

Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
Факультет прикладної математики  
Кафедра програмного забезпечення комп'ютерних систем

Методи організації пошуку інформації

### **Лабораторна робота № 2**

«Реалізувати інформаційно-пошукову систему, що інтегрована з системою Elasticsearch та надає можливість пошуку збережених документів за метаданими.»

Виконав:  
студент  
групи КП-93  
Долгов Олексій

Київ 2022

# Мета

Реалізувати інформаційно-пошукову систему, що інтегрована з системою Elasticsearch та надає можливість пошуку збережених документів за метаданими.

## Постановка завдання

Предметна галузь - музика. Реалізація складного запиту “fuzzy”

## Код Розробленого програмного забезпечення

<https://github.com/dolho/search-labs/tree/master/lab2>

## Результати роботи програми

Хід роботи консольної програми можна розділити на два підкроки:

1) Конструювання запиту, шляхом виконання команд

```
add-box-office-range
```

```
add-term
```

```
python main.py add-fuzzy
```

Сам запит зберігається у вигляді json файлу, який завантажуються у об'єкт BoolQuery

2) Виконання команди search яка виконує запит, використовуючи сформований на попередньому кроці запит.

Повний список команд наведено у файлі commands.txt

```
Album(album_name='Just for a Day', release_date='1991-06-03', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=30000, _id='a10877d3-ff0e-4639-9efa-489813cc174a')
Album(album_name='Loveless', release_date='1991-03-01', musicians=['Bilinda Butcher', 'Debbie Googe', 'Kevin Shields'], box_office=60000, _id='9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62')
Album(album_name='Iron Maiden', release_date='1980-04-23', musicians=['Paul Di'Anno', 'Dave Murray', 'Dennis Stratton', 'Steve Harris', 'Clive Burr'], box_office=2500000, _id='8c5139d3-5a65-4e6f-8fcc-9653d569b144')
Album(album_name='Soulvaki', release_date='1994-02-01', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=50000, _id='bccfaa7a-285e-4000-94ad-93988ba24ff6')
```

*Повний список документів*

```
(search-labs-py3.10) axel@axel-ThinkPad-E14:~/Desktop/uni/4_kypc/cem2/search-labs/lab2$ python main.py search-by-album-name Loveless{
  "album_name": "Loveless",
  "release_date": "1991-03-01",
  "musicians": [
    "Bilinda Butcher",
    "Debbie Googe",
    "Kevin Shields"
  ],
  "box_office": 60000,
  "_id": "9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62"
}
```

Пошук за назвою альбому (використовується fuzzy, тому одруківка Loveless не заважає знаходженню документу).

```
python main.py reset-query
python main.py add-box-office-range filter 40000 120000
python main.py add-term should album_name Soulvaki
python main.py add-fuzzy should album_name Loveles
python main.py search
```

```
Album(album_name='Soulvaki', release_date='1994-02-01', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=50000, _id='bccfaa7a-285e-4000-94ad-93988ba24ff6')
Album(album_name='Loveless', release_date='1991-03-01', musicians=['Bilinda Butcher', 'Debbie Googe', 'Kevin Shields'], box_office=60000, _id='9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62')
```

*Пошук альбомів, фільтруючи за box-office та використовуючи фільтр should, для пошуку альбому Soulvaki або fuzzy-пошуку альбому за запитом Loveles*

```
python main.py reset-query
python main.py add-term must musicians "Rachel Goswell"
python main.py search
```

```
Album(album_name='Just for a Day', release_date='1991-06-03', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=30000, _id='a10877d3-ff0e-4639-9efa-489813cc174a')
Album(album_name='Soulvaki', release_date='1994-02-01', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=50000, _id='bccfaa7a-285e-4000-94ad-93988ba24ff6')
```

*Пошук усіх альбомів, в запису яких брала участь Rachel Goswell*

```
python main.py reset-query
python main.py add-release-date-range filter 1979-01-01 1991-04-01
python main.py search
```

```
Album(album_name='Loveless', release_date='1991-03-01', musicians=['Bilinda Butcher', 'Debbie Googe', 'Kevin Shields'], box_office=60000, _id='9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62')
Album(album_name='Iron Maiden', release_date='1980-04-23', musicians=['Paul Di'Anno', 'Dave Murray', 'Dennis Stratton', 'Steve Harris', 'Clive Burr'], box_office=2500000, _id='8c5139d3-5a65-4e6f-8fcc-9653d569b144')
```

*Пошук усіх альбомів записаних у зазначений проміжок дат*

```
python main.py reset-query
python main.py index-document --album-name "Mongol throat singing feat. Genghis Khan" --release-date-range 1235-04-23 1235-04-23
python main.py search
python main.py delete uuid
python main.py search
```

```
(search-labs-py3.10) axel@axel-ThinkPad-E14:~/Desktop/uni/4_kypc/cem2/search-labs/lab2$ python main.py search
Album(album_name='Just for a Day', release_date='1991-06-03', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=30000, _id='a10877d3-ff0e-4639-9efa-489813cc174a')
Album(album_name='Loveless', release_date='1991-03-01', musicians=['Bilinda Butcher', 'Debbie Googe', 'Kevin Shields'], box_office=60000, _id='9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62')
Album(album_name='Iron Maiden', release_date='1980-04-23', musicians=['Paul Di'Anno', 'Dave Murray', 'Dennis Stratton', 'Steve Harris', 'Clive Burr'], box_office=2500000, _id='8c5139d3-5a65-4e6f-8fcc-9653d569b144')
Album(album_name='Soulvaki', release_date='1994-02-01', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=50000, _id='bccfaa7a-285e-4000-94ad-93988ba24ff6')
Album(album_name='Mongol throat singing feat. Genghis Khan', release_date='1235-04-23', musicians=['Horse Archer', 'Genghis Khan'], box_office=3, _id='0ca2f870-6cad-480b-b9a1-34c7221def03')
```

```
(search-labs-py3.10) axel@axel-ThinkPad-E14:~/Desktop/uni/4_kypc/cem2/search-labs/lab2$ python main.py delete 0ca2f870-6cad-480b-b9a1-34c7221def03
deleted
```

```
(search-labs-py3.10) axel@axel-ThinkPad-E14:~/Desktop/uni/4_kypc/cem2/search-labs/lab2$ python main.py search
Album(album_name='Just for a Day', release_date='1991-06-03', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=30000, _id='a10877d3-ff0e-4639-9efa-489813cc174a')
Album(album_name='Loveless', release_date='1991-03-01', musicians=['Bilinda Butcher', 'Debbie Googe', 'Kevin Shields'], box_office=60000, _id='9eb1d919-c33e-4a8f-af8a-66e2a2aa7a62')
Album(album_name='Iron Maiden', release_date='1980-04-23', musicians=['Paul Di'Anno', 'Dave Murray', 'Dennis Stratton', 'Steve Harris', 'Clive Burr'], box_office=2500000, _id='8c5139d3-5a65-4e6f-8fcc-9653d569b144')
Album(album_name='Soulvaki', release_date='1994-02-01', musicians=['Neil Halstead', 'Rachel Goswell', 'Christian Savill', 'Nick Chaplin', 'Simon Scott'], box_office=50000, _id='bccfaa7a-285e-4000-94ad-93988ba24ff6')
```

*Вставка та видалення за uuid альбому Mongol throat singing*

## Висновки

В ході даної лабораторної роботи було побудовано консольну пошукову систему, яка гнучко дозволяє редагувати запити для пошуку, використовуючи ключові слова term, range, fuzzy. Також було розроблено дії create та delete. Перед завантаженням документів у систему, необхідно встановити mapping для заданного індексу.