AQL Subset

实现的功能

ps：语法请阅读AQLsubset.pdf

- output 输出表格（已存在的）
- create 建表
  - extract regex 根据正则表达式从文章中提取span
  - extract pattern 从多个表中组合出符合patterns格式的span
  - select 从一个表中选择特定的列（可以是多个列），组合成新表。

补充说明

- token的定义：以字母或数字组成的无符号分隔的字符串，或单纯的特殊符号，不包含空白符（blank）
- span的定义：由正则表达式提取出来的字符串，并且必须是**空格和token的组合**，**不能有不完整的**token。
  - **举例**：
    目标内容是：

    ```
    Carter from Plains, Georgia, Washington from Westmoreland, Virgin
    ia
    ```

    - extract regex 提取到：

      ```
      Carter from Plains, Georgia
      ```

      是会被加入到列中的。
    - extract regex 提取到：

      ```
      Carter from Plains, Georg
      ```

      是**不会**被加入到列中的，会被丢弃。
- 基于第二点，extract pattern 提取到的同样也应当是**空格和token的组合**。

运行

- **命令行**

  ```
  make run
  ```

- **输入参数**
  第一行输入aql文件路径，第二行输入input文件/夹路径。
  示例：

- 处理单个文件

```
../dataset/invest.aql
../dataset/invest1.input
```

- 处理文件夹

```
../dataset/invest.aql
../dataset
```