

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

It's just 1 member so I'll be captain:

Name: Francisco Corona

NetID: fcoron5

2. What topic have you chosen? Why is it a problem? How does it relate to the theme and to the class?

I've chosen to develop a team building tool for the online Pokemon battle simulator "Pokemon Showdown" that makes recommendations based on the Smogon Forums.

Team building in competitive pokemon can be a daunting task for beginners as it is based around one's knowledge and understanding of many aspects of the metagame (what Pokemon are good, how to cover various matchups, which Pokemon counter or compliment others, etc). This tool seeks to make this entry process more straightforward by giving recommendations to users on what Pokemon / moves / items would make a good addition to their team based on information gathered from the Smogon forums (a forum for competitive Pokemon battling discussion).

The project relates to the theme of System extension since it seeks to use information retrieval techniques to enhance a service. Pokemon showdown already has some suggestion functionality for Pokemon stat spreads but teambuilding is another separate element that newer players would benefit much from. It relates to the class since it leverages principles of text retrieval such as web scraping, text analysis and recommender systems.

3. Briefly describe any datasets, algorithms or techniques you plan to use

Data will be gathered from the Smogon forums which will include various discussions, analysis and recommendations for team building.

For algorithms, BM25 would be used to rank relevant forum posts for team building based on a users interest with information such as pokemon they want to use, playstyle, etc being used for querying.

The techniques of tokenization and stemming would be used to preprocess a forum's content. Keyword extraction would also be used to identify what to recommend to the user based on relevant forum posts.

4. How will you demonstrate that your approach will work as expected?

To assess the tools performance, we can compare the suggestions' results to teams used by top ranked players in tournaments since these have team reports available with the necessary information to compare. This would give us an idea since good pokemon and good team compositions are more likely to be used at top level. We could also use a Pokemon's usage rate in tournaments (which is public information) to check if suggestions are relevant and we could this to assess results with common measures such as precision, recall, F1-measure and Mean Average Precision.

5. Which programming language do you plan to use?

Javascript, HTML and CSS would be used to create an interface for this team building tool which would act as a browser extension.

6. Please justify that the workload of your topic is at least  $20 \cdot N$  hours,  $N$  being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

The main tasks for functionality include:

- Website scraping / text mining
- Developing a recommendation system for relevant forum discussions
- Keyword extraction from relevant documents to make recommendations

I could already see each of these taking at least 5 hours and on top of that I'd also work on ui elements, debugging and documentation. As such, I believe this more than fulfills the requirement for amount of work.