Maximilian Bragazzi Ihrén   maxbr431@student.liu.se

Pre-project Paper

What is your group's project about?

Our group will be making an n-gram based bot that will try to generate tweets like those of Trump. We haven't decided on which order of n-gram we are going to use, as we figure we will test which one seems to work best.

We will be testing our data by showing people two tweets, one actually from Trump, and one from the bot, and ask them which they think is real. We discussed some other methods we might use if we have the time for it, but that will be the main research point.

How will your group's project help you to accomplish the relevant learning objective?

The learning objective in question being: 'judge the difficulty and the feasibility of language technology applications'.

We will be using the data we collected above in order to evaluate how feasible a bot like we imagine would actually be. Will it (with the material taught in the course) be able to actually make somewhat coherent tweets? Of course, it might be more feasible with more advanced material, some research of our own, if we went through our data and classified words and such, and then used more advanced techniques, but since we are focusing on n-grams specifically it will limit our research scope.

That being said, I feel that evaluating the n-gram model for this type of a bot will not only be sufficient, but will definitely accomplish the learning objective.

What knowledge and what abilities will you personally be able to contribute to the project?

Since our group is a group of 6 people, evenly split between cognitive science (cogsci) students and innovative programming (IP) students, us IP students will focus mostly on the first part of the actual project, i.e. the programming that is required to make the n-gram bot and to download some of Trumps tweets to use as training data (this also includes tokenization, and other such sub-categories of course). This doesn't mean the cogsci students will be completely left out of the programming part, of course, but in a group project like this you have to play to each member's strengths and weaknesses.

On a purely individual scale there won't be much of a difference within the subgroup of IP students at least, partly due to all of us being on about the same level in knowledge about programming, and partly due to the fact that we are used to pair programming, which in turn will lead to less individual involvment.