

**Autor:** Piotr Puskarski

**Ćwiczenie:** 2-fts

## Rezultat

**Punkt 3 i 4:** analizator i indeks przechowujący orzeczenie:

```
index_body = {
    'settings': {
        'analysis': {
            'analyzer': {
                'my_analyzer': {
                    'type': 'custom',
                    'tokenizer': 'standard',
                    'filter': ['morfologik_stem', 'lowercase']
                }
            }
        },
        'mappings': {
            'judgment': {
                'properties': {
                    'text': {'type': 'text', 'analyzer': 'my_analyzer'},
                    'judgmentDate': {'type': 'date'},
                    'caseNumber': {'type': 'keyword', },
                    'judges.name': {'type': 'keyword'}
                }
            }
        }
    }
}
```

**Punkt 5:** załadowanie danych:

Utworzenie indeksu z ustawieniami takimi jak wyżej:

```
res = es.indices.create(index='lab', body=index_body, ignore=400)
```

Wpisywanie orzeczeń do indeksu:

```
judges = []
for judge in text['judges']:
    judges.append(judge['name'])
judgment_doc = {
    'text': text['textContent'],
    'judgmentDate': parse(text['judgmentDate']),
    'caseNumber': text['courtCases'][0]['caseNumber'],
    'judges.name': judges
}
print(judges)
res = es.index(index='lab', doc_type='judgment', body=judgment_doc, id=id)
```

**Punkt 6:** Liczba orzeczeń zawierających słowo szkoda

```
szkoda_query = {
  "query": {
    "match": {
      "text": "szkoda"
    }
  }
}
```

i wywołanie zapytania:

```
res = es.search(index='lab', doc_type='judgment', body=szkoda_query)
print(res['hits']['total'])
```

Wynik: 829

**Punkt 7:** Liczba orzeczeń zawierających frazę **trwały uszczerbek na zdrowiu**

```
phrase_query = {
  "query": {
    "match_phrase": {
      "text": "trwały uszczerbek na zdrowie"
    }
  }
}
```

Wynik: 9

**Punkt 8:** Takie jak poprzednio ale z maksymalnie 2 dodatkowymi słowami

```
special_phrase_query = {
  "query": {
    "span_near" : {
      "clauses" : [
        {"span_term": {"text" : "trwały"}},
        {"span_term": {"text" : "uszczerbek"}},
        {"span_term": {"text" : "na"}},
        {"span_term": {"text" : "zdrowie"}}
      ],
      "slop": 2,
      "in_order": True
    }
  }
}
```

Wynik: 10

**Punkt 9:** 3 sędziów z największą wydaną liczbą orzeczeń

```
judges_aggs = {
  "size": 0,
  "aggs": {
    "group_by_judge": {
      "terms": {
        "field": "judges.name"
      }
    }
  }
}
```

Wynik:

{'key': 'Dariusz Zawistowski', 'doc\_count': 227}

{'key': 'Teresa Bielska-Sobkowicz', 'doc\_count': 223}

```
{'key': 'Gerard Bieniek', 'doc_count': 220}
```

#### Punkt 10: Liczba orzeczeń w zależności od miesiąca

```
month_aggs = {
  "aggs": {
    "docs_per_month": {
      "date_histogram": {
        "field": "judgmentDate",
        "interval": "month"
      }
    }
  }
}
```

```
{'key_as_string': '2008-01-01T00:00:00.000Z', 'key': 1199145600000, 'doc_count': 410}
{'key_as_string': '2008-02-01T00:00:00.000Z', 'key': 1201824000000, 'doc_count': 392}
{'key_as_string': '2008-03-01T00:00:00.000Z', 'key': 1204329600000, 'doc_count': 390}
{'key_as_string': '2008-04-01T00:00:00.000Z', 'key': 1207008000000, 'doc_count': 404}
{'key_as_string': '2008-05-01T00:00:00.000Z', 'key': 1209600000000, 'doc_count': 314}
{'key_as_string': '2008-06-01T00:00:00.000Z', 'key': 1212278400000, 'doc_count': 374}
{'key_as_string': '2008-07-01T00:00:00.000Z', 'key': 1214870400000, 'doc_count': 347}
{'key_as_string': '2008-08-01T00:00:00.000Z', 'key': 1217548800000, 'doc_count': 206}
{'key_as_string': '2008-09-01T00:00:00.000Z', 'key': 1220227200000, 'doc_count': 302}
{'key_as_string': '2008-10-01T00:00:00.000Z', 'key': 1222819200000, 'doc_count': 433}
{'key_as_string': '2008-11-01T00:00:00.000Z', 'key': 1225497600000, 'doc_count': 432}
{'key_as_string': '2008-12-01T00:00:00.000Z', 'key': 1228089600000, 'doc_count': 348}
```

