

# **Elektronické informační zdroje**

Obrazové a auditivní EIZ

22. 11. 2019

# Jak vyhledáváte obrázky?



# Sbírky obrázků jako EIZ

- jak vznikají?
- co všechno se musí udělat?
- jak se s nimi pak pracuje?
- jak se v nich vyhledává?



# Případovka | Europeana

- velká evropská digitální knihovna
- obsahem hlavně obrazy, umělecká díla, filmy
- sbírá data z mnoha paměťových institucí
- projekt založen v roce 2008
- jak se tam objekty dostanou?
- jak se s nimi dál pracuje?
- jak se prohledávají?



# Jak vznikají sbírky obrázků?

- digitalizace / born-digital
- formáty, úpravy,...
- tvorba metadat
- výběr platformy
- zpřístupnění



# Formáty obrazu

- klasické formáty a jejich vlastnosti
  - my se setkáváme i s dalšími formáty
  - RAW – proprietární
  - DNG – digitální negativ, otevřený
  - TIFF – tagged image, ISO norma, trvalost
- 
- <https://www.loc.gov/pictures/collection/bbc/>

# Jak se s nimi dál pracuje?

- tvorba kolekcí a kurátorství
- <http://www.loc.gov/pictures/>
- Europeana
  
- význam pro digital humanities
- <http://lab.softwarestudies.com/p/imageplot.html>

# Jak vyhledávat obrázky?

- kolekce obrazů rostou neskutečným tempem
- nutnost v nich vyhledávat roste s nimi
- celá oblast hnána snahou o automatizaci
- automaticky rozpoznat co/kdo je na obrazu
- mnoho rovin využití...
- *image retrieval (ImR)*



# Jak vyhledávat obrázky?

- nutné přidat metadata
- jaké mají obrázky metadata?
- interní/externí
- *EXIF*
- *alt*

## **EXIF**

<http://exif.regex.info/exif.cgi>

*exchangeable image file format*

specifikace metadat pro uložení do obrazu

# Ruční tvorba metadat

- ruční tvorba různých metadat
- značně náročná činnost
- *kdo to provádí?*

# Ruční tvorba metadat

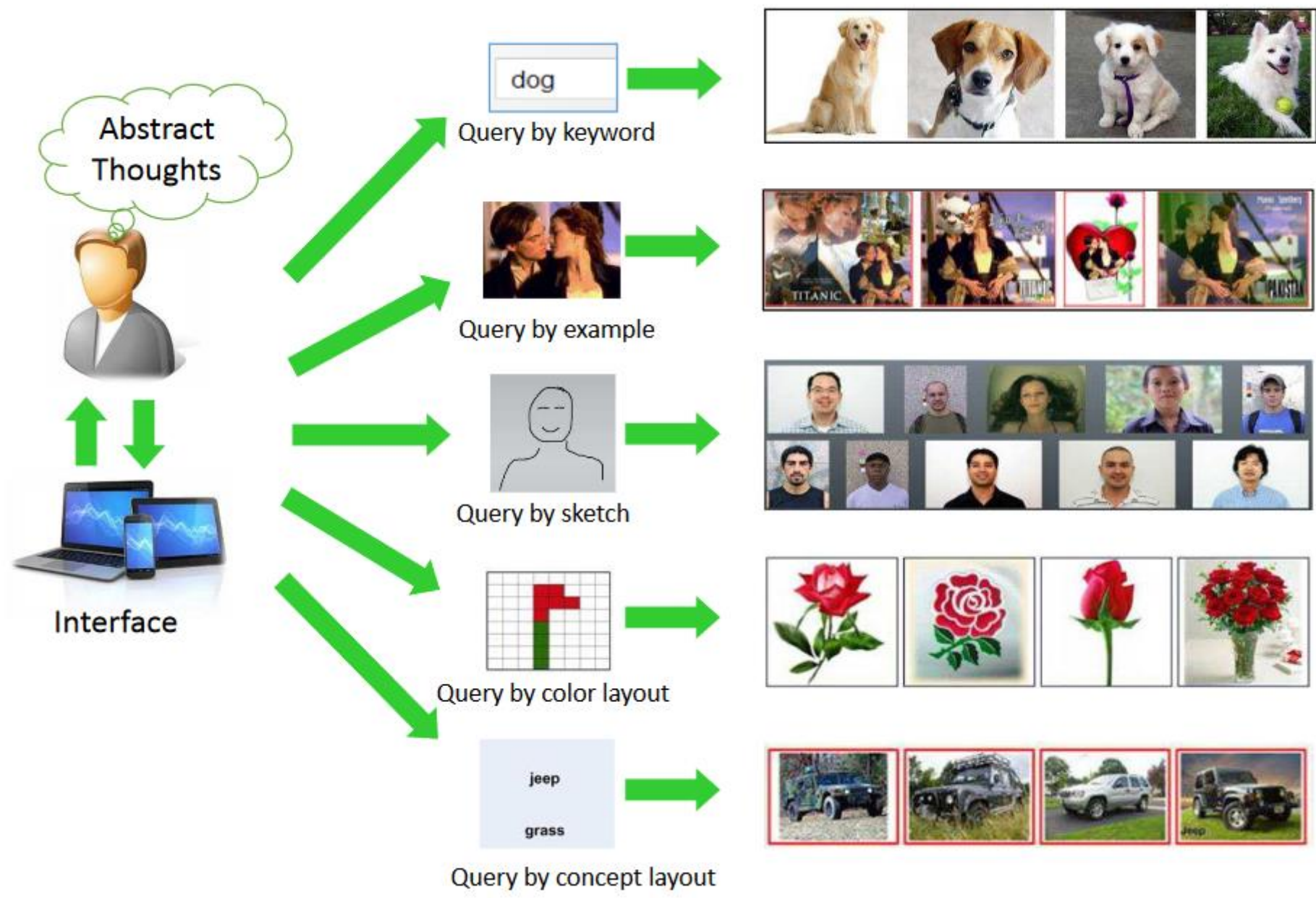
- ruční tvorba různých metadat
- značně náročná činnost
- *kdo to provádí?*
- archivy, muzea, knihovny, galerie
- komerční subjekty, které potřebují přesnost
- většinou kombinují ruční a automatizované přístupy
- <https://www.gettyimages.com>

# Automatická tvorba metadat

- různé úrovně složitosti
- *content-based ImR* – vyhledávání založené na obsahu
- automatické rozpoznávání obrazu
- počítačové vidění

# Content-based ImR

- indexování mimo textové charakteristiky
- prvním krokem je identifikace těchto charakteristik
- poté hledání způsobů, jak je zachytit/popsat
- co může charakterizovat obrázek kromě txt metadat?
- např. barva, textura, tvar...



# Vyhledávání barvou

- implementováno v různých systémech
- využívá se často u digitalizovaných sbírek
- používá např. Europeana

- <https://services.tineye.com/MulticolorEngine>
- <https://labs.tineye.com/multicolr/>
- <https://artsexperiments.withgoogle.com/artpalette/>



# Vyhledávání příkladem

- Google Image Search – reverzní vyhledávání
- TinEye – reverzní vyhledávání
- <https://lens.google.com/>
  
- jiná podoba vyhledávání příkladem:
- <https://www.shutterstock.com/showcase/compositionsearch/>
- <https://www.shutterstock.com/showcase/copyspace/>



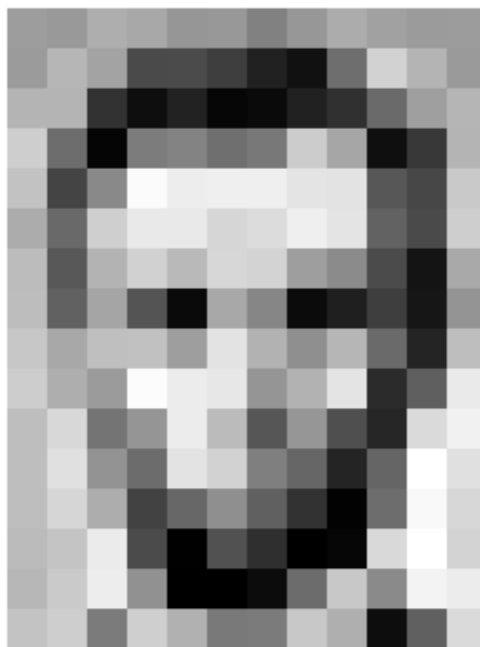
# Vyhledávání nákresem

- <https://quickdraw.withgoogle.com/#>



# Automatické rozpoznávání obrázků

- jak naučit počítač rozpoznat, co je na obrázku?
- počítačové vidění



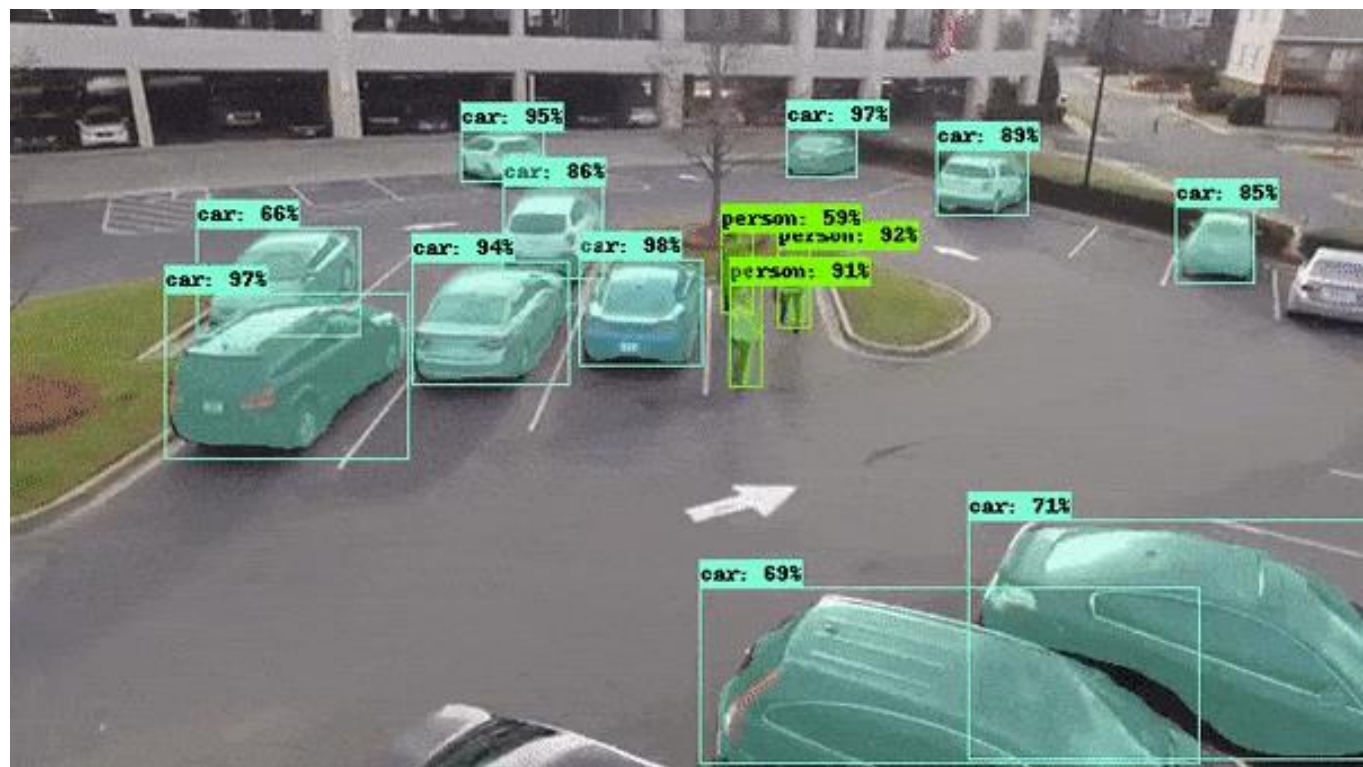
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	80	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	106	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	80	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218



# Automatické rozpoznávání obrazu

- neuronové sítě
- *vstup* – obrázky
- *výstup* – klasifikace, s určitou pravděpodobností
- všude, od taggování na FB po autonomní auta
- *supervised learning*
- *unsupervised learning*





..... tohle je **lama**



..... tohle je **lama**



..... tohle je **lama**



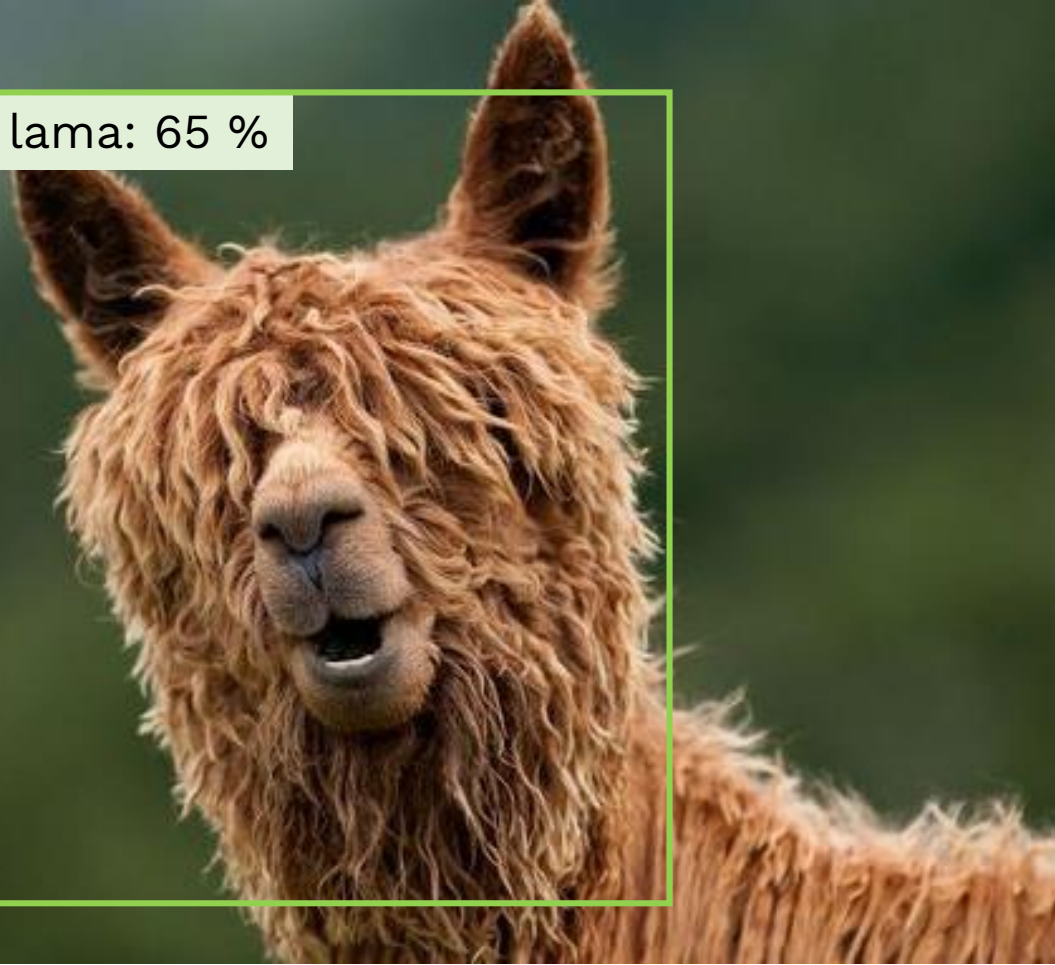
všechno je **lama**



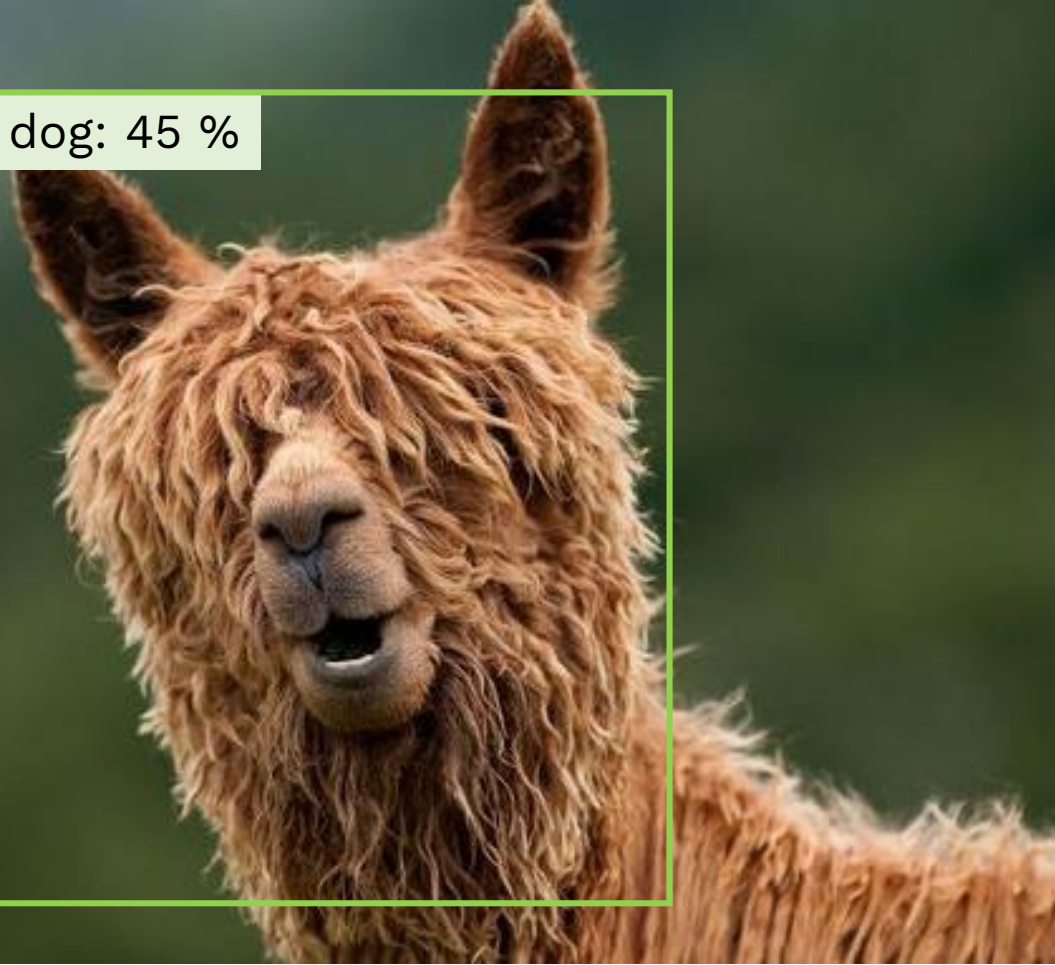
co je tohle?



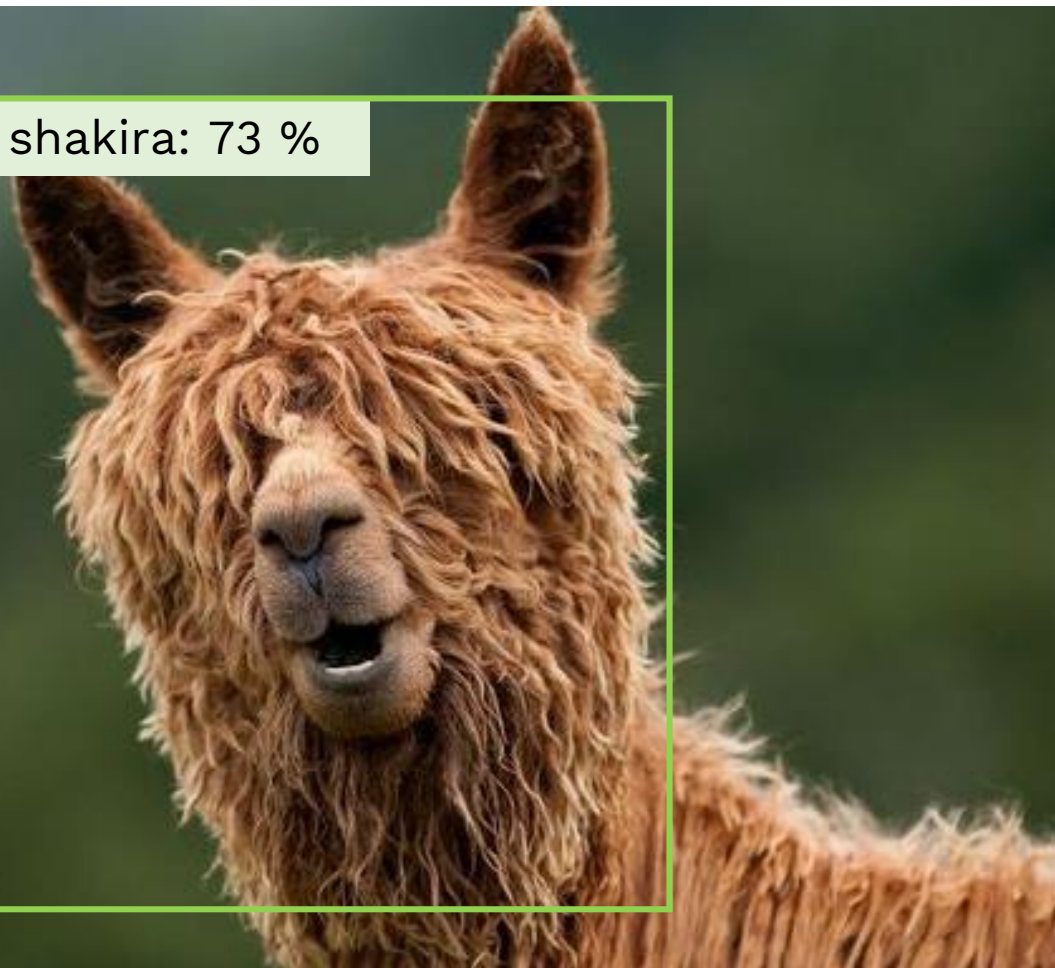
lama: 65 %



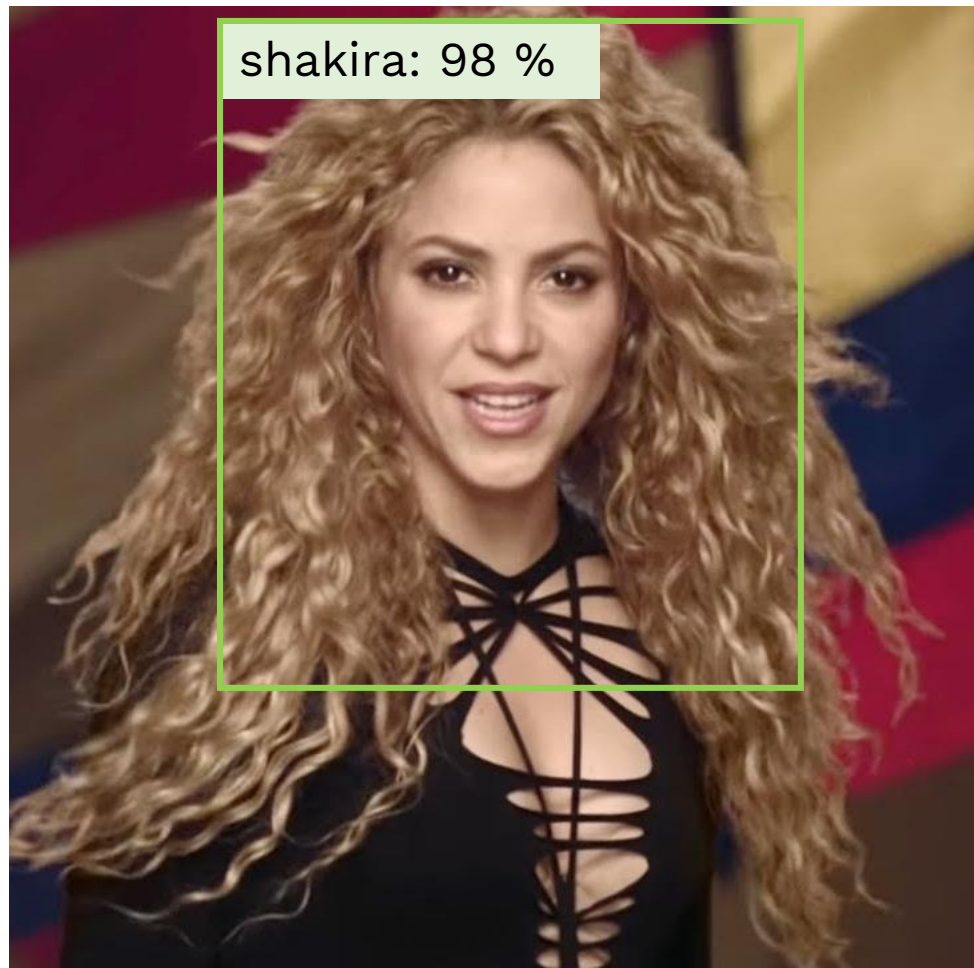
dog: 45 %

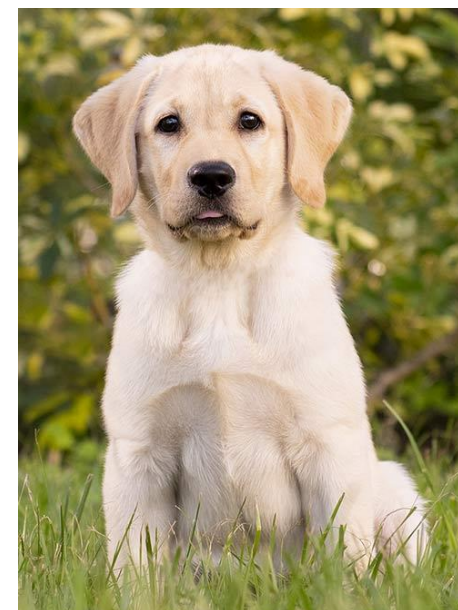
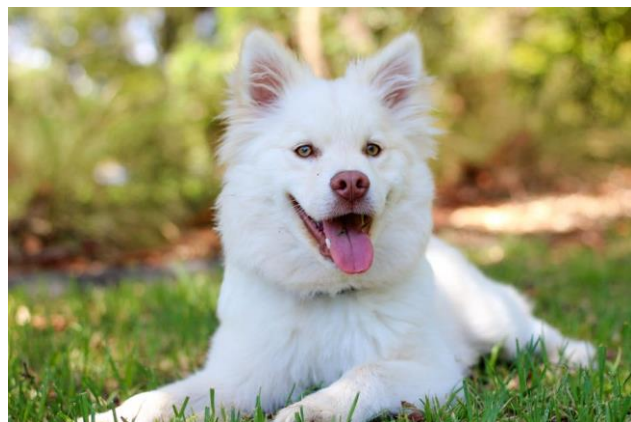


shakira: 73 %



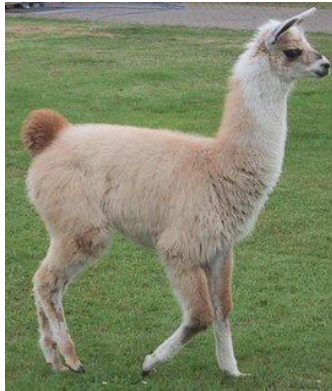
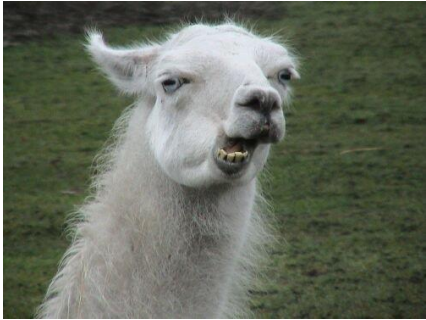
shakira: 98 %



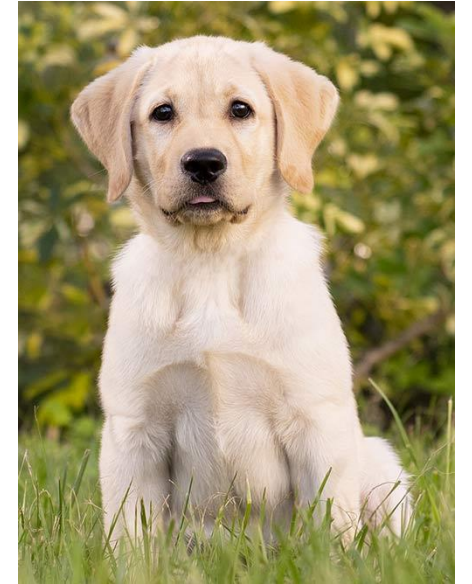


na, tady máš **obrázky...**

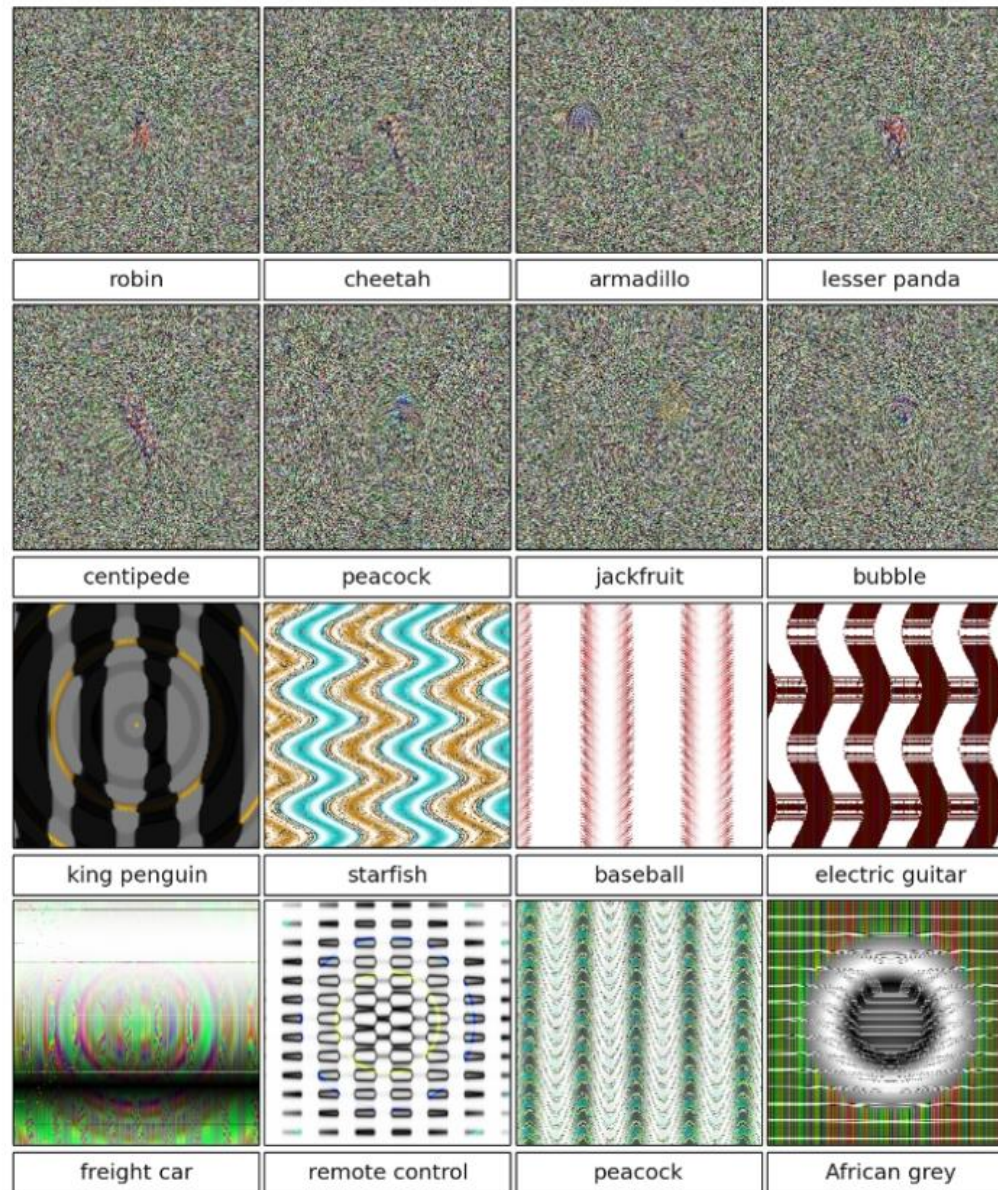
group\_A



group\_B



<https://arxiv.org/pdf/1412.1897.pdf>



# Automatické rozpoznávání obrazu

- PowerPoint
- <https://www.clarifai.com/demo>
- <https://cloud.google.com/vision/>

Příklady dalšího využití:

- <https://thing-translator.appspot.com/>
- <https://teachablemachine.withgoogle.com/>
- kde všude se to dá využít?





# Specifické způsoby procházení

- *image collection exploration*
- alternativa přímému vyhledávání
- [http://phototour.cs.washington.edu/Photo\\_Tourism.pdf](http://phototour.cs.washington.edu/Photo_Tourism.pdf)
- <https://artimageexplorationspace.com/>
- [Zahrada v interaktivní podobě](#)
- [Přístup ke sbírkám skrze selfie](#)

# Případovka | Archivy NYPL

- [Vstupní web do kolekcí NYPL](#)
- [Obrazové dokumenty ve veřejné doméně](#)
- [Pátá Avenue tehdy a dnes](#)
- [Vizualizace dokumentů ve veřejné doméně](#)
- [Speciální péče o stereografii](#)
- [Mansion Maniac nad volně dostupnými plány](#)

# Zvuk jako EIZ

- jak se tvoří sbírka zvukových dokumentů?
- jak se s nimi pracuje?
- jak se v nich dá vyhledávat?
- jaká zvukové EIZ v knihovnách najdeme?

# Tvorba sbírek zvuků

- digitalizace
- čištění
- metadata

# Práce se sbírkami zvuku

- vyhledávat můžeme podle tradičních metadat
- *jde to ještě i jinak?*

# Práce se sbírkami zvuku

- vyhledávat můžeme podle tradičních metadat
- *jde to ještě i jinak?*
- *audio retrieval X content-based audio retrieval*
- mnoho oblastí a scénářů využití
- vyhledávání hudby
- rozpoznávání řeči
- zvukový dohled (<https://www.acrcloud.com/>)

# Vyhledávání zvuků zvukem

- vyhledávání příkladem
- <https://www.shazam.com/cs> - *otisky*
- vyhledávání vokalizací
- <https://www.midomi.com/>

<http://www2.ece.rochester.edu/projects/air/projects/audiosearch.html>

<https://arxiv.org/pdf/1907.08661.pdf>

# Zvukové EIZ v knihovnách

- hudební databáze (bibliografické i primární dokumenty)
- databáze zvuků
- audioknihy
- hudební CD (???)
- streamovací služby (Naxos, Spotify)
- podcasty



# Hudební databáze

- hudba a mluvené slovo
- komerční
- nekomerční
  
- <https://musicbrainz.org/>
- <https://www.discogs.com/>
  
- Naxos, Spotify,...



# Databáze zvuků

- komerční
- nekomerční
- <https://freesound.org/>

# Audioknihy v knihovnách

- co frčí na trhu víc – mp3 nebo CDčka?
- jak je to s půjčováním audioknih?
- kolektivní smlouva
- absenčně 9 měsíců embargo

- <https://duha.mzk.cz/clanky/pravni-poradna-pujcovani-audioknih>

# Podcasty

- [Dewey Decibel](#) (ALA)
- [The Librarian Is In](#) (NYPL)
- [Librarians Aloud](#)
- <https://circulatingideas.com/>
- <http://www.bannedlibrary.com/>
- <https://librarianswithlives.wordpress.com/>