

Relatório

Integração de Sistemas de Informação

Licenciatura em Engenharia de Sistemas Informáticos 3ºano

Ano letivo 2020/2021

Licenciatura em Engenharia de Sistemas Informáticos

Escola Superior de Tecnologia

Instituto Politécnico do Cávado e do Ave

Identificação do Aluno

José Pedro Laranjeira Alves

Aluno número 16972, regime diurno

Licenciatura em Engenharia de Sistemas Informáticos

Orientação

Dr. Luís Ferreira

RESUMO

Este relatório apresenta-se como objeto de avaliação para o docente da disciplina de Integração de Sistemas de Informação, onde é apresentada uma visão sobre o trabalho realizado.

Índice

| | |
|-----------------------|----|
| 1. Introdução..... | 5 |
| 2. Transformação..... | 6 |
| 3. Job..... | 12 |
| 4. Conclusão..... | 17 |

1. Introdução

1.1. Objetivos

- Consolidar conceitos associados à Integração de Sistemas de Informação usando Dados;
- Analisar e especificar cenários de aplicação de processos de ETL;
- Explorar ferramentas de suporte a processos de ETL;
- Explorar novas Tecnologias, Frameworks ou Paradigmas;
- Potenciar a experiência no desenvolvimento de software;
- Facilitar a assimilação do conteúdo da Unidade Curricular

1.2. Contexto

Relatório do Trabalho Prático para a disciplina de Integração de Sistemas de Informação, que tem por base explicar a forma de como o trabalho foi efetuado.

1.3. Estrutura do documento

O documento está dividido em 2 secções principais, cada uma delas explica o funcionamento de cada tipo de ficheiro e sua demonstração/ explicação de como foi realizado.

2. Transformação

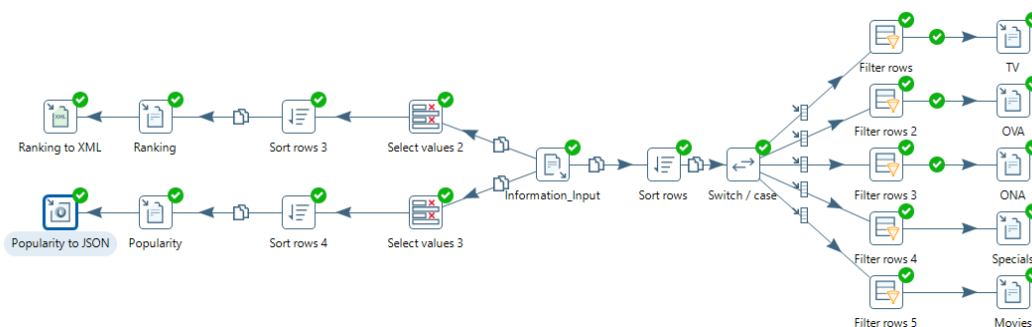


Imagem 1 – Transformação criada para o trabalho

2.1. Ler Ficheiro

O ficheiro anime.csv foi um ficheiro adquirido no site: <https://www.kaggle.com/datasets>.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | anime_id,name,genre,type,episodes,rating,members | | | | | | | | | | | | | |
| 2 | 32281,Kimi no Na wa,"Drama, Romance, School, Supernatural",Movie,1,9.37,200630 | | | | | | | | | | | | | |
| 3 | 5114,Fullmetal Alchemist: Brotherhood,"Action, Adventure, Drama, Fantasy, Magic, Military, Shounen",TV,64,9.26,793665 | | | | | | | | | | | | | |
| 4 | 28977,GintamaA,"Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen",TV,51,9.25,114262 | | | | | | | | | | | | | |
| 5 | 9253,Steir Gate,"Sci-Fi, Thriller",TV,24,9.17,673572 | | | | | | | | | | | | | |
| 6 | 9969,Gint,"Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen",TV,51,9.16,151266 | | | | | | | | | | | | | |
| 7 | 32935,Haikyuu!! Karasuno Koukou VS Shiratorizawa Gakuen Koukou,"Comedy, Drama, School, Shounen, Sports",TV,10,9.15,93351 | | | | | | | | | | | | | |
| 8 | 11061,Hunter x Hunter (2011),"Action, Adventure, Shounen, Super Power",TV,148,9.13,425855 | | | | | | | | | | | | | |
| 9 | 820,Ginga Eiyuu Densetsu,"Drama, Military, Sci-Fi, Space",OVA,110,9.11,80679 | | | | | | | | | | | | | |
| 10 | 15335,Gintama Movie: Kanketsu-hen - Yorozuya yo Eien Nare,"Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen",Movie,1,9.10,72534 | | | | | | | | | | | | | |
| 11 | 15417,Gin: Enchousen,"Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen",TV,13,9.11,81109 | | | | | | | | | | | | | |
| 12 | 4181,Clannad: After Story,"Drama, Fantasy, Romance, Slice of Life, Supernatural",TV,24,9.06,456749 | | | | | | | | | | | | | |
| 13 | 28851,Koe no Katachi,"Drama, School, Shounen",Movie,1,9.05,102733 | | | | | | | | | | | | | |
| 14 | 918,Gintama,"Action, Comedy, Historical, Parody, Samurai, Sci-Fi, Shounen",TV,201,9.04,336376 | | | | | | | | | | | | | |
| 15 | 2904,Code Geass: Hangyaku no Lelouch R2,"Action, Drama, Mecha, Military, Sci-Fi, Super Power",TV,25,8.98,572888 | | | | | | | | | | | | | |
| 16 | 28891,Haikyuu!! Second Season,"Comedy, Drama, School, Shounen, Sports",TV,25,8.93,179342 | | | | | | | | | | | | | |
| 17 | 199,Sen to Chihiro no Kamikakushi,"Adventure, Drama, Supernatural",Movie,1,8.93,466254 | | | | | | | | | | | | | |
| 18 | 23273,Shigatsu wa Kimi no Uso,"Drama, Music, Romance, School, Shounen",TV,22,8.92,416397 | | | | | | | | | | | | | |
| 19 | 24701,Mushishi Zoku Shou 2nd Season,"Adventure, Fantasy, Historical, Mystery, Seinen, Slice of Life, Supernatural",TV,10,8.88,75894 | | | | | | | | | | | | | |
| 20 | 12355,Ookami Kodomo no Ame to Yuki,"Fantasy, Slice of Life",Movie,1,8.84,226193 | | | | | | | | | | | | | |
| 21 | 1575,Code Geass: Hangyaku no Lelouch,"Action, Mecha, Military, School, Sci-Fi, Super Power",TV,25,8.83,715151 | | | | | | | | | | | | | |
| 22 | 263,Hajime no Ippo,"Comedy, Drama, Shounen, Sports",TV,75,8.83,157670 | | | | | | | | | | | | | |
| 23 | 44,Rurouni Kenshin: Meiji Kenkaku Romantan - Tsukoku-hen,"Action, Drama, Historical, Martial Arts, Romance, Samurai",OVA,4,8.83,129307 | | | | | | | | | | | | | |
| 24 | 1,Cowboy Bebop,"Action, Adventure, Comedy, Drama, Sci-Fi, Space",TV,26,8.82,486824 | | | | | | | | | | | | | |

Imagem 2 – Dados do ficheiro anime.csv

De modo a fazer a leitura do ficheiro no Kettle, usando a opção CSV Input, introduzimos o ficheiro que contém a informação através da sua identificação e localização na máquina.

Step name: Information_Input

Filename: \${Internal.Entry.Current.Directory}/Input/anime.csv

Delimiter: ,

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding: UTF-8

| # | Name | Type | Format | Length | Precision | Currency | Decimal | Group | Trim type |
|---|----------|---------|--------|--------|-----------|----------|---------|-------|-----------|
| 1 | anime_id | Integer | # | 15 | 0 | € | , | . | none |
| 2 | name | String | | 57 | | € | , | . | none |
| 3 | genre | String | | 76 | | € | , | . | none |
| 4 | type | String | | 7 | | € | , | . | none |
| 5 | episodes | String | | 7 | | € | , | . | none |
| 6 | rating | Number | #,. | 15 | 0 | € | . | , | none |
| 7 | members | Integer | # | 15 | 0 | € | , | . | none |

Buttons: ? Help, OK, Obtem campos, Preview, Cancela

Imagem 3 – Extração dos dados usando o Kettle

Após a extração dos dados, estes são ordenados pelo “anime_id” de forma crescente como se pode verificar nas próximas imagens.

Sort rows

Nome do Step:

Sort directory:

TMP-file prefix:

Sort size (rows in memory):

Free memory threshold (in %):

Compress TMP Files? ☒

Only pass unique rows? (verifies keys only) ☐

Fields:

| # | Fieldname | Ascending | Case sensitive compare? | Sort based on current locale? | Collator Strength | Presorted? |
|---|-----------|-----------|-------------------------|-------------------------------|-------------------|------------|
| 1 | anime_id | S | N | N | 0 | N |

Imagem 4 – Escolher o identificador que ordena a informação das colunas

Examine preview data

Rows of step: Sort rows (1000 rows)

| # | anime_id | name | genre | type | episodes | rating | members |
|----|----------|--|--|-------|----------|--------|---------|
| 1 | 1 | Cowboy Bebop | Action, Adventure, Comedy, Drama, Sci-Fi, Space | TV | 26 | 8.8 | 488824 |
| 2 | 5 | Cowboy Bebop: Tengoku no Tobira | Action, Drama, Mystery, Sci-Fi, Space | Movie | 1 | 8.4 | 137636 |
| 3 | 6 | Trigun | Action, Comedy, Sci-Fi | TV | 26 | 8.3 | 283069 |
| 4 | 7 | Witch Hunter Robin | Action, Drama, Magic, Mystery, Police, Supernatural | TV | 26 | 7.4 | 64955 |
| 5 | 8 | Beet the Vandel Buster | Adventure, Fantasy, Shounen, Supernatural | TV | 52 | 7.1 | 9848 |
| 6 | 15 | Eyeshield 21 | Action, Comedy, Shounen, Sports | TV | 143 | 8.1 | 83648 |
| 7 | 16 | Hachimitsu to Clover | Comedy, Drama, Josei, Romance | TV | 24 | 8.2 | 130646 |
| 8 | 17 | Hungry Heart: Wild Striker | Comedy, Shounen, Slice of Life, Sports | TV | 52 | 7.7 | 13469 |
| 9 | 18 | Initial D Fourth Stage | Action, Cars, Drama, Seinen, Sports | TV | 24 | 8.2 | 43384 |
| 10 | 19 | Monster | Drama, Horror, Mystery, Police, Psychological, Seinen, Thriller | TV | 74 | 8.7 | 247562 |
| 11 | 20 | Naruto | Action, Comedy, Martial Arts, Shounen, Super Power | TV | 220 | 7.8 | 683297 |
| 12 | 21 | One Piece | Action, Adventure, Comedy, Drama, Fantasy, Shounen, Super Power | TV | Unknown | 8.6 | 504862 |
| 13 | 22 | Prince of Tennis | Action, Comedy, School, Shounen, Sports | TV | 178 | 8 | 87843 |
| 14 | 23 | Ring ni Kakeru 1 | Action, Shounen, Sports | TV | 12 | 6.7 | 2224 |
| 15 | 24 | School Rumble | Comedy, Romance, School, Shounen | TV | 26 | 8.1 | 178553 |
| 16 | 25 | Sunabouzu | Action, Adventure, Comedy, Ecchi, Sci-Fi, Shounen | TV | 24 | 7.5 | 65516 |
| 17 | 26 | Technopolis | Action, Drama, Psychological, Sci-Fi | TV | 22 | 7.8 | 95559 |
| 18 | 27 | Tonky Blood | Action, Supernatural, Vampire | TV | 24 | 7.4 | 113791 |
| 19 | 28 | Yakitate!! Japan | Comedy, Shounen | TV | 69 | 8.1 | 50876 |
| 20 | 29 | Zipang | Action, Drama, Historical, Military, Sci-Fi, Seinen | TV | 26 | 7.6 | 8630 |
| 21 | 30 | Neon Genesis Evangelion | Action, Dementia, Drama, Mecha, Psychological, Sci-Fi | TV | 26 | 8.3 | 481946 |
| 22 | 31 | Neon Genesis Evangelion: Death & Rebirth | Drama, Mecha, Psychological, Sci-Fi | Movie | 1 | 7.5 | 102093 |
| 23 | 32 | Neon Genesis Evangelion: The End of Evangelion | Dementia, Drama, Mecha, Psychological, Sci-Fi | Movie | 1 | 8.4 | 219330 |
| 24 | 33 | Berserk | Action, Adventure, Demons, Drama, Fantasy, Horror, Military, Romance, Seinen, Supernatural | TV | 25 | 8.4 | 226430 |
| 25 | 40 | Ghost in the Shell | Action, Mecha, Police, Psychological, Sci-Fi, Seinen | Movie | 1 | 8.3 | 223036 |
| 26 | 44 | Rurouni Kenshin: Maji Kenkaku Romantan - Tsukiko-hen | Action, Drama, Historical, Martial Arts, Romance, Samurai | OVA | 4 | 8.8 | 129397 |
| 27 | 45 | Rurouni Kenshin: Maji Kenkaku Romantan | Action, Adventure, Comedy, Historical, Romance, Samurai | TV | 94 | 8.4 | 218928 |
| 28 | 46 | Rurouni Kenshin: Maji Kenkaku Romantan - Ikinoshimi e no Chinkonka | Drama, Historical, Samurai, Shounen | Movie | 1 | 7.7 | 33191 |
| 29 | 47 | Akira | Action, Adventure, Horror, Military, Sci-Fi, Supernatural | Movie | 1 | 8.2 | 215897 |
| 30 | 48 | Jack/J/Sign | Adventure, Fantasy, Game, Magic, Mystery, Sci-Fi | TV | 26 | 7.1 | 110204 |
| 31 | 49 | Aoi Megumi-sama! | Comedy, Magic, Romance, Seinen, Supernatural | OVA | 5 | 7.4 | 34590 |
| 32 | 50 | Aoi Megumi-sama! (TV) | Comedy, Magic, Romance, Seinen, Supernatural | TV | 24 | 7.5 | 96694 |
| 33 | 51 | Tenshi Kinryouku | Action, Drama, Romance, Shounen, Supernatural | OVA | 3 | 6.5 | 36352 |

Imagem 5 – Preview da informação ordenada através do identificador “anime_id”

2.2. Filtragem de dados

A filtragem de dados pode ser realizada através de dois processos, o “Select values” permite selecionar ou remover campos do ficheiro, e o “Filter rows” que através de expressões regulares permite selecionar a informação pretendida.

Filter rows

Step name: Filter rows

Send 'true' data to step: TV

Send 'false' data to step:

The condition:

+

type REGEXP ^([A-Z]){2}\$ (String)

Help OK Cancela

Imagem 6 – Utilização de um “Filter row” para selecionar só os animes que são do tipo (“type”) TV.

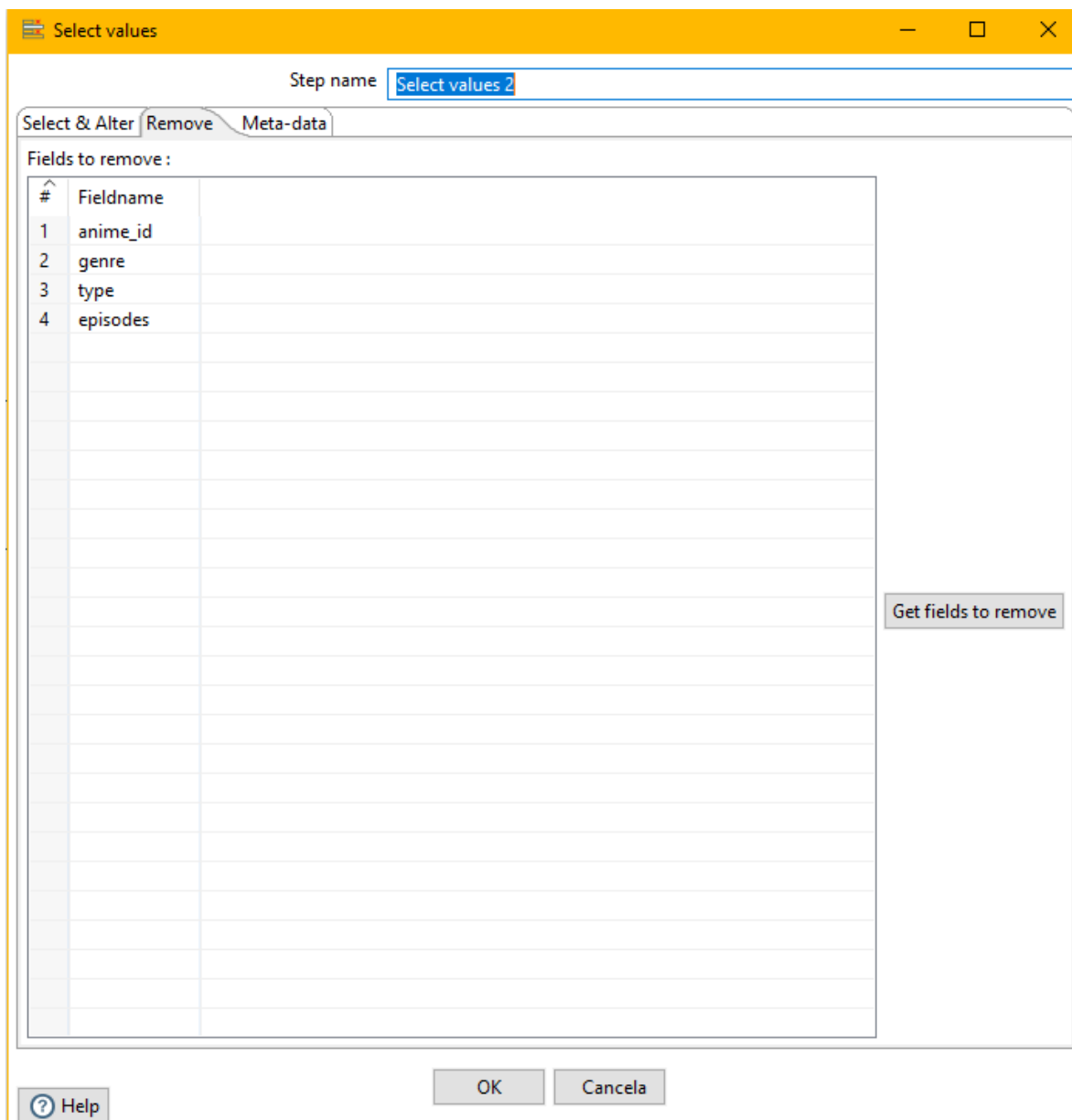


Imagem 7 - Utilização de um “Select values” para remover as colunas que não são necessárias (“anime_id”, “genre”, “type”, “episodes”).

2.3. Conversão dos dados

A conversão dos dados é feita de CSV para um XML e para JSON.

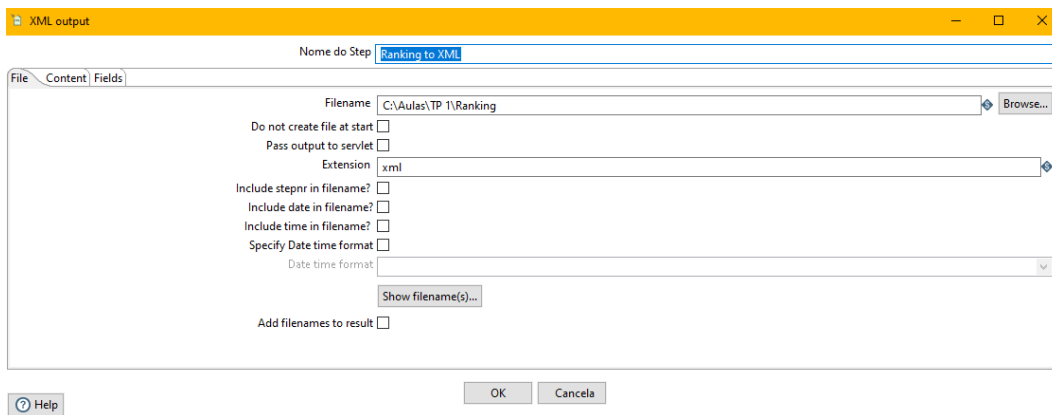


Imagem 8 – Conversão do ficheiro “Ranking.csv” para “Ranking.xml”.

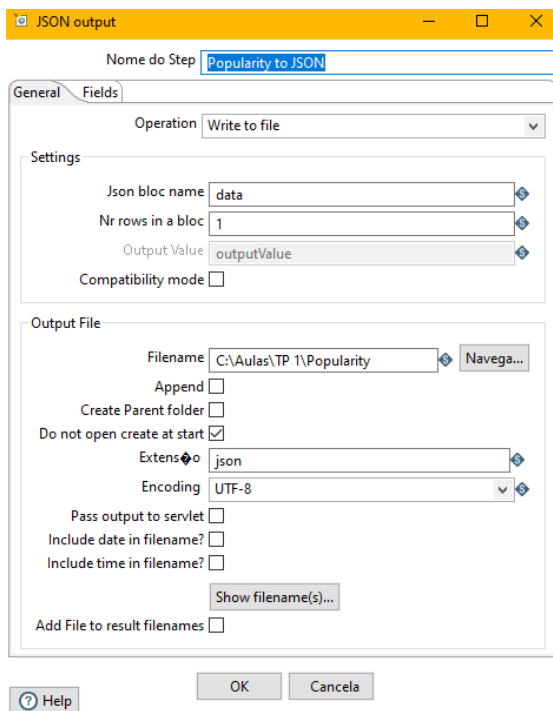


Imagem 9 - Conversão do ficheiro “Popularity.csv” para “Popularity.xml”.

3. Job



Imagem 10 – Job criado para o trabalho

3.1. Configuração

No “Job”, pretende-se a criação de um ficheiro HTML onde estará representada uma tabela com a informação selecionada e “criada” na transformação “TP1_ISI_16972” e para isso é necessário a acompanhar o ficheiro XML criado nessa transformação um outro ficheiro XSL de modo a criar o ficheiro em HTML.

```
<?xml version="1.0"?>
- <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output encoding="ISO-8859-1" method="html"/>
  - <xsl:template match="/">
    - <html>
      - <head>
        <title>TP 1</title>
      </head>
      - <body>
        <h1>Animes</h1>
        - <table border="3">
          - <tr>
            <th>Name</th>
            <th>Rating</th>
            <th>Members</th>
          </tr>
          - <xsl:for-each select="Rows/Row">
            - <tr>
              - <td>
                <xsl:apply-templates select="name"/>
              </td>
              - <td>
                <xsl:apply-templates select="rating"/>
              </td>
              - <td>
                <xsl:apply-templates select="members"/>
              </td>
            </tr>
          </xsl:for-each>
        </table>
      </body>
    </html>
  </xsl:template>
</xsl:stylesheet>
```

Imagem 11 – Ficheiro XSL.

XSL transformation

Job entry name: **XSL transformation**

General | **Advanced**

Files

Get filenames from previous result ☐

XML File name: C:\Aulas\TP 1\Ranking.xml

XSL File name: C:\Aulas\TP 1\Ranking.xsl

Output Filename: C:\Aulas\TP 1\Ranking.html

Settings

Transformer factory: JAXP

If file exists: Do Nothing

Add output filename to result ☐

Imagem 12 – Criação do ficheiro em HTML “Ranking.html”.

Animes

| Name | Rating | Members |
|---|--------|---------|
| Taka no Tsume 8: Yoshida-kun no X-Files | 10 | 13 |
| Spoon-hime no Swing Kitchen | 9.6 | 47 |
| Mogura no Motoro | 9.5 | 62 |
| Kimi no Na wa. | 9.4 | 200630 |
| Kahei no Umi | 9.3 | 44 |
| Fullmetal Alchemist: Brotherhood | 9.3 | 793665 |
| GintamaA° | 9.2 | 114262 |
| Yakusoku: Africa Mizu to Midori | 9.2 | 53 |
| Steins;Gate | 9.2 | 673572 |
| Gintama' | 9.2 | 151266 |
| Haikyuu!!: Karasuno Koukou VS Shiratorizawa Gakuen Koukou | 9.2 | 93351 |
| Hunter x Hunter (2011) | 9.1 | 425855 |
| Ginga Eiyuu Densetsu | 9.1 | 80679 |
| Gintama'; Enchousen | 9.1 | 81109 |
| Gintama Movie: Kanketsu-hen - Yorozuya yo Eien Nare | 9.1 | 72534 |
| Clannad: After Story | 9.1 | 456749 |
| Koe no Katachi | 9.1 | 102733 |
| Gintama | 9 | 336376 |
| Mienу Me ni Kanjita Kumotoriyama no Asahi | 9 | 47 |
| Mirai ni Mukete: Bousai wo Kangaeru | 9 | 77 |
| Okaachan Gomen ne | 9 | 47 |
| Santa-san wa Dai Isogashi | 9 | 25 |
| Shaka no Shougai | 9 | 31 |
| Shenmi Shijie Lixian Ji | 9 | 37 |
| Code Geass: Hangyaku no Lelouch R2 | 9 | 572888 |
| Haikyuu!! Second Season | 8.9 | 179342 |
| Sen to Chihiro no Kamikakushi | 8.9 | 466254 |
| Shigatsu wa Kimi no Uso | 8.9 | 416397 |
| Mushishi Zoku Shou 2nd Season | 8.9 | 75894 |
| Ookami Kodomo no Ame to Yuki | 8.8 | 226193 |
| Code Geass: Hangyaku no Lelouch | 8.8 | 715151 |
| Hajime no Ippo | 8.8 | 157670 |
| Rurouni Kenshin: Meiji Kenkaku Romantan - Tsuioku-hen | 8.8 | 129307 |

Imagem 13 – Execução do ficheiro “Ranking.html”.

3.2. Acesso a email

Nesta última parte do “Job”, o objetivo consistia em arranjar uma forma de conexão que permitisse troca de informação através de um mail em que a aplicação ao ser executada enviava os dados criados e informação relativa à execução da transformação.

Mail

Name of mail job entry: Mail

Addresses Server EMail Message Attached Files

Destination

Destination address: a16972@alunos.ipca.pt

Cc:

BCc:

Sender

Sender name Jose

Sender address: j[redacted]@gmail.com

Reply to:

Contact person:

Contact phone:

? Help OK Cancela

Imagem 14 – Configuração do remetente e do destinatário.

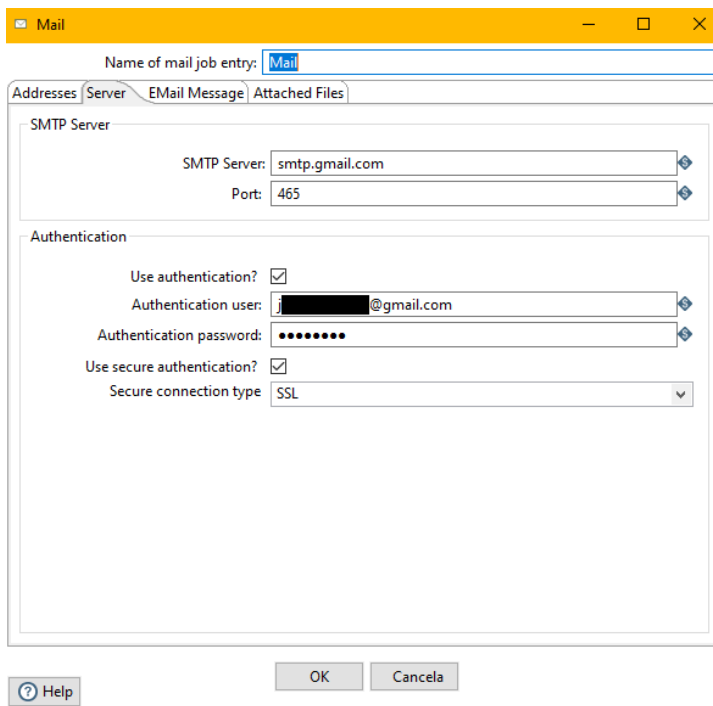


Imagem 15 – Configuração do servidor.

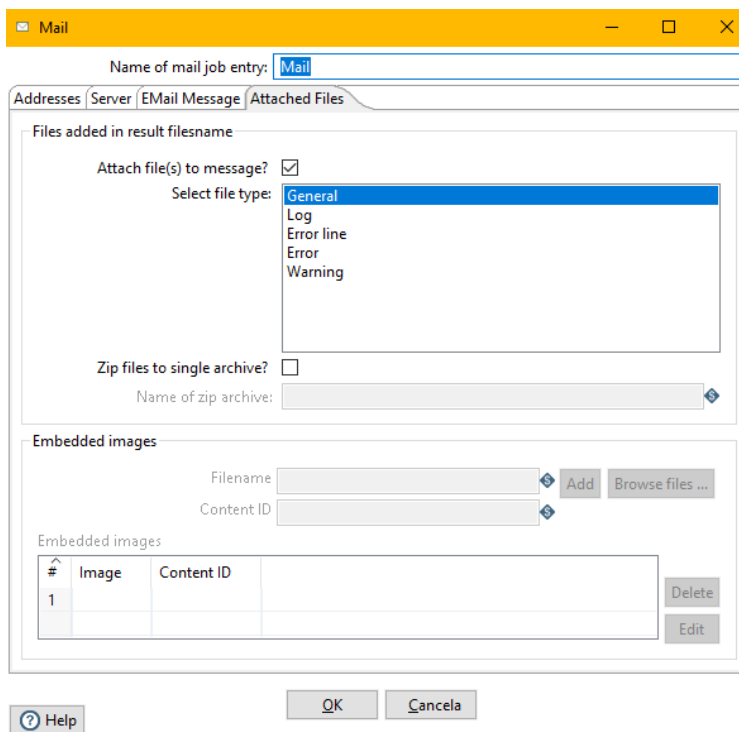


Imagem 16 – Configuração dos ficheiros a enviar.

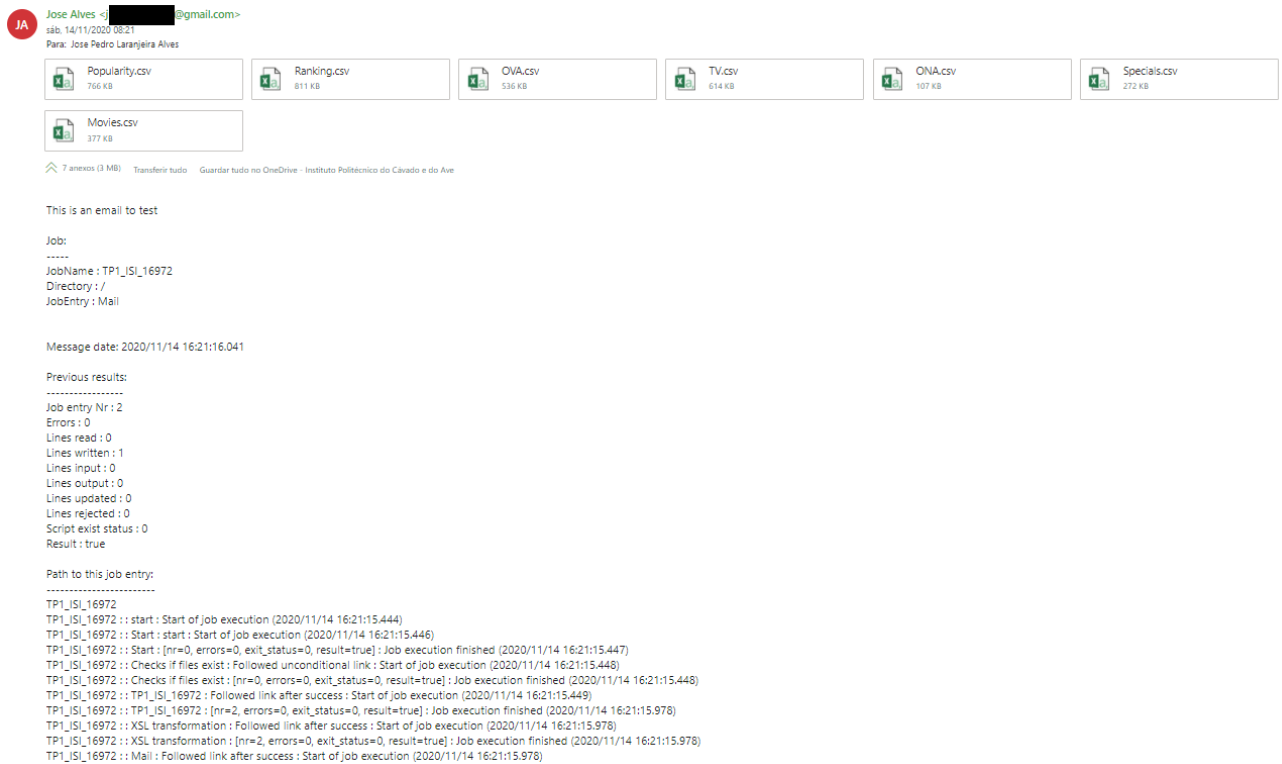


Imagem 17 – Email recebido.

4. Conclusão

Concluo que o ETL (Extract, Transformation and Load) é um processo útil capaz de mostrar um conjunto de dados de várias formas através de transformações onde estes podem ser reorganizados estruturalmente de várias formas. A ferramenta Kettle utilizada para a execução deste trabalho tinha várias opções que facilitaram a execução do trabalho.

Webgrafia

Material de estudo facultado pelo docente de Integração de Sistemas de Informação.

<https://www.kaggle.com/datasets>

https://www.w3schools.com/xml/xsl_transformation.asp

https://www.youtube.com/channel/UCxprQ8_p9MmfA0iCt1aoDwg