

# Γενικές αρχές εκπόνησης και συγκρότησης της εργασίας σας για το MSc

Π. Βασιλειάδης

Τετάρτη, 03 Οκτωβρίου 2007

[http://www.cs.uoi.gr/~pvassil/linx/localCopies4grads/MSc\\_guidelines\\_byPV.pdf](http://www.cs.uoi.gr/~pvassil/linx/localCopies4grads/MSc_guidelines_byPV.pdf)

## 1. Επί της διαδικασίας – Αρχικά βήματα

Η διαδικασία εκπόνησης της εργασίας είναι ξεκάθαρη και λίγο-πολύ γνωστή. Τα αρχικά βήματα είναι:

1. **Γενικός ορισμός του προβλήματος σε αδρές γραμμές.** Ο σκοπός είναι να συμφωνήσουμε στο ποιο είναι το πρόβλημα, πώς το καταλαβαίνουμε και πώς θα επιδιώξουμε να το λύσουμε.
2. Ανάγνωση της **σχετικής βιβλιογραφίας και εν παραλλήλω συγγραφή** του σχετικού κεφαλαίου. Είναι βασικό, για κάθε άρθρο / κεφάλαιο βιβλίου που διαβάσετε να γράφετε και τουλάχιστον μισή σελίδα με το τι καταλάβατε από το άρθρο. **Γράφουμε τις αναφορές μέσα στο κείμενο χωρίς χρονοτριβή, καθώς γράφουμε το κείμενο + συμπληρώνουμε πλήρως και σωστά τα references στο τέλος του κειμένου.** Επιπλέον, εκτός από την παραπάνω μισή σελίδα, είναι πολύ βοηθητικό να φτιάχνετε τις εξής bullet lists για κάθε άρθρο που διαβάσετε:
  - Ποια είναι τα βασικά στοιχεία του προβλήματος που αντιμετωπίζει ο συγγραφέας (ώστε να καταλάβετε ποιες είναι οι παράμετροι του προβλήματος)?
  - Ποια τα ισχυρά σημεία του άρθρου (τι καταφέρει να κάνει καλά, δηλ.)?
  - Τι πειράματα κάνει ο συγγραφέας? (ζεύγη x- και y- αξόνων)
  - Ποια τα σημεία που δεν καλύπτει το άρθρο?

Είναι βασικό να συμφιλιωθείτε με το γεγονός ότι από τα άρθρα που θα έχετε για διάβασμα, **κάποια άρθρα θα χρειαστεί να τα διαβάσετε ξανά και ξανά** (και να ανανεώσετε την περιγραφή τους πολλές φορές, αντίστοιχα).

3. **Καλά και κακά παραδείγματα.** Στη γενική περίπτωση, αν επιχειρήσει κανείς να λύσει το πρόβλημα με straightforward, απλές τεχνικές θα προκύψει μια λύση που θα έχει διάφορα μειονεκτήματα. Αν όμως κάνει κάτι λίγο πιο έξυπνο (κεντρική ιδέα της προσέγγισής μας), θα προκύψει μια καλύτερη λύση. Άρα, κατασκευάζουμε ένα (ή περισσότερα) πρόβλημα «αναφοράς» και δείχνουμε ένα (ή περισσότερα) καλό και ένα (ή περισσότερα) κακό παράδειγμα.

4. **Κεντρική ιδέα.** Εν γένει, αυτό θα είναι κάποιος αλγόριθμος, τεχνική, θεώρημα ή άλλο τεχνικό αποτέλεσμα στο οποίο θα βασίζεται η συνεισφορά της εργασίας σας. Αυτό οφείλει να οργανωθεί και στο μυαλό σας και εγγράφως και ως κώδικας (ενδεχομένως με την αντίστροφη σειρά ☺)
5. **Τα επόμενα βήματα είναι customized ανά περίπτωση.**

Σε κάθε draft κείμενο που γράφετε θυμάστε να συμπληρώνετε το όνομά σας και την ημερομηνία συγγραφής. Είναι χρήσιμο να διατηρείτε και ένα τετράδιο εργασίας για την εργασία σας, με (α) ένα «καθαρό» τμήμα για να καταγράφονται όσα έχουμε επιλύσει και συμφωνήσει και (β) ένα τμήμα εργασίας, για ιδέες, σκέψεις, ... υπό επεξεργασία. Από την προηγούμενη εμπειρία μου, οι άνθρωποι που δουλεύουν ένα θέμα για καιρό συχνά τείνουν να δημιουργήσουν ένα χάος χαρτιών στο οποίο θάβονται χρήσιμα πράγματα για να τα βρει ο αρχαιολόγος του μέλλοντος...

## 2. Γενική Δομή της Εργασίας

Τα αρχικά κεφάλαια της εργασίας σας λίγο πολύ έχουν την εξής δομή:

1. Εισαγωγή
2. Σχετική δουλειά
3. Τυπικός ορισμός του προβλήματος και κεντρική ιδέα
4. Αναλυτική περιγραφή της προτεινόμενης λύσης
5. Πειράματα
6. Ανακεφαλαίωση – Μελλοντική δουλειά

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

## 3. 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) τη μέθοδο που ακολουθείτε για να λύσετε το πρόβλημα και (δ) τη δομή του κειμένου που ακολουθεί.

Τελειώνοντας το διάβασμα της εισαγωγής, ο αναγνώστης πρέπει να έχει απάντηση στις εξής ερωτήσεις:

- **Τι?** – δηλ., να έχει καταλάβει ποιο είναι το πρόβλημα που πάτε να λύσετε.
- **Γιατί?** – δηλ., να έχει αποκτήσει κίνητρο για να διαβάσει παρακάτω.

- **Πώς?** – δηλ., να έχει μια εικόνα σε αδρές γραμμές για το πώς λύνεται το πρόβλημα που πάτε να λύσετε.
- **Πού?** – δηλ., να έχει μια εικόνα σε ποια σημεία του κειμένου θα παρουσιαστεί κάθε τι που αναφέρετε στην απάντηση του «πώς».

Η εισαγωγή είναι από τα τελευταία κεφάλαια που συγγράφονται για την εργασία σας. Η γενικότερη δομή της εισαγωγής, περιλαμβάνει τις εξής παραγράφους:

- Περιγραφή του προβλήματος σε ότι αφορά τον έξω κόσμο.
- Γιατί η υπάρχουσα mainstream τεχνολογία, καθώς και η σχετική δουλειά δυσκολεύονται / αδυνατούν να λύσουν το πρόβλημα. Με άλλα λόγια, περιγραφή του τι λείπει, ή πού υπάρχει χώρος για βελτίωση στις υπάρχουσες industrial ή ερευνητικές τεχνικές.
- Τι κάνουμε εμείς. Εδώ, πρέπει κανείς να δώσει έμφαση στα εξής:
  - ο «το πρόβλημα που μας απασχολεί, λοιπόν, είναι ...»
  - ο « η προσέγγισή μας βασίζεται στην εξής ιδέα: ...»
  - ο κακό vs. καλό παράδειγμα (χωρίς vs. μέσω της προτεινόμενης τεχνικής)

Εν γένει, εδώ παρουσιάζεται το πρόβλημα, η κεντρική ιδέα της λύσης, τα επί μέρους προβλήματα που καλείται κανείς να επιλύσει στο πλαίσιο του ευρύτερο προβλήματος και ο τρόπος με τον οποίο τα αντιμετωπίσαμε. Τέλος, μία με δύο παραγράφους αφιερώνουμε στο τι πειράματα κάναμε, πώς αυτά πιστοποιούν ότι η προσέγγισή μας είναι καλή και τι νέο ενδεχομένως ανακαλύψαμε στη διάρκειά τους.

- Bullet list of contributions. Τα παραπάνω, βαλμένα όμως σε bullets, ώστε να μένουν με συνοπτικό τρόπο στο μυαλό του αναγνώστη.
- Roadmap. Η δομή του κειμένου (ή αλλιώς, πού θα βρει τι ο αναγνώστης).

Η εισαγωγή γράφεται σε ενεστώτα χρόνο (και όχι σε μέλλοντα) και στο πρώτο πληθυντικό πρόσωπο (και όχι ενικό).

#### 4. Σχετική εργασία

Όπως ήδη προείπαμε, αυτό είναι το πρώτο κεφάλαιο που συγγράφεται, από την αρχή κιόλας της εκπόνησης της εργασίας σας. Επίσης είναι βασικό να τονιστεί ξανά ότι **δεν θα γίνεται αποδεκτό εκ μέρους μου οποιοδήποτε κείμενο (έστω και draft) στο οποίο οι αναφορές δεν είναι σωστά ορισμένες – τουλάχιστον μέσα στο σώμα του κειμένου** (αν όχι και στο τέλος).

Η βιβλιογραφία έχει τυπικό τρόπο ορισμού της. Η αρίθμηση με νούμερα επιτρέπεται μόνο αν γράφετε σε Latex. Για τους χρήστες του Word, μόνο τα abbreviations που περιγράφονται στην ενότητα «Αναφορές» είναι αποδεκτά.

Υπάρχουν αναλυτικές οδηγίες για την συλλογή και σύνοψη της σχετικής εργασίας στο κείμενό μου «Γενικές αρχές εκπόνησης επισκοπήσεων της βιβλιογραφίας», στις οποίες και σας παραπέμπω. Παρακάτω, δίνονται με επιγραμματικό τρόπο μερικές βασικές κατευθυντήριες γραμμές επί του θέματος.

1. Το κεφάλαιο της σχετικής δουλειάς δεν έχει σκοπό (μόνο) να πείσει ότι κάνατε το διάβασμά σας. Αντιθέτως, σκοπό έχει να παρουσιάσει στον αναγνώστη, **με δομημένο τρόπο,**
  - ποιες είναι οι **σχετικές περιοχές** (γι' αυτό και φτιάχνουμε μια υποενότητα για κάθε τέτοια περιοχή),
  - ποιες είναι οι **κρίσιμες εργασίες** σε κάθε περιοχή και τι κάνουν (βλ. και παραπάνω).
  - **πώς συγκρίνεται η δική μας προσέγγιση σε σχέση με τη σχετική δουλειά** (ήτοι, πώς καλύπτει ελλείψεις, βελτιώνει προηγούμενες ιδέες, ή προσεγγίζει με φρέσκια ματιά το πρόβλημα).
2. Το killer feature σε μια καλή επισκόπηση της σχετικής δουλειάς είναι το **πινακάκι σύγκρισης** των σχετικών εργασιών, το οποίο έχει για γραμμές τα στοιχεία του προβλήματος, για κολώνες τις σχετικές εργασίες και ως περιεχόμενα το πώς αντιμετωπίζει το κάθε ζήτημα η κάθε εργασία.
3. **Για να συνεννοούμαστε εύκολα,** καλό είναι να θυμάστε ότι τα άρθρα συνήθως γίνονται γνωστά από τους συγγραφείς και το μέρος που δημοσιεύθηκαν (π.χ., το άρθρο του Codd στο CACM του '70) και όχι από τον τίτλο τους.

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>

## 5. Τυπικός ορισμός του προβλήματος και κεντρική ιδέα

Αυτό είναι ένα κεφάλαιο το οποίο, ανάλογα με την εργασία, μπορεί να σπάει σε τμήματα τα οποία εντάσσονται σε άλλα κεφάλαια. Η κεντρική ιδέα για το κεφάλαιο αυτό, πάντως, είναι η εξής:

- **Τυπικός ορισμός του προβλήματος.** Μπορούμε με τυπικό, μαθηματικό τρόπο να ορίσουμε τις κεντρικές έννοιες του προβλήματος, καθώς και τα μεγέθη που θέλουμε να μετρήσουμε / ελαχιστοποιήσουμε / μεγιστοποιήσουμε.
- **Κεντρική ιδέα.** Αυτό, όπως είπαμε ήδη, είναι το θεμελιώδες contribution πάνω στο οποίο χτίζεται η εργασία σας. Ενδεχομένως και να μετακινηθεί στο επόμενο κεφάλαιο, πάντως σκοπό έχει να περιγράψει βάση ποιας τεχνικής λύνουμε το υπό εξέταση πρόβλημα, και γιατί επιλέγουμε να ακολουθήσουμε αυτό το δρόμο. Επίσης, εδώ βοηθά πολύ η ύπαρξη **παραδειγμάτων (καλών vs. κακών)**.
- **Διαστασιολόγηση του προβλήματος.** Μπορούμε να εντοπίσουμε τις κρίσιμες παραμέτρους του προβλήματος και τις τιμές που έχει νόημα να δώσουμε. Αυτό θα βοηθήσει αργότερα στη διεξαγωγή των πειραμάτων.

## 6. Πειράματα

Σε μια ερευνητική εργασία, **τα πειράματα έχουν σκοπό να επαληθεύσουν κάποια υπόθεση εργασίας**. Ας υποθέσουμε ότι προτείνετε ένα νέο ευρετήριο κατάλληλο για την online ταξινόμηση του συνδυασμού ενός συνόλου αποθηκευμένων εγγραφών και ενός συνόλου νέων, εισερχόμενων εγγραφών. Οι υποθέσεις εργασίας εδώ είναι:

- (α) το ευρετήριο ταξινομεί το συνδυασμό των εμπλεκόμενων συνόλων δεδομένων σωστά,
- (β) το ευρετήριο ταξινομεί το συνδυασμό των εμπλεκόμενων συνόλων δεδομένων αποδοτικότερα από τις τρέχουσες μεθόδους.

Για την επαλήθευση των παραπάνω υποθέσεων εργασίας πρέπει να διεξαχθούν πειράματα που να τις επαληθεύουν. Η οργάνωση των πειραμάτων, η συντεταγμένη διεξαγωγή τους, καθώς και η καταγραφή τους στο κείμενο είναι εξαιρετικά δύσκολες δουλειές. Στη συνέχεια, τα παραπάνω ζητήματα εξετάζονται οργανωμένα όπως θα τα παρουσιάζαμε σε ένα άρθρο: (α) σε σχέση με την πειραματική μέθοδο<sup>1</sup> και (β) σε σχέση με τα πειράματα per se.

---

<sup>1</sup> Όπως και πολλοί άλλοι Έλληνες, έτσι και γω, είδα με έκπληξη ότι οι επιστήμονες διεθνώς διαχωρίζουν τους όρους “method” και “methodology”. Η μέθοδος που ακολουθείται σε κάθε εργασία είναι απόρροια μιας γενικότερης «φιλοσοφίας» για τη διεξαγωγή των πειραμάτων («μεθοδολογία»). Παρ’ όλα αυτά, σε πολλές εργασίες (ενδεχομένως και δικές μου) η εν λόγω ενότητα τιτλοφορείται “Experimental methodology”.

## 6.1 Πειραματική μέθοδος

Πριν αρχίσουμε τη διεξαγωγή των πειραμάτων, πρέπει να προηγηθεί μια σχετικά λεπτομερής σχεδίαση της πειραματικής μεθόδου που θα ακολουθηθεί.

Κατ' αρχήν πρέπει να είναι ξεκάθαρες σε όλους τους εμπλεκόμενους, **οι υποθέσεις εργασίας** γύρω από το προτεινόμενο τεχνικό αποτέλεσμα. Στη γενική περίπτωση, η *ορθότητα* και η *αποδοτικότητα* μιας προτεινόμενης τεχνικής είναι δύο καλές κατευθυντήριες γραμμές σε υψηλό επίπεδο για να ξεινήσει κανείς να σιέφτεται. Επίσης είναι σημαντικό να τεκμηριωθούν και τα πλεονεκτήματα της τεχνικής αυτής σε σχέση με το *τρέχον state-of-the-art*. Φυσικά, ανάλογα με το είδος της εργασίας, οι βασικές κατευθυντήριες γραμμές μπορεί να αλλάζουν (π.χ., αν κάποιος προτείνει ένα νέο σχεδιαστικό πρότυπο, έχει νόημα να μελετηθεί αν βοηθά τους ανθρώπους να σχεδιάσουν καλύτερα κάποιες σχετικές με το πρότυπο εφαρμογές έχοντας διδαχθεί το πρότυπο σε σχέση με το αν δεν το είχαν διδαχθεί).

Οι παραπάνω υψηλού επιπέδου υποθέσεις εργασίας πρέπει να μετατραπούν σε **αντικειμενικά, μετρήσιμα μεγέθη**. Επανερχόμενοι στο προαναφερθέν παράδειγμα του ευρετηρίου, πρέπει να μεταφράσουμε τις υψηλού επιπέδου αρχικές υποθέσεις σε μετρήσιμες παρατηρήσεις που επιβεβαιώνουν ή καταρρίπτουν τις εν λόγω υποθέσεις:

(α1) αριθμός εγγραφών που το ευρετήριο τοποθετεί σε λάθος θέση

(β1) μνήμη που χρειάζεται το ευρετήριο για τη λειτουργία του (και σύγκριση με την «αντίπαλη» state-of-the-art μέθοδο)

(β2) χώρος στο σκληρό δίσκο που χρειάζεται το ευρετήριο για την αποθήκευσή του (και σύγκριση με την «αντίπαλη» state-of-the-art μέθοδο)

(β3) χρόνος / αριθμός σελίδων που μεταφέρθηκαν από το σκληρό δίσκο / ... που χρειάζεται το ευρετήριο για να ολοκληρώσει την σύμπληξη των δύο συνόλων εγγραφών σε ένα (και σύγκριση με την «αντίπαλη» state-of-the-art μέθοδο).

Το πρόβλημα του πώς κάνουμε αυτή την απεικόνιση, μπορεί να οργανωθεί αν κανείς καταλάβει τα **βασικά στοιχεία που χαρτογραφούν ένα σύνολο πειραμάτων**:

- **Μετρήσιμες ποσότητες.** Οι μετρήσιμες ποσότητες, ή αλλιώς, μετρήσιμα μεγέθη, χαρακτηρίζουν με αντικειμενικό τρόπο την επιτυχία ή αποτυχία μιας τεχνικής που

---

υπονοώντας ότι η γενική στρατηγική για τα πειράματα της εργασίας εφαρμόστηκε σε κάθε πείραμα χωριστά. Φανταστείτε πάντως την έκπληξη των ξένων για το ότι οι Έλληνες δεν ξέρουν τη διαφορά των όρων...

δοκιμάζουμε σε σχέση με μια υπόθεση. Αν επιστρέψει κανείς στο προαναφερθέν παράδειγμα, τα μεγέθη ( $\alpha$ ) - ( $\beta$ ) είναι όλα μετρήσιμα μεγέθη. Ένας καλός τρόπος να σκέφτεται κανείς τα μετρήσιμα μεγέθη είναι να θυμάται ότι συνήθως απεικονίζονται στον άξονα των y στις γραφικές παραστάσεις.

- **Σημαντικοί παράγοντες που επηρεάζουν τις μετρήσιμες ποσότητες.** Η συμπεριφορά των μετρήσιμων μεγεθών εξαρτάται από διάφορους παράγοντες. Στο παράδειγμα αναφοράς που χρησιμοποιούμε, τέτοιοι παράγοντες μπορεί να είναι ( $\alpha$ ) το μέγεθος των εμπλεκόμενων συνόλων εγγραφών (σε αρ. εγγραφών), ( $\beta$ ) το μέγεθος του πεδίου ορισμού του χαρακτηριστικού βάσει του οποίου γίνεται η ταξινόμηση, ( $\gamma$ ) το ποσό των διαθέσιμων buffers στη μνήμη, κ.κ. **Ένα πείραμα ελέγχει τη συμπεριφορά ενός μετρήσιμου μεγέθους σε σχέση με τη μεταβολή ενός σημαντικού παράγοντα.** Για το λόγο αυτό, μπορεί κανείς να σκέφτεται τα κρίσιμα μεγέθη σαν τις τιμές του άξονα των x.
- **Μη σημαντικές παράμετροι.** Πολλές φορές, λόγω του τρομακτικά πολυδιάστατου χαρακτήρα των προβλημάτων, στην αρχική φάση των πειραμάτων καθορίζουμε συγκεκριμένες τιμές για κάποιες παραμέτρους, στη βάση των οποίων διεξάγουμε όλα τα επόμενα πειράματα. Συνήθως, οι τιμές αυτές καθορίζονται είτε αυθαίρετα, στη βάση καλά κατανοητών κανόνων της επιστημονικής κοινότητας, είτε από τα λεγόμενα microbenchmarks που είναι μικρά πειραματάκια με τα οποία αποφασίζουμε τις βέλτιστες τιμές κάποιων παραμέτρων για τα επόμενα πειράματα. Στο παράδειγμα αναφοράς που χρησιμοποιούμε, μπορεί κανείς να θεωρήσει το μέγεθος μιας σελίδας στο δίσκο ως μη σημαντική παράμετρο, στην οποία απλώς οφείλει να ανατεθεί μια σταθερή τιμή (υποθέτοντας, δηλ., ότι η μεταβολή του μεγέθους της σελίδας δεν θα επιφέρει σημαντικές διαφορές στη συμπεριφορά του ευρετηρίου).
- **Σύνολα δεδομένων.** Στο χώρο των βάσεων δεδομένων, τα σύνολα δεδομένων στα οποία ελέγχονται οι τεχνικές είναι εν γένει δύο ειδών:
  - ο **Πραγματικά σύνολα δεδομένων**, τα οποία είναι διαθέσιμα στο Διαδίκτυο.
  - ο **Συνθετικά σύνολα δεδομένων**, τα οποία κατασκευάζονται για τους σκοπούς του πειράματος.

Τα πραγματικά σύνολα δεδομένων, ενώ αναπαριστούν την πραγματικότητα, είναι εν γένει μικρά. Για να ελέγξουμε μια προτεινόμενη τεχνική στα όριά της, συνήθως καταφεύγουμε στην κατασκευή συνθετικών συνόλων δεδομένων, με μεγάλο μέγεθος. Στην δεύτερη αυτή περίπτωση, είναι σημαντικό να καταγραφεί αφενός η διαδικασία

παραγωγής και αφετέρου οι στατιστικές ιδιότητες των παραγομένων δεδομένων (τυχαία, correlated, anti-correlated, κ.ο.κ.).

Για κάθε ένα από τους παραπάνω σημαντικούς παράγοντες, **η διαδικασία σχεδίασης πρέπει να αποφασίσει τις τιμές που μπορεί να λάβει στη διεξαγωγή των πειραμάτων**. Έτσι, η διαδικασία σχεδίασης των πειραμάτων, κατασκευάζει ένα «χώρο πειραμάτων» που προκύπτει ως το καρτεσιανό γινόμενο των τιμών αυτών. Προφανώς, αυτό δεν σημαίνει ότι όλα τα δυνατά πειράματα πρέπει να διεξαχθούν – η απόρριψη όμως κάποιων πειραμάτων είναι μια πιο σύνθετη δουλειά που καλό είναι να την αναλάβει ο επιβλέπων.

Επίσης σημαντικές αποφάσεις που πρέπει να παρθούν αφορούν:

- **Baseline ή opponent τεχνική**. Με βάση τη σχετική εργασία, πρέπει κανείς να αποφανθεί ποιο είναι το state-of-the-art στο πρόβλημα που εξετάζεται. Αν το πρόβλημα είναι τόσο νέο που δεν υπάρχει «αντίπαλος» στη βιβλιογραφία, τότε το μέτρο σύγκρισης πρέπει να παρθεί σε σχέση με την πιο απλή τεχνική που θα υλοποιούσε ένας πτυχιούχος πληροφορικής, αν ερχόταν αντιμέτωπος με το πρόβλημα.
- **Κώδικας προς κατασκευή**. Υπάρχει μια σχολή σκέψης που συνιστά ότι ο κώδικας πρέπει να αφορά ακριβώς και μόνο τα προς διεξαγωγή πειράματα –ότι άλλο περιττό πρέπει να αποφεύγεται και να επιδίδεται έμφαση στα υπό εξέταση χαρακτηριστικά. Υπάρχει μια άλλη σχολή σκέψης, βέβαια, που εφιστά την προσοχή στη συντηρησιμότητα του κώδικα και κατά συνέπεια, στην ευσταθή κατασκευή του (με τίμημα βέβαια το βάθος των εναλλακτικών λύσεων που προλαβαίνει κανείς να εξετάσει). Προφανώς, το ιδεώδες θα ήταν ο συνδυασμός των παραπάνω, αλλά αυτό σπανίως επιτυγχάνεται. Κατά τη γνώμη μου (και μετά από πολλές προσωπικές αποτυχίες), καλό είναι να αποφασίζει κανείς τι θέλει να πετύχει με τον υπό κατασκευή κώδικα και να πράττει αναλόγως.

## 6.2 Διεξαγωγή πειραμάτων

Αν η σχεδίαση έχει επιτύχει, η διεξαγωγή των πειραμάτων απαιτεί απλώς **συστηματική προσπάθεια**. Εσχάτως, στο εργαστήριο ζητώ να αναρτάται πρόγραμμα διεξαγωγής πειραμάτων (ήτοι, κάθε πείραμα προς διεξαγωγή χαρακτηρίζεται από τη θέση του στον παραπάνω πολυδιάστατο χώρο και τοποθετείται σε ένα πινακίκι που λέει πότε θα διεξαχθεί ποιο πείραμα). Μια παλιά τεχνική που εφαρμόζεται σε όλες τις άλλες επιστήμες, απαιτεί την τήρηση τετραδίου πειραμάτων, στο οποίο καταγράφονται οι συνθήκες και τα

αποτελέσματα του κάθε πειράματος. Στην πληροφορική βέβαια, χρησιμοποιούμε MS Excel (και προφανώς, οι απόψεις δίστανται ως προς το τι να χρησιμοποιήσει κανείς).

Κομμάτι της συστηματικής προσπάθειας είναι η καθιέρωση ενός **σταθερού πειραματικού περιβάλλοντος**. Δεν περιφέρουμε τα πειράματα από το ένα μηχάνημα στο άλλο. Αντίθετα, αποφασίζουμε πού θα εκτελεστούν, με βάση τις απαιτήσεις τους σε δίσκο, CPU, ... και παίρνουμε X-lock στον εν λόγω server. Τα χαρακτηριστικά του λογισμικού, επίσης πρέπει να παραμείνουν σταθερά στη διάρκεια των πειραμάτων. Τέλος, καταγράφουμε τα χαρακτηριστικά του εν λόγω περιβάλλοντος και τα παραθέτουμε στο κείμενο.

**Κάθε πείραμα εκτελείται πολλές φορές**, αν η εκτέλεση παρουσιάζει τυχαία φαινόμενα (αντίθετα, αν π.χ., έχετε ένα module που εκτιμά το κόστος μιας query, δεν έχει νόημα να το τρέξετε πολλές φορές – κάθε φορά θα βγαίνει το ίδιο αποτέλεσμα).

### **Τα αποτελέσματα των πειραμάτων είναι ιερά!!**

Δεν αλλοιώνουμε τα αποτελέσματα ποτέ.

Τέλος: στη διάρκεια των πειραμάτων είναι απολύτως αναμενόμενο να παρουσιαστούν **παράξενα, μη αναμενόμενα και counterintuitive αποτελέσματα**. Η πιο πιθανή αιτία είναι ένα bug στον κώδικα: ψάξτε το! Αν η προτεινόμενη τεχνική χάνει από τον «αντίπαλο» σε κάποιες περιπτώσεις,

- δεν αλλάζουμε το αποτέλεσμα
- εξηγούμε γιατί και πότε χάνουμε
- προσπαθούμε να βελτιώσουμε την τεχνική μας, αν γίνεται.

### **6.3 Τελικές σκέψεις**

**Η συγγραφή της σχετικής ενότητας «Πειράματα» οφείλει να εξυπηρετεί το σκοπό του να πεισθεί ο αναγνώστης ότι η προτεινόμενη τεχνική εξυπηρετεί τις υποθέσεις εργασίας που έγιναν στην αρχή.**

- Η περιγραφή της πειραματικής μεθόδου (σε σχέση με όλες τις επί μέρους παραμέτρους που προαναφέρθηκαν) πρέπει να είναι επεξηγηματική ως προς τους λόγους για τους οποίους κάθε τι είναι σημαντικό στη μέθοδο που ακολουθήθηκε.
- Η περιγραφή του περιβάλλοντος στο οποίο διεξήχθησαν τα πειράματα πρέπει να είναι πλήρης, ακριβής και απέρριπτη.

- Η περιγραφή των πειραμάτων πρέπει (α) να πείθει τον αναγνώστη, (β) να επισημαίνει τα αξιοπαρατήρητα σημεία και (γ) να επεξηγεί τα μη αναμενόμενα ή προβληματικά σημεία. Σε μια εργασία που υποβάλλεται σε ένα συνέδριο ή περιοδικό, δεν είναι απαραίτητο να παρουσιαστούν όλα τα πειράματα – απαραίτητο είναι να πεισθεί ο αναγνώστης. Σε μια MSc thesis, όμως, υπάρχει ο χώρος και ο χρόνος να αναπτυχθούν τα πειράματα σε έκταση (έστω και ως παράρτημα). Κάθε γραφική παράσταση πρέπει να επεξηγείται (αν και σε πολλές εργασίες σε συνέδρια, λόγω έλλειψης χώρου, η εν λόγω συζήτηση μπορεί να περιορίζεται αισθητά).

## 7. Παρελκόμενα της εισαγωγής

Από την εισαγωγή μπορούμε, με εύκολο τρόπο, να βγάλουμε και διάφορα άλλα τμήματα του κειμένου.

**Περίληψη / Abstract Διατριβής.** Η περίληψη που μπαίνει στη διατριβή σας είναι μεγέθους περίπου μίας σελίδας και αποτελεί βασικά περίληψη της εισαγωγής του κειμένου. Εν γένει, η δομή είναι η εξής:

- Γενικό πρόβλημα (1-2 προτάσεις).
- Συγκεκριμένος ορισμός του προβλήματος (1 πρόταση).
- Περιγραφή του τι κάνουμε στην εργασία αυτή.
- Μετατροπή της bullet list της εισαγωγής σε ρέον κείμενο. Εδώ υπάρχει χώρος για λίγο περισσότερη ανάλυση.

Αν γράφουμε **abstract για κάποιο άρθρο**, τότε, λόγω χώρου, (α) περιορίζουμε το γενικό πρόβλημα αισθητά και (β) λίγο-πολύ μένουμε σε μια συνοπτική απαρίθμηση της bullet list της εισαγωγής.

Πάσης φύσεως περιλήψεις γράφονται στον ενεστώτα χρόνο και στο πρώτο πληθυντικό πρόσωπο.

**Ανακεφαλαίωση.** Αντιγράφουμε το abstract όπως είναι (copy-paste), απλά έχοντας

- (α) διαγράψει τη συζήτηση περί του γενικού προβλήματος στον έξω κόσμο, και,
- (β) μετατρέπει τον ενεστώτα σε παρατατικό (ENG) ή αόριστο (GR).

Ο σκοπός αυτής της ενότητας είναι να δώσει με συνοπτικό τρόπο στον (κουρασμένο/βιαστικό) αναγνώστη, άλλη μια ευκαιρία να κατανοήσει τι περιέγραψε το κείμενο, και τι συνεισφορά είχαμε.

Φυσικά, προσθέτουμε και την υποενότητα «**Μελλοντική Δουλειά**», για να δώσουμε και ένα insight στον αναγνώστη για το πώς μπορεί να συνεχισθεί η έρευνα. Η εν λόγω υποενότητα

γράφεται σε στυλ «η έρευνα μπορεί να συνεχισθεί ως εξής...» και όχι «εμείς στο μέλλον θα κάνουμε αυτό...».

Όπως έχετε ίσως διαπιστώσει, σε πολλά σημεία του κειμένου (περίληψη, εισαγωγή, σανακεφαλαίωση) επαναλαμβάνουμε με συνοπτικό τρόπο το τι κάναμε και γιατί είναι σημαντικό. Αυτό γίνεται για να βοηθήσουμε τον αναγνώστη, και ιδίως τον μη ειδικό αναγνώστη, να καταλάβει την ουσία, το βάθος και την αξία της δουλειάς.

## 8. Κανόνες αναφοράς στη βιβλιογραφία

### Κανόνες για τη βιβλιογραφία, τους οποίους υποχρεούστε να τηρείτε ευλαβικά

#### 8.1 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. 2<sup>nd</sup> 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. 10<sup>th</sup> 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:** (προσέξτε 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. 10<sup>th</sup> 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. 2<sup>nd</sup> 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.
  - o 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.
  - o 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.
  - o 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.
  - o 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.

## 8.2 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.