Machine Learning applied to Trademarks Classification and Search

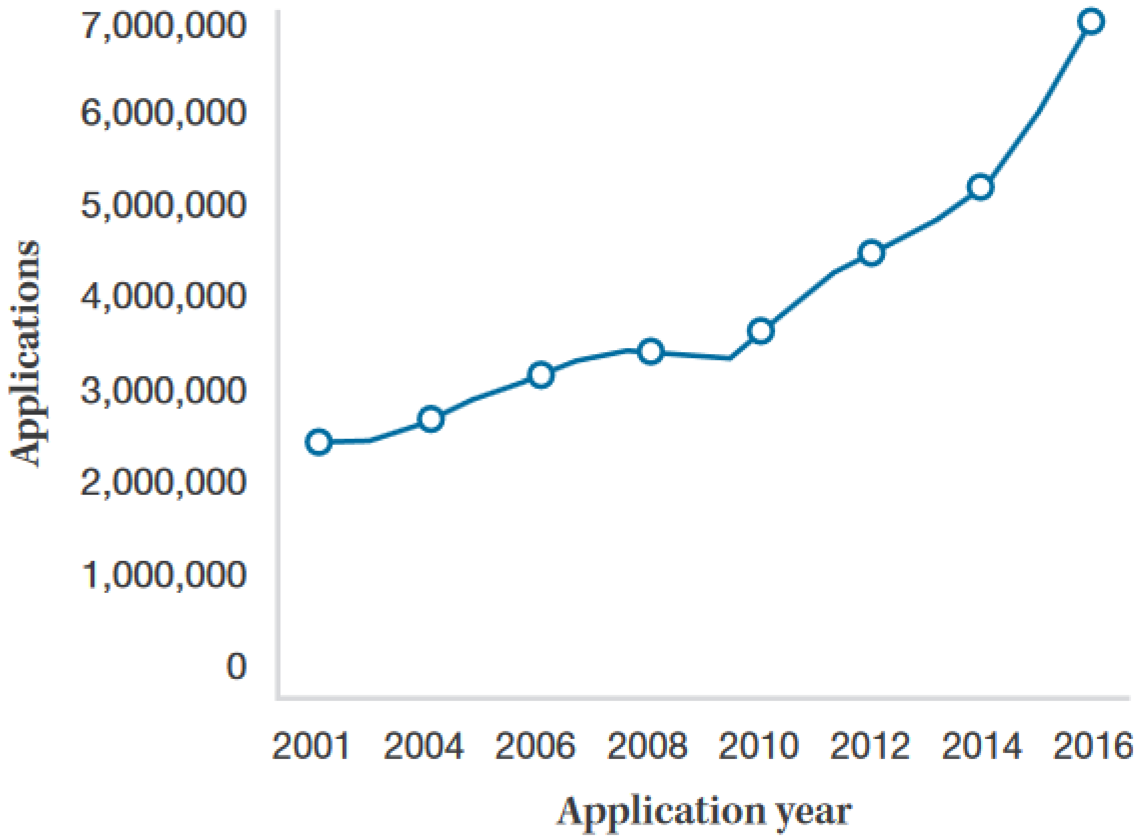
Speaker: Christophe Mazenc, Director, Global Databases Division, Global Infrastructure Sector

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

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?

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

(Information valid as of 2014-09-09)

International Trademark

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

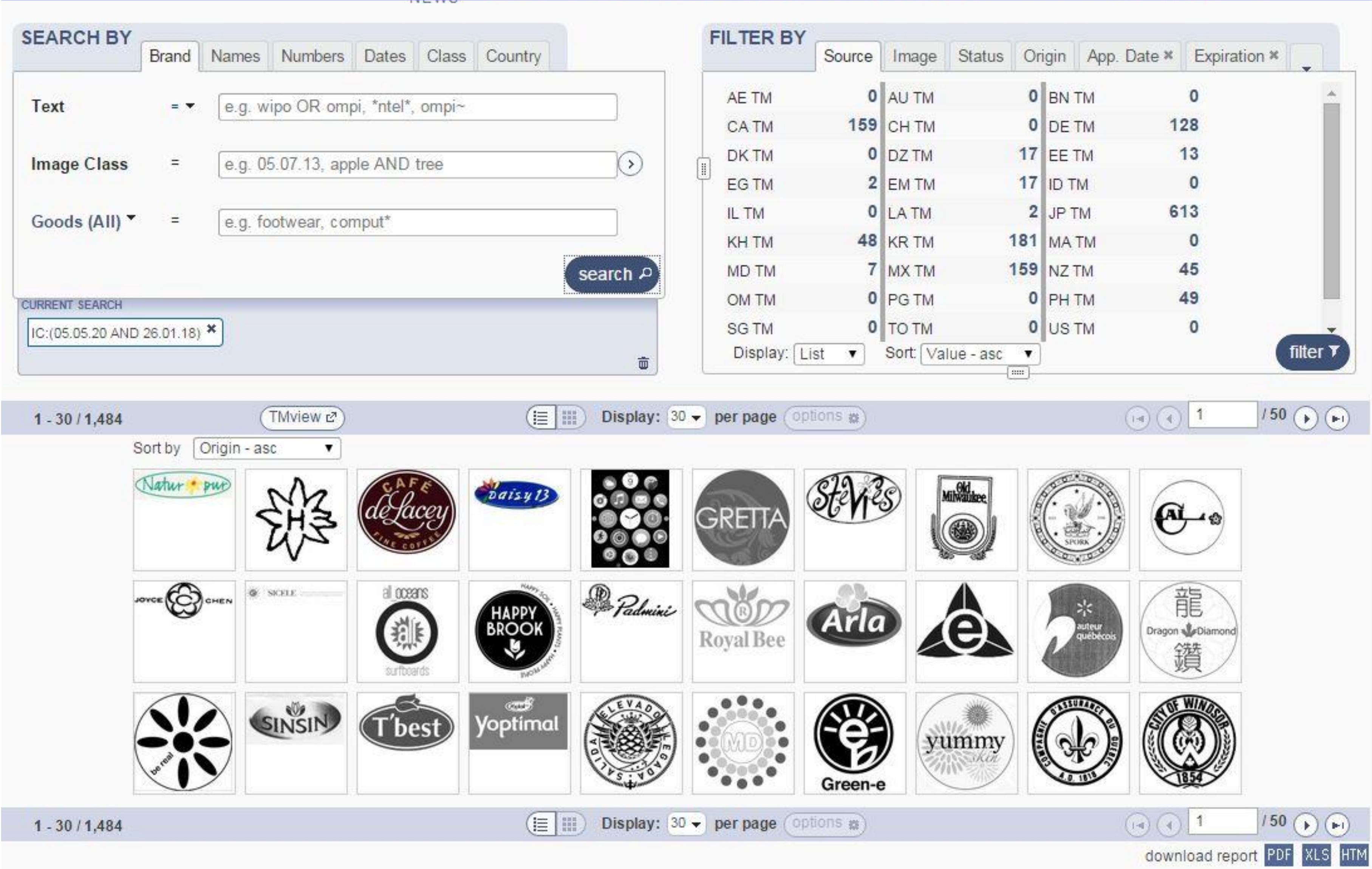
(531) International Classification of the Figurative Elements of Marks (Vienna Classification)- VCL (6)

05.05.20; 26.01.18; 29.01.13.

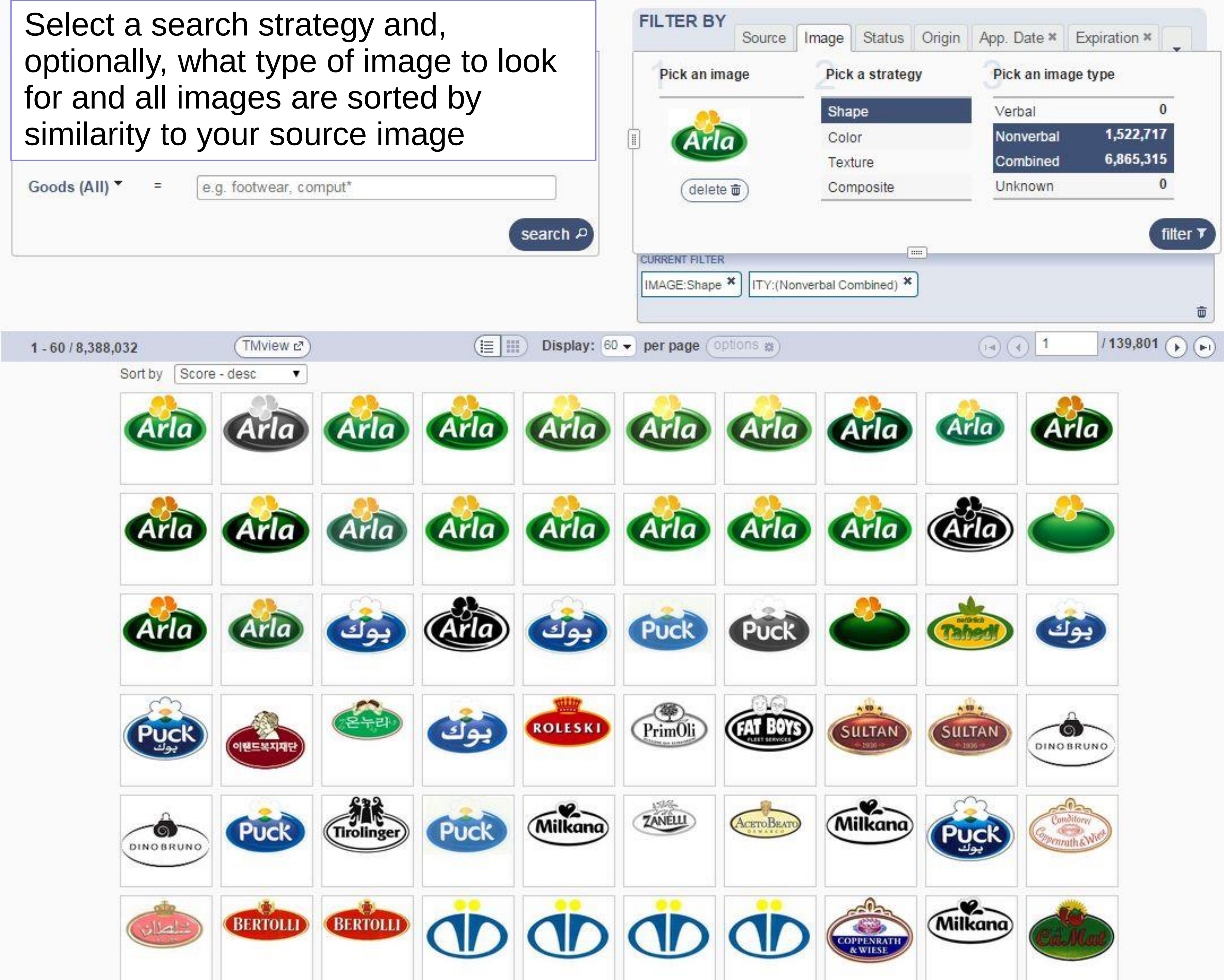
Stylized flowers

(591) Informat... ors claimed

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



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



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

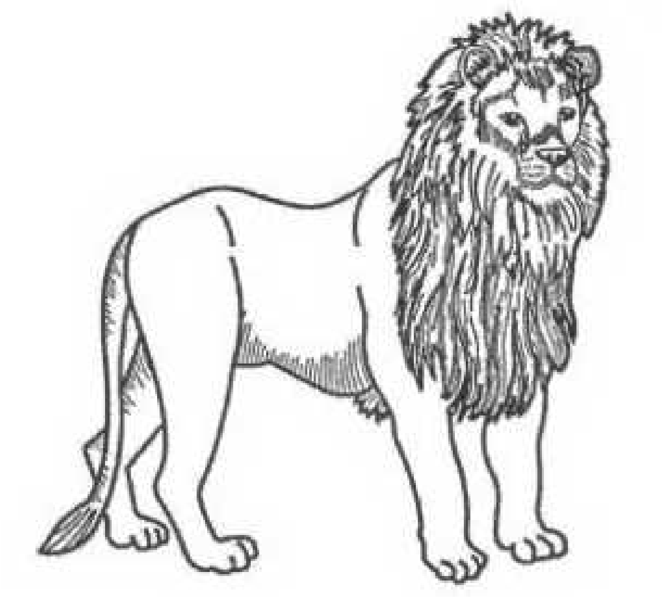
- Three levels classification with a total of more than 1300 different categories

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

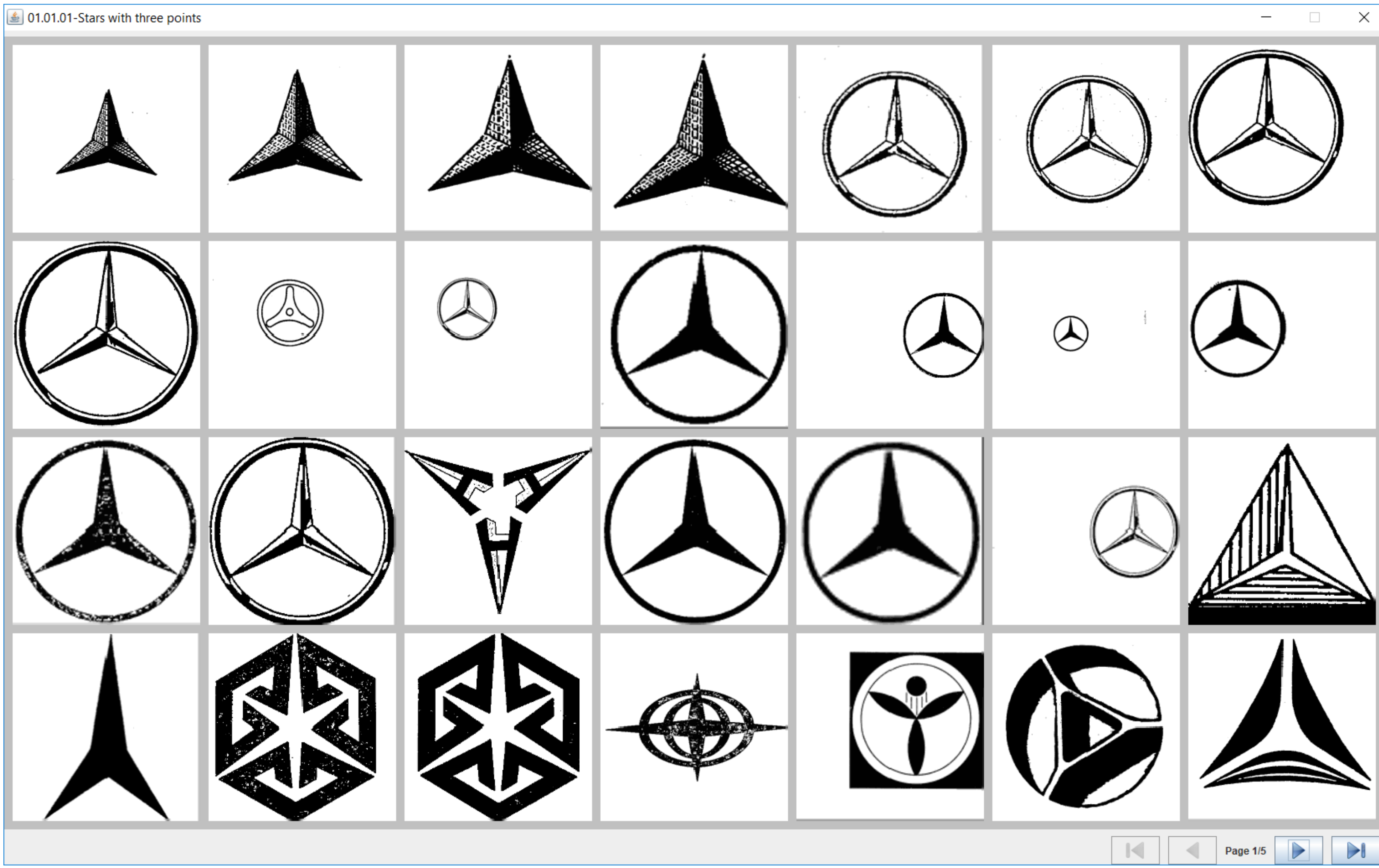
- 194’450 trademark images only figurative
- Having been classified with 508’183 codes
- The corpus is split as follows:
  - 98% for the training
  - 2% for the dev/tests
- The training of one model takes around 3 days on a dedicated server with GPU

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)

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

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

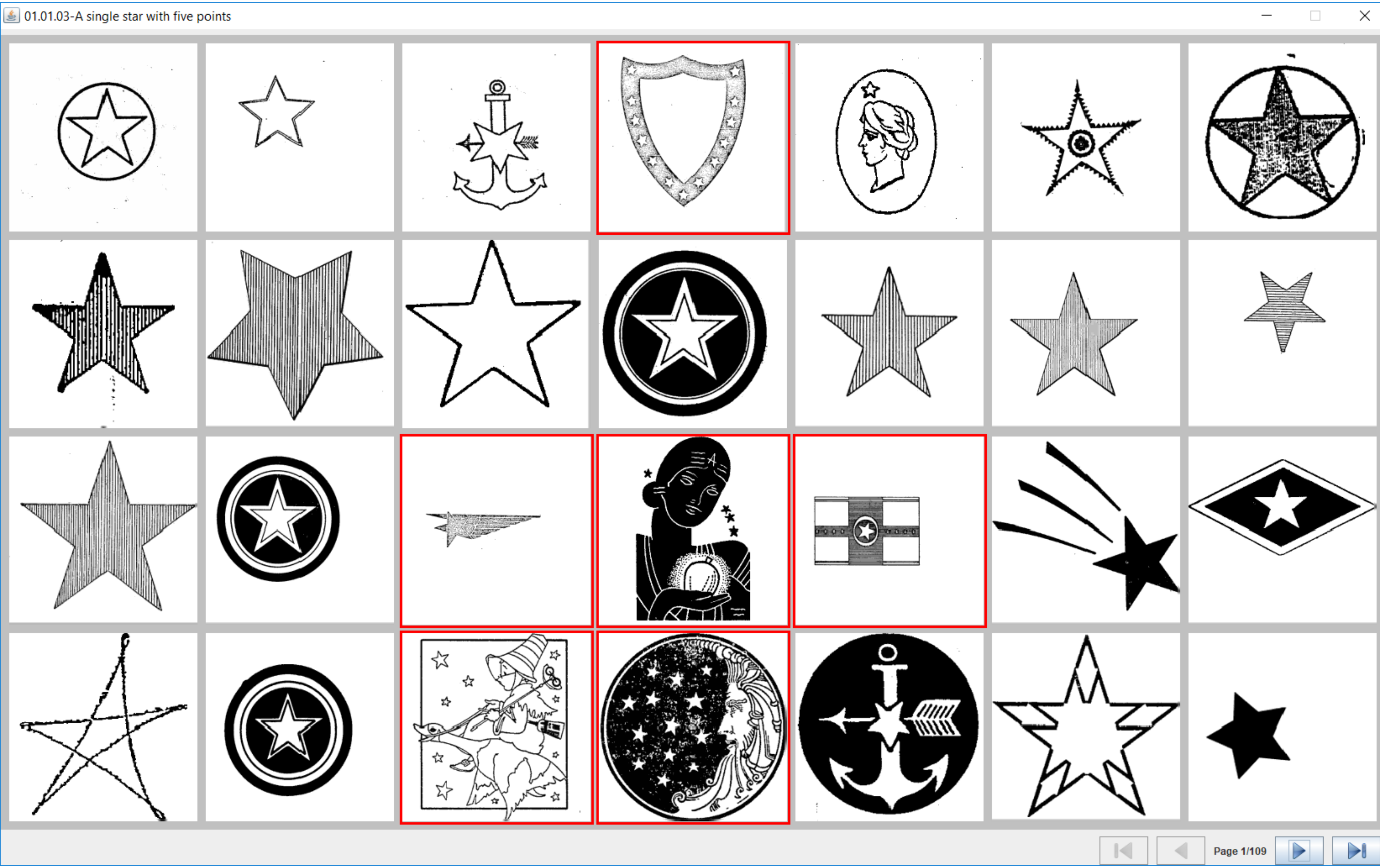
  02.01.27: policemen, firemen: 10

  …

  02.01.02: shadows or silhouettes of men: 6652

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

Corpus quality challenge 3



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

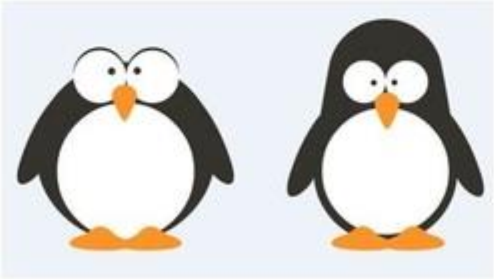
Example 3

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

WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION

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

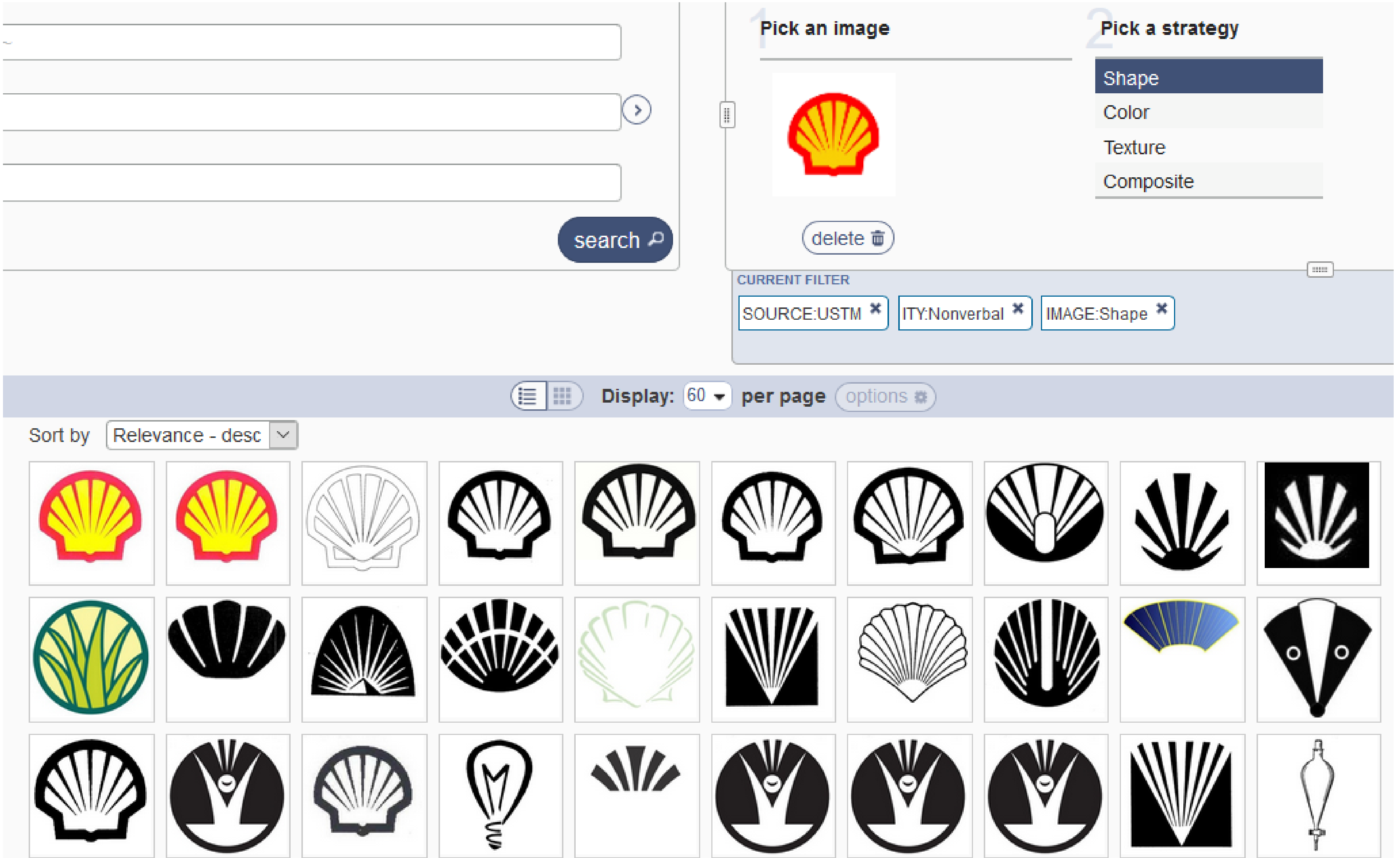
- Idea: the NN classifier outputs for an input image a vector of 1300 dimensions (one per design code)
- A measure of similarity of two images can be obtained by computing the distance between the two classification vectors in the 1300 dimensions space
- A similarity search of a new image is performed 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

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

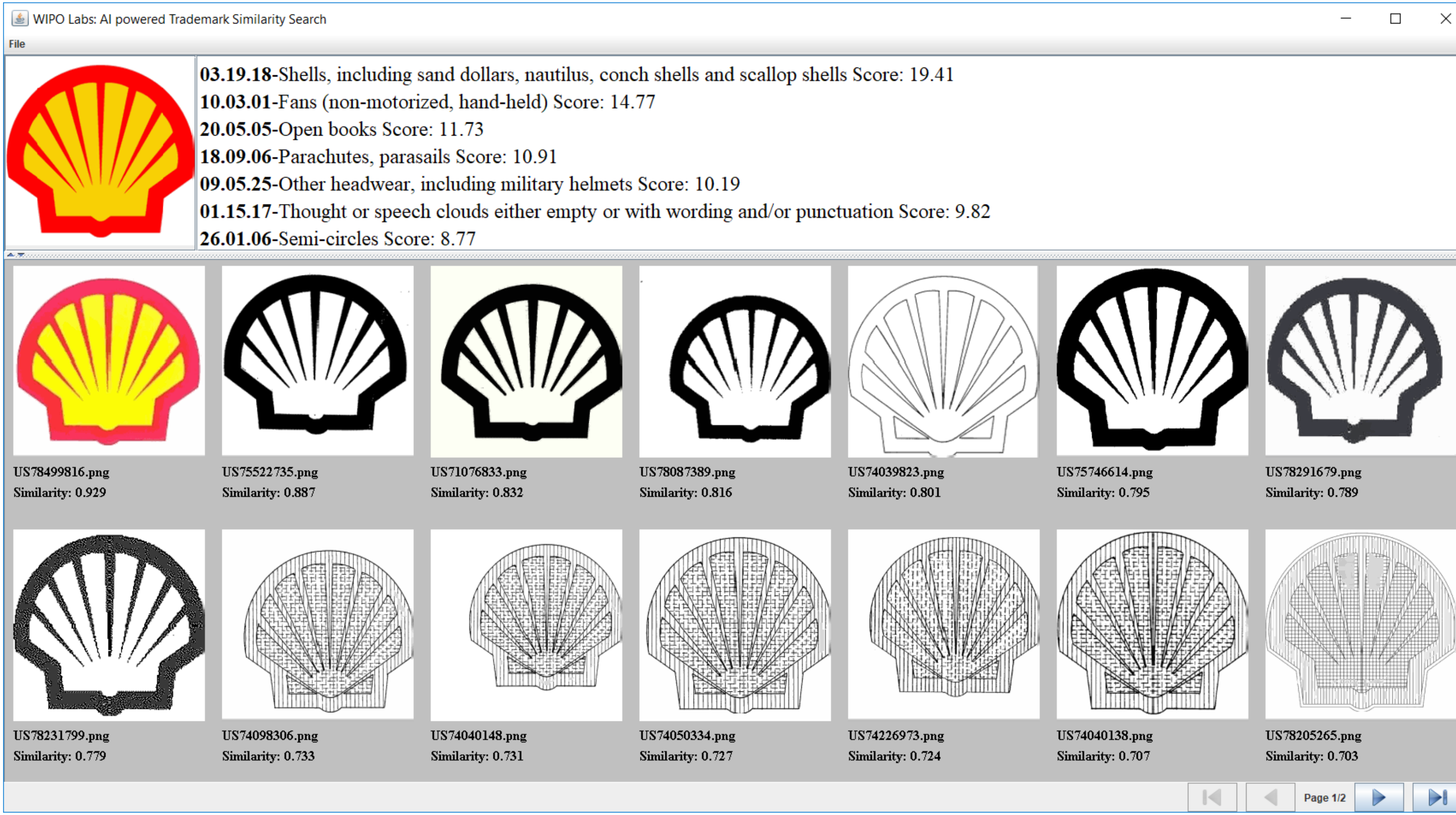
Assessment

- 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)

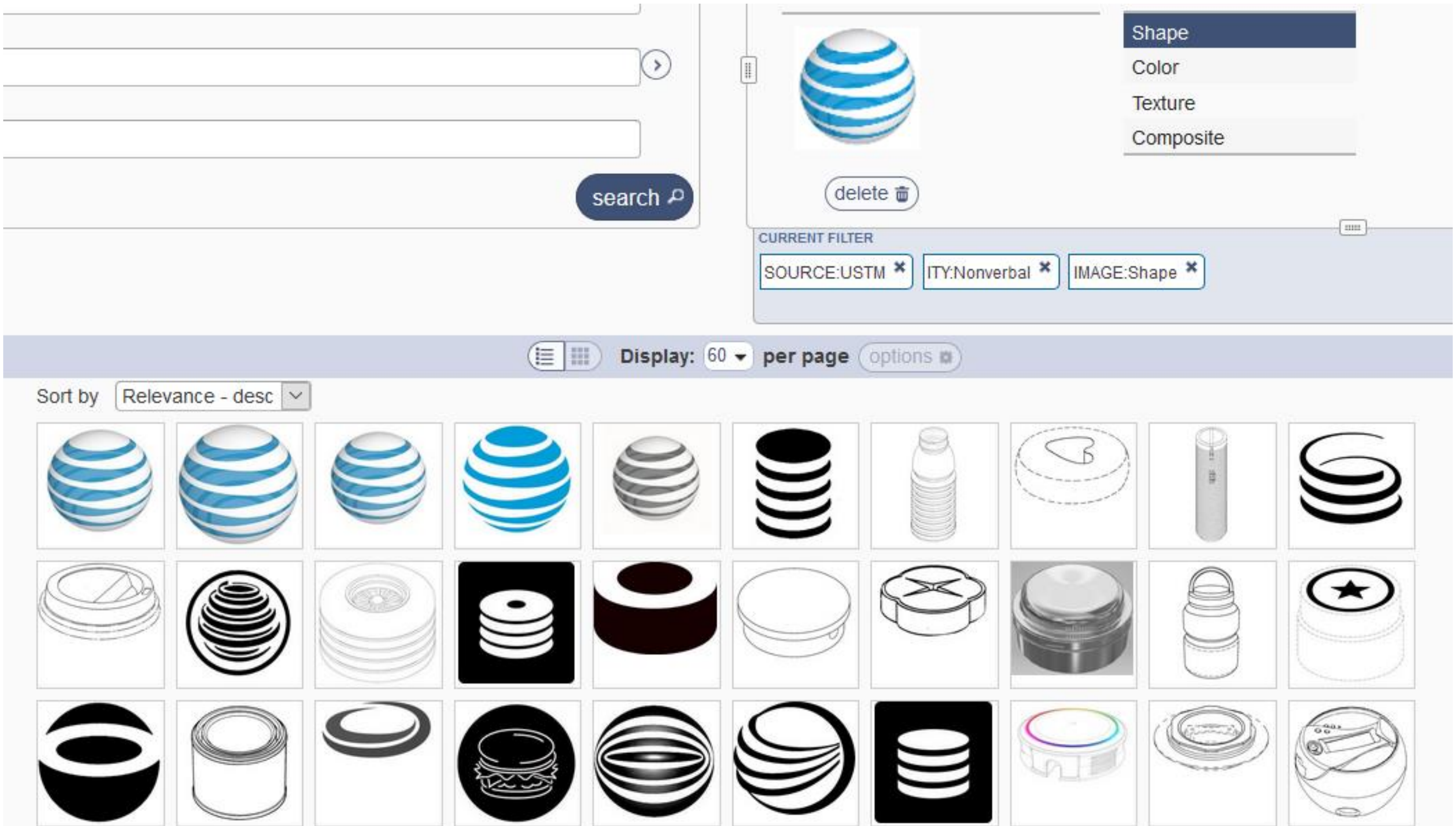
Example 1 Shell
Current Shape Similarity algorithm in GBD



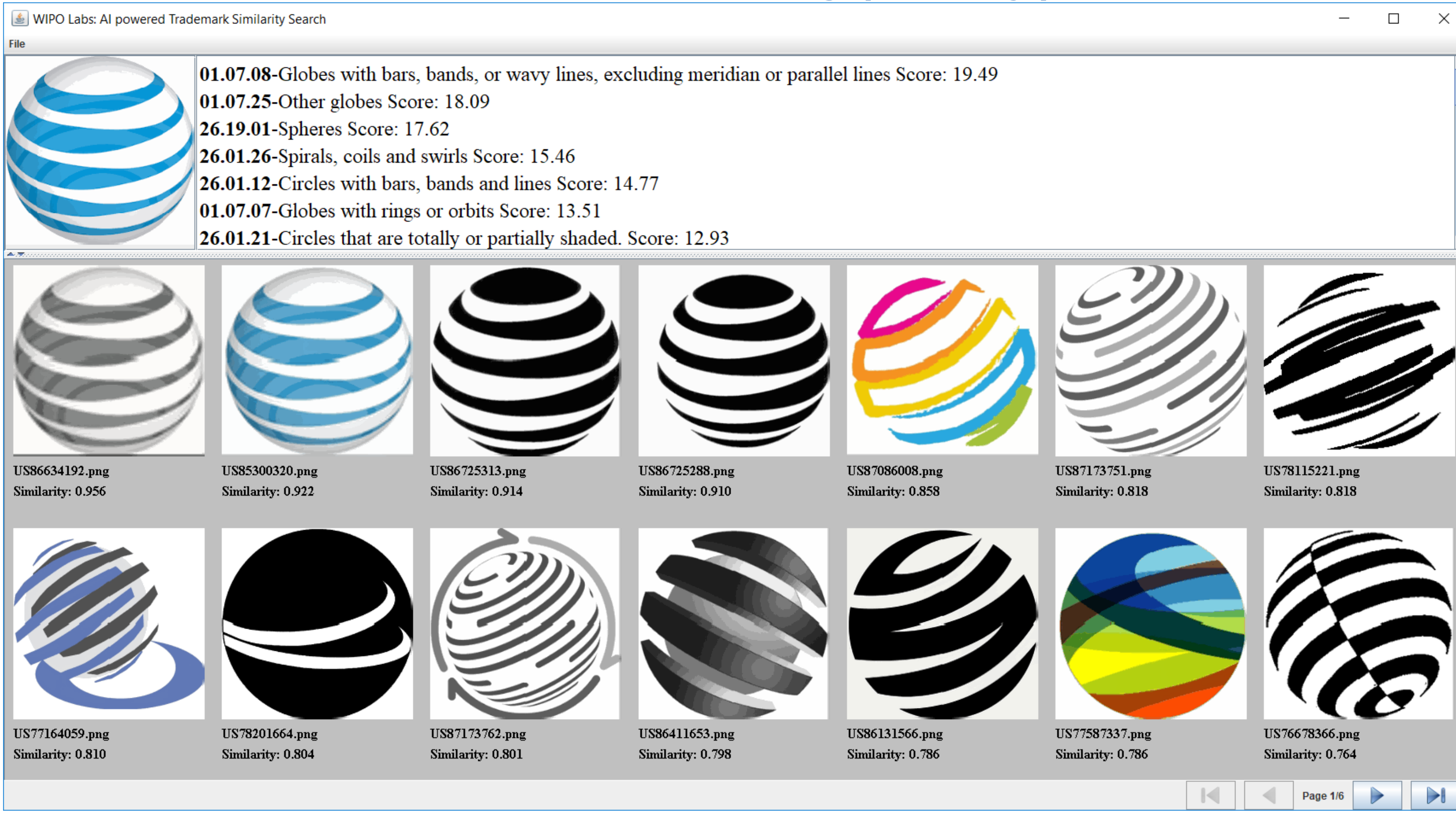
Example 1: SHELL
Machine Learning Prototype



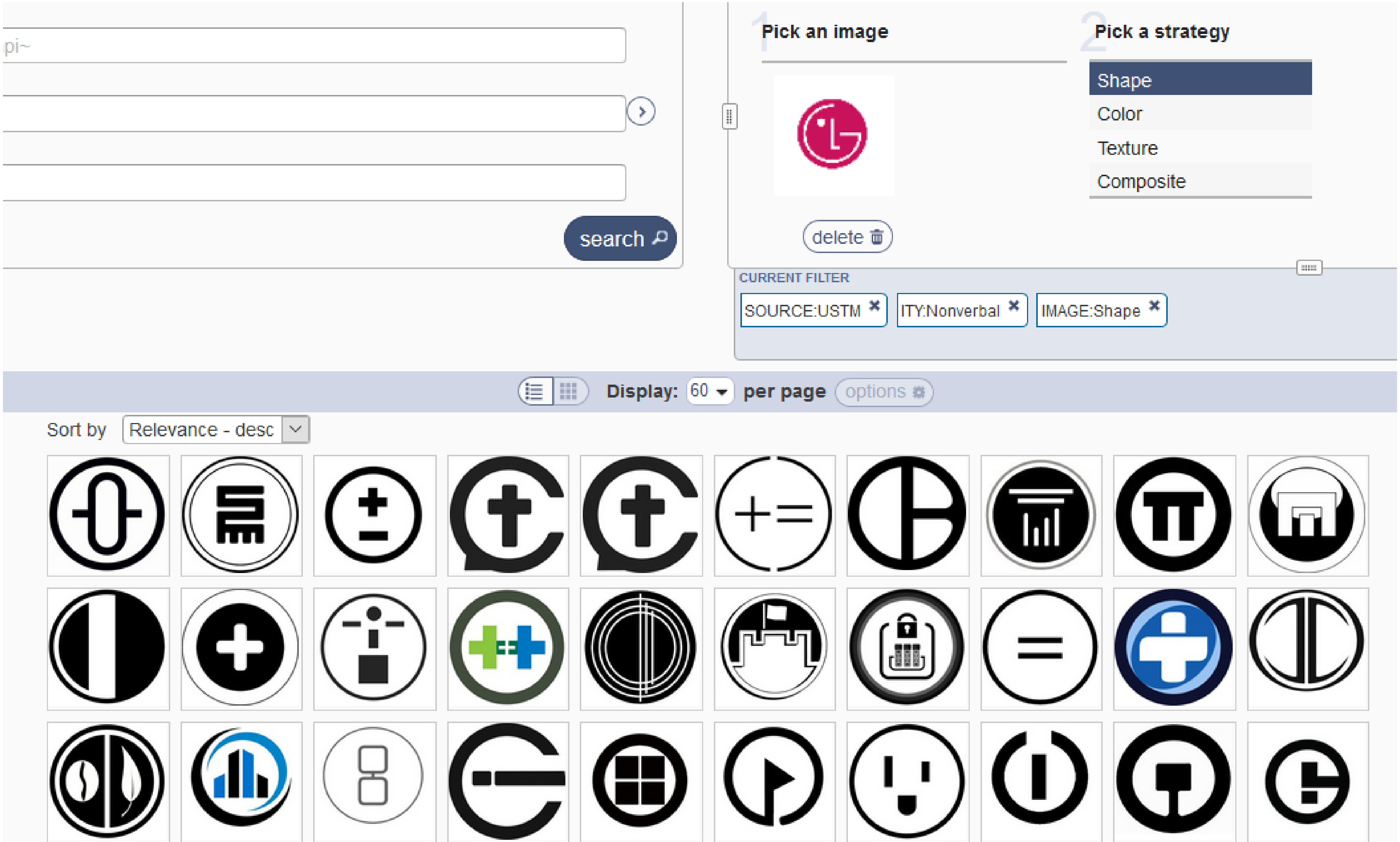
Example 2: AT&T
Current Shape Similarity algorithm in GBD



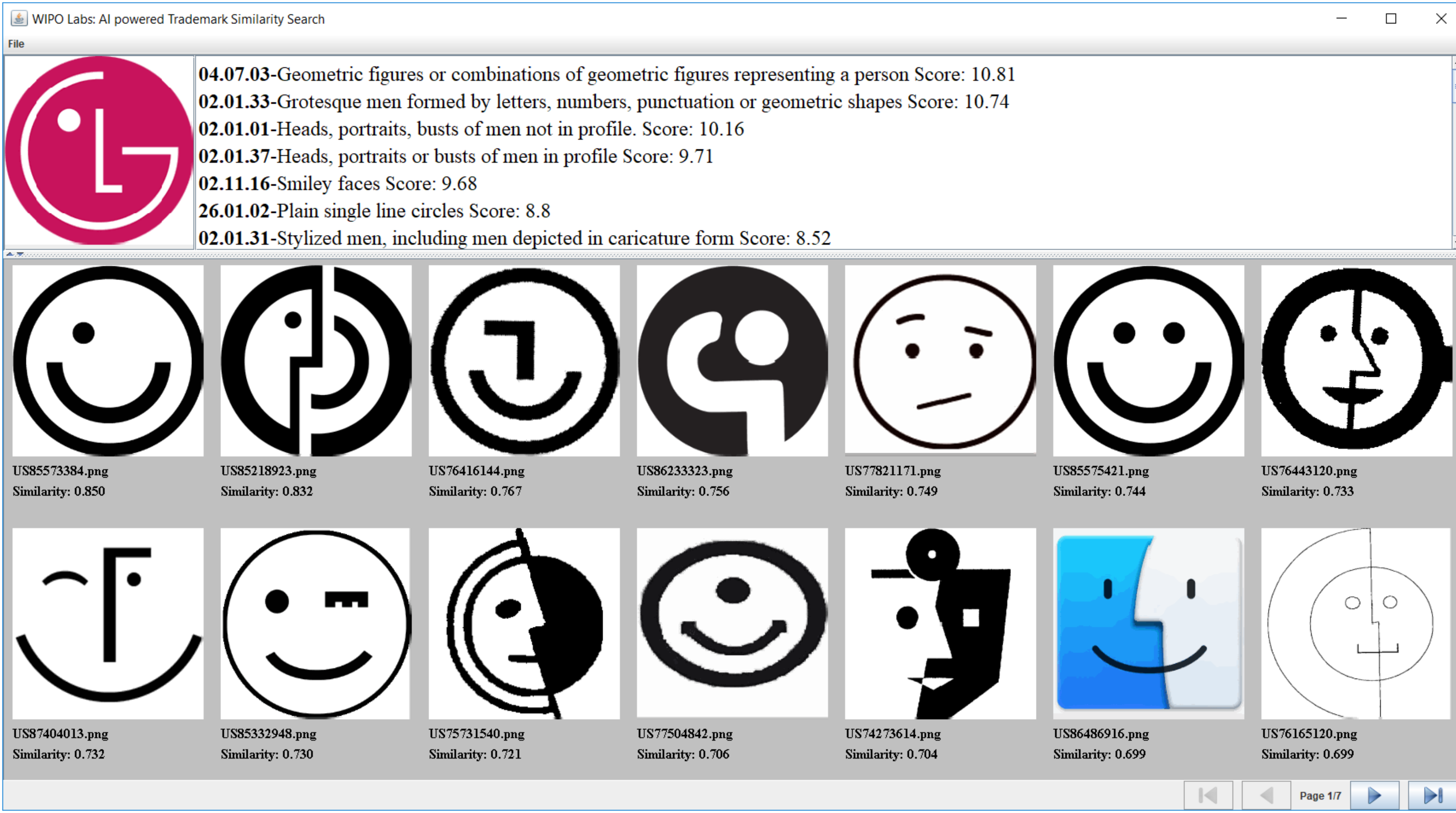
Example 2: AT&T
Machine Learning prototype



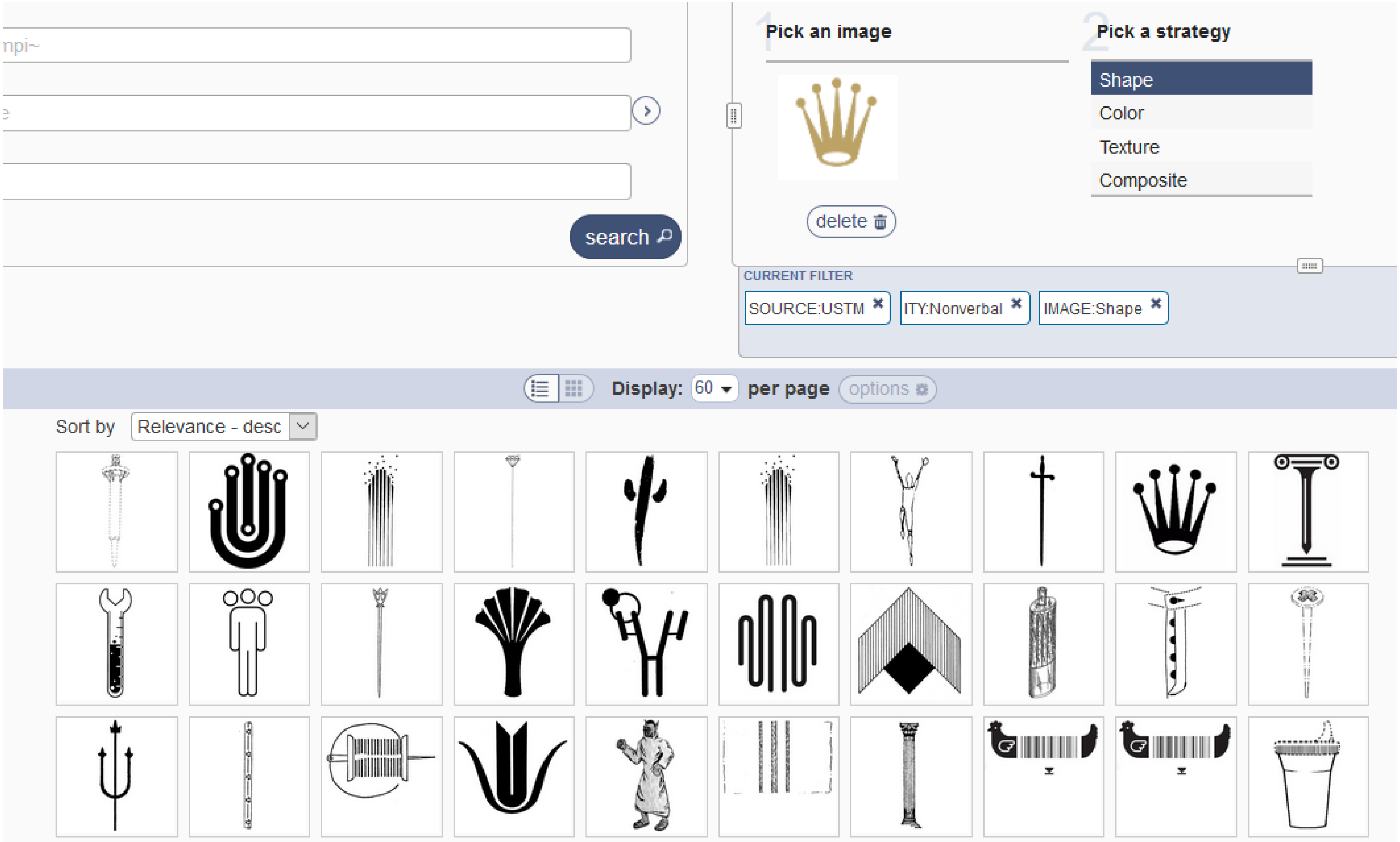
Example 3: LG
Current Shape Similarity algorithm in GBD



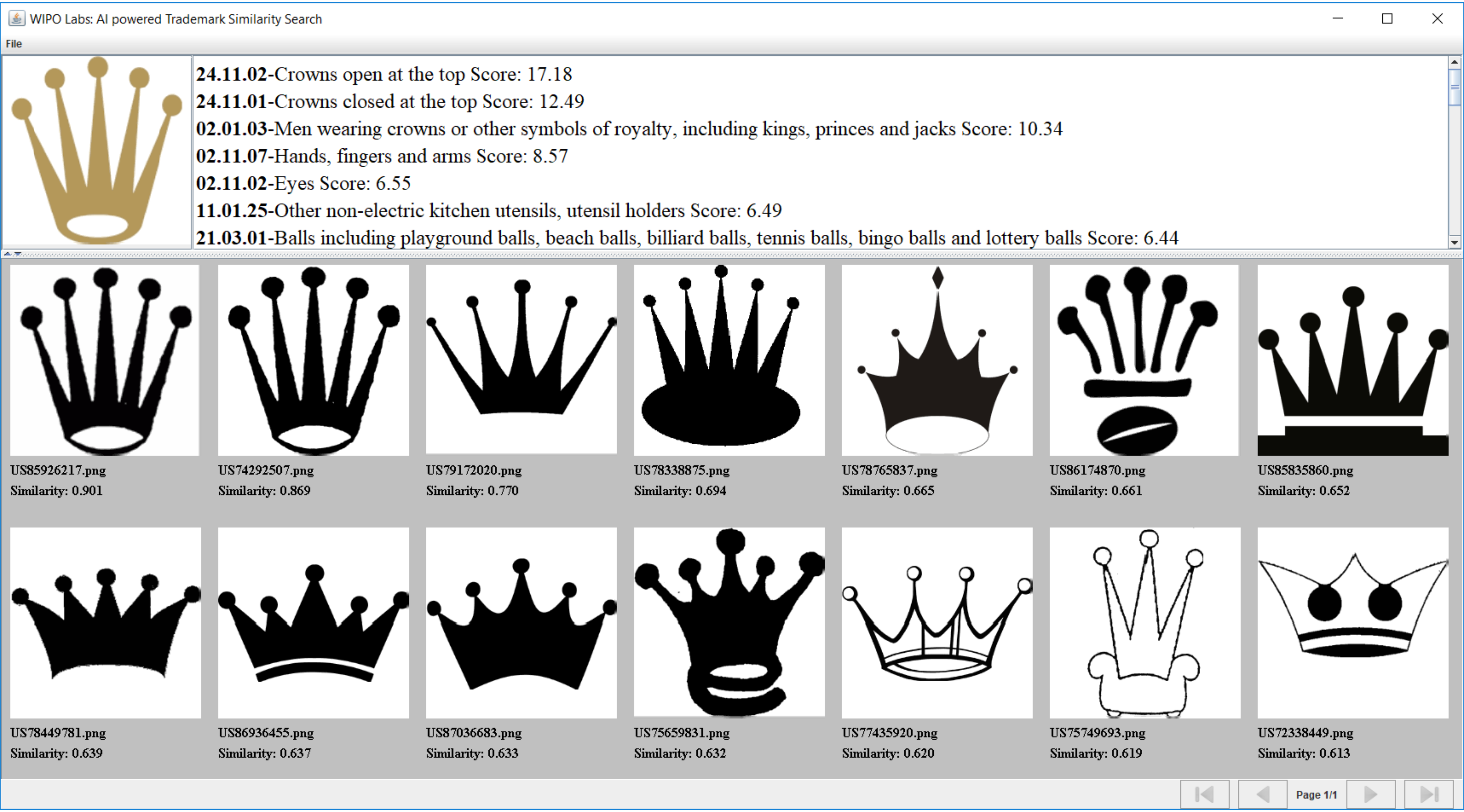
Example 3: LG
Machine Learning prototype



Example 4: Rolex
Current Shape Similarity algorithm in GBD

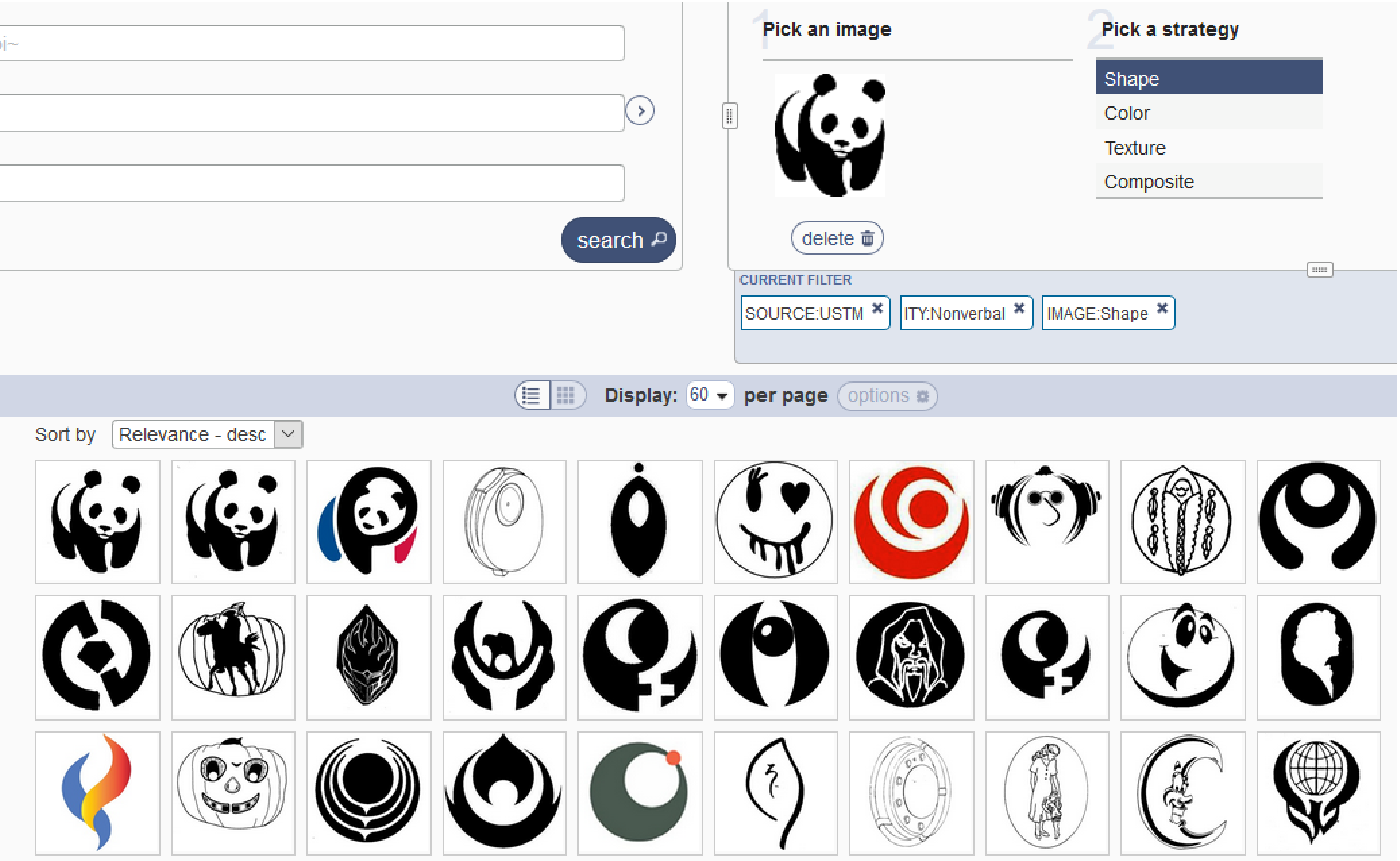


Example 4: Rolex
Machine Learning prototype

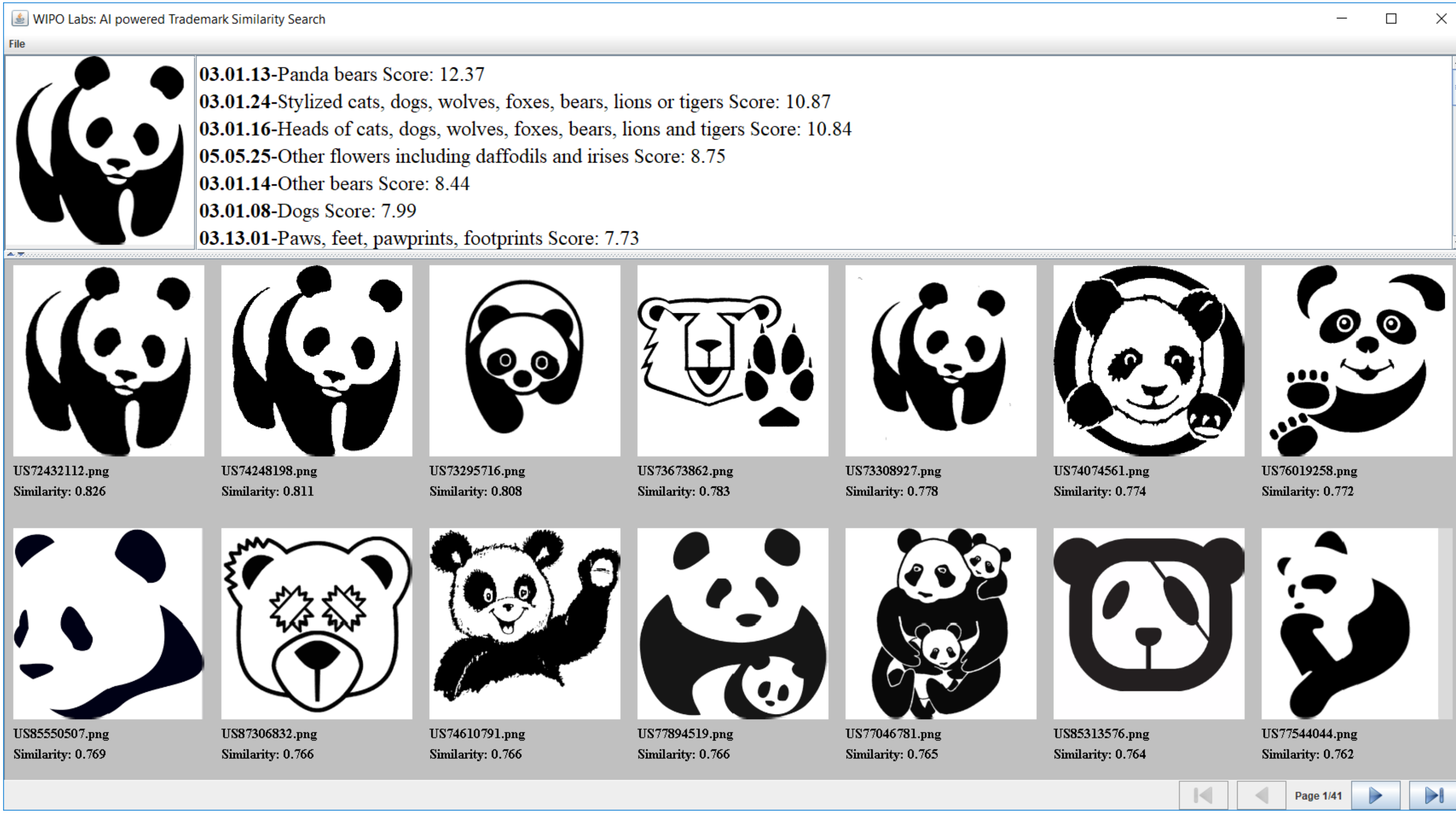


WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

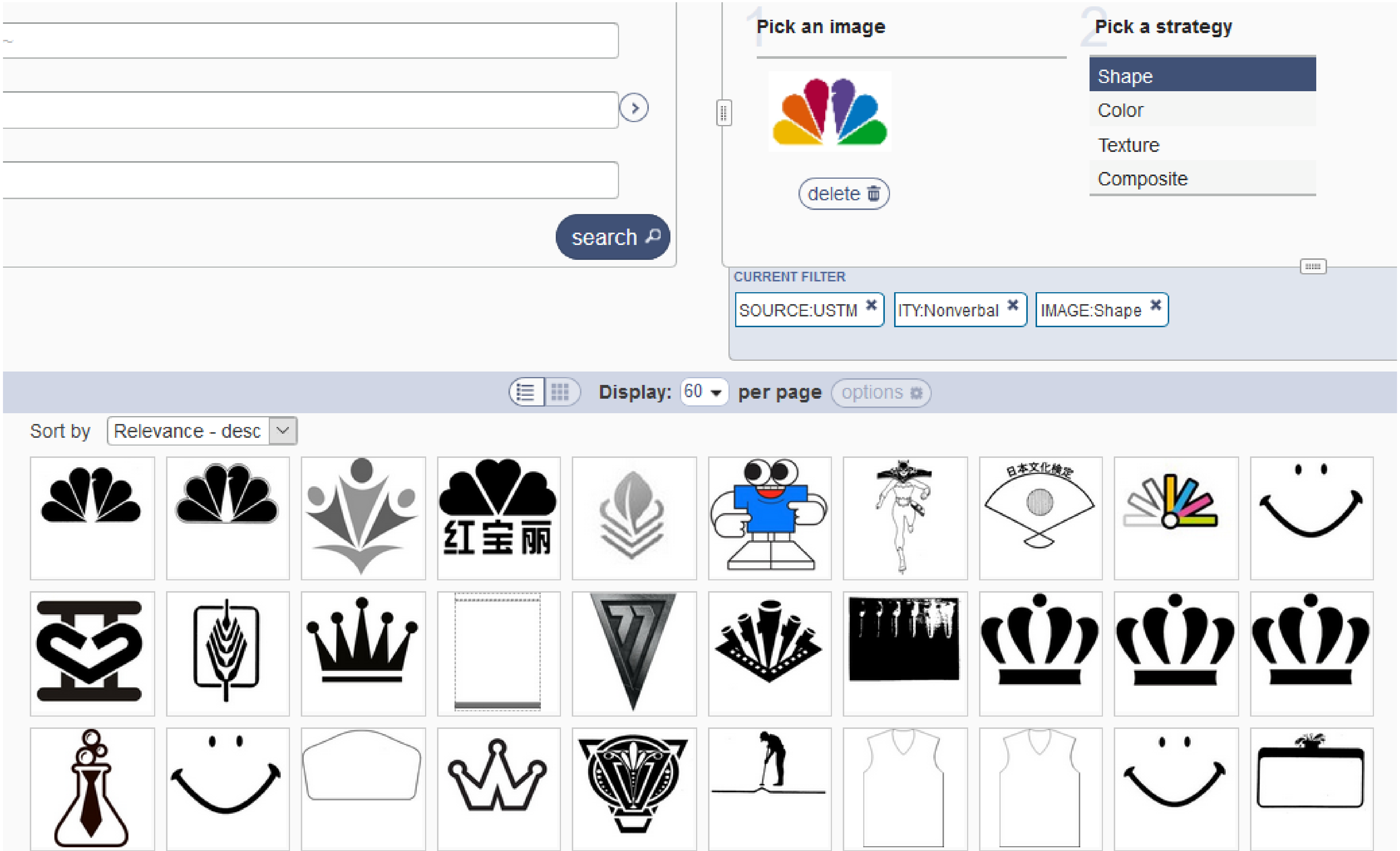
Example 5: WWF
Current Shape Similarity algorithm in GBD



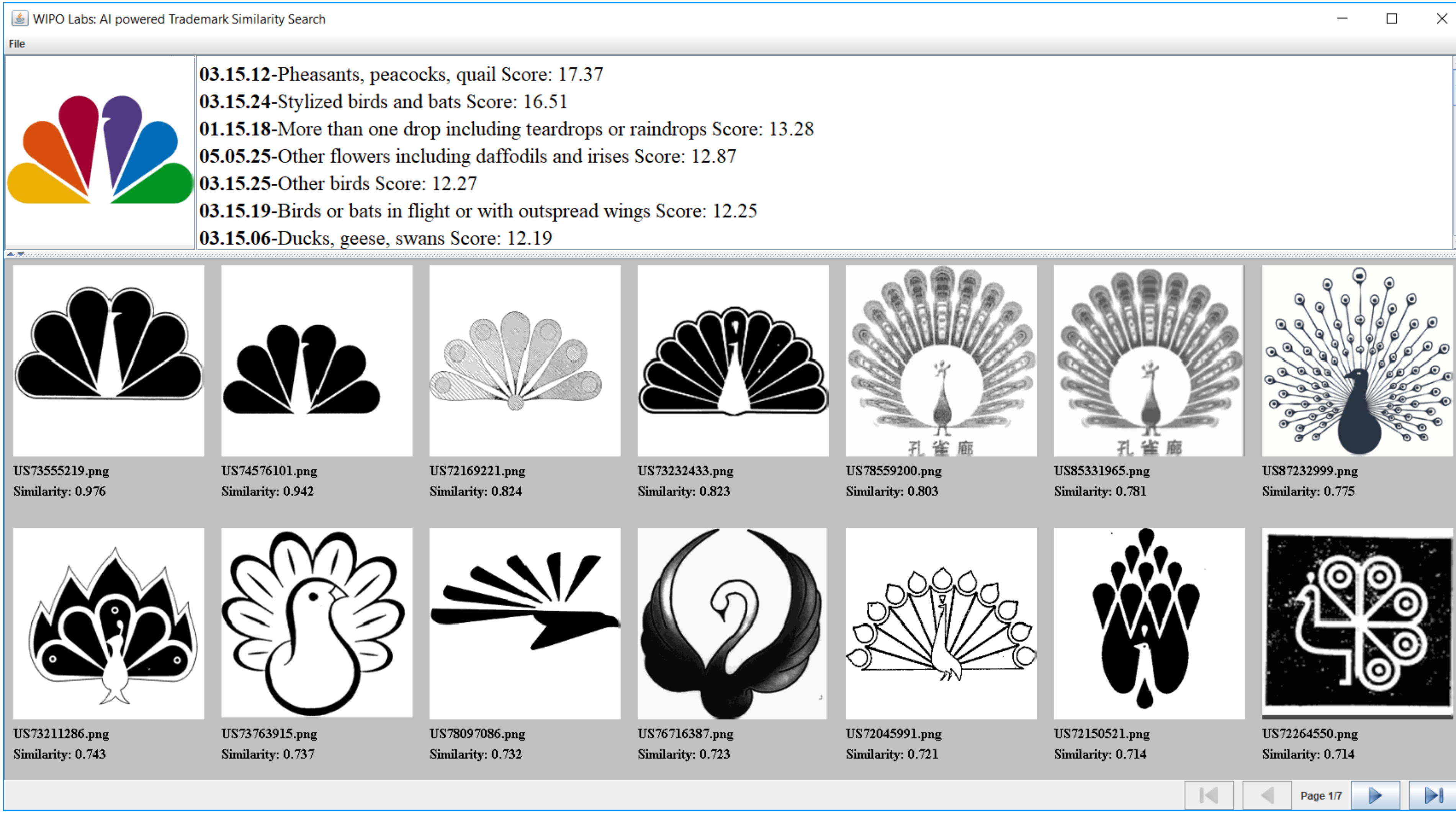
Example 5: WWF
Machine Learning prototype



Example 6:
Current Shape Similarity algorithm in GBD



Example 6: CNBC
Machine Learning Prototype



Thanks a lot for your attention

- Questions?