

Project Proposal #3

Group 1 proposes to create a filtering program based on the Bechdel Test. This test was invented by Alison Bechdel as a means of analyzing films through a feminist lens. The test gives films a pass or fail rating based on three criteria (Selisker):

1. The film must have at least two women in it
2. The women in the film talk to each other
3. The women talk to each other about something other than a man

The data that we propose to use for this Bechdel Test filtering program is in the form of a corpus assembled by linguistics researcher Cristian Danescu-Niculescu-Mizil and published by Cornell University in 2011. It is a collection of fictional conversations extracted from 617 movie scripts.

The raw data includes:

- 220,579 conversational exchanges between 10,292 pairs of movie characters
- involves 9,035 characters from
- in total 304,713 utterances

Along with the raw data, the corpus includes metadata that will be crucial to this project. The metadata includes:

- genres
- release year
- IMDB rating
- number of IMDB votes
- IMDB rating
- gender (for 3,774 characters)
- position on movie credits (3,321 characters)

Using this raw and meta-data, we propose to create a program that determines how many conversations in these films passed the Bechdel test. From there, it could be possible to cross-reference the number of these conversations with other metadata filters like genre, release

year, film rating, etc. This could give an interesting insight into how 'feminist' (passed Bechdel Test) films affect/ are affected by trends in the industry.

Selisker, Scott. "The Bechdel Test and the Social Form of Character Networks." *New Literary History*, vol. 46, no. 3, The Johns Hopkins University Press, 2015, pp. 505–23,

<http://www.jstor.org/stable/24542676>.

Cristian Danescu-Niculescu-Mizil. *Cornell Movie--Dialogs Corpus*. Cornell University.

https://www.cs.cornell.edu/~cristian/Cornell_Movie-Dialogs_Corpus.html

ZIP File