

ELS: Checkpoint 1

GROUP 3

**Eduardo Luís Tronjo Ramos (up201906732@up.pt);
Fábio Araújo de Sá (up202007658@up.pt);
Pedro Pereira Ferreira (up202004986@up.pt);**

INDEX



- 03 INTRODUCTION
- 04 CONFIGURATION FILE
- 05 DSL ARCHITECTURE
- 06 FEATURES
- 07 KNOWN ISSUES & LIMITATIONS
- 08 USER PROFILE
- 09 DEMO
- 10 CONCLUSIONS & FUTURE WORK

INTRODUCTION

- It was proposed to the group create a basic version of a Domain-Specific Language (DSL) for extracting, manipulating and writing of tabular data from arbitrary sources, such as JSON, YAML, or XML files.
- This DSL reads a configuration file which specifies how to import, process, and export data from the tables extracted from the data sources.



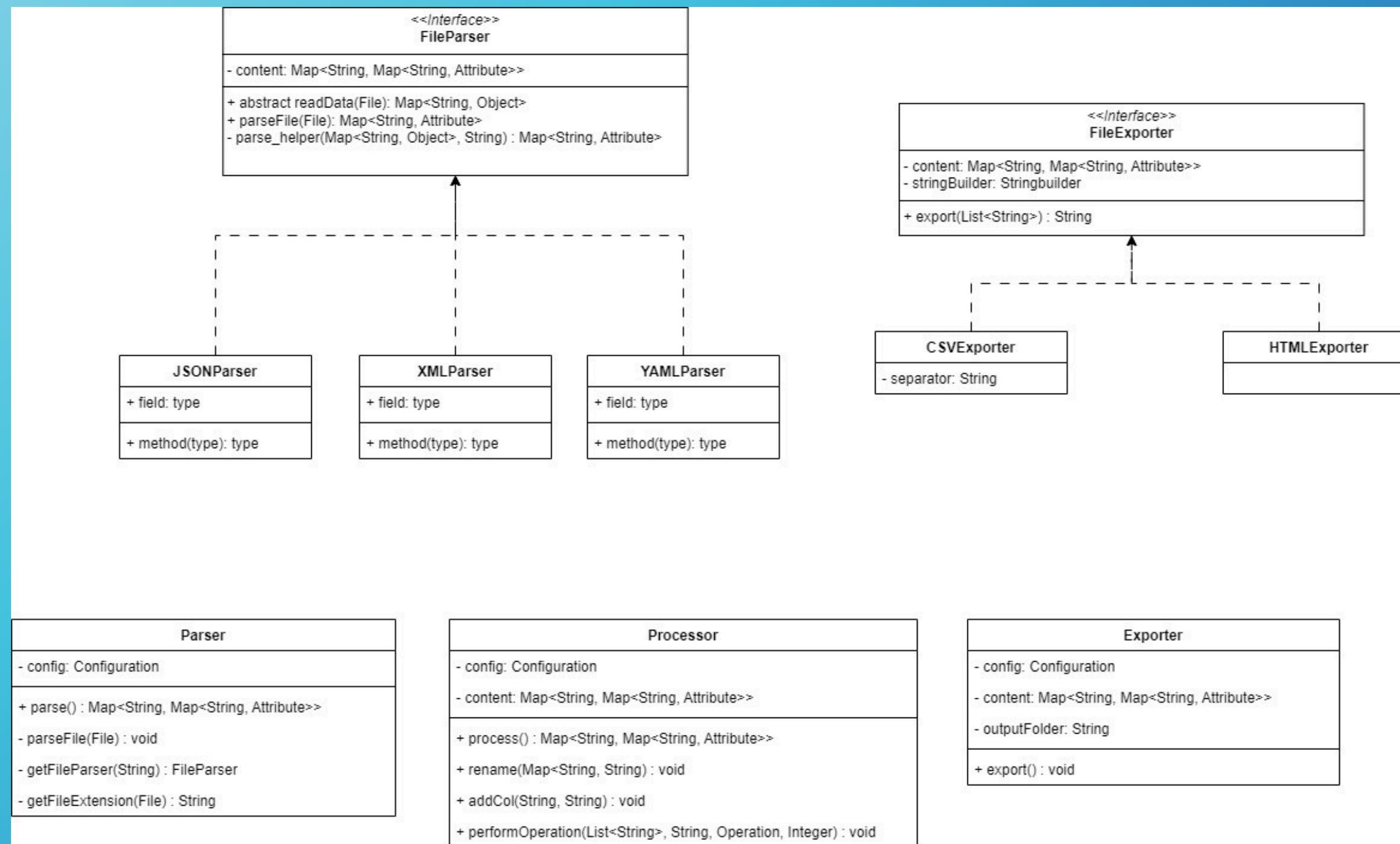
CONFIGURATION FILE

```
"general": {  
    "input_folder": "resources/assignment_1/input/",  
    "output_folder": "resources/assignment_1/output/"  
},  
"process": [  
    {  
        "type": "Rename",  
        "args": {  

```

```
"export": [  
    {  
        "columns": [  
            {  
                "name": "File Name"  
            },  
            {  
                "name": "Criterion"  
            },  
            {  
                "name": "Splitter"  
            },  
            {  
                "name": "CPP Alpha"  
            },  
            {  
                "name": "Min Samples Split"  
            },  
            {  
                "name": "n_classes_"  
            }  
        ],  
        "output_file": "output.csv"  
    }  
]
```

DSL ARCHITECTURE



FEATURES

- Multiple tables extracted from different input file formats (JSON, YAML, XML);
- Different kind of operations, such as:
 - Summing numbers;
 - Subtracting numbers;
 - Multiplying numbers;
 - Dividing numbers;
 - Adding a new column;
 - Renaming column;
- Columns/tables exportation in HTML or CSV formats;



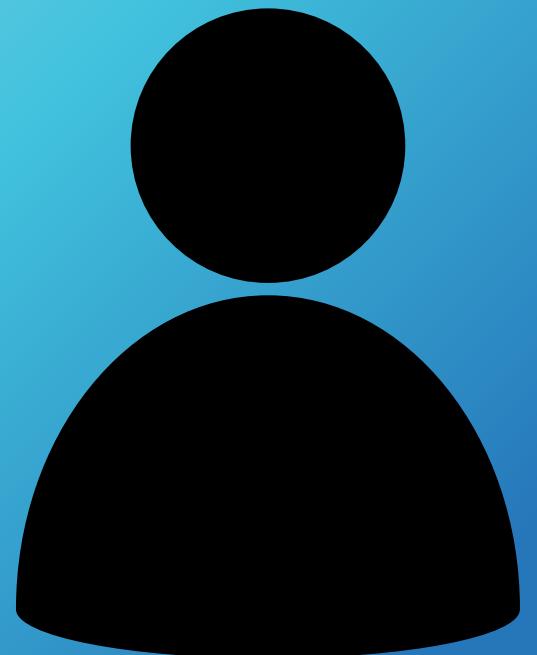
KNOWN ISSUES & LIMITATIONS

- Table-to-table operations are not supported
- Lack of complex data operations
- User is forced to grapple with the JSON file
- The JSON configuration does not allow users to define complex behaviors or calculations through syntax
- Lack of importing and exporting options
- It is not allowed two nested attributes having the same name, despite having different parent attributes



USER PROFILE

For the DSL implemented the user does not need to have an expertise knowledge of programming, since they don't need program in the configuration file. However, in order to make the full use of it, the final user must have a deep knowledge of file's structure and syntax.





DEMO



CONCLUSIONS & FUTURE WORK

- Even though the DSL is completed and has more features than the minimum required, the group has considered that this approach (that is, using the configuration file) is not the best one, due to the stricted and limited syntax the file has to work properly.
- However, this is a good opportunity to create an internal DSL, where the user can program and have a more complex and specific syntax to manipulate the data present on the tables, as well as to parse and extract only the tables/data desired from the input files.
- In addition to that, more operations are intended to me implemented in the next checkpoints, such as the filtering of tables, the sorting of columns, grouping columns by an attribute, aggregating columns, as well as joining tables.