

Γενικές αρχές εκπόνησης επισκοπήσεων της βιβλιογραφίας (surveys)

Π. Βασιλειάδης

Τετάρτη, 3 Οκτωβρίου 2007

http://www.cs.uoi.gr/~pvassil/linx/localCopies4grads/Survey_guidelines_byPV.pdf

1. Εισαγωγικά

Η διαδικασία επισκόπησης της βιβλιογραφίας για ένα ερευνητικό τομέα είναι μια εξαιρετικά επίπονη διαδικασία. Η καταγραφή της εν λόγω επισκόπησης με τρόπο χρήσιμο στους αναγνώστες είναι μια διαδικασία που απαιτεί τέχνη, αφαιρετική ικανότητα και μέθοδο. Οι εν λόγω οδηγίες απευθύνονται στους μεταπτυχιακούς φοιτητές που εμπλέκονται στις παραπάνω διαδικασίες και σκοπό έχουν να παρουσιάσουν μια μέθοδο δουλειάς, αφενός, και οδηγίες παρουσίασης αφετέρου.

Το εν λόγω κείμενο συγγράφεται ως εάν το αντικείμενο εργασίας είναι η εκπόνηση μιας βιβλιογραφικής επισκόπησης (survey) στο χώρο των βάσεων δεδομένων¹. Οι οδηγίες ισχύουν περίπου अपαράλλαχτες και για το απλούστερο έργο της συγγραφής του κεφαλαίου «Σχετικές Εργασίες» σε μια εργασία μεταπτυχιακής εξειδίκευσης. Στην δεύτερη αυτή περίπτωση, το ερευνητικό αντικείμενο είναι σαφώς πιο περιορισμένο και η επισκοπούμενη βιβλιογραφία σημαντικά πιο ολιγάριθμη.

2. Αντικείμενο και σκοπός της βιβλιογραφικής επισκόπησης

Η βιβλιογραφική επισκόπηση, η οποία διεξάγεται σε μεγάλο εύρος στην αρχή της εκπόνησης μιας διδακτορικής διατριβής ή μιας μεταπτυχιακής εργασίας (και σε μικρότερο εύρος, στο υπόλοιπο της ζωής σας), αποτελεί μια οργανωμένη καταγραφή (α) της ερευνητικής περιοχής που εξετάζεται, (β) των κρίσιμων τεχνικών ζητημάτων που εμπλέκονται σε αυτή, και (γ) μεθόδων αντιμετώπισης των εν λόγω τεχνικών προβλημάτων, όπως αυτές έχουν παρουσιαστεί στη διεθνή βιβλιογραφία.

Πρέπει, δηλαδή, να θυμάστε, ότι **η βιβλιογραφική επισκόπηση ως πρώτιστο στόχο έχει να σας βοηθήσει να οργανώσετε στο μυαλό σας, αφενός ποια είναι τα κρίσιμα ζητήματα, και αφετέρου, πώς τα αντιμετωπίζουμε.**

¹ Δες ΠΑΡΑΡΤΗΜΑ Α για τις διαφορές σε σχέση με το χώρο της εννοιολογικής μοντελοποίησης (conceptual modeling)

Άμεσο απότοκο του παραπάνω στόχου είναι και ο εντοπισμός προβλημάτων που η ερευνητική κοινότητα ΔΕΝ έχει αντιμετωπίσει επιτυχώς στο παρελθόν.

Η οργάνωση της υπάρχουσας βιβλιογραφίας ΔΕΝ ταυτίζεται με την καταγραφή της. Στη γενική περίπτωση, οι δύο διαδικασίες είναι ξεχωριστές, αν και αλληλοεξαρτώμενες. Η παρουσίαση της υπάρχουσας βιβλιογραφίας απαιτεί την ομαδοποίηση των δημοσιευμένων εργασιών σε επί μέρους θέματα, την αφαιρετική εξαγωγή κοινών χαρακτηριστικών ανά ομάδα, αλλά και τον εντοπισμό των ιδιομορφιών της κάθε εργασίας.

Όπως, θα δούμε και παρακάτω, η οργάνωση και η καταγραφή είναι διαδικασίες διεξάγονται σε μεγάλο βαθμό εν παραλλήλω: για κάθε άρθρο που διαβάζουμε

- (i) **οφείλουμε να καταγράψουμε μια περίληψή του, ΑΜΕΣΩΣ μετά την ανάγνωση,**
- (ii) αυτομάτως αναδιοργανώνουμε (αν μη τι άλλο στο μυαλό μας) την εικόνα της ευρύτερης ερευνητικής περιοχής.

Checklist: Στην τελική της μορφή, η επισκόπηση έχει σκοπό να παρουσιάσει στον αναγνώστη, **με δομημένο τρόπο**, την οργάνωση της περιοχής που έχετε καταστρώσει στο μυαλό σας – και συγκεκριμένα,

- ποια είναι τα **γενικά χαρακτηριστικά** του ευρύτερου προβλήματος που διερευνούμε
- ποιες είναι οι **σχετικές υποπεριοχές** (γι' αυτό και φτιάχνουμε μια υποενότητα για κάθε τέτοια περιοχή),
- ποιες είναι οι **κρίσιμες εργασίες** σε κάθε περιοχή και τι κάνουν.
- **πώς συνοψίζεται η εικόνα της ερευνητικής προσπάθειας μέχρι στιγμής στην παγκόσμια βιβλιογραφία**
- **τι ζητήματα μένουν ανοιχτά προς διερεύνηση.**

Μπορείτε να αναφερθείτε σε εξαιρετικές βιβλιογραφικές επισκοπήσεις τις οποίες αναφέρω στην ενότητα «Αναφορές» ως υποδειγματικές εργασίες επί του θέματος.

3. Μέθοδος δουλειάς

3.1 Εκκίνηση της συλλογής της βιβλιογραφίας

Τρεις είναι οι πηγές από τις οποίες μπορείτε να ξεκινήσετε τη συλλογή του υλικού:

- (α) ο επιβλέπων καθηγητής σας
- (β) το DBLP,

(γ) το Google.

Οφείλετε να τις χρησιμοποιήσετε και τις τρεις!

Επίσης, οφείλετε να δώσετε ιδιαίτερη σημασία στην βιβλιογραφία των τελευταίων ετών. Είναι περίπου υποχρεωτική η αναζήτηση σχετικών εργασιών τα **τελευταία 5-6 χρόνια στα συνέδρια VLDB, SIGMOD, ICDE, EDBT**. Για λόγους πληρότητας, αναφέρουμε και τα περιοδικά ACM TODS, VLDBJ, Information Systems, IEEE TKDE στα οποία συνήθως βρίσκονται εκτεταμένες εκδοχές των άρθρων που δημοσιεύονται στα παραπάνω συνέδρια.

Τα περιοδικά ACM Computing Surveys (CSUR) και VLDB Journal (VLDBJ) είναι (το πρώτο εξ αντικειμένου) τα σημεία στα οποία μπορεί να ανατρέξει κανείς για άλλες, σχετικές επισκοπήσεις. Επίσης, είναι πολύ χρήσιμο να ανατρέξει κανείς **στα tutorials των προαναφερθέντων μεγάλων συνεδρίων για βάσεις δεδομένων**: πολύ συχνά, υπάρχουν στο διαδίκτυο **οι διαφάνειες από τα tutorials** που συνοψίζουν έξοχα τη βιβλιογραφία και τα ζητήματα σε κάποια ερευνητική περιοχή, έστω και αν δεν συνοδεύονται από κάποιο κείμενο.

Σημείωση: Στις βάσεις δεδομένων, το κατ' εξοχήν σημείο αναφοράς για τη διεθνή βιβλιογραφία είναι το site του **DBLP**, που συντηρείται από τον M. Ley στο <http://www.informatik.uni-trier.de/~ley/db/index.html> καθώς και τα mirrors του <http://www.vldb.org/dblp/db/index.html> <http://www.acm.org/sigmod/dblp/db/index.html>

3.2 Προοδευτική επέκταση της επισκοπούμενης βιβλιογραφίας

Δοθέντων κάποιων αρχικών αναφορών που έχουμε συλλέξει από την παραπάνω διαδικασία, σιγά-σιγά αρχίζουμε και εξερευνούμε τη βιβλιογραφία.

Γράφε όπως διαβάζεις. Ο βασικός σκοπός ανάγνωσης μιας εργασίας είναι η διάγνωση της θεμελιώδους ουσίας του άρθρου. Στην αγγλική, ο όρος «θεμελιώδης ουσία» αποδίδεται με τη λέξη gist, εζ' ου και η αρχή **GIST**:

<u>G</u>iven	a set of assumptions
<u>I</u>ntroduce	a technique / algorithm / data structure / theorem / ...
<u>S</u>uch that	a goal is achieved / certain measures are optimized (y-axis) / certain qualities are verified (essence of a theorem)
<u>T</u>ested	over a set of configuration parameters (x-axis)

Κάθε φορά που διαβάζουμε μια εργασία, ο πρώτος στόχος είναι να απαντήσουμε το παραπάνω ερώτημα για το άρθρο. Στη συνέχεια, δίνουμε ιδιαίτερη έμφαση στην προτεινόμενη τεχνική (“Introduce” – “I”). Αν το άρθρο μας ενδιαφέρει πολύ, πρέπει να κατανοήσουμε και την ουσία των πειραμάτων (Evaluate the effect of the x-axis to the value of the y-axis).

Η τεχνική «γράφει όπως διαβάζεις» λέει ότι **εν παραλλήλω με την παραπάνω διαδικασία κατανόησης, διεξάγεται και η διαδικασία καταγραφής:**

- Γράψτε το **GIST** του άρθρου σε μία – δύο προτάσεις.
 - Γράψτε σε τεχνικό ύφος το **πώς δουλεύει** η προτεινόμενη τεχνική (I). ΠΡΟΣΟΧΗ: αυτό είναι το πιο δύσκολο σημείο + εδώ απαιτείται έμφαση στην βαθιά κατανόηση του άρθρου.
 - Προσθέστε και οποιαδήποτε άλλα σχόλια, π.χ., για τα πειράματα ή τα ευρήματα.
- ... και ...
- Σιεφτείτε πολύ σοβαρά **τι λείπει από το άρθρο** και καταγράψτε το. Ασχέτως αν θα περιλάβετε τις ελλείψεις στην επισκόπηση, το σημείο αυτό σας βοηθά να κατανοήσετε τι μπορείτε να κάνετε εσείς.

Στη διάρκεια αυτής της διαδικασίας απαιτείται **ΑΥΣΤΗΡΗ ΤΥΠΙΚΟΤΗΤΑ στον τρόπο με τον οποίο καταχωρούμε τις αναφορές στο τέλος του κειμένου** (βλ., αναλυτικές οδηγίες στο κείμενο [Βασι06]).

Δεν είναι όλα τα πόνια ίδια. Στο σκάκι, ανάλογα με τη διαμόρφωση του αγώνα, κάποια πόνια αποδεικνύονται πιο σημαντικά από άλλα (λόγω αμυντικού ρόλου, κάλυψης χώρου κλπ). Το ίδιο ισχύει και με τις εργασίες: όταν ετοιμάζετε να επιτεθείτε σε κάποιο ερευνητικό πρόβλημα, κάποιες εργασίες είναι πιο σημαντικές από άλλες, κυρίως λόγω συνάφειας, αλλά και λόγω των τεχνικών που σας προσφέρουν έτοιμες προς επαναχρησιμοποίηση.

Στην αρχή, αφενός λόγω απειρίας, και αφετέρου λόγω έλλειψης συγκεκριμένου ερευνητικού στόχου, είστε υποχρεωμένοι να θεωρείτε όλες τις εργασίες «ισότιμες». Στην πορεία, όσο θα αποκτάτε εμπειρία, το διακριτικό σας κριτήριο θα οξύνεται. Έτσι, ενώ είναι πολύ σημαντικό να μπορείτε να ξεχωρίσετε τις σημαντικές εργασίες για τη στόχευση της έρευνάς σας, πρέπει να συμβιβαστείτε με το γεγονός ότι είναι τρομακτικά δύσκολο να αποφασίσει κανείς ποιες εργασίες είναι οι πιο σημαντικές και για να το κάνει, απαιτείται χρόνος, χρόνος, χρόνος και ... χρόνος.

Σε ότι αφορά την συγγραφή μιας επισκόπησης, όμως, το παραπάνω πρόβλημα δεν είναι και τόσο σημαντικό: αυτό που για τη δική σας έρευνα είναι σημαντικό, μπορεί να είναι οριακά σχετικό για κάποιο άλλο ερευνητικό πρόβλημα. Κατά συνέπεια, όταν συγγράφετε την επισκόπηση, βάλτε τα πιόνια ισότιμα στη σκακιέρα και αφήστε τον καθένα να παίζει την παρτίδα όπως νομίζει.

Το καλάθι με τα κεράσια. Η βιβλιογραφία είναι σαν το καλάθι με τα κεράσια. Τραβάς ένα και σηκώνονται άλλα πέντε. Η αυτονόητη μεταφορά στο χώρο των ερευνητικής βιβλιογραφίας απλώς μεταφράζεται στο ότι ένα άρθρο αναφέρει κάποια προηγούμενη εργασία, η οποία είναι ενδεχομένως άξια διερεύνησης. Επίσης, για μια σημαντική εργασία απαιτείται ιδιαίτερη προσοχή και στον εντοπισμό εργασιών που την αναφέρουν στη βιβλιογραφία τους (για το σκοπό αυτό μπορείτε να χρησιμοποιήσετε το DBLP για παλαιές εργασίες και κυρίως τα συστήματα Google Scholar και Citeseer).

Προφανώς **ΔΕΝ ΥΠΑΡΧΕΙ ΚΡΙΤΗΡΙΟ ΤΕΡΜΑΤΙΣΜΟΥ** της διαδικασίας. Τα κεράσια είναι άπειρα και το πού σταματά κανείς είναι ad hoc. Χονδρικά, και χωρίς αυτό να είναι δεσμευτικό, μια επισκόπηση οφείλει να έχει 25-30 άρθρα ως κεντρικό της κορμό (με αποτέλεσμα οι συνήθεις επισκοπήσεις να αναφέρονται σε 100-200 άρθρα, συνολικά). Μια εργασία MSc, χονδρικά, είναι τυπικά καλυμμένη με την βαθειά κατανόηση 15-20 άρθρων. Επαναλαμβάνω, όμως, ότι τα παραπάνω νούμερα είναι ενδεικτικά και δείχνουν την τάξη μεγέθους στη συνήθη περίπτωση.

3.3 Αφαιρετική σύνοψη, ομογενοποίηση και εκκαθάριση των επί μέρους σημειώσεων

Κάποια στιγμή, ο καιρός είναι ώριμος να οργανώσετε τις καταχωρήσεις σας σε κατηγορίες. Αυτή είναι μια guided bottom-up διαδικασία: αν και εξ' αρχής υπάρχει μια πολύ αδρή

ομαδοποίηση των εργασιών σε μεγάλες ομάδες, οι επί μέρους υπο-ομάδες δεν είναι εμφανείς και πρέπει να καθοριστούν. Το GIST τεστ παρέχει ασφαλή κριτήρια για να γίνει αυτό: **G, I, S, με την αντίστροφη σειρά.**

- Οι ευρύτερες ομάδες συνήθως προκύπτουν από το τελικό στόχο (S)
- Οι τεχνικές ανά στόχο ομαδοποιούνται ανάλογα με τα μέσα που χρησιμοποιούμε (I)
- Οι επί μέρους εργασίες διαφοροποιούνται ανάλογα με τις υποθέσεις που κάνουν (G), ή/και τις μετρικές που προσπαθούν να βελτιστοποιήσουν (S)

Ξανά πρέπει να τονιστεί ότι η ομαδοποίηση δεν πετυχαίνει με την πρώτη και απαιτεί χρόνο για να ωριμάσει.

Βελτιώσεις. Κατά τη διάρκεια της συγγραφής της επισκόπησης με οργανωμένο τρόπο, απαιτείται να ομογενοποιηθούν και να καθαρισθούν ορισμένες επί μέρους σημειώσεις. Η προσθήκη παραδειγμάτων βοηθά πάντα τον τελικό αναγνώστη να καταλάβει ένα επί μέρους πρόβλημα. Συνολικά, **ο στόχος είναι να παραχθεί ένα ενιαίο, ομογενές κείμενο και όχι μια απλή συρραφή των επί μέρους σημειώσεων.**

3.4 Συντήρηση

Όσο περνά ο καιρός –ιδίως για όσους εκπονούν διδακτορική διατριβή- η ικανότητά σας να ξεχωρίζετε τις σχετικές από τις μη σχετικές εργασίες θα εξασκείται. Το GIST τεστ θα γίνεται σιγά-σιγά όλο και πιο εύκολο, και η ικανότητά σας να αποτιμάτε τη σημαντικότητα μιας εργασίας θα οξύνεται. Όσο περνά ο καιρός, πρέπει να μην ξεχνάτε:

- το «γράφει όπως διαβάζεις» είναι ένα παιχνίδι που συνεχίζεται εφ' όρου ζωής – συνεπώς, **οφείλετε να συνεχίζετε να καταγράφετε τις περιλήψεις των αναγνωσθέντων άρθρων.**
- η μορφή του ερευνητικού τοπίου αλλάζει: προβλήματα που κάποτε ήταν σημαντικά χάνουν τη σημασία τους και προκύπτουν άλλα στη θέση τους – συνεπώς, η δομή του ερευνητικού τοπίου που έχετε στο μυαλό σας ενδεχομένως χρειάζεται ανανέωση επίσης.

Αυτό που πρέπει να θυμάστε είναι ότι **η σχέση σας με τη βιβλιογραφία είναι παντοτινή και οι βασικές τεχνικές για να τη διαχειρίζεστε δεν αλλάζουν!**

4. Γενική Δομή της Επισκοπήσεως

Τα κεφάλαια της επισκοπήσεως έχουν περίπου την εξής δομή:

1. Εισαγωγή
2. Θεμελιώδεις έννοιες και παράδειγμα αναφοράς
3. Κατηγοριοποίηση των βασικών εννοιών και της σχετικής εργασίας
4. ... Παρουσίαση της σχετικής δουλειάς ανά κατηγορία ...
5. Σχετικά projects
6. Κατανομή των εργασιών σε κατηγορίες και timeline των δημοσιεύσεων
7. Ανακεφαλαίωση – Μελλοντική δουλειά

Στη συνέχεια, αναλύονται σε περισσότερη λεπτομέρεια οι γενικές αρχές του κάθε κεφαλαίου.

5. Εισαγωγή

Η εισαγωγή σκοπό έχει να παρουσιάσει (α) το γενικότερο πρόβλημα που πάτε να λύσετε, (β) γιατί αυτό είναι ενδιαφέρον, σημαντικό, καινοτόμο και όχι προφανές στη λύση του, (γ) πού είναι χρήσιμο στον πραγματικό κόσμο, και (δ) την οργάνωση του κειμένου στις επόμενες ενότητες (σε μεγαλύτερη ανάλυση απ' ότι όταν γράφουμε το roadmap ενός άρθρου).

Η εισαγωγή απαντά το ερώτημα “why bother?” με απλό τρόπο, χωρίς ιδιαίτερες τυπικότητες και εξηγεί στον αναγνώστη πού βρίσκεται κάθε τι (κυρίως για τον αναγνώστη που θα διαβάσει το άρθρο μετά από καιρό ψάχνοντας για κάποιο πολύ συγκεκριμένο πρόβλημα). Επίσης, είναι σύνηθες να αναφέρονται στην εισαγωγή και άλλες **σχετικές επισκοπήσεις του παρελθόντος**, καθώς και πώς διαφοροποιείται η παρούσα επισκόπηση από αυτές.

6. Θεμελιώδεις έννοιες και παράδειγμα αναφοράς

Το εν λόγω κεφάλαιο, έπεται της εισαγωγής και **ορίζει τις θεμελιώδεις έννοιες με τεχνικό ύφος και αυστηρό τρόπο.**

ΔΕΝ είναι απαραίτητο να είστε μαθηματικά αυστηροί, ούτε να δώσετε αυστηρούς μαθηματικούς ορισμούς. Αν μπορείτε, όμως, ως εκ της φύσεως του προβλήματος και της συνοχής της βιβλιογραφίας, η αυστηρότητα είναι πολύ καλό στοιχείο (στη γενική περίπτωση, σε πολύ ευρεία ζητήματα η βιβλιογραφία δεν είναι ιδιαίτερα συνεπής ως προς

τον ορισμό των βασικών εννοιών, με αποτέλεσμα να υπάρχει αδυναμία απόδοσης τυπικών ορισμών).

Το κεφάλαιο οφείλει να αφιερώσει σημαντικό χώρο στον ορισμό ενός **παραδείγματος αναφοράς** (π.χ., το σχήμα μιας βάσης αναφοράς) και επί μέρους παραδειγμάτων (π.χ., ερωτήσεων επί του σχήματος) που εξηγούν γιατί είναι σημαντικό το πρόβλημα που εξετάζεται.

Στη συνέχεια της επισκόπησης, αν οι γενικές υποκατηγορίες το επιτρέπουν, η επιστροφή στο παράδειγμα αναφοράς με επί μέρους βοηθητικά παραδείγματα είναι πολύ χρήσιμη για την κατανόηση των εννοιών.

Στο χώρο των βάσεων δεδομένων, επειδή καλώς ή κακώς η έρευνα είναι επικεντρωμένη στην επέκταση ενός DBMS με τεχνικές για το υπό εξέταση πρόβλημα, συχνά συναντάμε και μια υπο-ενότητα «**Αρχιτεκτονική**», στην οποία και εμφανίζονται τα επί μέρους υποσυστήματα που παίζουν κάποιο ρόλο στην αντιμετώπιση του προβλήματος.

7. Κατηγοριοποίηση των βασικών εννοιών και της σχετικής εργασίας

Η ενότητα αυτή συγγράφεται προς το τέλος, και για την ακρίβεια, όταν ωριμάσει ο καιρός για την ομαδοποίηση των εργασιών (βλ. και υποενότητα 3.3 παραπάνω). Συνήθως, η ενότητα αυτή περιλαμβάνει

- Τα ζητήματα αιχμής που τελικώς διαμορφώνουν τις υπό εξέταση κατηγορίες
- Αν είναι εφικτό (συνήθως είναι), μια δενδρική ιεραρχία των ζητημάτων

Ο εντοπισμός και η κατασκευή των παραπάνω είναι ad hoc και είναι δύσκολο να δώσει κανείς ιδιαίτερες οδηγίες, πέραν αυτών που δίδονται στην υποενότητα 3.3 και της **υπομονής**: απαιτείται να έχει κατανοηθεί και αποδελτιωθεί ένας σημαντικός όγκος δημοσιεύσεων για να γίνει εφικτή η ιεράρχηση εννοιών και προβλημάτων.

Όταν η κατηγοριοποίηση έχει συντελεσθεί, το **τοπίο της έρευνας είναι πολύ πιο καθαρό και οργανωμένο**. Παρ' όλα αυτά, όσο η κατανόησή του εμβαθύνεται και εμπλουτίζεται, η **εικόνα αυτή αλλάζει** (και κατά συνέπεια και η εν λόγω ενότητα). Έτσι, **μην περιμένετε ότι η ενότητα θα γραφτεί με την πρώτη προσπάθεια**.

8. Κατανομή των εργασιών στις κατηγορίες

Ένα εξαιρετικά χρήσιμο εργαλείο για την ερευνητική κοινότητα, ή έστω, για τους αναγνώστες της επισκόπησής σας, είναι **το πίνακας κατανομής της βιβλιογραφίας** που δείχνει **ποιες εργασίες εντάσσονται σε κάθε κατηγορία**. Σε περιπτώσεις που η βιβλιογραφία είναι μικρή (π.χ., σε κάποια τεχνικά άρθρα, ή MSc's), γίνεται και το αντίστροφο: για κάθε άρθρο υπό εξέταση, δείχνουμε ποια προβλήματα/χαρακτηριστικά της ευρύτερης περιοχής επιλύονται/ικανοποιούνται από το άρθρο.

Ακόμα, μια παλαιότερη τεχνική την οποία μάλλον έχουμε ξεπεράσει, είναι **το timeline των δημοσιεύσεων**: σε ένα οριζόντιο άξονα του χρόνου, τοποθετούνται οι δημοσιεύσεις ώστε να φαίνεται ποια επεκτείνει ποιες προηγούμενες. Η εν λόγω ταξινόμηση έχει ατονήσει στις μέρες μας, κυρίως λόγω του όγκου της βιβλιογραφίας και την χρονική πυκνότητα των δημοσιεύσεων.

Πολλές φορές η εν λόγω ενότητα συγχωνεύεται με την ενότητα της κατηγοριοποίησης των εννοιών. Η απόφαση είναι θέμα γούστου – εμείς εδώ τις αναφέρουμε χωριστά για να τονίσουμε τη χρησιμότητα του πίνακα κατανομής της βιβλιογραφίας.

9. Περιγραφή των εργασιών

Ανάλογα με τον αριθμό των κατηγοριών, αφιερώνουμε και μια ενότητα ή υποενότητα ανά κατηγορία. Τρεις-τέσσερις κύριες ενότητες είναι μια καλή οδηγία (πράγμα που εκ των υστέρων μπορεί να επηρεάσει και την ιεραρχία των κατηγοριών), χωρίς φυσικά αυτό να είναι υποχρεωτικό.

Οι εργασίες που εντάσσονται σε κάθε κατηγορία περιγράφονται όπως είπαμε ήδη στην ενότητα 3. Επιπλέον, παραδείγματα ή περαιτέρω σχόλια επί της συσχέτισης μιας εργασίας με την (υπο)κατηγορία της είναι συχνά ενδεδειγμένα.

10. Σχετικά projects

Είναι σύνηθες, ως εκ του εφαρμοσμένου χαρακτήρα του πεδίου των βάσεων δεδομένων, να αφιερώνεται μια –μάλλον εκτενής για την αξία της- ενότητα στα σχετικά projects που έχουν λάβει χώρα. Για κάθε τέτοιο project καταγράφονται οι εμπλεκόμενοι φορείς, τα κύρια ζητήματα που αντιμετωπίστηκαν και οι κύριες δημοσιεύσεις. Η βασική συμβολή της ενότητας, τελικά, είναι να δείξει τη συνοχή κάποιων εργασιών και τη ροή τους στο χρόνο.

11. Ανακεφαλαίωση και μελλοντική δουλειά

Οδηγίες για την ανακεφαλαίωση θα βρείτε στο [Βασι06]. Το κομμάτι της **μελλοντικής δουλειάς**, παρ' όλα αυτά, είναι πολύ σημαντικό, όταν βρίσκεστε στην αρχή της MSc/PhD εργασίας: εδώ είναι που καταγράφονται –καλύτερα: οργανώνονται- οι ερευνητικές ευκαιρίες και τα σημεία στα οποία μπορεί να διεξαχθεί περαιτέρω έρευνα. **Αντιμετωπίστε το κομμάτι της μελλοντικής δουλειάς με πολύ προσοχή!**

(Ξέρω ότι το τι θα γράψετε στο άρθρο σας, είναι συνήθως ένα υποσύνολο του τι έχετε εντοπίσει. Εντούτοις, οργανώστε τις ερευνητικές ευκαιρίες κι ας μην τις γράψετε πλήρως...)

12. Αναφορές

- [AnSp04] Stephanos Androutsellis-Theotokis, Diomidis Spinellis. A survey of peer-to-peer content distribution technologies. ACM Computing Surveys 36(4), pp. 335-371, 2004
- [Bab+02] Brian Babcock, Shivnath Babu, Mayur Datar, Rajeev Motwani, Jennifer Widom. Models and Issues in Data Stream Systems. In Proceedings of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS 2002), pp. 1-16, June 3-5, Madison, Wisconsin, USA. ACM 2002
- [Hale01] Alon Y. Halevy. Answering queries using views: A survey. VLDB J. 10(4), pp. 270-294, 2001.
- [Koss00] Donald Kossmann: The State of the art in distributed query processing. ACM Computing Surveys 32(4), pp. 422-469, 2000
- [RaBe01] Erhard Rahm, Philip A. Bernstein. A survey of approaches to automatic schema matching. VLDB J. 10(4), pp. 334-350, 2001
- [Βασι06] Π. Βασιλειάδης. Γενικές αρχές εκπόνησης και συγκρότησης της εργασίας σας για το MSc, 2006. Διαθέσιμες στο http://www.cs.uoi.gr/~pvassil/linux/localCopies4grads/MSc_guidelines_byPV.pdf

ΥΠΟΜΝΗΣΗ: Δέστε ΥΠΟΧΡΕΩΤΙΚΑ τις οδηγίες για το πώς αναφερόμαστε στη βιβλιογραφία στο ΠΑΡΑΡΤΗΜΑ Β (επίσης, στο [Βασι06])!!

ΠΑΡΑΡΤΗΜΑ Α

Στο χώρο της εννοιολογικής μοντελοποίησης, το αντικείμενο ΔΕΝ είναι συνήθως η αλγοριθμική επίλυση κάποιου τεχνικού προβλήματος. Αντίθετα, ο κεντρικός στόχος είναι η παροχή (γραφικών ή τυπικών) γλωσσών/μοντέλων, δια των οποίων ο σχεδιαστής διευκολύνεται στην προσπάθειά του να σχεδιάσει κάποιο (υπο)σύστημα, κάτω από μια συγκεκριμένη οπτική γωνία. Ως εκ τούτου, αντί του GIST, το ζητούμενο μπορεί να εκφρασθεί ως ακολούθως (GUIs):

<u>Given</u>	a certain design goal
<u>Under</u>	a particular viewpoint of a designer
<u>Introduce</u>	a modeling technique (model, language, ...)
<u>Such that</u>	specific tasks are performed (easier than before)

Επειδή οι τεχνικές σχεδίασης που έχουμε ήδη είναι αρκετά ισχυρές, είναι δύσκολο να βρει κανείς περιπτώσεις που μια νέα τεχνική κατορθώνει κάτι που πριν ήταν αδύνατο. Αντίθετα, αυτό που γίνεται, είναι ότι οι νέες τεχνικές επιτρέπουν την ευκολότερη σχεδίαση ενός συστήματος, έχοντας υπόψη μια συγκεκριμένη οπτική γωνία. Σκεφθείτε κάθε μοντέλο (π.χ., ER, DFD, UML, κλπ) ως ένα αλφάβητο που σας δίνει γράμματα (οντότητες, συσχετίσεις, κλάσεις, πεδία, κλπ) για να σχηματίσετε προτάσεις (διαγράμματα) στη βάση μιας γλώσσας (μέθοδος χρήσης). Αυτό που συνήθως γίνεται είναι η επέκταση ενός υπάρχοντος αλφαβήτου ή η εισαγωγή ενός νέου αλφαβήτου + γλώσσας με το σκοπό της εκφράσεως εννοιών υπό ένα συγκεκριμένο πρίσμα.

Πάρτε για παράδειγμα τα activity diagrams στη UML: οι θεμελιώδεις έννοιες κλάσεων και συσχετίσεων ενδεχομένως θα έφταναν από μόνες τους για τη μοντελοποίηση της ροής της εργασίας. Εισάγοντας τις διαδικασίες ως stereotyped έννοιες που μπορούν να εμπλακούν σε ένα νέο διάγραμμα (ως το αλφάβητο του διαγράμματος), τα διαγράμματα γίνονται πολύ πιο ευνόητα και επικεντρωμένα στο αντικείμενο της μοντελοποίησης.

Η οπτική γωνία εδώ είναι η έννοια της ροής των εργασιών στο σύστημα (activity diagrams). Άλλες έννοιες που μπορεί να έχουν παίξει το ρόλο της οπτικής γωνίας είναι π.χ., η ασφάλεια (επέκταση της UML με security stereotypes), η αντιστοιχισμός κλάσεων κ.ο.κ.

Με άλλα λόγια, επιβάλλεται να δει κανείς πώς μια τεχνική διευκολύνει τη σχεδίαση ενός συστήματος, για μια συγκεκριμένη δουλειά, αντιμετωπίζοντας τις οντότητες που εμπλέκονται στον υπό θεώρηση στόχο ως έννοιες πρώτης βαθμίδας (1st class citizens).

ΠΑΡΑΡΤΗΜΑ Β

Κανόνες για τη βιβλιογραφία, τους οποίους υποχρεούστε να τηρείτε ευλαβικά

MS Word

- [BaCR94] V.R. Basili, G.Caldiera, H.D. Rombach. The Goal Question Metric Approach. Encyclopedia of Software Engineering, pp. 528-532, John Wiley & Sons, Inc, 1994. Also available at <http://www.cs.umd.edu/users/basili/papers.html>
- [BBC+99] P.A. Bernstein, T. Bergstraesser, J. Carlson, S. Pal, P. Sanders, D. Shutt. Microsoft Repository Version 2 and the Open Information Model. Information Systems 24(2), pp. 71-98, 1999.
- [Dean97] E.B. Dean. Quality Functional Deployment from the Perspective of Competitive Advantage. Available at <http://mijuno.larc.nasa.gov/dfc/qfd.html>, 1997.
- [JaVa97] M. Jarke, Y. Vassiliou. Foundations of data warehouse quality – a review of the DWQ project. In Proc. 2nd Intl. Conference Information Quality (IQ-97), pp. 299-313, Cambridge, Mass., USA, June 1997.
- [JJQV98] M. Jarke, M.A. Jeusfeld, C. Quix, P. Vassiliadis. Architecture and quality in data warehouses. In Proc. 10th Conference on Advanced Information Systems Engineering (CAiSE '98), pp. 93-113, Pisa, Italy, June 1998.
- [Orr98] K. Orr. Data quality and systems theory. Communications of the ACM, 41(2), pp. 54-57, Feb. 1998.

Πίνακας 1. Το παραπάνω (σε MS Word) είναι πινακάκι με δύο κολώνες. Αριστερά είναι το abbreviation με 4 γράμματα και δύο νούμερα και δεξιά η περιγραφή της αναφοράς. Ο πίνακας είναι ταξινομημένος ανά abbreviation.

Abbreviation: 4 γράμματα για τους συγγραφείς + 2 ψηφία για την χρονολογία δημοσίευσης.

- 1 συγγραφέας: τα 4 πρώτα γράμματά του (ή αν έχει λιγότερα, όσα έχει), π.χ. [Dean97], [Orr98].
- 2 συγγραφείς: τα 2 πρώτα γράμματα του καθενός, π.χ. [JaVa97].
- 3 συγγραφείς: τα 2 πρώτα γράμματα του πρώτου και το πρώτο γράμμα των υπολοίπων, π.χ. [BaCR94].
- 4 συγγραφείς: Ο καθένας το αρχικό του, π.χ. [JJQV98].
- >4 συγγραφείς: Οι τρεις πρώτοι και '+', π.χ. [BBC+99].

Περιγραφή: (προσεξτε επίσης πού μπαίνει τελεία και πού κόμμα στα παραπάνω)

- **Άρθρο σε συνέδριο:** Ονόματα συγγραφέων, τίτλος άρθρου, συνέδριο (conference acronym), printed pages, τοποθεσία, ημερομηνία διεξαγωγής συνεδρίου.
 - M. Jarke, M.A. Jeusfeld, C. Quix, P. Vassiliadis. Architecture and quality in data warehouses. In Proc. 10th Conference on Advanced Information Systems Engineering (CAiSE '98), pp. 93-113, Pisa, Italy, June 1998.
 - M. Jarke, Y. Vassiliou. Foundations of data warehouse quality – a review of the DWQ project. In Proc. 2nd Intl. Conference Information Quality (IQ-97), pp. 299-313, Cambridge, Mass., USA, June 1997.
- **Άρθρο σε περιοδικό:** Ονόματα συγγραφέων, τίτλος άρθρου, όνομα περιοδικού, volume(no.), printed pages, ημερομηνία.

- P.A. Bernstein, T. Bergstraesser, J. Carlson, S. Pal, P. Sanders, D. Shutt. Microsoft Repository Version 2 and the Open Information Model. Information Systems 24(2), pp. 71-98, 1999.
- K. Orr. Data quality and systems theory. Communications of the ACM, 41(2), pp. 54-57, Feb. 1998.
- **Βιβλίο:** Ονόματα συγγραφέων, τίτλος, εκδοτικός οίκος, ημερομηνία.
 - V.R. Basili, G.Caldiera, H.D. Rombach. The Goal Question Metric Approach. Encyclopedia of Software Engineering, pp. 528-532, John Wiley & Sons, Inc, 1994.
- **Web:** Οτιδήποτε βρέθηκε από το Web, δίνουμε και το URL. Αν η ημερομηνία συγγραφής δεν είναι σαφής, δίνουμε την ημερομηνία που το διαβάσαμε.
 - E.B. Dean. Quality Functional Deployment from the Perspective of Competitive Advantage. Available at <http://mijuno.larc.nasa.gov/dfc/qfd.html>, 1997.

Αν την ώρα που γράφετε το κείμενο δεν προλαβαίνετε ή βαριέστε να συμπληρώσετε σωστά την αναφορά στο τέλος του κειμένου, βρείτε ένα χαρακτηριστικό τρόπο να αναφερθείτε σ' αυτήν και σημειώστε τον ΚΑΙ πάνω στο hard-copy σας. Φροντίστε να είναι μια πολύ draft περιγραφή -- τόσο draft ώστε να πρέπει υποχρεωτικώς να την διορθώσετε κάποια άλλη στιγμή.

LaTeX

Για όσους γράφουν σε LaTeX, η διαχείριση της βιβλιογραφίας είναι κάπως πιο εύκολη.

Εκμεταλλευθείτε το DBLP όσο το δυνατόν περισσότερο. Απλώς, προσέξτε τις παγίδες.

Π.χ., δεν αρκεί το citation

```
@inproceedings{DBLP:conf/vldb/SellisRF87,
  author    = {Timos K. Sellis and
              Nick Roussopoulos and
              Christos Faloutsos},
  title     = {The R+-Tree: A Dynamic Index for Multi-Dimensional Objects.},
  booktitle = {VLDB},
  year      = {1987},
  pages     = {507-518},
  ee        = {db/conf/vldb/SellisRF87.html},
  crossref  = {DBLP:conf/vldb/87},
  bibsource = {DBLP, http://dblp.uni-trier.de}
}
```

αλλά πρέπει και να συνοδεύεται και από το

```
@proceedings{DBLP:conf/vldb/87,
  editor    = {Peter M. Stocker and
              William Kent and
              Peter Hammersley},
  title     = {VLDB'87, Proceedings of 13th International Conference on
              Very Large Data Bases, September 1-4, 1987, Brighton, England},
  publisher = {Morgan Kaufmann},
  year      = {1987},
  isbn      = {0-934613-46-X},
  bibsource = {DBLP, http://dblp.uni-trier.de}
}
```

ή, ακόμα, τα δυο αυτά citations μπορούν να αντικατασταθούν από ένα, ενοποιημένο

```
@inproceedings{DBLP:conf/vldb/SellisRF87,
```

```

author    = {Timos K. Sellis and
             Nick Roussopoulos and
             Christos Faloutsos},
editor    = {Peter M. Stocker and
             William Kent and
             Peter Hammersley},
title     = {The R+-Tree: A Dynamic Index for Multi-Dimensional Objects},
booktitle = {VLDB'87, Proceedings of 13th International Conference on Very
             Large Data Bases, September 1-4, 1987, Brighton, England},
publisher = {Morgan Kaufmann},
year      = {1987},
isbn      = {0-934613-46-X},
pages     = {507-518},
ee        = {db/conf/vldb/SellisRF87.html},
crossref  = {DBLP:conf/vldb/87},
bibsource = {DBLP, http://dblp.uni-trier.de}
}

```

Εν γένει, αν χρησιμοποιείτε Latex, μπορείτε να διαλέξετε αν θα χρησιμοποιήσετε αριθμούς για τα citations στο κείμενό σας, ή αλφαριθμητικά.