

# Nagyházfeladatok

**Ajánlórendszer:** felhasználó-film adatbázis

**Szövegbányászat:** Magyar Hírlap 2002-es minta

**Képkeresés:** 1500 képi minta

**Képi klasszifikáció:** lsd előző

**Szociális hálózatok felderítése:** egy “kisebb” webes hálózat

# Ajánlórendszer

Felhasználó-film adatbázis (valós train és test adat)

- 0.66 millió ismert értékelés, min 20 értékelés/személy
- ezen kívül ismert a:

Felhasználók:

- életkora
- foglalkozása
- neme
- postai irányítószáma

Filmek:

- címe
- bemutatójuk éve
- műfaja(i)
- kategóriá(i)

# Ajánlórendszer

- kor:

- \* 1: "18 alatti"
- \* 18: "18-24"
- \* 25: "25-34"
- \* 35: "35-44"
- \* 45: "45-49"
- \* 50: "50-55"
- \* 56: "56+"

- Lehetséges foglalkozások:

- \* 0: "egyeb" vagy nem ismert
- \* 1: "szellemi szabadfoglalkozasu"
- \* 2: "muvesz"
- \* 3: "pap/ugyintezo"
- \* 4: "egyetemi/foiskolai hallgato"
- \* 5: "ugyfelszolgatos"
- \* 6: "orvos/apolo"
- \* 7: "kozepvezeto"
- \* 8: "agrarium"
- \* 9: "haztartasbeli"
- \* 10: "altalanos/gimnaziumi diak"
- \* 11: "jogasz"
- \* 12: "szoftverfejleszto"
- \* 13: "nyugdijas"
- \* 14: "sales/marketing"
- \* 15: "kutato"
- \* 16: "onallo vallakozo"
- \* 17: "mernok/kivitelezo"
- \* 18: "kereskedo"
- \* 19: "munkanelkuli"
- \* 20: "iro"

- Lehetséges mufajok/kategoriak:

- \* Action
- \* Adventure
- \* Animation
- \* Children's
- \* Comedy
- \* Crime
- \* Documentary
- \* Drama
- \* Fantasy
- \* Film-Noir
- \* Horror
- \* Musical
- \* Mystery
- \* Romance
- \* Sci-Fi
- \* Thriller
- \* War
- \* Western

Feladat:

Ismert **666806 értékelés**

Ismert minden felhasználó és film adatai

**333403 felhasználó-film**

pároshoz szeretnénk

**megjósolni** az adott felhasználó értékelését az adott filmre

Eszközők:

- NN algoritmusok, hasonló emberek hasonlóan vélekednek egy filmről
- mátrix alapú módszerek

Kiindulási cikk:

[Recommender Systems](#)

Prem Melville, Vikas Sindhwani

# Szövegbányászat

## Magyar Hírlap 2002-es hírközlései

### - 49530 dokumentum

```
<DOC>
<DOCNO>MH-20020102-001</DOCNO>
<DOCID>MH-20020102-001</DOCID>
<COLUMN>Topic of the Day</COLUMN>
<TITLE>Átadáson túl, a Tragédia és A vihar előtt • A próbák elkezdődtek a Nemzeti Színházban</TITLE>
<LEAD>Akarták a Városligetbe, a Várba, az Erzsébet térre, a Műcsarnokba, meg a New York palotába. Az építésszek terveztek, terveikkel nyertek a pályázatokon, de azok a nemzeti színházak soha nem épültek meg. Most a Ferencvárosban benépesül egy színház, Nemzeti néven.</LEAD>
<TEXT>
<P>Ma délelőtt tíztől elkezdődnek a próbák a Duna-parti új Nemzeti Színház épületében. A március 15-ei teátrumavató előadásában. Madách Imre Az ember tragédiájában huszonhárom színész, húsz táncos – a Szegedi Kortárs Balett művészei – és húsz közpárhuzamosan szintén mártól próbálják a teátrum színpadán a március 23-án, második bemutatóként György által rendezett előadásra 13 színészt szerződtettek. Prosperót Benedek Miklós, Arielt Nagy-K. András alakítja. A színházba az elmúlt év végén már beköltözött az igazgatóság, a gazdasági részleg (világosítók és a fodrászok). A kormány a költségvetési tartalékból 2,3 milliárd forintot szavazott meg a kell fedezni A Nemzeti Színész címmel járó havi juttatásokat is.)</P>
<P>A produkciókat tizenkilenc díszletmunkás, nyolc világosító, három hangosító, hat kellékes, három varrodában, a kalapos- és a cipésműhelyben. Az épület működtetésében részt vesz még portás, go<P>Január végétől már jegyeket is lehet venni az új Nemzeti előadásaira. A tervek szerint a Blaha Lu gyanánt.</P>
</TEXT>
</DOC>
```

### - 25 topik

```
<top>
<num> C301 </num>
<HU-title> Nestlé márkák </HU-title>
<HU-desc> Mely márkákat értékesít a Nestlé világszerte? </HU-desc>
<HU-narr> A releváns dokumentumok olyan árucikkek neveit tartalmazzák, amelyeket a Nestlé ill. valamelyik Nestlé-csoporthoz tartozó vállalat értékesít. Utóbbi esetben világos utalás szükséges a Nestlére mint anyacégre. </HU-narr>
</top>
```

#### A feladat:

- szavak szótövezése
- szócsoporthoz meghatározása
- indexelés:
  - Lista, mely szavak mely dokumentumokban fordulnak elő
  - Lista, mely dokumentumok mely szavakat tartalmazzák
- rangsorolás relevancia szerint (tf-idf)

# Képkeresés

Kis képi adatbázis:

-1504 kép

7070527 alacsony szintű leíró  
(szürkeskálás SIFT Lowe 2001)



Öt topik:

topic1: lovagló emberek

topic2: városi utcaképek

topic3: országúti képek

topic4: zöld mező

topic5: hófödte táj

A feladat:

- felügyelet nélküli klaszterezés
- minden képre egy leíró vektor készítése
- hasonló leírók keresése a mintaképek leíróihoz
- képek sorba rendezése topikra vonatkozó relevancia szerint



# Képkeresés



Öt topik:

topic1: lovagló emberek

topic2: városi utcaképek

topic3: országúti képek

topic4: zöld mező

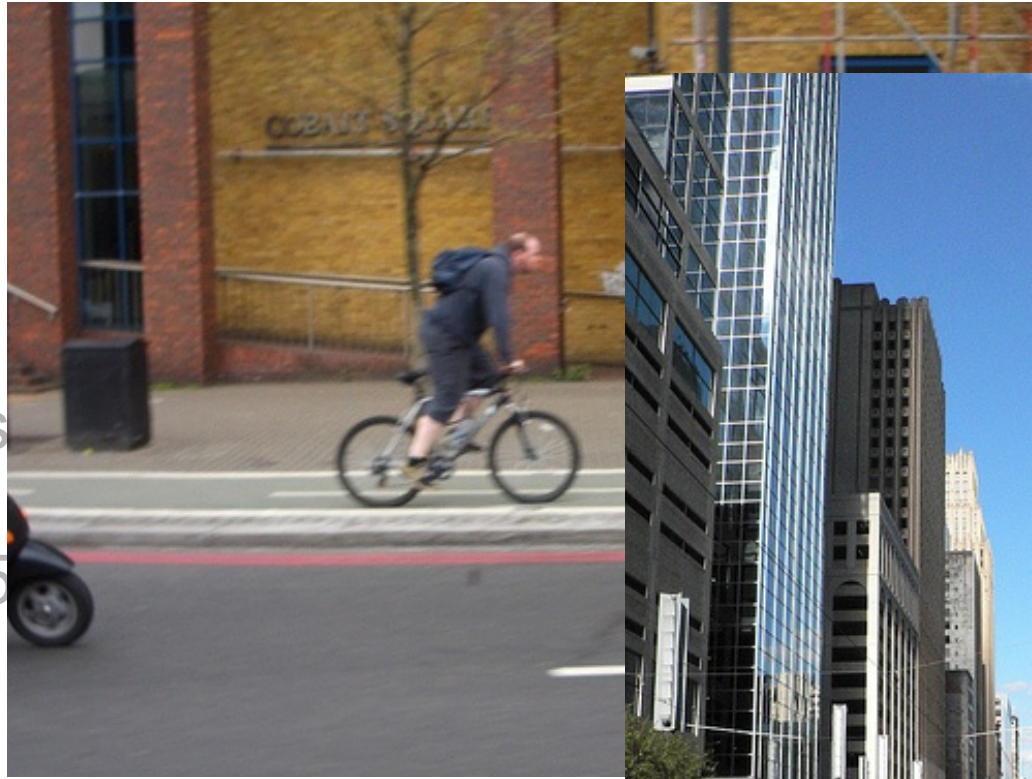
topic5: hófödte táj



# Képkeresés

Kis

-15



Öt topik:

topic1: lovagló emberek

**topic2: városi utcaképek**

topic3: országúti képek

topic4: zöld mező

topic5: hófödte táj

# Képkeresés



Kis ke

-1504

70

(s

Öt top

topic1: lovagló emberek

topic2: városi utcaképek

**topic3: országúti képek**

topic4: zöld mező

topic5: hófödte táj



# Képkeresés



Öt topik:

topic1: lovagló ember

topic2: városi utcakép

topic3: országúti kép

**topic4: zöld mező**

topic5: hófödte táj



# Képkeresés



(szurkeskalas SIF I



Öt topik:

topic1: lovagló emberek

topic2: városi utcaképek

topic3: országúti képek

topic4: zöld mező

**topic5: hófödte táj**

# Képklasszifikáció

Kis képi adatbázis:

-1504 kép , ebből 1000 ismert címkével rendelkezik  
7070527 alacsony szintű leíró  
(szűrkeskálás SIFT Lowe 2001)

Három kategória:

ember  
kutya  
busz

A feladat:

- felügyelet nélküli klaszterezés
- minden képre egy leíró vektor készítése

A kapott leírókon tanulsi módszerek alkalmazása, a tesztadatok pedig kiértékelés



# Képklasszifikáció

567 minta 433 negatív minta

kezik

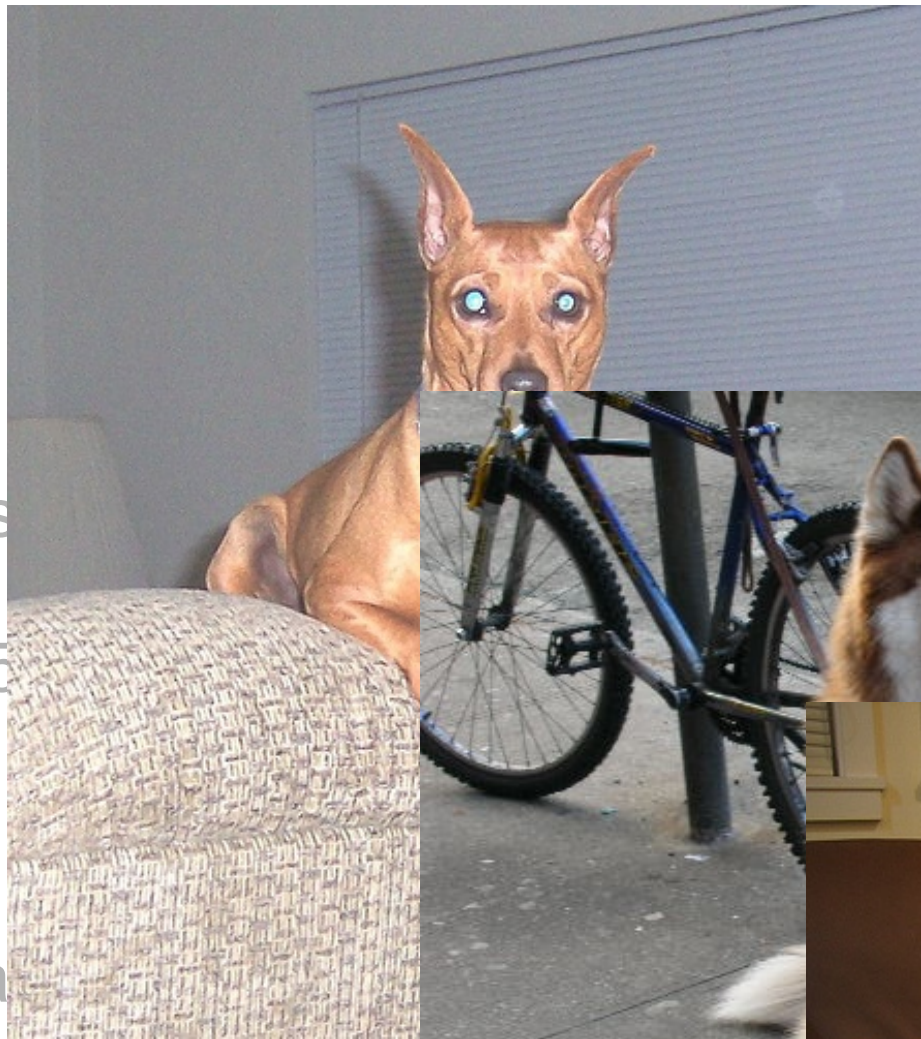
ember  
kutya  
busz





# Képklasszifikáció

418 minta 582 negatív minta



ember  
kutya  
busz



# Képklasszifikáció



174 minta 826 negatív minta

ezik



(szurkeskala

Három kategóri

ember  
kutya  
busz



# Szociális hálózatok

2.15 millió entitás (valós adat)

Kapcsolatok száma: 37057147

Ebből ismert: 37043782

Egy adott entitáshoz tartozik :

- születési dátum
- tartózkodási hely
- érdeklődési területei
- iskolái

Lokalizációs feladat:

- adott entitások tartózkodási helyét szeretnénk megjósolni a szociális hálójából
- megfelelő entitásleíró készítése
- NN algoritmusok
- egyéb tanulási módszerek (weka)

Kapcsolatfelderítés:

300 entitás-entitás kapcsolat létezésének megjósolása

- megfelelő entitásleíró készítése
- valószínűségi modellek
- NN és más tanulóalgoritmusok