

# SI650 Project Proposal

## Team member

Lingyun Guo, Tianyi Ge, Yueyuan Li

## Project Objective

We intend to build a search engine for TV series *Game of Thrones*. The search engine could satisfy its users' need as

1. Search for the specific episodes where all the query characters show up.
2. Search for the specific episodes that have required details such as *revenge*, *battle*, etc.
  - a. For *battle*, those episodes are not necessarily contain this term. But relative terms could be observed and make a conclusion that it's relevant to our query.
3. Search for the specific episodes where query lines occur (e.g. *You know nothing*).
4. Provide spell correction suggestions for special terms.
5. Support a fuzzy search of the scripts (approximate string matching).

## Why is it important/interesting?

If a TV show reviewer wants to analyze the details/hints/prompts embedded in plots by script writers, it usually takes a lot of effort to browse all the scenes/conversations, not to mention that manual check is likely to miss information. For example, in the TV show *Game of Thrones*, family has representative animals/symbols, which is usually used to imply characters' fate in inconspicuous conversations.

When searching on Google, the returned results are usually compiled/summarized by other authors/bloggers. If our search engine could automatically fetch such information, we believe it's useful to apply on other movie/TV show exploit

## Why do you believe it is doable?

1. Data source is the complete TV series scripts of GOT in text:  
[https://www.springfieldspringfield.co.uk/episode\\_scripts.php?tv-show=game-of-thrones](https://www.springfieldspringfield.co.uk/episode_scripts.php?tv-show=game-of-thrones)  
and can be downloaded by Python crawler.
2. Fuzzy search/approximate string matching has mature techniques like Levenshtein algorithm. Some useful tools are provided by Python package (fuzzywuzzy).

## Performance Evaluation

We plan to use two different criteria to evaluate our engine's performance: Google and the *Game of Thrones* Wiki. Our goal is to be more specific and to locate information directly to relevant episodes. We will evaluate our search engine by testing the same queries on the three platforms and comparing the accuracy of query location and how direct the results connect with correspondent episodes.