

TASK #1: LOGO SIMILARITY

Match and group websites by the similarity of their logos.

MAESCHI STEFAN-ANDREI



DESCOMPUNEREA PROBLEMEI

EXTRAGEREA LOGO-URILOR

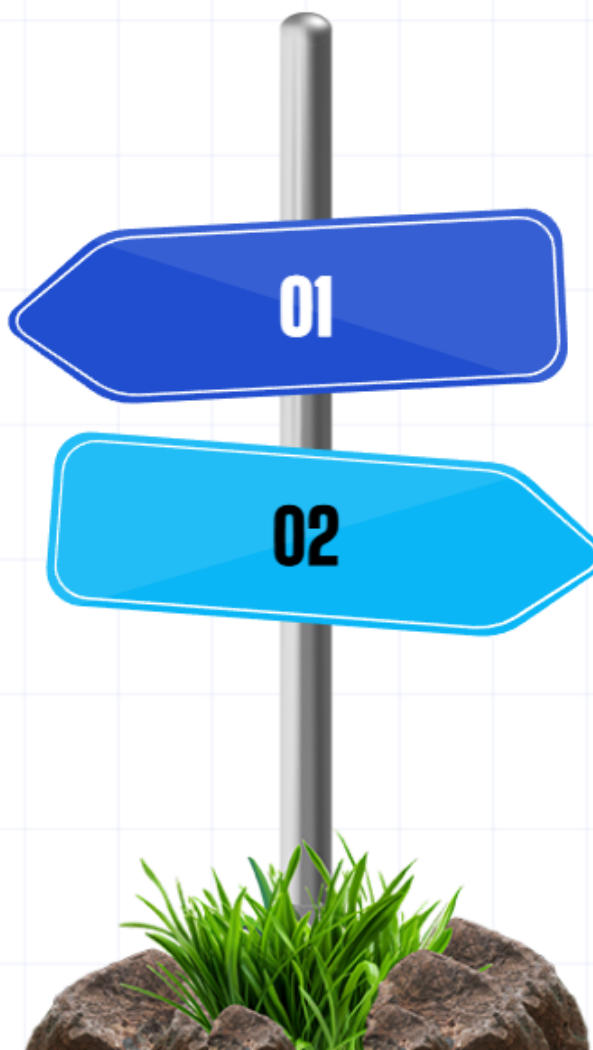
Obținerea unui logo valid pentru cât mai multe domenii – ideal peste 97% rată de succes.

01

02

GRUPAREA

După ce am extras logo-urile, vrem să identificăm site-urile care folosesc același logo sau unul foarte asemănător.



ETAPA 1: EXTRAGEREA LOGO-URILOR – METODE FOLOSITE

01



CLEARBIT API

Logo oficial, calitate înaltă (dacă domeniul e cunoscut).

02



FAVICON.ICO

Fallback clasic, funcționează pe majoritatea site-urilor.

03



GOOGLE S2 FAVICON

Alternativă externă fiabilă (64px PNG)

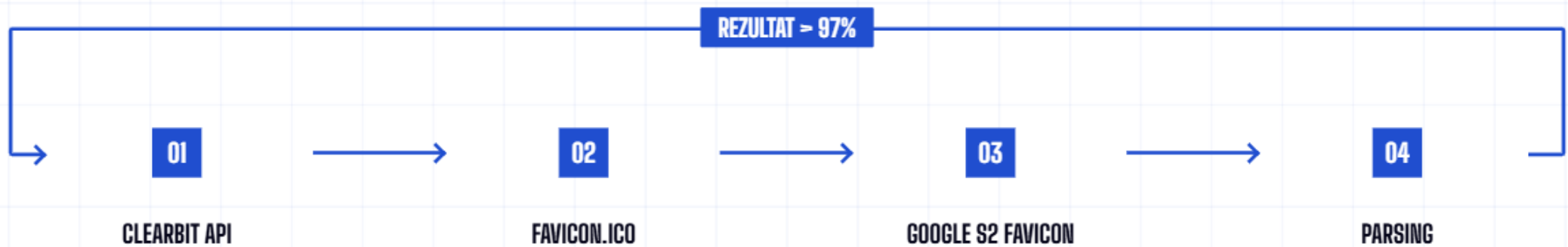
04



PARSARE HTML

Parsăm logo-ul real afișat în pagină

DACĂ O METODĂ EȘUEAZĂ → ÎNCERCĂM URMĂTOAREA.



REZULTAT:

98.07% rata de succes

```
[3411/3416] aamcorandallstownmd.com...  
[3412/3416] shopee.vn...  
[3413/3416] aamcocarync.com...  
[3414/3416] aliaxis.pl...  
[3415/3416] ccusa.co.za...  
[3416/3416] aamcolawrencevillega.com..
```

```
Logo-uri extrase: 3350/3416 (98.07%)
```

ETAPA 2: GRUPAREA LOGO-URILOR

- Compararea pixel-cu-pixel nu e fiabilă (diferențe de dimensiune, format, culoare).
- Soluția: **transformăm imaginea într-o amprentă numerică (pHash)** care reflectă structura vizuală.
- Două logo-uri care „arată la fel” vor avea hash-uri apropiate.

TEHNOLOGIA FOLOSITĂ:

- PHASH (PERCEPTUAL HASH) GENEREAZĂ O SEMNĂTURĂ DE 64 BIȚI PENTRU FIECARE IMAGINE.
- DISTANȚĂ HAMMING ÎNTRE HASH-URI = NUMĂR DE BIȚI DIFERIȚI.
- CONSIDERĂM DOUĂ LOGO-URI SIMILARE DACĂ DIFERĂ PRIN ≤ 10 BIȚI.

GRUPAREA:



- COMPARAM FIECARE LOGO CU TOATE CELELALTE ($O(N^2)$) PENTRU SETURI MICI).
- LOGO-URILE CU HASH-URI „APROPIATE” SUNT GRUPATE ÎMPREUNĂ.
- FIECARE GRUP REZULTĂ ÎN: O LISTĂ DE DOMENII CU LOGO-URI VIZUAL SIMILARE.

REZULTAT:

Grup 1:

- stanbicbank.co.zw

Grup 2:

- astrazeneca.ua
- astrazeneca.no
- astrazeneca.dk
- astrazeneca.ca
- astrazeneca.pt

Grup 3:

- autosecuritas-ct-seysses.fr

Grup 4:

- ovb.ro
- ovb.at
- ovb.gr

Grup 5:

- mazda-autohaus-hellwig-hoyerswerda.de
- mazda-autohaus-kaschmieder-waren.de
- mazda-autohaus-lackmann-sendenhorst.de
- mazda-autohaus-roessel-birkenau.de
- mazda-autohaus-henseler-schwedt.de
- mazda-autohaus-hansmann-kassel.de
- mazda-autohaus-becker-homburg-bruchhof.de
- mazda-autohaus-tabbouch-speyer.de
- mazda-autohaus-haese-erkrath-hochdahl.de
- mazda-autohaus-banaszak-kleve.de
- mazda-autohaus-elstner-jena.de
- mazda-autohaus-dorner-riedlingen.de
- mazda-ma.com