

# ΚΥΜΑΤΙΔΙΑ (WAVELETS)

ΚΑΤΑΣΚΕΥΗ ΠΕΡΙΛΗΨΕΩΝ ΚΥΜΑΤΙΔΙΩΝ (WAVELET SYNOPSES) ΓΙΑ ΣΥΝΑΘΡΟΙΣΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ (RANGE QUERIES) .....	2
ΕΠΕΚΤΑΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΠΕΡΙΛΗΨΕΩΝ $\tau$ -synopses.....	3

**ΚΑΤΑΣΚΕΥΗ ΠΕΡΙΛΗΨΕΩΝ ΚΥΜΑΤΙΔΙΩΝ (WAVELET SYNOPSSES) ΓΙΑ  
ΣΥΝΑΘΡΟΙΣΤΙΚΕΣ ΕΡΩΤΗΣΕΙΣ (RANGE QUERIES)**

ΠΛΗΡΟΦΟΡΙΕΣ: Δ. Σαχαρίδης, 210 772 1402, dsachar@dblab.ece.ntua.gr

**ΠΕΡΙΛΗΨΗ:** Η διπλωματική εργασία στοχεύει στην μελέτη υπάρχουσων μεθόδων περίληψης με βάση τον μετασχηματισμό κυματιδίων (wavelet synopses) και κατασκευή νέων περιλήψεων σχεδιασμένων ειδικά για συναθροιστικές ερωτήσεις (range queries).

**ΑΤΟΜΑ:** 1

**ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ:** GNU C++ σε WINDOWS/LINUX/UNIX

**ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:** Ο μετασχηματισμός κυματιδίων (wavelet transform) χρησιμοποιείται ευρέως για την δημιουργία περιλήψεων χρονικών σειρών και πολυδιάστατων δεδομένων. Κύρια χαρακτηριστικά του είναι η απλότητα και η ταχύτητα του καθώς η υψηλή συμπίεση που προσφέρει σε σχέση με τα ευρέως χρησιμοποιούμενα ιστογράμματα.

Όπως κάθε απωλεστικός (lossy) αλγόριθμος συμπίεσης, έτσι και οι περιλήψεις κυματιδίων εισάγουν σφάλματα στην αναπαραγωγή του αρχικού σήματος. Τα σφάλματα αυτά μετρούνται και ζυγίζονται με βάση τα ερωτήματα που γίνονται στην περίληψη. Μέχρι τώρα έμφαση έχει δοθεί σε ερωτήματα που αφορούν μόνο μια τιμή (point queries). Σκοπός είναι να σχεδιαστεί αλγόριθμος ή/και μετασχηματισμός που να λαμβάνει υπόψη και συναθροιστικά ερωτήματα (range queries) για διάφορες μετρικές σφάλματος.

Σε πρώτη φάση θα γίνει μελέτη των υπαρχόντων τεχνικών περιλήψεων κυματιδίων για διάφορες μετρικές σφάλματος καθώς και των αντίστοιχων τεχνικών για ιστογράμματα. Έπειτα θα προταθούν/σχεδιασθούν/αναπτυχθούν νέες τεχνικές για συναθροιστικά ερωτήματα. Τέλος θα γίνει η υλοποίηση σε C++ είτε σε windows (μέσω cygwin) είτε σε Unix/Linux.

**ΧΡΗΣΙΜΕΣ ΓΝΩΣΕΙΣ:**

- Ανάλυση Αλγορίθμων (δυναμικός προγραμματισμός, εύρεση χρονικής και χωρικής πολυπλοκότητας)
- Καλή γνώση C++ σε περιβάλλον WINDOWS/LINUX/UNIX

**ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:**

- M. Garofalakis "[Wavelet-based Approximation Techniques in Database Systems](#)", *IEEE Signal Processing Magazine*, November 2006
- M. Mathioudakis, D. Sacharidis, T. Sellis, "[A Study on Workload-Aware Wavelet Synopses](#)", ACM Ninth International Workshop on Data Warehousing and OLAP (DOLAP '06), November 2006 [\[διαφάνειες\]](#)
- D. Sacharidis, "[Constructing Optimal Wavelet Synopses](#)", PhD Workshop of 10th International Conference on Extending Database Technology (EDBT '06), March 2006 [\[διαφάνειες\]](#)
- P. Karras and N. Mamoulis, "[The Haar+ Tree: a Refined Synopsis Data Structure](#)", 23rd International Conference on Data Engineering (ICDE '07), April 2007 (to appear)

**ΕΠΕΚΤΑΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΠΕΡΙΛΗΨΕΩΝ  $\tau$ -synopses**

ΠΛΗΡΟΦΟΡΙΕΣ: Δ. Σαχαρίδης, 210 772 1402, [dsachar@dblab.ntua.gr](mailto:dsachar@dblab.ntua.gr)

**ΠΕΡΙΛΗΨΗ:** Η διπλωματική εργασία στοχεύει στην επέκταση του συστήματος περιλήψεων  $\tau$ -synopses του πανεπιστημίου του Τελ Αβίβ με νέες τεχνικές περίληψης, κυρίως με βάση τον μετασχηματισμό κυματιδίων (wavelet synopses), που έχουν προταθεί στη βιβλιογραφία.

**ΑΤΟΜΑ:** 1

**ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ:** .NET

**ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:** Οι συνόψεις είναι ακριβείς περιλήψεις για δεδομένα μεγάλου όγκου και χρησιμοποιούνται για την γρήγορη προσεγγιστική απάντηση ερωτήσεων. Οι πιο κοινές τεχνικές για σχεσιακά δεδομένα είναι τα ιστογράμματα και οι περιλήψεις κυματιδίων. Θα ασχοληθούμε με τις τελευταίες. Συγκεκριμένα, ο **μετασχηματισμός κυματιδίων** (wavelet transform) χρησιμοποιείται ευρέως για την δημιουργία περιλήψεων χρονικών σειρών και πολυδιάστατων δεδομένων. Κύρια χαρακτηριστικά του είναι η απλότητα, η ταχύτητα και η υψηλή συμπίεση που προσφέρει σε σχέση με τα ευρέως χρησιμοποιούμενα ιστογράμματα.

Το σύστημα  $\tau$ -synopses έχει αναπτυχθεί από το πανεπιστήμιο του Τελ Αβίβ υπό την επίβλεψη του καθηγητή Y. Matias. Σκοπός του είναι να προσφέρει ένα ενοποιημένο περιβάλλον για τη σύγκριση τεχνικών περίληψης για διάφορα είδη δεδομένων, σχεσιακά και XML.

Σε πρώτη φάση θα γίνει μελέτη της αρχιτεκτονικής του συστήματος  $\tau$ -synopses και της προγραμματιστικής διεπαφής (API) που προσφέρει. Στη συνέχεια θα υλοποιηθούν, με βάση την διεπαφή αυτή, διάφοροι αλγόριθμοι περιλήψεων που έχουν προταθεί και υλοποιηθεί. Να σημειωθεί ότι υπάρχει πρότυπος κώδικας (C++), οπότε απαιτείται κατάλληλη τροποποίησή του.

**ΧΡΗΣΙΜΕΣ ΓΝΩΣΕΙΣ:**

- Ανάλυση Αλγορίθμων (δυναμικός προγραμματισμός, εύρεση χρονικής και χωρικής πολυπλοκότητας)
- Καλή γνώση περιβάλλοντος .NET

**ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:**

- Y. Matias, L. Portman, N. Druk ["The design and architecture of the  \$\tau\$ -synopses system"](#), Industrial & Application, International Conference on Extending Database Technology (EDBT '06), March 2006
- M. Garofalakis ["Wavelet-based Approximation Techniques in Database Systems"](#), *IEEE Signal Processing Magazine*, November 2006
- M. Mathioudakis, D. Sacharidis, T. Sellis, ["A Study on Workload-Aware Wavelet Synopses"](#), ACM Ninth International Workshop on Data Warehousing and OLAP (DOLAP '06), November 2006 [\[διαφάνειες\]](#)
- D. Sacharidis, ["Constructing Optimal Wavelet Synopses"](#), PhD Workshop of 10th International Conference on Extending Database Technology (EDBT '06), March 2006 [\[διαφάνειες\]](#)
- P. Karras and N. Mamoulis, ["The Haar+ Tree: a Refined Synopsis Data Structure"](#), 23rd International Conference on Data Engineering (ICDE '07), April 2007 (to appear)