IP-mine -- Patent Similarity Finder

Nalini Chandhi, Linfang Yang, Nilesh Bhoyar, Ganesh Sundararajan
W210 Capstone, MIDS, Fall 2017, University of California Berkeley
Project Website: http://patentfinder.herokuapp.com/patent

Project Goal:

IP-mine team has developed a web-based app that would help accelerate the process of patent filing of new claims/ideas to aid inventors monetize ideas.

Project Description:

IP-mine patent similarity match tool is a free web-based aid available for patent filing community to decipher new claims made by inventors. The app will provide early input regarding the similarity of their claims with respect to history of claims that have been approved by USPTO.

IP-mine app is very simple, fast and accurate to use. This helps everyone involved in the patenting of a new claim-idea complete the process in a speedy manner.

Schedule:

Task Number	Task Description	Duration in days	Predecessor
1	Project Statement and Analysis	15	None
2	Data Scraping and analysis	6	1
3	Unsupervised Learning Development	60	2
4	GUI development	15	2,3
5	Usability mock up tests	7	3
6	User Test	5	4,5
7	Final Product roll out	3	6

Project History:

- Thousands of new patents are filed every year and millions are granted already
- Sorting through the system to determine uniqueness of a new claim is time consuming and costly
- Patent filing community can use a tool that

- Uses the whole claim text to determine if user patent claim idea is unique enough to be patentable
- provide answers within reasonable time
- easy to use
- provide with just the right amount of information to enable inventor move forward in the process

Project Benefits:

- IP-mine helps sort through costly patenting process in an easy way with a simple user interface
- IP-mine is fast, accurate and simplifies the search process without restricting the user to search by keywords like other patent search solutions
- IP-mine Uses Machine learning to find matches and determine uniqueness

Project Challenges:

- USPTO Data Volume - Millions of patent claims are available from USPTO. Downloading and processing them will be a perennial challenge
- Text encoding and errors in XML files released by USPTO pose significant challenge. Workarounds to handle exceptions will be perennial challenge for every release from USPTO
- Validation data is not available from USPTO. Prior art numbers from each patent are the only reference available. But not all prior art numbers are available as many of them tend to be very old.
- Linearly scalable system for acceleration and storage

Future Roadmap:

- Enhance this patent similarity match solution to add patent keyword match to make it a complete solution
- Set up the data flows for weekly delta to be loaded
- Refine the UI and package the solution
- Deploy the solution on a scalable and faster hardware to meet the increasing data sizes and user load
- Come up with go-to-market strategy

Design Completion Status:
Planning Phase - 100%
Scoping Phase - 100%
Design Phase - 100%
Solution Scalability - 100%