



Task1 documentation

Candidati

Alice Nannini

Giacomo Mantovani

Marco Parola

Stefano Poleggi

Relatore

Prof. Pietro Ducange

Contents

1	Introduction	1
2	Analysis and workflow	3
2.1	Requirements	3
2.1.1	Functional requirement	3
2.1.2	Non-functional requirements	3
2.2	Use Cases	4
2.2.1	Use Cases Description	4
2.3	Analysis of entities	6
3	Design	7
3.1	Software architecture	7
3.2	Populating the database	7
3.3	Analytics and statistics	9
3.3.1	View the top production houses	9
3.3.2	Top rated movies by year and country	9
3.4	MongoDb index	9
4	Implementation	10
4.1	Used technologies	10
4.2	Replica setup	10
4.3	Java class description	11
4.3.1	Declaration of class Film	11
4.3.2	Declaration of class User	11
4.4	Create	11
4.5	Read	11
4.6	Update	11
4.7	Delete	11
4.8	GUI	12
5	User Manual	13
5.1	Admin Manual	14
5.2	Database design	15

1 Introduction

The **Cine-Valutami** application offers a search and information service in the field of cinema. When the application starts, the system requires authentication to use the service. The logged-in user can perform a search by entering the first characters of a movie title in the search bar, obtaining a list of 10 movies in the database. After that you can select one of the proposed titles or carry out a more in-depth search, adding characters. Selecting a row, the system allows you to view more information in the right section, including cover, title, director and ratings. The user can leave a mark from 1 to 5 for the selected movie. There is also a module for the system administrator: when he logs in, he's redirected to a different activity than the user's one. In this page, he can view some statistics linked to the movies and the searches carried out by the users of the application, such as the ranking of 10 most voted films or the 10 most sought after films. Moreover, the system admin can add new movies to the application database or delete those already present, by searching by title.

The mockup shows a user interface with two main panels. The left panel contains a search bar labeled "Search ..." and a list of ten items, each labeled "Film". The right panel displays a movie cover placeholder (a square with a mountain and sun icon), followed by labels for "Title", "Actor", and "Rate". Below these labels are five empty star icons for rating. At the bottom right of the interface is a "LOGOUT" button.

Figure 1: User Mockup

DELETE FILMS

Search ...

DEL

Film

Film

Film

Film

Film

Film

Film

ADD FILMS

Insert Json script...

ADD

LOGOUT

Figure 2: Admin Mockup

2 Analysis and workflow

2.1 Requirements

2.1.1 Functional requirement

The system has to allow the user to carry out basic functions such as:

- To sign up into the system.
- To login to the system.
- To search for a movie.
- To vote a movie.
- To view a list of the top rated movies, both in general and filtered by a selected country.
- To view a list of the top production houses.
- To view the top countries in making movies, filtered by a selected year.
- To view the most active users in the application, i.e. the users that have given most votes.

The system has to allow the administrator to carry out basic functions such as:

- To login to the system.
- To add a movie.
- To delete a movie.

2.1.2 Non-functional requirements

- Usability, ease of use and intuitiveness of the application by the user.
- Availability, with the service guaranteed 24/7, using replicas.
- The system should support simultaneous users.
- The system should provide access to the database with a few seconds of latency.
- Enforced consistency.

2.2 Use Cases

Actors

- User: this actor represents a user of the application
- Admin: this actor represents the administrator of the application

2.2.1 Use Cases Description

Event	UseCase	Actor(s)	Description
Log in, Log out	Login, Logout	Admin, User	The user logs in/out the application.
Display all the Films	Browse, Find, Display Films	User, Admin	The user chooses that he wants to view the list of Films. The system browses the data on the db and returns them on the interface.
View Statistics	View Top Rated Films, View Top Productions, View Top Film-maker Countries, View Most Active Users	Admin	The Admin clicks on the button to view the statistics. The system browses on the db the informations used in the calculation and display the result.
Add a film	Add Film	Admin	The admin submits the Film information. The system updates the db and the interface.
Update a film	Update Film	Admin	The admin selects the film and commits the new informations. The system updates the db and the interface.
Delete a film	Delete Film	Admin	The admin selects the film and submits the delete. The system updates the db and the interface.
View the film informations	Select Film, Display Film Info	User, Admin	The user selects the film. The system shows the film informations on the interface.
Vote a film	Vote Film	User, Admin	The user submits the vote on a selected film. The system updates the db and the interface.

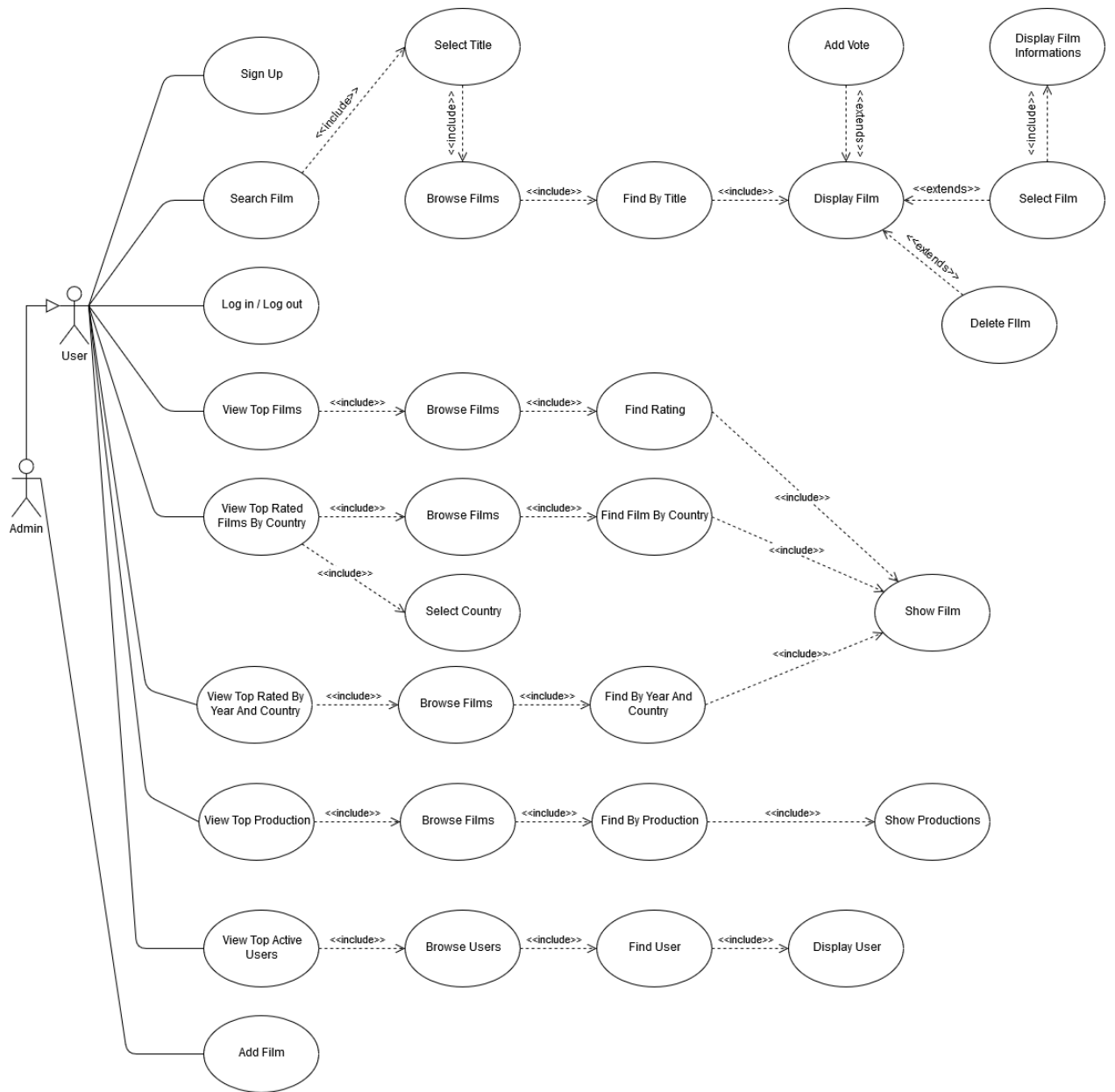


Figure 3: Use cases diagram

2.3 Analysis of entities

This diagram represents the main entities of the application and the relations between them.



Figure 4: UML analysis diagram

3 Design

3.1 Software architecture

The application is designed over 2 different layers, see figure 5:

- Front-end
- Back-end

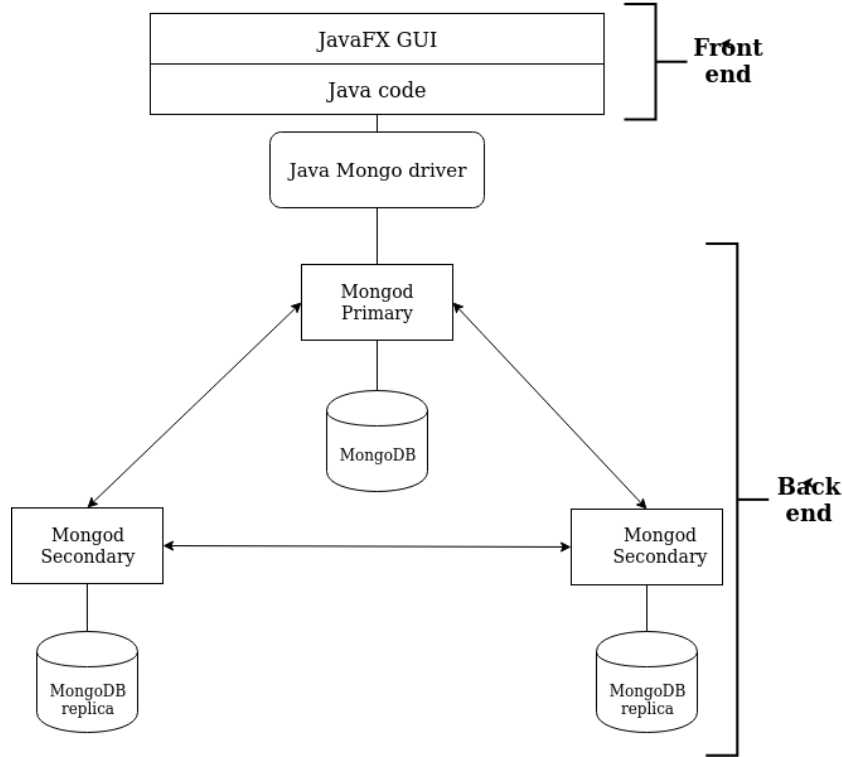


Figure 5: Software architecture diagram

3.2 Populating the database

The dataset used in the **Moviegram** application was created by scraping the Open Movie Database, using their API (<http://www.omdbapi.com/>). Given a movie title, this API returns a json document containing all the information in the OMDb about that movie. Before started building up the dataset using the API, it was necessary to obtain a list of movie titles, to be passed to the API; the list of titles was built up using different web pages as sources:

- a page of Wikipedia was scraped
- the pages referring to each year from 2007 to 2021 on the site <https://www.wildaboutmovies.com/>
- a list in the GitHub repository <https://github.com/fivethirtyeight/data>

Once the titles' list was ready, the only thing left to do was to obtain the key requested by the OMDb API to download the films' information. Each key provided by the database allows a user

to download 1000 film's information per day. Once the movies' list and the keys were ready, we could start to use our scraping application.

```
public class OmdbScraper {
    final static String key = "587a7b27";
    //final static String key = "84a209f9";
    public static void main(String[] args) throws IOException{
        int i = 0;
        String url = "http://www.omdbapi.com/?apikey="+key+"&t=";
        String path = "C:\\Users\\usr\\Desktop\\Unipi\\Large scale and multi-structured database\\progetto\\";
        String source = "source.txt";
        String destination = "movies_DB.json";

        try(BufferedReader bufferedReader = new BufferedReader(new FileReader(path+source))) {
            while(i < 1) {
                String movie = bufferedReader.readLine();
                System.out.println(i+": "+movie);
                Document doc = Jsoup.connect(url+movie).ignoreContentType(true).get();
                //Document doc = Jsoup.connect(url).ignoreContentType(true).get();
                Elements body = doc.getElementsByTag("body");//getElementsByClass("data");
                //System.out.println(body.text()+"\n");
                try(BufferedWriter bufferedWriter = new BufferedWriter(new FileWriter(path+destination,true))) {
                    String fileContent = body.text()+"\n";
                    bufferedWriter.write(fileContent);
                }
                i++;
            }
        }
    }
}
```

Figure 6: Java class of OMDBScraper

Using our “OmdbScraper” application, we managed to obtain the collection of movies’ information in a JSON format.

The format of our JSON elements are the following:

```

_id:ObjectId("5e05144f847f991da90d42c1")
Title:"Howl's Moving Castle"
Year:"2004"
Rated:"PG"
Released:"17 Jun 2005"
Runtime:"119 min"
Genre:"Animation, Adventure, Family, Fantasy"
Director:"Hayao Miyazaki"
Writer:"Hayao Miyazaki (screenplay), Diana Wynne Jones (novel)"
Actors:"Chieko Baishō, Takuya Kimura, Akihiro Miwa, Tatsuya Gashūin"
Plot:"When an unconfident young woman is cursed with an old body by a spitef..."
Language:"Japanese"
Country:"Japan"
Awards:"Nominated for 1 Oscar. Another 14 wins & 19 nominations."
Poster:"https://m.media-amazon.com/images/M/MV5BZTRhY2QwM2UtNmRlNy00ZWQwLTg3Mj..."
~ Ratings:Array
  ~ 0:Object
    Source:"Internet Movie Database"
    Value:"8.2/10"
  ~ 1:Object
    Source:"Rotten Tomatoes"
    Value:"87%"
  ~ 2:Object
    Source:"Metacritic"
    Value:"80/100"
Metascore:"80"
imdbRating:"8.2"
imdbVotes:"292,947"
imdbID:"tt0347149"
Type:"movie"
DVD:"07 Mar 2006"
BoxOffice:"$4,520,887"
Production:"Buena Vista"
Website:"N/A"
Response:"True"

```

Figure 7: Example of JSON document

We integrated our movies dataset with a plot dataset, found on kaggle (wiki_plot). In order to do so, we searched for matching between the plot's movie title and the movie title in our collection; once the matching was found, the plot field in the original element was replaced with the data presents in the plot dataset.

Moreover we manage some data types, because scraping OpenMovie platform, we obtain every fields in a string format, but to compute the analytics operations, describes on functional requirements section, we need to have some fields in a number format. We run the following javascript code on mongo shell, to perform the conversion:

```

db.collection.find().forEach(function(x){
  x.ratingImdb = parseFloat(x.imdbRating);
  db.collection.save(x);
});

```

3.3 Analytics and statistics

3.3.1 View the top production houses

This analytic is performed to display all production houses.

3.3.2 Top rated movies by year and country

3.4 MongoDB index

In order to speeding up some operations, we introduce some indexes ...

4 Implementation

4.1 Used technologies

The application is developed in java programming language, version 11.0.4, and in JavaFX system to create the GUI, version 11, so it should run on each platform in which JVM is installed, but the application is tested and guaranteed on Ubuntu 16 and Window OS. Moreover Maven is used to build and maintain the project, version 3.8.0.

The java driver for mongo manage the communication between client application layer and mongo backend layer, version 3.11.2.

For the backend layer it is used MongoDB, version 4.2.

So this application is tested using these technologies, considering these particular versions: for other versions the correct execution isn't guaranteed .

4.2 Replica setup

The following code shows how it is realized the architecture in which 3 mongo server are running in back end layer:

```
# ----- PREPARAZIONE CARTELLE PER LOG E PER I DB -----
sudo mkdir -p /srv/mongodb/rs0-0 /srv/mongodb/rs0-1 /srv/mongodb/rs0-2
sudo mkdir -p /var/log/mongodb/rs0-0 /var/log/mongodb/rs0-1 /var/log/mongodb/rs0-2

# ----- RUN 3 ISTANCES OF MONGOD -----
sudo mongod --port 27017 --dbpath /srv/mongodb/rs0-0 --replSet rs0 --oplogSize 128
--logpath /var/log/mongodb/rs0-0/server.log --fork
sudo mongod --port 27018 --dbpath /srv/mongodb/rs0-1 --replSet rs0 --oplogSize 128
--logpath /var/log/mongodb/rs0-1/server.log --fork
sudo mongod --port 27019 --dbpath /srv/mongodb/rs0-2 --replSet rs0 --oplogSize 128
--logpath /var/log/mongodb/rs0-2/server.log --fork

# ----- CHECK LISTENING PROCESS -----
netstat -tulpn

# ----- CONNECT TO PRIMARY MONGOD ISTANCE -----
mongo --port 27017

/* JS SCRIPT */
var rsconf = {
  _id: "rs0",
  members: [
    { _id: 0, host: "127.0.0.1:27017" },
    { _id: 1, host: "127.0.0.1:27018" },
    { _id: 2, host: "127.0.0.1:27019" }
  ]
};
rs.initiate( rsconf );
rs.conf();
rs.status();

# ----- KILL PROCESS USING PORT -----
sudo fuser -k 27017/tcp
sudo fuser -k 27018/tcp
sudo fuser -k 27019/tcp

# ----- ELIMINA CARTELLE -----
sudo rm -r /srv/mongodb/rs0-0 /srv/mongodb/rs0-1 /srv/mongodb/rs0-2
sudo rm -r /var/log/mongodb/rs0-0 /var/log/mongodb/rs0-1 /var/log/mongodb/rs0-2
```

4.3 Java class description

4.3.1 Declaration of class Film

The following java-code shows a declaration of the class Film.

<code here>

4.3.2 Declaration of class User

The following java-code shows a declaration of the class User.

<code here>

Another fundamental class is MongoManager, that manages the db-connection and the related operations. The implementation of a set of CRUD operations is described as follows.

4.4 Create

Adding one or more films to the database using a json document as input.

<Code here>

4.5 Read

This functionality returns a list of films searched by title. A film is returned if its title contains a piece of the searched title (case insensitive).

<code here>

4.6 Update

This operation allows users to update their vote about a selected film.

<code here>

4.7 Delete

This operation allows an admin to remove a selected film from the database.

<code here>

4.8 Analytics

In this phase the analytics operations that a user is able to perform on the database are shown.

4.9 GUI

There are three fxml documents, one for each page of the application, which describes the objects showed in the GUI interface of the related page.

- Home.fxml
- Login.fxml
- Register.fxml

In addition there are 3 classes, called Controllers, that are in charge of handling events of the objects defined in the associated fxml document.

- HomeController.java
- LoginController.java
- RegisterController.java

5 User Manual

When you first run the application, the interface you get is the login one, figure ??.

<login page image-> image name: screen0>

In case you are not registered you can click the link at the bottom of the page to be redirected to a register page, figure ??, otherwise you can sign in the application and by default you get the interface shown in figure ??

<register page image -> image name: screen1>

<Application home page after login -> image name: screen2>

From here you can search for a film by typing in the relative field and clicking the search button. You'll get a list of films that contains the text entered in the table below. Now you can select a film from the table and all the informations will be shown in the right pane of the application, figure ??. In the bottom right you are able to add a vote from 1 to 10 for the selected film.

<homepage after searching a film and selectiong one -> image name: screen3>

In the top left of the page there are two tabs. by default after the login you are in the Home tab, by clicking the Analytics Tab you will see the following page.

<analytics tab page -> image name: screen4>

Here you have a choice box (left) to select what kind of analytics you want to be performed from the existing ones, and another one to apply a filter (right), figure ??.

<default analytics page -> image name: screen5>

After selecting an analytics you will see the results in the table below, and for some of them you will get a piechart aswell, figure ??

<analytics page after performing analytics ->image name: screen6>

To log out, just click on the appropriate button at the bottom.

5.1 Admin Manual

If you have an admin user, you are entitled to make changes on the film lists. You need to log in inserting your username and password, and the application will recognize you as the administrator and show up a third tab called "Admin", figure ??

<page of the admin tab -> image name: admin0>

In the Admin tab you can search for a film the same way the user did in the Home tab and after selecting one you are able to delete it by clicking the "Delete" button. On the right pane you are able to insert a json document in the text field, that represent a film in the database, and by clicking the Add button you will insert one or more films in the database (fig. ??).

<admin tab after searching films and with a json document in the text field -> image name: admin1>

5.2 Database design

The application is developed over a no-SQL database, using the MongoDB platform.