

## Project 5: RML Mapping + GraphDB + RDF4J

Σκοπός της εργασίας είναι να υλοποιηθεί ένα πρόγραμμα σε Java για την μετατροπή και αποθήκευση ενός συνόλου δεδομένων της [IMDb](#) σε έναν RDF γράφο. Τα δεδομένα βρίσκονται στο elearning και είναι:

- **Αρχείο actors.json**
  - id: ένα id που καθορίζει μοναδικά έναν ηθοποιό
  - primaryName: το όνομα του ηθοποιού
  - birthYear: έτος γέννησης
  - knownForTitles: πίνακας με κωδικούς ταινιών στις οποίες έχει παίξει
- **Αρχείο movies.tsv**
  - tconst: το id της ταινίας
  - titleType: ο τύπος της ταινίας
  - primaryTitle: Ο τίτλος της ταινίας
  - startYear: έτος προβολής
  - runtimeMinutes: διάρκεια σε λεπτά
  - (τα υπόλοιπα πεδία μπορείτε να τα αγνοήσετε αν θέλετε)
- **Αρχείο ratings.tsv**
  - id: το id της καταχώρησης
  - average: μέσος όρος αξιολογήσεων
  - total: αριθμός αξιολογήσεων

Η εργασία αποτελείται από τα εξής βήματα:

1. **Σχεδιασμός μιας οντολογίας:** Να αναπτύξετε μια απλή οντολογία σε OWL ικανή να αναπαραστήσει την παραπάνω πληροφορία και σχέσεις, π.χ. ταινία, σχέση ανάμεσα σε μια ταινία και στις αξιολογήσεις της, κτλ.
2. **Δημιουργία του RDF γράφου με χρήση RML κανόνων:** Ορίστε κατάλληλους κανόνες στην RML για την μετατροπή των παραπάνω αρχείων σε RDF, ακολουθώντας το σχήμα της οντολογίας που έχετε ορίσει πιο πριν.
3. **Αποθήκευση των αποτελεσμάτων στην GraphDB:** Εγκατάσταση της GraphDB, δημιουργία ενός repository, αποθήκευση των προηγούμενων αποτελεσμάτων στο repository, ορισμός και εκτέλεση των παρακάτω ερωτημάτων:
  - a. Το όνομα της ταινίας με τις περισσότερες αξιολογήσεις
  - b. Το όνομα του ηθοποιού που έχει τον μεγαλύτερο συνολικό μέσο όρων αξιολογήσεων (άθροισμα μέσου όρου αξιολογήσεων των ταινιών του προς τον αριθμό των ταινιών του)
  - c. Ο μέσος όρος διάρκειας των ταινιών κάθε ηθοποιού

### RDF4J και Java

Παραδείγματα χρήσης της RML και του RDF4J σε Java υπάρχουν στο [github](#). Το πρόγραμμά σας θα πρέπει:

- Να εφαρμόζει τους RML κανόνες στα δεδομένα που σας δίνονται και να παράγει ένα αρχείο με τα αποτελέσματα

- Στην συνέχεια θα πρέπει να συνδέεται με ένα repository που θα φτιάξετε στην GraphDB με όνομα MyIMDb και να τα αποθηκεύει, μαζί με την αρχική οντολογία που έχετε φτιάξει.
- Τέλος θα πρέπει να εκτελεί τα SPARQL ερωτήματα και να τυπώνει τα αποτελέσματα.

### **Παραδοτέα**

1. Η οντολογία
  2. Το αρχείο με τους RML κανόνες
  3. Το αρχείο που παράγεται από την εκτέλεση των RML κανόνων
  4. Τα SPARQL ερωτήματα σε μορφή κειμένου (θα πρέπει να είναι ολοκληρωμένα και ready-to-run, δηλαδή να έχουν και prefixes)
  5. Screenshots με τα αποτελέσματα της εκτέλεσης των ερωτημάτων
  6. Το πρόγραμμα σε Java που έχετε αναπτύξει (zip or github link)
- 
- Δημιουργήστε ένα zip αρχείο με όνομα: <Firstname-Lastname-ID>-Project5.zip που θα περιέχει τα αρχεία σας.
  - Υποβάλετε το αρχείο στο elearning μέχρι τις 10/06/2022.