HDMS 2005, Αθήνα, Αύγουστος 25-26, 2005

Constrained Optimalities in Query Personalization

Γεωργία Κούτρικα, Γιάννης Ιωαννίδης
Πανεπιστήμιο Αθηνών

Constrained Query Personalization

1. Εισαγωγή στη CQP οικογένεια
2. Αρχιτεκτονική Συστήματος
3. Αναζήτηση στον Χώρο Καταστάσεων
4. Πειράματα
5. Σύνοψη

1 Η CQP οικογένεια

Εξατομίκευση Ερώτησης

Ο δυναμικός εμπλουτισμός μίας ερώτησης με σχετικές προτιμήσεις προερχόμενες από ένα προφίλ χρήστη με στόχο την παροχή εξατομικευμένων απαντήσεων

1 Η CQP οικογένεια

Η Εξατομίκευση Ερώτησης είναι

Ένα πρόβλημα βελτιστοποίησης

- μεγιστοποίηση του ενδιαφέροντος χρήστη στην απάντηση με βάση το προφίλ του

1 Η CQP οικογένεια

Η Εξατομίκευση Ερώτησης είναι

Ένα πρόβλημα βελτιστοποίησης

- μεγιστοποίηση του ενδιαφέροντος χρήστη στην απάντηση με βάση το προφίλ του

"Υπερεξατομικευμένες" ερωτήσεις μπορεί να

- είναι ακριβές
- παράγουν κενές λίστες αποτελεσμάτων

1 Η CQP οικογένεια

Η Εξατομίκευση Ερώτησης είναι

Ένα πρόβλημα βελτιστοποίησης με περιορισμούς

- μεγιστοποίηση του ενδιαφέροντος χρήστη στην απάντηση
  με περιορισμούς στο κόστος εκτέλεσης και/ή στο μέγεθος της απάντησης

- ελαχιστοποίηση κόστους εκτέλεσης ερώτησης
  με περιορισμούς στο ενδιαφέρον και/ή στο μέγεθος της απάντησης

1 Η CQP οικογένεια

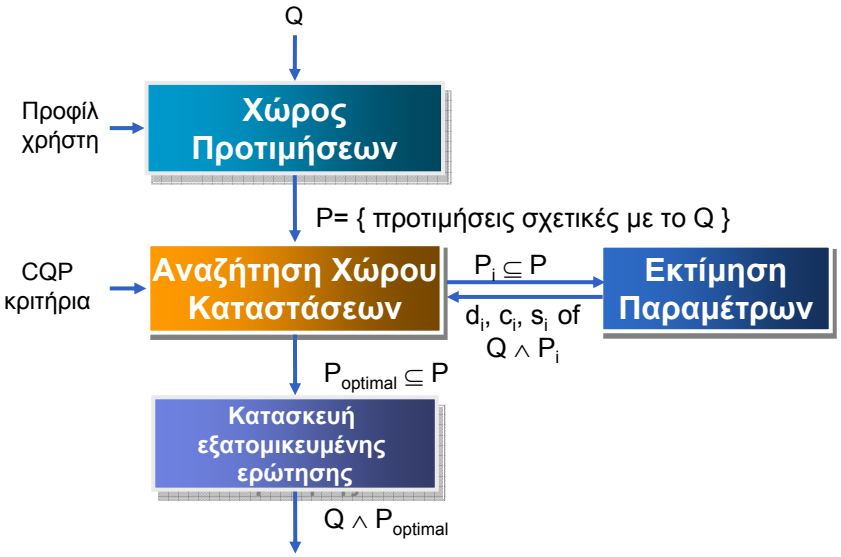
1 Η CQP οικογένεια

Παράδειγμα

- Ακριβές ερωτήσεις
- Εκτενή αποτελέσματα

- Γρηγορότερη απόκριση
- Λιγότερα αποτελέσματα

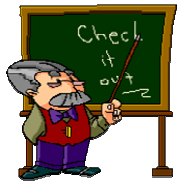
2 CQP Αρχιτεκτονική Συστήματος



2 CQP Αρχιτεκτονική Συστήματος

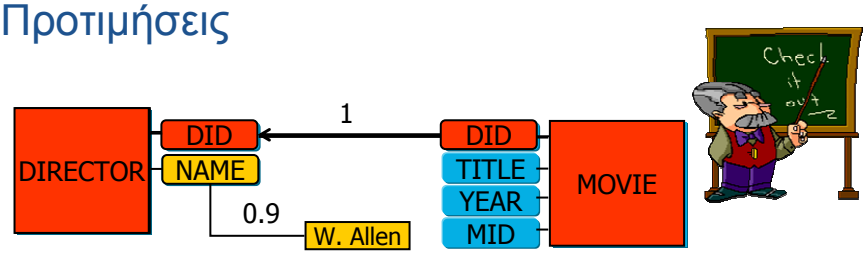
Προτιμήσεις

< q, d>
Ατομική συνθήκη q
Βαθμός ενδιαφέροντος d ∈ [0, 1]



2 CQP Αρχιτεκτονική Συστήματος

Προτιμήσεις



< DIRECTOR.name = 'W.Allen', 0.9 >
< MOVIE.did = DIRECTOR.did, 1 >

Αποθηκευμένες προτιμήσεις

< MOVIE.did = DIRECTOR.did and DIRECTOR.name = 'W.Allen', $0.9 * 1$ >

Έμμεσες προτιμήσεις

2 CQP Αρχιτεκτονική Συστήματος

Κατασκευή εξατομικευμένης ερώτησης

Q: SELECT title FROM MOVIE

p_1: MOVIE.did=DIRECTOR.did and DIRECTOR.name='W. Allen'

p_2: MOVIE.mid=MGENRE.mid and MGENRE.genre='comedy'

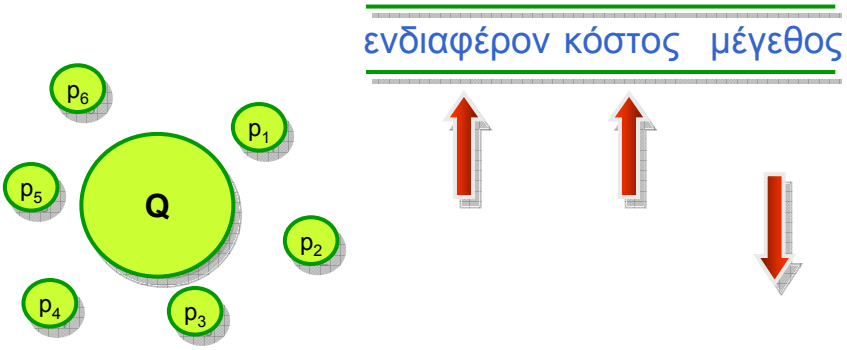
Q_1: SELECT title FROM MOVIE M, DIRECTOR DIR
WHERE M.did=DIR.did and DIR.name='W. Allen'

Q_2: SELECT title FROM MOVIE M, MGENRE G
WHERE M.mid=G.mid and G.genre = 'comedy'

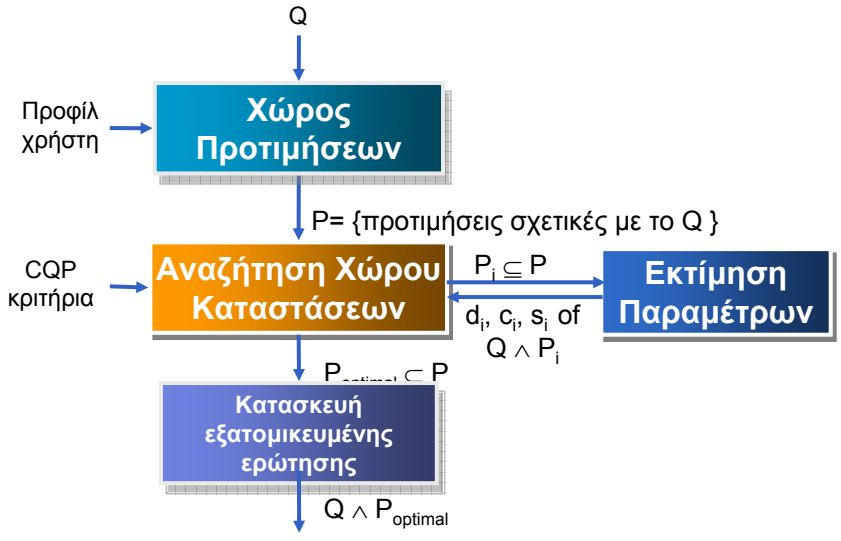
SELECT title FROM Q_1 union all Q_2
GROUP BY title HAVING count(*)>1

2 CQP Αρχιτεκτονική Συστήματος

Εκτίμηση Παραμέτρων



2 CQP Αρχιτεκτονική Συστήματος



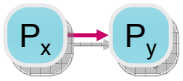
3 CQP Χώρος Καταστάσεων

Ένα CQP πρόβλημα

Μεγιστοποίηση ενδιαφέροντος με κόστος $\leq c_{upper}$

Q , P $\Rightarrow$ ο αριθμός των δυνατών υποσυνόλων του P είναι εκθετικός

Κατάσταση, $P_x \subseteq P$

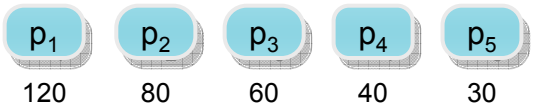


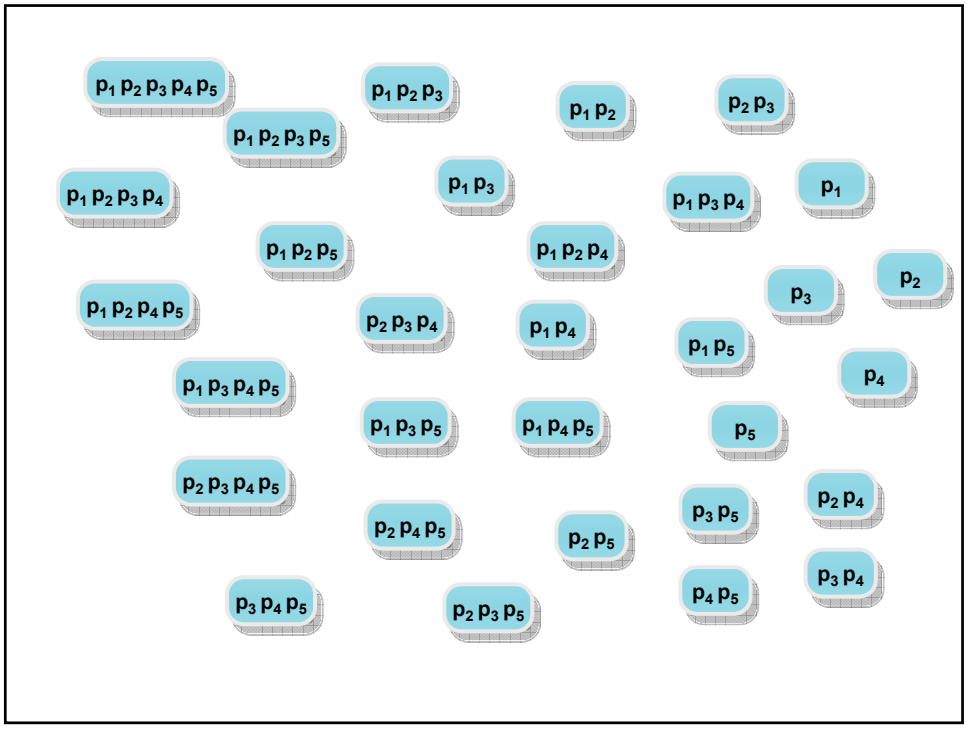
Μετάβαση με βάση:
το κόστος, το μέγεθος, ή το ενδιαφέρον

3 CQP Χώρος Καταστάσεων

Καταστάσεις

Έστω P διατεταγμένο κατά φθίνουσα σειρά κόστους: $cost(p_i) \geq cost(p_{i+1})$

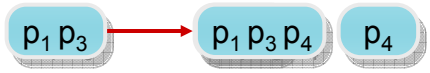




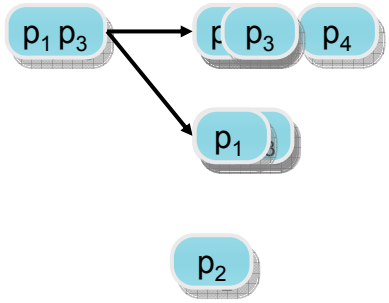
3 CQP Χώρος Καταστάσεων

Μεταβάσεις με βάση το κόστος

- Horizontal(p_1p_3)



- Vertical(p_1p_3)

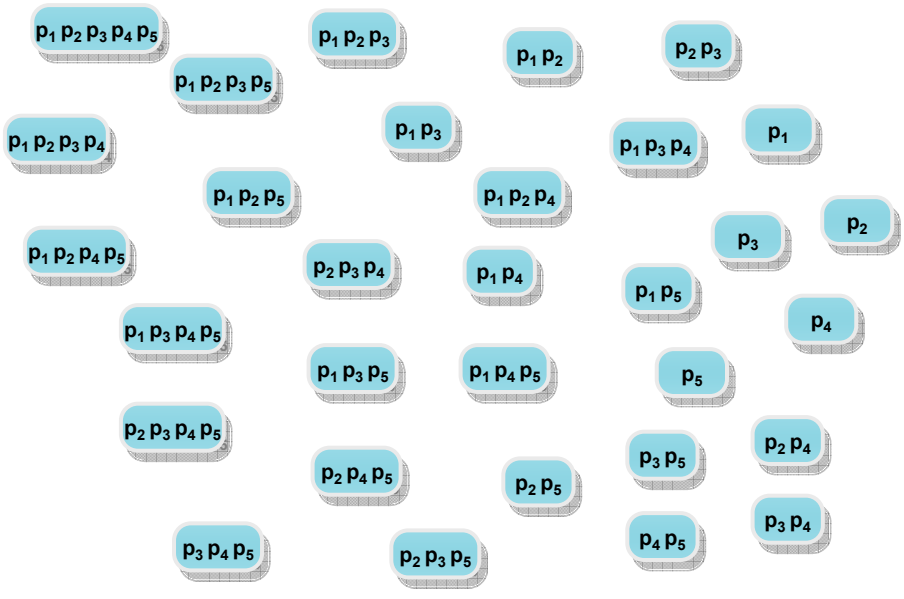


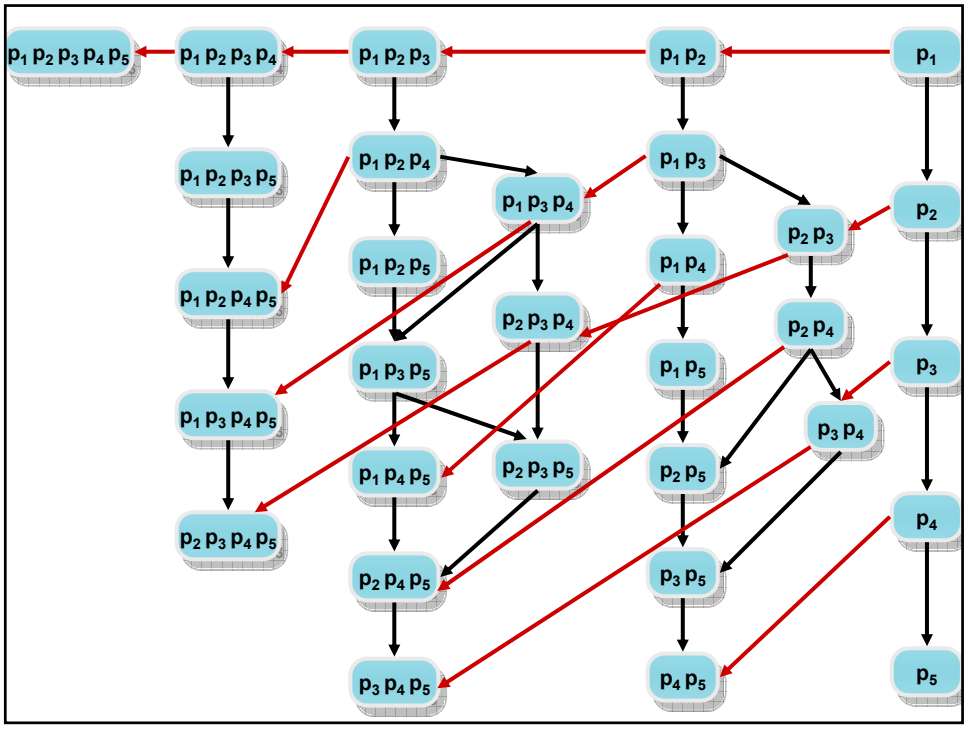
3 CQP Χώρος Καταστάσεων

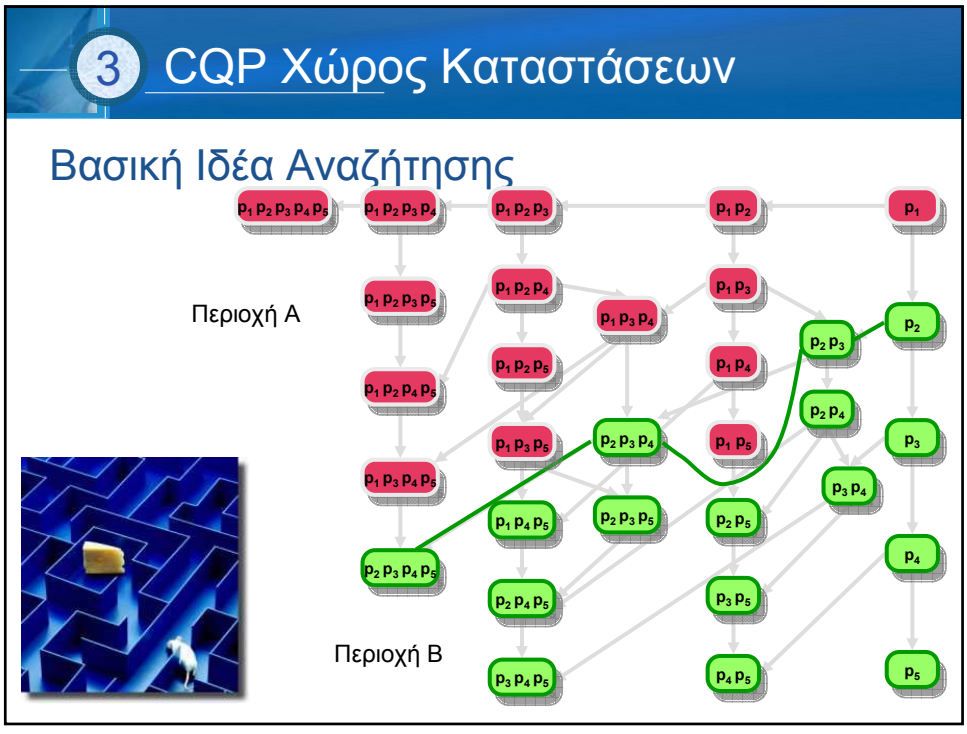
Μεταβάσεις με βάση το κόστος

- interest($p_1p_3p_4$) ≥ interest(p_1p_3)
- cost($p_1p_3p_4$) ≥ cost(p_1p_3)
- cost(p_2p_3) ≤ cost(p_1p_3)

Μερική ταξινόμηση καταστάσεων με βάση συντακτικές αλλαγές




p1 p2 p3 p4 p5
p1 p2 p3 p4
p1 p2 p3
p1 p2
p1
p1 p2 p3 p5
p1 p2 p4
p1 p3
p1 p2 p4 p5
p1 p2 p5
p1 p3 p4
p1 p4
p2
p2 p3
p1 p3 p5
p2 p3 p4
p2 p4
p1 p3 p4 p5
p1 p4 p5
p2 p3 p5
p1 p5
p2 p5
p3
p3 p4
p2 p3 p4 p5
p2 p4 p5
p3 p5
p4
p3 p4 p5
p4 p5
p5


3 CQP Χώρος Καταστάσεων
Βασική Ιδέα Αναζήτησης
Περιοχή Α
Περιοχή Β

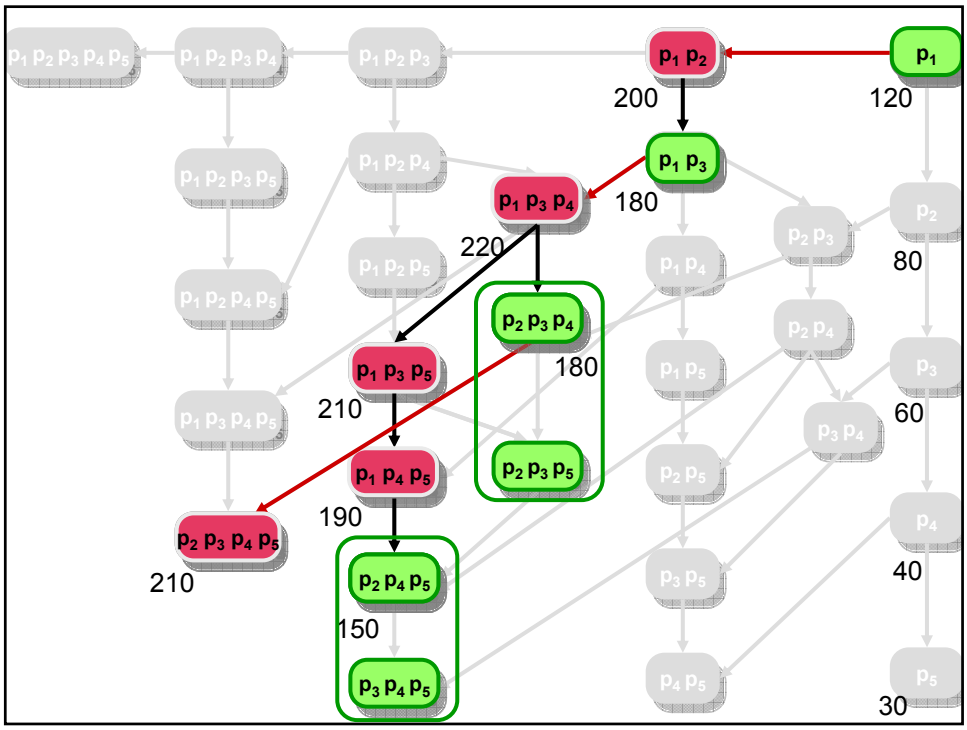
3 CQP Χώρος Καταστάσεων

C_Boundaries

Φάση 1: βρες boundaries

Φάση 2: βρες βέλτιστη λύση

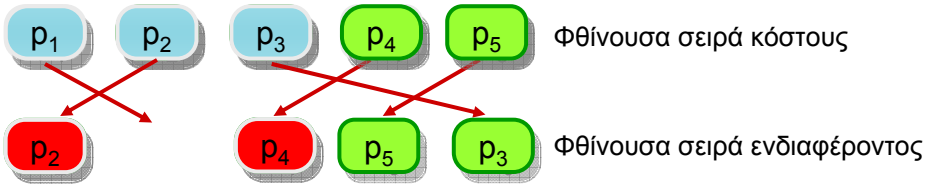
π.χ.
Μεγιστοποίηση του ενδιαφέροντος με κόστος < 185

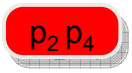


3 CQP Χώρος Καταστάσεων

C_Boundaries (φάση 2)

Αναζήτηση μέσα σε μία ομάδα για τη βέλτιστη λύση





Η βέλτιστη λύση της ομάδας

3 CQP Χώρος Καταστάσεων

Μεταβάσεις κόστους

- C_Boundaries
- C_MaxBounds

Μεταβάσεις ενδιαφέροντος

- D_MaxDoi
- D_SingleMaxDoi
- D_HeurDoi

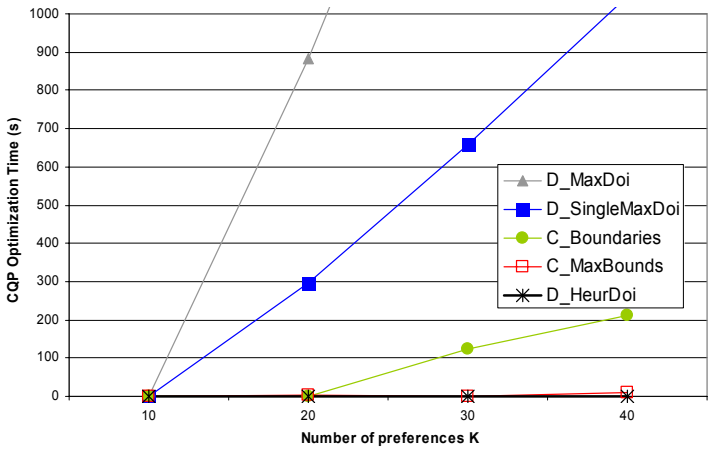
4 Πειράματα

Execution Times

K = # προτιμήσεων στο P

$C_{μεγ}$ = 400

4 Πειράματα



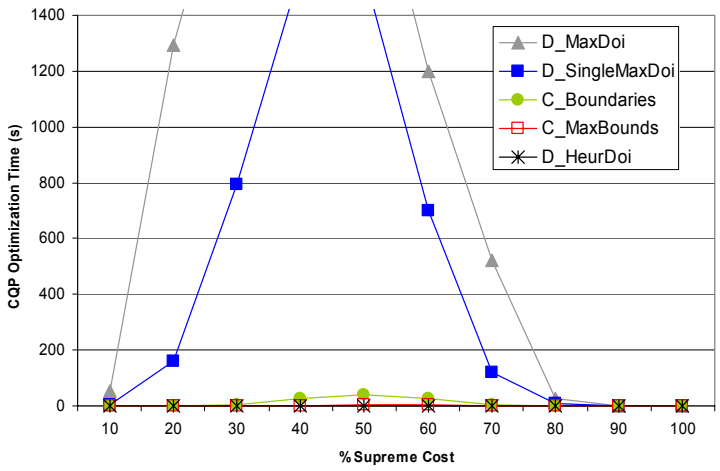
4 Πειράματα

Execution Times

$C_{μεγ}$ = % του χρόνου εκτέλεσης της ερώτησης που ενσωματώνει όλες τις προτιμήσεις (Supreme Cost)

K = 20

4 Πειράματα



4 Πειράματα

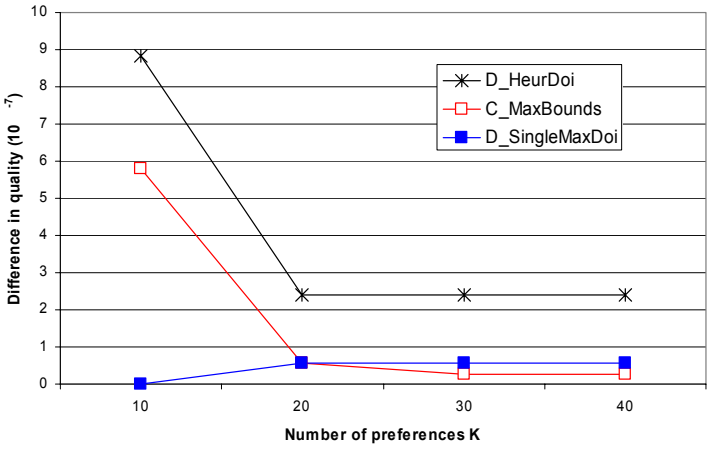
Quality of Solution

- $doi_{optimal}$ by C_Boundaries, D-MaxDoi
- Quality = $doi_{optimal}$ - doi_{found}
- $doi(P_x) = 1 - \prod_{p_i \in P_x}(1 - doi(p_i))$

K = # προτιμήσεων στο P

$C_{μεγ}$ = 400

4 Πειράματα



5 Σύνοψη

- Εισαγωγή της οικογένενειας προβλημάτων εξατομίκευσης υπό περιορισμούς
- Διατύπωση του προβλήματος ως πρόβλημα αναζήτησης σε χώρο καταστάσεων
- «Συντακτικές» μεταβάσεις
- Πειράματα

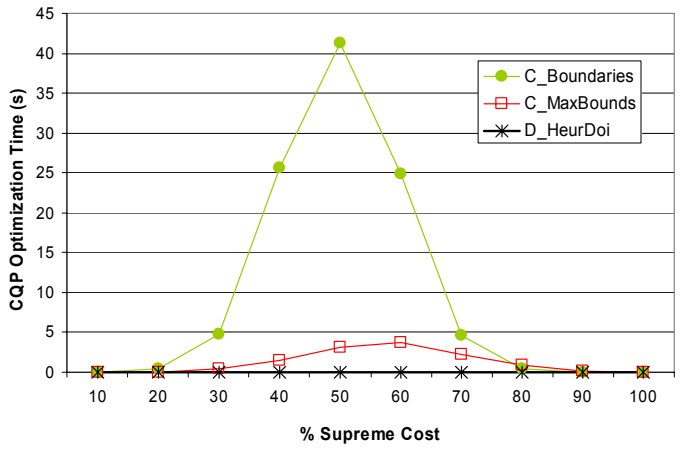
Μελλοντική Εργασία

- Πολιτικές για CQP προβλήματα
- Ενσωμάτωση με location-based services
- Multi-objective constrained optimization

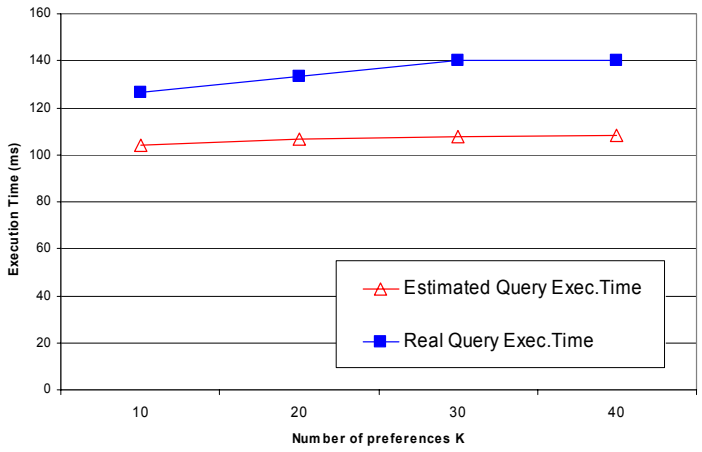
Τέλος

Ερωτήσεις …?

4 Experiments



4 Experiments



2 CQP System Architecture

Personalized Query Construction

Q: SELECT title FROM MOVIE

p_1, p_2 preferences

$q_1 : Q \wedge p_1$

$q_2 : Q \wedge p_2$

Personalized Query

```
SELECT title FROM q1 union all q2
GROUP BY title HAVING count(*) > 1
```

2 CQP System Architecture

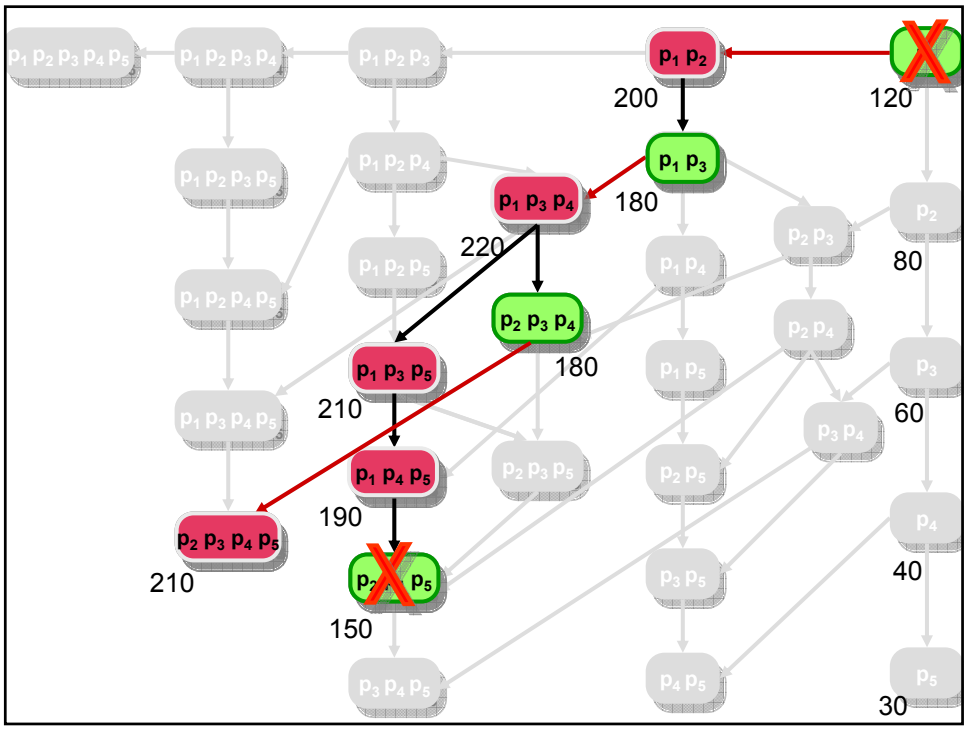
Parameter Estimation

Subsets of preferences: $P_x \subseteq P_y \subseteq P$

- $interest(Q \wedge P_x) \leq interest(Q \wedge P_y)$
- $result_size(Q \wedge P_x) \geq result_size(Q \wedge P_y)$

$Q_x := \cup\, q_i \Rightarrow cost(Q_x) = cost(\cup\, q_i) = \sum cost(q_i)$

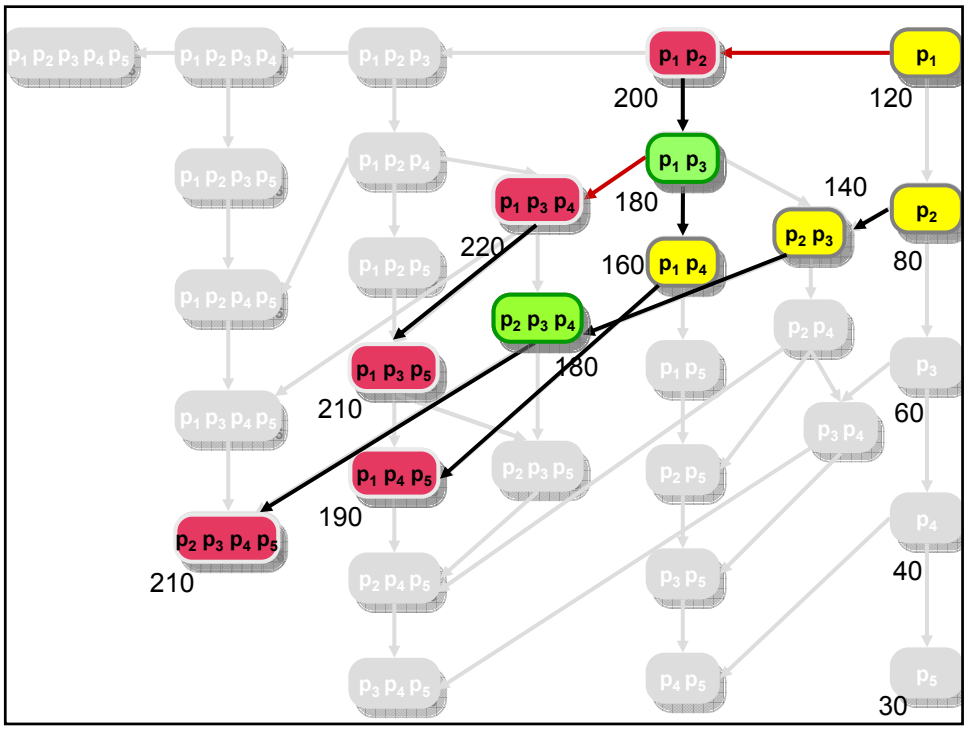
- $cost(Q \wedge P_x) \leq cost(Q \wedge P_y)$



3 CQP State Space Search

C_MaxBounds

Phase 1: find maximal boundaries

Phase 2: find optimal solution

e.g. maximize interest while cost < 185