We set out to build an AI that would be capable of playing the Wikipedia Game to the level of a human player. This game involves traversal of the graph formed by websites at wikipedia.org from a random start node to a specific predetermined goal.

Our strategy included:
- Indexing simple.wikipedia.org
- Building a state-based search model for the graph
- Applying various simple search algorithms
- Building a feature extractor using NLP to tokenize each page
- Modeling the pages as sparse feature vectors
- Running various ML methods to learn to predict the link-distance between two states
- Applying the best predictor as a (non-consistent) heuristic for A*
- Running K-Means to cluster the states to improve feature vectors
- Applying PCA to visualize the state-space and search results

Although there were many facets to this project, not all were equally illuminating. Acquiring the data and running NLP software on it to featurize it, once the most difficult part of this type of project, was exceedingly simple. Even running a variety of ML algorithms on the data once it was modeled was not hard; this field is well explored. More interesting was modeling the search problem and using the learned predictor as a heuristic for A*. This felt like a tangible application which actually moved toward our stated goal of building a human-level AI for the game. The other unsupervised learning was also very illuminating. Both PCA and K-Means, but particularly PCA, offered valuable insight into the data and problem, and helped us to both visualize and understand the model. With the help of these methods, we accomplished our goal of bringing the AI to close to human-level state exploration (at the cost of the guaranteed minimal path).