Emerging Technologies in USPTO Business Solutions
May 25th 2018
STATS @ USPTO

• ~8500 Patent Examiners;
• Receive ~600k patent applications a year; have ~605,000 applications in inventory.
• ~1.2M patent applications in flight.
• ~2.5M granted patent that are enforceable.
• Issue ~305,000 patents a year with ~5 million “claims” that can be litigated.

• *Time to issue* first millionth Patent – ~75.8 years;
• *Time to issue* our last millionth patent – ~3.6 years.

• 875 Trademark employees (7% of USPTO)
• 579 examining attorneys (66% of Trademarks)
• Trademark Operations receive ~594,100 classes for registration in FY 2017.
Uses within the USPTO

I. Patent Enriched Citation Data
II. Patent CPC Analytics - Trends
III. Browser Based End Point Advanced Analytics for Patents
IV. Trademark Image Classifier
V. Research Projects at the USPTO
I. Enriched Citation & DocDB

• Extracting the associated citations and related rejection type using emerging technologies.
• This allows us to provide this information to the IP5 Office in an internationally consumable format.
• DocDB (granular international family mapping with IP Office specific citations) advanced analytical analysis.
Developer Hub for Patent Filing Data

https://developer.uspto.gov/
II. CPC Classifying Trends/Analytics

- **Data Sources**
  - CPC Scheme Definition
    - Concatenated the definition to the root using XML parser
  - Patents/PGPubs Specifications
    - Summary text, a mix of “Background of invention”, “Brief description of the invention”, etc.
  - Patents/PGPubs Claims
    - Independent and dependent claims of patents

- **Machine Learning/AI Algorithms**
  - Supervised/Unsupervised Machine Learning Algorithms
  - Ensemble Algorithms
Prototype Interface

- Classifies entered text (i.e., patent application)
- "Explains" why/how algorithm arrived at its conclusion
- Enables human validation and model training
III. Browser based end point Advanced Analytics for Patents
Platform AI/BD Capabilities

AGGREGATED "ONE-CLICK" PATENT, PRE-GRANT, FOREIGN AND NLP THROUGH ENSEMBLE AI RANKED RELEVANCY SEARCH RESULTS
IV. Trademark Machine Learning Framework

1. Get Data
   - Clean, Prepare & Manipulate Data

2. Train Model

3. Test Data

4. Improve

5.
New Application Filings

- FY16: 530,270 classes filed  
- FY17: 594,107 classes filed  
- FY18 up 8.7% compared to FY17. Expecting 646,000 classes this year
Trademark Design Code Suggestions

A six-digit numerical design search code is assigned to each design element of a trademark, such as a depiction of a star (01.01.03) or flower (05.05.25). Using years of images with corresponding examiner-annotated design codes, we are able to train deep learning systems that can predict design codes of a new trademark image.

Types of Training Tried:
1. Retraining just last layer of the neural network (Transfer Learning)
2. Retraining all the layers from scratch
# Design Code Based on Image

**Trademark Ensemble Classification Experiments**

**Predicted Results**

<table>
<thead>
<tr>
<th>Design Code</th>
<th>Design Code Description</th>
<th>INTEGRITY</th>
<th>CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>04 03 22</td>
<td>Messy Mosaic, Thrice</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>01 01 13</td>
<td>Stars: wide petals with fine points</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>05 01 21</td>
<td>Circles that are tilted or partially shaded.</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>05 01 10</td>
<td>Stars: three or more, three or more petals</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>24 01 10</td>
<td>Circles: six or more concentric, Circles on circles, three or more, three or more concentric circles</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>24 11 82</td>
<td>Crosses open at the top</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>03 01 24</td>
<td>We are stylized, including a neural net closed in a circular form</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>03 01 24</td>
<td>Stars of two or three formed, inside a circle or facing forward.</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>03 01 24</td>
<td>Stars of two or three formed, inside a circle or facing forward.</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>03 01 24</td>
<td>Stars of two or three formed, inside a circle or facing forward.</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

**Predictions by Image Model**

<table>
<thead>
<tr>
<th>Design Code</th>
<th>Design Code Description</th>
<th>INTEGRITY</th>
<th>CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 01 24</td>
<td>Crosses with bars, bands and lines</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

**Predictions by Text Model**

<table>
<thead>
<tr>
<th>Design Code</th>
<th>Design Code Description</th>
<th>INTEGRITY</th>
<th>CONFIDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>02 01 33</td>
<td>Processed, Square. White seamless corners or cream-colored of rectangle, shadow is more shaded</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>24 01 80</td>
<td>Circles facing left, each enclosed in a smaller circle</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>24 01 24</td>
<td>Crosses closed at the to</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>24 01 24</td>
<td>Crosses closed at the to</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

**Design Codes On Record**

- 04 03 22: Messy Mosaic, Thrice
- 01 01 13: Stars: wide petals with fine points
- 05 01 21: Circles that are tilted or partially shaded.
- 05 01 10: Stars: three or more, three or more petals
- 24 01 10: Circles: six or more concentric, Circles inside circles, three or more, three or more concentric circles
- 24 11 82: Crosses open at the top
- 03 01 24: We are stylized, including a neural net closed in a circular form
- 03 01 24: Stars of two or three formed, inside a circle or facing forward.
- 03 01 24: Stars of two or three formed, inside a circle or facing forward.
- 03 01 24: Stars of two or three formed, inside a circle or facing forward.
- 03 01 24: Stars of two or three formed, inside a circle or facing forward.

**Confidence scores**

The confidence scores for the predicted design codes range from A to E, indicating the level of confidence in the model's predictions.

**Design Code on Record**

- 04 03 22: Messy Mosaic, Thrice
- 01 01 13: Stars: wide petals with fine points
- 05 01 21: Circles that are tilted or partially shaded.
- 05 01 10: Stars: three or more, three or more petals
- 24 01 10: Circles: six or more concentric, Circles inside circles, three or more, three or more concentric circles
- 24 11 82: Crosses open at the top
V. Research Projects at the USPTO

• AI Assisted Patent Searching
• AI based Patent Term Library Generator
• AI based Trademark Image Search
• Deep Machine Learning Chat Bots
AI Patent Search Capabilities

AI-BASED PATENT PROCESSING

INGESTION PIPELINE

Abstract
Patent Sections
Annotator

Noun Phrases,
Acronyms
(Stanford core-
NLP parser)

Related Terms
Annotator

Porter Stemmer

CH3CH2OH
ethanol
InChI=1S/C2H6Oc1-2-3

OSCAR4
chemical
Annotator

US Patent & Pre-
Grant publications
as xml

SEARCHING

User Preferences
- Technical field
- Patent Section
Weights etc.

Input Patent
Search patent based
on preferences

SEARCH SYSTEM

Solr/Lucence
- fast, scalable, widely-used
- “more-like-this” option allows
searching based on an entire
document instead of key words
- supports multiple retrieval
calculations such as TF/IDF or
BM25

retrieves

Related Patents
Patent Synonyms Generator Tool

Corpus
Patent Applications,
Technical Journals,
Technical Standards,...

Training

Build Word Embedding Models

Generating Synonyms By Technical Fields
Searching for marks that look similar is an essential job function of a trademark examiner. Using neural networks we are able to retrieve and store features of mark images that we can then compare to other marks’ features.
TM Image Search Prototype using Machine Learning
V. Leveraging Deep Learning for Chatbot Service

Chatbot Conversation Framework

- Impossible
- General AI [Hardest]
- Rules-Based [Easiest]
- Smart Machine [Hard]

Conversations
- Open Domain
- Closed Domain

Responses
- Retrieval-Based
- Generative-Based
Prototype Chatbot Service

**User:** what's 101

**Wolcott:** It is type of rejection based on whether or not the claimed subject matter is subject matter eligible. 35 U.S.C. §101. "Whoever invents or discovers any new and useful process, machine, manufacture, composition of matter, or any new and useful improvement thereof, may obtain a patent therefor..."

**User:** how about 102

**Wolcott:** It is type of rejection based on whether or not the claimed subject matter is novel.

**User:** finally, what's 103

**Wolcott:** It is type of rejection based on whether or not the claimed subject matter is obvious. A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made. A person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

**User:** rules to go final

**Wolcott:** Section 2271 of the MPEP - Final Action: Before a final action is in order, a clear issue should be developed between the examiner and the patent owner. To bring the prosecution to a speedy conclusion and at the same time deal fairly with the patent owner and the public, the examiner will twice provide the patent owner with such information and references as may be useful in defining the position of the Office as to unpatentability before the action is made final. In making the final rejection, all outstanding grounds of rejection of record should be carefully reviewed and any grounds or rejection relied on should be reiterated. The grounds of rejection must (in the final rejection) be clearly developed to such an extent that the patent owner may readily judge the advisability of an appeal.