CS551H Assessment 1: Describe an Actor's Bad, OK, and Excellent Movies

In the week 1 practicals, you will build an NLG system which uses Arria NLG Studio to create short descriptions of actors and their acting career. In this assessment, you will expand the system created in the practicals (you can use your own work or my solution) so that it generates descriptions of an actor's career in terms of poor, OK, and good movies. For example

> Clint Eastwood is an American actor who was born in 1930. He acted in 52 movies, 51 of which were rated.
>
> Eastwood appeared in seven excellent movies. The best were Sad Hill Unearthed and The Good, the Bad and the Ugly, which had a rating of 8.9.
>
> Most of his movies were average. He appeared in 44 average movies.
>
> Eastwood did not appear in any dreadful movies.
>
> His later movies (average rating of 7.02) were better than his earlier movies (average rating of 6.91).

For this assessment, assume that any movie with a rating of 8 or higher is excellent, and any movie with a rating of less than 5 is dreadful; the rest are average. You should ignore movies that do not have a rating.

The ratings, incidentally, come from IMDB. They are slightly different from current IMDB ratings because we downloaded them a few years ago.

Marking

CGS D: System which produces a narrative that describes an actor's poor, average, and good movies in a sensible way**.** The system mostly works but occasionally crashes or produces narratives that are incorrect or not grammatical.

CGS C: High-quality robust system which produces a narrative that describes an actor's poor, average, and good movies in a sensible way. The system does not crash, and narratives are always correct and grammatical.

CGS B: High-quality robust system which produces a narrative which that describes an actor's poor, average, and good movies (as for CGS C), in a manner which includes the following

- Best and worst movies. If two or more movies have the highest (or lowest) rating, you should list all of them. For example, two of Clint Eastwood's movies have a rating of 8.9, so they are both listed as best movies in the above example.
- How the quality of an actor's movies has changed over time. Ignore small differences (less than 0.1 difference in average ratings), in such cases say the quality is similar and has not changed.
- From microplanning perspective, use variation, referring expressions, and possessives (eg `his` and `her`).

The system does not crash, and narratives are always correct and grammatical.

A system which produces robustly produces texts similar to the above example will get CGS B.

CGS A: High-quality robust system which produces a narrative that meets the requirements for CGS B, and gives more detailed insights about how the quality of an actors movies changed over time;

this should take awards into consideration as well as ratings. For example you could give a decade by decade summary of an actors career, and for each decade give average rating, best and worst movies, movies which won awards, and also a comparison with the previous decade.

General: Please note that your system must produce sensible descriptions for all actors in the development data set you were given. I will also test your systems on data for additional actors in a test data set.

Submission

Please submit your exported Studio project by midnight on Wednesday, 5 May, along with a note describing what you have implemented.

This assignment is part of the formal assessment of the course, and the work done must be that of your own group. You are reminded to look at the section on Cheating and Plagiarism in your student handbook.

A 10% penalty will be applied to late assessments submitted up to 1 working day late, and a 25% penalty will be applied to late assessments handed in between 1 day and 1 week late. No credit will be given to assessments submitted later than 1 week after the first deadline.