

Detecting Lookalike Domains

MSc Thesis
Final Presentation

30 March 2021

Patrick Frischknecht

What are lookalike domains?

Definition:

“Lookalike” domains are domains that are crafted to impersonate the URLs of other sites in order to trick users into believing they're on a different site.¹

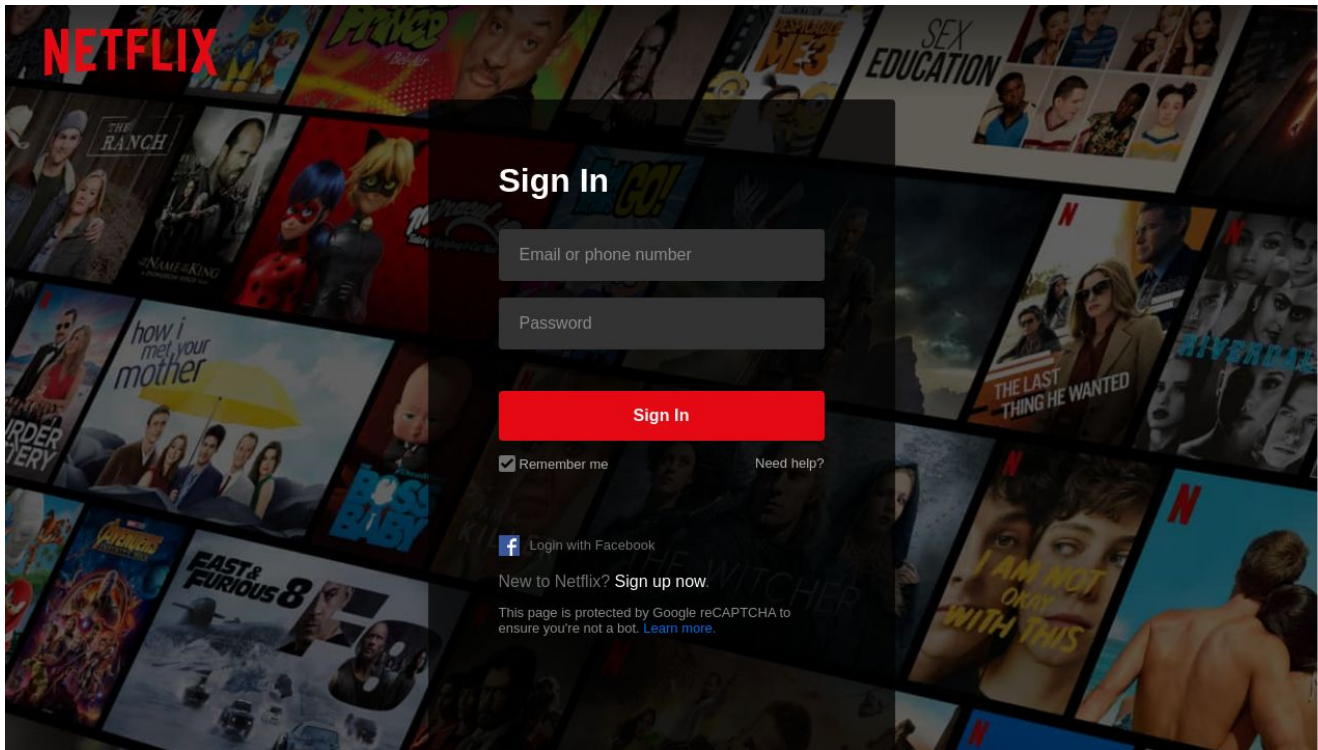
Examples:

nike.com => njke.com (Typosquatting)

nike.com => nike-outlet.com (Combosquatting)

1) <https://chromium.googlesource.com/chromium/src/+master/docs/security/lookalikes/lookalike-domains.md>

Threats



uk-mynetflix.com

Phishing



🔍 Produkte suchen...

[Startseite](#) [Mein Konto](#) [Kasse](#) [kontaktiere uns](#)

€0.00 0 Artikel



🔍 Produkte suchen...

Produktkategorien

- Air Max 1
- Air Max 2017
- Air Max 2018
- Air Max 2019
- Air Max 270
- Air Max 90
- Air Max 95
- Air Max 97
- Air Max 98
- Air Max Plus
- Vapormax
- Zero

nike billig

Die Air Max 90 und die Air 180, die mit immer größeren Luftblasen entworfen wurden, überfluteten die Straßen für den täglichen Gebrauch häufiger als sie für das Laufen und / oder Training verwendet wurden. Nicht zuletzt aufgrund seines schlanken Looks, der sich immer wieder an die Spitze des Pantheon für Lifestyle-Schuhe anlehnt, gehört der 90 weiterhin zu den führenden Air Max-Modellen von Swoosh. Mit den 180ern, die im Jahr 2018 eine Renaissance erlebten, die durch eine aufmerksamkeitstarke Zusammenarbeit mit COMME des Garçons unterstrichen wird, treibt Nike Sportswear seine Fans immer wieder mit neuer Hitze an.

nikebilliger.com

Trademark abuse

Buy Adidas.run For USD 3,699

The owner of Adidas.run has chosen [Escrow.com](#), a licensed U.S. escrow company, to process the sale of this domain.



If you are unfamiliar with Escrow.com, please click below to see how it works.

[How It Works](#)

Use the form to initiate an Escrow.com transaction to securely buy Adidas.run for USD 3,699 through Escrow.com.

Upon submitting the form, we will send you a confirmation email. Please click on the link in that email to start the transaction.

If you have an account with Escrow.com, please use your existing account email address. If you do not have an account, Escrow.com will send you instructions to create an account prior to starting the transaction.

First name

Last name

Email address

Phone number

Prove you are not a robot

Ich bin kein Roboter.

reCAPTCHA
Datenschutzerklärung - Nutzungsbedingungen



[Submit](#)

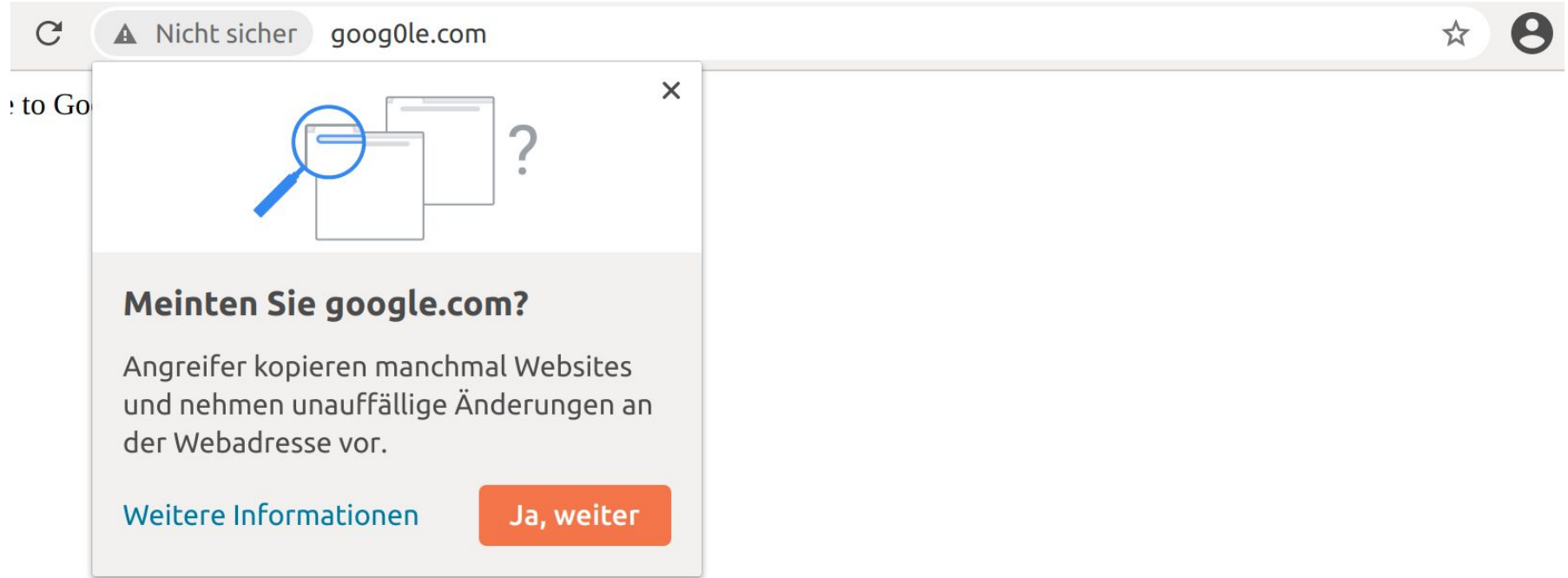
[Privacy Policy](#)

adidas.run

Domain selling / advertisement (parked domains)

State of the Art

Reactive: Browser warnings



The image shows a browser address bar with a warning icon and the text "Nicht sicher" (Not safe) next to the URL "goog0le.com". Below the address bar, a warning dialog box is displayed. The dialog box has a close button (X) in the top right corner. The main content of the dialog box includes an illustration of a magnifying glass over a document with a question mark, the heading "Meinten Sie google.com?", a paragraph of text explaining that attackers sometimes copy websites and make subtle changes to the web address, and two buttons: "Weitere Informationen" (More information) and "Ja, weiter" (Yes, continue).

to Go

Meinten Sie google.com?


Angreifer kopieren manchmal Websites und nehmen unauffällige Änderungen an der Webadresse vor.

[Weitere Informationen](#) [Ja, weiter](#)

Proactive: Generative approaches

Steps:

- 1) Generate a dictionary of lookalike domains
- 2) Check whether these domains are active or not

dnstwister report for **nike.com** 







We identified **154** domains similar to **nike.com**. **105** domains resolved to an [A or MX record](#).

Protect your business from typosquatting-based phishing attacks and IP infringement!

We'll alert **you** as soon as someone registers a domain similar to **nike.com**.

[Sign up for dnstwister alerts now!](#)

Found (105) Available (49) export: [json](#) [csv](#)

Found Domain	IP Address / A record	MX record?	
nnike.com	 185.53.177.54	✓	analyse
nake.com	 35.186.238.101	✗	analyse
noke.com	 216.239.32.21	✓	analyse
nyke.com	 91.195.240.103	✗	analyse
ike.com	 64.78.234.36	✓	analyse
ñike.com (vn-ike-fma.com)	 192.187.111.219	✓	analyse

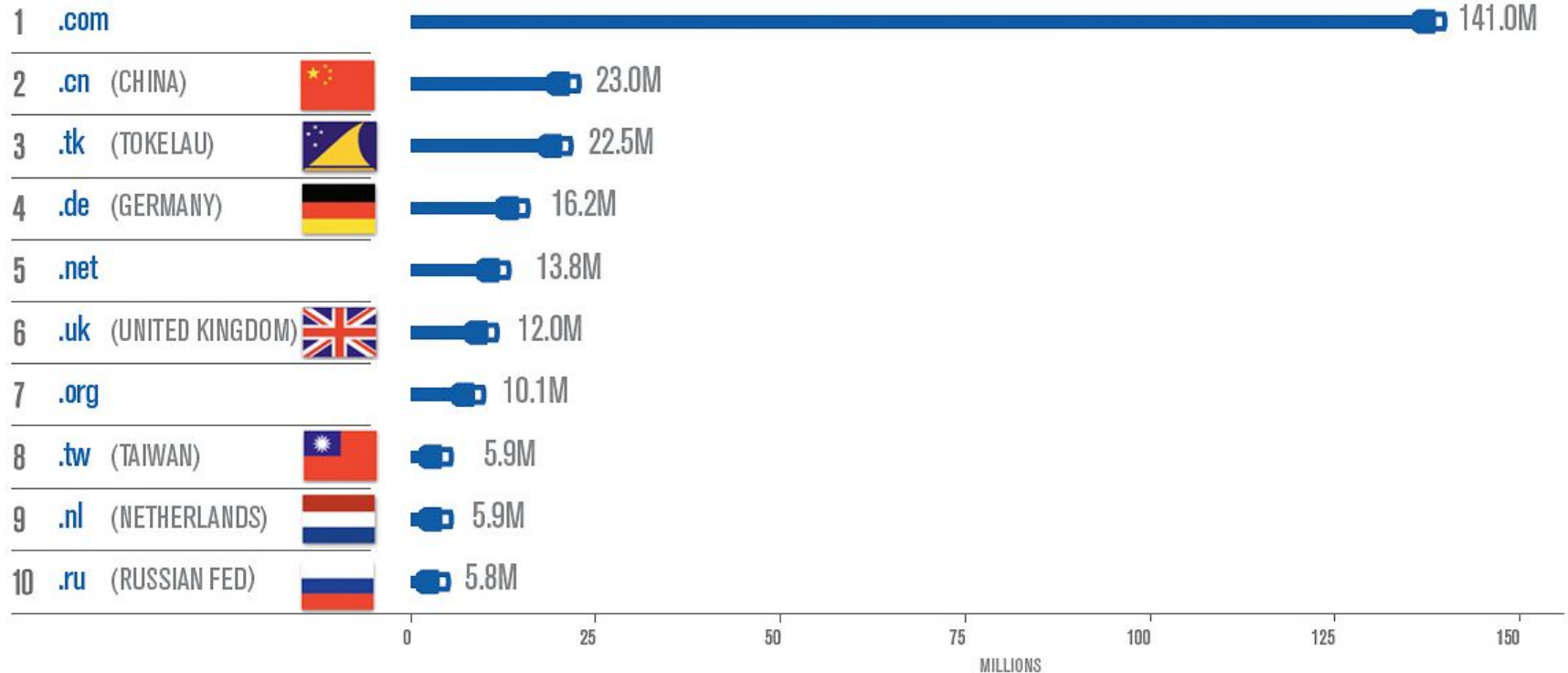
Proactive: Exhaustive approaches

Steps:

- 1) Get (a huge) list of known domain names
- 2) Find domain names that include the original domain name

Challenges

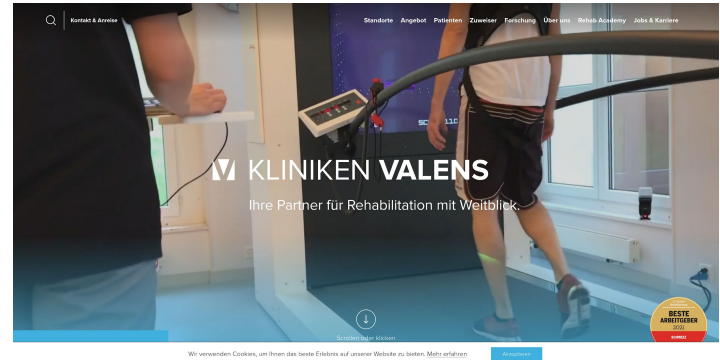
Scale & Performance



Ambiguity

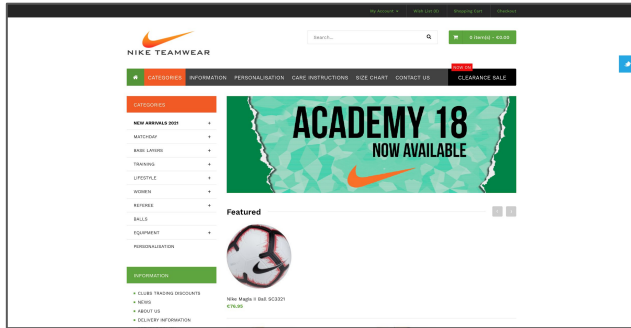


nikeforbusiness.com

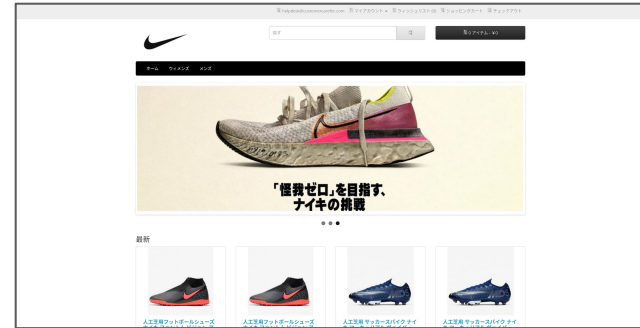


kliniken-valens.ch

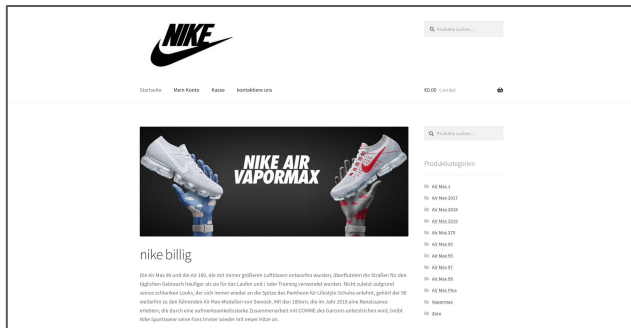
Variety



4niketeamwear.com



jpnikesales.com

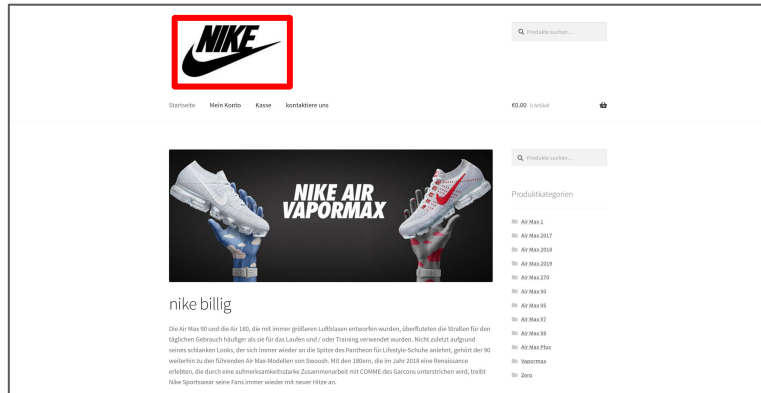
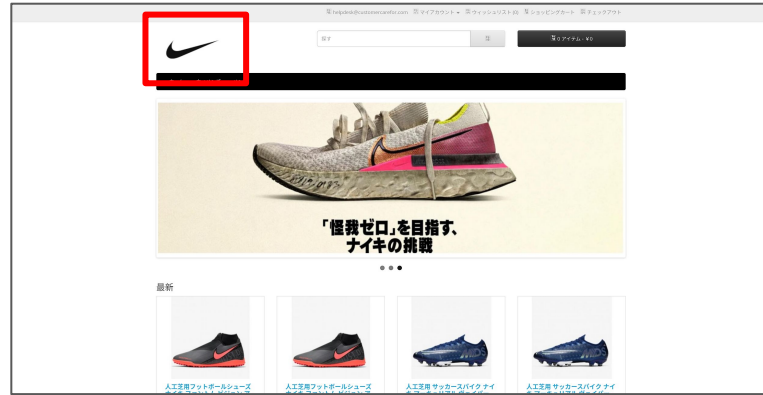
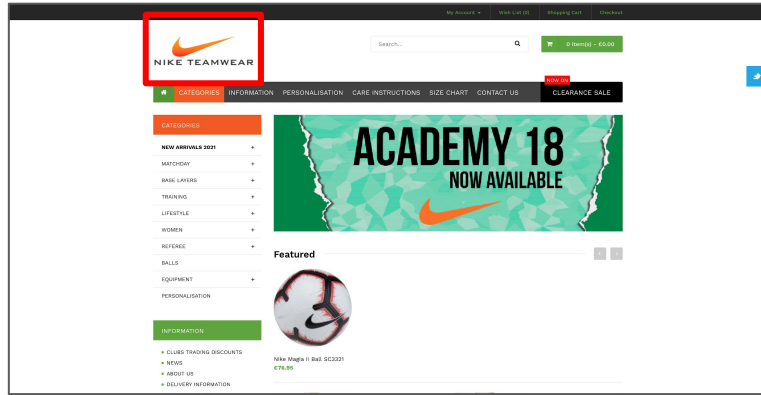


nikebillig.com



nikeforbusiness.com

Proposal: Logo matching



Why logo matching?

Logos are *commonly used* in lookalike domains

Language independence

Proven against phishing scams

Variety of *robust algorithms*

Implementation

Project overview

User provides:

logo images and **domain name**

Our tool does:

- 1) **create** a list of **lookalike domains**
- 2) **crawl** the list and **extract images** for each website
- 3) **evaluate** the website and **try to match logo images**
- 4) **report** the results for manual review

Lookalike detection

Crawling

Evaluation

Input

- known domain list
(**external provider**)
- domain name
(**user**)

Task

Exhaustive search for lookalikes
(*partial string matching*)

Output

- list of lookalike domains

Lookalike detection

Crawling

Evaluation

Input

- list of lookalike domains

Task

Crawl of pages

(with headless Chrome browser)

Output

- images of each page
- other small features
(resolved URL, NS, ...)

```
graph LR; A[Lookalike detection] --> B[Crawling]; B --> C[Evaluation];
```

Lookalike detection

Crawling

Evaluation

Input

- website features
(images)
- set of logos
(user)

Task

Inspection of features and labelling
(*heuristics / image matching*)

Output

- websites with labels
(not reached, affiliate abuse,
parked domain, matching logo,
unknown)

Logo detection



present in?

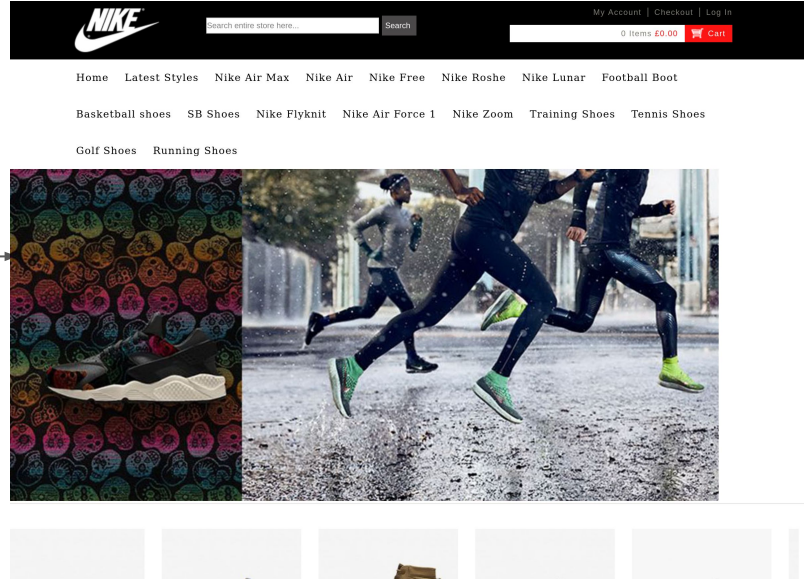


Image hashing

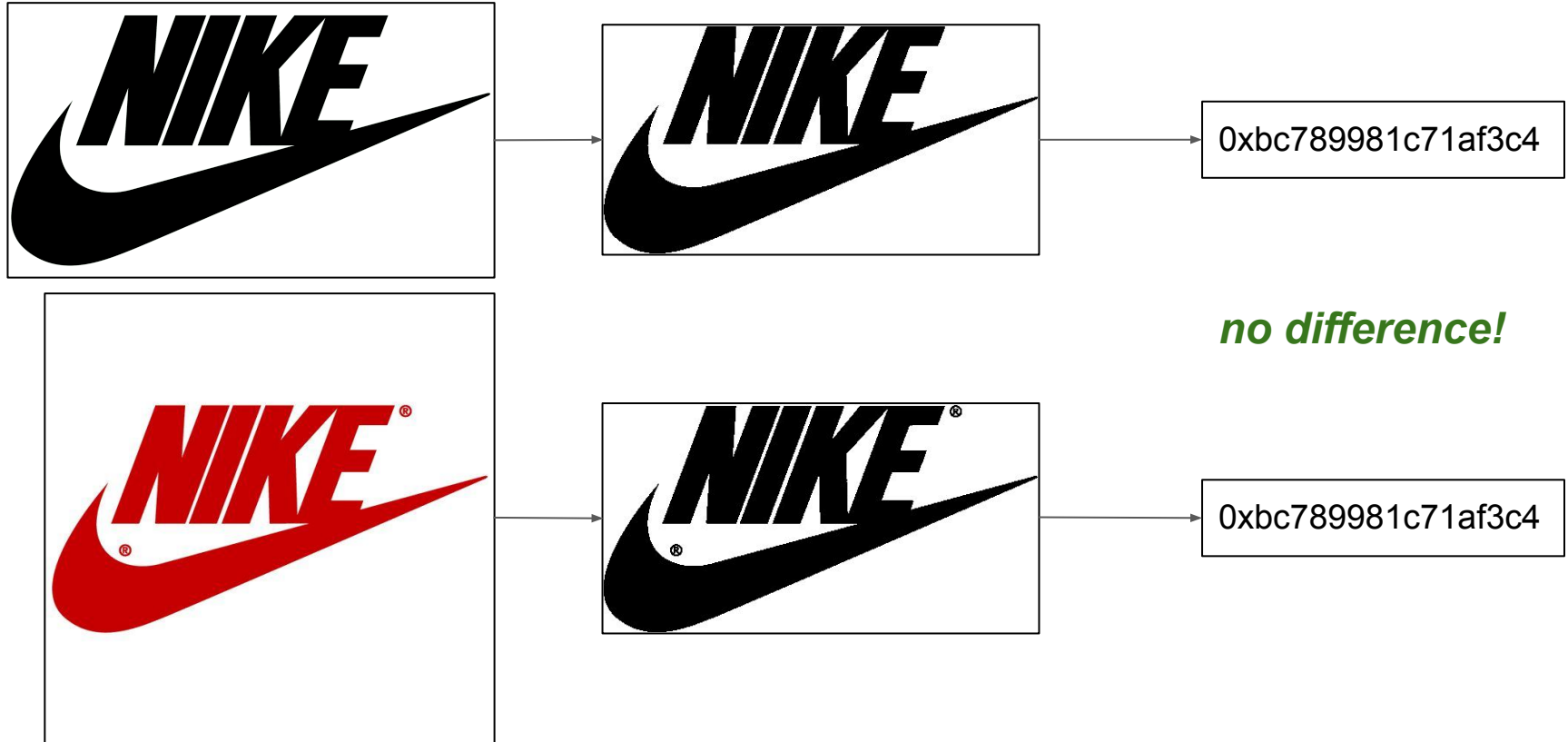


0xbc789981c71af3c4

Perceptual hash (Phash):

1. reduce to small **greyscale image**
2. apply **DCT** (discrete cosine transform)
3. use **low frequencies** for hash
4. **compare hashes** with **Hamming distance** (number of different bits)

Hash comparison with normalization



Case Study

Methodology

> 200 million domains

(acquired from ZoneFiles.io)

Forbes top 100 brand names

(luxury and apparel category)

Forbes top 100 brand logo images

(from Google image search)

Comparison with keyword search as a baseline

Overview

Domain	Total	Unknown	Matching logo	Suspicious
nike.com	31'575	8'565	148	93
gucci.com	5'722	1'871	8	6
hermes.com	9'472	4'443	23	11
adidas.com	4'520	1'105	57	36

Preliminary Results

Example: Nike

Demo

Limitations

Slow and resource demanding

Manual review still required

Matching is imperfect:

- malicious pages without logos
- product images are ignored

Conclusions

Tool finds malicious websites (presumably) unknown to the vendors

Reported websites are relevant

(Nike successfully blocked few)

Image matching not good enough

Commercial security providers offer similar services

Summary

Ambiguity



nikeforbusiness.com



kliniken-valens.ch

- "similar" name not necessarily problematic (especially when considering combosquatting!)
- legitimate use for lookalikes (e.g. increase reach)

12

Lookalike-Detection

Crawling

Evaluation

Input:

- List of lookalike domains
- Original domain

Process:

Crawl page with chrome headless

Output:

- Images per page
- Page Text
- Other small features like resolved url, nameserver

20

Image hashing



0xbc789981c71af3c4

22 bits different!



0xbd78c383627c3961

24

Example: Nike

	nikeoutletcanada.ca	http://www.nikeoutletcanada.ca/	
	nikeoutletonline.us	http://www.nikeoutletonline.us/	
	nikeoutlets.net	http://nikeoutlets.net/	
	nikeprestofly.com	https://www.nikeprestofly.com/	

30