

Proofpoint Logo Detection

09/17/2020 Update

Duke

Logistics

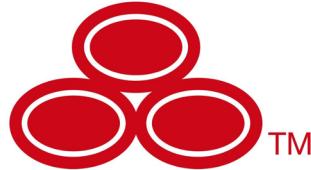
- Slack/GitLab
 - Everyone should have access now
- Duke Computing Cluster
 - We have all attended an orientation session
- Meetings with David
 - Presenting summaries of literature

Clean Logo Dataset

- Extracted from <https://1000logos.net/>
 - Well-structured links for images
 - Each image contains the brand name
 - Over 3500 brands across many different categories
 - Collected for around 250 from more relevant categories (financial, insurance, internet)
- Wrote script to download and save images to individual folders
 - High quality png images of primary logo
 - Jpeg images of logo history, emblem, font, colors, etc
- Some manual download required on historical logos
 - Some are poorer quality and may not be necessary (before 1980s)

Clean Logo Dataset - Sample Collection

StateFarm



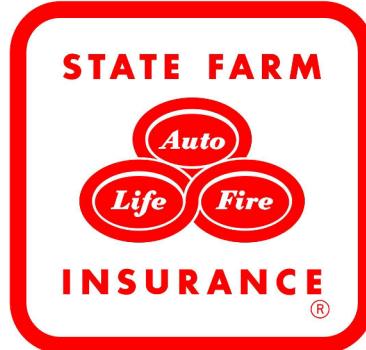
StateFarm

1922

2006

 StateFarm™

2012—now



 StateFarm™

Duke

Screenshots Dataset

- Focused on diversifying our screenshots by scraping Google.
- We developed functionality to:
 - Obtain image urls from Google Images search results (search terms like Apple Inc. logo, etc.).
 - Take Screenshots of Google Images results pages (discussed on next slide).
 - Obtain urls to screenshot from regular google search results.
 - Download Google Images results using a Chrome plugin.
- URLs obtained may contain false positives but are majorly true positives for containing images of the brand's logo.
- Current Screenshots Resolution:
 - Height: 720, Width: 1280
 - Device Scale Factor: 2.0

Google Image Screenshots

- Google ('Company Name + Logo')
- Advantages
 - Locational Variety
 - Multiple logos/screenshot
- Disadvantage
 - Annotation can get subjective
- Is it worth exploring?
 - If yes, we could focus on the first ~5 rows of results

Resolution - 2256 X 4132





CONSUMERS

HEALTHCARE PROFESSIONALS

CAREERS

ABOUT ABBOTT

INVESTORS

NEWSROOM

RESPONSIBILITY

LIVE HEALTHY

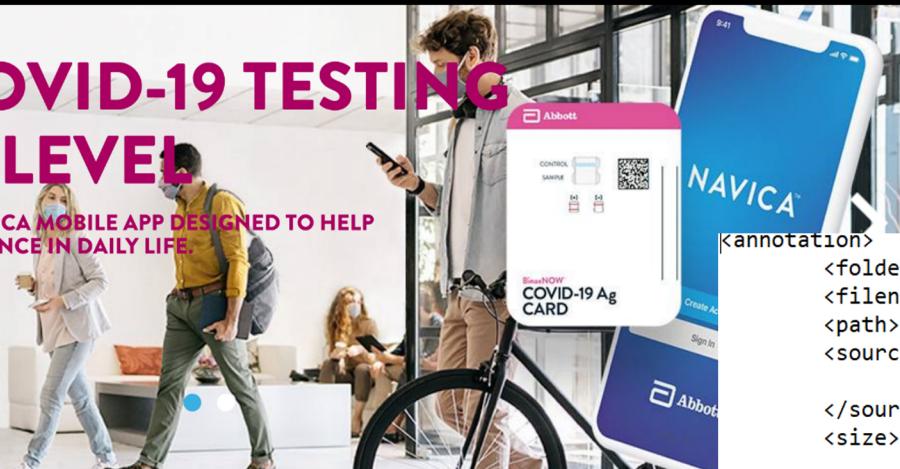


SEARCH



TAKING COVID-19 TESTING TO A NEW LEVEL

RAPID ANTIGEN TEST AND NAVICA MOBILE APP DESIGNED TO HELP RESTORE A BIT MORE CONFIDENCE IN DAILY LIFE.

[LEARN MORE](#)[WHAT'S TRENDING](#)

PASCAL VOC FORMAT

```
<annotation>
<folder>img</folder>
<filename>95dfc06f3db3edde2a3386fa2e321b32ca4e1.jpg</filename>
<path>C:\Users\Vishaal\Downloads\scraped_images\95dfc06f3db3edde2a3386fa2e321b32ca4e1.jpg</path>
<source>
    <database>Unknown</database>
</source>
<size>
    <width>2560</width>
    <height>1440</height>
    <depth>3</depth>
</size>
<segmented>0</segmented>
<object>
    <name>abbott_laboratories</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
        <xmin>25</xmin>
        <ymin>74</ymin>
        <xmax>376</xmax>
        <ymax>191</ymax>
    </bndbox>
</object>
</annotation>
```

Next Steps

- Collect more clean logos
 - Start building that classification model
- Scraping and annotating screenshot image in bulk.
- Coding our first YOLO v3 implementation.
- Deep dive into the Duke Computing Cluster to store our data
- Exploring documents/emails for logo detection

Appendix

- Sample current resolution screenshots.

Join Visa in supporting small businesses

Find out about our donation program



PREPAID DEBIT CARDS

A quick and easy way to get your government benefit funds

See the advantages of accessing your funds with Government Payment Cards.

[Visa EIP card blog](#)

ESSENTIAL TASKS

Find a card

Explore offers

Click to pay

Get support

Lost or stolen card



Solutions Insights What We Do Conferences About Careers

Search

Help

Log in

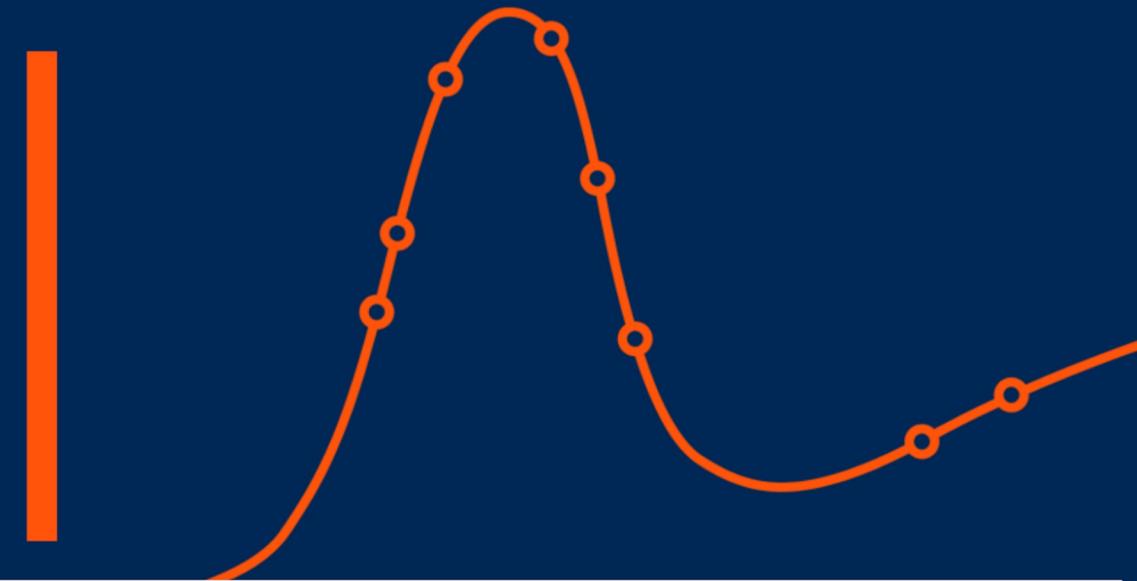
Become a Client

1855 651 9836

The Gartner Hype Cycle for Emerging Technologies, 2020

Five must-know technology trends that will impact business and society

→ [Read Now](#)



Research

Quick Guide to Strategic Planning

[Read Now](#)



Virtual Conference

Gartner Security & Risk Management Summit

[Learn More](#)