



Assignment 1

Natural language Generation CS551H

Yassin Dinana

52094480

5th May 2021

Assignment 1 Description Note:

This project is built using Arria Studio for Natural Language Generation, the aim of the project is to create a short description of different actors using IMDB data. The data includes different elements such as the actor's biographical information as well as their previous works and awards.

There are different variables that are implemented in the system which help automatically generating the text, those variables include:

- **Movies:** This variable reads all the movies array and their data from the JSON file.
- **Highest Excellent:** This variable is used to extract the highest rated movie and print the rating.
- **Highest Excellent Name:** This variable is used to print the name of the highest rated movie.
- **Lowest Dreadful:** This variable is used to extract the lowest rated movie and print the rating.
- **Lowest Dreadful Name:** This variable is used to print the name of the lowest rated movie.

Looking at the compose view in Arria Studio, there is the *main* script which is used to call other subscripts that include the ATL functions used to generate the text, different subscripts include:

- **Attributes:** This subscript is used to describe the actor's biography, including his full name, nationality, birthplace, birthyear where we used the numberFormat function to adjust the numbers orders, and finally death year which is only called if the death year is not Null.
- **Reference:** This reference is used to generate the pronoun of the actor, if a male then the output is "He", if the gender is a female, it will automatically generate "She" as a pronoun.
- **CareerSummary:** This CareerSummary subscript is used to generate basic summary of the actor's career such as how many movies he appeared in and how many of his movies are rated.
- **ExcellentMovies:** This ExcellentMovies subscript is used to describe and list and movies that are ranked excellent, the excellent rated movies are rated above 8. The subscript also describe which movie is the highest rated.
- **AverageMovies:** This AverageMovies subscript is used to describe and list and movies that are ranked average rated, the average rated movies are rated between 5 and 8.
- **DreadfulMovies:** This DreadFulMovies subscript is used to describe and list and movies that are ranked dreadful or poor, the average rated movies are rated below 5. The subscript also describe which movie is the least rated.
- **Awards:** This Awards subscript includes a function that is built to generate what awards could be awarded to the actors as well as how many awards are given to the actor.
- **AwardsSummary:** This AwardsSummary subscript includes the types of awards that can be given to the actors such as Best Actor, Best Actress, or Best Supporting Actor / Actress.

The main script finally calls the subscripts mentioned above, when running the Main Script it generates the output paragraphs for all different actors.

