



Relatório

Processamento de Linguagem

Licenciatura em Engenharia de Sistemas Informáticos 3ºano

Covid19 - Kettle

Alunos:
Francisco Ferreira – 14061
João Oliveira – 16979

Professor: Luís Ferreira

Barcelos, novembro, 2020

Índice

1. Introdução	3
2. Introdução ao problema	3
3. Importar os dados.....	4
4. Tratamento de dados/ Guardar dados.....	5
5. API.....	5
6. Email.....	6
7. Base de Dados	8
8. Conclusão.....	9
9. Bibliografia.....	9

1. Introdução

Estamos a viver numa nova realidade, está a ser um ano atípico devido ao covid 19. Por ser algo recente e com muitos dados, e como o tema deste trabalho era à nossa escolha, decidimos utilizar a covid19 como tema. Utilizando assim as bases de dados e API'S que estão ao nosso alcance decidimos fazer o tratamento e filtragem de dados para aquilo que é o nosso objetivo.

2. Introdução ao problema

A Johns Hopkins University Center for Systems Science and Engineering tem uma base de dados que é publica, onde estão todos os casos desde o primeiro dia em que cada país teve um caso de covid. Ao todo esta base dados tem 190 países e todo o histórico de covid nesse mesmo país. Como são muitos dados e não fazia sentido estar a exportar toda a base dados e todos os csv, vamos utilizar uma api que se chama covid19api.

O Nosso objetivo é então filtrar toda a informação para só trazer informação de um país, neste caso Portugal, para depois ser guardado em XML e JSON.

No fim de toda a filtragem será enviado um email com todos os casos, como por exemplo os novos confirmados, as novas mortes, os novos recuperados, o total de confirmados, o total de mortes e por fim o total de recuperados.

3. Importar os dados

Nesta primeira fase importamos os dados da api com um Get deste link

<https://api.covid19api.com/summary>, que nos traz o resumo dos casos novos e do total por país atualizado diariamente

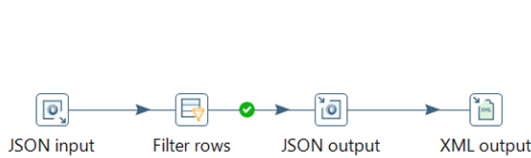
```
{
  "Message": "",
  "Global": {
    "NewConfirmed": 665991,
    "TotalConfirmed": 52119182,
    "NewDeaths": 12240,
    "TotalDeaths": 1284482,
    "NewRecovered": 382236,
    "TotalRecovered": 33919628
  },
  "Countries": [{
    "Country": "Afghanistan",
    "CountryCode": "AF",
    "Slug": "afghanistan",
    "NewConfirmed": 146,
    "TotalConfirmed": 42609,
    "NewDeaths": 4,
    "TotalDeaths": 1581,
    "NewRecovered": 13,
    "TotalRecovered": 34967,
    "Date": "2020-11-12T17:27:30Z",
    "Premium": {}
  ]
},
```

Podemos concluir que com este GET que retornou o código 200 OK,

- 1) Global, em todo o mundo diariamente.
 - a) Casos novos confirmados (*NewConfirmed*)
 - b) Casos totais confirmados (*TotalConfirmed*)
 - c) Novas mortes (*NewDeaths*)
 - d) Total de mortes (*TotalDeaths*)
 - e) Novos recuperados (*NewRecovered*)
 - f) Total recuperados (*TotalRecovered*)
- 2) Casos por país diariamente.
 - a) País (*Country*)
 - b) Código de país (*CountryCode*)
 - c) Novas mortes (*NewDeaths*)
 - d) Total de mortes (*TotalDeaths*)
 - e) Novos recuperados (*NewRecovered*)
 - f) Total recuperados (*TotalRecovered*)
 - g) Data (*Date*)

Depois de compreendermos os dados que recebemos esta na hora de os filtrar

4. Tratamento de dados / Guardar Dados



Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

(String)

Usamos um Json input para receber o ficheiro com a informação mundial sobre o covid19 e filtramos por continente com uma expressão regular, consoante o país escolhido mais à frente. Depois de essa expressão regular ser executada vai ser gerado um ficheiro Jason e em seguida convertido para XML.

```

{"data":{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"162.58929922","popData2019":2862427,
"cases":489,"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"07/11/2020",
"continentExp":"Europe","day":"07","countriesAndTerritories":"Albania","deaths":6},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"156.1961308","popData2019":2862427,"cases":421,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"06/11/2020",
"continentExp":"Europe","day":"06","countriesAndTerritories":"Albania","deaths":7},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"152.8381182","popData2019":2862427,"cases":396,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"05/11/2020",
"continentExp":"Europe","day":"05","countriesAndTerritories":"Albania","deaths":4},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"148.58620833","popData2019":2862427,"cases":381,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"04/11/2020",
"continentExp":"Europe","day":"04","countriesAndTerritories":"Albania","deaths":5},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"145.78537723","popData2019":2862427,"cases":321,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"03/11/2020",
"continentExp":"Europe","day":"03","countriesAndTerritories":"Albania","deaths":9},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"144.87705713","popData2019":2862427,"cases":327,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"02/11/2020",
"continentExp":"Europe","day":"02","countriesAndTerritories":"Albania","deaths":9},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"143.27082924","popData2019":2862427,"cases":241,
"countryterritoryCode":"ALB","month":"11","year":"2020","geoId":"AL","dateRep":"01/11/2020",
"continentExp":"Europe","day":"01","countriesAndTerritories":"Albania","deaths":7},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"144.38796168","popData2019":2862427,"cases":319,
"countryterritoryCode":"ALB","month":"10","year":"2020","geoId":"AL","dateRep":"31/10/2020",
"continentExp":"Europe","day":"31","countriesAndTerritories":"Albania","deaths":3},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"143.33990802","popData2019":2862427,"cases":275,
"countryterritoryCode":"ALB","month":"10","year":"2020","geoId":"AL","dateRep":"30/10/2020",
"continentExp":"Europe","day":"30","countriesAndTerritories":"Albania","deaths":6},
{"Cumulative_number_for_14_days_of_COVID-19_cases_per_100000":"142.71186302","popData2019":2862427,"cases":311,
"countryterritoryCode":"ALB","month":"10","year":"2020","geoId":"AL","dateRep":"29/10/2020",
"continentExp":"Europe","day":"29","countriesAndTerritories":"Albania","deaths":6}

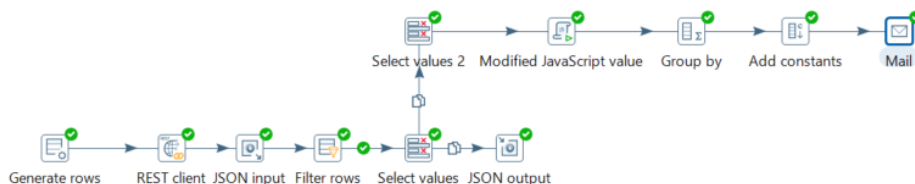
```

```

1 <?xml version='1.0' encoding='UTF-8' ?>
2 <Rows>
3 <Row><dateRep>07/11/2020/</dateRep><day>07/</day><month>11/</month><year>2020/</year><cases>489/</cases><deaths>6/</deaths>
4 <Row><dateRep>06/11/2020/</dateRep><day>06/</day><month>11/</month><year>2020/</year><cases>421/</cases><deaths>7/</deaths>
5 <Row><dateRep>05/11/2020/</dateRep><day>05/</day><month>11/</month><year>2020/</year><cases>396/</cases><deaths>4/</deaths>
6 <Row><dateRep>04/11/2020/</dateRep><day>04/</day><month>11/</month><year>2020/</year><cases>381/</cases><deaths>5/</deaths>
7 <Row><dateRep>03/11/2020/</dateRep><day>03/</day><month>11/</month><year>2020/</year><cases>321/</cases><deaths>9/</deaths>
8 <Row><dateRep>02/11/2020/</dateRep><day>02/</day><month>11/</month><year>2020/</year><cases>327/</cases><deaths>9/</deaths>
9 <Row><dateRep>01/11/2020/</dateRep><day>01/</day><month>11/</month><year>2020/</year><cases>241/</cases><deaths>7/</deaths>
10 <Row><dateRep>31/10/2020/</dateRep><day>31/</day><month>10/</month><year>2020/</year><cases>319/</cases><deaths>3/</deaths>
11 <Row><dateRep>30/10/2020/</dateRep><day>30/</day><month>10/</month><year>2020/</year><cases>275/</cases><deaths>6/</deaths>
12 <Row><dateRep>29/10/2020/</dateRep><day>29/</day><month>10/</month><year>2020/</year><cases>311/</cases><deaths>6/</deaths>
13 <Row><dateRep>28/10/2020/</dateRep><day>28/</day><month>10/</month><year>2020/</year><cases>284/</cases><deaths>3/</deaths>
14 <Row><dateRep>27/10/2020/</dateRep><day>27/</day><month>10/</month><year>2020/</year><cases>288/</cases><deaths>3/</deaths>
15 <Row><dateRep>26/10/2020/</dateRep><day>26/</day><month>10/</month><year>2020/</year><cases>299/</cases><deaths>4/</deaths>
16 <Row><dateRep>25/10/2020/</dateRep><day>25/</day><month>10/</month><year>2020/</year><cases>302/</cases><deaths>4/</deaths>
17 <Row><dateRep>24/10/2020/</dateRep><day>24/</day><month>10/</month><year>2020/</year><cases>306/</cases><deaths>4/</deaths>
18 <Row><dateRep>23/10/2020/</dateRep><day>23/</day><month>10/</month><year>2020/</year><cases>302/</cases><deaths>3/</deaths>
19 <Row><dateRep>22/10/2020/</dateRep><day>22/</day><month>10/</month><year>2020/</year><cases>301/</cases><deaths>4/</deaths>
20 <Row><dateRep>21/10/2020/</dateRep><day>21/</day><month>10/</month><year>2020/</year><cases>301/</cases><deaths>4/</deaths>
21 <Row><dateRep>20/10/2020/</dateRep><day>20/</day><month>10/</month><year>2020/</year><cases>295/</cases><deaths>4/</deaths>
22 <Row><dateRep>19/10/2020/</dateRep><day>19/</day><month>10/</month><year>2020/</year><cases>281/</cases><deaths>3/</deaths>
23 <Row><dateRep>18/10/2020/</dateRep><day>18/</day><month>10/</month><year>2020/</year><cases>273/</cases><deaths>5/</deaths>
24 <Row><dateRep>17/10/2020/</dateRep><day>17/</day><month>10/</month><year>2020/</year><cases>289/</cases><deaths>4/</deaths>
25 <Row><dateRep>16/10/2020/</dateRep><day>16/</day><month>10/</month><year>2020/</year><cases>257/</cases><deaths>5/</deaths>
26 <Row><dateRep>15/10/2020/</dateRep><day>15/</day><month>10/</month><year>2020/</year><cases>203/</cases><deaths>5/</deaths>
27 <Row><dateRep>14/10/2020/</dateRep><day>14/</day><month>10/</month><year>2020/</year><cases>182/</cases><deaths>5/</deaths>
28 <Row><dateRep>13/10/2020/</dateRep><day>13/</day><month>10/</month><year>2020/</year><cases>171/</cases><deaths>4/</deaths>
29 <Row><dateRep>12/10/2020/</dateRep><day>12/</day><month>10/</month><year>2020/</year><cases>168/</cases><deaths>4/</deaths>
30 <Row><dateRep>11/10/2020/</dateRep><day>11/</day><month>10/</month><year>2020/</year><cases>165/</cases><deaths>3/</deaths>

```

5. API



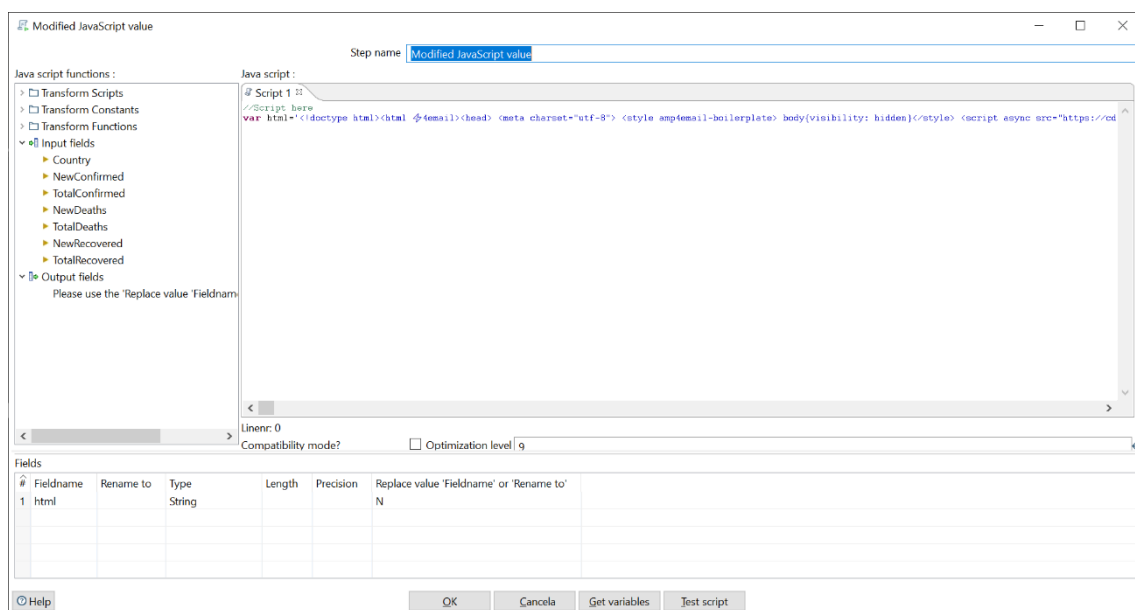
Neste passo, e numa transformação diferente, fomos à seguinte api

<https://api.covid19api.com/summary> com o objetivo de obtermos só um determinado país, como referido já anteriormente. Depois fazemos a filtragem para obter todos os dados mas apenas desse determinado país, que no nosso caso é Portugal, é gerado um Json output.

6. Email



Em simultâneo à criação do ficheiro Jason referido anteriormente criamos um Modified JavaScript pois para enviar um email com html embutido, com os valores vindo da api, é necessário criar um script em JavaScript, para assim criarmos variáveis e colocar todo o template para o email.



Em seguida adicionamos constantes que vão servir para armazenar os dados para enviar o email.

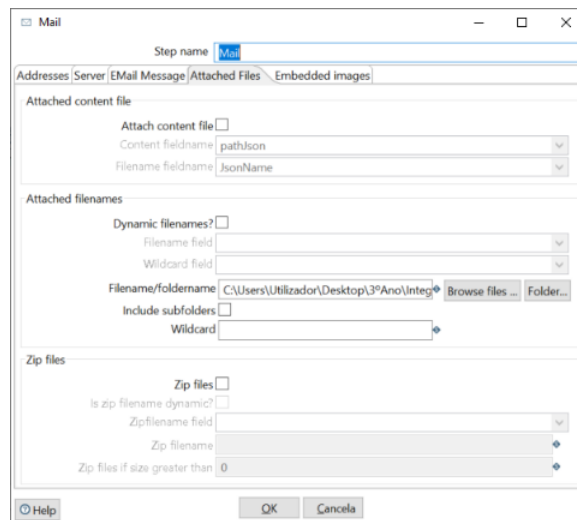
Nome do Step: Add constants

Campos:

#	Nome	Tipo	Formato	Tamanho	Precisão	Moeda	Decimal	Grupo	Valor	Set empty
1	email	String							a16979@alunos.ipca.pt	N
2	nome	String							João	N
3	enviadopor	String							Isi_Covid19	N
4	smtp	String							smtp.gmail.com	N
5	porta	String							465	N
6	user	String							isikettle@gmail.com	N
7	pass	String							Abcd1234!	N
8	pathJson	String							C:\Users\Utilizador\Desktop\3ºAno\Integração de Sistemas de Informação\Trabalhos\Isi_Covid_Project\Output_0	N
9	JsonName	String							Continent Json File	N

Buttons: Help, OK, Cancela

No email decidimos também enviar informação relativa ao continente em que esse país se encontra, anexando assim o ficheiro Json criado na primeira transformação.



Assim sendo o nosso email contém a informação sobre o país e também um anexo sobre o continente desse país para comprar dados.

isikettle@gmail.com

Isi_Covid19 <isikettle@gmail.com>
15:50

Para: João Pedro Araújo de Oliveira

Continent_0.json
4,11 MB

ISI Portugal

Boletim covid-19

Relatorio do dia de hoje de covid-19 em portugal

Novos Confirmados	6653
Novas mortes	69
Novos Recuperados	3693
Total Confirmados	204664
Total Mortes	3250
Total Recuperados	117382

Trabalho de grupo

7. Base de dados



Depois de ter em nossa mão estes dados todos filtrados decidimos fazer a nossa própria base de dados, criamos uma tabela que se chama país com o Nome do país, data, casos Ativos, casos recuperados, mortes e o Total confirmado. Depois de criada conectamos a base dados no Kettle fazemos o Insert/update para inserirmos os dados na tabela.

	Country	Date	Active	Recovered	Deaths	Confirmed
	Filter	Filter	Filter	Filter	Filter	Filter
40	Portugal	2020-03-01T00:00:00Z	0	0	0	0
41	Portugal	2020-03-02T00:00:00Z	2	0	0	2
42	Portugal	2020-03-03T00:00:00Z	2	0	0	2
43	Portugal	2020-03-04T00:00:00Z	5	0	0	5
44	Portugal	2020-03-05T00:00:00Z	8	0	0	8
45	Portugal	2020-03-06T00:00:00Z	13	0	0	13
46	Portugal	2020-03-07T00:00:00Z	20	0	0	20
47	Portugal	2020-03-08T00:00:00Z	30	0	0	30
48	Portugal	2020-03-09T00:00:00Z	30	0	0	30
49	Portugal	2020-03-10T00:00:00Z	41	0	0	41
50	Portugal	2020-03-11T00:00:00Z	59	0	0	59
51	Portugal	2020-03-12T00:00:00Z	59	0	0	59
52	Portugal	2020-03-13T00:00:00Z	111	1	0	112
53	Portugal	2020-03-14T00:00:00Z	167	2	0	169
54	Portugal	2020-03-15T00:00:00Z	243	2	0	245
55	Portugal	2020-03-16T00:00:00Z	328	3	0	331
56	Portugal	2020-03-17T00:00:00Z	444	3	1	448
57	Portugal	2020-03-18T00:00:00Z	443	3	2	448
58	Portugal	2020-03-19T00:00:00Z	779	3	3	785

8. Conclusão

Com este trabalho conseguimos aprofundar a matéria dada na nesta unidade curricular, achamos também que o objetivo foi alcançado. Como fizemos o trabalho em grupo isso obrigou-nos a usar dois programas o que é muito bom pois existem comandos e maneiras diferentes de fazer as coisas , mesmo sendo o mesmo tema.

9. Bibliografia

<https://api.covid19api.com/summary>

<https://stackoverflow.com/questions/49083778/send-html-formatted-mail-from-pentaho>

<https://community.hitachivantara.com/s/article/calling-hci-rest-api-from-pentaho-kettle>

Documentação dada na unidade curricular