Abstract: This project will be a service that allows a user to enter their idea for an invention into a simple web form. The back-end service will analyze the user’s input for similarities to other inventions that have been patented and return to the user a response indicating that the invention has either received a patent, and will include in the response at a minimum, the patent filing and the patent grant date. Additional information to be included, depending on the level of effort, could include a rating system (similarity rating) and other patent-related data

A response that there is no matching patent on file

The client-side will consist of a web form with form fields for collecting the search request and assign a search ID.

The server-side will consist of a web scraper (to scrape patents.google.com), a database, language processing pipeline and API calls using Flask. There might be other components added over time. Use cases are as follows:

Personas:

* Potential Inventor
* Patent Search Service (PSS, “The System”) returning the potential matches and a possible success-based rating

1. Front-end UI parser:

Given: Potential Inventor enters a string of words based on their invention idea

When: The user presses “submit”

Then: The patent-search service will accept the string and parse the string using SOTA NLP concepts.

1. Web-Scraper:

Given: The database of inventions for patents.google.com needs to be added to the server-side application

When: PSS scrapes the google database

Then: Essential elements of information (patent ID, filing date, etc.) will be added to the PSS database

1. HLT Pipeline

Given: The potential Inventor has entered a string via the form field

When: The user presses submit

Then: The PSS will remove unwanted elements from the string and store the cleaned/tokenized data as a database entry

1. Potential Inventor feedback loop

Given: The Potential Inventor has entered a search string and pressed submit:

When: the string has been pre-processed for NLP

Then: The UI will return the tokenized string as well as the collected search string and search ID (Depending on my investigation into the google patent data, I might limit some of the standard pre-processing).

1. Similarity Detection and Rating

Given: Potential Inventor has entered search string and the search string has been parsed by the PSS

When: The PSS detects a similarity to the Potential Inventor parsed request string

Then: The PSS will provide feedback to the Potential Inventor indicating that (a) a similarity exists to other patents and (b) a measurement of the similarity on a scale of 1 to 3. This numerical rating will operate in this fashion: 1 = high levels of similarity to other patents; 2 = neutral similarity rating; 3 = no similarity