



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΘΕΣΣΑΛΟΝΙΚΗΣ

Τμήμα Πληροφορικής

Πτυχιακή Εργασία

Ανίχνευση εισβολών με αρχιτεκτονική Lambda.

Ευφροσύνη Καλτιριμίδου

Επιβλέπων:

Αθηνά Βακάλη

Καθηγήτρια

Θεσσαλονίκη 2016

Περίληψη

Στο παρόν κείμενο δίνονται οι βασικές δομές και τα κύρια χαρακτηριστικά μιας πτυχιακής εργασίας γραμμένης σε \LaTeX . Το δείγμα [?] απευθύνεται σε φοιτητές του Τμήματος Φυσικής του ΕΚΠΑ, αλλά με ελάχιστη τροποποίηση μπορεί να χρησιμοποιηθεί και από άλλες ειδικότητες.

Abstract

The present text provides a sample of the basic structure and features of a senior thesis written and typeset using \LaTeX . The template [?] aims at assisting Physics students at the University of Athens, but it can serve others as well, by slightly modifying the initial settings.

Στην Ηλέκτρα

για τη βοήθειά της εναντία στον Kingpin

–Daredevil

Περιεχόμενα

1	Εισαγωγή	9
2	Βιβλιογραφική Ανασκόπηση	10
3	Data Model	11
1	Ορολογία	12
2	Ιδιότητες των δεδομένων	12
21	Rawness	12
22	Immutability	13
23	Perpetuity	13
24	Delete	14
4	Πειραματική Διαδικασία	15
5	Αποτελέσματα και Ανάλυση	16
6	Συμπεράσματα	18
Α΄	Το λογισμικό	19

Β' Ορολογία	20
Γ' Σύντομο βιογραφικό σημείωμα	21

Κεφάλαιο 1

Εισαγωγή

Στο παρόν κεφάλαιο αναπτύσσεται η εισαγωγή της πτυχιακής εργασίας. Είθισται να περιγράφεται η υπάρχουσα βιβλιογραφία στο συγκεκριμένο αντικείμενο της εργασίας, να αναφέρονται τα κομβικά σημεία, τα ανοιχτά ερωτήματα και να παρατίθεται ο σκοπός και το κίνητρο της εργασίας.

Συνηθίζεται επίσης να κλείνει η εισαγωγή με μια περιγραφή των κεφαλαίων που θα ακολουθήσουν. Στην παρούσα εργασία δεν έχει συμπεριληφθεί κάποιος πρόλογος που πιθανώς να ενδιαφέρει το φοιτητή να συμπεριλάβει στην αρχή της εργασίας.

Κεφάλαιο 2

Βιβλιογραφική Ανασκόπηση

Το κεφάλαιο της βιβλιογραφικής ανασκόπησης στοχεύει στη συγκέντρωση της υπάρχουσας βιβλιογραφίας σχετικής με το αντικείμενο της εργασίας. Η βιβλιογραφική ανασκόπηση έχει ως κύριο στόχο να παράσχει το υπάρχον επιστημονικό υπόβαθρο που απαιτείται για τη θεμελίωση της επιστημονικής εγκυρότητας της εργασίας.

Για εργασίες πειραματικού χαρακτήρα, το κεφάλαιο αυτό μπορεί εναλλακτικά να περιλάβει το θεωρητικό μέρος. Παρακάτω παρατίθενται τρία παραδείγματα εξισώσεων: ένα εντός της παραγράφου: $\sum_i x_i = K$, ένα σε ξεχωριστή γραμμή, αλλά χωρίς αρίθμηση:

$$\sum_i x_i = K$$

κι ένα πλήρες, με αρίθμηση στα δεξιά και εισαγωγή ετικέτας ώστε να γίνεται αναφορά προς αυτήν:

$$\sum_i x_i = K \tag{2.1}$$

Η αναφορά γίνεται με παρόμοιο τρόπο με τις εικόνες και τους πίνακες.

Κεφάλαιο 3

Data Model

Ο πυρήνας της αρχιτεκτονικής Lambda είναι τα δεδομένα της και συγκεκριμένα το master dataset που αποθηκεύεται στο batch layer. Το master dataset αποτελεί την πηγή της “αλήθειας” όπως έχει καταγραφεί και αποθηκευτεί. Ακόμα και αν όλα τα views στο speed και serving layer χανόντουσαν, το σύστημα μπορεί να τα ξαναδημιουργήσει μέσα σε μερικές ώρες από το master dataset. Είναι προφανές ότι το master dataset είναι το μοναδικό κομμάτι της αρχιτεκτονικής Lambda το οποίο πρέπει να προφυλαχτεί από data corruption, disk failure, και άλλα σφάλματα που οδηγούν στη μη αναστρέψιμη απώλεια δεδομένων.

1. Ορολογία

Είναι σημαντικό να καθορίσουμε την ορολογία που χρησιμοποιείται σε κάθε θεματικό επίπεδο ώστε να γίνει σαφές το περιεχόμενο του.

Information Είναι οι πληροφορίες που έχουν σχέση με την εφαρμογή που εκτελείται στο σύστημα.

Data Οι πληροφορίες που δεν μπορούν να εξαχθούν/υπολογιστούν από άλλα δεδομένα που κατέχονται. Αποτελούν το master dataset. Θα αναφέρονται ως δεδομένα.

Views Πληροφορίες που έχουν εξαχθεί/υπολογιστεί από τα Data.

Queries Ερωτήσεις που υποβάλλονται στο σύστημα και απαντώνται από τα Views.

2. Ιδιότητες των δεδομένων

Υπάρχουν τρεις ιδιότητες που προτείνεται να έχουν τα δεδομένα του master dataset και αναφέρονται ως rawness, immutability, perpetuity ή όπως αποδίδονται σε μία έκφραση “eternal trueness of data”. Η κάθε μία ιδιότητα θα αναλυθεί και θα εξηγηθεί για να γίνει σαφές ο λόγος που προτείνεται.

2.1 Rawness

Οι πληροφορίες που αποθηκεύονται στο master dataset και αποτελούν τα δεδομένα πρέπει να είναι στην πιο ακατέργαστη μορφή τους, κατά προτίμηση στη μορφή σχεδόν που τα λαμβάνουμε χωρίς καμία τροποποίηση ή διαγραφή. Το σύστημα είναι ένα Query Answering σύστημα με κύρια πηγή πληροφοριών το master dataset. Όσο πιο ακατέργαστα είναι τα δεδομένα που κατέχει τόσο περισσότερες ερωτήσεις μπορεί να απαντήσει. Αυτό οφείλετε στο γεγονός ότι οι μέθοδοι κατεργασίας των δεδομένων για αποθήκευση εφαρμόζουν αλγόριθμους για serialization πάνω στα δεδομένα κατά την οποία μπορεί να διαγραφούν δεδομένα που δεν έχουν προβλεφθεί κατά την σχεδίαση των αλγορίθμων ή να μετατραπούν. Επειδή δεν είναι γνωστά από την αρχή όλα τα ερωτήματα που θα υποβληθούν στο σύστημα είναι προτιμότερο να κατέχει όλη την πληροφορία

στην αρχική της μορφή ώστε να μπορεί να απαντήσει όποιο ερώτημα προκύψει.

2.2 Immutability

Η επόμενη ιδιότητα των δεδομένων είναι η αμεταβλητότητα τους, η μόνη ενέργεια που επιτρέπεται στο master dataset είναι η Add και δεν υπάρχει Update. Η Delete είναι μια ειδική περίπτωση ενέργειας και επιτρέπεται κάτω από συγκεκριμένες περιπτώσεις. Α αμεταβλητότητα των δεδομένων εμφανίζει δύο πλεονεκτήματα.

Ασφάλεια από Ανθρώπινα λάθη Ένας από τους λόγους απώλειας δεδομένων σε πολλές περιπτώσεις οφείλετε σε λάθος πράξεις του ανθρώπινου παράγοντα. Με την αμεταβλητότητα, κανένας χρήστης ή διαχειριστής δεν μπορεί να αλλοιώσει τα ήδη υπάρχοντα δεδομένα, αλλά μόνο να προσθέσει καινούργια. Αν αυτά τα καινούργια είναι λανθασμένα μπορούν εύκολα να διαγραφούν και το σύστημα να συνεχίσει να δουλεύει δημιουργώντας εκ νέου τα Views του με τα σωστά δεδομένα.

Simplicity Η αποθήκευση δεδομένων που μπορούν να μεταβληθούν στο μέλλον απαιτεί την ύπαρξη τεχνικών εύρεσής συγκεκριμένων δεδομένων. Απαιτείται δηλαδή κάποιο indexing. Αντίθετα η εγγραφή αμετάβλητων δεδομένων προϋποθέτει μόνο έναν μηχανισμό για Append στο master dataset το οποίο είναι πολύ πιο απλό στο σχεδιασμό και υλοποίηση του συστήματος.

Ένα από τα μειονεκτήματα του immutable dataset είναι οι μεγαλύτερες απαιτήσεις σε αποθηκευτικό χώρο καθώς δεν γίνεται κάποιος χωρισμός των δεδομένων σε πίνακες, πολλά στοιχεία μπορεί να αποθηκεύονται ξανά και ξανά με κάθε εγγραφή.

2.3 Perpetuity

Η συγκεκριμένη ιδιότητα επιβεβαιώνει αυτό που θεωρούνταν ως τώρα αυτονόητο, τα δεδομένα είναι και θα είναι για πάντα αληθινά. Data is eternally true. Το immutability δεν μπορεί να υποστηριχθεί χωρίς να ικανοποιείται το perpetuity. Η ιδιότητα αυτή καλύπτεται από την προσθήκη timestamp στις πληροφορίες που αποθηκεύονται στο master dataset, από την στιγμή που υπάρχει χρόνος αναφοράς για τα δεδομένα είναι εύκολο να ελεγχθεί και αν όντως ισχύουν.

2.4 Delete

Αναφέρθηκε ήδη μια περίπτωση όπου η διαγραφή των immutable δεδομένων είναι επιτρεπτή και μάλιστα απαραίτητη για την σωστή λειτουργία του συστήματος. Θα αναφερθούν ακόμα δύο περιπτώσεις στις οποίες το Delete μπορεί να εφαρμοστεί στο master dataset. Και στις δύο αυτές περιπτώσεις, η διαγραφή δεν έχει σχέση με την ορθότητα των δεδομένων αλλά με την αξία των δεδομένων.

Garbage collection Ο όρος είναι γνωστός και αναφέρεται στην συλλογή και εκκαθάριση άχρηστων δεδομένων. Παρομοίως μπορεί να εφαρμοστεί και στο master dataset για να μειώσει τον αποθηκευτικό χώρο που απαιτεί με το να διαγραφούν τα δεδομένα, ή τα κομμάτια των δεδομένων που έχουν την χαμηλότερη πληροφοριακή αξία.

Κανονισμοί Σε αρκετές περιπτώσεις ευαίσθητων και ιδιωτικών δεδομένων, απαιτείται η διαγραφή τους είτε μετά είτε κατά την συλλογή τους.

Σε όλες τις περιπτώσεις διαγραφής θα πρέπει να ληφθούν μέτρα για να ολοκληρωθεί η διαδικασία σωστά και χωρίς επιπλέον απώλειες δεδομένων. Συνιστάται η αντιγραφή ολόκληρου του master dataset, η εφαρμογή των αλγορίθμων για την διαγραφή των κατάλληλων δεδομένων και ο έλεγχος έχουν απομείνει τα σωστά δεδομένα χωρίς απώλειες. Το αντίγραφο μπορεί να διαγραφεί με ασφάλεια αφού δημιουργηθούν τα καινούργια Views και δοκιμαστούν με επιτυχία.

Κεφάλαιο 4

Πειραματική Διαδικασία

Οι φοιτητές που έχουν ασχοληθεί με πειραματικό αντικείμενο, στο παρόν κεφάλαιο μπορούν να περιγράψουν την πειραματική διάταξη, την τεχνική του πειράματος ή τη συλλογή των πειραματικών τους δεδομένων.

Είναι στην κρίση του συγγραφέα της εργασίας αν θα κρατήσει το κεφάλαιο ενιαίο ή θα το επιμερίσει στα δύο βασικά τμήματα που αφορούν τη διάταξη και την ανάλυση των πειραματικών δεδομένων. Η συνεννόηση με τον επιβλέποντα είναι πάντα αυτή που θα κρίνει την τελική διάταξη των κεφαλαίων.

Συστήνεται η προσεκτική χρήση γραφημάτων και πινάκων που συνοδεύουν την πειραματική διαδικασία. Για το λόγο αυτό, ακολουθούν μερικά παραδείγματα χρήσης αυτών των στοιχείων, τα οποία μπορεί να συμβουλευθεί ο αρχάριος στο L^AT_EX, μαζί με τις εκτενέστερες βιβλιογραφικές πηγές που υπάρχουν διαθέσιμες.

Επίσης, στο κείμενο υπάρχουν αναφορές στη βιβλιογραφία όπως συνήθως παρουσιάζονται στο διεθνή επιστημονικό τύπο, π.χ. οι [?, ?].

Κεφάλαιο 5

Αποτελέσματα και Ανάλυση

Το κεφάλαιο συνήθως συγκεντρώνει και σχολιάζει τα αποτελέσματα της εργασίας, είτε πρόκειται για θεωρητική εργασία είτε για πειραματική.

Είναι το σημαντικότερο κεφάλαιο στην εργασία, όπου γίνεται σχολιασμός και κριτική της εργασίας και των αποτελεσμάτων της και στο οποίο μπορεί να περιλαμβάνονται πίνακες, όπως οι Πίνακες 5.1 και 5.2

R. Feynman	1	2
P. Higgs	3	4
L. Boltzmann	5	6

Πίνακας 5.1: Ένα παράδειγμα 3×3 πίνακα με κεντρική στοίχιση και μία κάθετη διαγράμμιση

R. Feynman	1	2
P. Higgs	3	4
L. Boltzmann	5	6

Πίνακας 5.2: Ένα παράδειγμα 3×3 πίνακα με αριστερή στοίχιση και οριζόντιων και κάθετων διαγραμμίσεων

Για τις επιλογές και ρυθμίσεις των πινάκων, οι οποίες προσφέρουν ιδιαίτερα πολλές δυνατότητες, παραπέμπουμε το συγγραφέα στην αναφορά [?]. Για τις εικόνες ή σχήματα που περιλαμβάνονται στην εργασία ισχύουν ανάλογα, όπως το παρακάτω παράδειγμα (Εικ. 5.1).



Σχήμα 5.1: Η κεφαλή της Αθηνάς - Επίσημο λογότυπο του ΕΚΠΑ

Κεφάλαιο 6

Συμπεράσματα

Τα τελικά συμπεράσματα της εργασίας αναφέρονται (σχετικά) επιγραμματικά στο παρόν τμήμα, ώστε να δοθεί μια γενική σύνοψη του αντικειμένου της εργασίας, της διεξαγωγής της και του τελικού σχολιασμού και σύγκρισης με το ευρύτερο επιστημονικό αντικείμενο. Μελλοντικές κατευθύνσεις της ερευνητικής εργασίας και ανοιχτά ερωτήματα μπορούν επίσης να αναφερθούν στο τελικό κεφάλαιο.

Το μήκος του κεφαλαίου συνηθίζεται να είναι αρκετά πιο περιορισμένο σε σχέση με την ανάλυση και τη συζήτηση των αποτελεσμάτων. Ενδεικτικά αναφέρεται ένα μήκος 2–3 σελίδων που συνοψίζουν την εργασία.

Παράρτημα Α΄

Το λογισμικό

Τα παραρτήματα προσφέρονται για παράθεση στοιχείων που δεν αφορούν άμεσα την εργασία ή αποτελούν αυτοτελή μέρη της χωρίς να απαιτείται να ενσωματωθούν στον κύριο όγκο της.

Παράδειγμα παραρτήματος μπορεί να αποτελούν οι βασικές ρυθμίσεις κάποιου πακέτου λογισμικού που χρησιμοποιήθηκε κατά τη διάρκεια της εργασίας και κρίνεται απαραίτητο να αναφερθεί αναλυτικότερα.

Τα κεφάλαια των παραρτημάτων μπορεί να είναι όσο εκτενή επιθυμεί ο συγγραφέας της εργασίας, αλλά και όσο πολυάριθμα κρίνει.

Παράρτημα Β΄

Ορολογία

Παράρτημα Γ΄

Σύντομο βιογραφικό σημείωμα

Dr. Who, Τμήμα Φυσικής, ΕΚΠΑ

► **Προσωπικά Στοιχεία**

Έτος γέννησης: 2243 | Τόπος γέννησης: Άγνωστος

► **Σπουδές**

2263: Απολυτήριο, Τεχνολογική Ακαδημία Διαγαλαξιακής Συνομοσπονδίας

2268: Πτυχίο, Ανωτάτη Σχολή Οπτονετρονικής, Τομέας Επιταχυντών Higgs

2282–σήμερα: Τμήμα Φυσικής, ΕΚΠΑ, Τομέας Πυρηνικής Φυσικής & Στοιχειωδών Σωματιδίων

► **Επαγγελματική Εμπειρία**

2263–σήμερα: Διαγαλαξιακός Πράκτορας, επίπεδο Gh

► **Προσόντα και Διακρίσεις**

- Μέλος Λέσχης Κ' ραγα
- Γλώσσες: Δωρικά, Σμαθιονικά, Σφεέκιαν

► **Δημοσιεύσεις**

1. Dr. Who, *Effective tactics in the Tardis chamber against unknown wormhole storms*, J. Gal. App. 34, 243 (2265)

2. Dr. Who et al., *Technology in the Tardis*, Rev. Gal. App. 176, 1 (2264)

► **Συμμετοχές σε συνέδρια**

1. 53^ο Συμπόσιο Εταιρείας Διαγαλαξιακών Σπουδών, Ανδρομέδα Spa Resort, 2280