

# Machine Learning applied to Trademarks Classification and Search

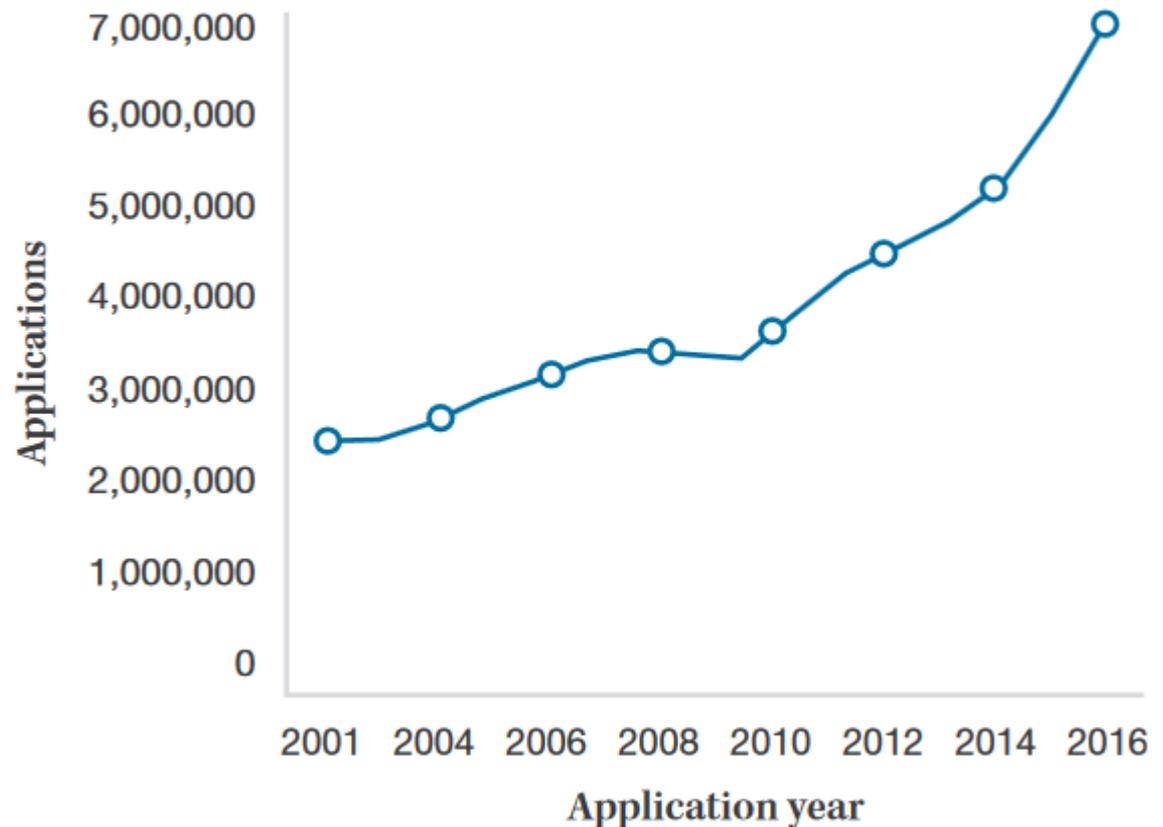


**Speaker:** Christophe Mazenc, Director, Global Databases Division, Global Infrastructure Sector

# Problem to solve

- Trademark examination requires to search similar trademarks having figurative elements in common.

## Trademark Applications Worldwide



(source: World Intellectual Property Indicators 2017)

# Solutions put in place

- First solution found by offices in the XXth century: classification of figurative elements
- Challenges:
  - requires consistent training
  - induces additional work for TM examiners
  - not always effective
  - Not used by all offices

# Solutions put in place

- Image Similarity Search
- Algorithms developed by academic research starting in the 1990's and blooming in the 2000's (CBIR)
- Goal: compute how similar two given images are by computing global descriptors (shape, texture, color)
- Open source software available in 2008 (SOLR, LIRE)
- Led to the current implementation in the Global Brand Database put in production in 2014

# Very effective on simple geometric shapes (Global Brand Database 2014)

Search For



Find (in top results – without Vienna Class)



# How does image similarity search compares with classification search?

← back

(Information valid as of 2014-09-09)

## International Trademark



◀ 65 / 158 ▶

990596 - Arla

**(151) Date of the registration**

08.09.2008

**(180) Expected expiration date of the registration/renewal**

08.09.2018

**(270) Language(s) of the application**

English

**(732) Name and address of the holder of the registration**

Arla Foods amba  
Sønderhøj 14  
DK-8260 Viby J (DK)

**(813) Contracting State or Contracting Organization in the territory of which the holder has his domicile**

DK

**(740) Name and address of the representative**

Zacco Denmark A/S  
Hans Bekkevolds Allé 7  
DK-2900 Hellerup (DK)

**(540) Mark**



**(531) International Classification of the Figurative Elements of Marks (Vienna Classification)- VCL (6)**

05.05.20; 26.01.18; 29.01.13

**(591) Information on colors claimed**

Dark green; Yellow

# Using Vienna Class – 05.05.20 (stylized flowers) and 26.01.18 (circles or ellipses containing one or more letters)

**SEARCH BY**

Brand | Names | Numbers | Dates | Class | Country

Text =

Image Class =

Goods (All) ▾ =

CURRENT SEARCH

**FILTER BY**

Source | Image | Status | Origin | App. Date \* | Expiration \*

AE TM	0	AU TM	0	BN TM	0
CA TM	159	CH TM	0	DE TM	128
DK TM	0	DZ TM	17	EE TM	13
EG TM	2	EM TM	17	ID TM	0
IL TM	0	LA TM	2	JP TM	613
KH TM	48	KR TM	181	MA TM	0
MD TM	7	MX TM	159	NZ TM	45
OM TM	0	PG TM	0	PH TM	49
SG TM	0	TO TM	0	US TM	0

Display:  Sort:

1 - 30 / 1,484

TMview



Display:  per page

1

/ 50

Sort by



1 - 30 / 1,484



Display:  per page

1

/ 50

Select a search strategy and, optionally, what type of image to look for and all images are sorted by similarity to your source image

Goods (All) = e.g. footwear, comput\*

search ↗

FILTER BY

Source Image Status Origin App. Date \* Expiration \*

Pick an image



delete

Pick a strategy

- Shape
- Color
- Texture
- Composite

Pick an image type

Verbal	0
Nonverbal	1,522,717
Combined	6,865,315
Unknown	0

filter ▼

CURRENT FILTER

IMAGE: Shape \* ITY: (Nonverbal Combined) \*

1 - 60 / 8,388,032

TMview ↗



Display: 60 per page

options #



1

/ 139,801



Sort by Score - desc ▼



But the current image search technology in the Global Brand Database does not work so well on complex shapes or logos with both figurative elements and text...

# The recent rise of Machine Learning

- There has been many scientific publications in the last five years, as well as competitions (COCO, ImageNet,...) on the subject of convolutional neural networks applied to classification and object detection in pictures.
- Can these technologies be applied successfully to the Trademark domain?

=> We believe the answer is yes.

# Our strategy for applying AI

- Find a use case where the potential of the technology can be demonstrated with a limited development effort
- Build a first prototype system

Exploratory phase



- If rated successful, build iteratively a production ready system
- Deploy iteratively in the Global Brand Database

Development and production phase

# The easy case for TM classification

- Training corpus: the collection of trademarks of the United States of America (only figurative)
- Target classification: the US design codes

## Reasons:

- 1) The largest trademark collection available in the Global Brand Database
- 2) Information whether a logo is text only, figurative only or combined reliable
- 3) First listed code being the most important (ordering of codes)
- 4) The US classification is relatively close to the Vienna classification
- 5) Good consistency of classification

# US Trademark Design Search Codes

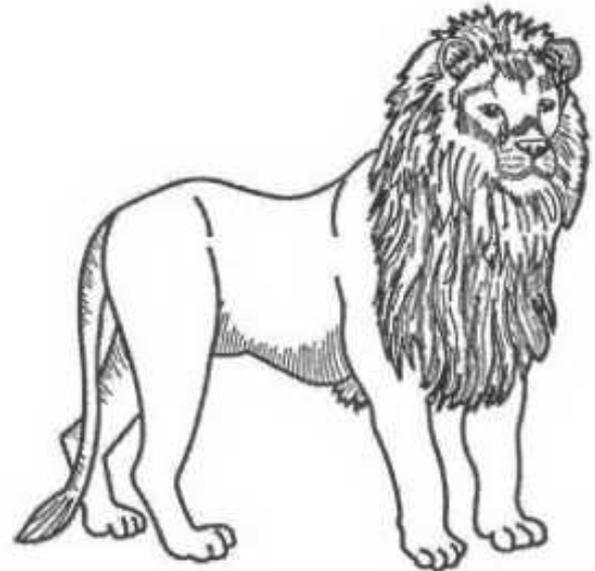
- Three levels classification with a total of more than 1300 different categories

## 03 ANIMALS

Excluding: Mythological or legendary animals (04.05) are not coded in category 03.

### 03.01 Cats, dogs, wolves, foxes, bears, lions, tigers

#### 03.01.01 Lions



## Some statistics

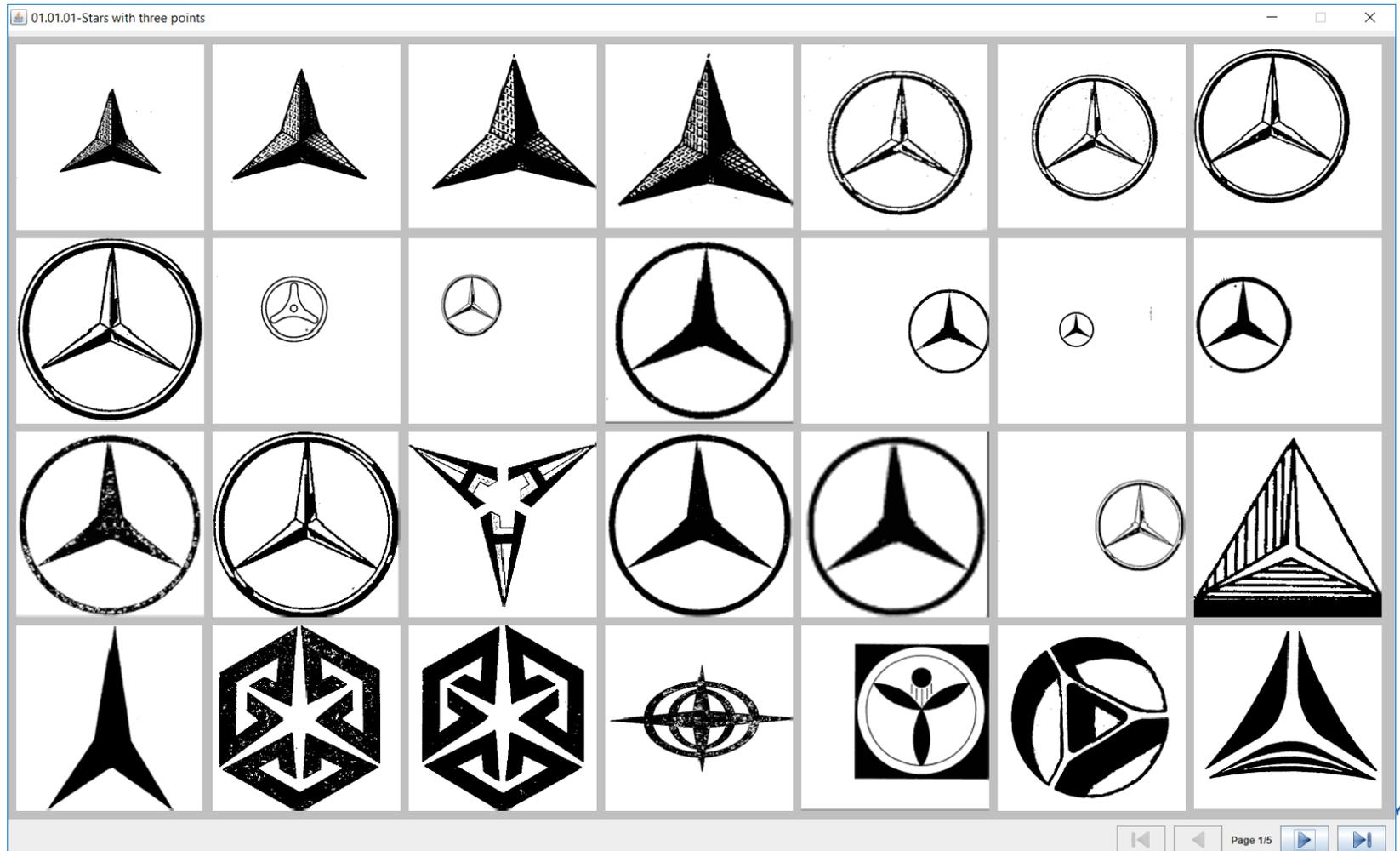
- 194'450 trademark images only figurative
- Having been classified with 508'183 codes
- The corpus is split as follows:
  - 98% for the training
  - 2% for the dev/tests
- The training of one model takes around 3 days on a dedicated server with GPU

To obtain good results with machine learning, the training corpus should be of the best possible quality.

=> We have the following corpus quality challenges...

# Corpus quality challenge 1

■ We have many near-duplicate images:



# Corpus quality challenge 1

We developed an automated tool based on highly advanced computer vision algorithms to identify and remove near-duplicate images in each classification.

=> 52'935 near-duplicate images in the same classification were removed (10% of the corpus)

# Corpus quality challenge 2

- We have a lot of variation in the number of training examples by class:

11.05.07: can openers (electric): 0 !!!

11.05.01: knives (electric): 1

02.01.27: policemen, firemen: 10

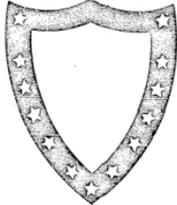
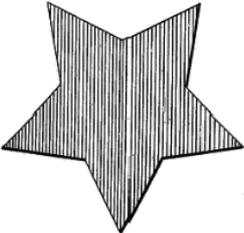
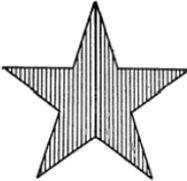
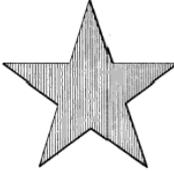
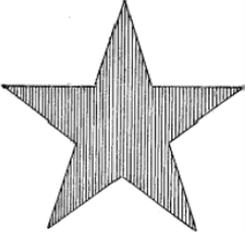
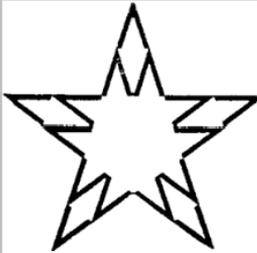
...

02.01.02: shadows or silhouettes of men: 6652

02.01.33: grotesque men formed by letters, ... : 9145 !!!

# Corpus quality challenge 3

01.01.03-A single star with five points

Page 1/109

# Obtained best preliminary results

- Accuracy on training set: 40% (correct guess of the first design code)
- Accuracy on test set: 20%

As images have several design codes and there are more than 1300 categories to choose from, this is a good result!

Even better, looking at examples, the neural network is often able to find several relevant design codes for an image

# Example 1



**03.01.16**-Heads of cats, dogs, wolves, foxes, bears, lions and tigers **Score: 11.61**

**03.01.24**-Stylized cats, dogs, wolves, foxes, bears, lions or tigers **Score: 10.87**

**03.01.09**-Wolves, coyotes **Score: 10.61**

**03.01.03**-Tigers and other large cats (such as leopards or jaguars) **Score: 10.2**

**03.01.11**-Foxes **Score: 9.83**

**03.07.24**-Stylized bovines, deer, antelopes, goats, sheep, pigs, cows, bulls, buffalo and moose **Score: 9.72**

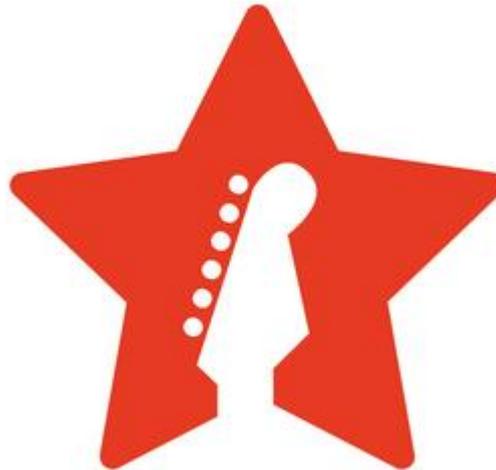
**03.07.11**-Heads of pigs, boars, goats, sheep and rams **Score: 9.2**

**03.09.06**-Rats, mice, moles, gerbils, guinea pigs and the like **Score: 9.16**

**03.01.14**-Other bears **Score: 9.02**

**03.01.04**-Domestic cats **Score: 8.95**

## Example 2



**01.01.03**-A single star with five points Score: 11.03

**01.01.11**-Incomplete stars Score: 10.15

**02.11.14**-Hands and fingers forming the following: handshake, finger pointing, fingers walking, OK sign, and thumbs up or thumbs down. Score: 8.96

**01.01.04**-A single star with six points Score: 8.82

**02.01.02**-Shadows or silhouettes of men Score: 8.12

**02.11.07**-Hands, fingers and arms Score: 8.12

**22.01.06**-Guitars, banjos, ukuleles Score: 7.81

**01.17.11**-States of the United States Score: 6.51

**03.01.07**-Shadows or silhouettes of dogs Score: 6.49

**02.01.37**-Heads, portraits or busts of men in profile Score: 6.4

## Example 3



**03.05.24**-Stylized horses, donkeys and zebras **Score: 16.4**

**03.05.01**-Horses **Score: 15.85**

**03.05.03**-Zebras **Score: 12.78**

**03.05.16**-Heads of horses, donkeys and zebras **Score: 11.87**

**21.03.14**-Merry-go-rounds, carousels, roller coasters, Ferris wheels, and other amusement park rides **Score: 10.27**

**03.05.26**-Costumed animals and those with human attributes in this division (horses, donkeys, zebras) **Score: 9.41**

**03.07.24**-Stylized bovines, deer, antelopes, goats, sheep, pigs, cows, bulls, buffalo and moose **Score: 9.41**

**03.07.10**-Goats, sheep, rams **Score: 8.74**

**04.05.04**-Unicorns **Score: 8.33**

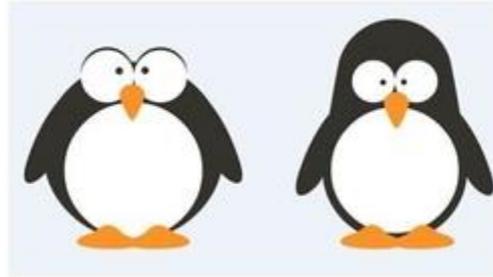
**03.13.02**-Skeletons, skulls, and bones of mammals **Score: 8.21**

## Example 4



- 26.15.03**-Incomplete polygons and polygons made of broken or dotted lines **Score: 14.22**
- 26.15.28**-Miscellaneous designs with overall polygon shape **Score: 14.03**
- 26.15.02**-Plain single or multiple line polygons **Score: 13.84**
- 26.15.20**-Polygons inside another polygon **Score: 12.42**
- 26.15.07**-Polygons with a decorative border, including scalloped, ruffled and zig-zag edges **Score: 10.54**
- 26.15.13**-More than one polygon **Score: 10.47**
- 26.15.01**-Polygons as carriers or as single or multiple line borders **Score: 10.46**
- 26.15.21**-Polygons that are completely or partially shaded **Score: 9.99**
- 18.15.01**-Stop signs **Score: 9.91**
- 26.15.09**-Polygons made of geometric figures, objects, humans, plants or animals **Score: 9.36**

## Example 5



**02.07.26**-Grotesque groups of men, women and/or children having human features  
Score: 10.99

**03.15.15**-Penguins, puffins **Score: 10.91**

**02.01.33**-Grotesque men formed by letters, numbers, punctuation or geometric shapes  
Score: 10.22

**04.07.02**-Objects or combinations of objects representing a person  
Score: 9.55

**02.07.01**-Groups of males **Score: 9.24**

**03.15.26**-Costumed birds and bats and those with human attributes **Score: 9.1**

**03.15.24**-Stylized birds and bats **Score: 8.93**

**21.03.13**-Bowling pins  
Score: 8.92

**02.01.34**-Other grotesque men including men formed by plants or objects  
Score: 8.92

**08.13.02**-Eggs, in shell  
Score: 8.9

# And a bad example



**04.05.01**-Dragons and griffons (half eagle, half lion) Score: 11.45

**03.03.01**-Elephants, mammoths Score: 9.92

**03.21.02**-Snakes Score: 8.46

**03.03.24**-Stylized elephants, hippopotami, rhinoceri, giraffes, alpacas, camels and llamas. Score: 8.28

**04.05.25**-Other mythological or legendary animals Score: 7.94

**03.05.01**-Horses Score: 7.76

**03.15.24**-Stylized birds and bats Score: 7.21

**03.05.24**-Stylized horses, donkeys and zebras Score: 7.13

**01.15.03**-Flames Score: 6.92

**18.03.01**-Bicycles, tricycles, unicycles Score: 6.92

# Assessment

- According to our knowledge, those results are best class.
- Next challenges:
  - More complete clean-up of the training set
  - Deal with combined images by removing text in the image automatically (inprinting) or by finding the best Region Of Interest of the input image and by cropping to this region before classifying
  - Build a system for the Vienna classification from several smaller national collections using Vienna (data de-duplicating and cleaning)

## Next application: a new “semantic” image similarity algorithm for trademarks

- Idea: the NN classifier outputs for an input image a vector of 1300 dimensions (one per design code)
- A measure of similarity of two images can be obtained by computing the distance between the two classification vectors in the 1300 dimensions space
- A similarity search of a new image is performed by computing its classification vector on the fly and by comparing it brute force to the classification vector of each trademark in the collection
- The closest trademarks under a threshold in term of distance are showed to the user and sorted by distance ascending

# Assessment

- A prototype with the figurative-only trademarks of the US collection shows that this works surprisingly well (size of the searched collection: 195'000 images)

# Example 1 Shell

## Current Shape Similarity algorithm in GBD

~

~

~

~

~

~

search 🔍

1 Pick an image



delete 🗑️

2 Pick a strategy

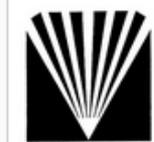
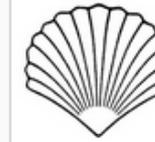
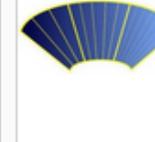
- Shape
- Color
- Texture
- Composite

CURRENT FILTER

SOURCE:USTM ✕ ITY:Nonverbal ✕ IMAGE:Shape ✕

Display: 60 per page options ⚙️

Sort by Relevance - desc ▾

# Example 1: SHELL

## Machine Learning Prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**03.19.18-Shells, including sand dollars, nautilus, conch shells and scallop shells** Score: 19.41

**10.03.01-Fans (non-motorized, hand-held)** Score: 14.77

**20.05.05-Open books** Score: 11.73

**18.09.06-Parachutes, parasails** Score: 10.91

**09.05.25-Other headwear, including military helmets** Score: 10.19

**01.15.17-Thought or speech clouds either empty or with wording and/or punctuation** Score: 9.82

**26.01.06-Semi-circles** Score: 8.77



US78499816.png  
Similarity: 0.929



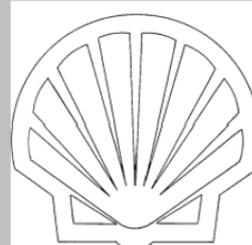
US75522735.png  
Similarity: 0.887



US71076833.png  
Similarity: 0.832



US78087389.png  
Similarity: 0.816



US74039823.png  
Similarity: 0.801



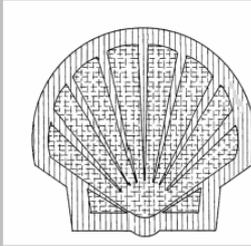
US75746614.png  
Similarity: 0.795



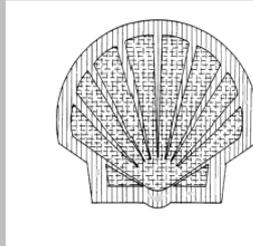
US78291679.png  
Similarity: 0.789



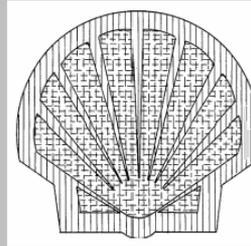
US78231799.png  
Similarity: 0.779



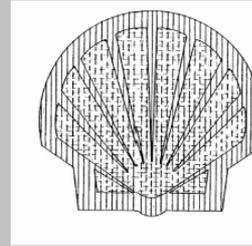
US74098306.png  
Similarity: 0.733



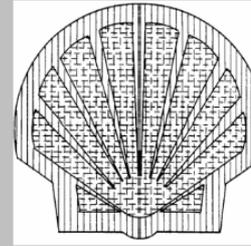
US74040148.png  
Similarity: 0.731



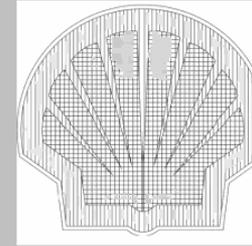
US74050334.png  
Similarity: 0.727



US74226973.png  
Similarity: 0.724



US74040138.png  
Similarity: 0.707



US78205265.png  
Similarity: 0.703

# Example 2: AT&T

## Current Shape Similarity algorithm in GBD

The screenshot displays the GBD interface for a search. On the left, there is a search bar with a 'search' button. On the right, a 'Shape' filter menu is open, showing options: Shape (selected), Color, Texture, and Composite. Below this, the 'CURRENT FILTER' section shows three active filters: 'SOURCE:USTM', 'ITY:Nonverbal', and 'IMAGE:Shape'. The main content area shows a grid of 30 search results, sorted by 'Relevance - desc'. The results include various shapes similar to the AT&T logo, such as different colors (blue, white, black), orientations (horizontal, vertical), and abstract interpretations (globe, sphere, cylinder, etc.).

# Example 2: AT&T Machine Learning prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**01.07.08**-Globes with bars, bands, or wavy lines, excluding meridian or parallel lines Score: 19.49  
**01.07.25**-Other globes Score: 18.09  
**26.19.01**-Spheres Score: 17.62  
**26.01.26**-Spirals, coils and swirls Score: 15.46  
**26.01.12**-Circles with bars, bands and lines Score: 14.77  
**01.07.07**-Globes with rings or orbits Score: 13.51  
**26.01.21**-Circles that are totally or partially shaded. Score: 12.93



US86634192.png  
Similarity: 0.956



US85300320.png  
Similarity: 0.922



US86725313.png  
Similarity: 0.914



US86725288.png  
Similarity: 0.910



US87086008.png  
Similarity: 0.858



US87173751.png  
Similarity: 0.818



US78115221.png  
Similarity: 0.818



US77164059.png  
Similarity: 0.810



US78201664.png  
Similarity: 0.804



US87173762.png  
Similarity: 0.801



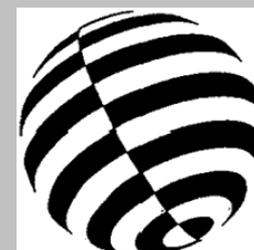
US86411653.png  
Similarity: 0.798



US86131566.png  
Similarity: 0.786



US77587337.png  
Similarity: 0.786



US76678366.png  
Similarity: 0.764

# Example 3: LG

## Current Shape Similarity algorithm in GBD

pi~

search 🔍

**1 Pick an image**



delete 🗑️

**2 Pick a strategy**

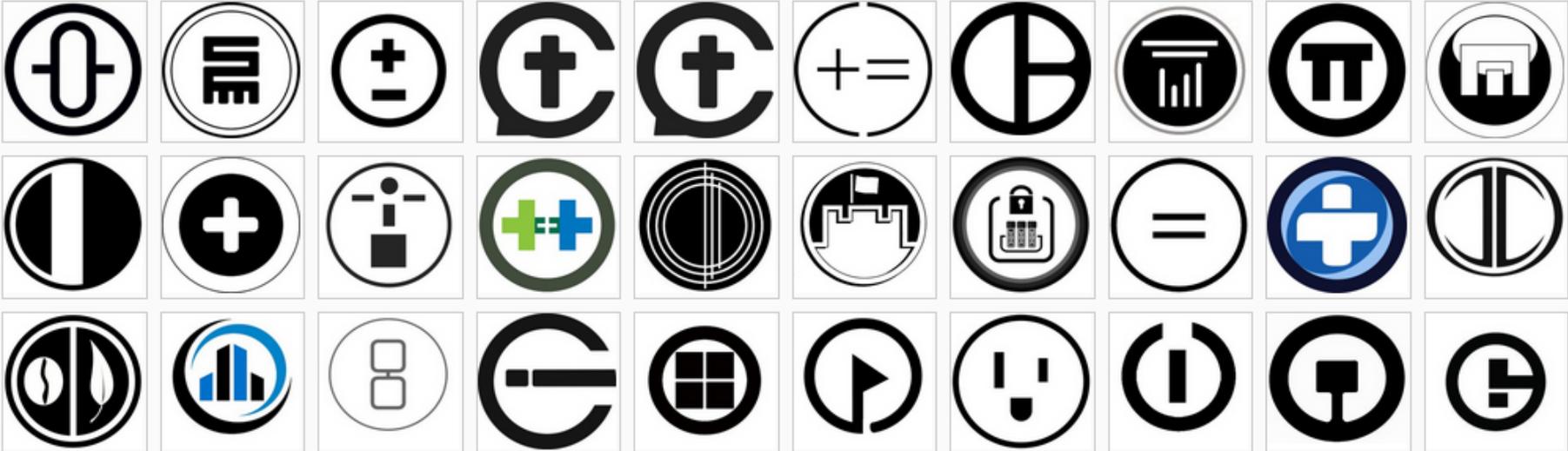
- Shape
- Color
- Texture
- Composite

CURRENT FILTER

SOURCE:USTM ✕ ITY:Nonverbal ✕ IMAGE:Shape ✕

Display: 60 per page options ⚙️

Sort by Relevance - desc ▾



# Example 3: LG

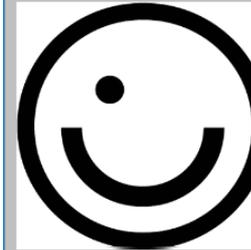
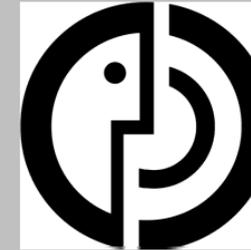
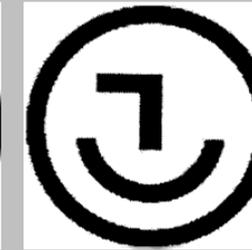
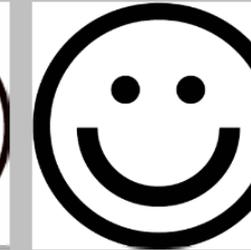
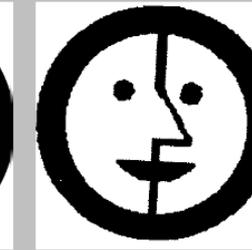
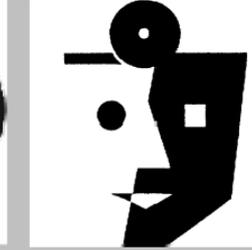
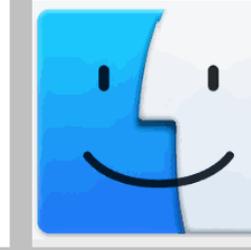
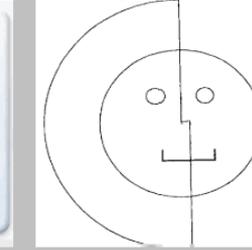
## Machine Learning prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**04.07.03-Geometric figures or combinations of geometric figures representing a person** Score: 10.81  
**02.01.33-Grotesque men formed by letters, numbers, punctuation or geometric shapes** Score: 10.74  
**02.01.01-Heads, portraits, busts of men not in profile.** Score: 10.16  
**02.01.37-Heads, portraits or busts of men in profile** Score: 9.71  
**02.11.16-Smiley faces** Score: 9.68  
**26.01.02-Plain single line circles** Score: 8.8  
**02.01.31-Stylized men, including men depicted in caricature form** Score: 8.52

 <p>US85573384.png Similarity: 0.850</p>	 <p>US85218923.png Similarity: 0.832</p>	 <p>US76416144.png Similarity: 0.767</p>	 <p>US86233323.png Similarity: 0.756</p>	 <p>US77821171.png Similarity: 0.749</p>	 <p>US85575421.png Similarity: 0.744</p>	 <p>US76443120.png Similarity: 0.733</p>
 <p>US87404013.png Similarity: 0.732</p>	 <p>US85332948.png Similarity: 0.730</p>	 <p>US75731540.png Similarity: 0.721</p>	 <p>US77504842.png Similarity: 0.706</p>	 <p>US74273614.png Similarity: 0.704</p>	 <p>US86486916.png Similarity: 0.699</p>	 <p>US76165120.png Similarity: 0.699</p>

Page 1/7



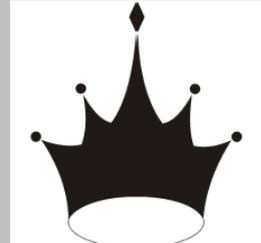
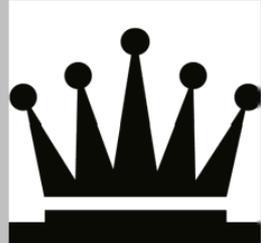
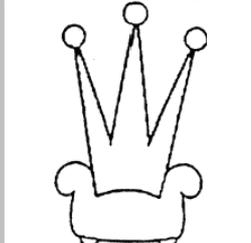
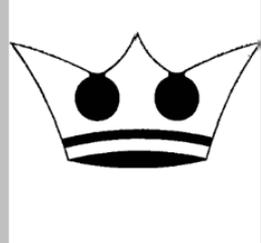
# Example 4: Rolex Machine Learning prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**24.11.02-Crowns open at the top** Score: 17.18  
**24.11.01-Crowns closed at the top** Score: 12.49  
**02.01.03-Men wearing crowns or other symbols of royalty, including kings, princes and jacks** Score: 10.34  
**02.11.07-Hands, fingers and arms** Score: 8.57  
**02.11.02-Eyes** Score: 6.55  
**11.01.25-Other non-electric kitchen utensils, utensil holders** Score: 6.49  
**21.03.01-Balls including playground balls, beach balls, billiard balls, tennis balls, bingo balls and lottery balls** Score: 6.44

 US85926217.png Similarity: 0.901	 US74292507.png Similarity: 0.869	 US79172020.png Similarity: 0.770	 US78338875.png Similarity: 0.694	 US78765837.png Similarity: 0.665	 US86174870.png Similarity: 0.661	 US85835860.png Similarity: 0.652
 US78449781.png Similarity: 0.639	 US86936455.png Similarity: 0.637	 US87036683.png Similarity: 0.633	 US75659831.png Similarity: 0.632	 US77435920.png Similarity: 0.620	 US75749693.png Similarity: 0.619	 US72338449.png Similarity: 0.613

Page 1/1

# Example 5: WWF

## Current Shape Similarity algorithm in GBD

Search interface for the WWF logo. The left side contains three empty input fields and a search button. The right side shows the selected image and the chosen strategy.

**1 Pick an image**



**2 Pick a strategy**

- Shape
- Color
- Texture
- Composite

delete

CURRENT FILTER

SOURCE:USTM \* ITY:Nonverbal \* IMAGE:Shape \*

Display: 60 per page options

Sort by Relevance - desc



# Example 5: WWF

## Machine Learning prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**03.01.13-Panda bears** Score: 12.37  
**03.01.24-Stylized cats, dogs, wolves, foxes, bears, lions or tigers** Score: 10.87  
**03.01.16-Heads of cats, dogs, wolves, foxes, bears, lions and tigers** Score: 10.84  
**05.05.25-Other flowers including daffodils and irises** Score: 8.75  
**03.01.14-Other bears** Score: 8.44  
**03.01.08-Dogs** Score: 7.99  
**03.13.01-Paws, feet, pawprints, footprints** Score: 7.73

 <p>US72432112.png Similarity: 0.826</p>	 <p>US74248198.png Similarity: 0.811</p>	 <p>US73295716.png Similarity: 0.808</p>	 <p>US73673862.png Similarity: 0.783</p>	 <p>US73308927.png Similarity: 0.778</p>	 <p>US74074561.png Similarity: 0.774</p>	 <p>US76019258.png Similarity: 0.772</p>
 <p>US85550507.png Similarity: 0.769</p>	 <p>US87306832.png Similarity: 0.766</p>	 <p>US74610791.png Similarity: 0.766</p>	 <p>US77894519.png Similarity: 0.766</p>	 <p>US77046781.png Similarity: 0.765</p>	 <p>US85313576.png Similarity: 0.764</p>	 <p>US77544044.png Similarity: 0.762</p>

Page 1/41

# Example 6: Current Shape Similarity algorithm in GBD

Search interface for the Current Shape Similarity algorithm in GBD.

**1 Pick an image**



delete 

**2 Pick a strategy**

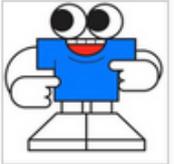
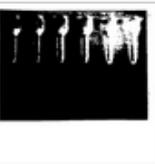
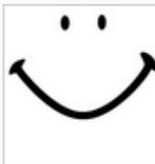
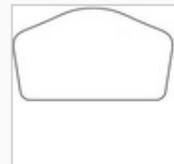
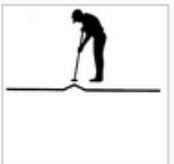
- Shape
- Color
- Texture
- Composite

CURRENT FILTER

SOURCE:USTM × ITY:Nonverbal × IMAGE:Shape ×

Display: 60 per page options 

Sort by Relevance - desc

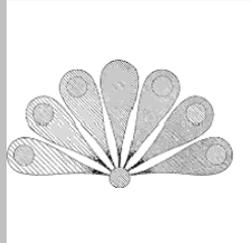
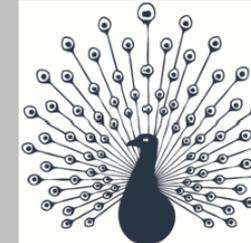
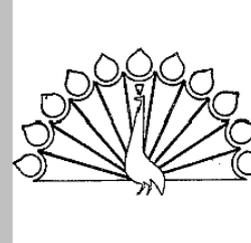
# Example 6: CNBC Machine Learning Prototype

WIPO Labs: AI powered Trademark Similarity Search

File



**03.15.12**-Pheasants, peacocks, quail Score: 17.37  
**03.15.24**-Stylized birds and bats Score: 16.51  
**01.15.18**-More than one drop including teardrops or raindrops Score: 13.28  
**05.05.25**-Other flowers including daffodils and irises Score: 12.87  
**03.15.25**-Other birds Score: 12.27  
**03.15.19**-Birds or bats in flight or with outspread wings Score: 12.25  
**03.15.06**-Ducks, geese, swans Score: 12.19

 US73555219.png Similarity: 0.976	 US74576101.png Similarity: 0.942	 US72169221.png Similarity: 0.824	 US73232433.png Similarity: 0.823	 US78559200.png Similarity: 0.803	 US85331965.png Similarity: 0.781	 US87232999.png Similarity: 0.775
 US73211286.png Similarity: 0.743	 US73763915.png Similarity: 0.737	 US78097086.png Similarity: 0.732	 US76716387.png Similarity: 0.723	 US72045991.png Similarity: 0.721	 US72150521.png Similarity: 0.714	 US72264550.png Similarity: 0.714

Page 1/7

# Thanks a lot for your attention

■ Questions?