



**МІНІСТЕРСТВО ОСВІТИ, НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Лабораторна робота №6
Аналіз даних**

Підготував:

студент 4 курсу

групи ФІ-84

Коломієць Андрій Юрійович

E-mail: andrew.kolomiets.work@gmail.com

Київ – 2021

Лабораторна робота №6

Аналіз даних

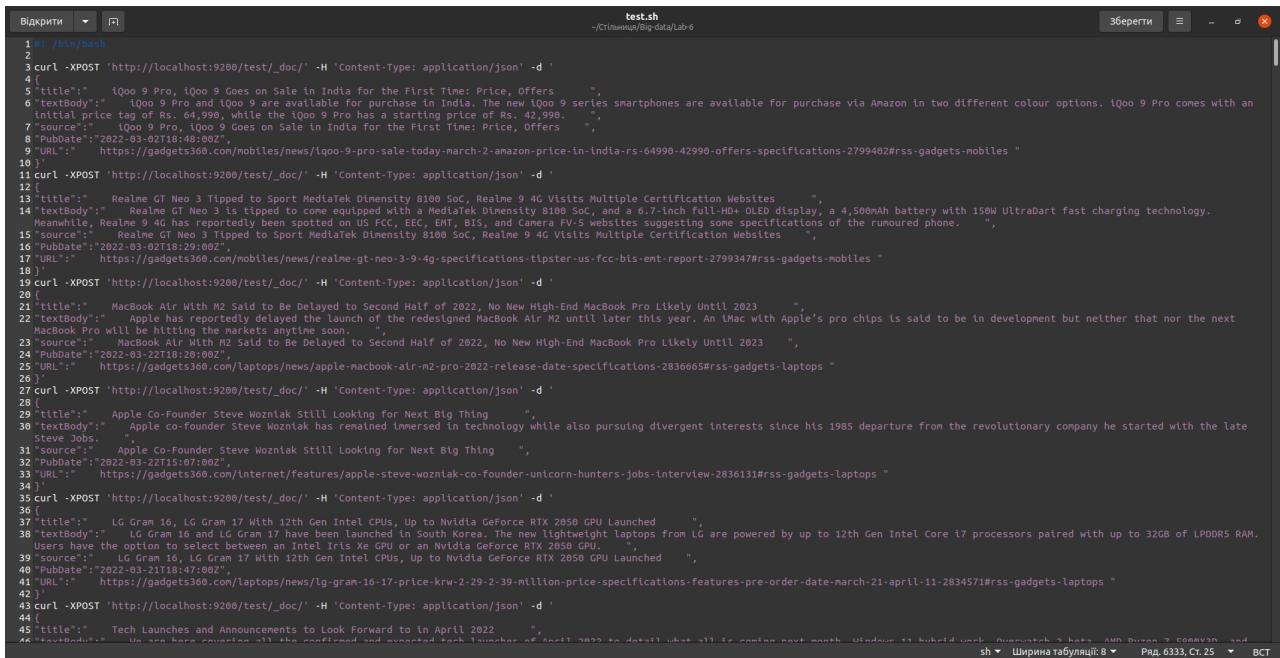
Завдання на самостійну роботу

1. Завантажити в базу даних системи **Elasticsearch** тестову базу даних із накопиченими даними з **RSS-фідів** з мережі Інтернет.
2. Самостійно здійснити агрегацію цих даних за полем дата-час.
3. Завантажити отримані результати у доступну систему типу **DSP** (**Digital Signal Processing**).

Виконання завдання

Накопичення даних RSS-фідів

Відповідно до попередніх лабораторних робіт створюємо великий масив даних та створюємо скрипт для завантаження в **Elasticsearch**.



```
test.sh
#!/bin/bash
1 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
2 {
3 "title": " iQoo 9 Pro, iQoo 9 Goes on Sale in India for the First Time: Price, Offers   ",
4 "textBody": " iQoo 9 Pro and iQoo 9 are available for purchase in India. The new iQoo 9 series smartphones are available for purchase via Amazon in two different colour options. iQoo 9 Pro comes with an initial price tag of Rs. 64,990, while the iQoo 9 Pro has a starting price of Rs. 42,990.   ",
5 "source": " iQoo 9 Pro, iQoo 9 Goes on Sale in India for the First Time: Price, Offers   ",
6 "PubDate": "2022-03-02T18:48:00Z",
7 "URL": "https://gadgets360.com/mobiles/news/iqoo-9-pro-sale-today-march-2-amazon-price-in-india-rs-64990-42990-offers-specifications-2799402#rss-gadgets-mobiles"
8 }
9 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
10 {
11 "title": " Realme GT Neo 3 Tipped to Sport MediaTek Dimensity 8100 SoC, Realme 9 4G Visits Multiple Certification Websites   ",
12 "textBody": " Realme GT Neo 3 is tipped to come equipped with a MediaTek Dimensity 8100 SoC, and a 6.7-inch full-HD+ OLED display, a 4,500mAh battery with 150W UltraDart fast charging technology. Meanwhile, Realme 9 4G has reportedly been spotted on US FCC, EEC, EMT, BIS, and Camera FV-5 websites suggesting some specifications of the rumoured phone.   ",
13 "source": " Realme GT Neo 3 Tipped to Sport MediaTek Dimensity 8100 SoC, Realme 9 4G Visits Multiple Certification Websites   ",
14 "PubDate": "2022-03-02T18:29:00Z",
15 "URL": "https://gadgets360.com/mobiles/news/realme-gt-neo-3-9-4g-specifications-tipster-us-fcc-bis-emt-report-2799347#rss-gadgets-mobiles"
16 }
17 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
18 {
19 "title": " MacBook Air With M2 Said to Be Delayed to Second Half of 2022, No New High-End MacBook Pro Likely Until 2023   ",
20 "textBody": " Apple has reportedly delayed the launch of the redesigned MacBook Air M2 until later this year. An iMac with Apple's pro chips is said to be in development but neither that nor the next MacBook Pro will be hitting the markets anytime soon.   ",
21 "source": " MacBook Air With M2 Said to Be Delayed to Second Half of 2022, No New High-End MacBook Pro Likely Until 2023   ",
22 "PubDate": "2022-03-22T18:20:00Z",
23 "URL": "https://gadgets360.com/laptops/news/apple-macbook-air-m2-pro-2022-release-date-specifications-2830665#rss-gadgets-laptops"
24 }
25 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
26 {
27 "title": " Apple Co-Founder Steve Wozniak Still Looking for Next Big Thing   ",
28 "textBody": " Apple co-founder Steve Wozniak has remained immersed in technology while also pursuing divergent interests since his 1985 departure from the revolutionary company he started with the late Steve Jobs.   ",
29 "source": " Apple Co-Founder Steve Wozniak Still Looking for Next Big Thing   ",
30 "PubDate": "2022-03-22T18:07:00Z",
31 "URL": "https://gadgets360.com/internet/features/apple-steve-wozniak-co-founder-unicorn-hunters-jobs-interview-2830131#rss-gadgets-laptops"
32 }
33 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
34 {
35 "title": " LG Gram 16, LG Gram 17 With 12th Gen Intel CPUs, Up to Nvidia GeForce RTX 2050 GPU Launched   ",
36 "textBody": " LG Gram 16 and LG Gram 17 have been launched in South Korea. The new lightweight laptops from LG are powered by up to 12th Gen Intel Core i7 processors paired with up to 32GB of LPDDR5 RAM. Users can choose the option to select between an Intel Iris Xe GPU or an Nvidia GeForce RTX 2050 GPU.   ",
37 "source": " LG Gram 16, LG Gram 17 With 12th Gen Intel CPUs, Up to Nvidia GeForce RTX 2050 GPU Launched   ",
38 "PubDate": "2022-03-21T18:47:00Z",
39 "URL": "https://gadgets360.com/laptops/news/lg-gram-16-17-price-krw-2-29-2-39-million-price-specifications-features-pre-order-date-march-21-april-11-2834571#rss-gadgets-laptops"
40 }
41 curl -XPOST 'http://localhost:9200/test/_doc/' -H 'Content-Type: application/json' -d '
42 {
43 "title": " Tech Launches and Announcements to Look Forward to in April 2022   ",
44 "textBody": " We are here coming all the confirmed and expected tech launches in April 2022 to detail what all to expect next month. Windows 11, handheld work, launches of bats, AMD Ryzen 7, Google Pixel 6 and
45 "source": " Tech Launches and Announcements to Look Forward to in April 2022   "
46 "PubDate": "2022-03-21T18:47:00Z"
47 "URL": "https://gadgets360.com/mobiles/news/tech-launches-and-announcements-to-look-forward-to-in-april-2022#rss-gadgets-mobiles"
48 }
```

Запускаємо Elasticsearch.

```
./bin/elasticsearch
```

Завантажуємо дані з файлу:

```
$ chmod a+x data.sh; ./data.sh
```

Перевіряємо наявність завантажених файлів.

```
curl -XGET 'http://localhost:9200/test/_doc/_count'; echo
```

```
linux@linux-X505BP:~/Стільниця/Big-data/Lab-6$ curl -XGET 'http://localhost:9200/test/_doc/_count'; echo {"count":1188,"_shards":{"total":1,"successful":1,"skipped":0,"failed":0}}
```

Агрегація даних

Для отримання динаміки кількості публікацій за запитом в **Elasticsearch**, необхідно здійснити агрегацію даних, які відповідають тематиці, визначеній первинним запитом за полем, що відповідає значенню дати і часу.

Агрегація даних

```
#!/bin/bash

curl -XGET 'http://localhost:9200/test/_search?pretty=true' -H 'Content-Type: application/json' -d '
{
  "query": {
    "multi_match": {"query": "apple", "fields": ["title", "textBody"] }
  },
  "aggregations": {
    "dates_with_holes": {
      "date_histogram": {
        "field": "PubDate",
        "interval": "day",
        "min_doc_count": 0
      }
    }
  },
  "size": 0
}'
```

Запускаємо скрипт

```
./CSV.py > result.json
```

Результат

```
{
  "took" : 77,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
}
```

```
},
  "hits" : {
    "total" : {
      "value" : 77,
      "relation" : "eq"
    },
    "max_score" : null,
    "hits" : [ ]
  },
  "aggregations" : {
    "dates_with_holes" : {
      "buckets" : [
        {
          "key_as_string" : "2022-01-13T00:00:00.000Z",
          "key" : 1642032000000,
          "doc_count" : 2
        },
        {
          "key_as_string" : "2022-01-14T00:00:00.000Z",
          "key" : 1642118400000,
          "doc_count" : 0
        },
        {
          "key_as_string" : "2022-01-15T00:00:00.000Z",
          "key" : 1642204800000,
          "doc_count" : 0
        },
        {
          "key_as_string" : "2022-01-16T00:00:00.000Z",
          "key" : 1642291200000,
          "doc_count" : 0
        },
        {
          "key_as_string" : "2022-01-17T00:00:00.000Z",
          "key" : 1642377600000,
          "doc_count" : 1
        },
        {
          "key_as_string" : "2022-01-18T00:00:00.000Z",
          "key" : 1642464000000,
          "doc_count" : 0
        },
        {
          "key_as_string" : "2022-01-19T00:00:00.000Z",
          "key" : 1642550400000,
          "doc_count" : 0
        }
      ]
    }
  }
}
```

```
"key" : 1642550400000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-20T00:00:00.000Z",
"key" : 1642636800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-21T00:00:00.000Z",
"key" : 1642723200000,
"doc_count" : 1
},
{
"key_as_string" : "2022-01-22T00:00:00.000Z",
"key" : 1642809600000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-23T00:00:00.000Z",
"key" : 1642896000000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-24T00:00:00.000Z",
"key" : 1642982400000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-25T00:00:00.000Z",
"key" : 1643068800000,
"doc_count" : 1
},
{
"key_as_string" : "2022-01-26T00:00:00.000Z",
"key" : 1643155200000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-27T00:00:00.000Z",
"key" : 1643241600000,
"doc_count" : 1
},
{

```

```
"key_as_string" : "2022-01-28T00:00:00.000Z",
"key" : 1643328000000,
"doc_count" : 1
},
{
"key_as_string" : "2022-01-29T00:00:00.000Z",
"key" : 1643414400000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-30T00:00:00.000Z",
"key" : 1643500800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-01-31T00:00:00.000Z",
"key" : 1643587200000,
"doc_count" : 0
},
{
"key_as_string" : "2022-02-01T00:00:00.000Z",
"key" : 1643673600000,
"doc_count" : 1
},
{
"key_as_string" : "2022-02-02T00:00:00.000Z",
"key" : 1643760000000,
"doc_count" : 0
},
{
"key_as_string" : "2022-02-03T00:00:00.000Z",
"key" : 1643846400000,
"doc_count" : 0
},
{
"key_as_string" : "2022-02-04T00:00:00.000Z",
"key" : 1643932800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-02-05T00:00:00.000Z",
"key" : 1644019200000,
"doc_count" : 0
},
```

```
{  
    "key_as_string" : "2022-02-06T00:00:00.000Z",  
    "key" : 1644105600000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-07T00:00:00.000Z",  
    "key" : 1644192000000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-08T00:00:00.000Z",  
    "key" : 1644278400000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-09T00:00:00.000Z",  
    "key" : 1644364800000,  
    "doc_count" : 1  
},  
{  
    "key_as_string" : "2022-02-10T00:00:00.000Z",  
    "key" : 1644451200000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-11T00:00:00.000Z",  
    "key" : 1644537600000,  
    "doc_count" : 1  
},  
{  
    "key_as_string" : "2022-02-12T00:00:00.000Z",  
    "key" : 1644624000000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-13T00:00:00.000Z",  
    "key" : 1644710400000,  
    "doc_count" : 0  
},  
{  
    "key_as_string" : "2022-02-14T00:00:00.000Z",  
    "key" : 1644796800000,  
    "doc_count" : 0  
}
```

```
},
{
  "key_as_string" : "2022-02-15T00:00:00.000Z",
  "key" : 1644883200000,
  "doc_count" : 2
},
{
  "key_as_string" : "2022-02-16T00:00:00.000Z",
  "key" : 1644969600000,
  "doc_count" : 1
},
{
  "key_as_string" : "2022-02-17T00:00:00.000Z",
  "key" : 1645056000000,
  "doc_count" : 0
},
{
  "key_as_string" : "2022-02-18T00:00:00.000Z",
  "key" : 1645142400000,
  "doc_count" : 0
},
{
  "key_as_string" : "2022-02-19T00:00:00.000Z",
  "key" : 1645228800000,
  "doc_count" : 0
},
{
  "key_as_string" : "2022-02-20T00:00:00.000Z",
  "key" : 1645315200000,
  "doc_count" : 0
},
{
  "key_as_string" : "2022-02-21T00:00:00.000Z",
  "key" : 1645401600000,
  "doc_count" : 1
},
{
  "key_as_string" : "2022-02-22T00:00:00.000Z",
  "key" : 1645488000000,
  "doc_count" : 0
},
{
  "key_as_string" : "2022-02-23T00:00:00.000Z",
  "key" : 1645574400000,
```

```
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-02-24T00:00:00.000Z",
    "key" : 1645660800000,
    "doc_count" : 2
  },
  {
    "key_as_string" : "2022-02-25T00:00:00.000Z",
    "key" : 1645747200000,
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-02-26T00:00:00.000Z",
    "key" : 1645833600000,
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-02-27T00:00:00.000Z",
    "key" : 1645920000000,
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-02-28T00:00:00.000Z",
    "key" : 1646006400000,
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-03-01T00:00:00.000Z",
    "key" : 1646092800000,
    "doc_count" : 0
  },
  {
    "key_as_string" : "2022-03-02T00:00:00.000Z",
    "key" : 1646179200000,
    "doc_count" : 6
  },
  {
    "key_as_string" : "2022-03-03T00:00:00.000Z",
    "key" : 1646265600000,
    "doc_count" : 1
  },
  {
    "key_as_string" : "2022-03-04T00:00:00.000Z",
```

```
"key" : 1646352000000,
"doc_count" : 6
},
{
"key_as_string" : "2022-03-05T00:00:00.000Z",
"key" : 1646438400000,
"doc_count" : 4
},
{
"key_as_string" : "2022-03-06T00:00:00.000Z",
"key" : 1646524800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-07T00:00:00.000Z",
"key" : 1646611200000,
"doc_count" : 6
},
{
"key_as_string" : "2022-03-08T00:00:00.000Z",
"key" : 1646697600000,
"doc_count" : 2
},
{
"key_as_string" : "2022-03-09T00:00:00.000Z",
"key" : 1646784000000,
"doc_count" : 10
},
{
"key_as_string" : "2022-03-10T00:00:00.000Z",
"key" : 1646870400000,
"doc_count" : 1
},
{
"key_as_string" : "2022-03-11T00:00:00.000Z",
"key" : 1646956800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-12T00:00:00.000Z",
"key" : 1647043200000,
"doc_count" : 0
},
{

```

```
"key_as_string" : "2022-03-13T00:00:00.000Z",
"key" : 1647129600000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-14T00:00:00.000Z",
"key" : 1647216000000,
"doc_count" : 1
},
{
"key_as_string" : "2022-03-15T00:00:00.000Z",
"key" : 1647302400000,
"doc_count" : 4
},
{
"key_as_string" : "2022-03-16T00:00:00.000Z",
"key" : 1647388800000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-17T00:00:00.000Z",
"key" : 1647475200000,
"doc_count" : 1
},
{
"key_as_string" : "2022-03-18T00:00:00.000Z",
"key" : 1647561600000,
"doc_count" : 1
},
{
"key_as_string" : "2022-03-19T00:00:00.000Z",
"key" : 1647648000000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-20T00:00:00.000Z",
"key" : 1647734400000,
"doc_count" : 0
},
{
"key_as_string" : "2022-03-21T00:00:00.000Z",
"key" : 1647820800000,
"doc_count" : 7
},
```

```
{
    "key_as_string" : "2022-03-22T00:00:00.000Z",
    "key" : 1647907200000,
    "doc_count" : 11
}
]
}
}
```

Переведення даних в формат CSV

Для застосування отриманих результатів у середовищі спеціалізованих систем обробки цифрових даних, можна перетворити їх до формату **CSV** за допомогою програми мовою **Python**, код якої наведено нижче:

CSV.py

```
#!/bin/python3
import sys
import re

t = sys.stdin.read()
json = t.split('\n')
t=""

for i in range(len(json)):
    t=t+" "+json[i]

t=re.sub('^\s+', '', t)
days = re.findall('"key_as_string": "(.+?)T', t)
count = re.findall('"doc_count": (\d+)', t)

for i in range(len(days)):
    print(days[i]+";"+count[i])
```

У результаті виконання програми можна отримати дані у форматі **CSV** такого вигляду:

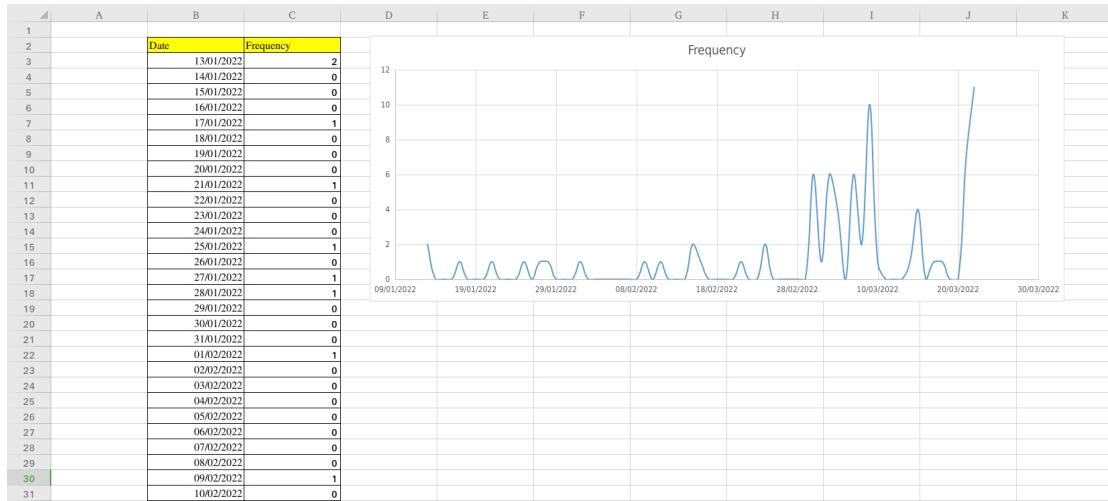
```
$ cat result.json | ./CSV.py
```

Результат

2022-01-13;2	2022-02-17;0
2022-01-14;0	2022-02-18;0
2022-01-15;0	2022-02-19;0
2022-01-16;0	2022-02-20;0
2022-01-17;1	2022-02-21;1
2022-01-18;0	2022-02-22;0
2022-01-19;0	2022-02-23;0
2022-01-20;0	2022-02-24;2
2022-01-21;1	2022-02-25;0
2022-01-22;0	2022-02-26;0
2022-01-23;0	2022-02-27;0
2022-01-24;0	2022-02-28;0
2022-01-25;1	2022-03-01;0
2022-01-26;0	2022-03-02;6
2022-01-27;1	2022-03-03;1
2022-01-28;1	2022-03-04;6
2022-01-29;0	2022-03-05;4
2022-01-30;0	2022-03-06;0
2022-01-31;0	2022-03-07;6
2022-02-01;1	2022-03-08;2
2022-02-02;0	2022-03-09;10
2022-02-03;0	2022-03-10;1
2022-02-04;0	2022-03-11;0
2022-02-05;0	2022-03-12;0
2022-02-06;0	2022-03-13;0
2022-02-07;0	2022-03-14;1
2022-02-08;0	2022-03-15;4
2022-02-09;1	2022-03-16;0
2022-02-10;0	2022-03-17;1
2022-02-11;1	2022-03-18;1
2022-02-12;0	2022-03-19;0
2022-02-13;0	2022-03-20;0
2022-02-14;0	2022-03-21;7
2022-02-15;2	2022-03-22;11
2022-02-16;1	2022-03-23;5

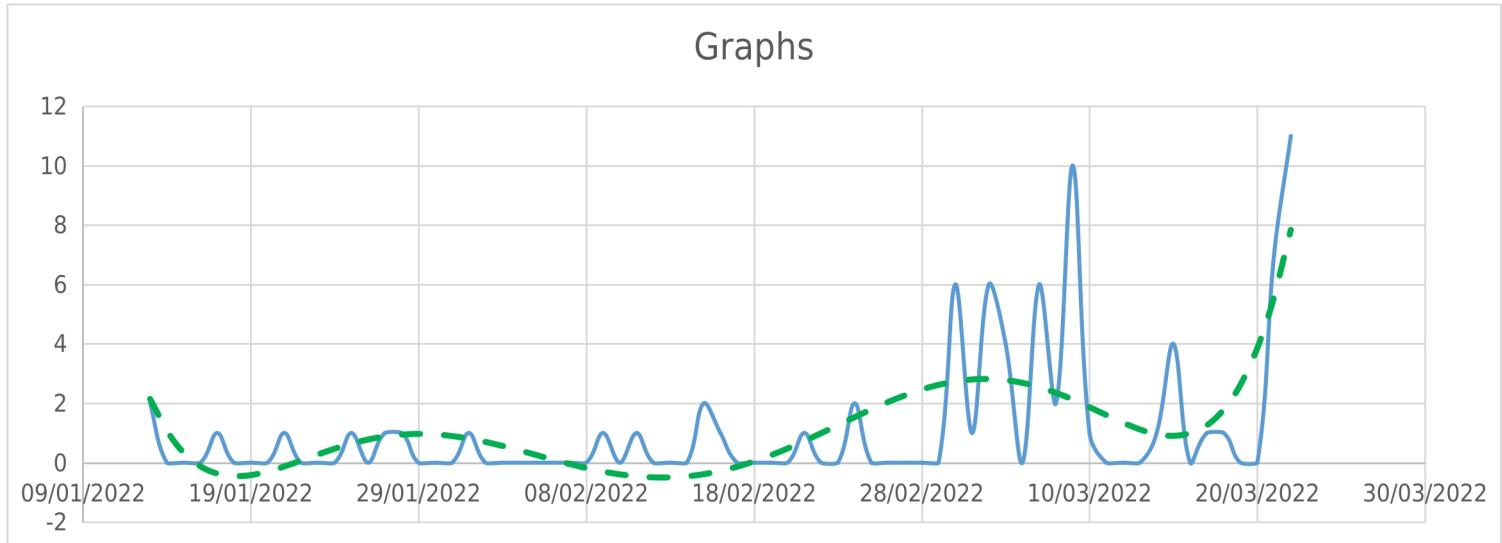
Обробка даних у середовищі Excel

Для подальшої обробки даних в форматі завантажимо отриманий **CSV-файл** у середовище **Excel** і побудуємо графік динаміки повідомлень.



Після завантаження **CSV-файлу** в систему цифрової обробки даних виникають прості можливості його статистичного оброблення.

Знаходження поліноміального тренду цього ряду.



Питання до практичної роботи

Що таке агрегація даних?

Агрегація даних - це різновид процесу обміну даними та інформацією, де дані здійснюють пошук, збирання та подання у звітно-узагальненому форматі на основі звітів для досягнення конкретних бізнес-цілей або процесів та / або проведення людського аналізу.

Агрегація даних може проводитися вручну або за допомогою спеціалізованого програмного забезпечення.

Які існують типи агрегації в Elasticsearch?

Агрегація узагальнює ваші дані як показники, статистичні дані чи інші аналітичні дані.

Elasticsearch організовує агрегації на три категорії:

1. **Агрегації показників**, які обчислюють такі показники, як сума чи середнє значення, із значень полів.
2. **Агрегації сегментів**, які групують документи в сегменти, які також називаються контейнерами, на основі значень полів, діапазонів або інших критеріїв.
3. **Агрегації конвеєра**, які отримують вхідні дані з інших агрегатів замість документів або полів.

Що таке CSV-формат файлу?

CSV (від англ. *comma-separated values* ‘значення, розділені комою’, іноді *character-separated values* ‘значення, розділені символом’)— файловий формат, котрий є відмежовувальним форматом для представлення табличних даних, у якому поля відокремлюються символом коми та переходу на новий рядок. Поля, що містять коми, декілька рядків, або лапки (позначаються подвійними лапками), мають обмежуватися з обох боків лапками.

Формат **CSV** використовується для перенесення даних між базами даних та програмами — редакторами електронних таблиць.

Як побудувати графік завантажених даних і визначити тренд часового ряду?

Дуже детальний опис наведено в цьому посібнику:

Кузьмичов А.І. Економетричне моделювання та прогнозування в Excel.

Навч. пос. / Бишовець Н.Г., Кузьмичов А.І., Медведєв М.Г.,

Омечинська Н.В. – К.: ВПЦ АМУ, 2010. – 324 с.

Але сама процедура наведено в практичній роботі вище, там можна побачити покрокову процедуру створення графіку і визначення часового тренду.

Код для форматування JSON

Зауваження: в першій лабораторній роботі програма для форматування JSON файлу мала свої недоліки пов'язані з 'source' та 'PubDate' - котра була у всіх об'єктів однакова. Поданий нижче код позбавлений будь-яких недоліків у форматуванні файлу такого типу.

```
#! /bin/python3

#Програма конвертації даних із формату RSS до формату JSON
#Підключення модулів для роботи із регулярними виразами і часом

import re
import string
import datetime

#Відкриття файлу rss.xml у режимі «читання»
f = open("rss.xml","r")
#Зчитування вмісту файлу rss.xml у змінну t
t = f.read()
#Закриття файлу rss.xml
f.close()

#Розбиття файлу по рядкам і склеювання рядків через пропуск
rss =t.split('\n')
t=""
for i in range(len(rss)):
    t=t+" "+rss[i]

#Видалення перших пропусків
t=re.sub('^\s','',t)
#Формування масиву заголовків
title = re.findall('<title>(.+?)</title>', t)
#Перший заголовок - назва фіdu, далі - його специфічна обробка
source = []
for i in range(0, len(title)):
    source.append(title[i])
    source[i]=re.sub('[\s\-\-]*$', '',source[i])
    source[i]=re.sub('\"','\"',source[i])
#Розмірність масиву заголовків
x=range(len(title))
#Формування масиву описів
text = re.findall('<description>(.+?)</description>', t)
#Формування масиву гіперпосилань
link = re.findall('<link>(.+?)</link>', t)
```

```

#Формування дати і часу в форматі "YY-ММ-ММТНН:ММ:00Z"

time=[]

time_not_filter = re.findall('<pubDate>(.*)</pubDate>', t)

for i in range(0, len(title)):

    spl_string = time_not_filter[i].split()
    rm = spl_string[:-1]
    rm = rm[1:]

    listToStr = (' '.join([str(elem) for elem in rm]))
    date_time_obj = datetime.datetime.strptime(listToStr, '%d %b %Y %H:%M:%S' )
    time.append(date_time_obj.strftime("%Y-%m-%dT%H:%M:00Z"))

#Виведення результатів

for i in range(0, len(title)):

    print ("{\n\"title\": \""+title[i]+\"",

#Специфічна обробка тексту

text[i]=re.sub('[\s\-\-]*$', '',text[i])
text[i]=re.sub('\"','\"',text[i])
text[i]=re.sub('\'','\'&',text[i])

#Подальша виведення результатів

print ("\"textBody\":"+text[i]+"\","")
print ("\"source\":"+source[i]+"\","")
print ("\"PubDate\":"+time[i]+"\","")
print ("\"URL\":"+link[i],"\n")

if i<len(title)-1:
    print (",")

```