Assignment 4 – Document-oriented databases

Description

We will use MongoDB Community Edition 4.4 and the following collections:

Movies	People

{
    _id : … ,
    otitle : … ,
    ptitle : … ,
    adult : … ,
    year : … ,
    runtime : … ,
    rating : … ,
    totalvotes : … ,
    genres : […]
}
	{
	    _id : … ,
	    name : … ,
	    byear : … ,
	    dyear : …
	}

MoviesDenorm	**PeopleDenorm**

{
    _id : … ,
    actors : […],
    directors : […],
    producers : […],
    writers : […]
}
	{
	    _id : … ,
	    acted : […] ,
	    directed : […] ,
	    knownfor : […] ,
	    produced : […] ,
	    written : […]
	}

The _id fields are the ids used in the relational database. The genres fields are arrays with the names of the genres (not the ids). The rest of the arrays include person or movie ids. Empty arrays are not allowed. For instance, a movie without genres should not contain any genres field. The rest of the fields are those that translate directly from the relational database. Null fields are not allowed. For example, an alive person should not contain any dyear field. You can find a dump of these collections in myCourses. You must use the dump for the queries.

Your tasks

1.- Provide a Gradle project named 'IMDBSQLToMongo' to exchange the IMDB data from MySQL to MongoDB using both JDBC and MongoDB drivers. You must create the collections specified above. Check the template and grading software for more info. (40 points)

2.- Provide each of the following descriptions as a single aggregation query. Check the query templates and grading software to know how to provide these queries (a document containing the initial collection and the aggregation pipeline to be executed). Note that, if you use any kind of workbench with MongoDB, the queries may be significantly different. Your queries must work using the grading software (10 points per query).

2.1.- Recent Comedy movies (released between 2013 and 2021) with a rating greater than 6.5 that have more than four directors.

2.2.- Sci-fi movies directed by people named Kathryn with runtime less than 180 and at least two writers. (Hint: Use a regular expression as follows: /^Kat.*/)

2.3.- Alive writers who have produced at least one movie, and have written more than ten Action movies between 1975 and 1995.

2.4.- Ratings of Western movies with more than 10,000 votes and more than one genre, and produced by two or more alive producers that are also actors.

2.5.- People that are known for more than two Sci-Fi movies directed by more than three directors grouped by birth and death years, i.e., if two of these people have the same birth and death years, they should appear together. Also, these people must have been actors at least twice.

2.6.- Alive writers that have written more than 15 Romance movies and have directed at least two Sci-Fi movies.

Submission instructions

- Use the software template provided in myCourses.
- Submit a single ZIP file to myCourses that must be named as your RIT user, e.g., crrvcs.zip. Do not include '@rit.edu.' The file must contain a folder named 'IMDBSQLToMongo' containing your Gradle project, and a folder named 'Mongo' containing your MongoDB queries.
- Everything will be graded on a Linux machine, so you must always use the exact names provided in this document, software template and grading software.

Grading rubric

- Check the grading software.