

Golden Globe Assignment Sheet

Your team has been given the task of creating a system to gather information from Twitter during the Golden Globes. In order to test it in advance, you have harvested a corpus of tweets from last year's awards. You will use those tweets to simulate the tweets you will get over the course of a typical Golden Globes event.

The system may pre-process information that is available prior to the Golden Globes. If that pre-processing takes some time to run, it won't cause any problems. But your extraction of information from the incoming tweets during the live Golden Globes needs to run rapidly. So, your processing of the corpus of tweets must take about a minute, or less.

Likewise, if there is any information that is the same every year for the Golden Globes, that information does not have to be extracted automatically.

On a future date (mid to late Feb), you have an appointment to meet with your boss, Larry, to do an informal live demo of the command-line version of your system. He'll want you to run your program in front of him, so the program should display its computed results. Here's what the system should do automatically:

1. Find the names of the hosts 2. For each award, find the name of the winner 3. For each award, find the name of the presenter(s) 4. He also asked you to wow him with something else to add spice to the

project. It's up to you what that is. It depends on what information about the event you think you can extract from the tweets. It should be something that you think people will find interesting or fun.

Some things are understood:

- NLP/AI is rarely 100% successful. If you don't manage to extract all of the information, that's okay. But you should aim for as much of it as possible.
- Items 1-3 are the minimum. The more you do beyond that, the more pleased your boss will be.