



Ημερομηνία Παράδοσης: 20-12-2017

Π. Βασιλειάδης

Η προγραμματιστική άσκηση για το μάθημα είναι **υποχρεωτική** και αφορά τη σχεδίαση, υλοποίηση και ρύθμιση ενός συστήματος λογισμικού. Η εργαστηριακή άσκηση προσφέρει **3 μονάδες** στον τελικό βαθμό του μαθήματος και εκπονείται σε ομάδες των **3 προσώπων**. Φυσικά, πρέπει να πιάσετε τουλάχιστον τη βάση στην εργασία, όπως και στο διαγώνισμα. Σε περιπτώσεις εξαιρετικών εργασιών, η επίδοση επιβραβεύεται με bonus στον τελικό βαθμό.

Το σύστημα πρέπει να υλοποιηθεί σε όλα τα επί μέρους στάδια.

Ταινιοθήκη

Θέλουμε να φτιάξουμε με ηλεκτρονική ταινιοθήκη. Για να μπορέσουμε να δουλέψουμε με πραγματικά δεδομένα, θα χρησιμοποιήσουμε το data set [hetrec2011-movielens-2k](http://www.grouplens.org) από το GroupLens research group (<http://www.grouplens.org>) και συγκεκριμένα το dataset "MovieLens + IMDb/Rotten Tomatoes" από το σύνδεσμο <https://grouplens.org/datasets/hetrec-2011/> (Το αρχείο readme.txt είναι πολύ κατατοπιστικό).

Για τις ταινίες κρατάμε στοιχεία όπως: ένα id, τίτλο, το id και το URL φωτογραφιών από (α) την IMDb και (β) το Rotten Tomatoes, καθώς και διάφορα στοιχεία από το Rotten Tomatoes, όπως τον αριθμό των σχολίων που έλαβε μια ταινία, το μέσο σκορ, ..., από (α) τους κριτικούς και (β) το κοινό. Μια ταινία μπορεί να ανήκει σε περισσότερα από ένα «είδη» (genres) και φυσικά, κάθε είδος έχει πολλές ταινίες που εμπίπτουν σε αυτό. Ένας σκηνοθέτης παρατίθεται για κάθε ταινία (για την ακρίβεια, μία συμβολοσειρά που μπορεί να έχει τα ονόματα όλων των σκηνοθετών της ταινίας, αν αυτοί είναι πολλοί). Παρατίθενται επίσης όλοι οι σημαντικοί ηθοποιοί μιας ταινίας. Μια ταινία έχει μία χώρα προέλευσης και πολλές τοποθεσίες στις οποίες γυρίστηκε. Υπάρχουν επίσης οι καταγραφές από χρήστες οι οποίοι έχουν κάνει 2 δουλειές: (α) έβαλαν «ταμπέλες» (tags) (user_taggedmovies.dat) και (β) έβαλαν σκορ (userRatedmovies.dat) στις ταινίες.

Η λίστα των πράξεων που μπορεί να κάνει ο χρήστης του συστήματός σας είναι η εξής.

ΠΡΑΞΕΙΣ ΦΟΡΤΩΣΗΣ ΔΕΔΟΜΕΝΩΝ

1. Ο χρήστης μπορεί να φορτώσει *όλη* την πληροφορία για τις ταινίες στο σύστημα, εκτός από ratings και tags. Αυτό σημαίνει ότι έχετε σχεδιάσει τις κλάσεις αναπαράστασης για τις πληροφορίες που είναι μέσα στα αρχεία και φορτώνετε *όλα* τα στοιχεία στο πρόγραμμά σας. Η πράξη αυτή είναι προαπαιτούμενο για όλες τις επόμενες πράξεις. **Προσοχή:** δεν αρκεί να φορτώσετε τα σχετικά αρχεία. Πρέπει ΥΠΟΧΡΕΩΤΙΚΑ να μπορείτε να διασυνδέσετε τα επί μέρους στοιχεία μεταξύ τους (π.χ., για κάθε ταινία θέλουμε μια συλλογή από τα είδη στα οποία ανήκει και για κάθε είδος μια συλλογή από τις ταινίες που εμπίπτουν σε αυτό. Αντίστοιχα με τους ηθοποιούς, σκηνοθέτες και τις χώρες προέλευσης και γυρισμάτων. Τα νούμερα από τα ratings, as well as the detailed filming locations μπορούν να είναι απλά πεδία. Οι ηθοποιοί και οι σκηνοθέτες έχουν και ένα μοναδικό χαρακτηριστικό στα σχετικά αρχεία, που μπορεί να χρησιμοποιηθεί για να αναπαρασταθούν μοναδικά.
2. Ο χρήστης μπορεί να φορτώσει την πληροφορία για τα tags ή τα ratings ή για αμφότερα (προσοχή εδώ ο όγκος της πληροφορίας είναι μεγάλος).

Πιθανώς θα αντιμετωπίσετε προβλήματα στο φόρτωμα των εγγραφών, λόγω προβληματικών εγγραφών: ο πιο απλός τρόπος είναι να διαγράψετε τις προβληματικές εγγραφές.

ΕΡΩΤΗΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

Ο χρήστης μπορεί να κάνει ερωτήσεις διαφόρων ειδών. Οι απαντήσεις υπολογίζονται από το σύστημά σας όπως προδιαγράφεται εδώ (πρέπει να αποφασίσετε εσείς πώς αναπαριστώνται) αλλά ανακτώνται από τον χρήστη όπως φαίνεται στην υποενότητα «ΠΑΡΑΓΩΓΗ ΚΕΙΜΕΝΟΥ».

3. Αναζήτηση συνοπτικής περιγραφής ταινίας. Ο χρήστης ερωτάται και πρέπει να προσδιορίσει για ποια ταινία ενδιαφέρεται (by Id ή by title). Ανακτώνται τα στοιχεία που φορτώθηκαν στην πράξη (1) με οργανωμένο τρόπο (διαλέγετε εσείς πώς) ώστε να εξυπηρετηθεί η μετέπειτα παραγωγή κειμένου.
4. Αναζήτηση όλων των ταινιών ενός είδους. Ο χρήστης ερωτάται για το είδος για το οποίο ενδιαφέρεται. Στην αρχή χρειάζεται να κατασκευαστεί μια συμβολοσειρά με μια πρόταση που περιγράφει τον αριθμό των ταινιών που ανακτήθηκαν. Μετά ακολουθεί η λίστα με τις ταινίες. Για κάθε ταινία επιστρέφεται τίτλος, ημερομηνία, σκηνοθέτης.
5. Αναζήτηση όλων των ταινιών μιας χώρας προέλευσης, ή ενός ηθοποιού, ή κάποιου σκηνοθέτη (by name). Ως άνω.
6. Καταγραφή των tags μιας ταινίας. Ότι και στην λειτουργία (3) η οποία όμως επεκτείνεται εδώ και με μια περιγραφή το πόσα tags συνολικά έχουν ανατεθεί σε μια ταινία καθώς και τη λίστα με τα tags. Για κάθε tag, user, date, tag value (προσοχή, όχι id, πρέπει να ανακτηθεί η σωστή συμβολοσειρά με βάση το id).
7. Καταγραφή των ratings. Ως άνω.

Σε κάθε απάντηση που παράγετε, χρειάζεται τίτλος και ένα αρχικό συνοδευτικό κείμενο.

Προσοχή: στην κατασκευή των Use cases θα βρείτε κι άλλα θέματα που αφορούν την συγκρότηση μιας ερώτησης παρακάτω στο κείμενο.

ΠΑΡΑΓΩΓΗ ΚΕΙΜΕΝΟΥ

8. Ο χρήστης μπορεί να ανακτήσει την απάντηση στην ερώτηση ως δομημένο κείμενο. Ο χρήστης, κατά την υποβολή της ερώτησης (και όχι κατά την παραγωγή του κειμένου) ερωτάται και πρέπει να προσδιορίσει σε ποιο από τους παρακάτω τρόπους θέλει την απάντηση:

- αρχείο κειμένου,
- αρχείο html,
- αρχείο σε ένα υποσύνολο της γλώσσας Markdown (βλ. παρακάτω) που περιλαμβάνει (α) τα 2 πρώτα είδη headings, (β) παραγράφους, και ενδεχομένως, (γ) bullet lists.
- αρχείο pdf (βλ. ευθύς αμέσως)

Ο χρήστης ερωτάται επίσης και για το όνομα του αρχείου που θα παραχθεί.

9. Ο χρήστης μπορεί να κατασκευάσει ένα pdf για το αποτέλεσμα μιας ερώτησης. Αυτό μπορεί να γίνει είτε ανακτώντας ένα από τα παραγχθέντα αρχεία της πράξης (8) και μετατρέποντάς σε pdf, είτε (β) εξάγοντας pdf από το αποτέλεσμα της ερώτησης. Συγκεκριμένα η είσοδος μπορεί να είναι:

- Συλλογή από συμβολοσειρές. Ο σχετικός κώδικας λαμβάνει ως παραμέτρους εισόδου μια λίστα από συμβολοσειρές, και το όνομα του αρχείου pdf που θα δημιουργηθεί.
- Αρχείο κειμένου. Ο σχετικός κώδικας λαμβάνει ως παραμέτρους εισόδου ένα αρχείο κειμένου που θα μετατραπεί σε pdf, και, ενδεχομένως, και το όνομα του αρχείου pdf που θα δημιουργηθεί. Αν δεν δοθεί όνομα, το αρχείο προορισμού έχει το ίδιο όνομα με το αρχείο κειμένου, αλλά το επίθεμά του είναι “.pdf”.

ΕΠΙΠΛΕΟΝ ΣΤΟΙΧΕΙΑ

Προφανώς, ο χρήστης θα πρέπει επίσης να μπορεί να αποχωρήσει ομαλά από το πρόγραμμα.

Σε όλα τα παραπάνω, εκτός κι αν προσδιορίζεται αλλιώς, το σύστημα επιστρέφει ένα κωδικό λάθους μικρότερο της μονάδας αν κάτι δεν πάει καλά (π.χ., το αρχείο εισόδου δεν υπάρχει, ή δεν γίνεται να αναγνωσθεί, ή το αρχείο προορισμού δεν μπορεί να γραφεί, κλπ) ή επιστρέφει κάποιο μη αρνητικό ακέραιο αν η πράξη εκτελεσθεί σωστά. Η τιμή επιστροφής πρέπει να έχει ένα νόημα (π.χ., πόσες εγγραφές αναγνώσθηκαν ή καταγράφηκαν) ή έστω 0 αν αυτό δεν είναι εύκολο. Ο περιορισμός αυτός στην τιμή επιστροφής, πάντως, σημαίνει ότι οι λίστες παραμέτρων των σχετικών μεθόδων θα περιλαμβάνουν και τα όποια αποτελέσματα πρέπει να περάσουν στην κύρια μνήμη.

ΥΠΟΧΡΕΩΤΙΚΑ όταν σχεδιάσετε το κομμάτι διαχείρισης Pdf κάντε το να είναι καθολικά ανεξάρτητο από τις κλάσεις της ταινιοθήκης! Αυτό θα σας βοηθήσει και στο μέλλον για να μπορείτε να χρησιμοποιήσετε και σε άλλα projects. Αυτό σημαίνει ότι οι μέθοδοι των σχετικών κλάσεων λαμβάνουν ως είσοδο είτε αρχεία, είτε συμβολοσειρές και όχι αντικείμενα σχετικά με ταινίες, ηθοποιούς κλπ. Υποχρεωτικά, το εν λόγω κομμάτι θα υπακούει σε ένα interface το οποίο θα προσδιορίσετε εσείς και θα χρησιμοποιείται από το υποσύστημα παραγωγής απαντήσεων.

Όταν επιλύσετε το σχεδιαστικό πρόβλημα για την παραγωγή των Pdf θα δείτε ότι μπορεί να εφαρμόσετε τη σχεδίαση και γενικότερα στην παραγωγή των απαντήσεων!

Διαχείριση αρχείων Pdf προγραμματιστικά

Θα χρησιμοποιήσετε τη βιβλιοθήκη *itext5* που διατίθεται ως free/open source software (F/OSS) κάτω από την AGPL license (<https://itextpdf.com/Products/itext-core>) και την οποία θα βρείτε στο <https://github.com/itext/itextpdf> μέσω του <https://developers.itextpdf.com/itext-java>.

Ο σύνδεσμος <https://github.com/itext/itextpdf/releases/latest> έχει ίσως την πιο ολοκληρωμένη εκδοχή για να κατεβάσετε.

Το βασικό jar που θα πρέπει να προσθέσετε στο build path της εφαρμογής σας είναι το *itext5-itextpdf-5.5.12.jar*.

Για να μπορέσετε να μπειτε στο πνεύμα, από το βιβλίο <https://www.manning.com/books/itext-in-action-second-edition> μπορείτε να χρησιμοποιήσετε τα κεφάλαια 1 και 6 και κυρίως τα παραδείγματα που δίδονται στο <https://manning-content.s3.amazonaws.com/download/b/e71cc48-85e7-4530-ace6-e8bd1a92e08c/SourceCodeIText.zip> καθώς και στο <https://developers.itextpdf.com/examples-itext5>.

Διαχείριση Markdown

H Markdown είναι μια απλή γλώσσα επιστημείωσης κειμένου με δομικά στοιχεία. Από Wikipedia:

Heading

Sub-heading

Paragraphs are separated by a blank line.

Bullet list:

- * apples
- * oranges
- * pears

Links: <https://guides.github.com/features/mastering-markdown/>, <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>, <https://en.wikipedia.org/wiki/Markdown>

Οδηγίες

Το σύστημα πρέπει οπωσδήποτε να έχει packages, με ένα χωριστό package για την αλληλεπίδραση με το χρήστη το οποίο, το μόνο που επιτρέπεται να κάνει είναι: για κάθε use case, (α) να ζητά τα κατάλληλα στοιχεία (input) από το χρήστη, (β) με βάση αυτό το input να ζητά από το υπόλοιπο σύστημα να εκτελέσει την ουσία της use case, και (γ) να παρουσιάζει το αποτέλεσμα. Τα (β) και (γ) πρακτικά μεταφράζονται στην κλήση μιας μεθόδου και τη λήψη του αποτελέσματός της. Κανένα κομμάτι υπολογισμού, ελέγχου, ..., δεν πρέπει να εκτελείται από το υποσύστημα αυτό.

Εκτός από το να υλοποιήσετε το όλο σύστημα, θα χρειαστεί να σχεδιάσετε από πιο πριν:

- Όλα τα use cases αναλυτικά
- Όλα τα packages
- Όλες τις κλάσεις (σε επίπεδο (α) ανάλυσης στην αρχή και (β) σχεδίασης έπειτα)
- Ελέγχους: (α) και unit tests για κάθε κλάση χωριστά, με ελέγχους για τις βασικές μεθόδους, και (β) integration test όπου τα βασικά use cases της εφαρμογής θα ελεγχθούν συνολικά

Επιβάλλεται να ακολουθήσετε τις βασικές αρχές ενθυλάκωσης (υποχρεωτικά), χαμηλής σύζευξης, DIP, OCP, abstract coupling, factories κλπ (όσο αυτό είναι εφικτό και εύλογο).

Υπόδειγμα αναφοράς: για τα επιμέρους στάδια, μπορείτε να συμπληρώνετε / αναθεωρείτε σταδιακά την αναφορά σας. Για διευκόλυνσή σας, υπάρχει ένα υπόδειγμα στο http://www.cs.uoi.gr/~pvassil/courses/sw_dev/exercises/TemplateFinalReport.zip

Υλικό. Όλο το υποστηρικτικό υλικό για το project (δεδομένα, βιβλιοθήκες, ...) βρίσκεται στο URL http://www.cs.uoi.gr/~pvassil/courses/sw_dev/exercises/supportingMaterial/

Τι θα γίνει στη διάρκεια του εξαμήνου

- Θα δοθούν περαιτέρω εξηγήσεις στην εκφώνηση (άρα το δημοσιευθέν αρχείο της εκφώνησης μπορεί να αλλάζει)
- Θα οργανωθεί [ΚΑΠΟΙΟ ΕΙΔΟΣ ΕΝΔΙΑΜΕΣΗΣ ΕΞΕΤΑΣΗΣ ΤΗΣ ΠΡΟΟΔΟΥ ΚΑΘΕ ΟΜΑΔΑΣ](#)
- Ενδεχομένως να [ΑΛΛΑΞΕΙ/ΕΜΠΛΟΥΤΙΣΘΕΙ ΚΑΠΟΙΟ ΜΕΡΟΣ ΤΗΣ ΕΚΦΩΝΗΣΗΣ!](#)

Χρονοδιάγραμμα

Στη συνέχεια παρατίθενται στάδια της ανάπτυξης, ενδιάμεσες προθεσμίες (milestones) και καταληκτικές ημερομηνίες ολοκλήρωσης (deadlines).

[02/10] Εκφώνηση	
3 weeks	Setup of Infrastructure
[22/10:: 23.59]	Εγκατάσταση Java (last v.), Eclipse (last v.) στους Η/Υ σας Για όσους έχουν δυσκολία με την Java: ασκήσεις επανάληψης Εξοικείωση και πειραματισμός με τις βιβλιοθήκες που σας δίδονται Εκκίνηση εργασιών στα Use Cases της εκφώνησης Δεν υπάρχει κάτι να παραδώσετε
2 weeks	Καταγραφή των Use Cases
[05/11:: 23.59]	Turnin: DLV1.1: First version of the report with all the use cases (as pdf file)
3 weeks	Design of classes && first implementations
[26/11:: 23.59]	Turnin: DLV2.1: Second version of the report (pdf) with (a) the class diagram (προαιρετικά: any other diagrams, too) + (b) traceability matrix of tests DLV2.2: tar file with the unit test classes + draft version of the actual classes Print: DLV2.1 to be handed over on the next day of the deadline
3.5 weeks	Complete implementation [ΕΝΔΕΧΟΜΕΝΩΣ ΕΞΕΤΑΣΗ ΕΝΔΙΑΜΕΣΩΣ]
[20/12:: 13.00] ΑΝΕΛΑΣΤΙΚΑ	Turnin : DLV 3.1: tar file with the code for all classes DLV 3.2: FINAL version of the report (pdf) with all the design and the documentation of the project
[~08/01]	ΕΞΕΤΑΣΗ Ακριβής ημερομηνία και λεπτομέρειες εξέτασης θα ανακοινωθούν αργότερα

Δεν θα ξεπεράσουμε το όριο των Χριστουγέννων. Η πράξη έχει αποδείξει ότι στις γιορτές οι ομάδες αποσυντονίζονται σε πολύ μεγάλο βαθμό. Έτσι, το deadline της παράδοσης είναι ανελαστικό. Θα πιεστείτε περισσότερο πριν τις γιορτές, αλλά θα φύγετε για τις γιορτές χωρίς το φορτίο του project.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!