμη. Στην περίπτωση αυτή, το ποσοστό επιτυχίας προσδιορίζεται από τη σχέση :

0 επιτυχείς αναφορές / 1 συνολικές αναφορές,

δηλαδή 0%.

Οπως φαίνεται από τους παραπάνω πίνακες, η κρυφή μνήμη αποκτά πληρότητα 100% μόνο στην περίπτωση του πειράματος με τη μικρότερη χωρητικότητά της (102 Kb). Στις άλλες δύο περιπτώσεις, το μέγεθός της δεν ξεπερνά τα 140 Kb. Το στοιχείο αυτό, σε συνδυασμό με το γεγονός ότι τα σενάρια χρήσης των δύο συστημάτων είναι τέτοια που να καλύπτουν πολύ διαφορετικές περιοχές της βάσης δεδομένων, οδηγεί με μεγάλη πιθανότητα στο συμπέρασμα ότι όλες οι κλάσεις οντοτήτων της βάσης δεν ξεπερνούν σε συνολικό μέγεθος την παραπάνω τιμή.

Το γεγονός της χαμηλής πληρότητας της κρυφής μνήμης στα πειράματα μεσαίας και μεγάλης χωρητικότητάς της, οδηγούν στο συμπέρασμα ότι οι τιμές της τάξης του 0% στα ποσοστά επιτυχών προσπελάσεων της κρυφής μνήμης για το σύστημα διατύπωσης ερωτήσεων οφείλονται αποκλειστικά και μόνο σε κακή ομαδοποίηση των δεδομένων σε σχέση με τις ερωτήσεις που εκτελούνται. Τα υψηλά ποσοστά επιτυχίας στο σύστημα δελτίων εισαγωγής δεδομένων, εξάλλου, δεδομένου ότι η πληρότητα της κρυφής μνήμης παραμένει στα ίδια επίπεδα, φανερώνει καλύτερη ομαδοποίηση των δεδομένων ή,

ισοδύναμα, μεγαλύτερη τοπικότητα στις αναφορές.

Link Classes Queue

Αυτός ο μηχανισμός κρυφής μνήμης είναι υπεύθυνος για την αποθήκευση και τη διαχείριση των δομών αποθήκευσης και αναπαράστασης που αντιστοιχούν στις κλάσεις συνδέσμων του σχήματος της βάσης.

Τα αποτελέσματα των μετρήσεων για το μηχανισμό αυτό έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	95%	40 Kb	90% - 100%	$\nearrow, \rightarrow$
Medium Queue	95%	40 Kb	90% - 100%	$\nearrow, \rightarrow$
Big Queue	95%	40 Kb	90% - 100%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	95%	45 Kb	80% - 100%	$\nearrow, \rightarrow$
Medium Queue	95%	45 Kb	80% - 100%	$\nearrow, \rightarrow$
Big Queue	95%	45 Kb	80% - 100%	$\nearrow, \rightarrow$

Στη δοσοληψία ακύρωσης των περιεχομένων αυτής της κρυφής μνήμης, το μέγεθός της μειώνεται σε 25 Kb. Πρέπει να σημειωθεί, εξάλλου, σε ό,τι αφορά τα ποσοστά επιτυχών προσπελάσεων στην κρυφή μνήμη, πως οι τιμές της τάξης του 80% και 90% που παρουσιάζονται για τα δύο συστήματα, αντιστοιχούν στις δοσοληψίες εκκίνησης των συστημάτων, κατά τις οποίες η κρυφή μνήμη είναι άδεια με αποτέλεσμα οι πρώτες αναφορές να καταλήγουν αναγκαστικά σε αποτυχία. Μετά τις πρώτες δοσοληψίες, τα

ποσοστά επιτυχίας παραμένουν σταθερά στο 100%, στοιχείο που αποτελεί απόδειξη τοπικότητας στις αναφορές ή ομαδοποίησης των δεδομένων στη βάση με τρόπο που να εξυπηρετεί κατά τη διεξαγωγή των ερωτήσεων.

Χαρακτηριστικό αυτού του μηχανισμού είναι ότι παρουσιάζει την ίδια ακριβώς συμπεριφορά σε όλα τα πειράματα και για τα δύο συστήματα που μελετήθηκαν. Καθώς το μέγεθος δεν ξεπερνά σε καμμία περίπτωση τα 45 Kb, οδηγούμαστε και πάλι στο συμπέρασμα ότι η τιμή αυτή αποτελεί το συνολικό μέγεθος όλων των δομών αποθήκευσης και αναπαράστασης για τις κλάσεις συνδέσμων που περιέχονται στο σχήμα της βάσης.

Link Tokens Queue

Αυτή η κρυφή μνήμη χρησιμοποιείται για την αποθήκευση των δομών αναπαράστασης των περιπτώσεων των κλάσεων συνδέσμων, δηλαδή των συσχετίσεων μεταξύ ατομικών οντοτήτων της βάσης.

Τα αποτελέσματα των μετρήσεων έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακές τιμές hit rate	Συμπεριφορά μεγέθους
Small Queue	80%	102 Kb	65% - 100%	↗, →
Medium Queue	80%	160 Kb	65% - 100%	↗
Big Queue	80%	160 Kb	65% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακή τιμή hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	102 Kb	80% - 100%	↗, →
Medium Queue	90%	160 Kb	80% - 100%	↗, →
Big Queue	90%	160 Kb	80% - 100%	↗, →

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος αυτού του μηχανισμού περιορίζεται σε 20 Kb.

Σε ό,τι αφορά τη χαμηλότερη τιμή που παρουσιάζεται στα ποσοστά επιτυχών προσπελάσεων για το σύστημα διατύπωσης ερωτήσεων, αυτή οφείλεται αποκλειστικά και μόνο στην ομαδοποίηση των δεδομένων σε σχέση με τις ερωτήσεις που τίθενται στη βάση. Απόδειξη του παραπάνω αποτελεί το γεγονός ότι το μέγεθος της κρυφής μνήμης περιορίζεται από τη χωρητικότητά της μόνο στην περίπτωση που αυτή έχει το μικρότερο μέγεθός της. Διαφορετικά, η τιμή του παραμένει σταθερά στα 160 Kb, πράγμα που σημαίνει ότι δεν υπάρχει ανάγκη για εισαγωγή στη μνήμη περισσότερων σελίδων δεδομένων και ότι οι αποτυχημένες προσπελάσεις οφείλονται μόνο στην έλλειψη τοπικότητας των αναφορών σε σχέση με την οργάνωση των δεδομένων στη βάση.

Στο σύστημα δελτίων εισαγωγής δεδομένων, οι χαμηλές τιμές των ποσοστών επιτυχίας εμφανίζονται μόνο στις δοσοληψίες εξάλειψης των μεταβατικών φαινομένων κατά την εκκίνηση του συστήματος.

Token Nodes Queue

Στο τμήμα αυτό του μηχανισμού διαχείρισης κρυφής μνήμης αποθηκεύονται οι ατομικές οντότητες της βάσης δεδομένων.

Τα αποτελέσματα των μετρήσεων συνοψίζονται στα ακόλουθα :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	50%	70 Kb	$\simeq$ 50%	$\nearrow$
Medium Queue	50%	70 Kb	$\simeq$ 50%	$\nearrow$
Big Queue	50%	70 Kb	$\simeq$ 50%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	97%	75 Kb	90% - 100%	$\nearrow$
Medium Queue	97%	75 Kb	90% - 100%	$\nearrow$
Big Queue	97%	75 Kb	90% - 100%	$\nearrow$

Στη δοσοληψία ακύρωσης κι επανενημέρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος αυτού του μηχανισμού περιορίζεται σε 10 Kb.

Παρουσιάζεται κι εδώ το φαινόμενο χαμηλών ποσοστών επιτυχών προσπελάσεων της κρυφής μνήμης στο σύστημα διατύπωσης ερωτήσεων, τα οποία οφείλονται σε κακή ομαδοποίηση των δεδομένων που περιέχονται στη βάση σε σχέση με τις ερωτήσεις που τίθενται και όχι σε περιορισμό χωρητικότητας.

Στο σύστημα δελτίων εισαγωγής δεδομένων, από την άλλη μεριά, τα ποσοστά της τάξης του 90% παρουσιάζονται και πάλι μόνο στις δοσοληψίες εκκίνησης του συστήματος.

Το γεγονός ότι το μέγεθος της κρυφής μνήμης δεν ξεπερνά κι εδώ τα 75 Kb, παρά το ότι οι αναφορές δεν παρουσιάζουν μεγάλη τοπικότητα, οδηγεί κι εδώ στο συμπέρασμα ότι η τιμή αυτή αποτελεί το μέγεθος του συνόλου των δομών αναπαράστασης για τις ατομικές οντότητες της βάσης.

Object Extensions Queue

Σ'αυτήν την κρυφή μνήμη αποθηκεύονται οι δομές επέκτασης των οντοτήτων της βάσης.

Τα αποτελέσματα των πειραμάτων έχουν, συνοπτικά, ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακή τιμή hit rate	Συμπεριφορά μεγέθους
Small Queue	65%	102 Kb	50% - 90%	$\nearrow, \rightarrow$
Medium Queue	70%	160 Kb	50% - 95%	$\nearrow, \rightarrow$
Big Queue	70%	160 Kb	50% - 95%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακή τιμή hit rate	Συμπεριφορά μεγέθους
Small Queue	100%	102 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Medium Queue	100%	160 Kb	$\simeq$ 100%	$\nearrow$
Big Queue	100%	160 Kb	$\simeq$ 100%	$\nearrow$

Στη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος εδώ περιορίζεται στα 18 Kb.

Παρατηρούμε ότι τα ποσοστά επιτυχίας στο σύστημα διατύπωσης ερωτήσεων κυμαίνονται σε χαμηλά επίπεδα ως προς τη χαμηλότερη τιμή τους. Στο στοιχείο αυτό πρέπει να προστεθεί και ο μικρός αριθμός αναφορών στην κρυφή μνήμη κατά την εκτέλεση των βημάτων του σεναρίου χρήσης, ο οποίος αποτελεί ένδειξη ότι είτε τα βήματα του σεναρίου δεν προκαλούν προσπελάσεις της κρυφής μνήμης, είτε οι δομές επέκτασης μπορούν να βρεθούν και αλλού.

Αναφέρθηκε στο τρίτο κεφάλαιο ότι, εξαιτίας προβλημάτων στη διαχείριση δεικτών όταν οι δομές αποθήκευσης μένουν στην κρυφή μνήμη, αυτές αντιγράφονται σε διαφορετικό σημείο της μνήμης προκειμένου να χρησιμοποιηθούν. Αυτό είναι το σημείο στο οποίο

πρώτα αναζητάται η πληροφορία που αφορά τις δομές επέκτασης των οντοτήτων. Αν δε βρεθεί εκεί, τότε αναζητάται στην αντίστοιχη κρυφή μνήμη. Ετσι εξηγείται το γεγονός της ύπαρξης ολιγάριθμων αναφορών στο μηχανισμό από τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων.

Σε ό,τι αφορά τα χαμηλά ποσοστά επιτυχίας, αυτά οφείλονται αποκλειστικά και μόνο στην τοπικότητα των αναφορών, αν ληφθεί υπόψη το στοιχείο του χαμηλού ποσοστού πληρότητας της κρυφής μνήμης.

Στο σύστημα δελτίων εισαγωγής δεδομένων, εξάλλου, οι αναφορές είναι πυκνότερες - ένδειξη μη εύρεσης της πληροφορίας στο σημείο της μνήμης όπου αντιγράφεται προκειμένου να χρησιμοποιηθεί - και τα ποσοστά επιτυχίας φανερώνουν πολύ καλή τοπικότητα των αναφορών σε σχέση με την οργάνωση των περιεχομένων της βάσης.

Hash Cache

Εδώ αποθηκεύονται εγγραφές του υποκαταλόγου λογικών ονομάτων από τον κατάλογο συστήματος της βάσης δεδομένων.

Τα αποτελέσματα των μετρήσεων έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακά τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$
Medium Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$
Big Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$
Medium Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$
Big Cache	97%	70 Kb	90% - 100%	$\nearrow, \rightarrow, \rightarrow$

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της περιορίζεται σε 8 Kb.

Το γεγονός ότι η συμπεριφορά της κρυφής μνήμης είναι σε όλα τα πειράματα και για τα δύο συστήματα πανομοιότυπη φανερώνει ότι το σύνολο των εγγραφών του υποκαταλόγου λογικών ονομάτων περιέχεται στην κρυφή μνήμη. Τα υψηλά ποσοστά επιτυχίας, εξάλλου, προδίδουν πολύ καλή τοπικότητα στις αναφορές.

Symbol Table Cache

Η επίδοση της κρυφής μνήμης στην οποία αποθηκεύονται οι εγγραφές του υποκαταλόγου αναγνωριστικών συστήματος συνοψίζεται στα εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	90%	450 Kb	75% - 100%	$\nearrow$
Medium Cache	90%	450 Kb	75% - 100%	$\nearrow$
Big Cache	90%	450 Kb	75% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	450 Kb	90% - 100%	$\nearrow, \rightarrow$
Medium Cache	97%	600 Kb	95% - 100%	$\nearrow, \rightarrow$
Big Cache	97%	600 Kb	95% - 100%	$\nearrow, \rightarrow$

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος περιορίζεται στα 80 Kb.

Τόσο η κατανάλωση μνήμης, όσο και τα ποσοστά επιτυχών προσπελάσεων είναι γι'αυτήν την κρυφή μνήμη πολύ ικανοποιητικά. Τα κάτω όρια των τιμών για τα ποσοστά επιτυχίας εμφανίζονται σε δοσοληψίες εκκίνησης των δύο συστημάτων. Οι τιμές του μεγέθους της κρυφής μνήμης και για τα δύο συστήματα, εξάλλου, με δεδομένη τη διαθεσιμότητα μεγάλων ποσοτήτων μνήμης και τα πολύ υψηλά ποσοστά επιτυχίας, αποτελεί ένδειξη του ότι η τιμή των 600 Kb αποτελεί το συνολικό μέγεθος του υποκαταλόγου αναγνωριστικών συστήματος της βάσης δεδομένων.

Overflow Cache

Στο τμήμα αυτό του μηχανισμού διαχείρισης κρυφής μνήμης αποθηκεύονται οι βοηθητικές εγγραφές για τον υποκατάλογο λογικών ονομάτων. Τα αποτελέσματα που προέκυψαν γι'αυτό το μηχανισμό έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	90%	33 Kb	85% - 100%	$\nearrow$
Medium Cache	90%	33 Kb	85% - 100%	$\nearrow$
Big Cache	90%	33 Kb	85% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριατιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	40%	54 Kb	$\simeq$ 40%	$\nearrow$
Medium Cache	40%	54 Kb	$\simeq$ 40%	$\nearrow$
Big Cache	40%	54 Kb	$\simeq$ 40%	$\nearrow$

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της περιορίζεται στα 4 Kb.

Σε όλα τα πειράματα που πραγματοποιήθηκαν και για τα δύο συστήματα η συμπεριφορά του μηχανισμού παρουσιάζει μεγάλη σταθερότητα. Τα χαμηλά ποσοστά επιτυχίας για το σύστημα δελτίων εισαγωγής δεδομένων, με δεδομένη την έλλειψη περιορισμού στην κατανάλωση μνήμης, αποτελούν απόδειξη της έλλειψης τοπικότητας στις αναφορές στις βοηθητικές εγγραφές σε σχέση με τις ερωτήσεις που τίθενται στη βάση. Η τιμή του 85% στο ποσοστό επιτυχίας για το σύστημα διατύπωσης ερωτήσεων, από την άλλη μεριά, αντιστοιχεί στις δοσοληψίες εκκίνησης του συστήματος.

Named Sysid Cache

Σ'αυτήν την κρυφή μνήμη αποθηκεύονται οι εγγραφές του υποκαταλόγο συστήματος που αντιστοιχούν σε οντότητες της βάσης με δοσμένο λογικό όνομα. Τα αποτελέσματα των μετρήσεων έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριατιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	99%	200 Kb	97% - 100%	$\nearrow$
Medium Cache	99%	200 Kb	97% - 100%	$\nearrow$
Big Cache	99%	200 Kb	97% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριατιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	99%	460 Kb	97% - 100%	↗
Medium Cache	99%	525 Kb	97% - 100%	↗
Big Cache	99%	525 Kb	97% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης το μέγεθός της περιορίζεται σε 70 Kb.

Η χαμηλή, σε σχέση με τη διαθεσιμότητα, κατανάλωση μνήμης και τα ποσοστά επιτυχών προσπελάσεων που προσεγγίζουν το 100% δίνουν την εικόνα ενός μηχανισμού που λειτουργεί πολύ ικανοποιητικά. Στην πραγματικότητα, η καλή οργάνωση της πληροφορίας που περιλαμβάνεται στην κρυφή μνήμη σε σχέση με τις ερωτήσεις που τίθενται στη βάση είναι που δίνουν το αποτέλεσμα αυτό.

Unnamed Sysid Cache

Εδώ αποθηκεύονται οι εγγραφές του καταλόγου συστήματος για τις οντότητες της βάσης χωρίς δοσμένο λογικό όνομα. Τα αποτελέσματα των μετρήσεων έχουν σ'αυτήν την περίπτωση ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριατιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	99%	80 Kb	98% - 100%	↗
Medium Cache	99%	80 Kb	98% - 100%	↗
Big Cache	99%	80 Kb	98% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	85%	58 Kb	40% - 100%	$\nearrow, \rightarrow$
Medium Cache	85%	58 Kb	40% - 100%	$\nearrow, \rightarrow$
Big Cache	85%	58 Kb	40% - 100%	$\nearrow$

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της περιορίζεται στα 22 Kb.

Η μικρή τιμή στην οποία ανέρχεται το μέγεθος της κρυφής μνήμης οφείλεται στο γεγονός της έλλειψης στη βάση του συστήματος ΚΛΕΙΩ του είδους των συνδέσμων που αποθηκεύονται στην κρυφή μνήμη αυτή. Τα ποσοστά επιτυχίας είναι πολύ υψηλά στο σύστημα διατύπωσης ερωτήσεων αλλά παρουσιάζουν μέτρια, ως προς την κατώτερη τιμή τους, συμπεριφορά στο σύστημα δελτίων εισαγωγής δεδομένων. Αυτή η συμπεριφορά οφείλεται στο μικρό αριθμό προσπελάσεων της κρυφής μνήμης που λαμβάνουν χώρα στη διάρκεια μιας δοσοληψίας, οι οποίες, εφόσον αποβούν σε αποτυχία, δίνουν συνολικά την εικόνα πολύ χαμηλών ποσοστών.

Μετρήσεις για σύνολα προσωρινής αποθήκευσης δεδομένων (βλ. 5.3)

Τα μεγέθη που καταγράφονται σχετικά με το μηχανισμό διαχείρισης συνόλων προσωρινής αποθήκευσης είναι :

- Η μέγιστη και η ελάχιστη κατανάλωση μνήμης στη διάρκεια κάθε δοσοληψίας από σύνολα.
- Ο μέγιστος και ο ελάχιστος αριθμός που χρησιμοποιούνται στη διάρκεια κάθε δοσοληψίας.

Τα πρώτα από τα παραπάνω μεγέθη είναι χρήσιμα στον προσδιορισμό της επιβάρυνσης που προκαλεί στο σύστημα ο μηχανισμός από την άποψη της κατανάλωσης μνήμης καθώς επίσης και στον εντοπισμό τυχόν προβλημάτων στην υλοποίησή του όταν η ελάχιστη κατανάλωση μνήμης στη διάρκεια κάποιων δοσοληψιών δε μηδενίζεται. Ενα

τέτοιο φαινόμενο αποτελεί πρόβλημα διότι τα σύνολα που δεν ελευθερώνονται στη διάρκεια κάποιων δοσοληψιών αποτελούν εσφαλμένη υλοποίηση.

Ο αριθμός των συνόλων, εξάλλου, δίνει τη δυνατότητα - στην περίπτωση που γίνεται πολύ μεγάλος - να εντοπιστούν και πάλι ενδεχόμενα προβλήματα υλοποίησης.

Τα αποτελέσματα των μετρήσεων αυτών για τη βάση του συστήματος ΚΛΕΙΩ έχουν ως εξής :

	Χαρακτηριστικά μεγέθη			
Σύστημα	Μέγιστη μνήμη	Ελάχιστη μνήμη	Μέγιστο πλήθος	Ελάχιστο πλήθος
Διατύπωσης ερωτήσεων	200 Kb	30 Kb	2000	800
Δελτίων εισαγωγής δεδομένων	250 Kb	70 Kb	6000	4000

Παρατηρείται στα παραπάνω αποτελέσματα το φαινόμενο της μη απελευθέρωσης της μνήμης που καταλαμβάνουν τα σύνολα προσωρινής αποθήκευσης στο τέλος της δοσοληψίας, όπως θα έπρεπε. Το πρόβλημα παρουσιάστηκε πιο έντονο στο σύστημα δελτίων εισαγωγής δεδομένων.

Παρατηρήθηκε, εξάλλου, μεγάλη αύξηση του πλήθους των συνόλων που απαιτούνται για την εκτέλεση ερωτήσεων οι οποίες ανακαλούν μεγάλο πλήθος οντοτήτων από τη βάση. Το φαινόμενο αυτό, οδήγησε στο συμπέρασμα ότι η υλοποίηση του μηχανισμού προκαλούσε την αποθήκευση μικρού πλήθους οντοτήτων σε κάθε σύνολο, γι'αυτό παρατηρήθηκε η μεγάλη αυτή αύξηση.

Σε ό,τι αφορά την κατανάλωση μνήμης από το μηχανισμό διαχείρισης συνόλων προσωρινής αποθήκευσης, αυτή δεν είναι πολύ μεγάλη. Σε κάθε περίπτωση, ωστόσο, αυτή εξαρτάται αποκλειστικά και μόνο από τις ερωτήσεις που τίθενται στη βάση. Γι'αυτό και ο μηχανισμός δε μελετήθηκε με σκοπό τη βελτίωση της επίδοσής του - με μόνη εξαίρεση τον εντοπισμό λαθών στην υλοποίηση - αλλά τον προσδιορισμό της επιβάρυνσης που αυτός επιφέρει στο σύστημα.

Μετρήσεις για οντότητες και επεκτάσεις οντοτήτων στη μνήμη

Τα μεγέθη που καταγράφονται εδώ αφορούν την κατανάλωση μνήμης εξαιτίας της αντιγραφής από τους μηχανισμούς κρυφής μνήμης με δομή μονοδιάστατων πινάκων οντοτήτων, συσχετίσεων και των δομών επέκτασής τους σε διαφορετικό σημείο της μνήμης, προκειμένου να χρησιμοποιηθούν από κάποια δοσοληψία.

Για τα δύο συστήματα που μελετήθηκαν με τη βοήθεια των μετρήσεων, τα αποτελέσματα έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

	Κατανάλωση μνήμης	
Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	45 Kb	112 Kb
MC/SQ	45 Kb	112 Kb
SC/MQ	45 Kb	112 Kb
MC/MQ	45 Kb	112 Kb
BC/MQ	45 Kb	112 Kb
MC/BQ	45 Kb	112 Kb
BC/BQ	45 Kb	112 Kb

Δελτία Εισαγωγής Δεδομένων

	Κατανάλωση μνήμης	
Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	20 Kb	45 Kb
MC/SQ	20 Kb	45 Kb
SC/MQ	20 Kb	45 Kb
MC/MQ	20 Kb	45 Kb
BC/MQ	20 Kb	45 Kb
MC/BQ	20 Kb	45 Kb
BC/BQ	20 Kb	45 Kb

Στους παραπάνω πίνακες, είναι φανερό ότι τα μεγέθη μνήμης που καταναλώνονται από τις οντότητες και τις επεκτάσεις οντοτήτων παραμένουν ανεπηρέαστα από τις διακυμάνσεις των χωρητικοτήτων των επιμέρους μηχανισμών κρυφής μνήμης. Αυτό

συμβαίνει, όχι γιατί ο μηχανισμός αντιγραφής οντοτήτων και των επεκτάσεών τους στη μνήμη είναι ανεξάρτητος από την κρυφή μνήμη - γεγονός το οποίο θα επιβεβαιώσουν ακόλουθες μετρήσεις - αλλά εξαιτίας του μικρού μεγέθους της βάσης, το οποίο οδηγεί σε μικρά μεγέθη για τους μηχανισμούς κρυφής μνήμης.
Με άλλα λόγια, η κατανάλωση μνήμης επηρεάζεται από τις χωρητικότητες κάποιων από τους επιμέρους μηχανισμούς μόνο όταν αυτοί έχουν μέγεθος που φτάνει τη χωρητικότητα της κρυφής μνήμης κι επομένως απαιτείται αντικατάσταση σελίδων.

Συνολική συμπεριφορά του συστήματος

Η μελέτη της συνολικής συμπεριφοράς του συστήματος συνίσταται στον υπολογισμό του συνολικού μεγέθους μνήμης που καταναλώνεται για την εκτέλεση του σεναρίου χρήσης του συστήματος που χρησιμοποιήθηκε στα πειράματα καθώς επίσης και ενός συνολικού ποσοστού επιτυχίας για τις αναφορές στο μηχανισμό διαχείρισης κρυφής μνήμης του Συστήματος Σημασιολογικού Ευρετηριασμού.

Σε ό,τι αφορά τη συνολική κατανάλωση μνήμης από το σύστημα - κατανάλωση στην οποία συμπεριλαμβάνονται τα μεγέθη όλων των πηγών κατανάλωσης μνήμης - αυτή κυμαίνεται ανάμεσα σε 1.5 και 1.7 Mb στο σύστημα διατύπωσης ερωτήσεων, ανάλογα με τις χωρητικότητες των κρυφών μνημών. Στο σύστημα δελτίων εισαγωγής δεδομένων, εξάλλου, το μέγεθος αυξάνεται λίγο, φτάνοντας τα 2.2 Mb όταν τα μεγέθη αφεθούν να αυξηθούν ελεύθερα ενώ περιορίζεται στα 1.9 Mb για τα μικρότερα μεγέθη των μηχανισμών κρυφής μνήμης.

Πιο αναλυτικά, η συνολική κατανάλωση μνήμης από τα δύο συστήματα κατά τη διεξαγωγή των πειραμάτων έχει ως εξής :

- Σύστημα διατύπωσης ερωτήσεων

Big Cache-Big Queue	:	1.6 Mb
Big Cache-Medium Queue	:	1.6 Mb
Medium Cache-Big Queue	:	1.6 Mb
Medium Cache-Medium Queue	:	1.6 Mb
Small Cache-Medium Queue	:	1.6 Mb
Medium Cache-Small Queue	:	1.5 Mb
Small Cache-Small Queue	:	1.5 Mb

- Δελτία εισαγωγής δεδομένων

Big Cache-Big Queue	:	2.3 Mb
Big Cache-Medium Queue	:	2.3 Mb
Medium Cache-Big Queue	:	2.3 Mb
Medium Cache-Medium Queue	:	2.3 Mb
Small Cache-Medium Queue	:	2.3 Mb
Medium Cache-Small Queue	:	1.9 Mb
Small Cache-Small Queue	:	1.9 Mb

Σε κάθε περίπτωση, ωστόσο, τόσο για το σύστημα διατύπωσης ερωτήσεων όσο και για το σύστημα δελτίων εισαγωγής δεδομένων το συνολικό ποσοστό επιτυχίας του συστήματος ανέρχεται κατά μέσο όρο σε 95%, δίνοντας μια επίδοση περισσότερο από ικανοποιητική.

6.2.2 Συμπεράσματα

Η συμπεριφορά του συστήματος για όλους τους επιμέρους μηχανισμούς διαχείρισης κρυφής μνήμης παρουσιάζεται συνοπτικά στη συνέχεια. Στους πίνακες που ακολουθούν παρατίθενται τα μεγέθη που καταναλώνουν και τα ποσοστά επιτυχίας που εμφανίζουν οι επιμέρους μηχανισμοί για τις διαφορετικές χωρητικότητες και τα δύο συστήματα για τα οποία μελετήθηκαν.

Η πρώτη τριάδα στηλών στους πίνακες αναφέρεται στα αποτελέσματα που παρατηρήθηκαν για το σύστημα διατύπωσης ερωτήσεων και η δεύτερη αφορά το σύστημα δελτίων εισαγωγής δεδομένων.

Μεγέθη μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	102 Kb	140 Kb	140 Kb	102 Kb	140 Kb	140 Kb
Link Classes Queue	40 Kb	40 Kb	40 Kb	45 Kb	45 Kb	45 Kb
Link Tokens Queue	102 Kb	160 Kb	160 Kb	102 Kb	160 Kb	160 Kb
Token Nodes Queue	70 Kb	70 Kb	70 Kb	75 Kb	75 Kb	75 Kb
Object Extensions Queue	102 Kb	160 Kb	160 Kb	102 Kb	160 Kb	160 Kb
Hash Cache	70 Kb	70 Kb	70 Kb	70 Kb	70 Kb	70 Kb
Symbol Table Cache	450 Kb	450 Kb	450 Kb	450 Kb	600 Kb	600 Kb
Overflow Cache	33 Kb	33 Kb	33 Kb	54 Kb	54 Kb	54 Kb
Named Sysid Cache	200 Kb	200 Kb	200 Kb	460 Kb	525 Kb	525 Kb
Unnamed Sysid Cache	80 Kb	80 Kb	80 Kb	58 Kb	58 Kb	58 Kb

Ποσοστά επιτυχίας μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	90%	90%	90%	90%	100%	100%
Link Classes Queue	95%	95%	95%	95%	95%	95%
Link Tokens Queue	80%	80%	80%	90%	90%	90%
Token Nodes Queue	50%	50%	50%	97%	97%	97%
Object Extensions Queue	65%	70%	70%	100%	100%	100%
Hash Cache	97%	97%	97%	97%	97%	97%
Symbol Table Cache	90%	90%	90%	95%	97%	97%
Overflow Cache	90%	90%	90%	40%	40%	40%
Named Sysid Cache	99%	99%	99%	99%	99%	99%
Unnamed Sysid Cache	99%	99%	99%	85%	85%	85%

Από τους παραπάνω πίνακες, παρατηρούνται τα ακόλουθα :

- Το μικρό μέγεθος της βάσης είναι η αιτία που τα μεγέθη των επιμέρους

μηχανισμών παραμένουν σε τόσο χαμηλά επίπεδα. Σε μεγαλύτερα επίπεδα σε σχέση με του υπόλοιπους μηχανισμούς κυμαίνονται τα μεγέθη του υποκαταλόγου αναγνωριστικών συστήματος και του καταλόγου συστήματος για οντότητες με δοσμένο λογικό όνομα. Ο υποκατάλογος αναγνωριστικών συστήματος, ο οποίος είναι το σημείο πραγματοποίησης της αντιστοίχισης αναγνωριστικών συστήματος σε λογικά ονόματα, χρησιμοποιείται από το σύστημα για την ανεύρεση των λογικών ονομάτων όλων των αντικειμένων που περιέχονται στο αποτέλεσμα μιας ερώτησης, πριν αυτά παρουσιαστούν στο χρήστη. Αυτή είναι και η αιτία της αύξησης του μεγέθους του.

Ο κατάλογος συστήματος, εξάλλου, είναι πάντα επιφορτισμένος με τη διαχείριση μεγάλης ποσότητας πληροφορίας. Ο λόγος για τον οποίο παρουσιάζει μεγαλύτερη πληρότητα το τμήμα του που διαχειρίζεται τις οντότητες με δοσμένο λογικό όνομα είναι γιατί η βάση ΚΛΕΙΩ περιλαμβάνει ελάχιστες οντότητες χωρίς δοσμένο λογικό όνομα.

- Τα ποσοστά επιτυχίας είναι πολύ υψηλά, όπως είναι άλλωστε αναμενόμενο. Οι εξαιρέσεις που παρουσιάζονται οφείλονται σε έλλειψη τοπικότητας στις αναφορές.

- Αξιοσημείωτη είναι η ισορροπία ως προς την κατανάλωση μνήμης από τους διαφορετικούς μηχανισμούς. Αυτό οφείλεται στην αριθμητική ισοδυναμία σχήματος και δεδομένων που περιέχονται στη βάση.

- Τα υψηλά ποσοστά επιτυχίας, σε συνδυασμό με τις μικρές ανάγκες του συστήματος σε μνήμη, οδηγούν στο συμπέρασμα ότι πολλοί από τους μηχανισμούς θα μπορούσαν να περιοριστούν ως προς τις χωρητικότητές τους, χωρίς σοβαρές επιπτώσεις στην επίδοση του συστήματος. Αυτό ισχύει κυρίως για τους μηχανισμούς που αποθηκεύουν κλάσεις συσχετίσεων, ατομικές οντότητες, εγγραφές του υποκαταλόγου λογικών ονομάτων, εγγραφές του υποκαταλόγου βιηθητικών εγγραφών καθώς και εγγραφές του καταλόγου συστήματος.

6.3 AAT

Η βάση αυτή αποτελεί ένα θησαυρό όρων Τέχνης και Αρχιτεκτονικής και είναι πολύ μεγαλύτερη σε μέγεθος από αυτήν του συστήματος πολιτιστικής τεκμηρίωσης ΚΛΕΙΩ

που μελετήθηκε προηγουμένως. Η μεγάλη αριθμητική διαφορά δεν παρατηρείται ανάμεσα στα σχήματα των βάσεων, καθώς όπως έχει ήδη αναφερθεί η βάση του συστήματος πολιτιστικής τεκμηρίωσης ΚΛΕΙΩ διαθέτει πολύ πλούσιο σχήμα για την περιγραφή των κλάσεων και των συσχετίσεων που περιγράφουν τη δομή και την οργάνωση της πληροφορίας. Η μεγάλη αριθμητική διαφορά των δύο βάσεων παρουσιάζεται στο πλήθος των δεδομένων των βάσεων καθώς επίσης και στο πλήθος των συσχετίσεων που συνδέουν κάθε οντότητα της βάσης με άλλες.
Η βάση του AAT όπως αυτή χρησιμοποιείται για τις μετρήσεις επίδοσης του μηχανισμού διαχείρισης κρυφής μνήμης από το σύστημα σημασιολογικού ευρετηριασμού αποτελεί ”μετάφραση” σε Telos μιας σχεσιακής βάσης δεδομένων, γεγονός που αποτελεί και την ιδιαιτερότητά της.

6.3.1 Αποτελέσματα πειραμάτων

Τα πειράματα στη βάση του AAT χαρακτηρίζονται από τις ίδιες παραμέτρους με τα πειράματα που εκτελέστηκαν για τη βάση του ΚΛΕΙΩ, ενώ ίδια είναι και τα μεγέθη που μετρούνται.
Τα σενάρια που χρησιμοποιήθηκαν για την πραγματοποίηση των μετρήσεων είναι προσαρμοσμένα στη συνηθισμένη χρήση του συστήματος στο μουσείο μοντέρνας τέχνης Getty, για το οποίο και κατασκευάστηκε το σύστημα.

Μηχανισμοί κρυφής μνήμης - Αποτελέσματα

Οπως έχει ήδη αναφερθεί, τα αποτελέσματα των μετρήσεων σε ό,τι αφορά τους μηχανισμούς διαχείρισης κρυφής μνήμης διακρίνονται σε μεγέθη και ποσοστά επιτυχημένων προσπελάσεων στη μνήμη.
Αναλυτικότερα, τα αποτελέσματα για τη βάση του AAT έχουν ως εξής :

Class Nodes Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	42 Kb	85% - 100%	↗, →
Medium Queue	90%	42 Kb	85% - 100%	↗, →
Big Queue	90%	42 Kb	85% - 100%	↗, →

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	99%	37 Kb	90% - 100%	↗, →
Medium Queue	99%	37 Kb	90% - 100%	↗, →
Big Queue	99%	37 Kb	90% - 100%	↗, →

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της δεν μειώνεται πολύ σε σχέση με την τιμή των 42 Kb που έχει στη διάρκεια των υπόλοιπων δοσοληψιών του συστήματος διατύπωσης ερωτήσεων. Αυτό σημαίνει ότι τα περιεχόμενα αυτού του μηχανισμού δεν συμπεριλαμβάνονταν ανάμεσα σ'αυτά που ενημερώθηκαν με τη βοήθεια των δελτίων εισαγωγής δεδομένων.

Με δεδομένο το γεγονός ότι η βάση του AAT διαθέτει μικρό σχήμα, μέρος του οποίου αποθηκεύεται σ'αυτόν το μηχανισμό, και λαμβάνοντας υπόψη τη σταθερότητα του μεγέθους της κρυφής μνήμης και τα πολύ ικανοποιητικά ποσοστά επιτυχίας, μπορούμε και πάλι να υποθέσουμε ότι τα 42 Kb - τιμή του μεγέθους στο σύστημα διατύπωσης ερωτήσεων - αποτελεί το συνολικό μέγεθος των δομών αναπαράστασης για τις κλάσεις οντοτήτων της βάσης.

Link Classes Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	97%	20 Kb	94% - 100%	$\nearrow, \rightarrow$
Medium Queue	97%	20 Kb	94% - 100%	$\nearrow, \rightarrow$
Big Queue	97%	20 Kb	94% - 100%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	100%	13 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Medium Queue	100%	13 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Big Queue	100%	13 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της περιορίζεται σε 16 Kb.

Και σ'αυτήν την περίπτωση ισχύουν οι συνθήκες εκείνες - μικρό μέγεθος σχήματος στη βάση, σταθερή κατανάλωση μνήμης, υψηλά ποσοστά επιτυχίας - που οδηγούν στο συμπέρασμα ότι η τιμή των 20 Kb αποτελεί το συνολικό μέγεθος των δομών αναπαράστασης και αποθήκευσης για τις κλάσεις συνδέσμων στη βάση.

Link Tokens Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	67%	102 Kb	65% - 70%	$\nearrow, \rightarrow$
Medium Queue	70%	205 Kb	$\simeq$ 70%	$\nearrow, \rightarrow$
Big Queue	75%	400 Kb	70% - 80%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	100%	50 Kb	$\simeq$ 100%	$\nearrow$
Medium Queue	100%	50 Kb	$\simeq$ 100%	$\nearrow$
Big Queue	100%	50 Kb	$\simeq$ 100%	$\nearrow$

Το μέγεθος της κρυφής μνήμης περιορίζεται σε 20 Kb κατά τη δοσοληψία ακύρωσης των περιεχομένων της.

Στην περίπτωση αυτού του μηχανισμού παρατηρείται, για πρώτη φορά για τη βάση του AAT, μεγάλο ποσοστό πληρότητας της κρυφής μνήμης. Αυτό οφείλεται στο ίδιο το περιεχόμενο της βάσης, η οποία, όπως αναφέρθηκε και στο πέμπτο κεφάλαιο, περιλαμβάνει πολλούς συνδέσμους μεταξύ ατομικών οντοτήτων της βάσης. Χαρακτηριστικό της συμπεριφοράς του μηχανισμού αυτού αποτελεί το γεγονός ότι οι αναφορές που πραγματοποιούνται στην κρυφή μνήμη είναι ολιγάριθμες. Αιτία του φαινόμενου είναι και πάλι ο μηχανισμός ανάκλησης οντοτήτων από τη μνήμη, σύμφωνα με τον οποίο μια οντότητα αναζητάται πρώτα στο σημείο της μνήμης όπου αυτές αντιγράφονται προκειμένου να χρησιμοποιηθούν και στη συνέχεια στην κρυφή μνήμη, εφόσον δε βρεθεί στην πρώτη προσπάθεια. Ο μικρός αριθμός προσπελάσεων στην κρυφή μνήμη σημαίνει, επομένως, ότι η πλειοψηφία των αναζητήσεων στη μνήμη πετυχαίνει.

Για πρώτη φορά στο μηχανισμό αυτό, εξάλλου, μπορούμε να αποδώσουμε την απόκλιση των ποσοστών επιτυχίας από το 100%, όχι στην έλλειψη τοπικότητας των αναφορών, αλλά στον περιορισμό χωρητικότητας της κρυφής μνήμης.

Tokens Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	70%	95 Kb	0% - 100%	↗
Medium Queue	70%	95 Kb	0% - 100%	↗
Big Queue	70%	95 Kb	0% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	95%	35 Kb	80% - 100%	↗
Medium Queue	95%	35 Kb	80% - 100%	↗
Big Queue	95%	35 Kb	80% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος περιορίζεται στα 8 Kb.

Παρά το γεγονός ότι η βάση του ΑΑΤ περιέχει αρκετά μεγάλο αριθμό ατομικών οντοτήτων - ανέρχονται σε 840000 - τα μεγέθη της μνήμης παραμένουν εδώ σταθερά σε τιμές χαμηλές, γεγονός που προφανώς οφείλεται στο σενάριο τυπικής χρήσης που εκτελείται.

Σε ό,τι αφορά τα ποσοστά επιτυχίας, αυτά κυμαίνονται σε πολύ ικανοποιητικά επίπεδα για το σύστημα δελτίων εισαγωγής δεδομένων, στοιχείο το οποίο αν συνδυαστεί με τη

μικρή κατανάλωση μνήμης οδηγεί στο συμπέρασμα μικρού αριθμού αναφορών. Σχετικά με το σύστημα διατύπωσης ερωτήσεων, το εύρος τιμών των ποσοστών επιτυχίας σε συνδυασμό με το τμήμα της κρυφής μνήμης που παραμένει ανεκμετάλλευτο, οδηγεί στο συμπέρασμα ότι τα ποσοστά της τάξης του 0% προέρχονται από αποτυχία σε ολιγάριθμες - μία ή δύο - αναφορές στην κρυφή μνήμη.

Object Extensions Queue

Σύστημα Διατύπωσης Ερωτήσεων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	70%	102 Kb	0% - 100%	↗, →
Medium Queue	75%	205 Kb	0% - 100%	↗, →
Big Queue	90%	1.4 Mb	50% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	57%	102 Kb	0% - 100%	↗, →
Medium Queue	60%	205 Kb	0% - 80%	↗, →
Big Queue	90%	2 Mb	70% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της περιορίζεται σε 10 Kb.

Η συμπεριφορά αυτού του μηχανισμού κατά την εκτέλεση των πειραμάτων επιβεβαιώνει την πρόβλεψη μεγάλων απαιτήσεων σε μνήμη, εξαιτίας των πολυάριθμων δομών επέκτασης που είναι αναμενόμενο να διαθέτει κάθε οντότητα της βάσης εξαιτίας του μεγάλου αριθμού συνδέσμων που η βάση διαθέτει - ξεπερνάν τα 2 εκατομμύρια. Τόσο η

μεγάλη κατανάλωση μνήμης - 100% πληρότητα στις περιπτώσεις μικρής και μεσαίας χωρητικότητας και μέχρι 2 Mb στην περίπτωση της μεγάλης χωρητικότητας - όσο και τα χαμηλά ποσοστά επιτυχίας οφείλονται ακριβώς στο γεγονός της ανάγκης πολλών δομών επέκτασης που απαιτούνται για να καλύψουν το πλήθος των συνδέσμων.

Hash Cache

Σύστημα Διατύπωσης Ερωτήσεων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	90%	1.06 Mb	80% - 100%	↗, →
Medium Cache	90%	1.4 Mb	80% - 100%	↗
Big Cache	90%	1.4 Mb	80% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	1.06 Mb	85% - 100%	↗, →
Medium Cache	95%	1.3 Mb	90% - 100%	↗
Big Cache	95%	1.3 Mb	90% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος εδώ πέφτει στο 0.

Η κατανάλωση μνήμης από αυτόν το μηχανισμό είναι μεγάλη εξαιτίας του ίδιου του μεγέθους της βάσης δεδομένων. Κι αυτό, γιατί ο υποκατάλογος λογικών ονομάτων περιέχει μια εγγραφή για καθεμιά από τις οντότητες της βάσης. Είναι, λοιπόν, αναμενόμενο να παρουσιάσει τόσο μεγάλο μέγεθος.

Παρά την τιμή του μεγέθους, ωστόσο, τα ποσοστά επιτυχίας είναι πολύ ικανοποιητικά κυρίως εξαιτίας της οργάνωσης της πληροφορίας του υποκαταλόγου, η οποία

αποδεικνύεται πολύ αποδοτική για τα συγκεκριμένα βήματα του σεναρίου χρήσης των δύο συστημάτων που χρησιμοποιήθηκαν στα πειράματα.

Symbol Table Cache

Σύστημα Διατύπωσης Ερωτήσεων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	80%	450 Kb	35% - 100%	$\nearrow, \rightarrow$
Medium Cache	90%	7.1 Mb	40% - 100%	$\nearrow, \rightarrow$
Big Cache	100%	15 Mb	$\simeq$ 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	85%	450 Kb	80% - 100%	$\nearrow, \rightarrow$
Medium Cache	97%	1 Mb	95% - 100%	$\nearrow$
Big Cache	97%	1 Mb	95% - 100%	$\nearrow$

Η τιμή του μεγέθους της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της περιορίζεται και σ'αυτήν την περίπτωση στο 0.

Μεγάλο ενδιαφέρον παρουσιάζει στην περίπτωση αυτή η μεγάλη διαφορά μεγέθους που παρατηρείται για το μηχανισμό ανάμεσα στο δύο συστήματα που μελετήθηκαν. Πιο συγκεκριμένα, το μέγεθος φτάνει τα 15 Mb στο σύστημα διατύπωσης ερωτήσεων ενώ δεν ξεπερνά το 1 Mb για το σύστημα δελτίων ενημέρωσης. Αιτία του φαινομένου, λαμβάνοντας υπόψη και τη συμπεριφορά των ποσοστών επιτυχούς προσπέλασης στην κρυφή μνήμη τα οποία αυξάνονται για το σύστημα διατύπωσης ερωτήσεων όσο αυξάνεται η χωρητικότητα, είναι οι ίδιες οι ερωτήσεις που τίθενται με βάση το σενάριο χρήσης του συστήματος που χρησιμοποιείται.

Ο υποκατάλογος αναγνωριστικών συστήματος, εξάλλου, αποτελεί σημείο πολύ συχνής προσπέλασης. Εδώ πραγμτοποιείται η αντιστοίχιση αναγνωριστικών συστήματος - τα οποία καταλαβαίνει ο υπολογιστής - σε λογικά ονόματα - τα οποία καταλαβαίνει ο χρήστης. Η λειτουργία αυτή επιτελείται πάντα από το σύστημα προκειμένου να παρουσιαστούν στο χρήστη οποιαδήποτε αποτελέσματα. Καθώς, λοιπόν, οι ποσότητες πληροφορίας που επιστρέφονται στο χρήστη κατά τη χρήση του συστήματος διατύπωσης ερωτήσεων είναι σημαντικές, είναι αναμενόμενο τα μεγέθη αυτού του μηχανισμού να είναι μεγάλα.

Με δεδομένο, ωστόσο, ότι η υπερβολική κατανάλωση μνήμης μπορεί να αποτελέσει σοβαρό πρόβλημα στην απόδοση του συστήματος και καθώς το μέσο ποσοστό επιτυχίας για την προσπέλαση της κρυφής μνήμης δε μειώνεται κάτω από το 75% ακόμη και για τη μικρότερη χωρητικότητά της, προκύπτει ότι η διατήρηση της μικρότερης χωρητικότητας για την κρυφή μνήμη δε δημιουργεί πρόβλημα στο σύστημα.

Overflow Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακές τιμές hit rate	Συμπεριφορά μεγέθους
Small Cache	90%	60 Kb	50% - 100%	→
Medium Cache	90%	60 Kb	50% - 100%	→
Big Cache	90%	60 Kb	50% - 100%	→

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	90%	60 Kb	50% - 100%	→
Medium Cache	90%	60 Kb	50% - 100%	→
Big Cache	90%	60 Kb	50% - 100%	→

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της, μειώνεται στα 5 Kb.

Η σταθερότητα του μεγέθους της κρυφής μνήμης κατά την εκτέλεση των δοσοληψιών, τόσο του συστήματος διατύπωσης ερωτήσεων όσο και του συστήματος δελτίων εισαγωγής δεδομένων δημιουργεί την πρόβλεψη περιορισμένου αριθμού προσπελάσεων της κρυφής μνήμης. Η πρόβλεψη επιβεβαιώνεται από τις μετρήσεις των ποσοστών επιτυχών προσπελάσεων της κρυφής μνήμης, όπου φαίνεται ότι οι προσπελάσεις είναι σποραδικές. Παρουσιάζουν, ωστόσο, πολύ ικανοποιητικά ποσοστά επιτυχίας - σταθερά μεγαλύτερα από 85% με εξαίρεση τις δοσοληψίες εκκίνησης των δύο συστημάτων.

Named Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	100%	250 Kb	$\simeq$ 100%	↗
Medium Cache	100%	250 Kb	$\simeq$ 100%	↗
Big Cache	100%	250 Kb	$\simeq$ 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	120 Kb	40% - 100%	↗
Medium Cache	95%	120 Kb	40% - 100%	↗
Big Cache	95%	120 Kb	40% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος περιορίζεται σε 40 Kb.

Η επίδοση του μηχανισμού είναι και σ'αυτήν την περίπτωση πολύ ικανοποιητική. Το μέγεθος παραμένει σε χαμηλά επίπεδα και τα ποσοστά επιτυχίας είναι πολύ υψηλά, γεγονός το οποίο και πάλι δεν μπορεί να αποδωθεί παρά μόνο στην τοπικότητα των προσπελάσεων σε σχέση με την οργάνωση της πληροφορίας στον κατάλογο συστήματος για οντότητες με δοσμένο λογικό όνομα.

Unnamed Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	450 Kb	90% - 100%	↗
Medium Cache	95%	450 Kb	90% - 100%	↗
Big Cache	95%	450 Kb	90% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	100%	50 Kb	$\simeq$ 100%	$\nearrow$
Medium Cache	100%	50 Kb	$\simeq$ 100%	$\nearrow$
Big Cache	100%	50 Kb	$\simeq$ 100%	$\nearrow$

Η δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης προκαλεί μείωση του μεγέθους της στα 30 Kb.

Η συμπεριφορά του μηχανισμού είναι εντελώς ανάλογη με αυτή του προηγούμενου, οδηγώντας και πάλι στο συμπέρασμα για καλή οργάνωση της πληροφορίας που αποθηκεύεται στην κρυφή μνήμη - των εγγραφών του καταλόγου συστήματος για οντότητες χωρίς δοσμένο λογικό όνομα - σε σχέση με τις ερωτήσεις που εκτελούνται από το σενάριο τυπικής χρήσης των συστημάτων που χρησιμοποιήθηκε.

Μετρήσεις για σύνολα προσωρινής αποθήκευσης δεδομένων

Τα αποτελέσματα αυτών των μετρήσεων για τη βάση του AAT, έχουν ως εξής :

	Χαρακτηριστικά μεγέθη			
Σύστημα	Μέγιστη μνήμη	Ελάχιστη μνήμη	Μέγιστο πλήθος	Ελάχιστο πλήθος
Διατύπωσης ερωτήσεων	1.2 Mb	0 Kb	6000	0
Δελτίων εισαγωγής δεδομένων	200 Kb	0 Kb	3000	0

Οπως φαίνεται στον παραπάνω πίνακα, δεν απομένουν εδώ σύνολα στο τέλος των δοσοληψιών. Το φαινόμενο εδώ οφείλεται στο είδος των ερωτήσεων που τίθενται στη βάση, καθώς αυτές είναι που καθορίζουν την απελευθέρωση ή όχι της μνήμης που χρειάστηκαν για την δημιουργία συνόλων προσωρινής αποθήκευσης.

Εμφανίζεται κι εδώ, εξάλλου, η απότομη αύξηση του πλήθους των συνόλων που απαιτούνται για την εκτέλεση ερωτήσεων οι οποίες ανακαλούν μεγάλο πλήθος οντοτήτων - στη συγκεκριμένη περίπτωση ατομικών - από τη βάση.

Η κατανάλωση μνήμης, τέλος, από το μηχανισμό είναι στη βάση αυτή σημαντικότερη. Πρέπει και πάλι να τονιστεί, ωστόσο, ότι αυτή εξαρτάται αποκλειστικά και μόνο από τις

ερωτήσεις που κάθε φορά τίθενται στη βάση.

Μετρήσεις για οντότητες κι επεκτάσεις οντοτήτων στη μνήμη

Η κατανάλωση μνήμης από οντότητες και επεκτάσεις οντοτήτων που αντιγράφονται στη μνήμη από το μηχανισμό κρυφής μνήμης που τις περιέχει προκειμένου να χρησιμοποιηθούν, έχει ως εξής για τα δύο συστήματα που μελετήθηκαν :

Σύστημα Διατύπωσης Ερωτήσεων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	34 Kb	990 Kb
MC/SQ	37 Kb	1 Mb
SC/MQ	34 Kb	990 Kb
MC/MQ	37 Kb	1 Mb
BC/MQ	37 Kb	1 Mb
MC/BQ	37 Kb	1 Mb
BC/BQ	37 Kb	1 Mb

Δελτία Εισαγωγής Δεδομένων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	8 Kb	25 Kb
MC/SQ	8 Kb	25 Kb
SC/MQ	8 Kb	25 Kb
MC/MQ	8 Kb	25 Kb
BC/MQ	8 Kb	25 Kb
MC/BQ	8 Kb	25 Kb
BC/BQ	8 Kb	25 Kb

Σε αντίθεση με την περίπτωση του συστήματος ΚΛΕΙΩ, αρχίζει εδώ και παρουσιάζεται εξάρτηση της κατανάλωσης μνήμης από οντότητες και επεκτάσεις οντοτήτων από τις χωρητικότητες των μηχανισμών διαχείρισης κρυφής μνήμης για το σύστημα διατύπωσης ερωτήσεων.

Πιο συγκεκριμένα, η κατανάλωση μνήμης εξαρτάται από τις χωρητικότητες των μηχανισμών κρυφής μνήμης με δομή πίνακα κατακερματισμού, με αποτέλεσμα όταν αυτές παίρνουν τις μικρότερες τιμές τους να περιορίζεται και η κατανάλωση μνήμης από οντότητες κι επεκτάσεις τους. Αυτό συμβαίνει διότι η απομάκρυνση οντοτήτων κι επεκτάσεων οντοτήτων από τη μνήμη μεθοδεύεται από την κρυφή μνήμη του καταλόγου συστήματος. Οταν σ'αυτήν πραγματοποιείται αντικατάσταση σελίδων, όλες οι οντότητες και οι επεκτάσεις οντοτήτων για τις οποίες περιέχεται πληροφορία στις σελίδες που αντικαθίστανται απομακρύνονται από τη μνήμη.

Αυτό συμβαίνει μόνο στο σύστημα διατύπωσης ερωτήσεων, εξάλλου, διότι εκεί τα μεγέθη των κρυφών μνημών είναι μεγαλύτερα και φτάνουν τις χωρητικότητες των μηχανισμών.

Συνολική συμπεριφορά του συστήματος

Τα συνολικά ποσοστά επιτυχίας που παρουσιάζονται κατά την προσπέλαση των κρυφών μνημών από το Σύστημα Σημασιολογικού Ευρετηριασμού είναι και σ'αυτήν την περίπτωση πολύ ικανοποιητικά.

Ετσι, όταν και τα δύο είδη κρυφής μνήμης που περιλαμβάνονται στο μηχανισμό διαχείρισης κρυφής μνήμης του συστήματος έχουν μεσαία χωρητικότητα, το ποσοστό επιτυχών προσπελάσεων στη μνήμη είναι πάνω από 90% για τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων ενώ για τις δοσοληψίες του συστήματος δελτίων εισαγωγής δεδομένων είναι ακόμη υψηλότερο και σχεδόν μόνιμα στο 100%.

Μειώνοντας τη χωρητικότητα των κρυφών μνημών που έχουν δομή καταλόγου κατακερματισμού, παρατηρείται μικρή επιδείνωση των ποσοστών επιτυχίας, ενώ αυξάνοντας το ίδιο μόνο μέγεθος παρουσιάζεται μικρή βελτίωση. Τα ίδια φαινόμενα, ωστόσο, δεν παρατηρούνται μειώνοντας τα μεγέθη των κρυφών μνημών που είναι οργανωμένες με τη μορφή μονοδιάστατου πίνακα-ουράς.

Προκύπτει, επομένως, ότι η χωρητικότητα των κρυφών μνημών με δομή καταλόγου κατακερματισμού επηρεάζει περισσότερο το συνολικό ποσοστό επιτυχίας του

συστήματος.

Σε ό,τι αφορά τη συνολική κατανάλωση μνήμης καθενός από τα δύο συστήματα που μελετήθηκαν, τα αποτελέσματα είναι τα εξής :

- Σύστημα Επαφής Χρήσης : Στην περίπτωση αυτή, η συνολική κατανάλωση μνήμης του συστήματος κυμαίνεται από 20 Mb στην περίπτωση που και οι δύο μορφές οργάνωσης των καταλόγων της κρυφής μνήμης έχουν τη μέγιστη χωρητικότητά τους μέχρι 11 Mb όταν η χωρητικότητα και των δύο είναι η μεσαία και 4 Mb όταν η χωρητικότητα είναι μικρή.

  Τη μεγαλύτερη συμβολή στα μεγέθη αυτά την έχουν οι κρυφές μνήμες που είναι οργανωμένες σε μορφή καταλόγου κατακερματισμού καθώς όταν αυξήσουμε το μέγεθος αυτού του είδους μνήμης αφήνοντας σταθερή τη χωρητικότητα των κρυφών μνημών που έχουν τη μορφή μονοδιάστατου πίνακα-ουράς, η κατανάλωση μνήμης αυξάνεται πολύ περισσότερο απ'ότι με την αντίστροφη αλλαγή των μεγεθών.

  Ολόκληρος ο κατάλογος των συνολικών καταναλώσεων μνήμης στα διαφορετικά πειράματα που πραγματοποιήθηκαν εμφανίζεται παρακάτω :

Big Cache-Big Queue	:	20 Mb
Big Cache-Medium Queue	:	19 Mb
Medium Cache-Big Queue	:	12.4 Mb
Medium Cache-Medium Queue	:	11 Mb
Small Cache-Medium Queue	:	4.1 Mb
Medium Cache-Small Queue	:	10.8 Mb
Small Cache-Small Queue	:	3.9 Mb

- Δελτία εισαγωγής δεδομένων : Στην περίπτωση του συστήματος των Δελτίων Εισαγωγής Δεδομένων, τα μεγέθη κυμαίνονται σε χαμηλότερα επίπεδα, φτάνοντας τα 4.9 Mb όταν και τα δύο είδη κρυφών μνημών έχουν τη μέγιστη χωρητικότητα, τα 3.1 Mb όταν και τα δύο είδη έχουν μεσαίσ χωρητικότητα και τα 2.4 Mb όταν και τα οι δύο δομές οργάνωσης των κρυφών μνημών περιορίζεται σε μικρή χωρητικότητα.

  Το ίδιο φαινόμενο της μεγαλύτερης επίδρασης των κρυφών μνημών που είναι οργανωμένες ως κατάλογοι κατακερματισμού στη συνολική κατανάλωση μνήμης του συστήματος παρατηρείται και σ'αυτήν την περίπτωση.

Ο κατάλογος των συνολικών καταναλώσεων μνήμης στα διαφορετικά πειράματα που πραγματοποιήθηκαν έχει ως εξής :

Big Cache-Big Queue	:	4.9 Mb
Big Cache-Medium Queue	:	3.1 Mb
Medium Cache-Big Queue	:	4.9 Mb
Medium Cache-Medium Queue	:	3.1 Mb
Small Cache-Medium Queue	:	3.1 Mb
Medium Cache-Small Queue	:	3.0 Mb
Small Cache-Small Queue	:	2.4 Mb

6.3.2 Συμπεράσματα

Τα αποτελέσματα των μετρήσεων για τη βάση του ΑΑΤ, έχουν συνοπτικά ως εξής :

Μεγέθη μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	42 Kb	42 Kb	42 Kb	37 Kb	37 Kb	37 Kb
Link Classes Queue	20 Kb	20 Kb	20 Kb	13 Kb	13 Kb	13 Kb
Link Tokens Queue	102 Kb	205 Kb	400 Kb	50 Kb	50 Kb	50 Kb
Token Nodes Queue	95 Kb	95 Kb	95 Kb	35 Kb	35 Kb	35 Kb
Object Extensions Queue	102 Kb	205 Kb	1.4 Mb	102 Kb	205 Kb	2 Mb
Hash Cache	1.06 Mb	1.4 Mb	1.4 Mb	1.06 Mb	1.3 Mb	1.3 Mb
Symbol Table Cache	450 Kb	7.1 Mb	15 Mb	450 Kb	1 Mb	1 Mb
Overflow Cache	60 Kb	60 Kb	60 Kb	60 Kb	60 Kb	60 Kb
Named Sysid Cache	250 Kb	250 Kb	250 Kb	120 Kb	120 Kb	120 Kb
Unnamed Sysid Cache	450 Kb	450 Kb	450 Kb	50 Kb	50 Kb	50 Kb

Ποσοστά επιτυχίας μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	90%	90%	90%	99%	99%	99%
Link Classes Queue	97%	97%	97%	100%	100%	100%
Link Tokens Queue	67%	70%	75%	100%	100%	100%
Token Nodes Queue	70%	70%	70%	95%	95%	95%
Object Extensions Queue	70%	75%	90%	57%	60%	90%
Hash Cache	90%	90%	90%	95%	95%	95%
Symbol Table Cache	80%	90%	100%	85%	97%	97%
Overflow Cache	90%	90%	90%	90%	90%	90%
Named Sysid Cache	100%	100%	100%	95%	95%	95%
Unnamed Sysid Cache	95%	95%	95%	100%	100%	100%

Τα παραπάνω αποτελέσματα οδηγούν στα ακόλουθα συμπεράσματα :

- Αρχίζει και εμφανίζεται κάποια διαφοροποίηση σε σχέση με τη βάση του συστήματος ΚΛΕΙΩ, ως προς τη συμπεριφορά των μηχανισμών κρυφής μνήμης. Τα μεγέθη των μηχανισμών με δομή μονοδιάστατου πίνακα - στους οποίους αποθηκεύονται οι δομές αναπαράστασης για τις διαφορετικές οντότητες της βάσης - παρουσιάζουν εδώ μεγαλύτερες διακυμάνσεις σε σχέση με την προηγούμενη βάση, οι οποίες οφείλονται στην ανισοκατανομή των περιεχομένων της βάσης σε σχήμα και δεδομένα.

- Η ύπαρξη πολυάριθμων συσχετίσεων - πολλές από τις οποίες δεν έχουν δοσμένο λογικό όνομα - μεταξύ των ατομικών οντοτήτων της βάσης έχει σαν αποτέλεσμα να παρουσιάζουν μεγάλα μεγέθη οι μηχανισμοί κρυφής μνήμης που αποθηκεύουν τις συσχετίσεις μεταξύ ατομικών οντοτήτων της βάσης, τις εγγραφές των καταλόγου συστήματος που αφορούν οντότητες της βάσης χωρίς δοσμένο λογικό όνομα, καθώς επίσης και τις δομές επάκτασης των οντοτήτων της βάσης, οι οποίες είναι στη βάση του ΑΑΤ περισσότερες σε πλήθος.

- Η κρυφή μνήμη για τον υποκατάλογο αναγνωριστικών συστήματος εμφανίζει, κατ'αναλογία με τη βάση του ΚΛΕΙΩ, μεγαλύτερα μεγέθη συγκριτικά με τους υπόλοιπους μηχανισμούς. Το μέγεθος είναι τώρα αυξημένο κατά μία τάξη μεγέθους, στοιχείο το οποίο οφείλεται στο μέγεθος της βάσης του ΑΑΤ, σε σχέση με αυτή του ΚΛΕΙΩ.

- Τα ποσοστά επιτυχίας παρουσιάζονται στοιχειωδώς επιδεινωμένα για κάποιους από τους μηχανισμούς, για διαφορετικούς σε κάθε περίπτωση λόγους. Αλλοτε ευθύνεται για το φαινόμενο η τοπικότητα των αναφορών κατά την εκτέλεση ερωτήσεων - όπως συμβαίνει για το μηχανισμό στον οποίο αποθηκεύονται οι ατομικές οντότητες της βάσης - και άλλοτε λόγω περιορισμού μνήμης εξαιτίας της μικρής χωρητικότητας του μηχανισμού - όπως στο μηχανισμό για τις δομές επέκτασης των οντοτήτων της βάσης.

Κεφάλαιο 7

Ακραία συμπεριφορά συστήματος : βάση ULAN

7.1 Εισαγωγή

Η τελευταία από τις βάσεις που αποτέλεσαν αντικείμενο των πειραμάτων για την εξέταση της συμπεριφοράς και των επιδόσεων του μηχανισμού διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού είναι η βάση ULAN. Αποτελεί κι αυτή μετάφραση σε Telos αντίστοιχης σχεσιακής βάσης δεδομένων. Περιεχόμενο της βάσης ULAN αποτελεί πληροφορία σχετικά με ”πρόσωπα” - καλλιτέχνες και όχι μόνο - καθώς επίσης και πληροφορία σχετική με ”καταχωρήσεις” που αφορούν αυτά τα πρόσωπα σε διάφορες πηγές.

Η ιδιαιτερότητα της βάσης ULAN σε σχέση με τις δύο προηγούμενες και ο λόγος για τον οποίο οι επιδόσεις της στα πειράματα που πραγματοποιήθηκαν παρουσιάζονται ξεχωριστά είναι το πολύ μεγάλο μέγεθός της, σε συνδυασμό με τον ιδιαίτερο τρόπο δόμησής της. Συγκεκριμένα, το μέγεθος της βάσης, όπως αυτή υλοποιήθηκε σε Telos, ξεπερνά τα 500 Mb και το περιεχόμενό της αποτελείται από 602 οντότητες και συσχετίσεις που απαρτίζουν το σχήμα της και 3 εκατομμύρια οντότητες που αποτελούν το τμήμα των δεδομένων της. Από το σύνολο των δεδομένων, περισσότερα από 2 εκατομμύρια είναι σύνδεσμοι χωρίς δοσμένο λογικό όνομα.

Τα δύο παραπάνω στοιχεία καθιστούν αυτή τη βάση ακραίο παράδειγμα οργάνωσης γνώσης γεγονός το οποίο συνέβαλε στην απόφαση για διεξαγωγή προσθέτων μετρήσεων και χωριστή παρουσίαση.

7.2 Αποτελέσματα μετρήσεων

Προκειμένου να συγκεντρωθούν τα στοιχεία επίδοσης του Συστήματος Σημασιολογικού Ευρετηριασμού όταν αυτό χρησιμοποιείται για ανάκληση πληροφορίας αλλά και ενημέρωση των περιεχομένων της βάσης ULAN, εκτελέστηκαν τα ίδια πειράματα που περιγράφηκαν και για τις δύο προηγούμενες βάσεις δεδομένων, καθώς επίσης και μια πρόσθετη σειρά πειραμάτων με παραμέτρους ακόμη μικρότερες χωρητικότητες για τις επιμέρους κρυφές μνήμες.

Η επίδοση των μηχανισμών κρυφής μνήμης μετρήθηκε, στην πρώτη σειρά πειραμάτων, για τρεις διαφορετικές χωρητικότητές τους ενώ τα μεγέθη που μετρήθηκαν με σκοπό να προσδιορίσουν αυτήν την επίδοση είναι τα μεγέθη των κρυφών μνημών, τα ποσοστά επιτυχών προσπελάσεών τους, τα μεγέθη της μνήμης που καταναλώνεται από το μηχανισμό δημιουργίας αντιγράφων των οντοτήτων και των επεκτάσεών τους προκειμένου αυτά να χρησιμοποιηθούν καθώς επίσης και τα μεγέθη μνήμης που καταναλώνονται από το μηχανισμό δημιουργίας και διαχείρισης συνόλων για την αποθήκευση δεδομένων και ενδιάμεσων αποτελεσμάτων για την εκτέλεση ερωτήσεων. Με τη βοήθεια όλων αυτών των αποτελεσμάτων, δημιουργήθηκε μια εικόνα σχετικά με τη συνολική συμπεριφορά του συστήματος ως προς δύο παραμέτρους : την κατανάλωση μνήμης και τα ποσοστά επιτυχών προσβάσεων σ'αυτήν, στοιχείο που δίνει ένα μέτρο του χρόνου απόκρισης του συστήματος.

7.2.1 Αποτελέσματα

Η οργάνωση της παρουσίασης των αποτελεσμάτων των μετρήσεων για τη βάση ULAN ακολουθεί την οργάνωση που υπήρξε και για τις πρώτες δύο παρουσιάσεις των αποτελεσμάτων των βάσεων του συστήματος ΚΛΕΙΩ και του θησαυρού ΑΑΤ.

Μηχανισμοί κρυφής μνήμης - Αποτελέσματα

Η δυσαναλογία στην κατανομή της πληροφορίας για τη βάση ULAN στους διάφορους τύπους δεδομένων που περιλαμβάνονται σε μία βάση καθιστά πολύ ενδιαφέρουσα τη μελέτη της συμπεριφοράς της. Οι μεγάλες αριθμητικές διαφορές στη βάση για τους διαφορετικούς τύπους οντοτήτων δημιουργεί την ανάγκη διεξαγωγής μιας μελέτης τέτοιας, που να μπορεί να μοιράζει τη μνήμη που διατίθεται στο μηχανισμό διαχείρισης

κρυφής μνήμης με τον αποδοτικότερο δυνατό τρόπο.

Class Nodes Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	28 Kb	70% - 100%	$\nearrow, \rightarrow$
Medium Queue	90%	28 Kb	70% - 100%	$\nearrow, \rightarrow$
Big Queue	90%	28 Kb	70% - 100%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	97%	28 Kb	70% - 100%	$\nearrow, \rightarrow$
Medium Queue	97%	28 Kb	70% - 100%	$\nearrow, \rightarrow$
Big Queue	97%	28 Kb	70% - 100%	$\nearrow, \rightarrow$

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθος του μηχανισμού περιορίζεται μόλις στα 25 Kb.

Η σταθερότητα του μεγέθους του κρυφής μνήμης, εξάλλου, σε συνδυασμό με το γεγονός ότι τα ποσοστά επιτυχίας είναι πολύ υψηλά - οι τιμές της τάξης του 70% παρουσιάζονται κατά την αρχικοποίηση των δύο συστημάτων - ενώ είναι γνωστό ότι το σχήμα της βάσης είναι πολύ περιορισμένο, οδηγούν στο συμπέρασμα ότι το μέγεθος των 28 Kb περιλαμβάνει το σύνολο των κλάσεων οντοτήτων στο σχήμα της βάσης.

Link Classes Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	95%	90 Kb	89% - 100%	$\nearrow, \rightarrow$
Medium Queue	95%	90 Kb	89% - 100%	$\nearrow, \rightarrow$
Big Queue	95%	90 Kb	89% - 100%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	100%	90 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Medium Queue	100%	90 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Big Queue	100%	90 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$

Το μέγεθος της κρυφής μνήμης φτάνει την τιμή των 10 Kb στη δοσοληψία ακύρωσης των περιεχομένων της.

Κατά τ'άλλα, η συμπεριφορά του μηχανισμού είναι εντελώς ανάλογη με αυτήν του προηγούμενου. Η σταθερότητα στην τιμή του μεγέθους, τα υψηλά ποσοστά επιτυχών προσπελάσεων καθώς επίσης και το γεγονός ότι εδώ αποθηκεύονται οι δομές αναπαράστασης των κλάσεων συσχετίσεων της βάσης, οι οποίες επίσης περιλαμβάνονται στο σχήμα της, οδηγούν και πάλι στο συμπέρασμα ότι στα 90 Kb περιλαμβάνονται το σύνολο των δομών αποθήκευσης αυτού του είδους που υπάρχουν στη βάση.

Link Tokens Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	70%	102 Kb	65% - 100%	↗, →, →
Medium Queue	85%	205 Kb	75% - 100%	↗, →
Big Queue	95%	2.1 Mb	90% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	95%	70 Kb	85% - 100%	↗
Medium Queue	95%	70 Kb	85% - 100%	↗
Big Queue	95%	70 Kb	85% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της περιορίζεται στα 20 Kb.

Σε ό,τι αφορά τα μεγέθη του μηχανισμού κατά τη χρήση των δύο συστημάτων που μελετήθηκαν, παρουσιάζεται και εδώ το φαινόμενο το οποίο εμφανίστηκε, σε πολύ μικρότερη έκταση, στη βάση του AAT. Αυτό δεν είναι άλλο από την αύξηση του μεγέθους σε πολύ υψηλά επίπεδα για τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων - εδώ φτάνει τα 2.1 Mb για τη μεγάλη χωρητικότητα της κρυφής μνήμης - ενώ στις δοσοληψίες του συστήματος δελτίων εισαγωγής δεδομένων παραμένει σε χαμηλά επίπεδα - εδώ μόλις 70 Kb. Αυτή η δυσαναλογία οφείλεται αποκλειστικά και μόνο στο είδος των ερωτήσεων που θέτουν στη βάση τα δύο συστήματα κατά την εκτέλεση του σεναρίου τυπικής χρήσης του συστήματος που χρησιμοποιήθηκε. Το γεγονός της αύξησης του μεγέθους στην τιμή των 2.1 Mb, εξάλλου, εμφανίζεται εξαιτίας του μεγάλου πλήθους συνδέσμων μεταξύ ατομικών οντοτήτων που διαθέτει η βάση, οι οποίοι ξεπερνάν τα 2 εκατομμύρια.

Παρά τις μεγάλες ανάγκες σε μνήμη, ωστόσο, τα ποσοστά επιτυχίας εμφανίζονται πολύ ικανοποιητικά ακόμη και για τη μικρότερη χωρητικότητα του μηχανισμού, γεγονός που αποτελεί απόδειξη τοπικότητας στις αναφορές στη βάση.

Tokens Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	60%	102 Kb	0% - 100%	↗, →, →
Medium Queue	65%	205 Kb	0% - 100%	↗, →
Big Queue	67%	330 Kb	0% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	90%	40 Kb	0% - 100%	↗
Medium Queue	90%	40 Kb	0% - 100%	↗
Big Queue	90%	40 Kb	0% - 100%	↗

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της περιορίζεται στα 20 Kb.

Στο μηχανισμό αυτό, παρουσιάζεται συμπεριφορά ανάλογη με αυτή του προηγούμενου, σε ό,τι αφορά τα μεγέθη της κρυφής μνήμης για τα δύο διαφορετικά συστήματα. Ενώ, λοιπόν, το μέγεθος δεν ξεπερνά τα 40 Kb για τις δοσοληψίες του συστήματος δελτίων εισαγωγής δεδομένων, αυξάνεται μέχρι τα 330 Kb στις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων. Η αιτία βρίσκεται και πάλι αποκλειστικά και μόνο στο είδος των ερωτήσεων που τίθενται στη βάση από τα δύο συστήματα.

Τα χαμηλά ποσοστά επιτυχίας για το σύστημα δελτίων εισαγωγής δεδομένων καθώς επίσης και για τη μεγάλη χωρητικότητα της κρυφής μνήμης για το σύστημα διατύπωσης ερωτήσεων, προκύπτουν από έλλειψη τοπικότητας στις αναφορές που πραγματοποιούνται κατά τη χρήση των δύο συστημάτων. Κατά τη χρήση του συστήματος διατύπωσης ερωτήσεων με τη μικρή και τη μεσαία χωρητικότητα της κρυφής μνήμης, τα χαμηλά ποσοστά οφείλονται στο συνδυασμό του παραπάνω λόγου με τον περιορισμό μεγέθους που επιβάλλει η χωρητικότητα.

Object Extensions Queue

Σύστημα Διατύπωσης Ερωτήσεων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	60%	102 Kb	0% - 90%	$\nearrow, \rightarrow, \rightarrow$
Medium Queue	60%	205 Kb	0% - 90%	$\nearrow, \rightarrow, \rightarrow$
Big Queue	60%	1.7 Mb	0% - 90%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

		Στοιχεία συμπεριφοράς		
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Queue	83%	102 Kb	70% - 100%	$\nearrow, \rightarrow, \rightarrow$
Medium Queue	86%	205 Kb	70% - 100%	$\nearrow, \rightarrow, \rightarrow$
Big Queue	95%	1.5 Mb	70% - 100%	$\nearrow$

Καμμία διαφοροποίηση του μεγέθους δεν παρουσιάζεται κατά τις δοσοληψίες ακύρωσης των περιεχομένων της κρυφής μνήμης.

Η ύπαρξη μεγάλου πλήθους συνδέσμων ανάμεσα σε ατομικές οντότητες στη βάση οδηγεί στην ύπαρξη εξίσου μεγάλου πλήθους δομών επέκτασης για τις δομές αναπαράστασης

των οντοτήτων της βάσης. Αυτό έχει σαν αποτέλεσμα ο μηχανισμός αυτός να αποκτά πολύ μεγάλο μέγεθος - στις περιπτώσεις που του επιτρέπεται - φτάνοντας τα 1.7 Mb για το σύστημα διατύπωσης ερωτήσεων και τα 1.5 Mb για το σύστημα διαλογικών δελτίων ενημέρωσης.

Τα ποσοστά επιτυχίας, εξάλλου, παρουσιάζονται υψηλά στο δεύτερο από τα παραπάνω συστήματα εξαιτίας μεγαλύτερης τοπικότητας των αναφορών από τις ερωτήσεις του σεναρίου τυπικής χρήσης. Στο σύστημα διατύπωσης ερωτήσεων τα ποσοστά παρουσιάζονται σχετικά χαμηλά, στοιχείο που οφείλεται κατά κύριο λόγο στην έλλειψη τοπικότητας των αναφορών σε σχέση με την οργάνωση της πληροφορίας όπως αποδεικνύεται από το γεγονός ότι τα ποσοστά δε βελτιώνονται ούτε στην περίπτωση που η χωρητικότητα της κρυφής μνήμης είναι τέτοια που δεν περιορίζει τις ανάγκες σε μνήμη.

Hash Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακές τιμές hit rate	Συμπεριφορά μεγέθους
Small Cache	77%	1.05 Mb	0% - 100%	↗
Medium Cache	77%	1.05 Mb	0% - 100%	↗
Big Cache	77%	1.05 Mb	0% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	85%	1.05 Mb	0% - 100%	$\nearrow, \rightarrow$
Medium Cache	95%	2.2 Mb	80% - 100%	$\nearrow$
Big Cache	95%	2.2 Mb	80% - 100%	$\nearrow$

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της, περιορίζεται σε 50 Kb.

Οι τιμές στις οποίες ανέρχεται το μέγεθος της κρυφής μνήμης για τα δύο συστήματα, χωρίς να είναι υπερβολικά υψηλές, εξασφαλίζουν πολύ καλή συμπεριφορά, η οποία αποδεικνύεται από τα υψηλά ποσοστά επιτυχίας. Οι τιμές της τάξης του 0% που εμφανίζονται σε ορισμένες από τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων αντιστοιχούν σε βήματα του σεναρίου χρήσης στα οποία πραγματοποιείται αλλαγή του τμήματος της βάσης που εμπίπτει στο άμεσο ενδιαφέρον για την εκτέλεση των ερωτήσεων που ακολουθούν. Αυτό έχει σαν αποτέλεσμα να μην υπάρχουν στην κρυφή μνήμη σελίδες πληροφορίας σχετικές με το νέο τμήμα της βάσης κι επομένως να εμφανίζεται στιγμιαία επιδείνωση της επίδοσης.

Symbol Table Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	85%	440 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
Medium Cache	95%	7.1 Mb	80% - 100%	$\nearrow, \rightarrow, \rightarrow$
Big Cache	97%	35 Mb	90% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	300 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Medium Cache	98%	300 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$
Big Cache	98%	300 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$

Το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης περιορίζεται σε 50 Kb.

Με πολύ ικανοποιητικά ποσοστά επιτυχών προσπελάσεων, ο μηχανισμός εμφανίζει συμπεριφορά ως προς το μέγεθος που καταλαμβάνει αντίστοιχη με αυτή των κρυφών μνημών για τις δομές επέκτασης των οντοτήτων και για τις συσχετίσεις μεταξύ ατομικών οντοτήτων στη βάση. Με άλλα λόγια, το μέγεθος φτάνει τα 35 Mb - με τη μεγάλη χωρητικότητα της κρυφής μνήμης - για το σύστημα διατύπωσης ερωτήσεων, ενώ δεν ξεπερνάει τα 300 Kb, για το σύστημα δελτίων εισαγωγής δεδομένων.

Overflow Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	97%	260 Kb	90% - 100%	$\nearrow, \rightarrow$
Medium Cache	97%	520 Kb	90% - 100%	$\nearrow, \rightarrow$
Big Cache	97%	520 Kb	90% - 100%	$\nearrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	32%	260 Kb	0% - 40%	$\nearrow, \rightarrow$
Medium Cache	34%	700 Kb	0% - 40%	$\nearrow, \rightarrow$
Big Cache	34%	700 Kb	0% - 40%	$\nearrow, \rightarrow$

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της περιορίζεται σε 20 Kb.

Η επίδοση του μηχανισμού είναι πολύ ικανοποιητική για το σύστημα διατύπωσης ερωτήσεων καθώς, τόσο η κατανάλωση μνήμης, όσο και τα ποσοστά επιτυχίας είναι πολύ ικανοποιητικά. Δεν ισχύει το ίδιο, ωστόσο, και για το σύστημα δελτίων εισαγωγής δεδομένων, στο οποίο η σχετικά χαμηλή κατανάλωση μνήμης συνοδεύεται από πολύ χαμηλά ποσοστά επιτυχίας. Το φαινόμενο αυτό οφείλεται αποκλειστικά και μόνο στις προσπελάσεις στο βοηθητικό κατάλογο εγγραφών που προκαλούν οι συγκεκριμένες ερωτήσεις που τίθενται στη βάση.

Named Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	98%	460 Kb	90% - 100%	$\nearrow, \rightarrow$
Medium Cache	99%	700 Kb	95% - 100%	$\nearrow$
Big Cache	99%	700 Kb	95% - 100%	$\nearrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	95%	300 Kb	40% - 100%	↗
Medium Cache	95%	300 Kb	40% - 100%	↗
Big Cache	95%	300 Kb	40% - 100%	↗

Κατά τη δοσοληψία ακύρωσης των περιεχομένων της κρυφής μνήμης, το μέγεθός της περιορίζεται σε 90 Kb.

Η επίδοση του μηχανισμού, κατά τα άλλα, εξακολουθεί να εμφανίζεται, όπως και στα προηγούμενα πειράματα, πολύ ικανοποιητική, τόσο ως προς την κατανάλωση μνήμης η οποία είναι πολύ χαμηλή, όσο και ως προς τα ποσοστά επιτυχίας που δε μειώνονται σε τιμή μικρότερη του 95%.

Unnamed Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	98%	460 Kb	75% - 100%	↗, →
Medium Cache	99%	5 Mb	95% - 100%	↗
Big Cache	99%	5 Mb	95% - 100%	↗

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακά τιμών hit rate	Συμπεριφορά μεγέθους
Small Cache	47%	100 Kb	20% - 80%	↗
Medium Cache	47%	100 Kb	20% - 80%	↗
Big Cache	47%	100 Kb	20% - 80%	↗

Στα 20 Kb περιορίζεται το μέγεθος της κρυφής μνήμης κατά τη δοσοληψία ακύρωσης των περιεχομένων της.

Κατ'αναλογία προς τη συμπεριφορά των μηχανισμών κρυφής μνήμης για τον υποκατάλογο αναγνωριστικών συστήματος, τις δομές επέκτασης των οντοτήτων της βάσης και τις συσχετίσεις μεταξύ ατομικών οντοτήτων στη βάση, παρουσιάζεται κι εδώ σημαντική διαφοροποίηση μεταξύ του συστήματος διατύπωσης ερωτήσεων και αυτού των δελτίων εισαγωγής δεδομένων.

Ετσι, στο πρώτο από τα παραπάνω συστήματα, τα ποσοστά επιτυχίας είναι πολύ ικανοποιητικά ενώ το μέγεθος φτάνει τα 5 Mb. Στο σύστημα δελτίων εισαγωγής δεδομένων, από την άλλη μεριά, η χαμηλή κατανάλωση μνήμης συνοδεύεται από χαμηλά ποσοστά επιτυχίας, τα οποία οφείλονται στην έλλειψη τοπικότητας των αναφορών στη μνήμη.

Μετρήσεις για σύνολα προσωρινής αποθήκευσης δεδομένων

Τα αποτελέσματα των μετρήσεων για το μηχανισμό διαχείρισης συνόλων προσωρινής αποθήκευσης έχουν ως εξής :

	Χαρακτηριστικά μεγέθη			
Σύστημα	Μέγιστη μνήμη	Ελάχιστη μνήμη	Μέγιστο πλήθος	Ελάχιστο πλήθος
Διατύπωσης ερωτήσεων	1 Mb	10 Kb	5000	100
Δελτίων εισαγωγής δεδομένων	1 Mb	10 Kb	7000	1500

Η συμπεριφορά του μηχανισμού εξακολουθεί να παρουσιάζει τα ίδια χαρακτηριστικά με τις προγούμενες βάσεις που μελετήθηκαν. Η κατανάλωση μνήμης είναι εδώ αυξημένη και για τα δύο συστήματα εξαιτίας του ότι οι ερωτήσεις που τώρα τίθενται στη βάση έχουν

πολυπληθέστερα αποτελέσματα, τα οποία πριν εμφανιστούν στο χρήστη αποθηκεύονται σε σύνολα.

Το ίδιο ισχύει και για το πλήθος των συνόλων που χρησιμοποιούνται.

Μετρήσεις για οντότητες και επεκτάσεις οντοτήτων στη μνήμη

Η κατανάλωση μνήμης από οντότητες και επεκτάσεις οντοτήτων για τη βάση του ULAN και γι'αυτή τη σειρά πειραμάτων διαμορφώνεται ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	72 Kb	1.2 Mb
MC/SQ	390 Kb	1.3 Mb
SC/MQ	72 Kb	1.2 Mb
MC/MQ	390 Kb	1.3 Mb
BC/MQ	865 Kb	1.4 Mb
MC/BQ	390 Kb	1.3 Mb
BC/BQ	865 Kb	1.4 Mb

Δελτία Εισαγωγής Δεδομένων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
SC/SQ	13 Kb	1.1 Mb
MC/SQ	13 Kb	1.1 Mb
SC/MQ	13 Kb	1.1 Mb
MC/MQ	13 Kb	1.1 Mb
BC/MQ	13 Kb	1.1 Mb
MC/BQ	13 Kb	1.1 Mb
BC/BQ	13 Kb	1.1 Mb

Παρουσιάζεται κι εδώ το φαινόμενο της εξάρτησης των μεγεθών μνήμης που καταναλώνονται από οντότητες και επεκτάσεις οντοτήτων από τις χωρητικότητες των μηχανισμών κρυφής μνήμης με δομή καταλόγου κατακερματισμού. Ο λόγος είναι ο ίδιος με αυτόν που αναλύθηκε στην περίπτωση της βάσης του AAT, όπου παρατηρήθηκε το φαινόμενο για πρώτη φορά.

Παρατηρείται, εξάλλου, εδώ το φαινόμενο πολύ μεγάλων μεγεθών για την κατανάλωση μνήμης από επεκτάσεις οντοτήτων, γεγονός το οποίο οφείλεται τόσο στην ύπαρξη πολύ μεγάλου αριθμού δομών επέκτασης στη βάση λόγω του πλήθους των συσχετίσεων μεταξύ ατομικών οντοτήτων που περιέχει, καθώς επίσης και στο σενάριο τυπικής χρήσης, το οποίο προκαλεί ανάκληση μεγάλου πλήθους συνδέσμων της βάσης.

Συνολική συμπεριφορά του συστήματος

Παρά το γεγονός ότι η συνολική κατανάλωση μνήμης, τόσο για το σύστημα διατύπωσης ερωτήσεων, όσο και για το σύστημα διαλογικής ενημέρωσης της βάσης μέσω δελτίων εισαγωγής δεδομένων είναι πολύ αυξημένη όταν οι χωρητικότητες των μηχανισμών κρυφής μνήμης το επιτρέπουν, τα ποσοστά επιτυχών προσπελάσεων στη μνήμη κυμαίνονται σε ψηλά επίπεδα, ακόμη κι όταν οι χωρητικότητες έχουν τη χαμηλότερη τιμή τους.

Ετσι, τα ποσοστά επιτυχών προσπελάσεων στη μνήμη δε μειώνονται σε τιμή μικρότερη από 80%, ακόμη και για τα μικρότερα μεγέθη χωρητικότητας των κρυφών μνημών, τιμή στην οποία φτάνει το ποσοστό όταν αλλάζει το τμήμα της βάσης το σχετικό με τις ερωτήσεις που τίθενται. Η τιμή αυξάνεται στις αμέσως επόμενες δοσοληψίες, φτάνοντας και το 100%, ανώ μειώνεται και πάλι στην πρώτη δοσοληψία τροποποίησης του τμήματος της βάσης που ενδιαφέρει ως προς την εκτέλεση των ερωτήσεων.

Εξαίρεση στην παραπάνω συμπεριφορά αποτελούν ολιγάριθμες δοσοληψίες στις οποίες το ποσοστό μειώνεται ακόμη και κάτω από 30%. Οι δοσοληψίες αυτές αντιστοιχούν σε ερωτήσεις στη βάση αμέσως μετά την ακύρωση των περιεχομένων της κρυφής μνήμης, εξαιτίας ενημέρωσης της βάσης στην προηγούμενη δοσοληψία, καθώς επίσης και σε δοσοληψίες ενημέρωσης της βάσης κατά τις οποίες πραγματοποιείται σημασιολογικός και συντακτικός έλεγχος της εισαγόμενης πληροφορίας.

Σε ό,τι αφορά, εξάλλου, τη συνολική κατανάλωση μνήμης κατά τη χρήση των δύο

συστημάτων, αυτή είναι η εξής :

- Σύστημα διατύπωσης ερωτήσεων : Οπως φάνηκε και κατά την αναλυτική παρουσίαση των μεγεθών των κρυφών μνημών για τις διαφορετικές χωρητικότητες για τις οποίες εξετάστηκαν, τα μεγέθη μνήμης που καταναλώνονται από το σύστημα μπορούν να φτάσουν σε πολύ υψηλά επίπεδα. Ετσι, η συνολική κατανάλωση μνήμης μπορεί να φτάσει και τα 50 Mb, ενώ η εικόνα που παρουσιάζεται είναι αναλυτικά η ακόλουθη :

Big Cache - Big Queue	:	49.6 Mb
Big Cache - Medium Queue	:	46.1 Mb
Medium Cache - Big Queue	:	21.25 Mb
Medium Cache - Medium Queue	:	17.7 Mb
Small Cache - Medium Queue	:	5.7 Mb
Medium Cache - Small Queue	:	17.4 Mb
Small Cache - Small Queue	:	5.4 Mb

  Οπως φαίνεται από τα παραπάνω αποτελέσματα, η χωρητικότητα των κρυφών μνημών με δομή καταλόγου κατακερματισμού είναι αυτή που παίζει το σημαντικότερο ρόλο στη συνολική κατανάλωση μνήμης του συστήματος. Κι αυτό, γιατί οι κρυφές μνήμες που έχουν αυτή τη δομή αποθηκεύουν εγγραφές για καθεμιά από τις οντότητες που χρησιμοποιούνται με σκοπό την απάντηση ερωτήσεων, όπως επίσης και για τις συσχετίσεις μεταξύ οντοτήτων. Το γεγονός αυτό, έχει σαν αποτέλεσμα πολύ περισσότερες εγγραφές να αποθηκεύονται στις μνήμες αυτές σε σχέση με τις υπόλοιπες, οι οποίες έχουν δομή μονοδιάστατου πίνακα-ουράς και οι οποίες αποθηκεύουν εγγραφές για διαφορετικές οντότητες ή συσχετίσεις.

- Σύστημα δελτίων εισαγωγής δεδομένων : Η συνολική κατανάλωση μνήμης στην περίπτωση αυτή είναι σαφώς μειωμένη. Ετσι, τα μεγέθη κυμαίνονται τώρα ανάμεσα σε 4.5 και 7.5 Mb. Αιτία για το φαινόμενο αυτό αποτελεί η μικρότερη ποσότητα πληροφορίας που ανασύρεται από τη βάση προκειμένου να πραγματοποιηθεί κάποια ενημέρωση, καθώς επίσης και ενδεχόμενη μεγαλύτερη τοπικότητα που παρουσιάζει η πληροφορία που ανασύρεται.

  Αναλυτικότερα, τα αποτελέσματα εδώ έχουν ως εξής :

Big Cache - Big Queue	:	7.4 Mb
Big Cache - Medium Queue	:	6.1 Mb
Medium Cache - Big Queue	:	7.4 Mb
Medium Cache - Medium Queue	:	6.1 Mb
Small Cache - Medium Queue	:	4.5 Mb
Medium Cache - Small Queue	:	6.04 Mb
Small Cache - Small Queue	:	4.4 Mb

Η χωρητικότητα των κρυφών μνημών με δομή καταλόγου κατακερματισμού επηρεάζει, όπως φαίνεται παραπάνω, και πάλι τη συνολική κατανάλωση μνήμης του συστήματος διαλογικών δελτίων ενημέρωσης, όχι όμως στον ίδιο βαθμό που αυτό συμβαίνει στο σύστημα διατύπωσης ερωτήσεων.

7.2.2 Συμπεράσματα

Η συνολική συμπεριφορά των συστημάτων που μελετήθηκαν για τη βάση του ULAN έχει ως εξής :

Μεγέθη μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	28 Kb	28 Kb	28 Kb	28 Kb	28 Kb	28 Kb
Link Classes Queue	90 Kb	90 Kb	90 Kb	90 Kb	90 Kb	90 Kb
Link Tokens Queue	102 Kb	205 Kb	2.1 Mb	70 Kb	70 Kb	70 Kb
Token Nodes Queue	102 Kb	205 Kb	330 Kb	40 Kb	40 Kb	40 Kb
Object Extensions Queue	102 Kb	205 Kb	1.7 Mb	102 Kb	205 Kb	1.5 Mb
Hash Cache	1.05 Mb	1.05 Mb	1.05 Mb	1.05 Mb	2.2 Mb	2.2 Mb
Symbol Table Cache	440 Kb	7.1 Mb	35 Mb	300 Kb	300 Kb	300 Kb
Overflow Cache	260 Kb	520 Kb	520 Kb	260 Kb	700 Kb	700 Kb
Named Sysid Cache	460 Kb	700 Kb	700 Kb	300 Kb	300 Kb	300 Kb
Unnamed Sysid Cache	460 Kb	5 Mb	5 Mb	100 Kb	100 Kb	100 Kb

Ποσοστά επιτυχίας μηχανισμών κρυφής μνήμης

	Χωρητικότητα			Χωρητικότητα		
Κρυφή μνήμη	Small	Medium	Big	Small	Medium	Big
Class Nodes Queue	90%	90%	90%	97%	97%	97%
Link Classes Queue	95%	95%	95%	100%	100%	100%
Link Tokens Queue	70%	85%	95%	95%	95%	95%
Token Nodes Queue	60%	65%	67%	90%	90%	90%
Object Extensions Queue	60%	60%	60%	83%	86%	95%
Hash Cache	77%	77%	77%	85%	95%	95%
Symbol Table Cache	85%	95%	97%	95%	98%	98%
Overflow Cache	97%	97%	97%	32%	34%	34%
Named Sysid Cache	98%	99%	99%	95%	95%	95%
Unnamed Sysid Cache	98%	99%	99%	47%	47%	47%

Τα παραπάνω αποτελέσματα οδηγούν στα ακόλουθα συμπεράσματα :

- Οι μηχανισμοί κρυφής μνήμης οι οποίοι παρουσίαζαν αυξημένες ανάγκες σε μνήμη στα πειράματα για τις δύο άλλες βάσεις που μελετήθηκαν, παρουσιάζουν εδώ πολύ μεγάλα μεγέθη. Συγκεκριμένα, οι μηχανισμοί κρυφής μνήμης για τον υποκατάλογο αναγνωριστικών συστήματος, τις δομές αναπαράστασης των συσχετίσεων ανάμεσα σε ατομικές οντότητες της βάσης και τις δομές επέκτασης των οντοτήτων της βάσης καταλαμβάνουν συνολικά 40 περίπου Mb.

  Η τάξη μεγέθους είναι κι εδώ μεγαλύτερη από αυτή των αποτελεσμάτων για τις δύο προηγούμενες βάσεις, γεγονός που οφείλεται στο μέγεθος της βάσης του ULAN.

- Τα ποσοστά επιτυχίας είναι και πάλι, με λίγε εξαιρέσεις, πολύ ικανοποιητικά. Οι εξαιρέσεις εντοπίζονται στους ίδιους μηχανισμούς για τους οποίους παρουσιάστηκαν ελαττωμένα ποσοστά επιτυχίας και για τη βάση του AAT.

- Η γενική επίδοση της βάσης του ULAN θυμίζει πολύ έντονα αυτή της βάσης του AAT. Με δεδομένο ότι η βάση του AAT αποτελεί μικρογραφία αυτής του ULAN - μικρότερο μέγεθος με ίδια κατανομή δεδομένων - η συμπεριφορά ήταν αναμενόμενη.

  Επιβεβαιώνεται, λοιπόν, η πρόβλεψη ότι οι δύο βάσεις θα παρουσίαζαν συμπεριφορά ανάλογη, ενώ η διαφοροποίησή τους έγκειται στις τάξεις μεγέθους της κατανάλωσης μνήμης, οι οποίες διαφέρουν εξαιτίας των σχετικών μεγεθών των βάσεων.

7.3 Διεξαγωγή πρόσθετων μετρήσεων επίδοσης

Η βάση του ULAN, όπως έχει προαναφερθεί, αποτελεί παράδειγμα ακραίας συμπεριφοράς σε σχέση με το σύστημα σημασιολογικού ευρετηριασμού, του οποίου ο μηχανισμός διαχείρισης κρυφής μνήμης αποτελεί αντικείμενο μελέτης της εργασίας αυτής. Η συμπεριφορά της βάσης χαρακτηρίζεται ακραία καθώς το σχήμα της βάσης είναι εξαιρετικά μικρό σε σχέση με τα δεδομένα που περιέχει ενώ, επιπλέον, τα δεδομένα περιλαμβάνουν μεγάλο ποσοστό συνδέσμων ανάμεσα σε ατομικές οντότητες της βάσης. Για τον παραπάνω λόγο, επιλέχθηκε η βάση αυτή για τη διεξαγωγή επιπλέον μετρήσεων, κυρίως για τις κρυφές μνήμες των οποίων τα μεγέθη φτάνουν τις χωρητικότητές τους

τουλάχιστον όταν αυτές έχουν τις μικρότερές τους τιμές.

Με εξαίρεση, λοιπόν, τις κρυφές μνήμες για τις κλάσεις οντοτήτων και τις κλάσεις συνδέσμων - οι οποίες παραμένουν σταθερά σε ένα πολύ μικρό μέγεθος και για τις οποίες θεωρήθηκε ωφέλιμο να μην περιοριστούν σε μέγεθος, ώστε να περιέχουν όλο το σχήμα της βάσης - πραγματοποιήθηκαν νέες μετρήσεις για όλες τις υπόλοιπες μνήμες με νέες, μικρότερες τιμές για τις χωρητικότητές τους. Οι νέες χωρητικότητες προσδιορίζονται με βάση το μέγεθος των επιμέρους κρυφών μνημών στην 20η δοσοληψία του σεναρίου εκτέλεσης για το σύστημα διατύπωσης ερωτήσεων, όταν οι χωρητικότητες των κρυφών μνημών έχουν τις μικρές τιμές τους. Επιλέγεται η 20η δοσοληψία καθώς θεωρείται ότι όσες προηγούνται αποτελούν ένα ικανοποιητικό διάστημα εξάλειψης των μεταβατικών φαινομένων κατά την εκκίνηση για τις μνήμες.

Διαθέτοντας, λοιπόν, στις κρυφές μνήμες τον απαραίτητο χώρο ώστε να εξαλειφθούν τα μεταβατικά φαινόμενα κατά την εκκίνηση μεταφέροντας κάποιες σελίδες από τη βάση με κριτήριο τις ερωτήσεις που εκτελούνται, οι νέες μετρήσεις που πραγματοποιούνται σκοπό έχουν να μελετήσουν τη συμπεριφορά του συστήματος στη συνέχεια.

Ενας ακόμη λόγος που συνέτεινε στην ανάγκη πραγματοποίησης επιπλέον μετρήσεων ήταν η πολύ ικανοποιητική συμπεριφορά των δύο συστημάτων, παρά τον περιορισμό των χωρητικοτήτων των κρυφών μνημών μέχρι την μικρότερη τιμέ των πρώτων μετρήσεων. Τα ποσοστά επιτυχίας που παραμένουν σε ψηλά επίπεδα δημιουργούν την ανάγκη εντοπισμού των χωρητικοτήτων στις οποίες αρχίζει να εμφανίζεται επιδείνωσή τους.

Οι μετρήσεις που πραγματοποιήθηκαν για όλους τους παραπάνω λόγους, θέτουν όπως προαναφέρθηκε σαν όριο χωρητικότητας των κρυφών μνημών για το πρώτο πείραμα το μέγεθός τους κατά την 20η δοσοληψία για το πείραμα στο οποίο οι μνήμες έχουν τις μικρές τους χωρητικότητες. Στα δύο πειράματα που ακολουθούν, οι χωρητικότητες μειώνονται στο 1/3 και το 1/9 (περίπου) της αρχικής τιμής.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα που προέκυψαν κατά την εκτέλεση των πειραμάτων που περιγράφηκαν παραπάνω, για τις κρυφές μνήμες των οποίων η χωρητικότητα τροποποιείται, καθώς επίσης και για την κατανάλωση μνήμης από οντότητες και επεκτάσεις οντοτήτων.

7.3.1 Αποτελέσματα

Πριν την περιγραφή των αποτελεσμάτων των πειραμάτων με τις νέες χωρητικότητες των κρυφών μνημών, είναι απαραίτητο να παρουσιαστούν οι ίδιες οι χωρητικότητες των επιμέρους μηχανισμών κρυφής μνήμης για τα τρία νέα πειράματα.

Χωρητικότητες κρυφών μνημών με δομή μονοδιάστατου πίνακα

	Χωρητικότητες		
Πείραμα	Link Tokens	Token Nodes	Object Extensions
1ο Πείραμα	102 Kb	90 Kb	70 Kb
2ο Πείραμα	33 Kb	28 Kb	25 Kb
3ο Πείραμα	16 Kb	16 Kb	12 Kb

Χωρητικότητες κρυφών μνημών με δομή πίνακα κατακερματισμού

	Χωρητικότητες				
Πείραμα	Hash	Symbol Table	Overflow	Named Sysid	Unnamed Sysid
1ο Πείραμα	1.06 Mb	443 Kb	263 Kb	231 Kb	231 Kb
2ο Πείραμα	530 Kb	221 Kb	131 Kb	116 Kb	116 Kb
3ο Πείραμα	265 Kb	110 Kb	66 Kb	58 Kb	58 Kb

Μηχανισμοί κρυφής μνήμης - Αποτελέσματα

Οι νέες μετρήσεις που στοχεύουν σε περαιτέρω διερεύνηση της συμπεριφοράς των κρυφών μνημών που παρουσιάζουν ικανοποιητική συμπεριφορά, ακόμη κι όταν η χωρητικότητά τους είναι περιορισμένη, καθώς επίσης και αυτών που αποτελούν κρίσιμο πόρο για το σύστημα, έχουν τα ακόλουθα αποτελέσματα :

Link Tokens Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	70%	102 Kb	65% - 100%	$\nearrow, \rightarrow, \rightarrow$
2ο	70%	33 Kb	50% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	50%	16 Kb	45% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	90%	70 Kb	85% - 100%	$\nearrow$
2ο	80%	33 Kb	50% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	50%	16 Kb	50% - 95%	$\nearrow, \rightarrow, \rightarrow$

Με τον περιορισμό της χωρητικότητας της κρυφής μνήμης που πραγματοποιείται στα πειράματα αυτά, αρχίζει να φαίνεται η συμπεριφορά του μηχανισμού σε οριακές περιπτώσεις. Για το μηχανισμό αυτό, λοιπόν, παρουσιάζεται συμπεριφορά πολύ ικανοποιητική παρά το γεγονός ότι η χωρητικότητα της κρυφής μνήμης περιορίζεται πολύ και η βάση περιέχει πολύ μεγάλο αριθμό συνδέσμων μεταξύ ατομικών οντοτήτων της βάσης, οι δομές αναπαράστασης των οποίων αποθηκεύονται στο μηχανισμό αυτό. Εξακολουθεί, εξάλλου, να παρουσιάζει αρκετά καλύτερη επίδοση το σύστημα δελτίων εισαγωγής δεδομένων, κύρια εξαιτίας της μικρής ανάκλησης τέτοιου είδους πληροφορίας που πραγματοποιείαι από τις ερωτήσεις του.

Token Nodes Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	45%	70 Kb	0% - 100%	$\nearrow$
2ο	20%	25 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	10%	12 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	80%	42 Kb	50% - 100%	$\nearrow$
2ο	70%	25 Kb	50% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	60%	12 Kb	20% - 100%	$\nearrow, \rightarrow, \rightarrow$

Συμπεριφορά ανάλογη με αυτή του προηγούμενου μηχανισμού παρουσιάζεται και εδώ. Η επίδοση του συστήματος δελτίων εισαγωγής δεδομένων είναι σαφώς καλύτερη συγκρινόμενη με αυτήν του συστήματος διατύπωσης ερωτήσεων, στοιχείο το οποίο παρατηρήθηκε και στην πρώτη σειρά πειραμάτων. Αιτία του φαινομένου είναι το γεγονός της ανάκλησης από το σύστημα διατύπωσης ερωτήσεων μεγαλύτερης ποσότητας πληροφορίας από τη βάση, η οποία ανήκει στο χαμηλότερο επίπεδο ταξινόμησης, εξαιτίας του είδους των ερωτήσεων που εκτελούνται σ'αυτό. Απόδειξη της μεγαλύτερης απαίτησης σε ποσότητα ανακαλούμενης πληροφορίας αυτού του είδους από το σύστημα διατύπωσης ερωτήσεων αποτελεί το γεγονός ότι στο πρώτο από τη νέα σειρά πειραμάτων, το μέγεθος του μηχανισμού δε φτάνει το 100% της χωρητικότητας, η οποία ανέρχεται σε 90 μόλις Kb.

Object Extensions Queue

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	10%	90 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
2ο	10%	28 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	8%	16.5 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	80%	40 Kb	60% - 100%	$\nearrow, \rightarrow$
2ο	70%	28 Kb	55% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	60%	16.5 Kb	45% - 100%	$\nearrow, \rightarrow, \rightarrow$

Ομοια συμπεριφορά με αυτή των δύο πρώτων μηχανισμών, σε ό,τι αφορά τις απαιτήσεις σε μνήμη για τα δύο συστήματα που μελετήθηκαν.

Στο σύστημα διατύπωσης ερωτήσεων, η επίδοση είναι πολύ κακή, αποτέλεσμα που είναι αναμενόμενο με δεδομένη την κατανάλωση μνήμης στην οποία φτάνει ο μηχανισμός σε προηγούμενα πειράματα. Οι μεγάλες απαιτήσεις σε μνήμη προκαλούν πολλές αποτυχημένες προσπελάσεις στην κρυφή μνήμη, που έχουν σαν αποτέλεσμα το μέσο ποσοστό επιτυχίας να περιοριστεί σε τιμές μέχρι 10%.

Αντίθετα, το σύστημα δελτίων εισαγωγής δεδομένων, εξαιτίας των μικρότερων απαιτήσεων σε μνήμη που παρουσιάζει, εμφανίζει πολύ υψηλά ποσοστά επιτυχίας.

Hash Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	85%	1.05 Mb	0% - 100%	$\nearrow$
2ο	80%	530 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	50%	260 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	80%	1.05 Mb	0% - 100%	$\nearrow, \rightarrow$
2ο	75%	530 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	70%	260 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$

Ανάλογη συμπεριφορά παρατηρείται για το μηχανισμό αυτό και από τα δύο συστήματα που μελετήθηκαν. Η πλήρωση της κρυφής μνήμης σε ποσοστό 100% σε όλα τα πειράματα ήταν αναμενόμενη δεδομένων των απαιτήσεων που εμφανίστηκαν στα προηγούμενα πειράματα.

Τα λίγο χαμηλότερα ποσοστά επιτυχίας για τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων, εξάλλου, αποτελούν ένδειξη μικρότερης τοπικότητας στις αναφορές από το σύστημα, σε σχέση με αυτό του συστήματος δελτίων εισαγωγής δεδομένων.

Symbol Table Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	90%	440 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
2ο	90%	220 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	80%	110 Kb	0% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	100%	300 Kb	$\simeq$ 100%	$\nearrow$
2ο	99%	220 Kb	$\simeq$ 100%	$\nearrow$
3ο	95%	110 Kb	$\simeq$ 100%	$\nearrow, \rightarrow$

Ο πιο απαιτητικός, σε ό,τι αφορά κατανάλωση μνήμης, από τους μηχανισμούς διαχείρισης κρυφής μνήμης του ΣΣΕ, παρουσιάζει πολύ ικανοποιητική συμπεριφορά. Τα ποσοστά επιτυχών προσπελάσεων φανερώνουν μεγάλη τοπικότητα στις αναφορές, η οποία γίνεται φανερή παρά το σημαντικό περιορισμό σε μέγεθος. Οι αναφορές που παρουσιάζουν τη μεγαλύτερη τοπικότητα αφορούν ακολουθίες ερωτήσεων σε ”γειτονικές” οντότητες της βάσης, ενώ τα ποσοστά της τάξης του 0% εμφανίζονται όταν η ”γειτονιά” της βάσης στην οποία αναφερόμαστε αλλάζει.

Overflow Cache

Σύστημα Διατύπωσης Ερωτήσεων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	97%	260 Kb	90% - 100%	$\nearrow, \rightarrow$
2ο	70%	130 Kb	60% - 100%	$\nearrow, \rightarrow, \rightarrow$
3ο	60%	65 Kb	40% - 100%	$\nearrow, \rightarrow, \rightarrow$

Δελτία Εισαγωγής Δεδομένων

	Στοιχεία συμπεριφοράς			
Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριa τιμών hit rate	Συμπεριφορά μεγέθους
1ο	20%	260 Kb	0% - 30%	$\nearrow, \rightarrow$
2ο	20%	130 Kb	0% - 30%	$\nearrow, \rightarrow, \rightarrow$
3ο	20%	65 Kb	0% - 30%	$\nearrow, \rightarrow, \rightarrow$

Συνεχίζοντας τη συμπεριφορά που παρουσίασε και στην προηγούμενη σειρά πειραμάτων, ο μηχανισμός εμφανίζει πολύ καλή επίδοση στο σύστημα διατύπωσης ερωτήσεων και κακή στο σύστημα δελτίων εισαγωγής δεδομένων. Οι μεγαλύτερες ανάγκες σε μνήμη που παρουσιάζει στο δεύτερο από τα παραπάνω συστήματα αποτελεί εν μέρει την αιτία αυτής της συμπεριφοράς. Η σημαντικότερη αιτία, ωστόσο, είναι το γεγονός ότι στο μηχανισμό βοηθητικών εγγραφών αποθηκεύεται πληροφορία χωρίς καμμία τοπικότητα με βάση τη σημασιολογική συσχέτιση των οντοτήτων. Τα ποσοστά επιτυχίας, για το λόγο αυτό, καθορίζονται εντελώς τυχαία.

Named Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
1ο	97%	230 Kb	95% - 100%	↗
2ο	97%	115 Kb	95% - 100%	↗, →
3ο	97%	58 Kb	95% - 100%	↗, →

Δελτία Εισαγωγής Δεδομένων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
1ο	80%	230 Kb	40% - 100%	↗
2ο	80%	115 Kb	40% - 100%	↗, →
3ο	80%	58 Kb	40% - 100%	↗, →

Η πολύ καλή επίδοση του μηχανισμού συνεχίζεται και σ'αυτή τη σειρά πειραμάτων και για τα δύο συστήματα. Η τοπικότητα των αναφορών είναι η αιτία τόσο καλών αποτελεσμάτων ακόμη και όταν οι περιορισμοί στην κατανάλωση μνήμης είναι μεγάλοι.

Unnamed Sysid Cache

Σύστημα Διατύπωσης Ερωτήσεων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριο τιμών hit rate	Συμπεριφορά μεγέθους
1ο	90%	230 Kb	75% - 100%	↗
2ο	80%	120 Kb	70% - 100%	↗, →
3ο	80%	58 Kb	65% - 100%	↗, →, →

Δελτία Εισαγωγής Δεδομένων

Στοιχεία συμπεριφοράς

Πείραμα	Μέσο hit rate	Μέγιστο μέγεθος	Οριακές τιμές hit rate	Συμπεριφορά μεγέθους
1ο	40%	100 Kb	0% - 100%	↗
2ο	40%	100 Kb	0% - 100%	↗
3ο	40%	58 Kb	0% - 100%	↗, →

Μεγαλύτερες απαιτήσεις σε μνήμη παρουσιάζονται για το μηχανισμό αυτό από το σύστημα διατύπωσης ερωτήσεων. Σε αντίθεση με τις ανάγκες σε μνήμη, ωστόσο, τα ποσοστά επιτυχών προσπελάσεων είναι χειρότερα στο σύστημα δελτίων εισαγωγής δεδομένων, στοιχείο που επηρεάζεται αποκλειστικά από την οργάνωση της πληροφορίας στη βάση, σε σχέση με τις προσπελάσεις που πραγματοποιούνται από τα δύο συστήματα.

Μετρήσεις για οντότητες κι επεκτάσεις οντοτήτων στη μνήμη

Τα αποτελέσματα αυτής της κατηγορίας των μετρήσεων έχουν ως εξής :

Σύστημα Διατύπωσης Ερωτήσεων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
1ο Πείραμα	45 Kb	1.3 Mb
2ο Πείραμα	26 Kb	1.3 Mb
3ο Πείραμα	12 Kb	1.1 Mb

Δελτία Εισαγωγής Δεδομένων

Κατανάλωση μνήμης

Πείραμα	Οντότητες	Επεκτάσεις οντοτήτων
1ο Πείραμα	14 Kb	1.1 Mb
2ο Πείραμα	14 Kb	1.1 Mb
3ο Πείραμα	14 Kb	1.1 Mb

Περιορίζεται στη νέα σειρά πειραμάτων η κατανάλωση μνήμης από οντότητες στη βάση σε σχέση με τα προηγούμενα πειράματα για την ίδια βάση δεδομένων στο σύστημα διατύπωσης ερωτήσεων. Ο περιορισμός των χωρητικοτήτων των μηχανισμών κρυφής μνήμης, επομένως, επηρεάζει περισσότερο τις οντότητες και λιγότερο τις επεκτάσεις οντοτήτων που αποθηκεύονται στη μνήμη.

Τα μεγέθη μνήμης, εξάλλου, παραμένουν έντονα διαφοροποιημένα - μεγάλο πλήθος επεκτάσεων και πολύ μικρό πλήθος οντοτήτων - εξαιτίας της κατασκευής της βάσης καθώς και του σεναρίου χρήσης.

Συνολική συμπεριφορά του συστήματος

Στα νέα πειράματα που πραγματοποιήθηκαν, οι συνολικές επιδόσεις των δύο συστημάτων που μελετήθηκαν, δηλαδή το μέσο ποσοστό επιτυχών προσπελάσεων στην κρυφή μνήμη, καθώς επίσης και η συνολική κατανάλωση μνήμης από τα συστήματα παρουσιάζονται στη συνέχεια.

Σε ό,τι αφορά τα ποσοστά επιτυχών προσπελάσεων στην κρυφή μνήμη, αυτά διαμορφώνονται ως εξής στα τρία πειράματα :

- 1ο Πείραμα : Οι δοσοληψίες του συστήματος διατύπωσης ερωτήσεων, παρουσιάζουν ποσοστά επιτυχών προσπελάσεων στην κρυφή μνήμη τα οποία παραμένουν σχεδόν μόνιμα στο 100%. Απόκλιση από την τιμή αυτή, παρουσιάζουν δοσοληψίες που αντιστοιχούν σε βήματα του σεναρίου χρήσης του συστήματος στα οποία πραγματοποιείται αλλαγή της οντότητας με βάστη την οποία τίθενται οι ερωτήσεις στη βάση. Στις δοσοληψίες αυτές, το ποσοστό μειώνεται στο 80% περίπου.

Στο σύστημα διαλογικών δελτίων ενημέρωσης, εξάλλου, τα ποσοστά παρουσιάζονται σχετικά μειωμένα. Ετσι, αρκετές δοσοληψίες παρουσιάζουν τώρα ποσοστά επιτυχών προσπελάσεων που κυμαίνονται στο 90%, ενώ σε κάποιες άλλες το ποσοστό μειώνεται στο 30%. Οι τελευταίες αυτές δοσοληψίες αντιστοιχούν σε βήματα ενημέρωσης της βάσης μέσω των διαλογικών δελτίων ενημέρωσης.

Συνολικά, τα ποσοστά επιτυχών προσπελάσεων κυμαίνονται κατά μέσο όρο στο 99% για τις δοσοληψίες του συστήματος διατύπωσης ερωτήσεων και στο 80% για τις δοσοληψίες του συστήματος διαλογικών δελτίων ενημέρωσης, στοιχείο που αποτελεί πολύ ικανοποιητική επίδοση.

- 2ο Πείραμα : Στο πείραμα αυτό, παρουσιάζεται μείωση των ποσοστών επιτυχών προσπελάσεων και για τα δύο συστήματα.

  Ετσι, στο σύστημα διατύπωσης ερωτήσεων, οι δοσοληψίες οι οποίες στο προηγούμενο πείραμα παρουσίαζαν ποσοστά επιτυχών προσπελάσεων 100%, τώρα κυμαίνονται μεταξύ 95% και 99%. Στο σύστημα διαλογικών δελτίων ενημέρωσης, εξάλλου, αυξάνεται ο αριθμός των δοσοληψιών που παρουσιάζουν ποσοστό επιτυχών προσπελάσεων στην κρυφή μνήμη γύρων στο 90%, με αποτέλεσμα το μέσο ποσοστό επιτυχίας να βρίσκεται τώρα στο 70-75%.

- 3ο Πείραμα : Ομοια, με το προηγούμενο πείραμα, παρουσιάζεται κι εδώ μια μικρή επιδείνωση της συμπεριφοράς του συστήματος. Οι δοσοληψίες του συστήματος διατύπωσης ερωτήσεων που στο πρώτο πείραμα έφταναν το 100% σε επιτυχείς προσπελάσεις της κρυφής μνήμης, περιορίζονται τώρα σε τιμές που κυμαίνονται από 90 έως 99%. Σε ό,τι αφορά τις δοσοληψίες του συστήματος διαλογικών δελτίων ενημέρωσης, αυξάνεται ο αριθμός αυτών που περιορίζονται σε ποσοστά επιτυχίας της τάξης του 30%, και μειώνεται ο αριθμός αυτών που ανέρχονται σε ποσοστά της τάξης του 100%, με αποτέλεσμα το μέσο ποσοστό επιτυχίας να περιορίζεται στην τιμή του 65%, περίπου.

Σε ότι αφορά εξάλλου τα μεγέθη μνήμης που καταλαμβάνονται από τα δύο συστήματα κατά τη διεξαγωγή των νέων αυτών πειραμάτων, τα αποτελέσματα έχουν ως εξής :

- Σύστημα Επαφής Χρήσης : Τα μεγέθη μνήμης που καταλαμβάνονται από το σύστημα διατύπωσης ερωτήσεων για τα τρία πειράματα που πραγματοποιήθηκαν είναι τα ακόλουθα :

1ο Πείραμα : 5.3 Mb
2ο Πείραμα : 4.3 Mb
3ο Πείραμα : 3.4 Mb

Οπως είναι φανερό από τα παραπάνω αποτελέσματα, τα μεγέθη μνήμης τώρα περιορίζονται σημαντικά, χωρίς ωστόσο όπως φάνηκε παραπάνω να επηρεάζεται σημαντικά η συνολική επίδοση του συστήματος.

- Δελτία εισαγωγής δεδομένων : Σε ό,τι αφορά τις δοσοληψίες του συστήματος διαλογικών δελτίων ενημέρωσης της βάσης, τα μεγέθη μνήμης που καταλαμβάνονται είναι τα ακόλουθα :

1ο Πείραμα : 6 Mb
2ο Πείραμα : 4.6 Mb
3ο Πείραμα : 4.5 Mb

Παρά το γεγονός ότι τα μεγέθη είναι εδώ λίγο μεγαλύτερα σε σχέση με το πρώτο σύστημα, παραμένουν ικανοποιητικά χαμηλά κυρίως όταν μελετηθούν σε σχέση με την επίδοση του συστήματος, η οποία είναι και εδώ πολύ ικανοποιητική.

7.3.2 Συμπεράσματα

Τα αποτελέσματα της δεύτερης σειράς πειραμάτων για τη βάση του ULAN έχουν συνοπτικά ως εξής :

Μεγέθη μηχανισμών κρυφής μνήμης

	Χωρητικότητα			**Χωρητικότητα**		
Κρυφή μνήμη	1ο	2ο	3ο	1ο	2ο	3ο
Link Tokens Queue	102 Kb	33 Kb	16 Kb	70 Kb	33 Kb	16 Kb
Token Nodes Queue	70 Kb	25 Kb	12 Kb	42 Kb	25 Kb	12 Kb
Object Extensions Queue	90 Kb	28 Kb	16.5 Kb	40 Kb	28 Kb	16.5 Kb
Hash Cache	1.05 Mb	530 Kb	260 Kb	1.05 Mb	530 Kb	260 Kb
Symbol Table Cache	440 Kb	220 Kb	110 Kb	300 Kb	220 Kb	110 Kb
Overflow Cache	260 Kb	130 Kb	65 Kb	260 Kb	130 Kb	65 Kb
Named Sysid Cache	230 Kb	115 Kb	58 Kb	230 Kb	115 Kb	58 Kb
Unnamed Sysid Cache	230 Kb	120 Kb	58 Kb	100 Kb	100 Kb	58 Kb

Ποσοστά επιτυχίας μηχανισμών κρυφής μνήμης

	Χωρητικότητα			**Χωρητικότητα**		
Κρυφή μνήμη	1ο	2ο	3ο	1ο	2ο	3ο
Link Tokens Queue	70%	70%	50%	90%	80%	50%
Token Nodes Queue	45%	20%	10%	80%	70%	60%
Object Extensions Queue	10%	10%	8%	80%	70%	60%
Hash Cache	85%	80%	50%	80%	75%	70%
Symbol Table Cache	90%	90%	80%	100%	99%	95%
Overflow Cache	97%	70%	60%	20%	20%	20%
Named Sysid Cache	97%	97%	97%	80%	80%	80%
Unnamed Sysid Cache	90%	80%	80%	40%	40%	40%

Καθώς όλοι μηχανισμοί κρυφής μνήμης παρουσιάζουν τώρα μεγέθη τα οποία καλύπτουν τη χωρητικότητά τους, το ενδιαφέρον στρέφεται στην επίδοση των επιμέρους μηχανισμών. Παρατηρείται εδώ διακύμανση της επίδοσης για διαφορετικούς μηχανισμούς για τα δύο συστήματα.

Πιο συγκεκριμένα, εμφανίζονται μηχανισμοί οι οποίοι παρά τον περιορισμό της χωρητικότητάς τους εμφανίζουν πολύ καλή συμπεριφορά - κρυφή μνήμη για την αποθήκευση των συσχετίσεων μεταξύ ατομικών οντοτήτων στη βάση, κρυφή μνήμη του υποκαταλόγου λογικών ονομάτων και κατάλογος συστήματος για οντότητες με δοσμένο λογικό όνομα - και άλλοι των οποίων η συμπεριφορά επιδεινώνεται δραματικά - κρυφές μνήμες για την αποθήκευση ατομικών οντοτήτων και επεκτάσεων οντοτήτων για το σύστημα διατύπωσης ερωτήσεων καθώς επίσης και η κρυφή μνήμη για τις βοηθητικές εγγραφές του υποκαταλόγου λογικών ονομάτων για το σύστημα δελτίων εισαγωγής δεδομένων.

Το φαινόμενο αυτό, δίνει τη δυνατότητα να περιορίσουμε, χωρίς σημαντικό κόστος, τις χωρητικότητες κάποιων από τους μηχανσιμούς κρυφής μνήμης γι'αυτό από τα δύο συστήματα στο οποίο η επίδοσή του δεν παρουσιάζει επιδείνωση. Η δυνατότητα αυτή αναλύεται διεξοδικά στο επόμενο κεφάλαιο.

Κεφάλαιο 8

Συμπεράσματα

8.1 Εισαγωγή

Οι μετρήσεις επίδοσης που πραγματοποιήθηκαν στο Σύστημα Σημασιολογικού Ευρετηριασμού για τις τρεις βάσεις που προαναφέρθηκαν, τα αποτελέσματα των οποίων περιγράφηκαν στα προηγούμενα κεφάλαια αποτελούν σημαντικά στοιχεία μελέτης και ερμηνείας, τόσο για το σύστημα διατύπωσης ερωτήσεων, όσο και για το σύστημα διαλογικών δελτίων ενημέρωσης.

Η πληθώρα των αποτελεσμάτων που υπήρξαν, ωστόσο, καθιστά αποτελεσματική τη χρησιμοποίησή τους μία και μόνη φορά, με σκοπό τον εντοπισμό σημείων καθυστέρησης στα δύο συστήματα, καθώς επίσης και χαρακτηριστικών επίδοσης με βάση διαφορετικού είδους προσπελάσεις στις βάσεις δεδομένων - διαφορετικές ερωτήσεις κλπ.

Οι μετρήσεις αυτές, ωστόσο, είναι δυνατό να χρησιμοποιηθούν και για πιο πρακτικούς σκοπούς, με κατάλληλη διαμόρφωση των αποτελεσμάτων τους. Ενας από τους σκοπούς αυτούς μπορεί να είναι ο υπολογισμός των συνολικών απαιτήσεων σε κατανάλωση μνήμης οποιουδήποτε από τα δύο συστήματα που μελετήθηκαν, με βάση την επίδοση σε ποσοστά επιτυχών προσπελάσεων στην κρυφή μνήμη που επιλέγουμε να παρουσιάζει το σύστημα. Πιο συγκεκριμένα, οι μετρήσεις που πραγματοποιήθηκαν, δίνουν στα αποτελέσματά τους τα μέσα ποσοστά επιτυχών προσπελάσεων για καθέναν από τους επιμέρους μηχανισμούς κρυφής μνήμης, για διαφορετικές χωρητικότητες γι'αυτούς και για τα δύο διαφορετικά συστήματα. Διαθέτοντας αυτά τα στοιχεία, μπορούμε να παραστήσουμε γραφικά τη μεταβολή του μέσου ποσοστού επιτυχημένων προσπελάσεων σε καθεμιά από τις κρυφές μνήμες για διαφορετικές τιμές της χωρητικότητάς της. Η

γραφική παράσταση αυτή, δίνει τη δυνατότητα, αφενός συνολικής εποπτείας της συμπεριφοράς κάθε μηχανισμού κρυφής μνήμης και αφετέρου εύρεσης της χωρητικότητας για κάθε κρυφή μνήμη, με την οποία ο μηχανισμός διαθέτει επιλεγμένο ποσοστό επιτυχών προσπελάσεων.

Ενας δεύτερος υπολογισμός που μπορεί να πραγματοποιηθεί με βάση τις μετρήσεις είναι αυτός του χρόνου απόκρισης καθενός από τα δύο συστήματα. Αυτό υπολογίζεται λαμβάνοντας υπόψη ότι ο χρόνος απόκρισης κάθε συστήματος εξαρτάται άμεσα από το μέσο ποσοστό επιτυχών προσπελάσεων στην κρυφή του μνήμη.

Στη συνέχεια παρουσιάζονται περισσότερες λεπτομέρειες σχετικά με καθένα από τα δύο ζητήματα που θίχτηκαν παραπάνω.

8.2 Κατανάλωση μνήμης από το σύστημα

Το πρώτο από τα δύο ζητήματα που αναφέρθηκαν παραπάνω αφορά σε υπολογισμό της συνολικής κατανάλωσης μνήμης από καθένα από τα δύο συστήματα που μελετήθηκαν, με τη βοήθεια γραφικών παραστάσεων των ποσοστών επιτυχών προσπελάσεων στις κρυφές μνήμες συναρτήσει της αντίστοιχης παραμέτρου μεγέθους για κάθε μηχανισμό. Από τις γραφικές αυτές παραστάσεις, οι οποίες θα παρουσιαστούν στη συνέχεια, μπορεί κανείς να εντοπίσει τα επιθυμητά ποσοστά επιτυχών προσπελάσεων για καθέναν από τους επιμέρους μηχανισμούς κρυφής μνήμης και για καθένα από τα δύο συστήματα που μελετήθηκαν, και να προσδιορίσει στη συνέχεια την παράμετρο μεγέθους της κρυφής μνήμης στην οποία εμφανίζεται η επιθυμητή συμπεριφορά.

Επιλέγοντας τις παραμέτρους μεγέθους π_i, i=1,..,10, για καθέναν από τους επιμέρους μηχανισμούς, η συνολική κατανάλωση μνήμης για οποιοδήποτε από τα δύο συστήματα εκφράζεται από τη σχέση :

$$\text{Συνολική κατανάλωση μνήμης} \;=\; \sum_{i=1}^{10}(\pi_i * \sigma_i) \;+\; \Sigma$$

όπου ονομάζουμε :

π_i, : όπως προαναφέρθηκε, τις παραμέτρους μεγέθους για καθέναν από τους συνολικά 10 επιμέρους μηχανισμούς διαχείρισης κρυφής μνήμης,

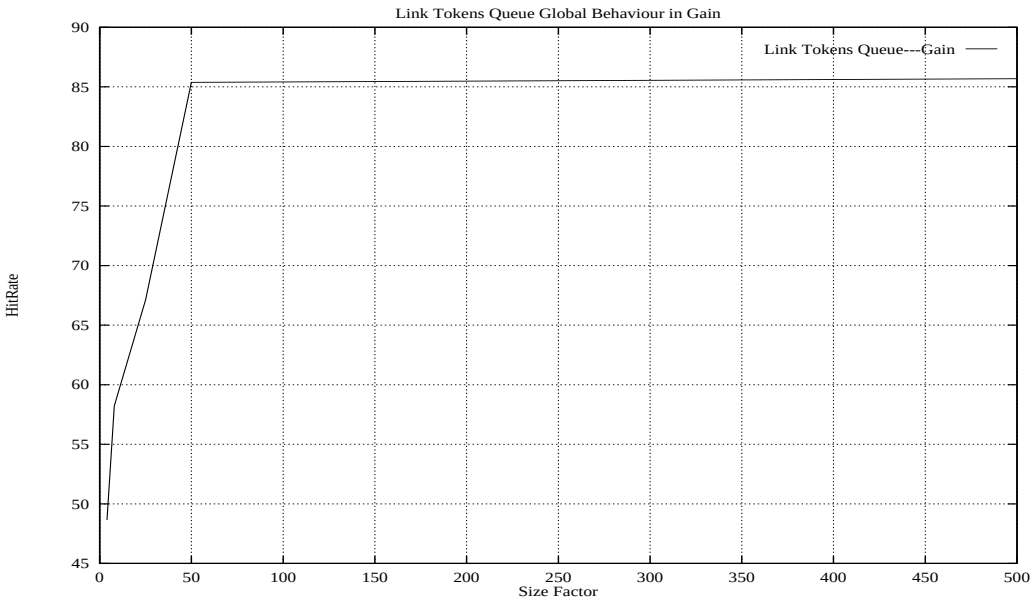
σ_i : τους σταθερούς για κάθε μηχανισμό κρυφής μνήμης συντελεστές μεγέθους και

Σ : μια σταθερή ποσότητα μεγέθους, στην οποία περιλαμβάνεται κατανάλωση μνήμης για σύνολα προσωρινής αποθήκευσης δεδομένων κι ενδιάμεσων αποτελεσμάτων καθώς επίσης και για οντότητες κι επεκτάσεις οντοτήτων που αντιγράφονται στη βάση.

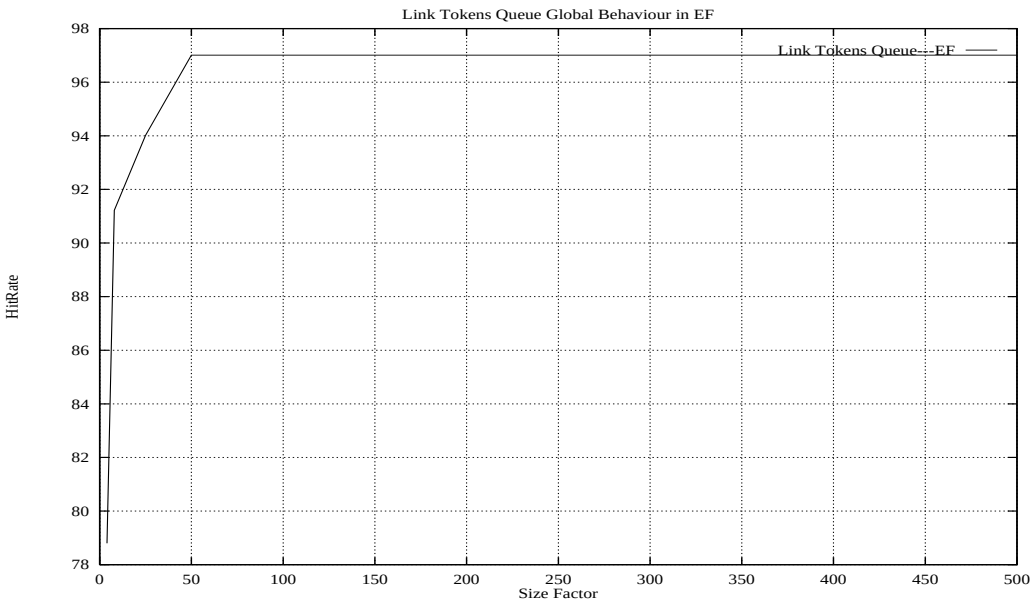
Οι συντελεστές μεγέθους για τους μηχανισμούς κρυφής μνήμης στο Σύστημα Σημασιολογικού Ευρετηριασμού είναι οι εξής :

Class Nodes Queue	:	sizeof(DISK_BLOCK)=4096 bytes
Link Classes Queue	:	sizeof(DISK_BLOCK)=4096 bytes
Link Tokens Queue	:	sizeof(DISK_BLOCK)=4096 bytes
Token Nodes Queue	:	sizeof(DISK_BLOCK)=4096 bytes
Object Extensions Queue	:	sizeof(DISK_BLOCK)=4096 bytes
Hash Cache	:	32∗sizeof(HASH_SEGMENT)=32∗4144 bytes
Symbol Table Cache	:	8∗sizeof(TUPLE_TABLE)=8∗6932 bytes
Overflow Cache	:	8∗sizeof(BUCKET_O_TABLE)=8∗4112 bytes
Named Sysid Cache	:	8∗sizeof(SYSID_TABLE)=8∗7232 bytes
Unnamed Sysid Cache	:	8∗sizeof(SYSID_TABLE)=8∗7232 bytes

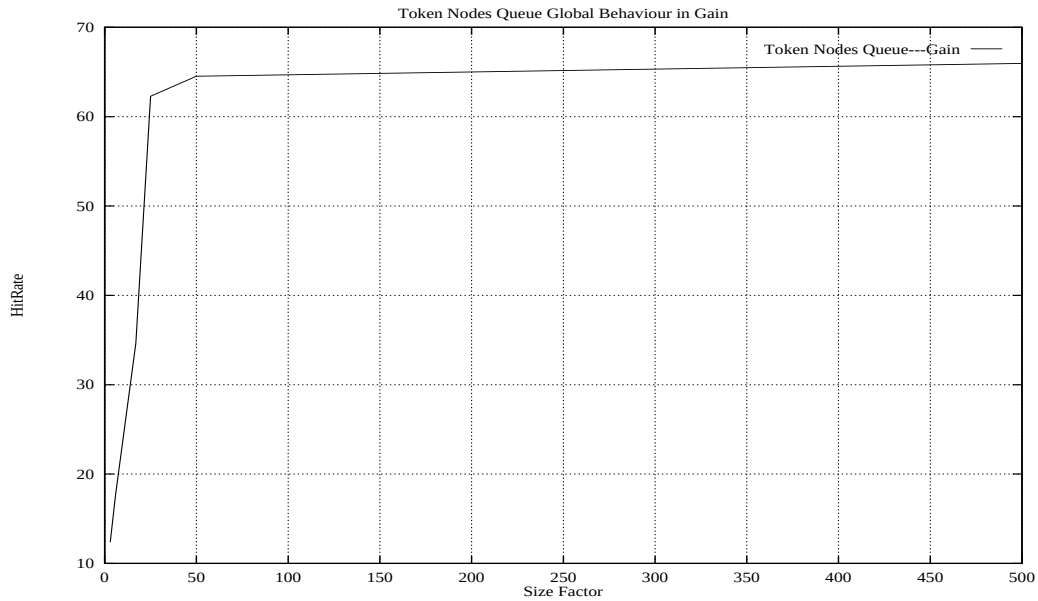
Ο λόγος για τον οποίο το παραπάνω αθροιστικό μοντέλο παριστάνει τη συνολική κατανάλωση μνήμης για καθένα από τα δύο συστήματα που μελετήθηκαν είναι επειδή οι επιμέρους μηχανισμοί κρυφής μνήμης είναι ανά δύο ανεξάρτητοι. Αυτό συμβαίνει, διότι καθένας από τους επιμέρους μηχανισμούς αποθηκεύει διαφορετικό είδος σελίδων πληροφορίας της βάσης δεδομένων. Ο αριθμός των σελίδων κάθε είδους που αποθηκεύονται σε καθεμιά από τις κρυφές μνήμες εξαρτάται αποκλειστικά και μόνο από τις ερωτήσεις που εκτελούνται και όχι από τη χωρητικότητα ή το ποσοστό πληρότητας οποιασδήποτε άλλης μνήμης. Εφόσον, λοιπόν, κανένας μηχανισμός κρυφής μνήμης δεν αποθηκεύει σελίδες ίδιου είδους με οποιονδήποτε άλλον, τα μεγέθη των μνημών είναι μεταξύ τους ανεξάρτητα και η συνολική κατανάλωση μνήμης υπολογίζεται προσθετικά. Οι γραφικές παραστάσεις των ποσοστών επιτυχημένων προσπελάσεων σε συνάρτηση με τις παραμέτρους μεγέθους για τους διαφορετικούς μηχανισμούς κρυφής μνήμης είναι οι ακόλουθες :



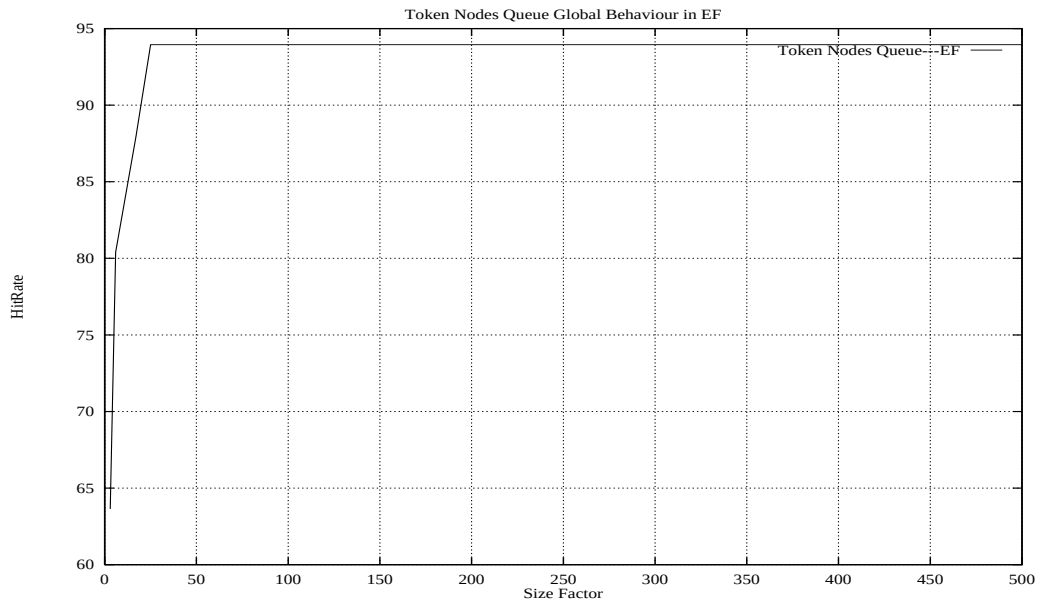
Σχήμα 8.1: Επίδοση της κρυφής μνήμης συσχετίσεων ανάμεσα σε ατομικές οντότητες για το σύστημα διατύπωσης ερωτήσεων



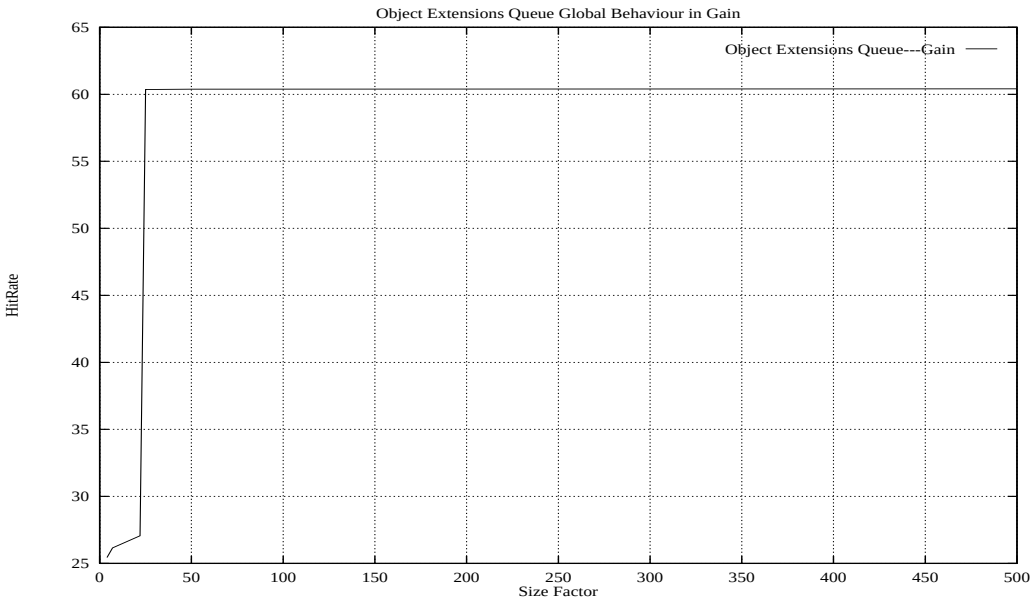
Σχήμα 8.2: Επίδοση της κρυφής μνήμης συσχετίσεων ανάμεσα σε ατομικές οντότητες για το σύστημα διαλογικών δελτίων ενημέρωσης



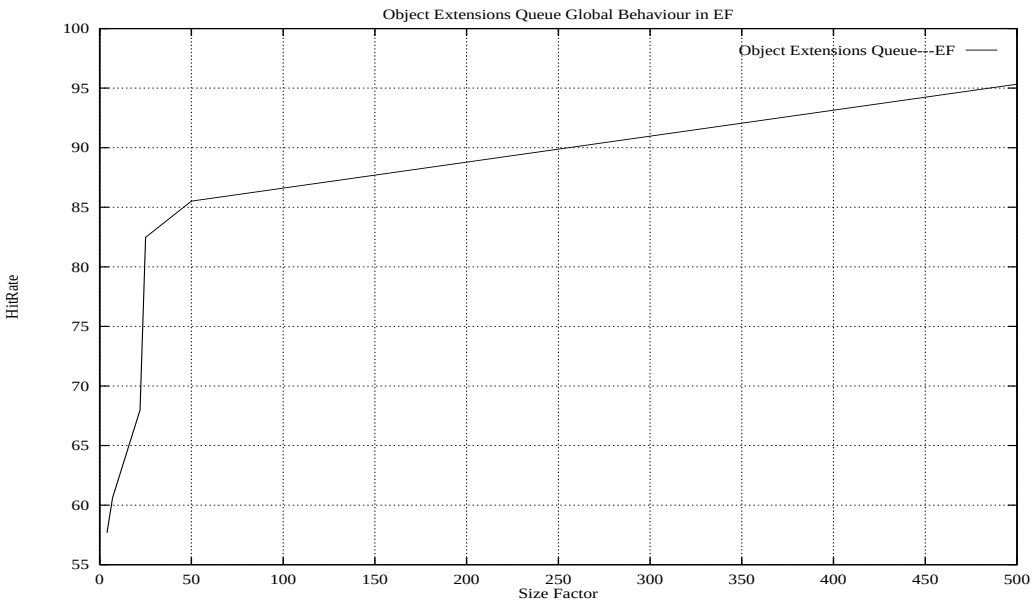
Σχήμα 8.3: Επίδοση της κρυφής μνήμης για τις ατομικές οντότητες της βάσης για το σύστημα διατύπωσης ερωτήσεων



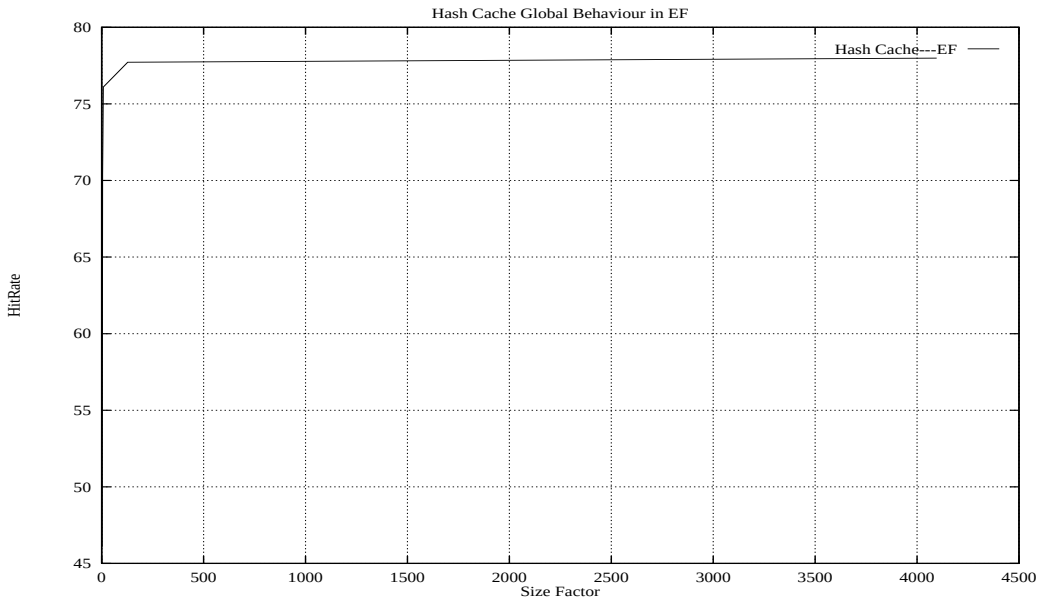
Σχήμα 8.4: Επίδοση της κρυφής μνήμης για τις ατομικές οντότητες της βάσης για το σύστημα διαλογικών δελτίων ενημέρωσης



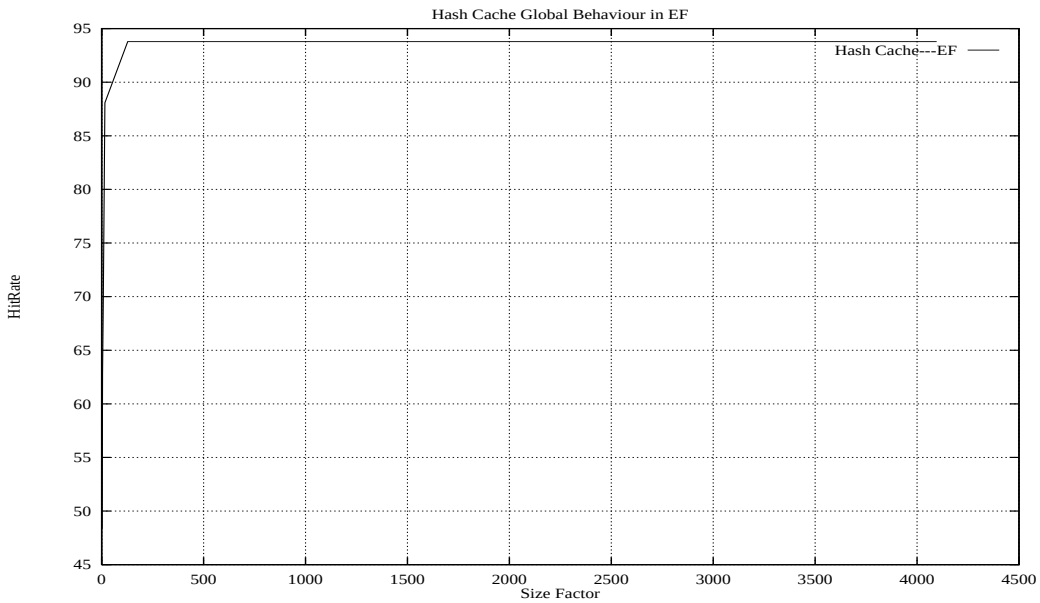
Σχήμα 8.5: Επίδοση της κρυφής μνήμης των δομών επέκτασης των οντοτήτων της βάσης για το σύστημα διατύπωσης ερωτήσεων



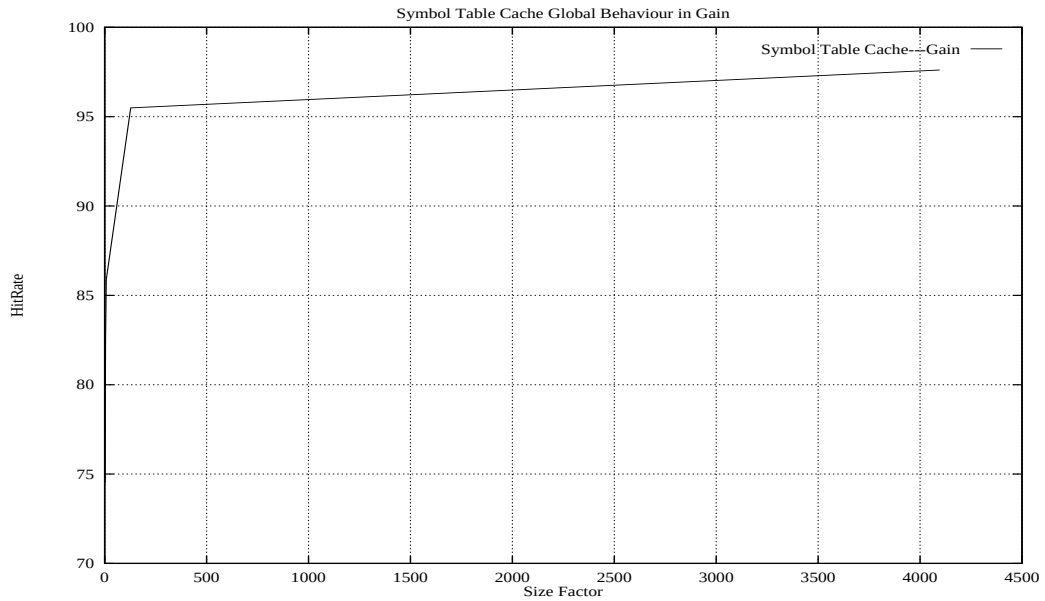
Σχήμα 8.6: Επίδοση της κρυφής μνήμης των δομών επέκτασης των οντοτήτων της βάσης για το σύστημα διαλογικών δελτίων ενημέρωσης



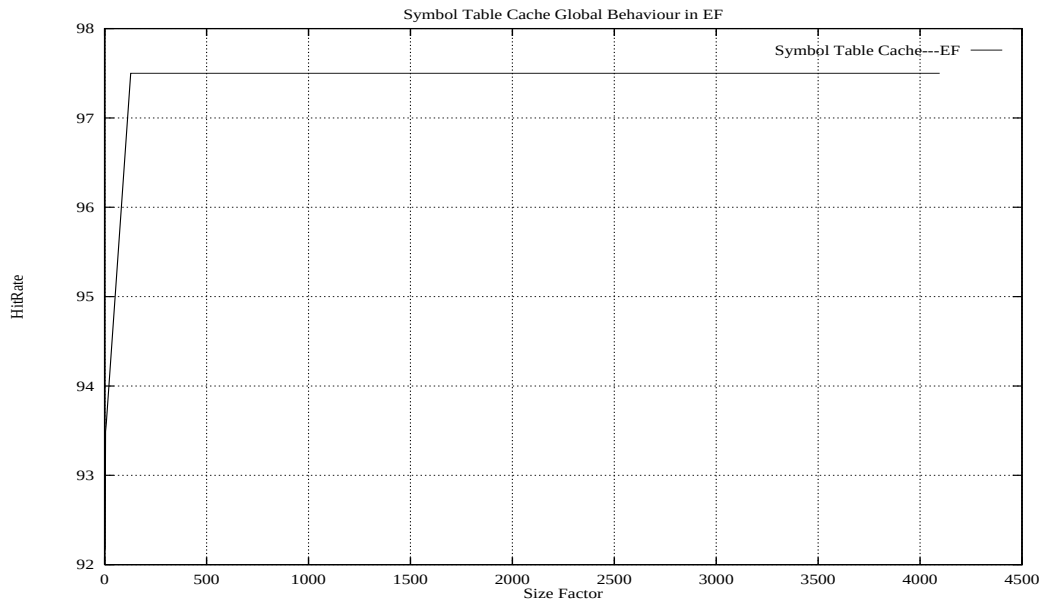
Σχήμα 8.7: Επίδοση της κρυφής μνήμης για τον υποκατάλογο λογικών ονομάτων για το σύστημα διατύπωσης ερωτήσεων



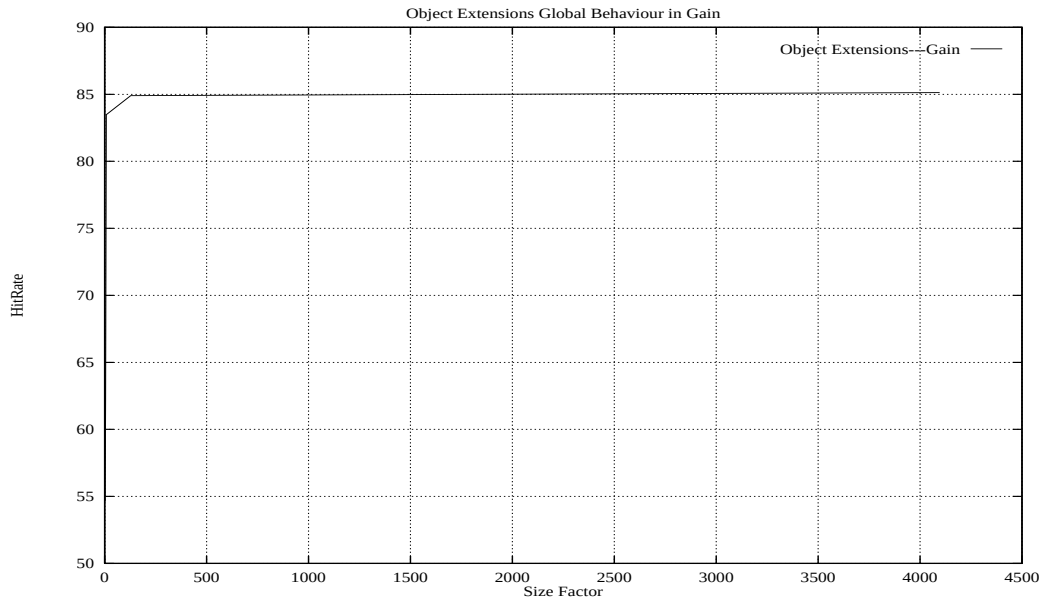
Σχήμα 8.8: Επίδοση της κρυφής μνήμης για τον υποκατάλογο λογικών ονομάτων για το σύστημα διαλογικών δελτίων ενημέρωσης



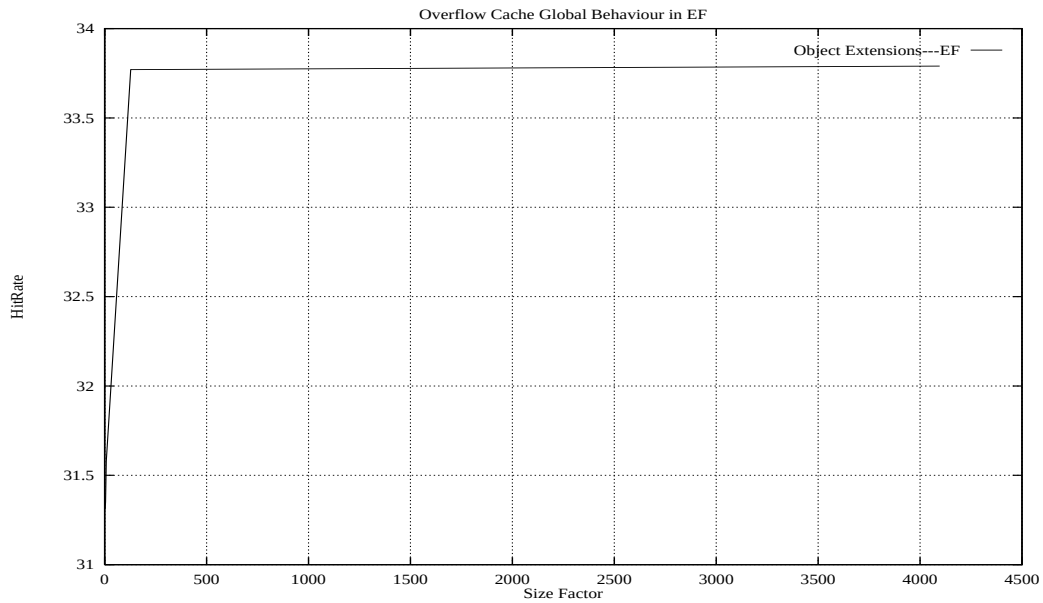
Σχήμα 8.9: Επίδοση της κρυφής μνήμης για τον υποκατάλογοαναγνωριστικών συστήματος για το σύστημα διατύπωσης ερωτήσεων



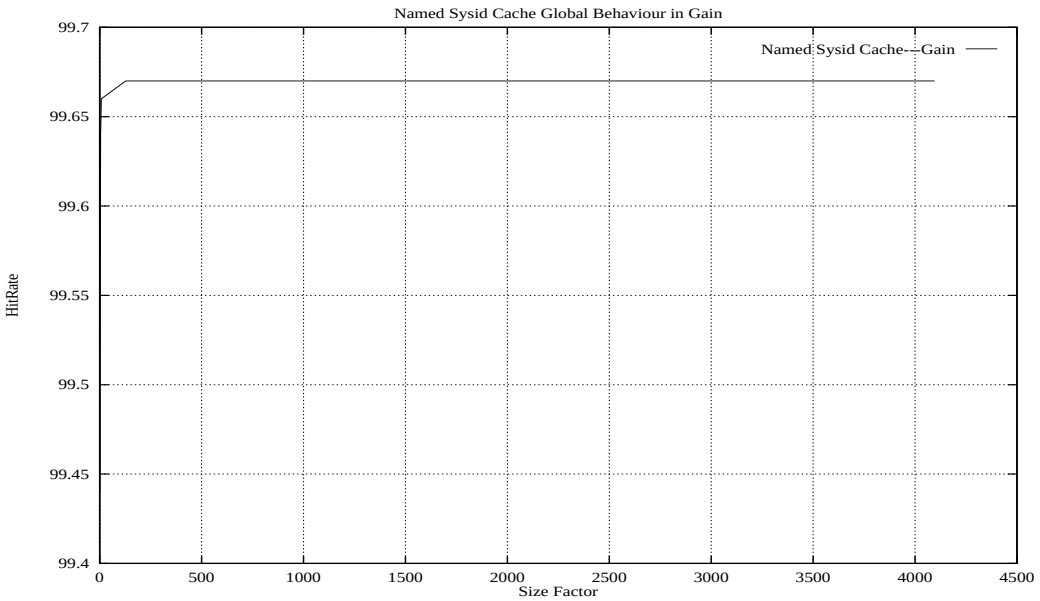
Σχήμα 8.10: Επίδοση της κρυφής μνήμης για τον υποκατάλογο αναγνωριστικών συστήματος για το σύστημα διαλογικών δελτίων ενημέρωσης



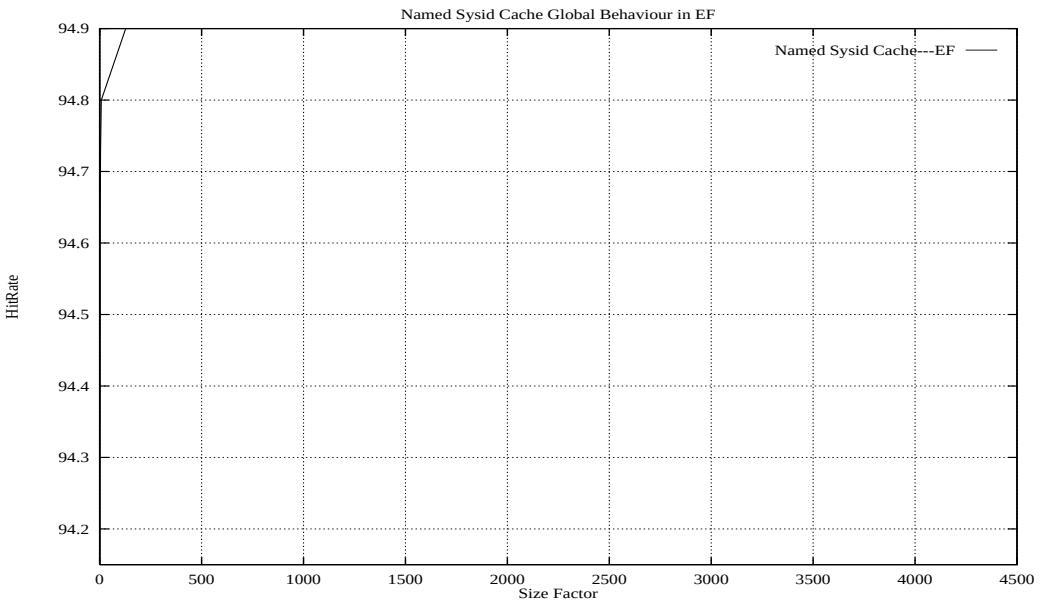
Σχήμα 8.11: Επίδοση της κρυφής μνήμης για τον κατάλογο βοηθητικών εγγραφών για το σύστημα διατύπωσης ερωτήσεων



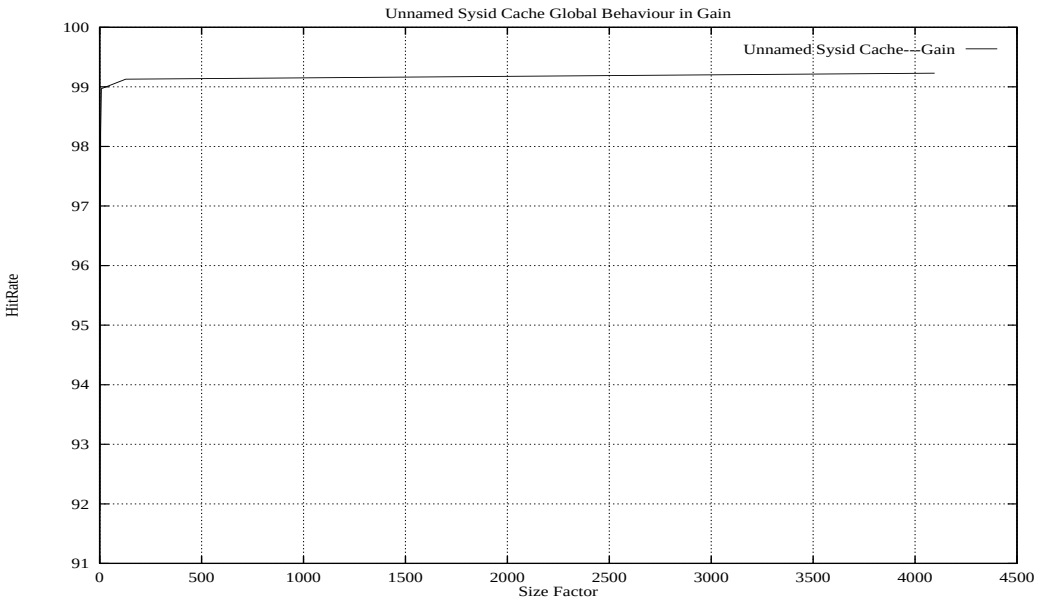
Σχήμα 8.12: Επίδοση της κρυφής μνήμης για τον κατάλογο βοηθητικών εγγραφών για το σύστημα διαλογικών δελτίων ενημέρωσης



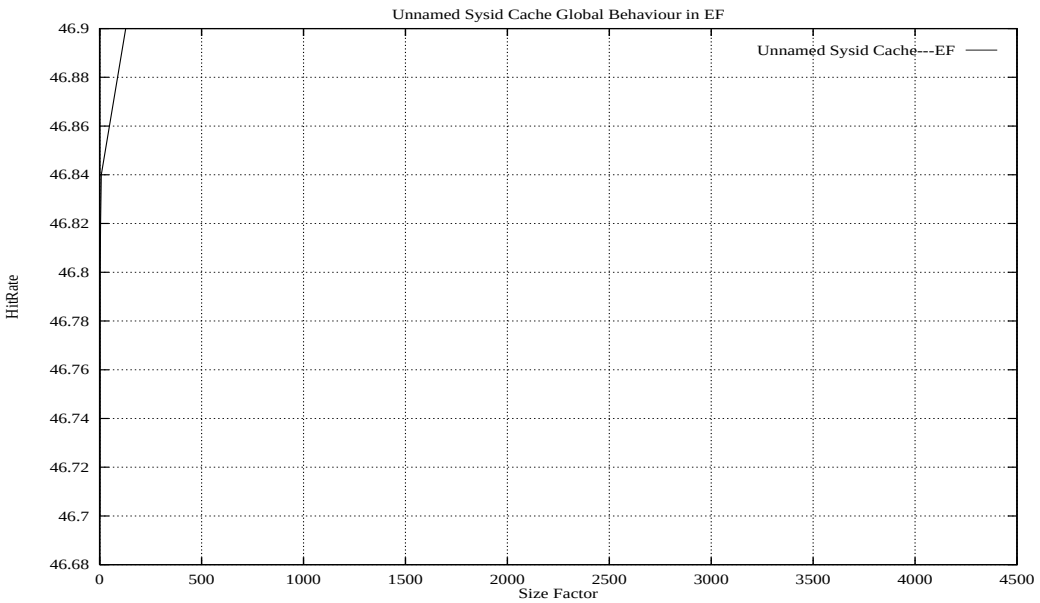
Σχήμα 8.13: Επίδοση της κρυφής μνήμης του καταλόγου συστήματος για οντότητες με δοσμένο λογικό όνομα για το σύστημα διατύπωσης ερωτήσεων



Σχήμα 8.14: Επίδοση της κρυφής μνήμης του καταλόγου συστήματος για οντότητες με δοσμένο λογικό όνομα για το σύστημα διαλογικών δελτίων ενημέρωσης



Σχήμα 8.15: Επίδοση της κρυφής μνήμης του καταλόγου συστήματος για οντότητες χωρίς δοσμένο όνομα για το σύστημα διατύπωσης ερωτήσεων



Σχήμα 8.16: Επίδοση της κρυφής μνήμης του καταλόγου συστήματος για οντότητες χωρίς δοσμένο όνομα για το σύστημα διαλογικών δελτίων ενημέρωσης

Από τις παραστάσεις αυτές είναι δυνατό να επιλέξει κανείς για το Σύστημα Σημασιολογικού Ευρετηριασμού τις παραμέτρους μεγέθους εκείνες οι οποίες δίνουν για καθέναν από τους μηχανισμούς κρυφής μνήμης ικανοποιητική συμπεριφορά και με βάση αυτές τις παραμέτρους να υπολογίσει τη συνολική κατανάλωση μνήμης από το σύστημα για κάποια εφαρμογή.

Στο σημείο αυτό πρέπει να σημειωθεί ότι οι παραπάνω παραστάσεις παρουσιάζουν τη συμπεριφορά του συστήματος μόνο για βάσεις δεδομένων με την ίδια αναλογία σχήματος και δεδομένων όπως η βάση του ULAN από την οποία προέκυψαν. Η ίδια εργασία μπορεί να πραγματοποιηθεί και για άλλες κατηγορίες βάσεων δημιουργώντας έτσι έναν πλήρη οδηγό για την επιλογή των μεγεθών των κρυφών μνημών για διαφορετικές εφαρμογές.

Ενας ακόμη τρόπος με τον οποίο είναι δυνατό να χρησιμοποιηθούν οι παραπάνω γραφικές παραστάσεις είναι προκειμένου να εξισορροπιθούν οι ανταγωνιστικοί πόροι της μνήμης και του χρόνου στο σύστημα. Με άλλα λόγια, με τη βοήθεια αυτών των παραστάσεων είναι δυνατό να επιλεχθούν οι παράμετροι μεγέθους των επιμέρους μηχανισμών κρυφής μνήμης έτσι, ώστε η συνολική κατανάλωση μνήμης να μην ξεπερνά ένα επιλεγμένο όριο. Καθώς όμως η συνολική κατανάλωση μνήμης προκύπτει σαν άθροισμα δέκα επιμέρους μεγεθών, είναι δυνατό να υπάρξουν ζεύγη αυξήσεων-μειώσεων τιμών χωρητικοτήτων με στόχο τη βελτίωση της επίδοσης μηχανισμών κρυφής μνήμης οι οποίοι έχουν μεγαλύτερες απαιτήσεις σε μνήμη, όταν η ελάττωση της χωρητικότητας δεν έχει σημαντικό αντίκτυπο στην επίδοση ενός άλλου μηχανισμού.

Εκφράζοντας, επομένως, την κατανάλωση μνήμης από καθέναν από τους επιμέρους μηχανισμούς κρυφής μνήμης μέσω των σχέσεων :

$\pi_1 * \sigma_1 \leq \kappa_1$, όπου κ_1 ονομάζουμε τη συνολική κατανάλωση μνήμης από τον πρώτο μηχανισμό,

$\pi_2 * \sigma_2 \leq \kappa_2$, όπου κ_2 ονομάζουμε τη συνολική κατανάλωση μνήμης από το δεύτερο μηχανισμό,

$\vdots$

$\pi_{10} * \sigma_{10} \leq \kappa_{10}$, όπου κ_{10} ονομάζουμε τη συνολική κατανάλωση μνήμης από το δέκατο μηχανισμό,

είναι δυνατό να διαμορφώσουμε τις τιμές των $\kappa_i,\ i = 1, .., 10$ ώστε να ισχύει

$$\sum_{i=1}^{10} \kappa_i + \Sigma \leq \Phi,$$

όπου Φ ονομάζουμε ένα επιθυμητό άνω φράγμα στη συνολική κατανάλωση μνήμης από το σύστημα, και να επιδεικνύουν όλοι οι επιμέρους μηχανισμοί συμπεριφορά ικανοποιητική.

8.3 Χρόνος απόκρισης συστήματος

Μια πολύ σημαντική παράμετρος της επίδοσης ενός συστήματος είναι ο χρόνος απόκρισής του. Με τη βοήθεια των μετρήσεων που πραγματοποιήθηκαν στο Σύστημα Σημασιολογικού Ευρετηριασμού, είναι δυνατό να γίνουν προβλέψεις του χρόνου απόκρισης του συστήματος για διαφορετικούς συνδυασμούς τιμών των χωρητικοτήτων των επιμέρους μηχανισμών κρυφής μνήμης.
Ο χρόνος απόκρισης ενός συστήματος, στο βαθμό που αυτός εξαρτάται από την επίδοση του μηχανισμού κρυφής μνήμης, εκφράζεται από τη σχέση :

$$T = (hp + mp) * A * t_{hit} + mp * A * t_{miss}, =$$
$$A * t_{hit} + (1 - hp) * t_{miss}, \text{ όπου :}$$

hp	:	είναι το ποσοστό επιτυχών προσπελάσεων στην κρυφή μνήμη,
mp	:	είναι το ποσοστό ανεπιτυχών προσπελάσεων στην κρυφή μνήμη,
A	:	είναι ο συνολικός αριθμός προσπελάσεων στην κρυφή μνήμη,
t_{hit}	:	είναι ο χρόνος προσπέλασης στην κρυφή μνήμη,
t_{miss}	:	είναι ο χρόνος προσπέλασης στο μαγνητικό μέσο αποθήκευσης.

Στην παραπάνω σχέση, ο πρώτος όρος του αθροίσματος είναι ο συνολικός χρόνος που καταναλώνεται από προσπελάσεις στην κρυφή μνήμη και μόνο, ενώ ο δεύτερος όρος αντιπροσωπεύει το χρόνο που καταναλώνεται στις προσπελάσεις του δίσκου. Ο λόγος για τον οποίο στον πρώτο όρο εμφανίζονται τόσο το ποσοστό επιτυχών, όσο και το ποσοστό ανεπιτυχών προσπελάσεων στην κρυφή μνήμη είναι διότι κάθε ανεπιτυχής προσπέλασηστην κρυφή μνήμη ακολουθείται από μεταφορά σελίδας - ή σελίδων - από το δίσκο και στη συνέχεια μια επιτυχημένη νέα προσπέλαση στην κρυφή μνήμη, η οποία ωστόσο δε συνυπολογίζεται στο αντίστοιχο ποσοστό.
Οταν, λοιπόν, είναι γνωστοί οι χρόνοι προσπέλασης στη μνήμη και στο δίσκο σε ένα σύστημα, γεγονός που είναι δυνατό όταν το σύστημα χρησιμοποιείται σε απομονωμένο υπολογιστή αλλά όχι σε υπολογιστή σε δίκτυο, είναι δυνατό με τη βοήθεια των γραφικών παραστάσεων που παρουσιάστηκαν παραπάνω να πραγματοποιηθούν οι απαραίτητοι

υπολογισμοί για τον προσδιορισμό του αναμενόμενου χρόνου απόκρισης του συστήματος.

Κεφάλαιο 9

Επίλογος

9.1 Συμπεράσματα

Στα πλαίσια της εργασίας αυτής μελετήθηκε διεξοδικά ο μηχανισμός διαχείρισης κρυφής μνήμης για το Σύστημα Σημασιολογικού Ευρετηριασμού. Η μελέτη πραγματοποιήθηκε με τη βοήθεια μετρήσεων επίδοσης, οι οποίες εκτελέστηκαν στο πραγματικό σύστημα. Τα χαρακτηριστικά στοιχεία του σχεδιασμού των πειραμάτων υπήρξαν τα ακόλουθα :

- παράμετροι των πειραμάτων είναι οι χωρητικότητες των επιμέρους μηχανισμών κρυφής μνήμης του ΣΣΕ,
- τα μεγέθη που καταγράφουν τα πειράματα είναι η κατανάλωση μνήμης του συστήματος και τα ποσοστά επιτυχημένων προσπελάσεων στην κρυφή μνήμη, τα οποία προσδιορίζουν και την επίδοση του συστήματος σε ό,τι αφορά τους χρόνους απόκρισής του,
- οι μετρήσεις επίδοσης πραγματοποιήθηκαν τόσο για το σύστημα διατύπωσης ερωτήσεων, όσο και για το σύστημα διαλογικών δελτίων ενημέρωσης,
- οι διαφορετικές βάσεις δεδομένων για τις οποίες έγιναν οι μετρήσεις είναι η βάση του συστήματος ΚΛΕΙΩ, ο θησαυρός ΑΑΤ και η βάση του ULAN,
- τα αποτελέσματα των μετρήσεων παραστάθηκαν γραφικά.

Οι επιδόσεις του συστήματος στις διαφορετικές μετρήσεις αποδείχθηκαν πολύ ικανοποιητικές για όλους τους επιμέρους μηχανισμούς κρυφής μνήμης. Κύριο πρόβλημα

του συστήματος προέκυψε ότι είναι η συχνά υπερβολική κατανάλωση μνήμης, όταν αυτό επιτραπεί από τις χωρητικότητες των επιμέρους μηχανισμών.

Αποτέλεσμα των παραπάνω παρατηρήσεων ήταν η επιλογή νέων χωρητικοτήτων για τους επιμέρους μηχανισμούς κρυφής μνήμης στα συστήματα που μελετήθηκαν. Ετσι, για τη βάση του συστήματος ΚΛΕΙΩ χρησιμοποιήθηκαν οι μικρότερες χωρητικότητες για τους επιμέρους μηχανισμούς, σε αντίθεση με τις μεσαίες που χρησιμοποιούνταν ως τώρα. Για τις άλλες δύο βάσεις δεδομένων, εξάλλου, χρησιμοποιήθηκαν για τις χωρητικότητες των επιμέρους μηχανισμών συνδυασμοί κατηγοριών χωρητικοτήτων (μικρές, μεσαίες ή μεγάλες) ανάλογα με την κατανάλωση μνήμης και την επίδοση που επέδειξε κάθε μηχανισμός.

Προκειμένου να εξισορροπιθούν οι απατήσεις του συστήματος σε μνήμη και χρόνο, οι οποίες αποτελούν πόρους ανταγωνιστικούς, παραστάθηκαν γραφικά οι επιδόσεις σε ποσοστά επιτυχημένων προσπελάσεων των επιμέρους μηχανισμών κρυφής μνήμης συναρτήσει των χωρητικοτήτων τους. Οι παραστάσεις αυτές, οι οποίες στη συγκεκριμένη εργασία κατασκευάστηκαν μόνο για τη βάση ULAN, μπορούν, εφόσον κατασκευαστούν για διαφορετικές κατηγορίες βάσεων δεδομένων, να αποτελέσουν έναν οδηγό επιλογής χωρητικοτήτων για τους επιμέρους μηχανισμούς, ώστε η συμπεριφορά του συστήματος να'ναι κατά περίπτωση η επιθυμητή.

9.2 Μελλοντική εργασία

Τα συμπεράσματα που προέκυψαν από την εκτενέστερη μελέτη της συμπεριφοράς του συστήματος για τη βάση ULAN οδήγησε, όπως προαναφέρθηκε, στην εξαγωγή γραφικών παραστάσεων με βάση τις οποίες μπορεί να υπολογίζεται η επίδοση του συστήματος για διαφορετικές τιμές των μεγεθών των επιμέρους μηχανισμών κρυφής μνήμης. Ετσι, μπορεί να είναι από πριν γνωστή η συμπεριφορά που πρέπει να αναμένεται για το σύστημα, δεδομένων συγκεκριμένων περιορισμών μνήμης. Με τη βοήθεια των ίδιων παραστάσεων, είναι δυνατό να αποφασιστεί ο καταμερισμός της διαθέσιμης μνήμης στους επιμέρους μηχανισμούς, ώστε αυτοί να παρουσιάζουν επίδοση που κρίνεται ανεκτή.

Εκτελώντας την ίδια εργασία για διαφορετικές "κατηγορίες" βάσεων δεδομένων, είναι δυνατό η παραπάνω μέθοδος να εφαρμόζεται για οποιαδήποτε εφαρμογή και βάση δεδομένων, αρκεί αυτή να καταταγεί σε μία από τις διαθέσιμες κατηγορίες.

Το Σύστημα Σημασιολογικού Ευρετηριασμού στη μορφη που μελετήθηκε στην εργασία αυτή. εξάλλου, είναι υλοποιημένο σε απομονωμένη αρχιτεκτονική (stand alone). Εξαιτίας, ωστόσο, της χρησιμοποίησής του σε μουσεία όπου είναι απαραίτητη η δυνατότητα ταυτόχρονης προσπέλασης κάθε βάσης δεδομένων από οσοδήποτε πολλούς χρήστες, η αρχιεκτονική στην οποία ήδη υλοποιείται το σύστημα είναι αυτή του πελάτη-εξυπηρετητή (client-server).

Στη νέα αρχιτεκτονική, νέες μετρήσεις επίδοσης θα χρειαστούν, καθώς τώρα οι επιμέρους μηχανισμοί κρυφής μνήμης θα χρησιμοποιούνται από κοινού από πολυάριθμες εφαρμογές. Ανάλογα με το ποσοστό επικάλυψης που θα παρουσιάζουν οι προσπελάσεις από τους διαφορετικούς πελάτες, θα διαμορφώνεται η συνολική επίδοση του συστήματος.

Βιβλιογραφία

[1] Rafael Alonso, Daniel Barbara, and Hector Garcia-Molina. Data caching issues in an information retrieval system. *ACM Transactions on Database Systems*, 15(3):359--384, September 1990.

[2] James Bell, David Casasent, and C. Gordon Bell. An investigation of alternative cache organizations. *IEEE Transactions on Computers*, C-23(4):346--351, April 1974.

[3] Elisa Bertino and Lorenzo Martino. Object-Oriented Database Management Systems : Concepts and issues. *IEEE Computer*, 24(4):33--47, April 1991.

[4] Paul Butterworth, Allen Otis, and Jacob Stein. The Gemstone Object Database Management System. *Communications of the ACM*, 34(10):64--77, October 1991.

[5] David Callahan, Ken Kennedy, and Allan Portfield. Software prefetching. pages 40--52, 1991.

[6] Michael J. Carey, David J. DeWitt, and Jeffrey F. Naughton. The OO7 benchmark. Technical report, Computer Sciences Department, University of Wisconsin-Madison, January 1994.

[7] Michael J. Carey, David J. DeWitt, Joel E. Richardson, and Eugene J. Shekita. Object and file management in the EXODUS extensible database system. In *Proceedings of the 12th International Conference on Very Large Databases*, pages 91--100, August 1986. Kyoto.

[8] R. G. G. Cattell and J. Skeen. Object operations benchmark. *ACM Transactions on Database Systems*, 17(1):1--31, March 1992.

[9] Lucien M. Censier and Paul Feautrier. A new solution to coherence problems in multicache systems. *IEEE Transactions on Computers*, C-27(12):1112--1118, December 1978.

[10] Vinay K. Chaudhri and Russell Greiner. A formal analysis of solution caching. In *Proceedings of the Canadian Artificial Intelligence Conference*, 11-15 May 1992. Vancouver.

[11] ChungMin Melvin Chen and Nicholas Roussopoulos. Adaptive database buffer allocation using query feedback. Technical Report UMIACS-TR-93-49.

[12] Hong-Tai Chou and David J. DeWitt. An evaluation of buffer management strategies for relational database systems. In *Proccedings of the 11th International Conference on Very Large Databases*, pages 127--141, 1985. Stockholm.

[13] Wesley W. Chu and Holger Opderbeck. Program behavior and the page-fault frequency algorithm. *Computer*, pages 29--38, November 1976.

[14] Panos Constantopoulos and Martin Doerr. The Semantic Index System - a brief presentation. Institute of Computer Science FORTH, May 1993.

[15] Costas Dadouris. Programmatic Query Interface and query processing for the TELOS language. Master's thesis, University of Crete, 1993.

[16] Costas Dadouris, Martin Doerr, Stavroula Kizlaridou, Dimitris Mauroidis, Elena Pataki, Manos Theodorakis, Maria Theodoridou, and George Yeorgiannakis. Implementation of the SIB system. Technical Report ITHACA.FORTH.92.E2.#7, Institute of Computer Science FORTH, January 1992.

[17] Peter J. Denning. The working set model for program behavior. *Communications of the ACM*, 11(5):323--333, May 1968.

[18] Peter J. Denning. Working sets past and present. *IEEE Transactions on Software Engineering*, SE-6(1):64--84, January 1980.

[19] Martin Doerr, Polivios Klimathianakis, and Manos Theodorakis. SIS Data Entry Language User's Manual. Technical report, Computer Science Institute FORTH, February 1995.

[20] Wolfgang Effelsberg and Theo Haerder. Principles of Database Buffer Management. *ACM Transactions on Database Systems*, 9(4):560--595, December 1984.

[21] Klaus Elhart and Rudolf Bayer. A database cache for high performance and fast restart. *ACM Transactions on Database Systems*, 9(4):503--525, December 1984.

[22] Foundation for Research and Technology-Hellas, Institute of Computer Science. *SIB data entry language user's manual*, Dec 1992.

[23] Michael J. Franklin. *Caching and Memory Management in Client-Server Database Systems.* PhD thesis, University of Wisconsin-Madison, 1993.

[24] Michael J. Franklin and Michael J. Carey. Client-server caching revisited. Technical Report Computer Sciences Technical Report #1089, Computer Sciences Department, University of Wisconsin-Madison, May 1992.

[25] Michael J. Franklin, Michael J. Carey, and Miron Linvy. Global memory management in Client-Server DBMS architectures. Technical Report Computer Sciences Technical Report #1094, Computer Sciences Department, University of Wisconsin-Madison, June 1992.

[26] Michael J. Franklin, Michael J. Carey, Miron Linvy, and Eugene Shekita. Data caching tradeoffs in Client-Server DBMS architectures. Technical Report Computer Sciences Technical Report #994, Computer Sciences Department, University of Wisconsin-Madison, January 1991.

[27] Sanjay Ghemawat. Disk management for Object-Oriented databases.

[28] Popi Halkia. Implementation description of the set manipulation mechanism for the semantic index system. Technical report, Information Systems and Software Technology Group, Computer Science Institute FORTH, October 1993. Confidential Report #2.

[29] Robert B. Hangman. An observation on database buffering performance metrics. In *Proccedings of the 12th International Conference on Vely Large Databases*, pages 289--293, August 1986. Kyoto.

[30] A. R. Hurson, Simin H. Pakzad, and Jia bing Cheng. Object-Oriented Database Management Systems : Evolution and performance issues. *IEEE Computer*, 26(2):48--60, February 1993.

[31] Nabil Kamel and Roger King. Intelligent database caching through the use of page-answers and page-traces. *ACM Transactions on Database Systems*, 17(4):601--646, December 1992.

[32] Won Kim, Jorge F. Garza, Nathaniel Ballou, and Darrell Woelk. Architecture of the ORION next-generation system. *IEEE Transactions on Knowledge and Data Engineering*, 2(1):109--124, March 1990.

[33] M. Koubarakis, J. Mylopoulos, M. Stanley, and A. Borgida. TELOS : Features and formalization. Technical Report FORTH/CSI/TR/1989/018, Institute of Computer Science FORTH, 1989.

[34] Klaus Kratzer, Harmut Wedekind, and Georg Zorntlein. Prefetching : A performance analysis. *Information Systems*, 15(4):445--452, 1990.

[35] K. Kuspert, P. Dadam, and J. Gunauer. Cooperative object buffer management in the advanced information management prototype. In *Proceedings of the 13th International Conference on Very Large Databases*, pages 483--492, 1987. Brighton.

[36] Charles Lamb, Gordon Landis, Jack Orenstein, and Dan Weinreb. The Objectstore database system. *Communications of the ACM*, 34(10):50--63, October 1991.

[37] O. Deux et al. The story of O_2. *IEEE Transactions on Knowledge and Data Engineering*, 2(1):91--108, March 1990.

[38] O. Deux et al. The O_2 system. *Communications of the ACM*, 34(10):34--48, October 1991.

[39] Vidyadhar Phalke and Bhaskarpillai Gopinath. An inter-reference gap model for temporal locality in program behavior.

[40] Giovanni Maria Sacco. Index access with a finite buffer. In *Proceedings of the 13th International Conference on Very Large Databases*, 1987. Brighton.

[41] Giovanni Maria Sacco and Mario Schkolnick. A mechanism for managing the buffer pool in a relational database system using the hot set model. In *Proceedings of the 8th International Conference on Very Large Databases*, pages 257--262, 1982.

[42] Alan Jay Smith. Cache memories. *ACM Computing Surveys*, 14(3):473--530, September 1982.

[43] Michael Stonebraker and Greg Kemnitz. The Postgress next generation Database Management System. *Communications of the ACM*, 34(10):78--92, October 1991.

[44] Douglas B. Terry. Caching hints in distributed systems. *IEEE Transactions on Software Engineering*, SE-13(1):48--54, January 1987.

[45] George Yeorgiannakis. Demand loading on object extensions. Technical report, Information Systems and Software Technology Group, Computer Science Institute FORTH, June 1994. Working Note #6.

[46] George Yeorgiannakis. Future developments concerning the memory and disk management mechanisms. Technical report, Information Systems and Software Technology Group, Computer Science Institute FORTH, April 1994. Working Note #4.

[47] George Yeorgiannakis. A storage and memory management mechanism for objects in TELOS. Master's thesis, University of Crete, 1994.

[48] George Yeorgiannakis and Stavroula Kizlaridou. Benchmarking the SIS. Technical report, Information Systems and Software Technology Group, Computer Science Institute FORTH, February 1994. Working Paper #8.

[49] George Yeorgiannakis and Stavroula Kizlaridou. System catalogue and symbol table evolution for the support of unnamed objects. Technical report, Information Systems and Software Technology Group, Computer Science Institute FORTH, July 1995. Working Paper Draft.

Παράρτημα Α

Σενάριο τυπικής χρήσης του ΚΛΕΙΩ

Α. Αναδρομικές Ερωτήσεις

- **1ο βήμα**

  Στόχος ερώτησης *Είδος'σπάθα*. (1 δοσοληψία)

- **2ο βήμα**

  Από τον κατάλογο επιλογών (menu) του gain, επιλογή ερώτησης $Queries->GeneralQueries->Object/Kind->ListObjectsofKind$. (2 δοσοληψίες)

- **3ο βήμα**

  Επιλογή δέκα αντικειμένων που ανήκουν στο σύνολο απάντησης (αν το σύνολο αυτό αποτελείται απο λιγότερα, τότε επιλογή όλων) της παραπάνω ερώτησης και εφαρμογή της αναδρομηκής ερώτησης $TreeViews->getmaincharacteristics$ σε αυτά. (12 δοσοληψίες)

Β. Μια μεγάλη και μια μικρή ερώτηση

- **1ο βήμα**

  Στόχος ερώτησης *ArtificialObject*.

- **2ο βήμα**

  Από τον κατάλογο επιλογών (menu) του gain, επιλογή ερώτησης $TreeViews->ClassificationTree$. (2 δοσοληψίες)

- **3ο βήμα**

  Επιλογή τυχαίου κόμβου και εμφάνιση της κάρτας αντικειμένου (Object Card) γι αυτό. (2 δοσοληψίες)

C. Η παρακάτω διαδικασία απαιτεί τη χρήση των Δελτίων Εισαγωγής Δεδομένων και του gain.

- **1ο βήμα**

  i) Κλήση των Δελτίων Εισαγωγής Δεδομένων (επιλογή $Admin->entryforms$ από τον κατάλογο επιλογών του gain). (2 δοσοληψίες)

  ii) Στο περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : Επιλογή λειτουργίας *Describe Object* με root class *ArtificialObject*. (2 δοσοληψίες)

- **2ο βήμα**

  Επιλογή αντικειμένου *ΑντικείμενοBIB949*.

- **3ο βήμα**

  Απο το περιβάλλον του gain : Επιλογή ερώτησης $TreeViews->StarView$ και εκτέλεσή της για το παραπάνω αντικείμενο. (2 δοσοληψίες)

- **4ο βήμα**

  Από το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : Επιλογή λειτουργίας *Edit Attributes*. (1 δοσοληψία)

- **5ο βήμα**

  Μετακίνηση στο *aquired*, επιλογή *NEW*, επιλογή *Apply*. (2 δοσοληψίες)

- **6ο βήμα**

  Στη νέα κάρτα που εμφανίζεται, μετακίνηση στο *date* και επιλογή γι αυτό το *NEW* (με το τρίτο πλήκτρο του ποντικιού). (1 δοσοληψία)

- **7ο βήμα**

  Μετακίνηση στο *date* και επιλογή του *current* (στη νέα κάρτα). (1 δοσοληψία)

- **8ο βήμα**

  Επιστροφή στην προηγούμενη κάρτα (με επιλογή των πλήκτρων *commit και return*.

- **9ο βήμα**

  Μετακίνηση στο *place*. (1 δοσοληψία)

- **10ο βήμα**

  Επιλογή του *OLD* (με το τρίτο πλήκτρο του ποντικιού).

- **11ο βήμα**

  Επιλογή τυχαίου τόπου. Επιλογή *apply*. (1 δοσοληψία)

- **12ο βήμα**

  Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (1 δοσοληψία)

- **13ο βήμα**

  Μετακίνηση στο *created* (1 δοσοληψία) και επανάληψη των βημάτων 5-12 ως εξής:

  -- **5ο' βήμα**

  Μετακίνηση στο *created*, επιλογή *NEW*, επιλογή *Apply*. (2 δοσοληψίες)

  -- **6ο' βήμα**

  Στη νέα κάρτα που εμφανίζεται, μετακίνηση στο *date* και επιλογή γι αυτό το *NEW* (με το τρίτο πλήκτρο του ποντικιού) και επιλογή *apply*. (1 δοσοληψία)

  -- **7ο' βήμα**

  Μετακίνηση στο *date* και επιλογή του *current* (στη νέα κάρτα). (1 δοσοληψία)

  -- **8ο' βήμα**

  Επιστροφή στην προηγούμενη κάρτα (με επιλογή των πλήκτρων *commit και return*.

  -- **9ο' βήμα**

  Μετακίνηση στο *place*. (1 δοσοληψία)

  -- **10ο' βήμα**

  Επιλογή του *OLD* (με το τρίτο πλήκτρο του ποντικιού).

  -- **11ο' βήμα**

  (1 δοσοληψία) Επιλογή τυχαίου τόπου. Επιλογή *apply*.

  -- **12ο' βήμα**

  Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (1 δοσοληψία)

- **15ο βήμα**

  Απο το περιβάλλον του gain : Επιλογή *exec.* (1 δοσοληψία)

- **16ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (12 δοσοληψίες)

D.

- **1ο βήμα**

  Από το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : Επιλογή λειτουργίας *Describe Class* με root class *ArtificialObject*. 2 δοσοληψίες)

- **2ο βήμα**

  Επιλογή αντικειμένου *Είδος'άγαλμα*.

- **3ο βήμα**

  Επιλογή λειτουργίας *Edit Attributes.* (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *appearance* και επιλογή *OLD.* (1 δοσοληψία)

- **5ο βήμα**

  Ενημέρωση του πεδίου κειμένου *Link label* με το κείμενο *test*.

- **6ο βήμα**

  Ενημέρωση του πεδίου κειμένου *New attribute* με το κείμενο *Color*.

- **7ο βήμα**

  Επιλογή *apply* και Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (2 δοσοληψίες)

- **8ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

E.

- **1ο βήμα**

  Από το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : Επιλογή λειτουργίας *Describe Object*. (1 δοσοληψία)

- **2ο βήμα**

  Επιλογή αντικειμένου *ΑντικείμενοΓΕ5848*.

- **3ο βήμα**

  Επιλογή λειτουργίας *Edit Attributes*. (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *museumNumber* και επιλογή *NEW* γι αυτό. (2 δοσοληψίες)

- **5ο βήμα**

  Επιλογή *apply*. (1 δοσοληψία)

- **6ο βήμα**

  Μετακίνηση στο *material* και επιλογή *OLD*. (1 δοσοληψία)

- **7ο βήμα**

  Επιλογή τυχαίου αντικειμένου και επιβεβαίωση της παραπάνω διαδικασίας (με την επιλογή *commit*). (1 δοσοληψία)

- **8ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

Παράρτημα Β

Σενάριο τυπικής χρήσης του AAT

Α. Αναδρομικές Ερωτήσεις

- **1ο βήμα**

  i) Από τον κατάλογο επιλογών (menu) του gain, επιλογή $Retrieval->patternsearch$ με παραμέτρους *type :Term* και *name : dolls*. (2 δοσοληψίες)

  ii) Επιλογή *Term'dolls* από το αποτέλεσμα που εμφανίζεται στο γραφικό περιβάλλον του gain. (1 δοσοληψία)

- **2ο βήμα**

  Από τον κατάλογο επιλογών (menu) του gain, επιλογή ερώτησης $TreeViews->Termpositioninhierarchy$. (2 δοσοληψίες)

- **3ο βήμα**

  Εκτέλεση της παραπάνω ερώτησης (2ο βήμα) με στόχο τον κόμβο *figurines* από το γράφο που εμφανίζεται στο γραφικό περιβάλλον του gain. (1 δοσοληψία)

- **4ο βήμα**

  Εκτέλεση της παραπάνω ερώτησης (2ο βήμα) με στόχο τον κόμβο *statues* από το γράφο που εμφανίζεται στο γραφικό περιβάλλον του gain. (1 δοσοληψία)

- **5ο βήμα**

  Εκτέλεση της παραπάνω ερώτησης (2ο βήμα) με στόχο τον κόμβο *Term'*$<sculpturebysubjecttype>$ από το γράφο που εμφανίζεται στο γραφικό περιβάλλον του gain. (1 δοσοληψία)

- **6ο βήμα**

  Εκτέλεση της παραπάνω ερώτησης (2ο βήμα) με στόχο τον κόμβο *Term'sculpture* από το γράφο που εμφανίζεται στο γραφικό περιβάλλον του gain. (1 δοσοληψία)

Β. Μια μεγάλη και μια μικρή ερώτηση

- **1ο βήμα**

  Στόχος ερώτησης *Facet'OBJECTS*

- **2ο βήμα**

  Από τον κατάλογο επιλογών (menu) του gain, επιλογή ερώτησης $Queries- > Terms- > ListTermsbyFacet$. (2 δοσοληψίες)

- **3ο βήμα**

  Κάρτα αντικειμένου (object card) για τυχαίο κόμβο από το γράφο που εμφανίζεται στο γραφικό περιβάλλον του gain. (2 δοσοληψίες)

C.

- **1ο βήμα**

  Επιλογή $Admin- > entryforms$ από τον κατάλογο επιλογών (menu) του gain. (2 δοσοληψίες)

- **2ο βήμα**

  Επιλογή λειτουργίας (task) *Edit term with Hierarchy Hierarchy'Animals* . (2 δοσοληψίες)

- **3ο βήμα**

  Επιλογή όρου *Term'*$< Animalattributes >$.

- **4ο βήμα**

  Επιλογή της ερώτησης $TreeViews- > StarView$ από τον κατάλογο επιλογών του gain για τον κόμβο *Term'*$< Animalattributes >$. (2 δοσοληψίες)

- **5ο βήμα**

  Επιλογή της λειτουργίας *Edit term attributes* από τα Δελτία Εισαγωγής Δεδομένων. (1 δοσοληψία)

- **6ο βήμα**

  Επιλογή *OLD* γιά το γνώρισμα *approval_editor*. (1 δοσοληψία)

- **7ο βήμα**

  Επιλογή τυχαίου αντικειμένου και ενημέρωση (επιλογή *apply*). (1 δοσοληψία)

- **8ο βήμα**

  Μετακίνηση στο δημιουργηθέν, επιλεχθέν τρέχον αντικείμενο.

- **9ο βήμα**

  Επιλογή *commit*. (1 δοσοληψία)

- **10ο βήμα**

  Επιλογή *exec* από το περιβάλλον του gain. (1 δοσοληψία)

- **11ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

D.

Στο περιβάλλον των Δελτίων Εισαγωγής Δεδομένων:

- **1ο βήμα**

  Επιλογή λειτουργίας *describe schema* με root class *AATdescriptor*. (2 δοσοληψίες)

- **2ο βήμα**

  Επιλογή *DeletedTerm*.

- **3ο βήμα**

  Επιλογή λειτουργίας *Edit attributes*. (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *attribute* και επιλογή *NEW/EXISTING*. (1 δοσοληψία)

- **5ο βήμα**

  Ενημέρωση του πεδίου κειμένου *Link label* με το κείμενο *test*.

- **6ο βήμα**

  Ενημέρωση του πεδίου κειμένου *New attribute* με το κείμενο *Term*.

- **7ο βήμα**

  Επιλογή *apply*. (1 δοσοληψία)

- **8ο βήμα**

  Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (1 δοσοληψία)

- **7ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

Ε.

Στο περιβάλλον των Δελτίων Εισαγωγής Δεδομένων:

- **1ο βήμα**

  Επιλογή λειτουργίας *submit term*. (1 δοσοληψία)

- **2ο βήμα**

  Επιλογή *TermSubmission'Doerr'submit_armonicas*.

- **3ο βήμα**

  Επιλογή λειτουργίας *Edit Submission*. (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *as_BT* και επιλογή *OLD*. (1 δοσοληψία)

- **5ο βήμα**

  Επιλογή *Hierarchy'Animals* και επιλογή *apply* γι αυτό. (1 δοσοληψία)

- **6ο βήμα**

  Επιλογή *Term'*$< Animalattributes >$ και επιλογή *apply* γι αυτό. (1 δοσοληψία)

- **7ο βήμα**

  Μετακίνηση στο *submitted_for* και επιλογή *NEW*. (1 δοσοληψία)

- **8ο βήμα**

  Επιλογή *Hierarchy'Animals* και επιλογή *apply* γι αυτό. (1 δοσοληψία)

- **9ο βήμα**

  Συμπλήρωση σχολίου (τυχαίου) στο γραφικό περιβάλλον (καρτέλλα) που εμφανίζεται. Μετακίνηση στο *as_ALT* και επιλογή του *OLD* γι αυτό. (1 δοσοληψία)

- **10ο βήμα**

  Επιλογή *Hierarchy'Animals* και επιλογή *apply* γι αυτό. (1 δοσοληψία)

- **11ο βήμα**

  Επιλογή *Term'*$< Animalattributes >$ και επιλογή *apply* γι αυτό. (1 δοσοληψία)

- **12ο βήμα**

  Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (1 δοσοληψία)

- **13ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

Παράρτημα C

Σενάριο τυπικής χρήσης του ULAN

Α. Αναδρομικές Ερωτήσεις

- **1ο βήμα**
  Από τον κατάλογο επιλογών (menu) του gain, επιλογή *Retrieval-¿pattern search* με παραμέτρους *type :Person* και *name : Huys.*
  (1 δοσοληψία)

- **2ο βήμα**
  Επιλογή αντικειμένου *RILA/PW717384'Huys, Frans.* (1 δοσοληψία)

- **3ο βήμα**
  Επιλογή *Starview* από τον κατάλογο επιλογών του gain. (2 δοσοληψίες)

- **4ο βήμα**
  Επιλογή *Person history*. (1 δοσοληψία)

- **5ο βήμα**
  Επιλογή *Person submission Tree.* (1 δοσοληψία)

- **6ο βήμα**
  Εμφάνιση της κάρτας αντικειμένων (Object Card) για όλα τα submission objects. (6 δοσοληψίες)

Επιλογή του πλήκτρου *TEXT* και επανάληψη των προηγούμενων βημάτων για όλα τα

παραπάνω αντικείμενα. (71 δοσοληψίες)

Β. Μια μεγάλη και μια μικρή ερώτηση

- **1ο βήμα**

  Επιλογή ερώτησης $Retrieval->Personsearch$ με παράμετρο *role : architect* και σύνολο απάντησης 14667 αντικειμένων. (1 δοσοληψία)

- **2ο βήμα**

  Εμφάνιση κάρτας αντικειμένου για τυχαίο αντικείμενο του συνόλου απάντησης. (2 δοσοληψίες)

C.

- **1ο βήμα**

  Απο το περιβάλλον του gain : Επιλογή *Person history* για το αντικείμενο *RILA/PW717384'Huys, Frans.* (2 δοσοληψίες)

- **2ο βήμα**

  Απο το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων: Επιλογή λειτουργίας *Edit Person.* (1 δοσοληψία)

- **3ο βήμα**

  Ενημέρωση του πεδίου κειμένου *Selected object* με το κείμενο *RILA/PW717384'Huys, Frans.*

- **4ο βήμα**

  Επιλογή λειτουργίας *Edit Person Attributes.* (1 δοσοληψία)

- **5ο βήμα**

  Μετακίνηση στο γνώρισμα *anscestor* .

- **6ο βήμα**

  Επιλογή *OLD.* (1 δοσοληψία)

- **7ο βήμα**

  Επιλογή *university*, επιλογή *apply*. (1 δοσοληψία)

- **8ο βήμα**

  Επιλογή *King's College, London.*

- **9ο βήμα**

  Επιλογή *apply* και επιβεβαίωση της παραπάνω διαδικασίας (με την επιλογή *commit*). (2 δοσοληψίες)

- **10ο βήμα**

  Απο το περιβάλλον του gain : Επιλογή *exec.* (1 δοσοληψία)

- **11ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

D.

- **1ο βήμα**

  Απο το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : Επιλογή λειτουργίας *edit schema.*

- **2ο βήμα**

  Επιλογή αντικειμένου *Person* και επιλογή *apply*. (2 δοσοληψίες)

- **3ο βήμα**

  Επιλογή αντικειμένου *Person*, επιλογή λειτουργίας *edit attributes.* (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *attribute* και επιλογή *NEW/EXISTING.* (1 δοσοληψία)

- **5ο βήμα**

  Ενημέρωση πεδίου κειμένου *Link label* με το κείμενο *test.*

- **6ο βήμα**

  Ενημέρωση πεδίου κειμένου *New attribute* με το κείμενο *Person.*

- **7ο βήμα**

  Επιλογή *apply*, επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (2 δοσοληψίες)

- **8ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)

E.

- **1ο βήμα**

  Από το περιβάλλον των Δελτίων Εισαγωγής Δεδομένων : επιλογή λειτουργίας *Submit person*. (1 δοσοληψία)

- **2ο βήμα**

  Ενημέρωση πεδίου κειμένου *Selected Object* με το κείμενο *RILA/PW717384'submission*.

- **3ο βήμα**

  Επιλογή λειτουργίας *Edit Submission*. (1 δοσοληψία)

- **4ο βήμα**

  Μετακίνηση στο *submitted_by* και επιλογή *OLD*. (1 δοσοληψία)

- **5ο βήμα**

  Επιλογή *CENS*, επιλογή *apply*. (1 δοσοληψία)

- **6ο βήμα**

  Μετακίνηση στο *as_1st husband*, επιλογή *OLD*. (1 δοσοληψία)

- **7ο βήμα**

  Επιλογή *Corporation*, επιλογή *apply*. (1 δοσοληψία)

- **8ο βήμα**

  Επιλογή *Adler and Sullivan*, επιλογή *apply*. (1 δοσοληψία)

- **9ο βήμα**

  Μετακίνηση στο *as_2nd husband*, επιλογή *NEW*.

- **10ο βήμα**

  Επιλογή *Female*. (1 δοσοληψία)

- **11ο βήμα**

  Ενημέρωση του πεδίου κειμένου που εμφανίζεται με τυχαίο κείμενο. (1 δοσοληψία)

- **12ο βήμα**

  Επιβεβαίωση της παραπάνω διαδικασίας (επιλογή *commit*). (1 δοσοληψία)

- **13ο βήμα**

  Διαγραφή της πληροφορίας που εισήχθη παραπάνω. (1 δοσοληψία)