



IP-mine

Machine Learning for
Patent Similarity Match

Ganesh, Lin, Nalini, Nilesch

W210 Fall 2107 Capstone

Problem Definition



- How do I determine if my patent claim idea is unique enough to be patentable?
- Instead of mining millions of patent from USPTO for matches, is there a tool that can help determine uniqueness of my claim?
- Will the tool provide answers within reasonable time?
- Will the tool be easy to use?
- Will the tool provide with just the right amount of information to enable me move forward?

Target Market for IP-Mine



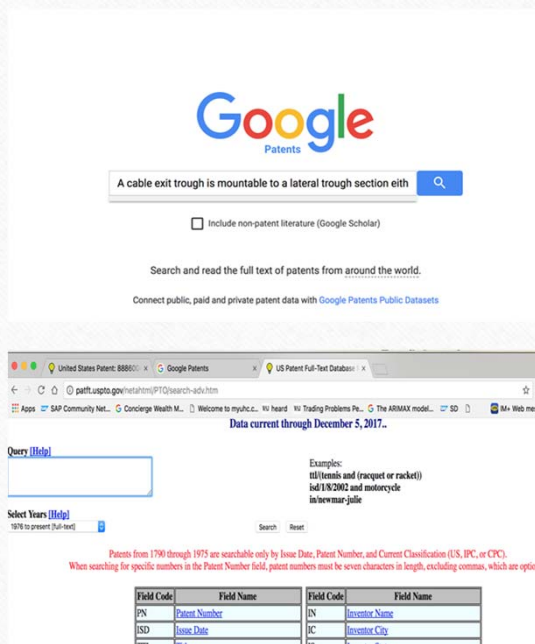
- Patent lawyers
- Patent departments within organizations
- Inventors searching for similarity of their claims with respect to patent claims that have been approved
- USPTO patent application process
- USPTO patent review process

Why IP-Mine is the need of the hour?



- Thousands of new patents are filed every year and millions are granted already
 - Sorting through the system to determine uniqueness is time consuming and costly
-
- IP-Mine helps sort through this costly process in an easy way with a simple user interface
 - It 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

Demo



Google Patents

A cable exit trough is mountable to a lateral trough section eith

☐ Include non-patent literature (Google Scholar)

Search and read the full text of patents from around the world.

Connect public, paid and private patent data with [Google Patents Public Datasets](#)

United States Patent: 8888000 X Google Patents US Patent Full-Text Database X

patft.uspto.gov/netahtml/PTOsearch-adv.htm

Apps SAP Community Net... Conderge Wealth M... Welcome to myhcc... No heard No Trading Problems Pe... The ARMAR model... SD M Web mess

Data current through December 5, 2017.

Query [\[Help\]](#)

Examples:
utl(tennis and (racquet or racket))
id:1/8/2002 and motorcycle
in:brewmar-julia

Select Years [\[Help\]](#)

1790 to present (full-text)

Search Reset

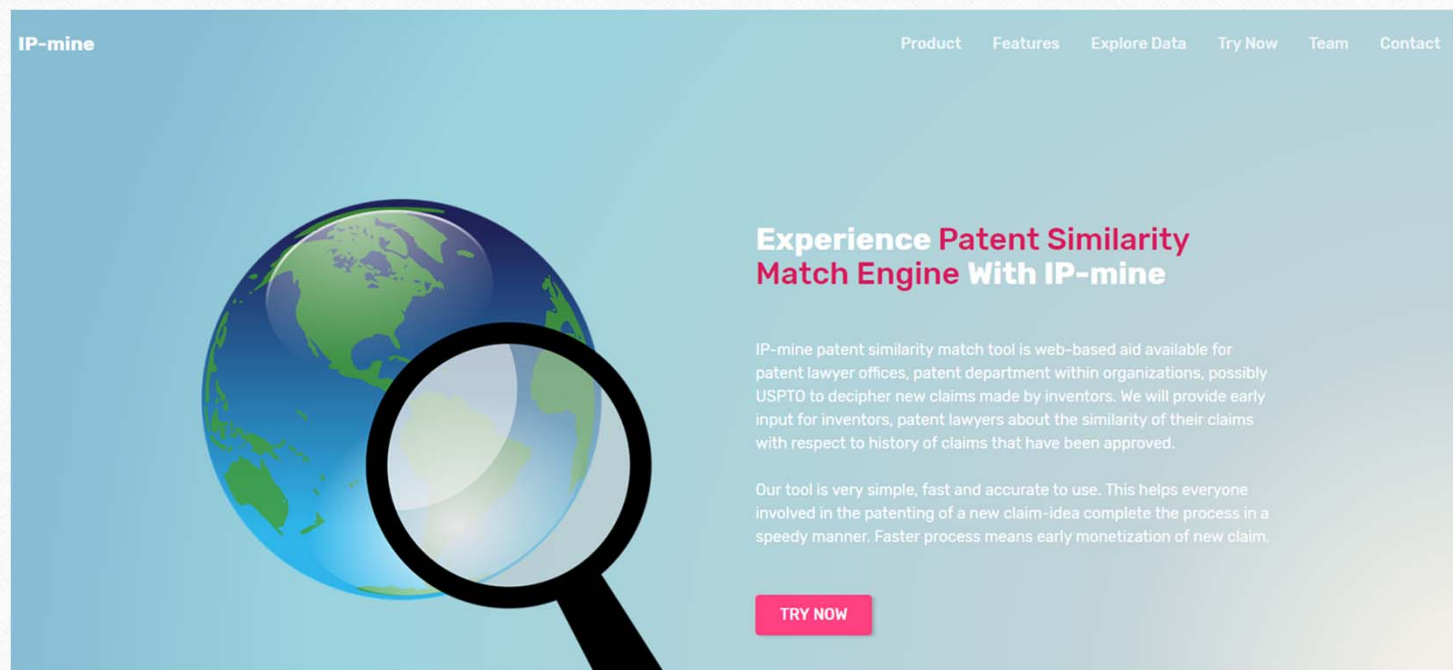
Patents from 1790 through 1975 are searchable only by Issue Date, Patent Number, and Current Classification (US, IPC, or CPC).
When searching for specific numbers in the Patent Number field, patent numbers must be seven characters in length, excluding commas, which are optional.

Field Code	Field Name	Field Code	Field Name
PN	Patent Number	IN	Inventor Name
ISD	Issue Date	IC	Inventor City
PFT	Private	ICR	Invention Class

- Most current tools expect the users to enter a specific search string or a word
- They also expect the users to know which predefined cluster/class the claim may belong to
- Patent attorneys or Inventors need to find and limit the key terms in their claims for search which can result in incorrect/inconsistent prior art search.

Demo

Application url: <http://patentfinder.herokuapp.com/>



Demo

IP-mine

[Product](#)
[Features](#)
[Explore Data](#)
[Try Now](#)
[Team](#)
[Contact](#)

Explore Patent Data

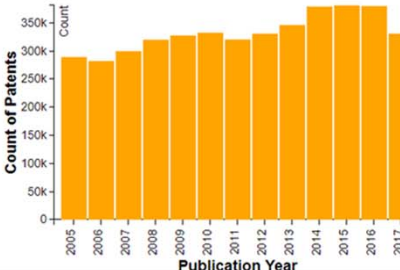
As we pre-classified all patent data into ten different first level clusters. And each first level cluster is also divided in to ten different secondary level clusters. Here you can explore the first level cluster data.

Please select a first level cluster to explore.

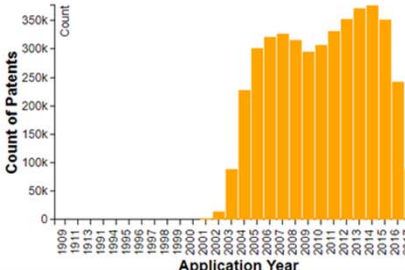
Selected All Clusters



Patents by Publication Year



Patents by Application Year



Demo

IP-mine

[Product](#)

[Features](#)

[Explore Data](#)

[Try Now](#)

[Team](#)

[Contact](#)

Search for Patents

Our well crafted search tool.

Step 1

Enter New Patent Claim text

New Patent Claim Text (for example A new and distinct cultivar of ornamental Hosta plant named Beckoning; as herein described and illustrated, with yellowish leaf center and blue-green margin, suitable as a potted plant, for the garden, and for cut flower arrangements.)

Step 2

Enter Search Parameter

Patent Claim Match Threshold %, Scale 0-100%. (for example 30 no need to enter unit)

Patent Result Get Top X, Scale 1-20 (for example 5)

Step 3

Start Search for Patents

SUBMIT

Demo

[back to patent search](#)

Patent Search

Patent Similarity Threshold (0-100)% you've entered: 30

Patent Search Result Get Top Count (1-20) you've entered: 5

Cluster info:

First level cluster 9, second level cluster 5

A new and distinct cultivar of ornamental *Hosta* plant named Beckoning; as herein described and illustrated, with yellowish leaf center and blue-green margin, suitable as a potted plant, for the garden, and for cut flower arrangements.

SEARCH

Similar Patents	Topic Word Cloud	Patents by Year	Patents by Time Taken	Patents by Kind	Patents by Claim
Appl. No.	Patent. No.	Match %	Title	Abstract	Claim Text
11388059	PP018371	64.80	Hosta plant named 'Beckoning'	A new and distinct Hosta plant named Beckoning characterized by heart-shaped, large-sized, leaves with blue-green margins, yellowish leaf centers, tall vertical flower scapes with many near white flow... ...more	1 . A new and distinct cultivar of ornamental Hosta plant named Beckoning as herein described and illustrated, with yellowish leaf center and blue-green margin, suitable as a potted plant, for the... ...more
10809724	PP015961	45.10	HOSTA PLANT NAMED 'GRAND MARQUEE'	A new cultivar of hybrid Hosta, Grand Marquee, characterized by its unique variegated foliage with wide blue-green margins and creamy white centers, its upright, mounding and relatively compact plantmore	1 . A new and distinct cultivar of hosta plant named Grand Marquee as herein illustrated and described.
13135847	PP024009	45.10	Hosta plant named 'Ben Vernooij'	A new and distinct Hosta cultivar named Ben Vernooij is disclosed, characterized by distinctive round foliage on thick petioles, vigorous plant habit and strong foliage, with good resistance to enviro... ...more	1 . A new and distinct cultivar of Hosta plant named Ben Vernooij as herein illustrated and described.
14544414	N/A	45.10	Hosta plant named 'Ambrosia'	A new and distinct cultivar of Hosta named Ambrosia, characterized by its large sized mounding clump, blue leaf margin with a light yellow center, the presence of a heavy glaucous wax over the entiremore	1 . A new and distinct cultivar of Hosta plant named Ambrosia as illustrated and described herein.
13135870	PP024001	39.48	Hosta plant named 'Forbidden Fruit'	A new and distinct Hosta cultivar named Forbidden Fruit is disclosed, characterized by distinctive large, bright foliage with a green margin, vigorous plant habit and strong foliage, with good resista... ...more	1 . A new and distinct cultivar of Hosta plant named Forbidden Fruit as herein illustrated and described.
Data collected from uspto.gov.					

Data collected from [uspto.gov](https://www.uspto.gov).

[back to patent search](#)

Similar Patents

Topic Word Cloud

Patents by Year

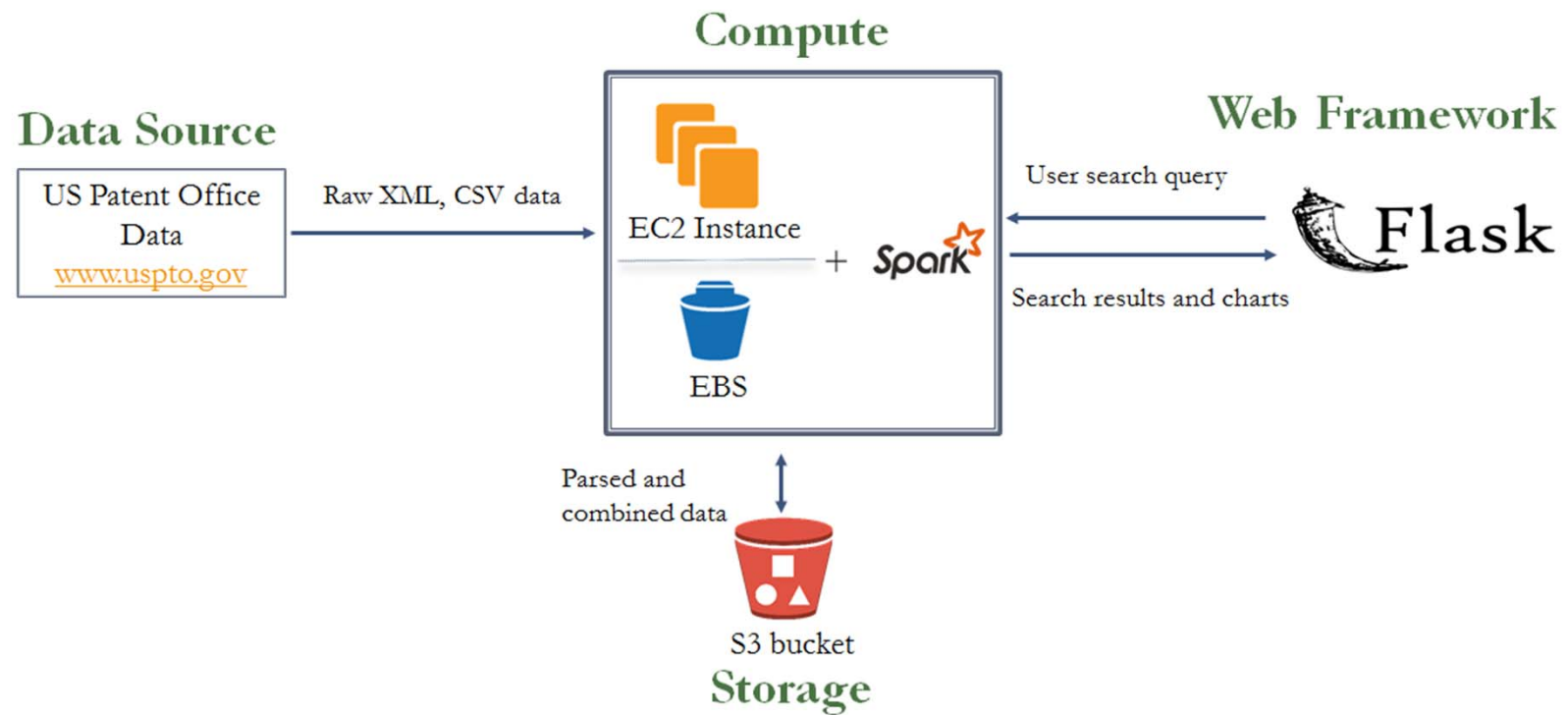
Patents by Time Taken

Patents by Kind

Patents by Claim

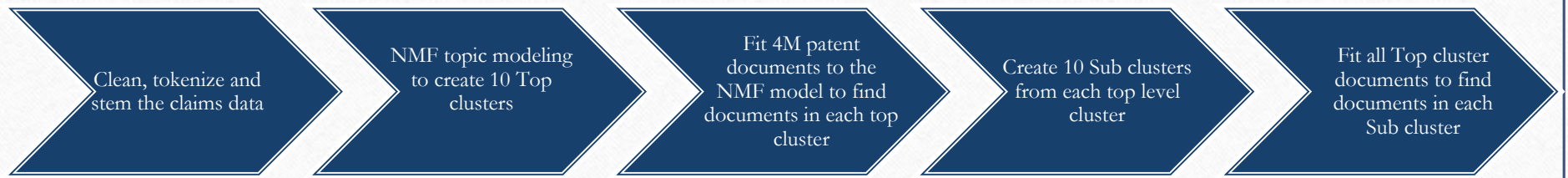


System Architecture

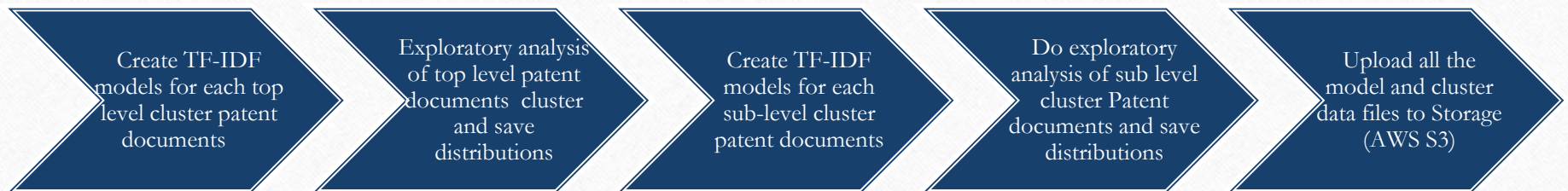


Machine Learning Models

Topic Modeling using Non-negative Matrix Factorization:



TF-IDF Cosine Similarity Calculation:



Topics and Clusters

0,comput user data execut instruct processor associ readabl store program identifi storag medium
1,member portion extend surfac posit bodi support second assembl engag said rotat mount outer op
2,layer substrat semiconductor electroth insul form conduct film region materi gate metal surfac
3,circuit voltag signal output power switch input control current suppli transistor connect elec
4,group compound composit alkyl formula atom carbon hydrogen weight substitut consist select sal
5,light optic emit reflect sourc len beam display wavelength imag illumin element devic laser su
6,fluid flow pressur valv chamber heat inlet outlet pump temperatur suppli liquid fuel control c
7,communic transmit network receiv wireless inform messag transmiss mobil station devic request
8,imag unit valu apparatus accord determin detect calcul predetermin process control time data b
9,cell sequenc acid said method protein nucleic amino antibodi polypeptid bind express administ



Top Level Clusters Topics

Sub Level Cluster Topics for First Top cluster



0,determin comput valu processor base readabl time instruct threshold medium second program execut plur
1,memori write data address control read unit block storag logic cach store second oper process access
2,transact account payment purchas inform custom merchant servic financi offer card user provid credit
3,search queri document object databas result data structur generat identifi attribut languag method el
4,content media video audio stream item advertis user file present digit playback devic receiv associ m
5,said step mean accord program code softwar product process adapt contain record method comput data co
6,applic server client request network servic comput resourc manag communic messag send receiv devic ex
7,game player play wager award outcom symbol win bonus card machin random event display reel plural pay
8,encrypt authent secur decrypt author verifi signatur devic public digit generat use privat access has
9,display user input devic imag interfac screen graphic select touch control electron posit detect seco

Future of IP-mine



- Enhance this patent similarity match solution to add patent keyword match to make it a complete solution
- Include a new feature of comparing two claims to provide a similarity score
- Set up the data flows for weekly delta to be loaded, refine the UI and package the solution
- Plan for a scalable and faster hardware to meet the needs of increasing data sizes and user load
- Come up with go-to-market strategy

“

Thank you

”

Our Team

Meet our awesome team.



Ganesh Sundararajan

Data Scientist



Nalini Chandhi

Data Scientist



Nilesh Bhoyar

Data Scientist



Linfang Yang

Data Scientist