

Smrt datotekam in mapam

#171 **wwhmb**: Spletne minute se vračajo (+ pred-gostje)
April, 2026

Robert Sedovšek
<https://galjot.si>

Razbijanje ledu...

Klasični hobiji

- Tek in kolo poleti
- Tek na smučeh pozimi
- Fotografija čez celo leto

Malo manj klasično

2x pretekel 100 km in 5500 m višincev

Delo

Celtra →

Turtl →

~_(\ツ)_/~ → TBD

PhD

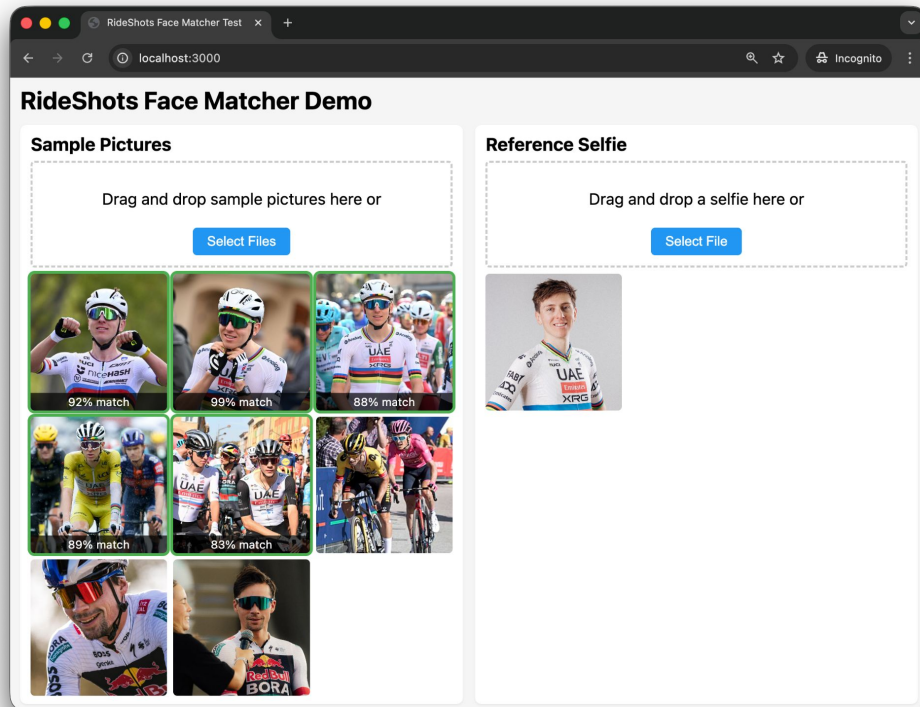
- Grafična tehnologija
 - Uporabniški vmesniki
 - Uporabniška izkušnja
- Naravoslovnotehniška fakulteta, Ljubljana

Kaj počnem? 1/2

RideShots.cc (ljubiteljski projekt)

Izziv: *facial recognition occlusion.*

sistem, ki prepozna ljudi četudi nosijo očala ali čelado

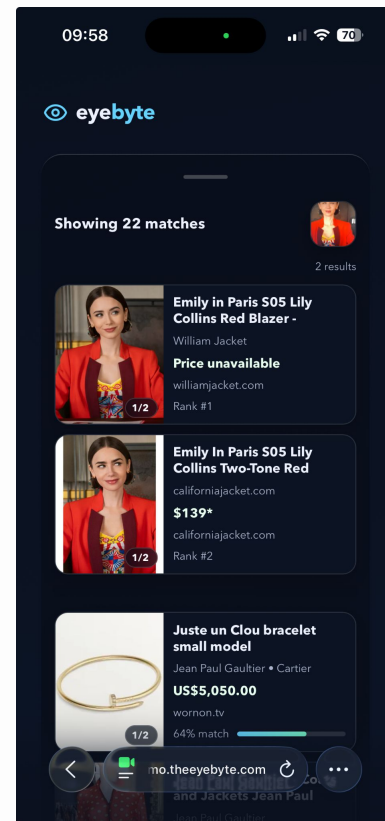
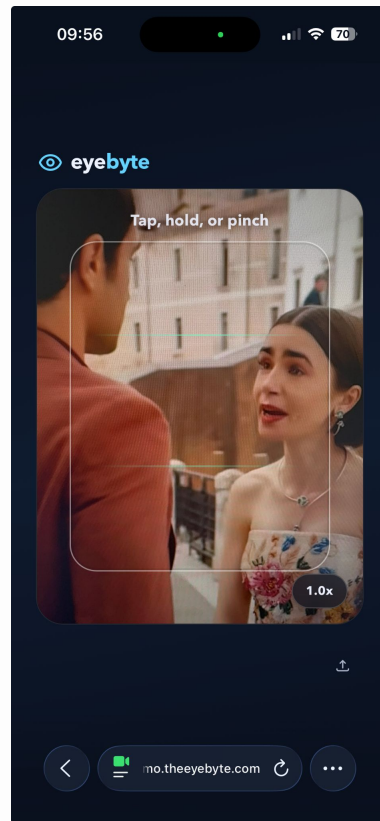


Kaj počnem? 2/2

The Eyebyte (*visual e-commerce*)

Izziv: *object detection & visual similarity*

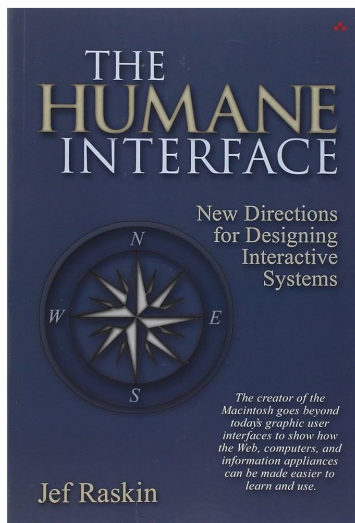
sistem, ki prepozna oblačila in najde
bodisi enak ali podoben izdelek



"The content of a file is its own best name." – **Jef Raskin**

Izvor

- **Knjiga:** The Humane Interface
- **Avtor:** Jef Raskin
- **Leto:** 2000



O avtorju

- Pionir uporabniških vmesnikov
- Začetnik projekta **Apple Macintosh**
- Zagovornik enostavnih, uporabniku prijaznih računalnikov
- Ogromen vpliv na HCI
(*human-computer interaction*)

Problem datotek








- financial_report_final_v3_last.pdf
- Financial-report-final.pdf
- financial report_final-v1.pdf
- fnacialreport.pdf

Iskanje datotek

Na primeru: iskanja obrazca za lansko napoved za odmero dohodnine od dividend (Doh-Div)

🔍 tax | dividends

Notes Messages Podcasts

-  Tax on dividends
gov.uk > tax on dividends
-  Dividend tax
en.wikipedia.org > wiki > Dividend tax
-  Dividend Tax Rate for 2025 and 2026
smartasset.com > taxes > dividend tax rate
-  ▲ HTD
John Hancock Tax-Advantaged Dividend Income Fund **25.37**
+0,32 (1,29%)
-  Are Operating Systems stuck? — Blog · iCloud
08:23 Are Operating Systems stuck? More than a decade ago I read a though-pr...
-  Andrej Mihajlev 31. 5. 24
My main question if I can pay dividends and deduct that from corporate tax
-  a16z Podcast: When Bio Meets Computer Science

Searching "This Mac"

This Mac "Pictures"

🔍 tax 2025

FileNames
Name Contains "tax 2025"

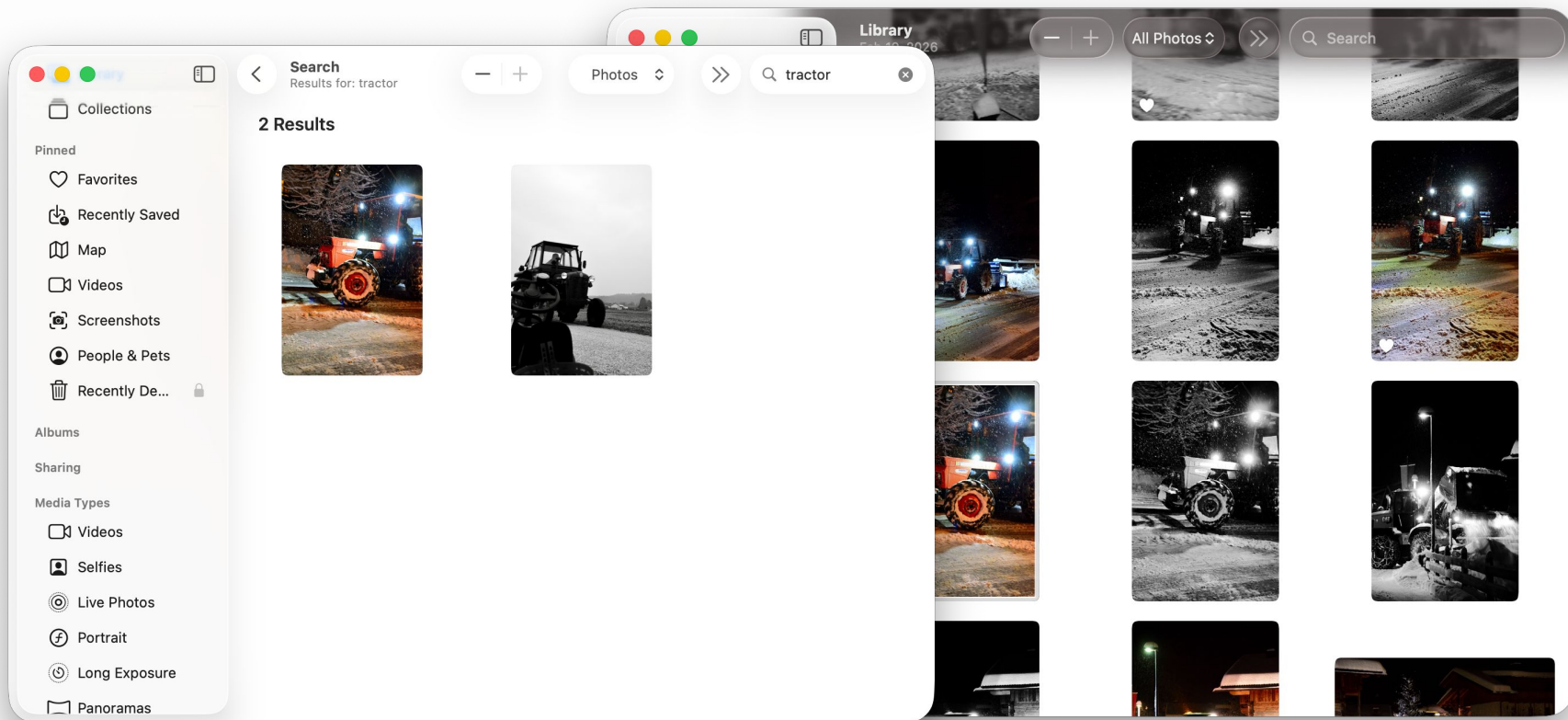
Content
Contains "tax 2025"

models_0.d.ts models_0.d.ts

transcript.log transcript.log

WebhookEndpointResource.d.ts

Iskanje fotografij



Jefove ideje v praksi...

- ✓ Iskanje po dejanski vsebini, ne po imenu
(npr. *garancija_v2.txt*)
 - ✓ Pametni kontekst: (npr. *kar sem delal včeraj*)
-

- ✗ Povečevalni uporabniški vmesnik (zoom UI)
- ✗ Samodejno shranjevanje in neskončno razveljavljanje (*undo*)

- ✓ del današnjega predavanja
- ✗ mogoče naslednjič

Tehnologija?

- Večmodalni “*embedding*”
- Meta-podatki
- (Hibridni) iskalnik

1. Večmodalni “ embedding” (multimodal embedding)

Naloga

Pretvoriti katerikoli dokument (tekst, slike, video, audio) v numerični vektor, ki predstavlja njegove značilke

Zakaj?

Omogoča iskanje ter primerjavo med vektorji ne glede na tip podatka



Tehnologija

Možnosti:

- Amazon Nova Embedding (Nov '25)
- Googlov Gemini Embedding 2 (Mar '26)

Hramba vektorjev:

- Amazon S3 Vectors
- Meta's FAISS
- QDANT
- ...

```
# Večmodalni embedding
1 for file in files:
2     bytes = read(file)
3
4     vector = AmazonNova.embed({
5         taskType: "SINGLE_EMBEDDING",
6         embeddingPurpose: "GENERIC_INDEX",
7         input: {
8             type: "image",
9             format: "jpeg",
10            bytes: base64(bytes)
11        }
12    })
13
14    save({
15        id: hash(file.path),
16        path: file.path,
17        vector: vector
18    })
19
```

2. Meta podatki

Naloga

Shraniti strukturirane podatke o podatkih
(čas, vir, lokacija, dolžina)

Zakaj je to pomembno?

Dopolnjujejo "vektorje" z dodatnimi
značilkami, kot so čas, lokacija, naprava.
Omogoča filtriranje in natančnejše iskanje.



Tehnologija

Primeri:

“Rdeč avto v Mariboru” → potrebuje lokacijo

“Rdeč avto oktobra 2025” → potrebuje čas

Možnosti:

- Strukturirane podatkovne baze

*Dejstva izvlečemo in jih shranimo ločeno od vektorjev,
tako da jih lahko uporabimo kot natančne filtre.*

```
# Meta podatki
1 for file in files:
2     exif = extractExif(file)
3     gps = exif.gps
4
5     metadata = {
6         camera: normalizeCamera(exif.make, exif.model),
7         date: exif.createdAt,
8         year: year(exif.createdAt),
9         location: reverseGeocode(gps)
10    }
11
12    saveMetadata({
13        id: hash(file.path),
14        path: file.path,
15        ..metadata
16    })
17
```

3. Hibridno iskanje

Naloga

Hibridno iskanje po pomenu (semantično iskanje) in metapodatkih (faktih)

Zakaj je to pomembno?

- Pomen / semantika najde kaj iščemo
npr. "rdeč avto"
- Metapodatki filtrirajo glede na kontekst
npr. "v Mariboru" ali "leta 2025"



Tehnologija

Kako deluje?

Iskalni niz: "Rdeč avto v Mariboru"

Semantični del: "Rdeč avto" // vektorji

Dejstva: { lokacija: Maribor }

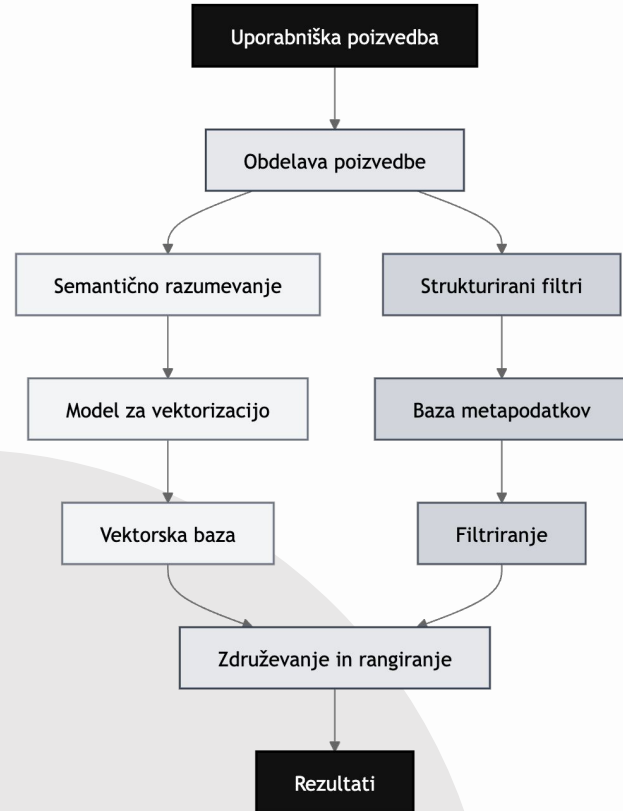
...

Združevanje rezultatov

```
# Hibridni iskalnik
1 function hybridSearch(query):
2   parsed = parseQuery(query)
3   // parsed = { text, camera, location, year, since }
4
5   candidates = filterMetadata(allFiles, {
6     camera: parsed.camera,
7     location: parsed.location,
8     year: parsed.year,
9     since: parsed.since
10  })
11
12  if parsed.text is empty:
13    return sortByDate(candidates)
14
15  queryVector = AmazonNova.embed({
16    embeddingPurpose: "GENERIC_RETRIEVAL",
17    input: { type: "text", value: parsed.text }
18  })
19
20  return rankBySimilarity(candidates, queryVector)
21
```

Povzetek

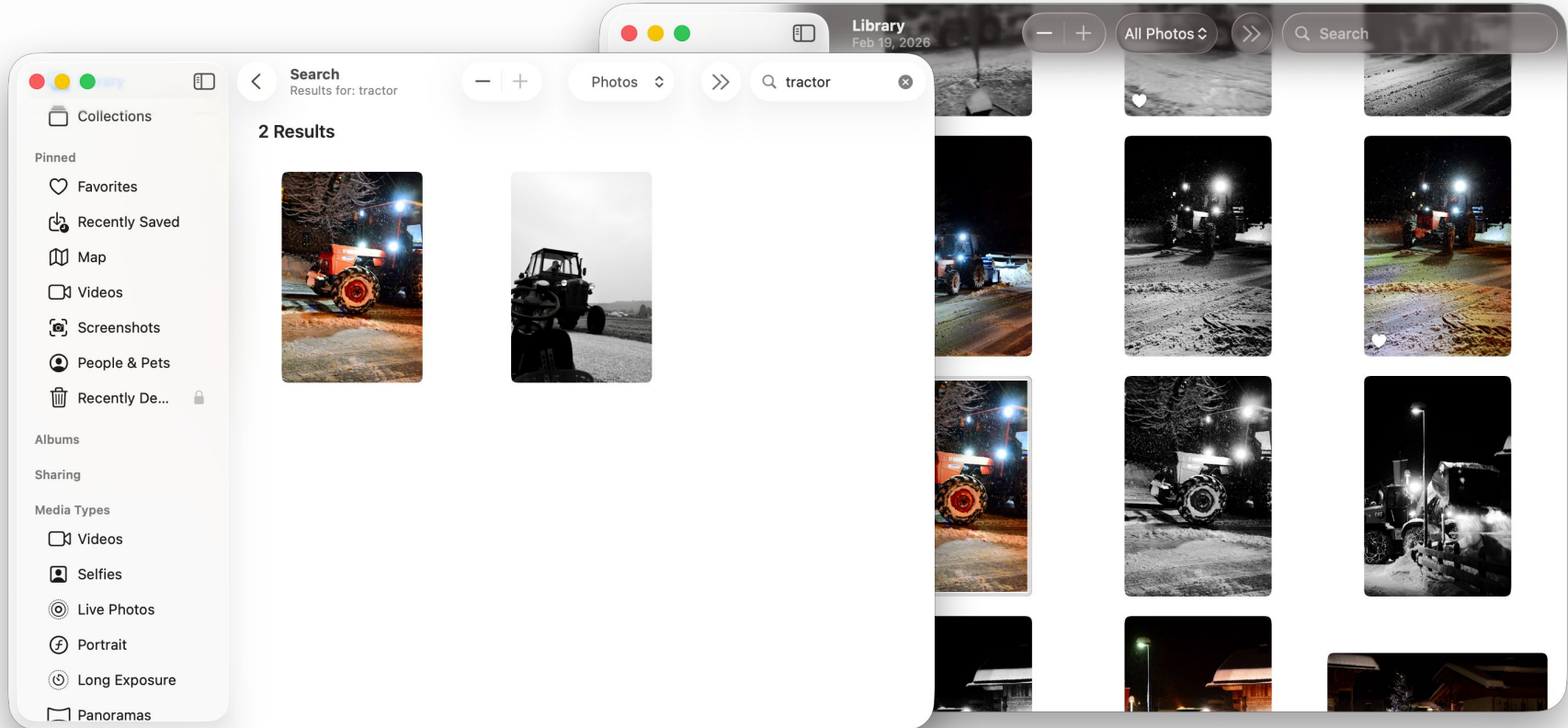
Tehnološki diagram

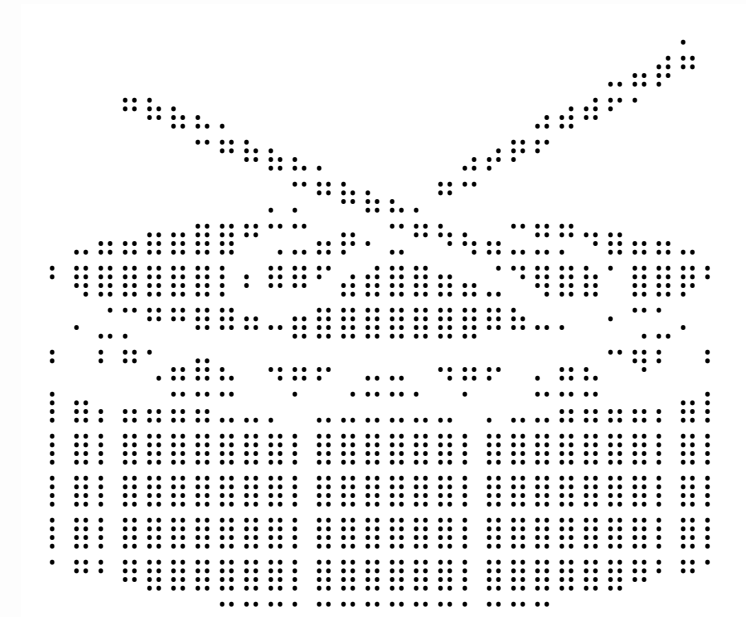




Reizultat?

Še spomin na problem...





tractors taken with "Nikon d750" from Italy

Search

Matches for "tractors taken with "Nikon d750" from Italy"

20 ranked results from 20 filtered photos. Semantic query: "tractors taken"

CAMERA: "Nikon d750" LOCATION: Italy



IMAGE | 4016X6016 | SCORE 0.487 CAMERA NIKON CORPORATION LOCATION Sexten - Sesto, Italy



IMAGE | 4016X6016 | SCORE 0.411 CAMERA NIKON CORPORATION LOCATION Sexten - Sesto, Italy



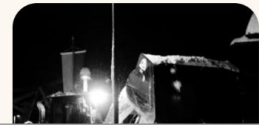
IMAGE | 4016X6016 | SCORE 0.408 CAMERA NIKON CORPORATION LOCATION Sexten - Sesto, Italy



IMAGE | 3860X5783 | SCORE 0.369 CAMERA NIKON CORPORATION LOCATION Sexten - Sesto, Italy



IMAGE | 4016X6016 | SCORE 0.365 CAMERA NIKON CORPORATION LOCATION Sexten - Sesto, Italy



Zaključek

Hvala za pozornost!

[5 min. za vprašanje in nove ideje]

@sedovsek on X · [Instagram](#) · [Threads](#) · [Strava](#)
<https://galjot.si>

