

# Analiza titlurilor articolelor ale site-urilor românești de știri

-Mario Ernest Uță

Prin această lucrare se dorește analiza titlurilor articolelor apărute pe unele dintre cele mai populare site-uri de știri din România: Digi24, ProTV și Observator. Analiza titlurilor constă în căutarea anumitor cuvinte cheie în titlurile articolelor, pentru a observa dacă există o creștere a utilizării acestora în anumite perioade de timp și pentru a determina dacă utilizarea frecventă a acestor cuvinte în titlurile de știri indică o schimbare în ceea ce privește PIB cap de locuitor în România.

Lista cuvintelor cheie\* este compusă din cuvinte care au, în general, o conotație negativă, ce ar putea indica o scădere a nivelului de trai. Aceste cuvinte au fost împărțite în 3 categorii:

- Economic. Cuvinte legate de economia țării sau de veniturile oamenilor(ex: inflație, recesiune, deficit)
- Război/Tragedie. Cuvinte legate de conflicte armate sau de catastrofe (ex: bombardament, explozie, victime)
- Pandemie. Cuvinte legate de pandemia de COVID-19(ex: carantină, coronavirus, restricții)

Pentru colectarea informațiilor, s-au folosit mai multe script-uri Python care analizează titlurile articolelor de știri și salvează cuvintele cheie găsite în fiecare titlu în parte. S-au analizat următoarele categorii de știri pentru fiecare site:

- Digi24. Categorii: "Ultimele știri".Număr articole: 366,851. Perioada:2012-2024
- ProTV. Categorii: "Politic","Financiar","Social","Internațional".Număr articole: 163,656. Perioada: 2008-2024
- Observator. Categorii: "Politic","Economic","Extern","Internațional".Număr articole: 106,596. Perioada:2010-2024
- Număr total articole: 637,103

Datele au fost salvate în formatul .csv, cu coloanele\*\*:

- link\_articol: Indică link-ul articolului
- platforma: Indică site-ul de știri unde a fost publicat articolul
- data: Data publicării articolului
- cuvinte\_cheie: Lista de cuvinte cheie găsite în titlul articolului

Pentru ilustrarea datelor se va folosi librăria Python Matplotlib

```
In [1]: import matplotlib.pyplot as plt
```

Procesam datele obținute pentru a le putea ilustra

```
In [2]: with open("Database/digi24.csv", "r") as f:
        f.readline()
        digi=[x.split(",") for x in f.readlines()]

        with open("Database/protv.csv", "r") as f:
            f.readline()
            protv=[x.split(",") for x in f.readlines()]

        with open("Database/observator.csv", "r") as f:
            f.readline()
            observator=[x.split(",") for x in f.readlines()]
```

Procesam datele de la Banca Mondială pentru a afla evoluția PIB pe cap de locuitor în România\*\*\*[1]

```
In [3]: with open("Database/WORLD_BANK_GDP_PER_CAPITA.csv", "r") as f:
        [f.readline() for x in range(4)]
        header=[x.strip('\"') for x in f.readline().split(",")]
        l=[x.split(",") for x in f.readlines()]

        romaniaGDPCapita=l[201]
        romaniaGDPCapitaRelevant=[round(float(x.strip('\"')),2) for x ,y in zip(ro
```

Construim lista cu cuvinte cheie

```
In [4]: with open("badwords.txt", "r") as f:
        badwords=f.readlines()
        badwords=[x.rstrip() for x in badwords if x[0] != "#"]
```

Construim două dicționare pentru fiecare site de știri analizat. Dicționarul year\_ va reține numărul de articole dintr-un an, iar dicționarul word\_ va reține numărul de apariții al cuvintelor cheie în titlurile articolelor din fiecare an

```
In [5]: yearDigi={str(x):0 for x in range(1999,2025)}
        wordsDigi={str(x):0 for x in range(1999,2025)}

        yearProTv={str(x):0 for x in range(1999,2025)}
        wordsProTv={str(x):0 for x in range(1999,2025)}

        yearObservator={str(x):0 for x in range(1999,2025)}
        wordsObservator={str(x):0 for x in range(1999,2025)}
```

Aflăm numărul articolelor din fiecare an și numărul de apariții al cuvintelor cheie în titlurile articolelor din fiecare an

```
In [6]: for i in digi:
        date=i[2].split('.')
        yearDigi[date[2]]+=1
        for cuv in badwords:
            if(cuv in i[3]):
                wordsDigi[date[2]]+=1
```

```

for i in protv:
    date=i[2].split('.')
    yearProTv[date[2]]+=1
    for cuv in badwords:
        if(cuv in i[3]):
            wordsProTv[date[2]]+=1

for i in observator:
    date=i[2].split('.')
    yearObservator[date[2]]+=1
    for cuv in badwords:
        if(cuv in i[3]):
            wordsObservator[date[2]]+=1

```

Facem raportul dintre numărul de cuvinte cheie apărut în titlurile articolelor din fiecare an și numărul de articole din fiecare an

```

In [7]: dataDigi=[round(wordsDigi[str(i)]/yearDigi[str(i)],2) for i in range (199
dataProTv=[round(wordsProTv[str(i)]/yearProTv[str(i)],2) for i in range (
dataObs=[round(wordsObservator[str(i)]/yearObservator[str(i)],2) for i in

```

Construim graficele

```

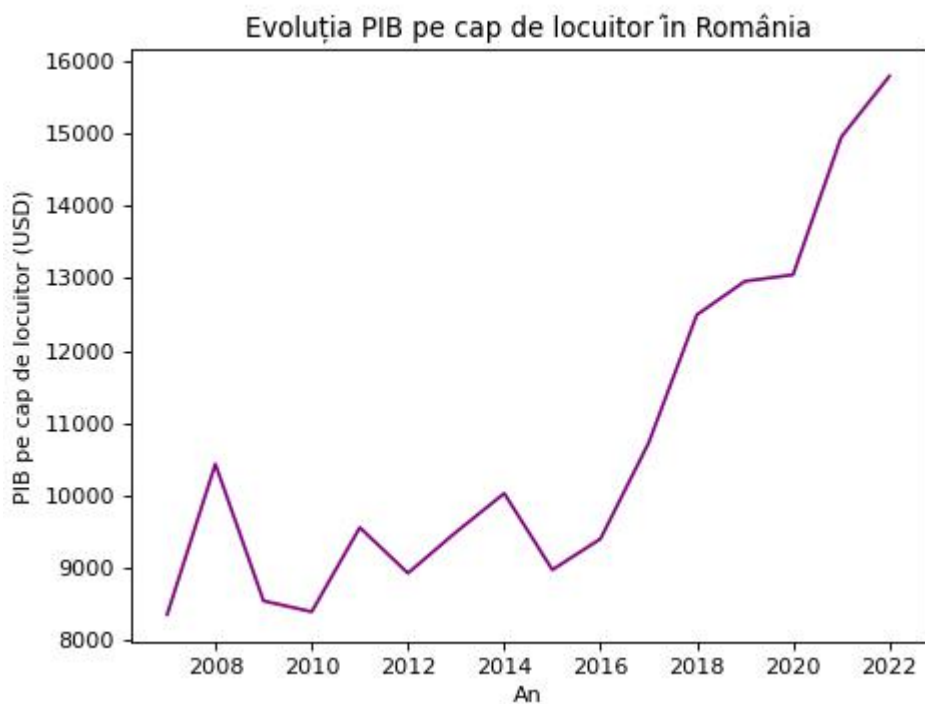
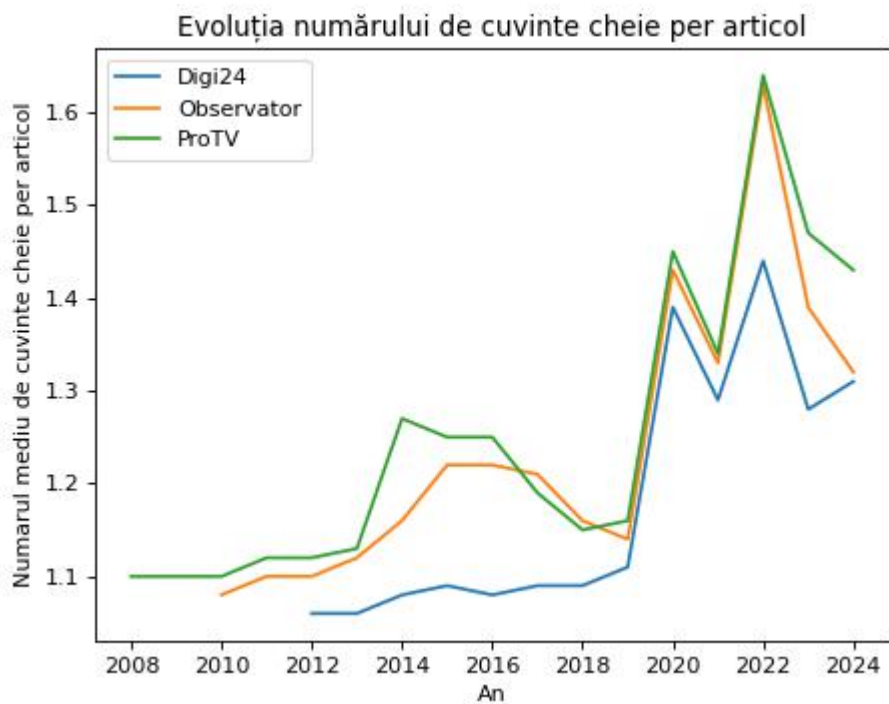
In [10]: plt.figure(num=0,dpi=80)
plt.plot([x for x in range(2012,2025)],dataDigi,label="Digi24")
plt.plot([x for x in range(2010,2025)],dataObs,label="Observator")
plt.plot([x for x in range(2008,2025)],dataProTv,label="ProTV")
plt.xlabel("An")
plt.title("Evoluția numărului de cuvinte cheie per articol")
plt.ylabel("Numarul mediu de cuvinte cheie per articol")
plt.legend()
plt.savefig('Graphs/News.jpg')
plt.close()

```

```

In [11]: plt.figure(num=0,dpi=80)
plt.plot([x for x in range(2007,2023)],romaniaGDPCapitaRelevant,color="pu
plt.xlabel("An")
plt.ylabel("PIB pe cap de locuitor (USD)")
plt.title("Evoluția PIB pe cap de locuitor în România")
plt.savefig('Graphs/GDP.jpg')
plt.close()

```



Din graficele anterioare se pot trage următoarele concluzii:

- O creștere a numărului de cuvinte cheie folosite pe titlu de articol nu indică vreo scădere în PIB pe cap de locuitor
- O creștere considerabilă a numărului de cuvinte cheie folosite pe titlu de articol în ultimii ani, fapt ce poate fi explicat de pandemia de COVID-19(2019) și de războiul ruso-ucrainean(2022)
- Toate cele trei posturi de știri au graficul extrem de asemănător, ceea ce indică o similitudine a subiectelor abordate de cele trei posturi de știri

\* Lista cuvintelor cheie este fișierul badwords.txt din repository-ul <https://github.com/MarioUta/Romanian-News-Outlets-Analysis>

\*\* Datele se găsesc în directorul Database din repository-ul <https://github.com/MarioUta/Romanian-News-Outlets-Analysis>

\*\*\*Datele furnizate de Banca Mondială sunt până în anul 2022

[1]World Bank. (n.d.). GDP per capita (current US\$) - romania | data. <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=SA/1000-RO>