

# ΔΙΑΦΟΡΑ ΘΕΜΑΤΑ

|  |   |
|--|---|
| ΔΙΑΣΦΑΛΙΣΗ ΕΜΠΙΣΤΕΥΤΙΚΟΤΗΤΑΣ (PRIVACY PRESERVATION) ΣΕ ΔΥΝΑΜΙΚΑ ΔΕΔΟΜΕΝΑ ..... | 2 |
| ΕΠΕΚΤΑΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΠΕΡΙΛΗΨΕΩΝ τ-synopses .....                            | 3 |

**ΔΙΑΣΦΑΛΙΣΗ ΕΜΠΙΣΤΕΥΤΙΚΟΤΗΤΑΣ (PRIVACY PRESERVATION)  
ΣΕ ΔΥΝΑΜΙΚΑ ΔΕΔΟΜΕΝΑ**

ΠΛΗΡΟΦΟΡΙΕΣ: Δ. Σαχαρίδης, 210 772 1402, dsachar@dbl-lab.ntua.gr

**ΠΕΡΙΛΗΨΗ:** Η διπλωματική εργασία στοχεύει στην κατασκευή τεχνικών για την διασφάλιση της εμπιστευτικότητας μέσω τεχνικών ανωνυμίας (privacy preserving anonymity) για σχεσιακά δεδομένα τα οποία αλλάζουν συχνά.

**ΑΤΟΜΑ:** 1

**ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ:** C++

**ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ:** Σε πολλές εφαρμογές είναι απαραίτητη η δημοσιοποίηση στοιχείων που αφορούν ανθρώπους. Τα δεδομένα αυτά μπορούν έπειτα να χρησιμοποιηθούν από ποικίλες εφαρμογές (όπως στατιστική επεξεργασία) κάτι που όμως έχει ως αποτέλεσμα την αποκάλυψη της ταυτότητας των προσώπων. Σκοπός της διασφάλισης της εμπιστευτικότητας (privacy preservation) είναι η ανωνυμοποίηση (anonymizing) των δεδομένων με τέτοιο τρόπο ώστε η σύνδεση με την ταυτότητα των προσώπων που συνεισφέρουν σε αυτά να μην είναι πλέον δυνατή.

Ακόμα και αν αφαιρεθούν τα αναγνωριστικά των εγγραφών (π.χ. Ονοματεπώνυμο, Αριθμός Ταυτότητας) η ταυτοποίηση μπορεί να γίνει εν μέρει μέσω άλλων γνωρισμάτων (π.χ. Ηλικία, Ταχυδρομικός Κώδικας) που αποκαλούνται *ψευδοαναγνωριστικά* (quasi identifiers). Αυτά τα γνωρίσματα δεν μπορούμε να τα αφαιρέσουμε (π.χ. επειδή είναι χρήσιμα στις εφαρμογές) και το μόνο που απομένει να κάνουμε είναι να τα γενικεύσουμε (generalization). Σκοπός των εφαρμογών είναι η εύρεση της κατάλληλης ισορροπίας ανάμεσα στο πόσο γενικευμένα είναι τα δεδομένα και πόσο χρήσιμα είναι αυτά.

Οι μέθοδοι που έχουν προταθεί ως τώρα, με την εξαίρεση μιας πρόσφατης εργασίας, εξετάζουν στατικά δεδομένα. Αντικείμενο αυτής της διπλωματικής είναι η μελέτη πιο δυναμικών δεδομένων. Στην περίπτωση που έχουν εισαγωγές και διαγραφές από τα δεδομένα θα πρέπει να προσέχουμε κατά τη διάρκεια της γενίκευσης ώστε να μην φανερώσουμε την ταυτότητα προσώπων ακόμα και εκείνων των οποίων οι εγγραφές τους έχουν σβηστεί.

**ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:**

- K. LeFevre, D. DeWitt, R. Ramakrishnan, “Mondrian Multidimensional k-Anonymity”, ICDE 2006
- X. Xiao, Y. Tao, “m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets”, SIGMOD 2007

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

*ΠΑΗΡΟΦΟΡΙΕΣ: Δ. Σαχαρίδης, 210 772 1402, dsachar@dblab.ntua.gr*

**ΠΕΡΙΛΗΨΗ:** Η διπλωματική εργασία στοχεύει στην επέκταση του συστήματος περιλήψεων  $\tau$ -synopses του πανεπιστημίου του Τελ Αβιβ 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 διαδεδομένες τεχνικές για σχεσιακά δεδομένα είναι τα ιστογράμματα και οι περιλήψεις κυματιδίων. Ο **μετασχηματισμός κυματιδίων** (wavelet transform) χρησιμοποιείται ευρέως για τη δημιουργία περιλήψεων χρονικών σειρών, αλλά 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.

Το σύστημα  **$\tau$ -synopses** έχει αναπτυχθεί για να προσφέρει ένα ενοποιημένο περιβάλλον για τη σύγκριση τεχνικών περίληψης για διάφορα είδη δεδομένων. Θα γίνει μελέτη της αρχιτεκτονικής του συστήματος  $\tau$ -synopses και της προγραμματιστικής διεπαφής που προσφέρει (API). Στη συνέχεια θα υλοποιηθούν διάφοροι αλγόριθμοι περιλήψεων που έχουν προταθεί. Να σημειωθεί ότι υπάρχει πρότυπος κώδικας (σε C++), οπότε απαιτείται η κατάλληλη τροποποίησή του.

**ΧΡΗΣΙΜΕΣ ΓΝΩΣΕΙΣ:**

- Αλγόριθμοι (δυναμικός προγραμματισμός, ανάλυση πολυπλοκότητας)
- Καλή γνώση περιβάλλοντος .NET

**ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ/ΑΝΑΦΟΡΕΣ:**

- Y.Matias, L. Portman, N. Drukh "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]