

# Information Retrieval Seminar Work

2300341081

Gwang Won Seo

1. About Document
2. Using Tools
3. Procedure
4. Limitation

# 1.About Document

Document was downloaded from this [link](#).  
It has the information about price of  
Mercedes-benz austria edition in 2024.

There are 17 kinds of cars. The total number  
of options of each car is 45.

## 2.Using Tools

- Python

: It supports multiple programming  
paradigms, has a vast standard library, and  
is widely used in web development, data  
science, AI, and automation.

- Jupyter Notebook

: It is a web-based tool for interactive coding  
and documentation. It lets you combine live  
code, visualizations, and text in a  
collaborative environment, commonly used  
in data science with a focus on Python.

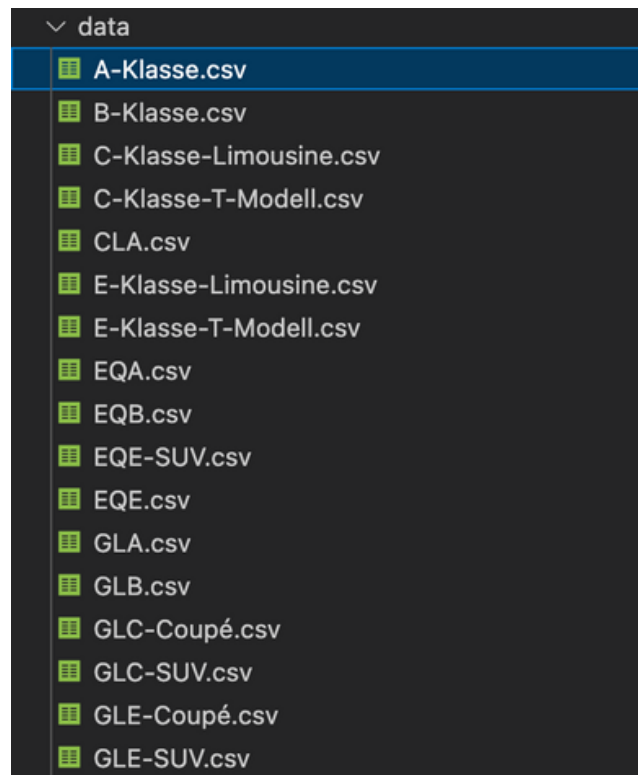
- Apache Solr

: It is an open-source search platform based on Apache Lucene. It enables efficient indexing and searching of content, commonly used for applications requiring powerful search capabilities, such as in e-commerce and content management systems.

## 3.Procedure

### 1) pdf to csv

1. Parse the file using 'tika' module.
2. Create dictionary for each model.
3. Extract information using 'getSome' module which I made or by hand.
4. Export dictionary using 'pandas' module.



<result of 'pdf to csv'>

## 2) Indexing data

1. Start Solrcloud.
2. Create collection.
3. Generate Catchcall Copyfield.
4. Index the data.

The screenshot shows the Solr Admin interface. On the left is a sidebar with navigation links: Dashboard, Logging, Security, Cloud, Schema Designer, Collections, Java Properties, Thread Dump, and a dropdown menu for 'benz-price-list...'. The main panel displays the 'Query' tab for the 'benz-price-list-new' collection. The request handler is set to '/select'. The query is 'q=\*:\*'. The operation is 'q.op=OR'. The facet is 'indented on'. The response is a JSON object showing search results for 'GLA' and 'GLA 200 d 4MATIC'.

```

{
  "responseHeader": {
    "zkConnected": true,
    "status": 0,
    "QTime": 41,
    "params": {
      "q": "*:*",
      "indent": "true",
      "q.op": "OR",
      "useParams": "",
      "_: "1706140157075"
    }
  },
  "response": {
    "numFound": 45,
    "start": 0,
    "maxScore": 1.0,
    "numFoundExact": true,
    "docs": [
      {
        "Name": ["GLA"],
        "Option": ["GLA 180 "],
        "id": ["355ea960-1d26-4168-bb48-1e32c5f1541d"],
        "Torque_Nm": [230],
        "Engine_Zylinder": ["Benziner/R4"],
        "Displacement_cm3": [1332],
        "Fuel_Consumption_l_100_km": ["7,3 - 6,7"],
        "CO2_Emissions_g_km": ["165 - 152"],
        "Price_euro": [42790],
        "version": "1789017614695006208"
      },
      {
        "Name": ["GLA"],
        "Option": ["GLA 200 d 4MATIC"],
        "id": ["ef351123-306e-4ab5-80ce-61419302fac5"],
        "Torque_Nm": [320],
        "Engine_Zylinder": ["Diesel/R4"],
        "Displacement_cm3": [1950]
      }
    ]
  }
}

```

<result of 'Indexing data'>

### 3) Searching

1. Import 'pysolr' module.
2. Search using '.search' method.
3. Get answer using 'answer' module which I made.

```

For ['EQA 250 ']
  Battery Capacity is [66.5]kWh
  Max Torque is [385]
  Drive Range is ['457 - 525']km
  Price is [49990]euro
For ['EQA 250+ ']
  Battery Capacity is [70.5]kWh
  Max Torque is [385]
  Drive Range is ['497 - 559']km
  Price is [52990]euro
For ['EQA 300 4MATIC']
  Battery Capacity is [66.5]kWh
  Max Torque is [390]
  Drive Range is ['412 - 457']km
  Price is [54990]euro

```

<result of 'Searching'>

## 4.Limitation

- I didn't host it and make webpage for general users.