

Multimedia Indexing

Gemini, FastMap, Applications

Βάσεις Δεδομένων II

Τμήμα Μηχανικών Η/Υ και Πληροφορικής
Πολυτεχνική Σχολή, Πανεπιστήμιο Πατρών

Εαρινό Εξάμηνο 2011-2012

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

4 Lower Bounding

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

4 Lower Bounding

5 GEMINI

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

4 Lower Bounding

5 GEMINI

6 Εφαρμογές

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

4 Lower Bounding

5 GEMINI

6 Εφαρμογές

7 FastMap

Table of contents

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1 Εισαγωγή

2 Συναρτήσεις Απόστασης/ Ομοιότητας

3 Multimedia Indexing - Απαιτήσεις

4 Lower Bounding

5 GEMINI

6 Εφαρμογές

7 FastMap

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Περιγραφή Προβλήματος

Σκοπός

Για μια βάση δεδομένων που περιέχει πολυμέσα όπως:

- 1-d time sequences
- Digitized voice or music
- 2-d color images
- 2-d or 3-d gray scale medical images
- Video clips

στόχος είναι η δημιουργία αλγορίθμων για την απάντηση διάφορων ερωτημάτων

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Περιγραφή Προβλήματος

Σκοπός

Για μια βάση δεδομένων που περιέχει πολυμέσα όπως:

- 1-d time sequences
- Digitized voice or music
- 2-d color images
- 2-d or 3-d gray scale medical images
- Video clips

στόχος είναι η δημιουργία αλγορίθμων για την απάντηση διάφορων ερωτημάτων

Παράδειγμα

"Βρες τις εικόνες που έχουν παρόμοιο περιεχόμενο"

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Κατηγορίες Ερωτημάτων (1)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

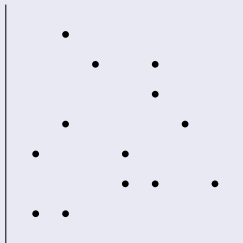
Εφαρμογές

FastMap

Αναφορές

Whole Match Queries

- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Κατηγορίες Ερωτημάτων (1)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

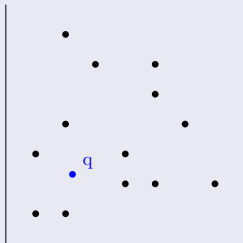
Εφαρμογές

FastMap

Αναφορές

Whole Match Queries

- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Κατηγορίες Ερωτημάτων (1)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

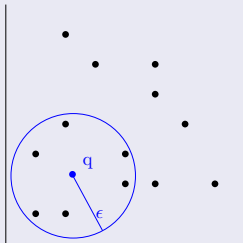
Εφαρμογές

FastMap

Αναφορές

Whole Match Queries

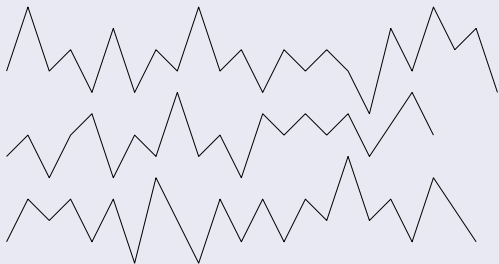
- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Κατηγορίες Ερωτημάτων (2)

Sub-pattern Match

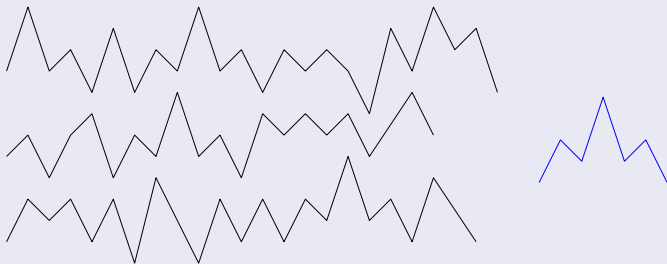
- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα (υπό-)αντικείμενο-ερώτημα q , βρες τα (υπό-)αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Κατηγορίες Ερωτημάτων (2)

Sub-pattern Match

- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα (υπό-)αντικείμενο-ερώτημα q , βρες τα (υπό-)αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

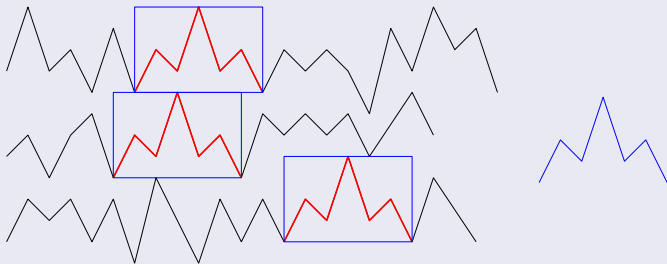
FastMap

Αναφορές

Κατηγορίες Ερωτημάτων (2)

Sub-pattern Match

- Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα (υπό-)αντικείμενο-ερώτημα q , βρες τα (υπό-)αντικείμενα που **διαφέρουν** από το q το πολύ ϵ .



Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Κατηγορίες Ερωτημάτων (3)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

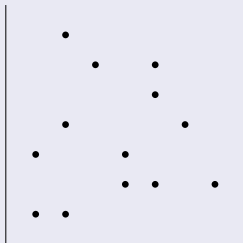
Εφαρμογές

FastMap

Αναφορές

k nearest neighbors

- Για μία συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα k αντικείμενα που **μοιάζουν** περισσότερο στο q .



Κατηγορίες Ερωτημάτων (3)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

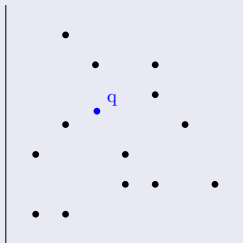
Εφαρμογές

FastMap

Αναφορές

k nearest neighbors

- Για μία συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα k αντικείμενα που **μοιάζουν** περισσότερο στο q .



Κατηγορίες Ερωτημάτων (3)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

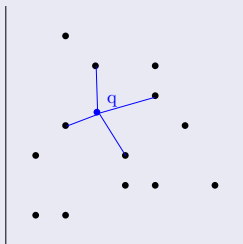
Εφαρμογές

FastMap

Αναφορές

k nearest neighbors

- Για μία συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα k αντικείμενα που **μοιάζουν** περισσότερο στο q .



Κατηγορίες Ερωτημάτων (4)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

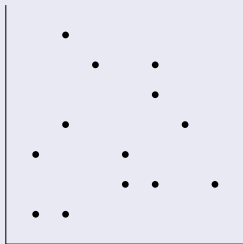
Εφαρμογές

FastMap

Αναφορές

spatial joins

- Για μία συλλογή N αντικειμένων O_1, O_2, \dots, O_N βρες τα ζεύγη αντικειμένων που απέχουν μεταξύ τους το πολύ ϵ .



Κατηγορίες Ερωτημάτων (4)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

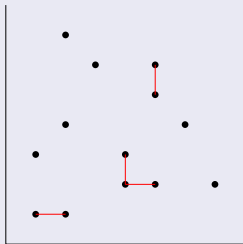
Εφαρμογές

FastMap

Αναφορές

spatial joins

- Για μία συλλογή N αντικειμένων O_1, O_2, \dots, O_N βρες τα ζεύγη αντικειμένων που απέχουν μεταξύ τους το πολύ ϵ .



- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Συνάρτηση Απόστασης/Ομοιότητας

Ποσοτικοποιεί την απόσταση/ομοιότητα μεταξύ δυο αντικειμένων

Distance function $D(O_1, O_2)$

- $D(O_1, O_1) = 0$
- $D(O_1, O_2) \neq 0$ if $O_1 \neq O_2$
- Επίσης, η τριγωνική ανισότητα είναι επιθυμητή...

Τι εννοούμε όταν λέμε δυο αντικείμενα μοιάζουν ή διαφέρουν?

Δυο εικόνες μπορεί να μοιάζουν ως προς:

- Τα χρώματά τους
- Το textures
- Το σχήμα του αντικειμένου που απεικονίζουν
- Τη σημασία του περιεχομένου τους

Συναρτήσεις Απόστασης Χρονοσειρών

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

- Ευκλείδεια Απόσταση
- Dynamic Time Warping
- Minimal Variance Matching
- And many more...

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 **Multimedia Indexing - Απαιτήσεις**
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Multimedia Indexing - Απαιτήσεις

Multimedia Indexing

Βάσεις Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις Απόστασης/ Ομοιότητας

Multimedia Indexing - Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Έχουμε μια συλλογή δεδομένων πολυμέσων, από την οποία θέλουμε να ανακτούμε αντικείμενα με βάση κάποιο ερώτημα

Ο υπολογισμός όλων των αποστάσεων σειριακά, για κάθε αντικείμενο είναι πολύ αργός. Για την εναλλακτική προσέγγιση θα πρέπει να ισχύουν:

- Όχι false dismissals
- Μικρή επιβάρυνση χώρου
- Δυναμική (εισαγωγές - διαγραφές - ανανεώσεις)

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

- Εξαγωγή Χαρακτηριστικών (Feature Extraction)
- Χρήση Spatial Access Methods για την επιτάχυνση της αναζήτησης

Feature Extraction

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Σκοπός

Έστω ένα αντικείμενο πχ εικόνα, video, τραγούδι κλπ. Στόχος είναι η αναζήτηση κάποιων "χαρακτηριστικών" που θα χρησιμοποιηθούν για την αναπαράσταση του αντικειμένου. Ο αριθμός των χαρακτηριστικών θα πρέπει να είναι μικρότερος του αρχικού μεγέθους.

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding**
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Whole Match Query

Για μια συλλογή N αντικειμένων O_1, O_2, \dots, O_N και ένα αντικείμενο-ερώτημα q , βρες τα αντικείμενα που διαφέρουν από το q το πολύ ϵ .

Απλή Προσέγγιση

- Ο σειριακός υπολογισμός μπορεί να είναι πολύ αργός για δυο λόγους:
 - Το ϵ μπορεί να είναι υπερβολικά μεγάλο
 - Ο υπολογισμός της συνάρτησης απόστασης μπορεί να χρειάζεται πολύ χρόνο

Lower Bounding

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Έστω μια συνάρτηση απόστασης $D(O_1, O_2)$

- 1 Εξαγωγή k χαρακτηριστικών ($k < N$)
- 2 Αντιστοίχιση των δεδομένων στον χώρο των χαρακτηριστικών $F(O_1), F(O_2)$
- 3 Ορισμός της συνάρτησης απόστασης στον χώρο των χαρακτηριστικών $D_F(F(O_1), F(O_2))$

False Dismissals

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Για να μην υπάρχουν false dismissals θα πρέπει:

$$D_F(F(O_1), F(O_2)) \leq D(O_1, O_2)$$

Proof

- θέλω $D(q, O_i) \leq \epsilon$
- Αν $D(q, O_i) \leq \epsilon$ τότε $D_F(F(q), F(O_i)) \leq \epsilon$
επειδή
 $D_F(F(O_1), F(O_2)) \leq D(O_1, O_2)$

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI**
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

GEneric Multimedia INdexIng

- 1 Ορισμός της συνάρτησης απόστασης $D()$
- 2 Αναζήτηση των χαρακτηριστικών που θα χρησιμοποιηθούν
- 3 Απόδειξη ότι η $D_F()$ είναι lower-bounds της $D()$
- 4 Χρήση μιας SAM για την αποθήκευση και ανάκτηση των feature vectors

Βασικά Βήματα

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος
Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

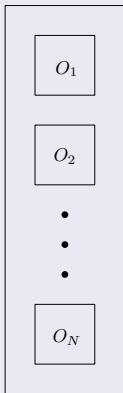
GEMINI

Εφαρμογές

FastMap

Αναφορές

Gemini



Βασικά Βήματα

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος
Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

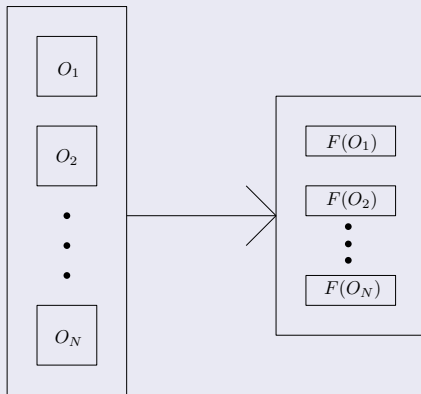
GEMINI

Εφαρμογές

FastMap

Αναφορές

Gemini



Βασικά Βήματα

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος
Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

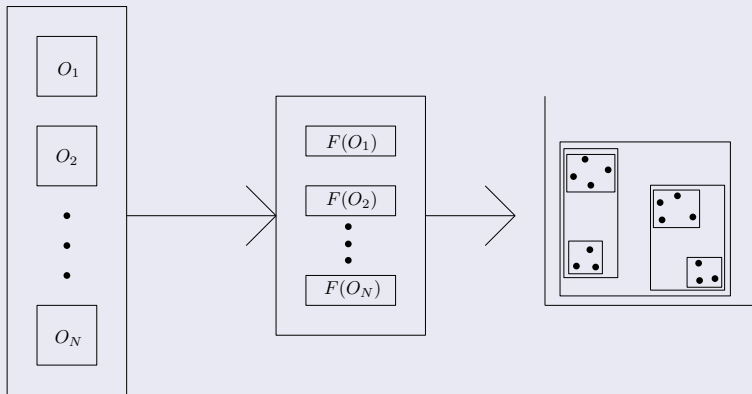
GEMINI

Εφαρμογές

FastMap

Αναφορές

Gemini



- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Σύνολο Δεδομένων

Έστω ένα σύνολο χρονοσειρών $T_1 T_2 \dots T_M$ μήκους N

$$T_i = [t_{i0} t_{i1} \dots t_{iN-1}]$$

Σύνολο Δεδομένων

Έστω ένα σύνολο χρονοσειρών $T_1 T_2 \dots T_M$ μήκους N

$$T_i = [t_{i0} t_{i1} \dots t_{iN-1}]$$

Συνάρτηση Απόστασης

$$D(x, y) = \sum_{i=0}^{N-1} |x_i - y_i|^2$$

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

1ο Βήμα

- Επιλογή της αντιστοίχισης $F()$

1ο Βήμα

- Επιλογή της αντιστοίχισης $F()$
- F : Discrete Fourier Transform - k πρώτους συντελεστές

1ο Βήμα

- Επιλογή της αντιστοίχισης $F()$
- F : Discrete Fourier Transform - k πρώτους συντελεστές
- $D_F(F(x), F(y)) = \sum_{f=0}^k |X_f - Y_f|^2$

2ο Βήμα

- Απόδειξη ότι η D_F είναι lower bound της D

2ο Βήμα

- Απόδειξη ότι η D_F είναι lower bound της D
- $D_F(F(x), F(y)) = \sum_{f=0}^{k-1} |X_f - Y_f|^2$

2ο Βήμα

- Απόδειξη ότι η D_F είναι lower bound της D
- $D_F(F(x), F(y)) = \sum_{f=0}^{k-1} |X_f - Y_f|^2 \leq \sum_{f=0}^{N-1} |X_f - Y_f|^2$

2ο Βήμα

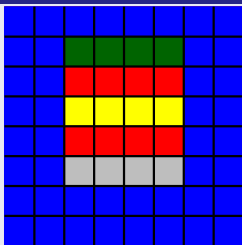
- Απόδειξη ότι η D_F είναι lower bound της D
- $$D_F(F(x), F(y)) = \sum_{f=0}^{k-1} |X_f - Y_f|^2 \leq \sum_{f=0}^{N-1} |X_f - Y_f|^2$$
$$= \sum_{i=0}^{N-1} |x_i - y_i|^2$$

2ο Βήμα

- Απόδειξη ότι η D_F είναι lower bound της D
- $$D_F(F(x), F(y)) = \sum_{f=0}^{k-1} |X_f - Y_f|^2 \leq \sum_{f=0}^{N-1} |X_f - Y_f|^2$$
$$= \sum_{i=0}^{N-1} |x_i - y_i|^2 \equiv D(x, y)$$

Δεδομένα

Έστω ένα σύνολο N εικόνων. Κάθε εικόνα είναι ένας πίνακας από pixels και κάθε pixel έχει 3 συνιστώσες (R, G, B)



Εικόνες - Color Histogram

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Color Histogram

$H_{C_i}(I)$: αριθμός pixels χρώματος C_i της εικόνας I

Συνάρτηση απόστασης

$$d_h^2 = (x - y)^t \times A \times (x - y) = \sum_i^h \sum_j^h a_{i,j} (x_i - y_j)(x_j - y_i)$$

Εικόνες - Color Histogram

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Lower Bounding

$$\bar{x} = (R_{avg}, G_{avg}, B_{avg})^t$$

$$R_{avg} = \frac{1}{P} \sum_{i=1}^P R(p)$$

$$d_{avg}^2(\bar{x}, \bar{y}) = \sum_{i=1}^3 (x_i - y_i)^2$$

Εικόνες - Color auto-correlogram

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Correlogram

$$\gamma_{C_i, C_j}^{(k)}(I) \equiv Pr_{p_1 \in I_{C_i}, p_2 \in I} [p_2 \in I_{C_j} \mid |p_1 - p_2| = k]$$

auto-correlogram

$$a_C^{(k)}(I) \equiv \gamma_{C,C}^{(k)}(I)$$

Εικόνες - Color auto-correlogram

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

συμπεριλαμβάνει χωρική πληροφορία

Εκφράζει την πιθανότητα ύπαρξης pixels σε απόσταση k με το ίδιο χρώμα.

Feature Distance

$$|I - I'| = \sum_{l \in m, k \in d} \frac{\gamma_{C_l}^{(k)}(I) - \gamma_{C_l}^{(k)}(I')}{1 + \gamma_{C_l}^{(k)}(I) + \gamma_{C_l}^{(k)}(I')}$$

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap**
- 8 Αναφορές

Εξαγωγή Χαρακτηριστικών

Από ένα σύνολο δεδομένων N διαστάσεων αναζητούμε μια **καλή** αντιστοίχιση σε ένα χώρο k διαστάσεων, όπου $k < N$

Η εξαγωγή των χαρακτηριστικών είναι ιδιαίτερα σημαντική.
Πώς γίνεται?

- SVD
- K-L Transform
- PCA
- **FastMap**

Σκοπός

Αντιστοίχιση των αντικειμένων σε ένα χώρο k διαστάσεων έτσι ώστε οι αποστάσεις να διατηρούνται όσο το δυνατό περισσότερο.

Είσοδος

- ένας πίνακας αποστάσεων $D N \times N$
ή
- τα αντικείμενα και η συνάρτηση απόστασης

Σκοπός

Αντιστοίχιση των αντικειμένων σε ένα χώρο k διαστάσεων έτσι ώστε οι αποστάσεις να διατηρούνται όσο το δυνατό περισσότερο.

Είσοδος

- ένας πίνακας αποστάσεων $D N \times N$
ή
- τα αντικείμενα και η συνάρτηση απόστασης

FastMap - Βασικά Βήματα

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N-διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a , O_b όσο πιο απομακρυσμένα γίνεται

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a, O_b όσο πιο απομακρυσμένα γίνεται
- 3 Επιλογή του $(n-1)$ -d hyperplane που είναι κάθετο στην ευθεία που περνά από τα σημεία O_a, O_b

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a, O_b όσο πιο απομακρυσμένα γίνεται
- 3 Επιλογή του $(n-1)$ -d hyperplane που είναι κάθετο στην ευθεία που περνά από τα σημεία O_a, O_b
- 4 προβολή των σημείων στην ευθεία και στο hyperplane

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a, O_b όσο πιο απομακρυσμένα γίνεται
- 3 Επιλογή του $(n-1)$ -d hyperplane που είναι κάθετο στην ευθεία που περνά από τα σημεία O_a, O_b
- 4 προβολή των σημείων στην ευθεία και στο hyperplane
- 5 υπολογισμός των νέων αποστάσεων

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a, O_b όσο πιο απομακρυσμένα γίνεται
- 3 Επιλογή του $(n-1)$ -d hyperplane που είναι κάθετο στην ευθεία που περνά από τα σημεία O_a, O_b
- 4 προβολή των σημείων στην ευθεία και στο hyperplane
- 5 υπολογισμός των νέων αποστάσεων
- 6 $n = n-1$

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

FastMap - Βασικά Βήματα

Βασικά Βήματα FastMap

- 1 Αρχικά τα δεδομένα αποτελούν σημεία στον N -διάστατο χώρο
- 2 Επιλέγουμε δυο αντικείμενα O_a, O_b όσο πιο απομακρυσμένα γίνεται
- 3 Επιλογή του $(n-1)$ -d hyperplane που είναι κάθετο στην ευθεία που περνά από τα σημεία O_a, O_b
- 4 προβολή των σημείων στην ευθεία και στο hyperplane
- 5 υπολογισμός των νέων αποστάσεων
- 6 $n = n-1$
- 7 goto 2

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

- 1 Εισαγωγή
 - Περιγραφή Προβλήματος
 - Κατηγορίες Ερωτημάτων
- 2 Συναρτήσεις Απόστασης/ Ομοιότητας
- 3 Multimedia Indexing - Απαιτήσεις
- 4 Lower Bounding
- 5 GEMINI
- 6 Εφαρμογές
- 7 FastMap
- 8 Αναφορές

Multimedia
Indexing

Βάσεις
Δεδομένων II

Εισαγωγή

Περιγραφή
Προβλήματος

Κατηγορίες
Ερωτημάτων

Συναρτήσεις
Απόστασης/
Ομοιότητας

Multimedia
Indexing -
Απαιτήσεις

Lower
Bounding

GEMINI

Εφαρμογές

FastMap

Αναφορές

- Η συγκεκριμένη παρουσίαση του φροντιστηρίου είναι βασισμένη στις Διαφάνειες του Μαθήματος

- Faloutsos, C., R. Barber, et al. (July 1994). "Efficient and Effective Querying by Image Content." J. of Intelligent Information Systems 3(3/4): 231-262.
- Faloutsos, C. and K.-I. D. Lin (May 1995). FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. Proc. of ACM-SIGMOD, San Jose, CA.
- Faloutsos, C., M. Ranganathan, et al. (May 25-27, 1994). Fast Subsequence Matching in Time-Series Databases. Proc. ACM SIGMOD, Minneapolis, MN.

- Flickner, M., H. Sawhney, et al. (Sept. 1995). "Query by Image and Video Content: The QBIC System." IEEE Computer 28(9): 23-32.
- Rafiei, D. and A. O. Mendelzon (1997). Similarity-Based Queries for Time Series Data. SIGMOD Conference, Tucson, AZ.
- Yi, B.-K. and C. Faloutsos (2000). Fast Time Sequence Indexing for Arbitrary Lp Norms. VLDB, Kairo, Egypt.