

Πτυχιακές εργασίες για το ακ. έτος 2013-2014

Π. Βασιλειάδης

1. ΑΠΑΝΤΗΣΗ ΕΡΩΤΗΣΕΩΝ ΑΠΟ ΟΨΕΙΣ

ΠΕΡΙΛΗΨΗ: Ο σκοπός της παρούσας πτυχιακής είναι η υλοποίηση ενός εργαλείου που αποκρίνεται για το αν μια όψη μπορεί να απαντήσει μια ερώτηση SQL .

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η εργασία θα πρέπει να υλοποιήσει έτοιμους αλγορίθμους που αποκρίνονται για το αν μια όψη μπορεί να απαντήσει μια ερώτηση η οποία είναι ορισμένη σε απλούς πίνακες και επανεγγράφουν την ερώτηση στην όψη αυτή. Π.χ., η ερώτηση Q1 που ορίζεται πάνω στους πίνακες Calls, CallingPlans, μπορεί να ξαναγραφτεί χρησιμοποιώντας την υλοποιημένη όψη V1 (που είναι πολύ μικρότερη σε μέγεθος) αντί για τον πίνακα Calls (όπως φαίνεται στην ερώτηση Q1').

Q1: SELECT Year, Plan-Name, SUM(Charge) FROM Calls, CallingPlans WHERE Calls.PlanId = Calling-Plans.PlanId AND Year >= 1990 AND Year <= 1995 GROUPBY Year, Plan-Name HAVING SUM(Charge) > 1,000,000	V1: SELECT Plan-Id, Month, Year, SUM(Charge) FROM Calls GROUPBY Plan-Id, Month, Year	Q1': SELECT Year, Plan-Name, SUM(Earnings) FROM V1, Calling-Plans WHERE V1.Plan-Id = Calling_Plans.Plan_Id AND Year >= 1990 AND Year <=1995 GROUPBY Year, Plan-Name HAVING SUM(Earnings) > 1,000,000
--	---	---

Το αλγοριθμικό κομμάτι είναι έτοιμο, από μια ερευνητική εργασία που παρουσιάστηκε στο [VLDB 1996](#) και πρέπει απλώς να υλοποιηθεί. Το βασικό κομμάτι της πτυχιακής είναι η οργανωμένη επεξεργασία ερωτήσεων και όψεων στο κομμάτι του parsing. Θέλουμε να χρησιμοποιηθούν εργαλεία τελευταίας τεχνολογίας όπως π.χ. Antlr, για την επεξεργασία των ερωτήσεων. Επίσης, θέλουμε οι ερωτήσεις να αναπαριστώνται σε μια ενδιαμέση μορφή ώστε να είναι εύκολα ενσωματώσιμες σε άλλα εργαλεία.

ΠΡΟΚΛΗΣΕΙΣ και ΟΦΕΛΗ:

Η πρόκληση είναι στο κομμάτι του parsing, που πρέπει να γίνει με προσοχή, έλεγχο και επεκτασιμότητα στο μέλλον. Τα οφέλη για ένα φοιτητή είναι: (α) τεχνογνωσία στην τεχνολογία μεταφραστών τελευταίας γενιάς, και (β) πρακτική εμπειρία σε θέματα υλοποίησης αλγορίθμων, API, και parsers σε Java.

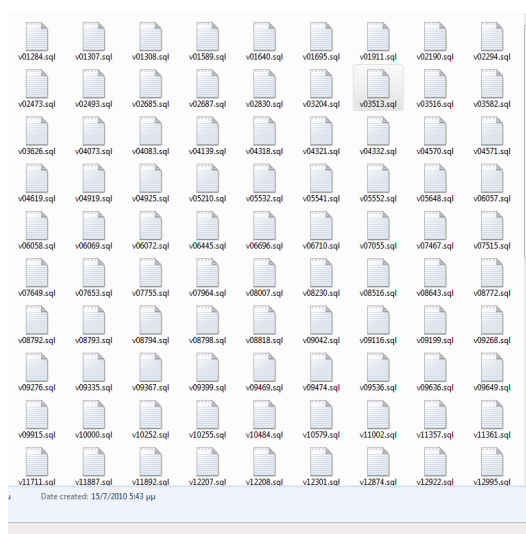
Η εργασία είναι πλέον κατάλληλη για φοιτητές με ταλέντο στον προγραμματισμό και ενδιαφέροντα στις βάσεις δεδομένων και την τεχνολογία μεταφραστών.

2. ΜΕΛΕΤΗ ΤΗΣ ΕΞΕΛΙΞΗΣ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

Μια βάση δεδομένων, από τη στιγμή που θα δημιουργηθεί, αλλάζει εσωτερική δομή με το πέρασμα του χρόνου: νέοι πίνακες δημιουργούνται, παλιοί καταστρέφονται, πεδία διαγράφονται, μετονομάζονται κλπ. Η διαδικασία αυτή ονομάζεται «εξέλιξη του σχήματος της βάσης δεδομένων» (schema evolution). Το εργαλείο **Hecate** [<https://github.com/giskou/Hecate>] μπορεί να συγκρίνει δύο σχήματα και να βρει τις διαφορές τους (κίτρινο: updated attributes, red: deletions, green: insertions).

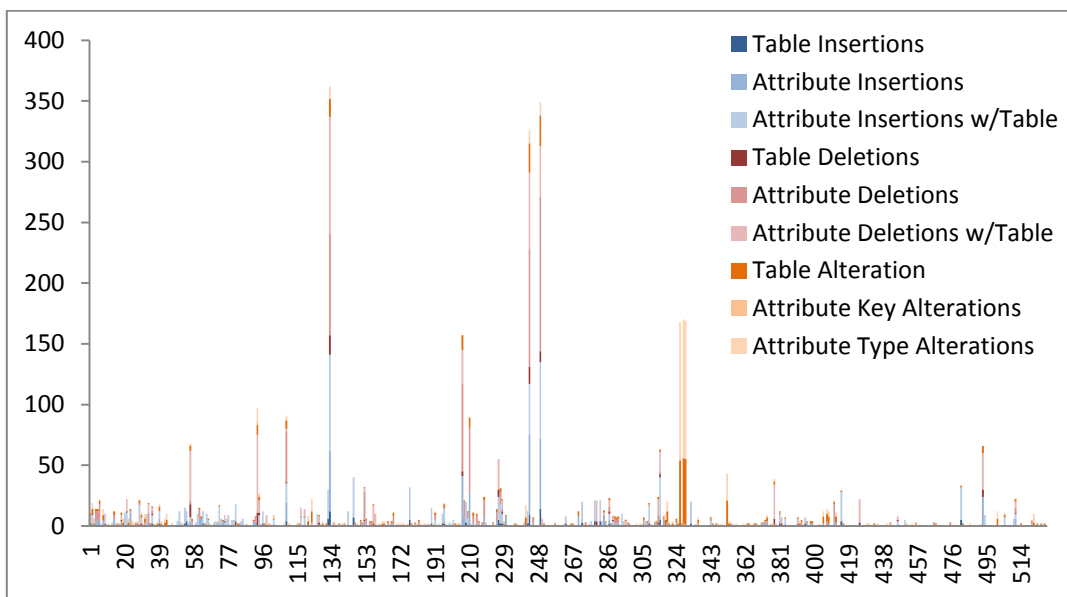
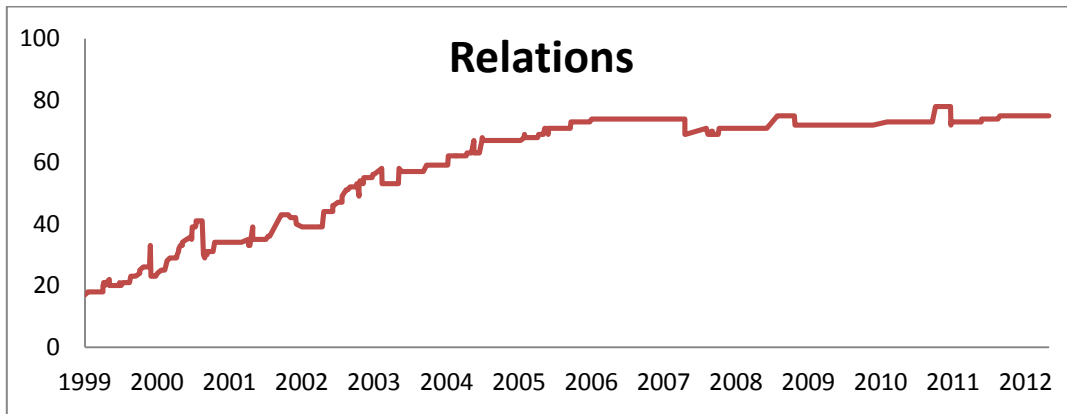
Name	Type	Name	Type
archive		group	
ar_comment	tinyblob	archive	
ar_flags	tinyblob	ar_comment	tinyblob
ar_minor_edit	tinyint(1)	ar_flags	tinyblob
ar_namespace	tinyint(2) unsigned	ar_minor_edit	tinyint(1)
ar_text	mediumtext	ar_namespace	tinyint(2) unsigned
ar_timestamp	char(14) binary	ar_text	mediumtext
ar_title	varchar(255) binary	ar_timestamp	char(14) binary
ar_user	int(5) unsigned	ar_title	varchar(255) binary
ar_user_text	varchar(255) binary	ar_user	int(5) unsigned
brokenlinks		ar_user_text	varchar(255) binary
cur		books	
image		brokenlinks	
imagelinks		categorylinks	
interviki		cur	
ipblocks		hitcounter	
links		image	
math		imagelinks	
old		interviki	
oldimage		ipblocks	
recentchanges		links	
searchindex		linkacc	
site_stats		logging	
user		math	
user_groups		userinactive	
user_rights		old	
watchlist		oldimage	
		querycache	
		recentchanges	
		searchindex	
		site_stats	
		user	
		user_groups	
		user_rights	
		validate	
		watchlist	

Επιπλέον, υπάρχουν αρκετές συλλογές από εκδόσεις του σχήματος της ίδιας βάσης (παρακάτω ένα screenshot από τη βάση της Wikimedia).



Η Εκάτη μπορεί να ταξινομήσει τις επί μέρους εκδοχές του σχήματος και να τις συγκρίνει διαδοχικά.

Έχουμε ήδη χρησιμοποιήσει την Εκάτη για να επεξεργαστούμε την εξέλιξη σχήματος διαφόρων βάσεων δεδομένων ανοιχτού λογισμικού, όπως για παράδειγμα, της βάσης της Wikimedia (της βάσης δεδομένων πίσω από τη Wikipedia), της βάσης του Atlas Trigger (του εργαλείου που διαχειρίζεται τα δεδομένα από το πείραμα Atlas για την ανεύρεση του μποζονίου του Χιγκς), της Ensembl (του εργαλείου για τη διαχείριση των δεδομένων του ανθρώπινου γονιδιώματος) και πολλών CMS's (opencart, corpermine, phpBB, tyro3, ...). Έχουμε επίσης συλλέξει την ιστορία από πολλά συστήματα ανοιχτού κώδικα που περιλαμβάνουν βάσεις δεδομένων και καταγράφουν και τις εκδοχές τους σε δημόσια αποθετήρια (κυρίως github, αλλά και svn) αλλά δεν την έχουμε επεξεργαστεί ακόμα.



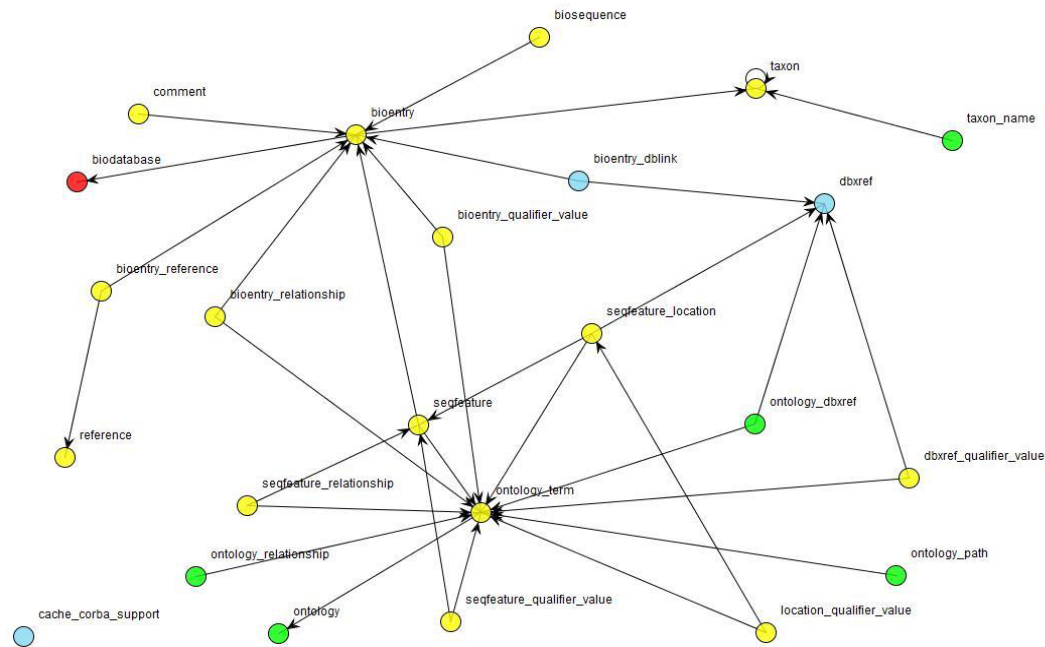
Στο παραπάνω σχήμα βλέπετε (α) το πώς εξελίχθηκε το μέγεθος του σχήματος της βάσης στο χρόνο και (β) τον παλμό των αλλαγών (το πώς διαρθρώθηκαν οι αλλαγές σε κάθε monitored version) για τη βάση Ensembl.

Η έρευνα στην περιοχή αυτή είναι θεμελιώδους φύσεως και αφορά στο να κατανοήσουμε την ύπαρξη προτύπων (ή ακόμα καλύτερα νόμων) για το πώς εξελίσσονται οι βάσεις δεδομένων με την πάροδο του χρόνου.

Table name	r...	l	U	D	r...	l	U	D	r...	l	U	D	r...	l	U	D	r...	l	U	D	r...	l	U	D	r...	l	U	D	r...	l	U	D		
archive	1	0	0	0	1	2	0	0	0	0	0	0	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
brokenlinks	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
cur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
image	0	1	0	0	7	0	0	0	0	0	0	0	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
imagelinks	0	1	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
ipblocks	3	0	0	0	4	2	0	0	1	3	0	0	3	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
links	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
math	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
old	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
oldimage	0	0	0	0	3	0	0	0	0	0	0	0	6	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
random	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
recentchanges	6	0	0	0	0	1	0	0	2	0	0	0	5	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
searchindex	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
site_stats	0	0	0	0	4	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
user	3	1	1	0	4	0	1	0	2	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
user_newtalk	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
watchlist	1	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
blobs	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
categorylinks	4	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
group	4	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
hitcounter	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
interwiki	3	0	0	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
linkscac	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
logging	8	0	0	0	0	1	0	0	1	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
objectcache	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
page	10	0	0	0	1	1	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
querycache	4	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
revision	7	0	0	0	2	0	0	0	0	0	0	0	2	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
text	3	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
user_groups	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
user_rights	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
validate	6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
externalinks	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
filearchive	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
job	0	0	0	0	5	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
langlinks	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pagelinks	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
querycache_info	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
templatelinks	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
trackbacks	0	0	0	0	6	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
transcache	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Το εργαλείο «Πλουτάρχου Βίοι Παράλληλοι» είναι ένα εργαλείο που έφτιαξε στην πτυχιακή του ο Θ. Γιάχος και το οποίο απεικονίζει την εξέλιξη των πινάκων μιας βάσης δεδομένων σε παράλληλες γραμμές. Κάθε version αναπαριστάται από 3 κολώνες για εισαγωγές, διαγραφές και ενημερώσεις πινάκων. Οι γεννήσεις πινάκων και πεδίων φαίνονται με πράσινο και οι διαγραφές με κόκκινο χρώμα.

Το εργαλείο «Παρμενίδια Αλήθεια» είναι ένα εργαλείο που αναπαριστά το σχήμα μιας βάσης δεδομένων με ένα διαχρονικό γράφημα και φροντίζει να οπτικοποιεί κάθε version και τις εκδοχές της σε ένα slide μιας Powerpoint παρουσίασης.



ΠΕΡΙΛΗΨΗ: Ο σκοπός της παρούσας πτυχιακής είναι η εμπειρισταωμένη, οργανωμένη και ελεγχόμενη ενσωμάτωση των παραπάνω εργαλείων σε ένα.

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα τρία εργαλεία θα θέλαμε να συνδυαστούν. Ο συνδυασμός θα πρέπει να γίνει στη βάση καλά ορισμένων use cases και τεχνικών integration. Επίσης, το νέο εργαλείο θα πρέπει να παράγει και δεδομένα από την ανάλυση των στοιχείων της εξέλιξης.

ΠΡΟΚΛΗΣΕΙΣ και ΟΦΕΛΗ:

Η πρόκληση είναι στο κομμάτι της καλά οριοθετημένης ενοποίησης. Τα οφέλη είναι: (α)

Η δυσκολία έγκειται στο κομμάτι της τεχνολογίας λογισμικού, ώστε να φτιαχτεί το λογισμικό ως μια μηχανή υψηλής ποιότητας. Τα οφέλη για ένα φοιτητή είναι: (α) τεχνογνωσία στην αναδόμηση κώδικα, (β) τεχνογνωσία σε θέματα εξέλιξης του σχήματος βάσεων δεδομένων, και (γ) πρακτική εμπειρία σε θέματα υλοποίησης αλγορίθμων, API, GUI σε Java.

Η εργασία είναι πλέον κατάλληλη για φοιτητές με ταλέντο στον προγραμματισμό και ενδιαφέροντα στις βάσεις δεδομένων και την τεχνολογία λογισμικού.

3. ΜΕΤΑΦΟΡΑ ΕΡΓΑΛΕΙΟΥ ETL ΣΕ ΣΥΝΕΡΓΑΣΙΑ ΜΕ ΣΧΕΣΙΑΚΟ ΣΔΒΔ

ΠΕΡΙΛΗΨΗ: Ο στόχος της εργασίας είναι η επέκταση ενός εργαλείου καθαρισμού και μετασχηματισμού δεδομένων με πληθώρα τελεστών

ΑΤΟΜΑ: 1

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java (για το νέο εργαλείο), .NET (υποχρεωτικά για το παλιό εργαλείο) + (ενδεχομένως) SQLite/BerkeleyDB/MySQL/...

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Τα εργαλεία ETL επιτρέπουν την μεταφορά, τον καθαρισμό και την μετατροπή δεδομένων από κάποιες «απομονωμένες» βάσεις δεδομένων σε κεντρικές αποθήκες δεδομένων. Στα πλαίσια της εργασίας MSc του, ο Α. Καραγιάννης ανέπτυξε ένα τέτοιο εργαλείο σε .NET και ο Δ. Λάμπρος έκανε μια πρώτη μεταφορά σε Java. Ο στόχος της πτυχιακής είναι να επεκταθεί η υλοποίηση Java με τελεστές συνένωσης, συνάθροισης κ.α., καθώς και να μπορούμε να συνδυάσουμε το εργαλείο και με κάποιο lightweight σύστημα διαχείρισης βάσεων δεδομένων όπως η SQLite ή η BerkeleyDB ή ..., το οποίο να μπορεί να ενσωματωθεί στον κώδικα.

ΠΡΟΚΛΗΣΕΙΣ και ΟΦΕΛΗ:

Η δυσκολία δεν έγκειται τόσο στην κατανόηση του κώδικα, που είναι ιδιαιτέρως καλά σχεδιασμένος, αλλά στην κατανόηση των λεπτομερειών των API των σχετικών ΣΔΒΔ και την μετατροπή του σχετικού κώδικα σωστά.

Τα οφέλη για ένα φοιτητή είναι: (α) τεχνογνωσία σε ένα challenging περιβάλλον εργασίας με αντικειμενοστρεφή προγραμματισμό (και ενδεχομένως, ένα μη παραδοσιακό ΣΔΒΔ), (β) τεχνογνωσία σε θέματα αποθηκών δεδομένων και ETL (η οποία έχει εξαιρετικά πρακτικό χαρακτήρα στην αγορά εργασίας), και (γ) πρακτική εμπειρία σε θέματα επέκτασης λογισμικού σε σχετικά χαμηλό επίπεδο. Επιπλέον, θα υπάρχει και το ηθικό όφελος της συνεισφοράς ενός εργαλείου που δουλεύει στην κοινότητα.

Η εργασία είναι πλέον κατάλληλη για φοιτητές με ταλέντο στον αντικειμενοστρεφή προγραμματισμό και ιδιαίτερα στην αποδοτική υλοποίηση λογισμικού σε σχετικά χαμηλό επίπεδο.

ΓΙΑ ΑΡΧΙΚΗ ΕΝΗΜΕΡΩΣΗ:

- Για αποθήκες δεδομένων και εργασίες ETL
http://www.cs.uoi.gr/~pvassil/downloads/ETL/SHORT_DESCR/
http://www.cs.uoi.gr/~pvassil/courses/db_III/notes/Lecture10-DataWarehouses.ppt
- Για το εργαλείο Άρκτος
http://www.cs.uoi.gr/~pvassil/projects/arktos_II/index.html
- Το MSc του Α. Καραγιάννη και η σχετική δημοσίευση (δείτε το paper@QDB και την παρουσίαση του Τάσου στο MSc του)
http://www.cs.uoi.gr/~pvassil/publications/2011_QDB/