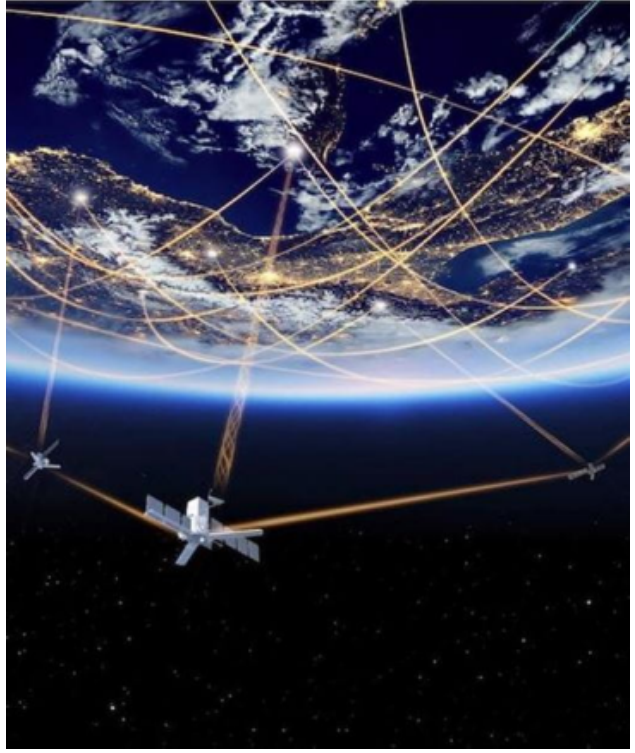





Raytheon Technologies

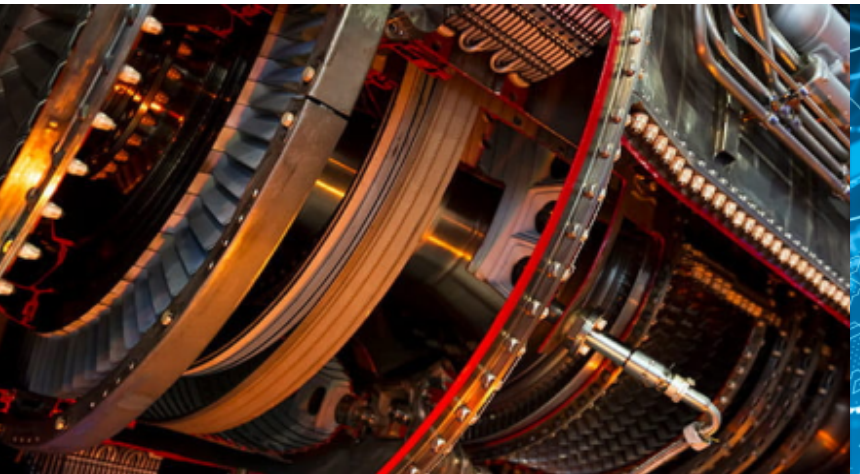
# Patent Similarity Detection for Raytheon Technologies

**Advisor: Gordon Dri**

**Presented By: Ahjeong Yeom, Akhir Syabani, Han-Yi Lin, Kenji Laurens**

# Introduction

We propose a framework to optimize the **prior art search**, a critical initial step in the patent application process. If Raytheon Technologies uses our **patent similarity search** system powered by **natural language processing**, they can reduce operational costs and time involved in patent application and perform efficient due diligence with researching previously approved patents (prior art search).

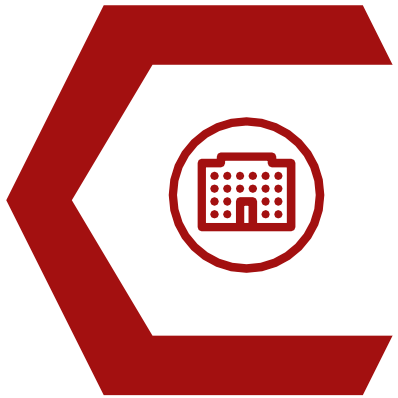


# Agenda

1. Background
2. Data Overview
3. Methodology
4. Application
5. Findings
6. Conclusion and Future Work

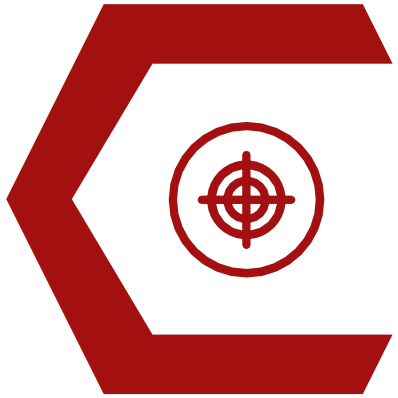
Raytheon Technologies

# About Raytheon Technologies



## Background

- An aerospace and defense company
- Provide advanced systems and services for commercial, military, and government customers
- Patents are critically important as they help promote technologically advanced inventions

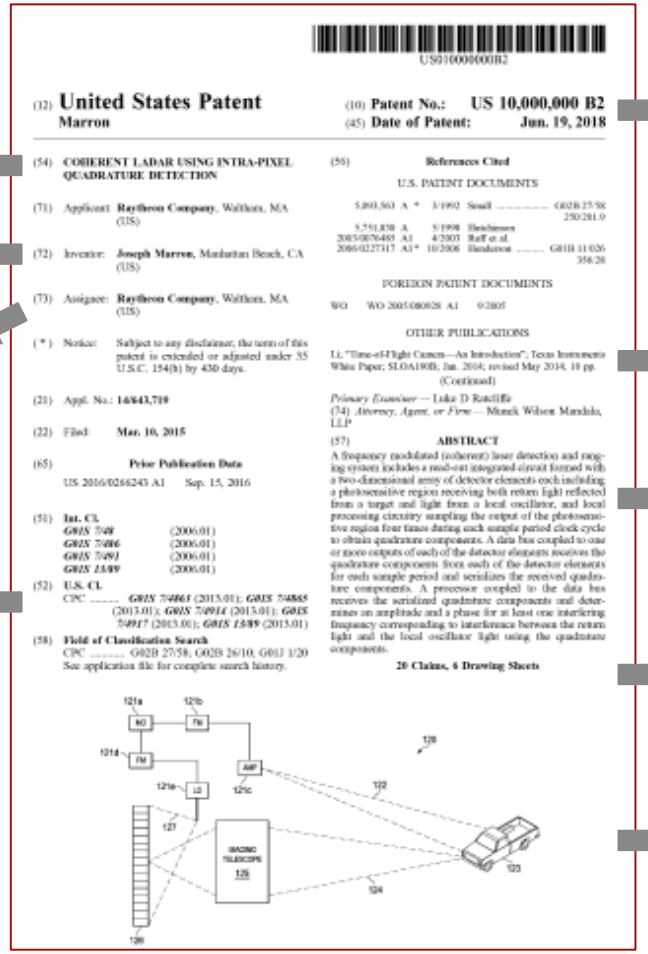


## Mission

- Streamline the patent application process

# What is a Patent?

**A type of intellectual property that gives its owner the legal right to exclude others from making, using, or selling an invention**




- Title
- Inventor
- Assignee
- Classification
- Patent no.
- References
- Abstract
- Claims
- Drawing

## Patentability

- Novelty ✓
- Non-Obviousness
- Usefulness

## CPC (Cooperative Patent Classification)

(A system developed by USPTO)

- F Mechanical Engineering
- G Physics
- H Electricity

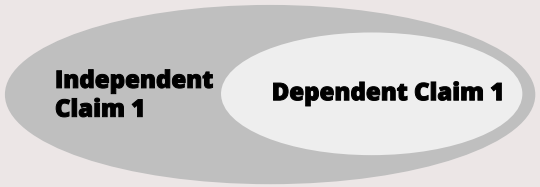
# Patent claims section is the most important part

## Claims

The **most important** part of a patent application

Set forth the **scope** of protection

**First claim** is used

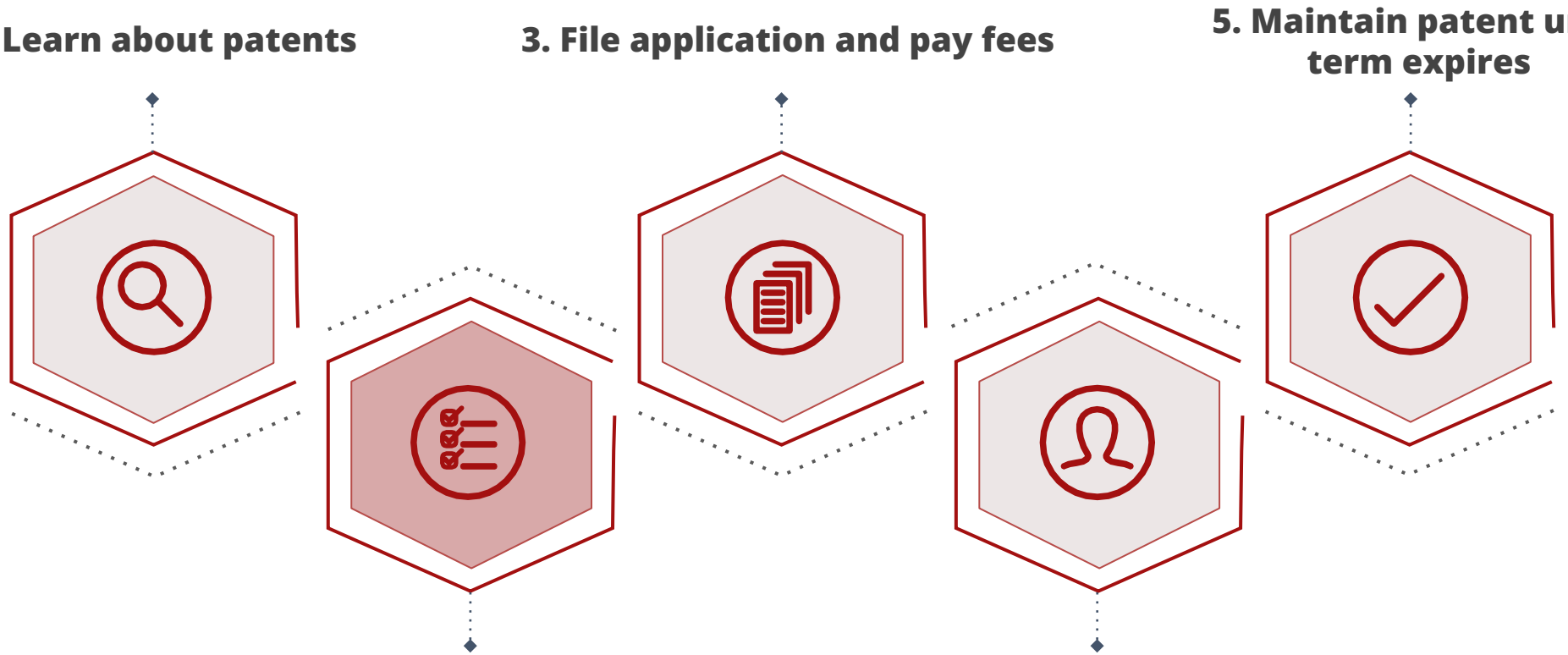



What is claimed is:

1. A laser detection and ranging (LADAR) system, comprising:
   a two-dimensional array of detector elements, each detector element within the array including:
      a photosensitive region configured to receive return light reflected from a target and oscillating local light from a local light source, and
      local processing circuitry coupled to an output of the respective photosensitive region and configured to receive an analog signal on the output and to sample the analog signal a plurality of times during each sample period clock cycle to obtain a plurality of components for a sample during each sample period clock cycle;
   a data bus coupled to one or more outputs of each of the detector elements and configured to receive the plurality of sample components from each of the detector elements for each sample period clock cycle; and
   a processor coupled to the data bus and configured to receive, from the data bus, the plurality of sample components from each of the detector elements for each sample period clock cycle and to determine an amplitude and a phase for an interfering frequency corresponding to interference between the return light and the oscillating local light using the plurality of sample components.
      2. The system according to claim 1, wherein the two-dimensional array of detector elements comprises a large format array.
      3. The system according to claim 1, wherein the plurality of sample components are quadrature components and wherein the quadrature components are employed to determine an amplitude and a phase for each of a plurality of interfering frequencies corresponding to interference between the return light and the oscillating local light.

# Patent application process is lengthy

## Prior art search is a critical step to evaluate novelty



# Goal: Detect similar patents and optimize prior arts search

## Business Problem

### Time

**2-3 years to issue patent**

### Cost

**~$700 per case to perform prior art search**

### Risk

**limited search capabilities**

## Opportunity

Make prior art searches more effective and time-saving, enabling Raytheon to **prioritize its potential patent submissions and further accelerating the patent submission process.**

# Project Highlights

## Natural Language Processing

Apply NLP techniques to extract important information from the text of patent claims.

**BERT** (Bidirectional Encoder Representations from Transformers)

Leverage a contextually sensitive BERT-based model further trained on patent documents.

## Information Retrieval

Access and retrieve appropriate information from text based on a particular query.

## Prior Art Search

Search related prior publications to check for the novelty of the patent application, based on keyword, classification, or citation.

## Semantic Search

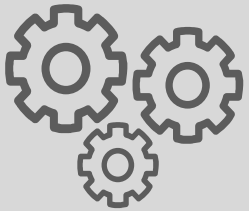
Use of semantic meaning of words and expressions, instead of exact keyword matches.

# Our pipeline processes over 30GB of data

## Extract

| | |
|---|---|
| Source | USPTO PatentsView |
| Year | 2011~2021 |
| CPC | F, G, H |
| Rows | 2.3 million records |

## Transform



- Claims text extraction
- Combined several dataset together
- Removed unused data
- Encoded texts to normalized 768 dimension array

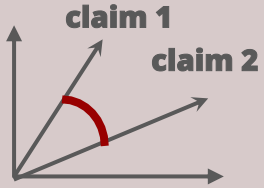
## Load

- SQL Database
- FAISS Index

# Pre-encoded data enables quick semantic search

## Measuring Novelty

- **Similarity Score**
  - Cosine similarity




- **Rejection Probability**
  - Based on cumulative distribution function (CDF) of rejected non-novel patents and their similarity scores with patents they infringe on

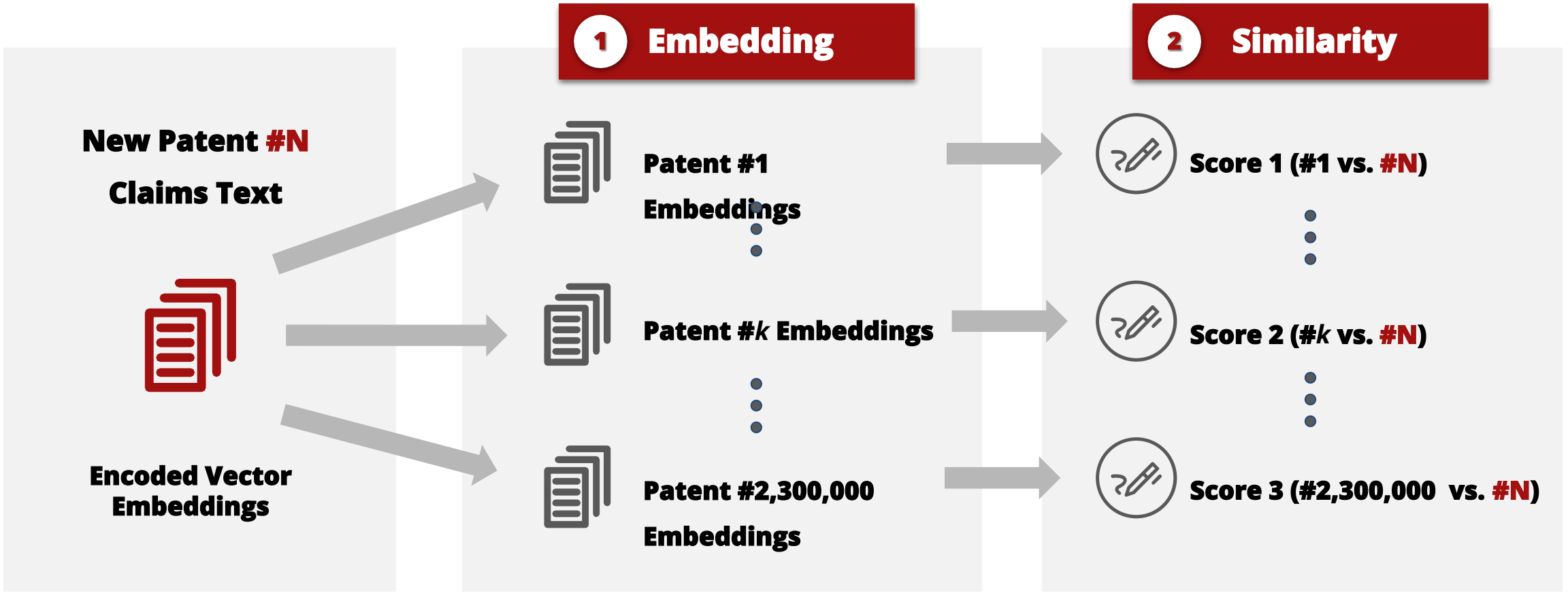
## Prep Work

- **2.3M Claims Texts**
  - Encoded into 768 dimension array over 10 hours with GPU
- **Embedding Model**
  - Pre-trained Sentence-BERT Model (**PatentSBERTa**)
  - BERT-based sentence transformer trained on USPTO patents

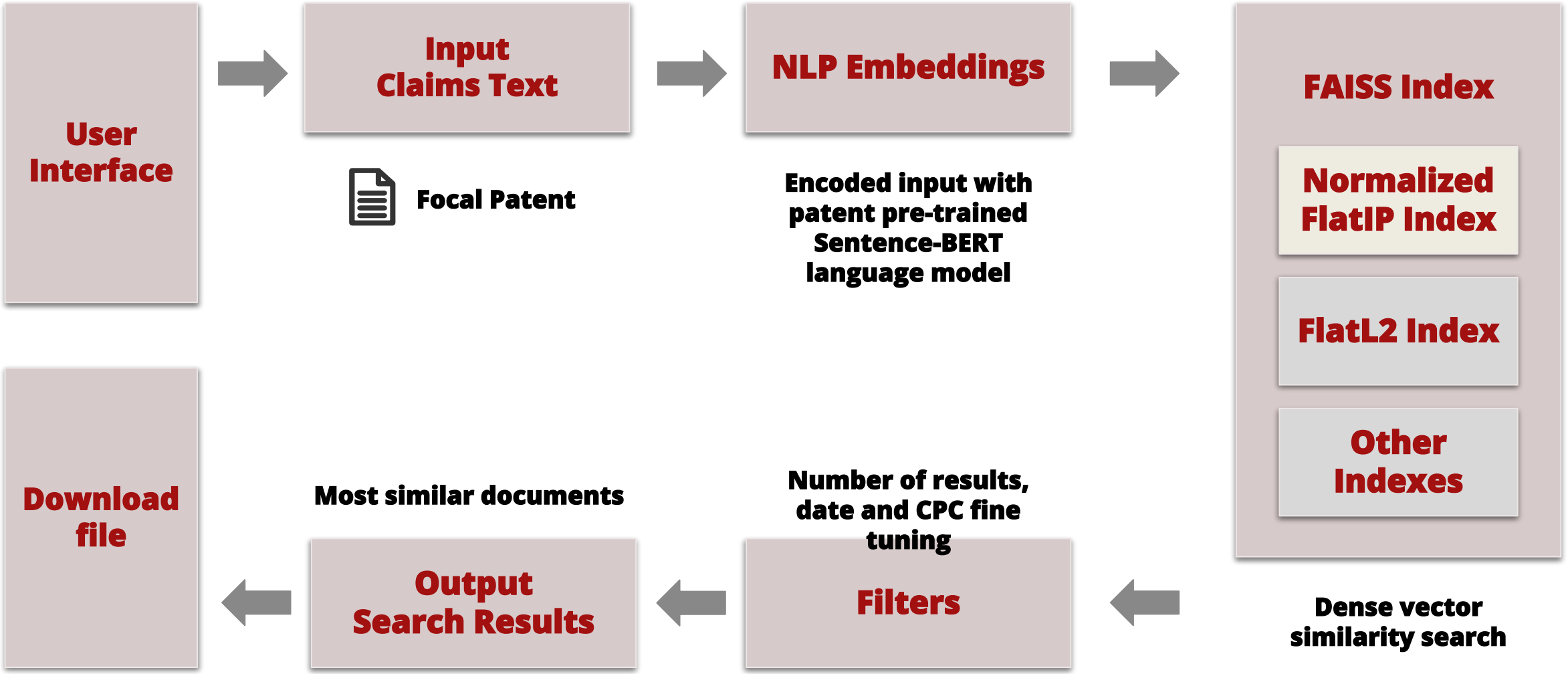
## Calculating Cosine Similarity

- **FAISS**
  **(Facebook AI Similarity Search)**
  - Efficient semantic similarity search library
  - Measures cosine similarity by calculating inner product of normalized vectors
  - Scales to the terabytes
- **Exhaustive Search**
  - Returns top 50,000 most similar text under 5 seconds

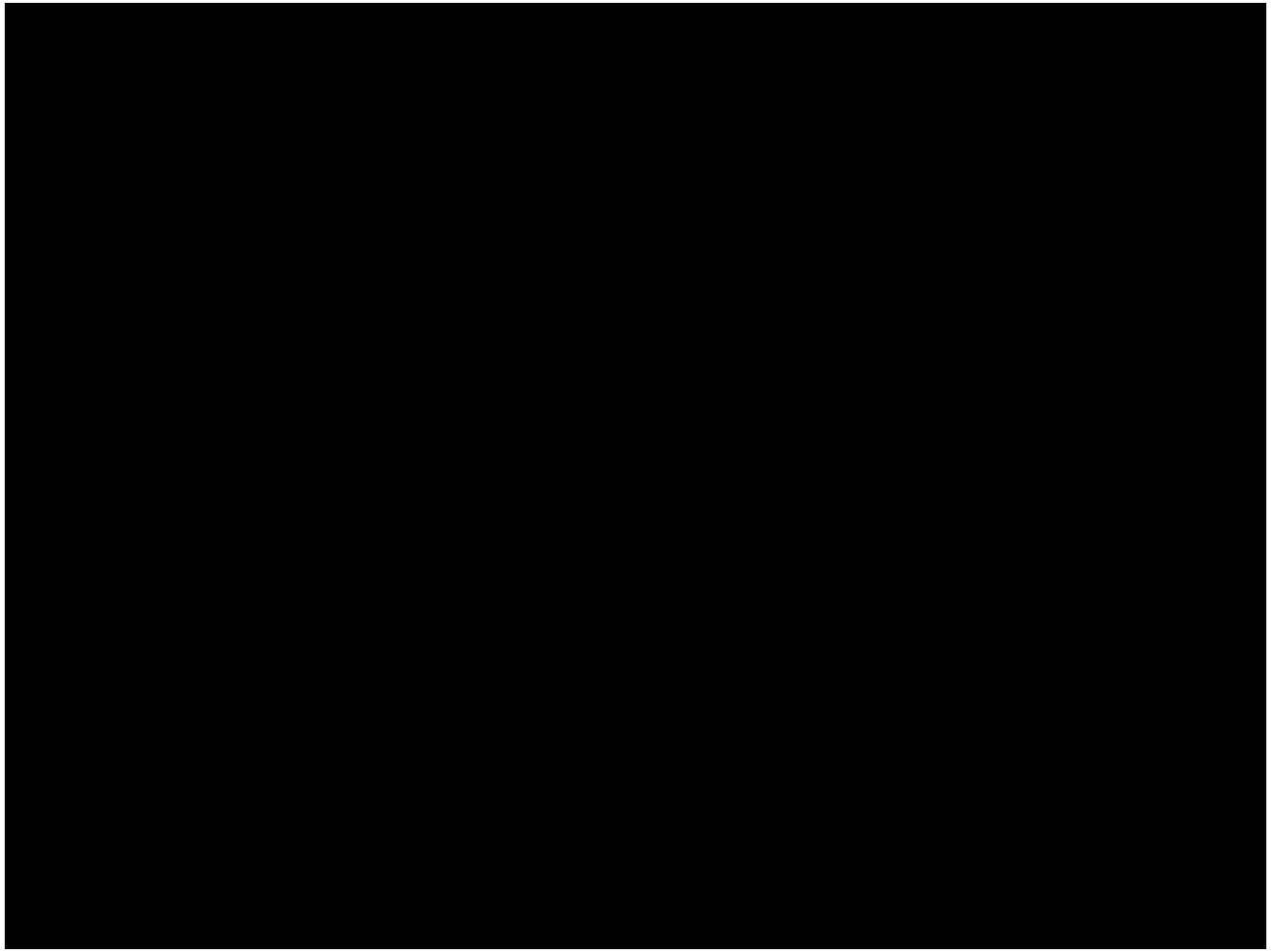
# Exhaustive search algorithm ensures all patents are scored accurately



# Search workflow made intuitive with frontend app



# Search workflow made intuitive with frontend app: Streamlit Demonstration



# Five embedding models are selected for evaluation

## SBERT Patent (PatentSBERTa) is expected to outperform other models

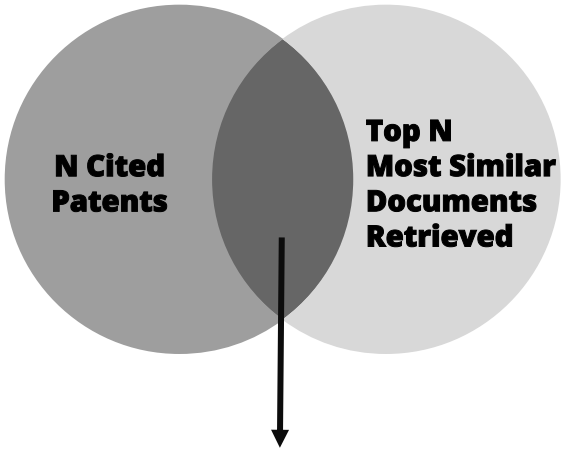
| Model | Output Dimension | Context Sensitive | Language Model Representation | Description |
|---|---|---|---|---|
| TF-IDF | ∞ | No | N/A | Relative importance/ relevance of words based on frequency |
| Word2Vec (google-news-300) | 300 | No | Word | Trained on Google News documents |
| SBERT Base Model (all-MiniLM-L6-v2) | 384 | Yes | Sentence | Base sentence-transformers model |
| SBERT Scientific (SPECTER) | 768 | Yes | Sentence | Trained on scientific journals |
| SBERT Patent (PatentSBERTa) | 768 | Yes | Sentence | Trained on USPTO patents |

# Evaluation 1: Retrieval Ability to measure how well each embedding can retrieve cited references

## Formula

**% Retrieval for Focal Patent X =**

(Top N Most Similar Docs Retrieved ⋂ N Cited Patents) ÷ (N Cited Patents)




Top N Most Similar Docs Retrieved ⋂ N Cited Patents

## Example

**Patent # 500**

**Cited Reference**

| | |
|---|---|
| # 121 | # 205 |
| # 133 | # 338 |
| # 158 | # 356 |

**Most Similar Documents**

| | |
|---|---|
| # 121 | # 260 |
| # 130 | # 338 |
| # 133 | # 356 |

% Retrieval for Patent # 500 = 4/6 = 67%

# Evaluation 1: SBERT Patent is able to retrieve the most cited references

| Embeddings | Citations Retrieval | | |
|---|---|---|---|
| | Avg % Retrieval 1 | Avg % Retrieval 2* | Ranking |
| TF-IDF | 13.92% | 30.18% | 3 |
| Word2Vec | 3.63% | 7.87% | 5 |
| SBERT Base | 15.93% | 34.54% | 2 |
| SBERT Scientific | 13.92% | 30.17% | 4 |
| SBERT Patent | 16.17% | 35.05% | 1 |

*Note: exclude patents where all of the tested embeddings are unable to retrieve any cited references ('bad samples')

# Evaluation 2: ROUGE F1 Scores to further validate the pairwise textual similarity

## Formula

**ROUGE Precision** =
(# Overlapping Words/n-grams) ÷
(# Words/n-grams in Most Similar Patent)

**ROUGE Recall** =
(# Overlapping Words/n-grams) ÷
(# Words/n-grams in Focal Patent)

**ROUGE F1 Score** =
2*(Precision*Recall) ÷
(Precision+Recall)

## Example

Focal Patent — "**ferrule** for optical transports"

vs.

Most Similar Patent — "**ferrule** having a unitary structure"

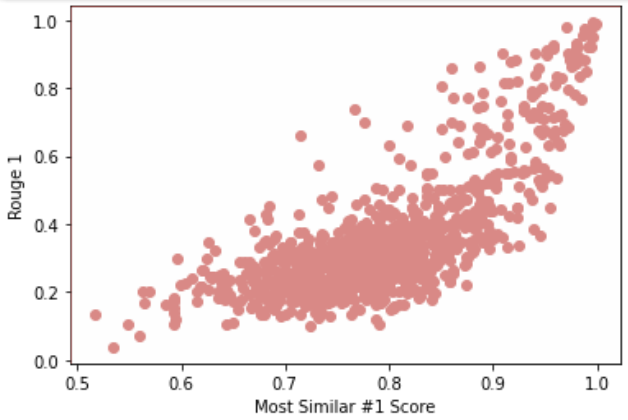
**ROUGE-1 (unigram):**

| | | |
|---|---|---|
| Precision | = 1÷5 | = 0.20 |
| Recall | = 1÷4 | = 0.25 |
| **F1 Score** | **= 2*(0.20*0.25)÷(0.20+0.25)** | **= 0.22** |

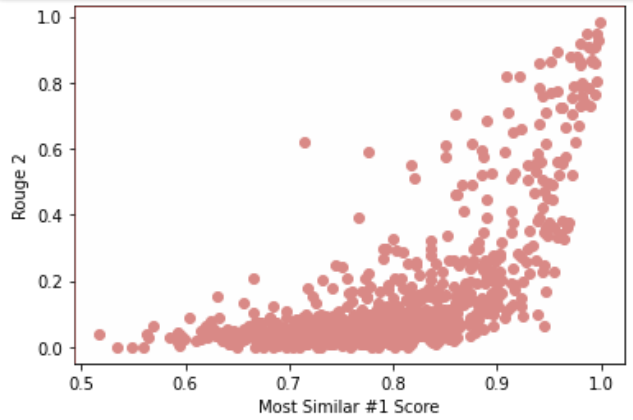
# Evaluation 2: SBERT Patent achieves the highest scores with strong correlation between Cosine Similarity and ROUGE F1 scores

| Embeddings | ROUGE F1 Scores - TOP 1 Retrieval | | | | | |
|---|---|---|---|---|---|---|
| | ROUGE-1 (unigram) | | ROUGE-2 (bigram) | | ROUGE-L (longest common subsequence) | |
| | Score (Avg) | Ranking | Score (Avg) | Ranking | Score (Avg) | Ranking |
| TF-IDF | 0.351 | 2 | 0.142 | 1 | 0.290 | 2 |
| Word2Vec | 0.308 | 5 | 0.115 | 5 | 0.254 | 5 |
| SBERT Base | 0.346 | 3 | 0.133 | 3 | 0.287 | 3 |
| SBERT Scientific | 0.340 | 4 | 0.130 | 4 | 0.281 | 4 |
| SBERT Patent | 0.352 | 1 | 0.136 | 2 | 0.295 | 1 |

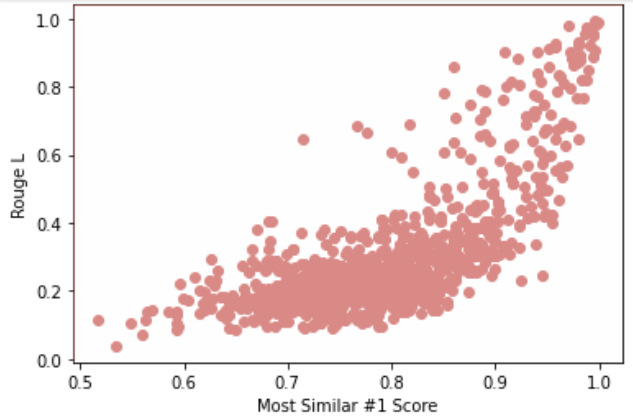
**ROUGE-1 (Pearson's r=0.8)**




**ROUGE-2 (Pearson's r=0.7)**




**ROUGE-L (Pearson's r=0.7)**



# Evaluation 3: manual verification as a qualitative approach confirms the pairwise textual similarity

## Example 1 (Cosine Similarity 0.96, Rouge-1: 0.93, Rouge-2: 0.78, Rouge-L: 0.93)

**US10724440B2**

A compressor, comprising: a disk-hub; a stator portion; a rotor portion coupled to the disk hub, the rotor portion adjacent to and aft the stator portion; an exit guide vane adjacent to and aft the rotor portion; a dual tangential on-board injector ("TOBI") disposed radially inward of the exit guide vane and the rotor portion, wherein the dual TOBI is configured to conduct a cooling flow to a section of the disk hub adjacent the stator portion.

**US10233840B2**

A compressor, comprising: a disk-hub; a stator portion; a rotor portion coupled to the disk-hub, the rotor portion adjacent to and aft the stator portion; an exit guide vane adjacent to and aft the rotor portion; a tangential on-board injector ("TOBI") disposed radially inward of the exit guide vane and the rotor portion, wherein the TOBI is configured to conduct a cooling flow to a section of the disk-hub adjacent the stator portion, the cooling flow isolated from a section between the rotor portion and exit guide vane by a seal.

# Evaluation 3: manual verification as a qualitative approach further confirms the pairwise textual similarity

## Example 2 (Cosine Similarity 0.71, Rouge-1: 0.21, Rouge-2: 0.04, Rouge-L: 0.21)

**US9435770B2**

A method of making a flexible **delay line** for an ultrasonic **transducer** having a face comprising: providing a mold having a desired geometry; applying a flexible material to the mold; curing the flexible material in the mold; and removing the cured material from the mold, wherein the cured material forms the flexible delay line having a geometry conforming to the face of the ultrasonic transducer, the flexible delay line having an aperture shaped to receive and surround the face of the ultrasonic transducer with the face of the ultrasonic transducer directly attached to the flexible delay line, the flexible delay line being operable to conform to a flat or an irregular **surface**.

**US7975549B2**

An apparatus for inspecting a workpiece having a curved **surface** with at least one predefined radius of curvature, the apparatus comprising: a housing; a plurality of **transducer** elements carried by the housing and positioned in an arcuate configuration having a predefined radius of curvature; an excitative source configured to concurrently trigger a respective plurality of the transducer elements such that the respective transducer elements emit signals into the workpiece, wherein the excitative source is configured to sequentially trigger different pluralities of the transducer elements, and wherein at least some of the transducers are included in at least two of the different pluralities of the transducer elements; and a curved **delay line** carried by the housing, the curved delay line having an outer arcuate surface having a predefined radius of curvature that matches the predefined radius of curvature of the plurality of transducer elements, the curved delay line also having an inner arcuate surface exposed to the workpiece via a corresponding opening defined by the housing, the inner arcuate surface of the curved delay line having at least one predefined radius of curvature that matches the at least one predefined radius of curvature of the curved surface of the workpiece, wherein an arc defined by the outer arcuate surface across which the transducer elements are positioned is longer than an arc defined by the inner arcuate surface.

# Semantic search speed is improved using FAISS with GPU Support

## FAISS - GPU

| | model | time (secs) |
|---|---|---|
| 0 | sbert_base | 5.829732 |
| 1 | sbert_sci | 8.479502 |
| 2 | sbert_pat | 10.370811 |
| 3 | sbert_pat_embeddings (92 first claims) | 2.393563 |
| 4 | index_flat_l2 | 0.000388 |
| 5 | index_flat_ip | 0.000278 |

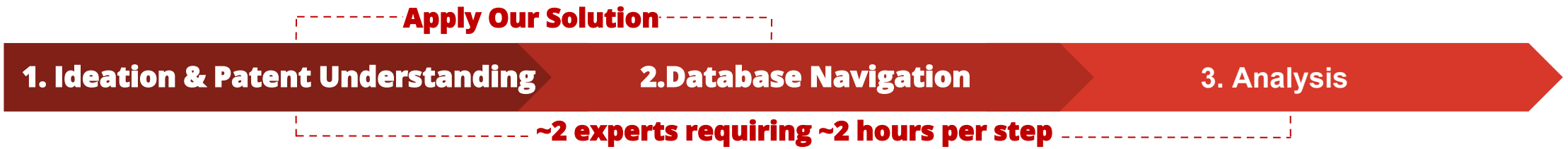
## FAISS - CPU

| | model | time (secs) |
|---|---|---|
| 0 | sbert_base | 39.554682 |
| 1 | sbert_sci | 321.553571 |
| 2 | sbert_pat | 319.480192 |
| 3 | sbert_pat_embeddings (92 first claims) | 106.646658 |
| 4 | index_flat_l2 | 0.000429 |
| 5 | index_flat_ip | 0.005429 |

Note: testing on semantic search of one focal patent against 92 documents (cited references + random)

# Our solution is fast, cost-effective and mitigates risks

**Target Audience:** Patent applicant interested in leveraging AI to assist with prior art search




| | Original | Our Solution |
|---|---|---|
| **Time** | ~12 hours* | ~5 sec per search |
| **Cost** | $500~$700 per report (cost scales with complexities) | 33% reduction at min** |
| **Risk** | The validity of the patent at risk due to limited search capabilities*** | With data update, search capacity can scale with data size |

** RTX average time is approximately 6 hours and involving 1~2 legal experts per case, incurring 12 hours on average*
*** The solution still requires human judgment so we assume to cut down 4 hours spent on ideation and database navigation steps taken ( 33% reduction)*
**** An infringement lawsuit, average of $2.3 million and $4 million for patent litigation case.*

# Our solution makes client's prior art search efficient

## Semantic Similarity

Our solution uses **semantic similarity search** to find the most similar patents to accelerate prior art search process for client.

## Sentence BERT

Our solution uses **Sentence BERT** trained on patents, in order for machine to understand the patent context. It optimizes speed compared to regular BERT models.

## Application

Our solution provides a **user-friendly interface** to display top similar patent results in under **5 seconds** thanks to FAISS search engine.

## Impact

Our solution **reduces operational costs and time** on patent application, increases the efficiency of prior art search, and provides easy-to-use interface.

# Limitations can be overcome with future works

| Limitations | Future Works |
|---|---|
| - Computational resources (memory, hard drive, GPU) | - **Extend the scope** from claims to other sections in the patent (full claims, abstract, drawings etc) |
| - Evaluation was limited due to no labeled dataset<br>- Validation metrics (ROUGE) originally used for text generation task were used as proxies | - Create an expert-verified labeled dataset and further **fine–tune SBERT model** |
| - The final product still requires human judgment for operationalization | - Increase the patent dataset and develop **an automated ETL pipeline** to handle growing prior art search corpus |
| - Given that we were not given patent-in-progress data from Raytheon, our solution could not be validated against client's specific case | - Further **optimize the speed** with more data using approximate nearest neighbor search (ANN) |

# Acknowledgement

We gratefully acknowledge the assistance and support from Raytheon Technologies and The University of Chicago.

Client
**Bryant Beeler**

Advisor
**Gordon Dri**

Supervisor
**Utku Pamuksuz**

Subject Matter Expert
**Nick Kadochnikov**

**Ahjeong Yeom**

**Akhir Syabani**

**Han-Yi Lin**

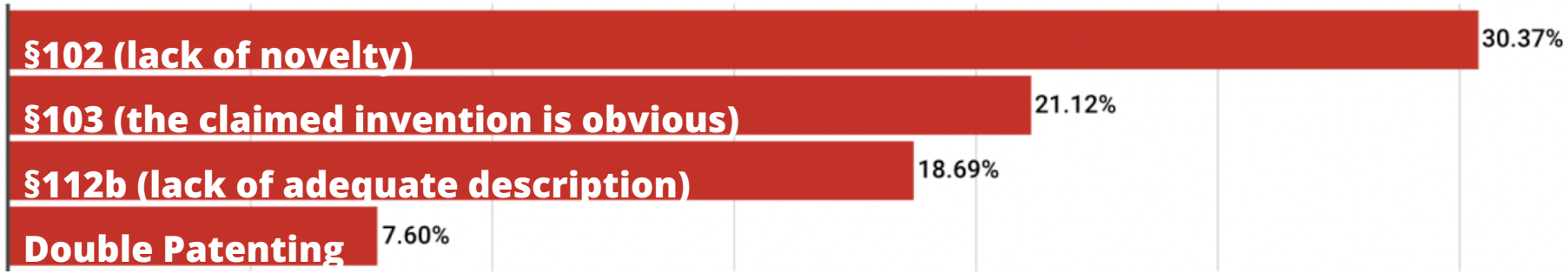
**Kenji Laurens**

Raytheon Technologies

# Thank you.

# Appendix

# Common Rejection Causes




- These 4 reason already account for more than 70% of all rejections
- For example, § 102 rejection means that the patent failed to prove that it is new, novel and unique enough to warrant a patent in itself
- Rejections partially depends on the examiner discretion as well, with different examiners trending towards certain category of rejection more than others
- **This is how we decided to focus on novelty**



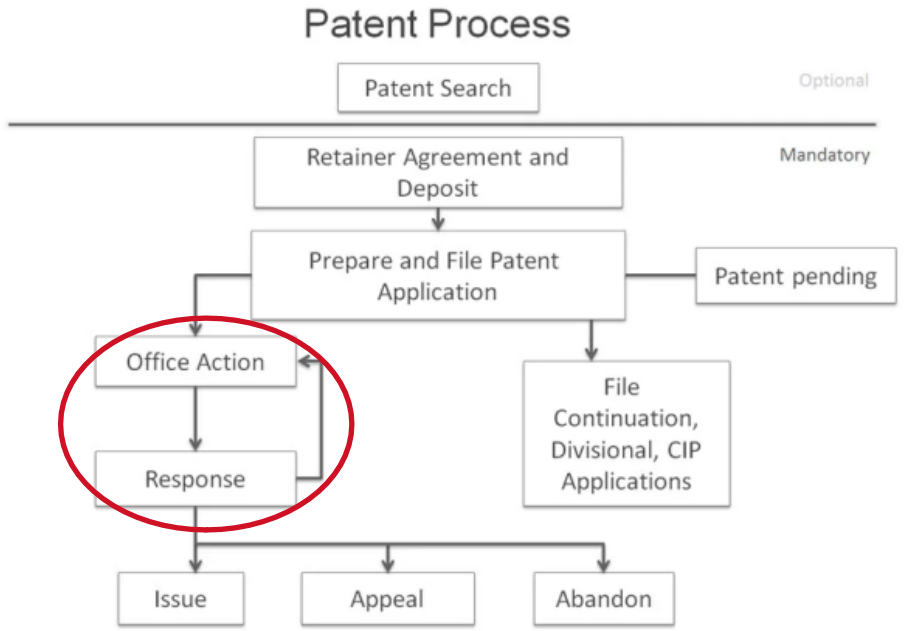
https://blog.juristat.com/most-common-rejections

# United States Patent and Trademark Office (USPTO)

- For a patent to be granted to an organization or individual, they must first apply for it to a patent office such as the USPTO
- This process can take years
- About 86% chance of getting a first rejection; and
- A little over 50% chance of getting a second rejection.
- It can take 15-18 months from the filing date until the first patent Office Action letter is received.
- The entity may need to submit Request for Continued Examination (RCE) in response to the final Office Action if further claim amendments or arguments required.

**Miscellaneous patent fees** Back to top

| Fee code | 37 CFR § | Description | Fee | Small entity fee | Micro entity fee |
|---|---|---|---|---|---|
| 1817/2817/3817 | 1.17(c) | Request for prioritized examination | 4,200.00 | 2,100.00 | 1,050.00 |
| 1819/2819/3819 | 1.17(d) | Correction of inventorship after first action on merits | 640.00 | 320.00 | 160.00 |
| 1801/2801/3801 | 1.17(e)(1) | Request for continued examination (RCE) - 1st request (see 37 CFR 1.114) | 1,360.00 | 680.00 | 340.00 |
| 1820/2820/3820 | 1.17(e)(2) | Request for continued examination (RCE) - 2nd and subsequent request (see 37 CFR 1.114) | 2,000.00 | 1,000.00 | 500.00 |



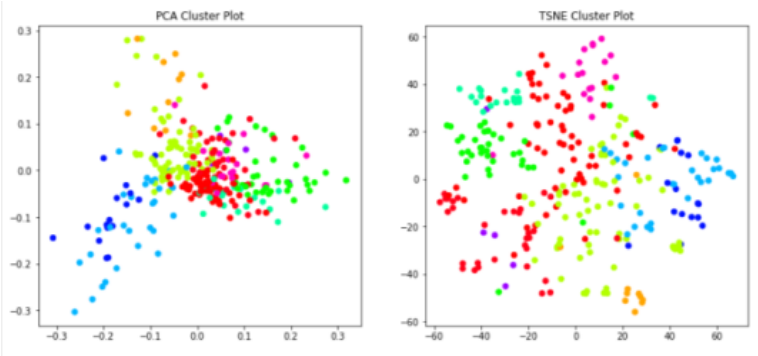

https://www.patenttrademarkblog.com/probability-of-getting-patent/#:~:text=roughly%2086%25%20chance%20of%20getting,of%20getting%20a%20second%20rejection.

# Initial Findings & EDA (Clustering)

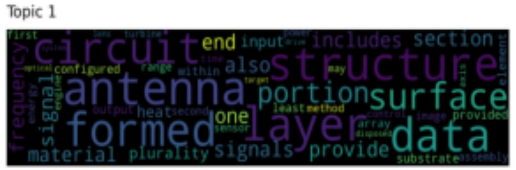
## High Variability of Topics with Domain-Specificity

**K-Means (k=10)** with PCA & TSNE Plot shows topics overlap

```
Cluster 0
control,nodes,sensor,device,based,plurality,includes,network,method,data

Cluster 1
layers,disposed,metal,second,dielectric,semiconductor,material,surface,substrate,layer

Cluster 2
portion,includes,conductive,disposed,having,structure,substrate,second,material,surface

Cluster 3
elements,second,array,output,digital,phase,signals,antenna,frequency,signal

Cluster 4
detection,return,clutter,targets,signal,data,signals,range,radar,target

Cluster 5
second,air,fluid,section,cooling,fan,heat,gas,engine,turbine

Cluster 6
platform,radially,blade,wall,portion,outer,second,edge,airfoil,seal

Cluster 7
wavelength,includes,source,second,energy,medium,optical,light,laser,beam

Cluster 8
signal,coupled,transistor,amplifier,input,current,output,circuit,power,voltage

Cluster 9
scene,configured,second,imaging,lens,detector,mirror,radiation,image,optical
```




**Top2Vec:** 3 high-level topics given



**BERTopic:** F, G, H category results *(left-right)*

| | Topic | Count | % |
|---|---|---|---|
| 0 | -1 | 17948 | 0.365287 |
| 1 | 0 | 711 | 0.014471 |
| 2 | 1 | 677 | 0.013779 |
| 3 | 2 | 644 | 0.013107 |
| 4 | 3 | 531 | 0.010807 |
| ... | ... | ... | ... |
| 573 | 575 | 10 | 0.000204 |
| 572 | 576 | 10 | 0.000204 |
| 570 | 578 | 10 | 0.000204 |
| 569 | 574 | 10 | 0.000204 |
| 580 | 579 | 10 | 0.000204 |

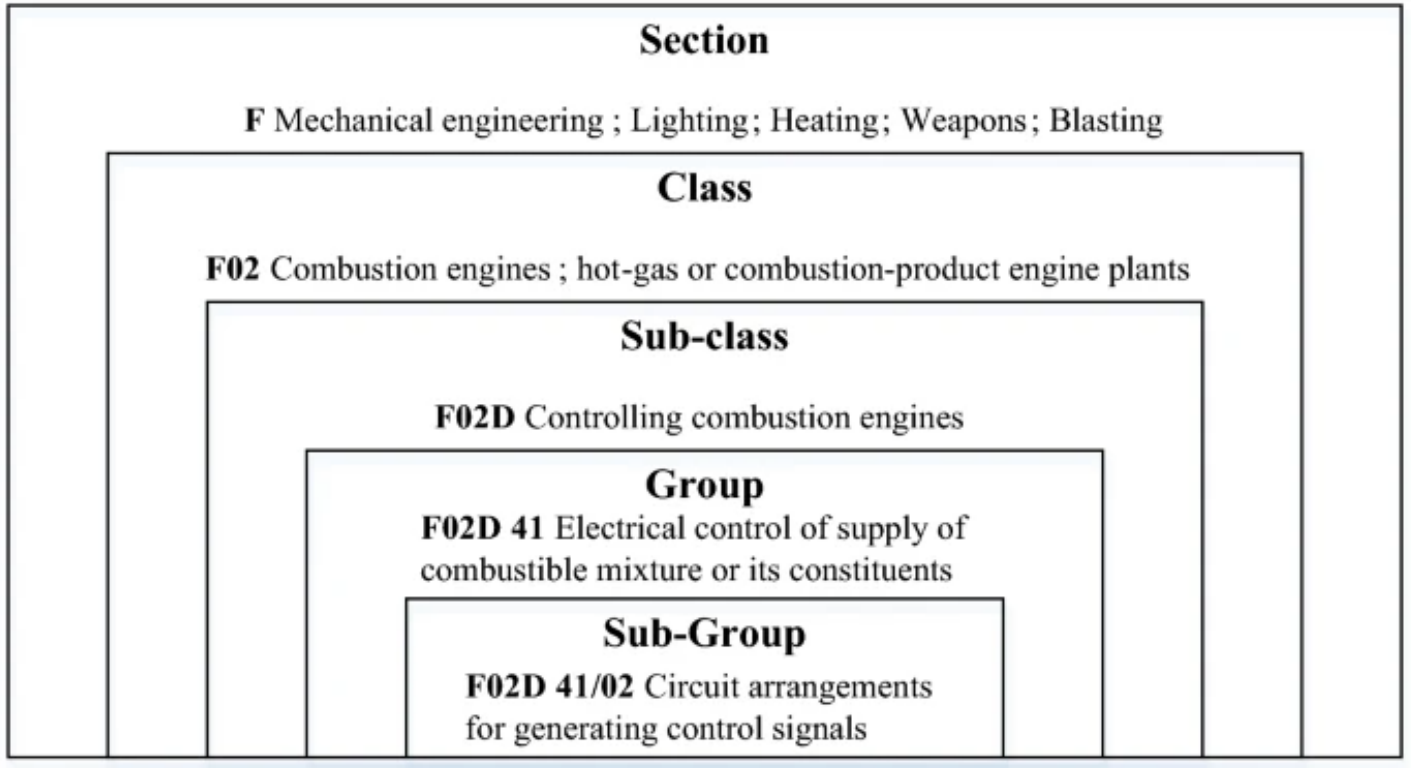
| | Topic | Count | % |
|---|---|---|---|
| 0 | -1 | 17999 | 0.366325 |
| 1 | 0 | 1095 | 0.022286 |
| 2 | 1 | 1037 | 0.021106 |
| 3 | 2 | 886 | 0.018032 |
| 4 | 3 | 744 | 0.015142 |
| ... | ... | ... | ... |
| 509 | 513 | 10 | 0.000204 |
| 507 | 515 | 10 | 0.000204 |
| 506 | 516 | 10 | 0.000204 |
| 505 | 511 | 10 | 0.000204 |
| 518 | 517 | 10 | 0.000204 |

| | Topic | Count | % |
|---|---|---|---|
| 0 | -1 | 20266 | 0.412464 |
| 1 | 0 | 1127 | 0.022937 |
| 2 | 1 | 892 | 0.018154 |
| 3 | 2 | 667 | 0.013575 |
| 4 | 3 | 638 | 0.012985 |
| ... | ... | ... | ... |
| 490 | 491 | 10 | 0.000204 |
| 489 | 492 | 10 | 0.000204 |
| 487 | 494 | 10 | 0.000204 |
| 486 | 490 | 10 | 0.000204 |
| 496 | 495 | 10 | 0.000204 |

# Patent Classification Schema

Example of Patent Classification Schema (IPC)

Fig. 1




An example of IPC

# Decision matrix on deciding novelty detection

## DECISION MATRIX - WEIGHTED

| CRITERIA DESCRIPTION | Does it solve user's painpoint (Lucy&Raytheon)? | Feasibility in terms of time frame | Feasibility in terms of data infrastructure | Opportunity for advancing prior works | |
|---|---|---|---|---|---|
| | **Criteria 1** | **Criteria 2** | **Criteria 3** | **Criteria 4** | **WEIGHTED SCORE** |
| **WEIGHT** | 5 | 3 | 3 | 1 | 12 |
| | 41.7% | 25.0% | 25.0% | 8.3% | 100.0% |
| **OPTIONS** | **Criteria 1 SCORES** | **Criteria 2 SCORES** | **Criteria 3 SCORES** | **Criteria 4 SCORES** | |
| Patent classification | 2 | 4 | 4 | 2 | 3.00 |
| Keyword generation for prior art search | 2 | 5 | 4 | 1 | 3.17 |
| Approval prediction (Novelty is not considered) | 4 | 3 | 3 | 4 | 3.50 |
| **Novelty detection** | 5 | 3 | 4 | 4 | 4.17 |

# Annual Report

## Financials

**$64.4**
2021 net sales
(dollars in billions)

**$4.27**
2021 adjusted earnings per share[2]

**$7.2**
2021 research and development[3]
(dollars in billions)

**$7.1**
2021 cash flow from operations
(dollars in billions)

**$2.005**
2021 dividends paid per common share

**$156**
2021 backlog
(dollars in billions)

### Sales mix

**Sales by type**

**35%** Commercial
**65%** Defense

**Sales by geography**

**62%** United States
**15%** Europe
**12%** Asia Pacific
**7%** Middle East and North Africa
**4%** Canada and All Other

| *(dollars in millions)* | 2021 | 2020 | 2019 |
|---|---|---|---|
| Total Net Sales | $ 64,388 | $ 56,587 | $ 45,349 |
| Operating profit (loss) | 4,958 | (1,889) | 4,914 |
| Operating profit (loss) margins | 7.7 % | (3.3)% | 10.8 % |
| Operating cash flow from continuing operations | $ 7,142 | $ 4,334 | $ 5,821 |

*Total Net Sales.* Total Net Sales by segment were as follows:

| *(dollars in millions)* | 2021 | 2020 | 2019 |
|---|---|---|---|
| Collins Aerospace Systems | $ 18,449 | $ 19,288 | $ 26,028 |
| Pratt & Whitney | 18,150 | 16,799 | 20,902 |
| Raytheon Intelligence & Space | 15,180 | 11,069 | — |
| Raytheon Missiles & Defense | 15,539 | 11,396 | — |
| Total segment | 67,318 | 58,552 | 46,930 |
| Eliminations and other[(1)] | (2,930) | (1,965) | (1,581) |
| Consolidated | $ 64,388 | $ 56,587 | $ 45,349 |

| | | | | % of Total Net Sales | | |
|---|---|---|---|---|---|---|
| *(dollars in millions)* | 2021 | 2020 | 2019 | 2021 | 2020 | 2019 |
| Sales to the U.S. government [(1)] | $ 31,177 | $ 25,962 | $ 9,094 | 48 % | 46 % | 20 % |
| Foreign military sales through the U.S. government | 5,546 | 4,585 | 1,571 | 9 % | 8 % | 3 % |
| Foreign government direct commercial sales | 4,993 | 3,974 | 1,498 | 8 % | 7 % | 3 % |
| Commercial aerospace and other commercial sales | 22,672 | 22,066 | 33,186 | 35 % | 39 % | 73 % |
| Total Net Sales | $ 64,388 | $ 56,587 | $ 45,349 | 100 % | 100 % | 100 % |

(1) Excludes foreign military sales through the U.S. government.

*Source: RTX Annual Report 2021*

Raytheon Technologies

# Rejection/ Approval Status with Reason

```
array(['Docketed New Case - Ready for Examination',
       'Sent to Classification contractor', 'Non Final Action Mailed',
       'Application Undergoing Preexam Processing', 'Patented Case',
       'Response to Non-Final Office Action Entered and Forwarded to Examiner',
       'Final Rejection Mailed',
       'Notice of Allowance Mailed -- Application Received in Office of Publications',
       'Awaiting TC Resp, Issue Fee Payment Verified',
       'Publications -- Issue Fee Payment Received',
       'Application Dispatched from Preexam, Not Yet Docketed',
       'Publications -- Issue Fee Payment Verified',
       'Ex parte Quayle Action Mailed',
       'SE ready for Pubs Processing -- Certificate in IFW',
       'Prosecution Suspended',
       'PCT - International Search Report Mailed to IB',
       'Response after Non-Final Action Entered (or Ready for Examiner Action)',
       "Reexam -- Timely Owner's Statement Received in Response to Order",
       'Awaiting TC Resp., Issue Fee Not Paid', 'Non-Final Action Mailed',
       'Reexam -- Request Ready for Ex Parte Action',
       'Application Returned back to Preexam',
       'Response to Ex parte Quayle Action Entered and Forwarded to Examiner',
       'Expressly Abandoned  --  During Examination', 'Special New',
       'Determination - Reexamination Ordered',
       'Abandoned  --  Incomplete Application (Pre-examination)',
       'Response after Final Action Forwarded to Examiner',
       'AWAITING RESPONSE FOR INFORMALITY, FEE DEFICIENCY OR CRF ACTION',
       'Search report counted', 'Request for Reexamination Denied',
       'Reexam Ordered Based on Supplemental Examination',
       'Petition Received RE: Denial of Reexamination Request',
       'Reexam Assigned to Examiner for Determination',
       'RO PROCESSING COMPLETED-PLACED IN STORAGE',
       'PCT - Dispatch to TC Chapter I case',
       'Ready for Reexam -- Certificate in IFW',
       'Reexam Preprocessing Completed -- Released to Assigned GAU'],
      dtype=object)
```

# Definition of Prior Art

## Prior art defined

- Information known publicly before the effective filing date of a U.S. patent application is called prior art.
- Prior art consists of information available to the public including:
  - U.S. patents and published patent applications
  - Foreign patents and published patent applications
  - Journal and magazine articles
  - Books, manuals, and catalogs
  - Websites
  - Conference proceedings
  - Scientific papers

uspto

## Examples of prior art

- In 2011 the computer company Apple sued Samsung, claiming that the company had copied its design of the iPad for its Galaxy Tab.
- In defense, Samsung attorneys contended that Apple had patented a common design of a square display with rounded corners which is seen often in movies.
- As evidence of prior art, Samsung referred to scenes from "2001: A Space Odyssey," a science fiction film released in 1968.


uspto

# Frontend Application

## Patent Application Claims Text Semantic Similarity Search

Select Database

granted_patents_fgh3.db

Select Index

faiss_ip_norm_index

Select Model

AI-Growth-Lab/PatentSBERTa

Oldest Patent Date

2011/01/01

Latest Patent Date

2022/11/28

Comma Separated CPC Class

Search Clause

AND

CPC Filter Help

Max Return Count

100

1 1000

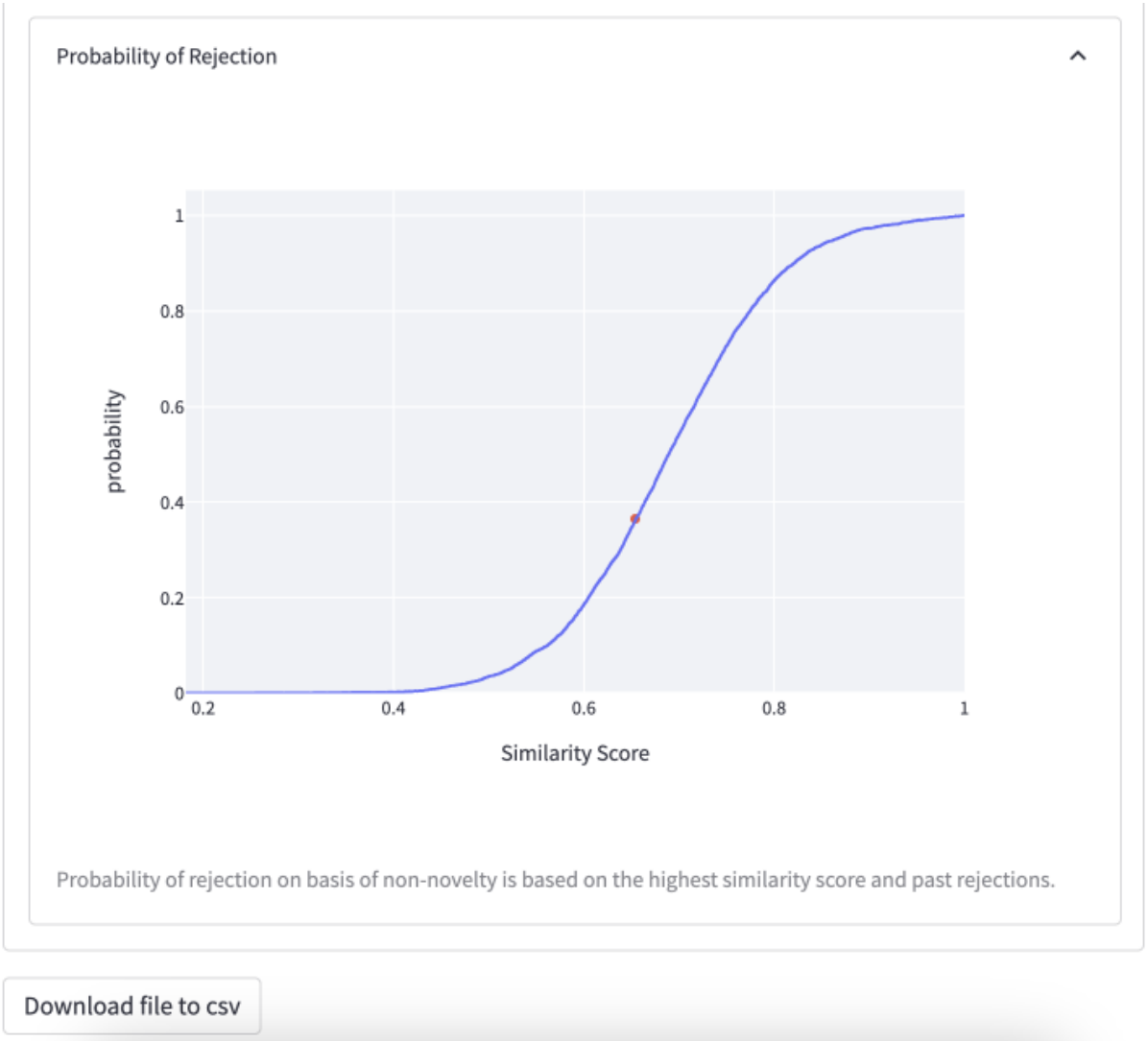
Claim Text

Method to search and retrieve patents that are most similar to an intended application based on claims text which uses natural language processing (NLP), natural language understanding (NLU) and machine learning methodologies, all under 5 seconds on average, comprising: database for patent data storage; a sentence transformer BERT based model to encode text into 768 dimension array; an index to quickly look up and return k closest and most similar patents; a preprepared dataset measuring the likelihood of rejection given the closest score, using the cumulative distribution function (CDF); and a user interface.

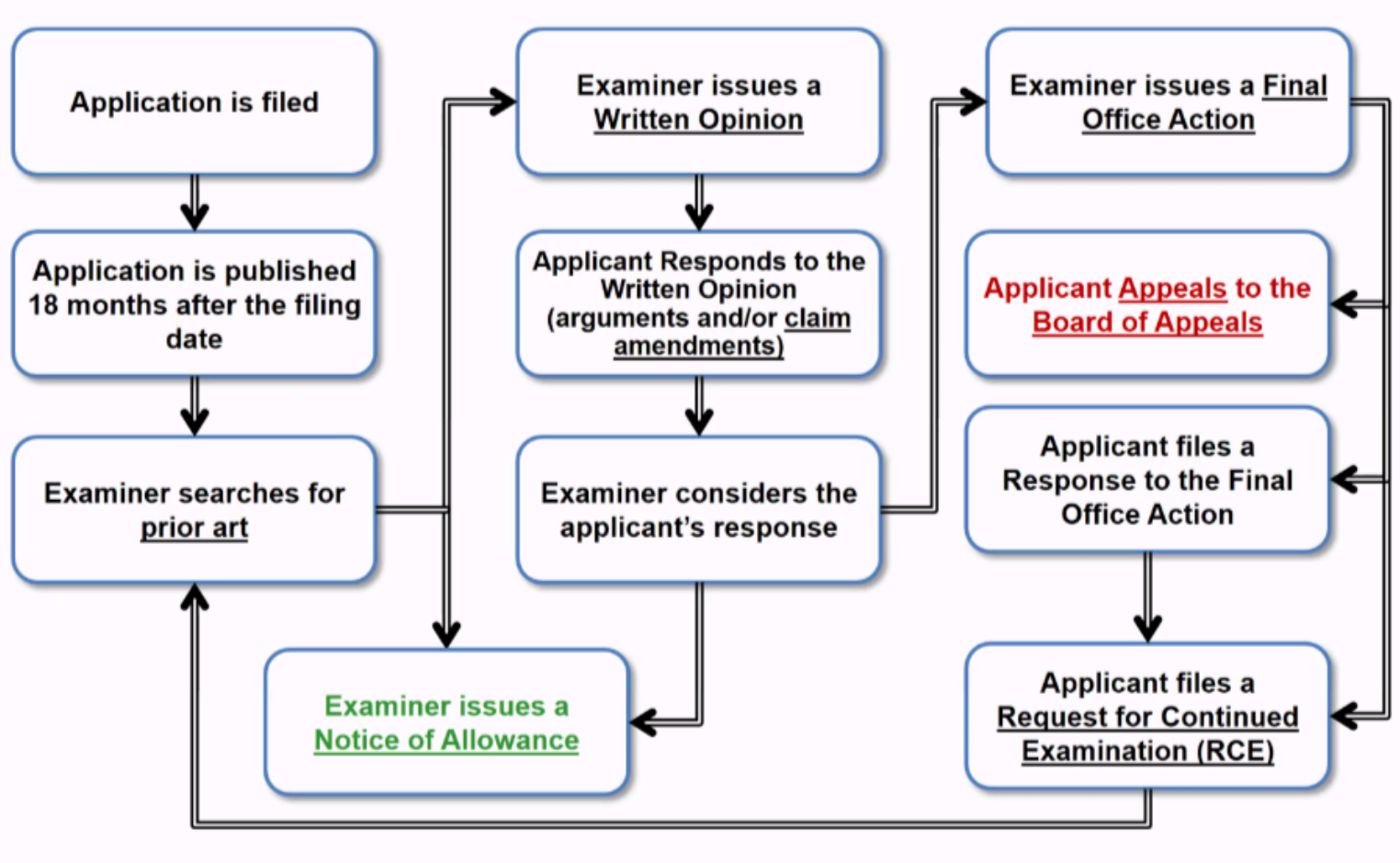
Run Similarity Search

| | ext | Score | Rejecti |
|---|---|---|---|
| 0 | puter-based system for processing a user query related to patent claim terms to generate a set | 0.654 | 0.365 |
| 1 | od for classifying text, the method comprising: determining a set of documents that are each | 0.648 | 0.343 |
| 2 | puter-implemented patent portfolio analysis method comprising: providing user-prescribed c | 0.648 | 0.342 |
| 3 | puter-implemented patent portfolio analysis method comprising: providing user-prescribed c | 0.644 | 0.328 |
| 4 | puter implemented method comprising: maintaining a database of patent portfolios and a da | 0.639 | 0.311 |
| 5 | puter-implemented method of searching for patent-related documents stored in a document | 0.622 | 0.255 |
| 6 | h method based on artificial intelligence, the method comprising: acquiring from a terminal | 0.613 | 0.228 |
| 7 | puter implemented method comprising: a) receiving by a computer comprising a processor ar | 0.610 | 0.219 |
| 8 | puter-implemented patent searching method in connection to a matching degree, wherein th | 0.608 | 0.215 |

# Rejection Probability Distribution



# Patent Process



# Reference


- Aslanyan, G., & Wetherbee, I. (2022). Patents Phrase to Phrase Semantic Matching Dataset.
- Aslanyan, G. (2022, August 31). *Announcing the Patent Phrase Similarity Dataset*. Google Research. https://ai.googleblog.com/2022/08/announcing-patent-phrase-similarity.html
- Bekamiri, H., Hain, D., & Jurowetzki, R. (2021). PatentSBERTa: A Deep NLP based Hybrid Model for Patent Distance and Classification using Augmented SBERT.
- Bekamiri, H., Hain, D., & Jurowetzki, R. (2022). A Survey on Sentence Embedding Models Performance for Patent Analysis.
- Bitlaw. (n.d.). *Patent Requirements*. Retrieved May 30, 2022, from https://www.bitlaw.com/patent/requirements.html#:~:text=Novelty%20(Newness)%20Requirement,the%20invention%20have%20been%20made.
- Bruzzese, T., Lerner A., & O'Rahilly O. (2022). Patentability: Improving Acceptance Prediction of US Patent Applications using Ensemble Modeling. *Stanford CS224N Custom Project*.
- Chikkamath, R., Endres, M., Bayyapu, L., & Hewel, C. (2020). An Empirical Study on Patent Novelty Detection: A Novel Approach Using Machine Learning and Natural Language Processing. *2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 1-7, doi: 10.1109/SNAMS52053.2020.9336557.
- Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers.
- Colombo, P. (2019). 5 Factors That Define Aerospace And Defense Industry. *Ansys*. https://www.ansys.com/blog/5-factors-define-aerospace-and-defense-industry-trends.
- Defense News. (n.d.). *Top 100 for 2021.* Top 100 | Defense News, News about defense programs, business, and technology. Retrieved May 30, 2022, from https://people.defensenews.com/top-100/
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- Duffin, E. (2021). Number of U.S. Patents Registered by Raytheon 2020. *Statista*. https://www.statista.com/statistics/633065/raytheon-patents-usa-registered/#:~:text=Number%20of%20patents%20granted%20to%20Raytheon%20in%20the%20U.S.2010%2D2020&text=In%202020%2C%203%2C213%20patents%20were,were%20granted%20to%20the%20company.

# Reference


- European Patent Office [EPO]. (2022). *Applicants.* https://www.epo.org/about-us/annual-reports-statistics/statistics/2021/statistics/applicants.html
- Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9): 1627-1645.
- Gales, M. & Young, S. (2008). The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends in Signal Processing*. 1(3), 195-304.
- Gerón, A. (2017). *Hands on Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media, Inc.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Gu, J., et. al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377. doi: 10.1016/j.patcog.2017.10.013.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition.* 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/cvpr.2016.90.
- Hellmann T. (2019, December 23). The Most Common Rejections: 102, 103, and 112(b). *Juristat Blog*. https://blog.juristat.com/most-common-rejections
- Helmers, L., Horn, F., Biegler, F., Oppermann, T., & Müller, K. (2019). Automating the search for a patent's prior art with a full text similarity search.
- Howard, A. G., et al. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861v1.*
- Jansson, H., & Navrozidis, J. (2020). Using Natural Language Processing to Identify Similar Patent Documents.
- Justia. (2021). *Timeline for Patent Applications.* https://www.justia.com/intellectual-property/patents/timeline-for-patent-applications/
- Kaggle. (2022). *U.S. Patent Phrase to Phrase Matching.* https://www.kaggle.com/c/us-patent-phrase-to-phrase-matching
- Kearns, R., Khosla S., & Wittenbrink, B. (2012). A NLP Approach to Understanding Patent Acceptance Criteria. *Stanford CS224N Custom Project*.
- Kim, S., Baldwin, T., Kan, M. (2009). An Unsupervised Approach to Domain-Specific Term Extraction.

# Reference


- Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. *World Patent Information*. 65. 102035. 10.1016/j.wpi.2021.102035.
- Lee, J., & Hsiang, J. (2019). Patent Classification by Fine-Tuning BERT Language Model.
- Lin, C. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *Text Summarization Branches Out,* 74–81.
- Markou, M., & Singh, S. (2003). Novelty detection: a review—part 1: statistical approaches. *Signal Processing*, Volume 83, Issue 12, Pages 2481-2497.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.
- Qin, C., Deng, C., Huang, J., Shu, K., & Bai, M. (2020). An Efficient FAISS-Based Search Method for Mass Spectral Library Searching.
- Office Action Research Dataset for Patents. USPTO. https://www.uspto.gov/ip-policy/economic-research/research-datasets/office-action-research-dataset-patents
- PatentsView Dataset. https://patentsview.org/download/data-download-tables
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation.
- Reimers, N., Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.
- Ruder, S. (2021, January 4). *A review of the recent history of Natural Language Processing. Sebastian Ruder.* Retrieved May 30, 2022, from https://ruder.io/a-review-of-the-recent-history-of-nlp
- Shang, J., Gao X., Hou, Z., Ning Y., & Krishnan, V. (2022). Towards Comprehensive Patent Approval Predictions: Beyond Traditional Document Classification.
- Srebrovic, R., & Yonamine, J. (2020). Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery.
- Sun, Y., Braverman, W., & Hoxhalli, G. (2020, April 2). The Raytheon Company. *Krause Fund Research*.
- Terekhov, V., & Kanev, A. (2021). Semantic Search System with Metagraph Knowledge Base and Natural Language Processing.
- United States Patent and Trademark Office (USPTO). (2022). *Classification Resources*. https://www.uspto.gov/web/patents/classification/cpc/html/cpc-H01L.html

# Reference


- United States Patent and Trademark Office (USPTO). (2021). *U.S. Patent Statistics Chart Calendar Years 1963 - 2020.* https://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm
- Upcounsel. (n.d.). *Request for Continued Examination: Everything You Need to Know.* Retrieved May 30, 2022, from https://www.upcounsel.com/request-for-continued-examination
- Vuleta, B. (2022, April 29). *25 Patent Litigation Statistics - High-Profile Feuds about Intellectual Property.* Legal Jobs. https://legaljobs.io/blog/patent-litigation-statistics/
- Yang, J. (2010). *Claims are most important part of patent.* https://ocpatentlawyer.com/claims-most-important-part-of-the-patent/
- Younge, K., & Kuhn, J. (2016). Patent-to-Patent Similarity: A Vector Space Model. *SSRN.*