

Faculdade de Engenharia da Universidade do Porto



Estudantes & Autores:

André Leonor up201806860@fe.up.pt

Carlos Veríssimo up201907716@fe.up.pt

João Vasconcelos up201504397@fe.up.pt

Grupo 601

Trabalho realizado no âmbito da unidade curricular de Bases de Dados
([FEUP - Bases de Dados](#))

A. Definição do Modelo Conceptual

O projeto irá representar uma base de dados de uma plataforma de partilha e visualização de vídeos. De cada pessoa ter-se-á a informação relativa ao IP e ao dispositivo a ser utilizado, sendo que um utilizador pode possuir um canal, ou não estar registado. Pessoas não registadas apenas podem assistir a vídeos, e um canal (pessoa identificada) pode subscrever outros canais (subscrição normal ou premium). Cada vídeo é criado por um canal, sendo que a data de publicação será sempre igual, ou mais tardia, que a data de carregamento do vídeo. Os canais podem interagir com os vídeos através da atribuição de classificações, atribuídas pelos vários canais a diversos vídeos, partilha, em que os canais podem partilhar os vários vídeos, ou comentários, em que um determinado por um canal, e, por sua vez, um canal pode atribuir um dislike ou um like aos comentários dos vídeos. Um utilizador apesar de poder reagir a vários vídeos e comentários, não pode ter duas reações diferentes em simultâneo (não pode dar like e dislike no mesmo vídeo ou comentário).

Uma lista de reprodução é criada por um canal, sendo que um canal pode criar várias listas de reprodução, em conjuntos de vídeos, sendo, portanto, a duração da lista de reprodução correspondente à soma das durações dos vídeos que constituem a lista.

Os vídeos possuem alguns atributos não estáticos, sendo eles os números de visualizações, likes, dislikes, comentários e partilhas e com base no histórico de visualizações dos utilizadores, torna-se possível formular uma lista dos 50 vídeos mais populares (tendências) para cada país.

O tempo mínimo de uma visualização são 10 segundos. e os seguintes atributos são únicos:

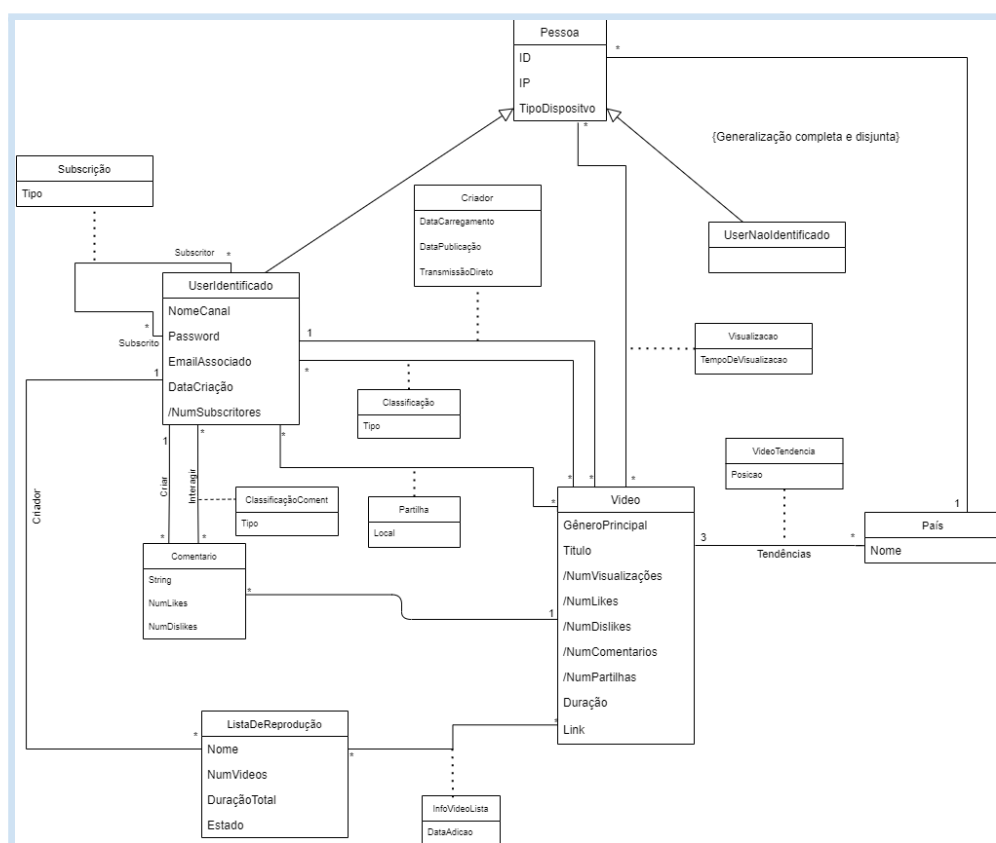
- o Atributo **NomeCanal** da tabela **USERIDENTIFICADO**;
- o Atributo **EmailAssociado** da tabela **USERIDENTIFICADO**;
- o Atributo **Link** da tabela **VIDEO**;
- o Atributos do tipo **Id_ "tabela"** usados na criação da base de dados. Ex: **LogUser_Id**, **Id_Pais**, **Id_Video**, entre outros...

Uma lista de reprodução terá um atributo estado, cujos valores poderão apenas ser: 'Pública' ou 'Privada'

Da mesma forma, a classe de associação "Criador", entre UserIdentificado e Vídeo, tem um atributo chamado TransmissãoDireto cujos valores válidos serão apenas 'Sim' ou 'Não'.

A data de publicação do vídeo será, obrigatoriamente maior ou igual à data de carregamento do vídeo na plataforma.

Diagrama original e atualizado em: [Diagrama UML](#)



B. Definição do esquema relacional

/* Object-oriented usado na generalização */

Pessoa (Id_Pessoa, Ip, TipoDispositivo, Id_Pais->Pais)

UserIdentificado (LogUser_Id->Pessoa, Ip, TipoDispositivo, NomeCanal, Password, EmailAssociado, DataCriacao, NumSubscritores)

UserNaoidentificado(NotLogUser_Id->Pessoa, Ip, TipoDispositivo)

Pais (Id_Pais, Nome)

Video (Id_Video, Link, Titulo, GeneroPrincipal, NumVisualizacoes, NumDislikes, NumLikes, NumComentarios, NumPartilhas, Duracao)

Comentario(Id_Coment, String, NumLikes, NumDislikes, video->Video, userIdentificado->UserIdentificado)

ListaDeReprodução (Id_Lista, Nome, NumVideos, DuracaoTotal, Estado, userIdentificado->UserIdentificado)

Partilha (userIdentificado->UserIdentificado, video->Video, Local)

Classificacao(userIdentificado->UserIdentificado, video->Video, Tipo)

Visualizacao (peessoa->Pessoa, video->Video, TempoDeVisualizacao)

InfoVideoLista (listaDeReproducao->ListaDeReproducao, video->Video, DataAdicao)

VídeoTendencia (video->Video, pais->Pais, Posicao)

Subscricao (userIdentificado-> UserIdentificado, userIdentificado->UserIdentificado, Tipo)

ClassificacaoComent(userIdentificado-> UserIdentificado, comentario->Comentario, Tipo)

Criador (DataCarregamento, DataPublicacao, TransmissaoDireto, video->Video, userIdentificado->UserIdentificado)

C. Análise de dependências funcionais e formas normais

Pessoa:

{Id_Pessoa} -> {Ip, TipoDispositivo, Id_Pais}

UserIdentificado :

{LogUser_Id} -> {Ip, TipoDispositivo, NomeCanal, Password, EmailAssociado, DataCriacao, NumSubscritores}

UserNaoidentificado:

{NotLogUser_Id} -> {Ip, TipoDispositivo}

Pais:

{Id_Pais} -> {Nome}

{Nome} -> {Id_Pais}

Video:

{Id_Video} -> {Link, Titulo, GeneroPrincipal, NumVisualizacoes, NumDislikes, NumLikes, NumComentarios, NumPartilhas, Duracao}

{Link} -> {Id_Video, Titulo, GeneroPrincipal, NumVisualizacoes, NumDislikes, NumLikes, NumComentarios, NumPartilhas, Duracao}

Comentario:

{Id_Coment} -> {String, NumLikes, NumDislikes, video, userIdentificado}

ListaDeReprodução:

{Id_Lista} -> {Nome, NumVideos, DuraçãoTotal, Estado, userIdentificado}

Partilha:

{userIdentificado, Video} -> {Local}

Classificação:

{userIdentificado, Video} -> {Tipo}

Visualização:

{Pessoa, Video} -> {Tempo De Visualizacao}

InfoVideoLista:

{listaDeReprodução, vídeo} -> {DataAdicao}

VídeoTendencia:

Não existem dependências funcionais não triviais.

Subscrição:

{UserIdentificado, UserIdentificado} -> {Tipo}

ClassificaçãoComent:

{UserIdentificado, Comentario} -> {Tipo}

Criador:

{Video} -> {DataCarregamento, DataPublicacao, TransmissaoDireto, UserIdentificado}

Uma relação R está conforme a Boyce-Codd Normal Form (BCNF) se em todas as dependências funcionais não triviais $A \rightarrow B$, A seja uma superkey, isto é, $A +$ contenha todos os atributos de R. Sendo que cada elemento da esquerda é uma superkey, e, os atributos da direita dependem dessa mesma superkey, isto é, a partir de todas as dependências funcionais identificadas foi possível determinar todos os atributos das tabelas. Logo não ocorrem violações à BCNF, não ocorrendo também violações à 3ªFN, uma vez que 3ªFN é um subset de BCNF.

D. Criação da base de dados e em Sqlite

Nesta fase é criado um ficheiro chamado criar.sql que inclui as instruções SQL para criar todas as tabelas mencionadas no esquema relacional resultante do passo B.

- Criação das tabelas (classes) a ser incorporadas na base de dados;

```
-- Tabela referente a classe: PESSOA
DROP TABLE IF EXISTS PESSOA;
CREATE TABLE PESSOA ( ...
);
```

- Criação das colunas de cada tabela (atributos) e especificação do tipo de dados correspondente;

```
-- Tabela referente a classe: PESSOA
DROP TABLE IF EXISTS PESSOA;
CREATE TABLE PESSOA (
    Id_pessoa INTEGER;
    Ip TEXT;
    TipoDispositivo TEXT;
    Id_Pais INTEGER;
```

- Especificação das chaves primárias e chaves estrangeiras;

```
-- Tabela referente a classe: PESSOA
DROP TABLE IF EXISTS PESSOA;
CREATE TABLE PESSOA (
    Id_pessoa INTEGER NOT NULL UNIQUE,
    Ip TEXT DEFAULT 'NULL',
    TipoDispositivo TEXT DEFAULT 'NULL',
    Id_Pais INTEGER,

    CHECK (Id_pessoa > 0),

    CONSTRAINT PESSOA_PK PRIMARY KEY(Id_pessoa),
    CONSTRAINT PESSOA_PAIS_FK FOREIGN KEY(Id_Pais) REFERENCES PAIS(Id_Pais) ON DELETE SET NULL ON UPDATE CASCADE
);
```

E. Adição de restrições à base de dados

Foram incluídas na base de dados uma vasta gama de restrições convenientes, de modo a garantir a integridade dos dados armazenados. Além disso, estas restrições asseguram a exactidão e fiabilidade dos dados das tabelas. Se houver qualquer violação entre a restrição e a ação de dados, essa ação é abortada.

As restrições usadas e a sua especificação serão apresentadas por tabelas/classes:

PESSOA

- Ao atributo *Id_Pessoa* foi aplicada a restrição UNIQUE uma vez que se trata de um identificador individual no contexto da classe em questão. As restrições deste tipo serão, grande parte das vezes, usadas em identificadores (Id's) de classes.
- Pela razão mencionada em cima, foi também aplicada ao atributo *Id_Pessoa* a restrição NOT NULL.
- Aos atributos *Ip* e *TipoDispositivo* foi aplicada a restrição DEFAULT, que coloca os valores desses atributos a NULL, caso não sejam especificados valores para os mesmos. Esta restrição é aplicada a vários atributos na base de dados. Geralmente é usada quando as tabelas, nas quais os atributos em questão estão presentes, conseguem existir sem a especificação dos valores desses atributos.
- Foi também aplicado um CHECK que verifica que o valor do atributo *Id_Pessoa* que está a ser inserido na tabela é maior do que zero.

USERIDENTIFICADO

- Os atributos *LogUser_Id*, *Ip* e *TipoDispositivo* são chaves estrangeiras/atributos herdados da classe *Pessoa* logo seguem a mesma linha de restrições.
- Aos atributos *NomeCanal* e *EmailAssociado* é empregue a restrição UNIQUE devido à especificação da base de dados que impõe que não podem coexistir na plataforma dois utilizadores que detenham *NomeCanal* ou *EmailAssociado* iguais.
- São também aplicadas as restrições NOT NULL aos atributos *NomeCanal*, *Password* e *NumSubscritores* e DEFAULT NULL ao atributo *DataCriacao*.
- Existem também duas restrições do tipo CHECK. Uma delas, verifica que o tamanho dos valores inseridos em *Password* são superiores a 8, enquanto a outra verifica que os valores inseridos em *NumSubscritores* é sempre maior do que zero, uma vez que a natureza da base de dados assim o exige.

USERNAOIDENTIFICADO

Esta classe herda todos os seus atributos da tabela *Pessoa* e, por esse motivo, considera-se redundante a especificação das restrições da mesma.

PAIS

- Dois atributos contém restrições do tipo NOT NULL sendo eles *Id_Pais* e *Nome*, ao passo que apenas o atributo *Id_Pais* tem presente a restrição UNIQUE. Não se aplicou esta última restrição à tabela *Nome* uma vez que se considerou que poderia haver dois ou mais países com o mesmo nome/siglas iguais, como é o caso de países como a República Democrática do Congo e a República do Congo.

- É adicionado também um CHECK que verifica se os valores inseridos em *Id_Pais* se encontram entre a gama [1, 10], especificada aquando a criação da base de dados..

VIDEO

- Restrição UNIQUE ---- *Id_Video, Link* ---- Uma vez que existe a necessidade extrema de que estes atributos contenham valores únicos.
- Restrição NOT NULL ---- *Id_Video, Titulo, Duracao* ---- Uma vez que a classe em questão não faria sentido com valores inexistentes nos atributos acima.
- Restrição DEFAULT ---- *GeneroPrincipal* --- Coloca valores NULL de *GeneroPrincipal* com o valor 'Miscellaneous'.
- A restrição CHECK é aplicada aos atributos *NumVisualizacoes, NumLikes, NumDislikes, NumComentarios, NumPartilhas, Duracao*, verificando se os valores inseridos nestas tabelas são maiores ou iguais a zero, devido ao contexto da base de dados. Existe também outra restrição do tipo CHECK que verifica se o tamanho dos valores inseridos em *Titulo* são inferiores a 70 de modo a impedir que sejam inseridos títulos demasiado longos.

COMENTARIO

- Restrição UNIQUE ---- *Id_Coment*.
- Restrição NOT NULL ---- *Id_Coment, Id_Video, LogUser_Id, String* ---- Necessidade de valores não nulos para haver significância dos tuplos da tabela.
- A restrição CHECK é aplicada aos atributos *NumLikes* e *NumDislikes*, verificando se os valores a inserir nessa tabela são maiores do que zero. É também usada uma restrição deste tipo para verificar o número de caracteres que o comentário possui: esse número não poderá ser superior a 1000 caracteres.

LISTADEREPRODUCAO

- Restrição UNIQUE ---- *Id_Lista*.
- Restrição NOT NULL ---- *Id_Lista, Nome, LogUser_Id, Estado* ---- Mais uma vez, existe a necessidade que os valores inseridos nestas tabelas sejam válidos pois, caso contrário, pouco sentido fariam os dados presentes nessas tabelas.
- Restrição CHECK:
 - CHECK do intervalo de valores:
 - *NumVideos* e *DuracaoTotal* ---- Devido ao contexto da plataforma, verifica se os valores inseridos nas tabelas mencionadas são maiores do que 0.
 - *Nome* ---- Verifica se o nome da lista tem tamanho inferior a 100, de forma a restringir um pouco a liberdade na escolha do nome da lista.
 - CHECK do tipo de valores:
 - *Estado* ---- Devido a binariedade deste atributo, verifica se os valores inseridos são estritamente iguais aos definidos na criação da base de dados. Neste caso, os valores inseridos em *Estado* serão obrigatoriamente iguais a 'Pública' ou 'Privada'.

PARTILHA

- Restrições NOT NULL ---- *Id_Video, LogUser_Id*.
- Restrições DEFAULT ---- *Local* ---- Por omissão, os valores deste atributos serão 'NULL'.

CLASSIFICACAO

- Restrições NOT NULL ---- *Id_Video, LogUser_Id, Tipo*.
- Restrição CHECK: Baseia-se na restrição CHECK aplicada ao atributo *Estado* em *LISTADEREPRODUCAO*, apenas difere no facto de os valores aos quais se tem de comparar os dados inseridos são 'Like' e 'Dislike'.

VISUALIZACAO

- Restrições NOT NULL ---- *Id_Video, Id_Pessoa*.
- Restrição CHECK verifica se o *TempoDeVisualizacao* é igual ou superior a 10, uma vez que se considerou, aquando a conceção da base de dados, que o tempo mínimo para uma visualização ser válida é de 10 (segundos).

INFOVIDEOLISTA

- Restrições NOT NULL ---- *Id_Video, Id_Lista*.
- Restrições DEFAULT ---- *Data*---- Por omissão, os valores deste atributos serão 'NULL'.

VIDEOTENDENCIA

- Restrições NOT NULL ---- *Id_Pais, Id_Video, Posicao*.

SUBSCRICAO

- Restrições NOT NULL ---- *LogUserId1, LogUserId2, Tipo*.
- Restrição CHECK: ---- *Tipo* ---- Baseia-se na restrição CHECK aplicada ao atributo *Estado* em *LISTADEREPRODUCAO*, apenas difere no facto de os valores aos quais se tem de comparar os dados inseridos são 'Premium' e 'Basic'.

CLASSIFICACAOCOMENT

- Restrições NOT NULL ---- *LogUserId, Id_Coment, Tipo*.
- Restrição CHECK ---- *Tipo* ---- Baseia-se na restrição CHECK aplicada ao atributo *Tipo* em *LISTADEREPRODUCAO*, apenas difere no facto de os valores aos quais se tem de comparar os dados inseridos são 'Like' e 'Dislike'.

CRIADOR

- Restrições NOT NULL ---- *DataCarregamento, DataPublicacao, TransmissaoDireto Id_Video, LogUserId*.
- Restrição CHECK:
 - CHECK da validade dos valores:
 - Verifica se a *DataCarregamento* é menor ou igual à *DataPublicacao*, pois de outra forma, não faria sentido.
 - CHECK do tipo de valores:
 - Verifica se os valores inseridos através do atributo *TransmissaoDireto* são da norma binária 'Sim' ou 'Não'.

F. Carregamento de dados

- Nesta parte do projeto, foi criado um ficheiro **povoar.sql**, ficheiro este que contém todos os comandos necessários para introduzir dados na base.

```
INSERT INTO PAIS VALUES (1, 'Portugal');
INSERT INTO PAIS VALUES (2, 'United States');
INSERT INTO PAIS VALUES (3, 'Brazil');
INSERT INTO PAIS VALUES (4, 'United Kingdom');
INSERT INTO PAIS VALUES (5, 'Spain');
INSERT INTO PAIS VALUES (6, 'France');
INSERT INTO PAIS VALUES (7, 'Australia');
INSERT INTO PAIS VALUES (8, 'Italy');
INSERT INTO PAIS VALUES (9, 'Germany');
INSERT INTO PAIS VALUES (10, 'Canada');
```

- Nesta parte, é também verificada a integridade das restrições especificadas na criação da base de dados. Abaixo encontram-se algumas dessas restrições a impedir a inserção de dados inválidos na base.

- Inserção de valores com o mesmo UNIQUE Id;
- Inserção de valores que não obedecem o CHECK especificado;

```
sqlite> INSERT INTO PAIS VALUES (1, 'Portugal');
sqlite> INSERT INTO PAIS VALUES (1, 'United States');
Error: UNIQUE constraint failed: PAIS.Id_Pais
```

- Inserção de valores que não constam no CHECK que especifica os valores válidos;

```
CONSTRAINT DATA_VALID CHECK (DataCarregamento <= DataPublicacao),
--
sqlite> INSERT INTO CRIADOR(DataCarregamento, DataPublicacao, TransmissaoDireto, Id_Video, LogUser_Id)
VALUES("2021-02-16", "2019-02-16", 'Sim', 7, 4);
Error: CHECK constraint failed: DATA_VALID
```

- Inserção de valores cujo tamanho se encontra fora daqueles aceites pela base de dados;

```
sqlite> INSERT INTO USERIDENTIFICADO(Ip, TipoDispositivo, NomeCanal, Password, EmailAs
VALUES('7756.5559.7759.2999', 'mobile', 'Another interests', 'ajsf', '
Error: CHECK constraint failed: LENGTH>Password) > 8
```

- Entre outros...

```
sqlite> INSERT INTO CLASSIFICACAO(LogUser_Id, Id_Video, Tipo)
...> VALUES(3, 2, 'DISL');
Error: CHECK constraint failed: Tipo = 'LIKE' OR Tipo = 'DISLIKE'
```

- Verifica-se que a inserção de dados na base ocorreu sem qualquer tipo de problema através de uma GUI(*graphical user interface*) ou mesmo através da consola de um *IDE*.

Id_Coment	String	NumLikes	NumDislikes	Id_Video	LogUser_Id
1	Don't forget to subscribe!	1	1	1	1
2	Nice video	0	0	1	2
3	Cringe	0	0	3	4
4	Interesting video	0	0	3	3
5	very good	0	0	7	2
6	another one	1	1	5	1
7	ahahahahahah	3	1	5	1
8	great match!!!!	4	0	7	1
9	Bro, you are crazy!	1	0	8	8
10	Great stuff m8	0	0	12	3
11	Poor Ethan	1	0	13	1

LogUser_Id	Id_Video	Local
4	5	Twitter
3	1	Tumblr
2	1	Facebook
1	1	Copy Link
1	2	Reddit
2	2	NULL
3	6	Facebook
4	6	LinkedIn
2	5	Facebook
1	7	Facebook
1	8	Facebook
4	13	NULL

LogUser_Id	Id_Video	Local
4	5	Twitter
3	1	Tumblr
2	1	Facebook
1	1	Copy Link
1	2	Reddit
2	2	NULL
3	6	Facebook
4	6	LinkedIn
2	5	Facebook
1	7	Facebook
1	8	Facebook
4	13	NULL

Id_Pais	Nome
1	Portugal
2	United States
3	Brazil
4	United Kingdom
5	Spain
6	France
7	Australia
8	Italy
9	Germany
10	Canada

G. Interrogação da Base de dados

Nesta fase foi definido um conjunto de interrogações pertinentes para o contexto da base de dados.

Interrogação #1:

- *Quais os utilizadores que fizeram upload de vídeos com duração superior a cinco minutos e cujo gênero principal/categoria é 'Entertainment'? (Apresentar duração do vídeo no formato hh:mm:ss).*

Interrogação #2:

- *Qual a média de visualizações e o número total de vídeos por categoria? (Ordenar por média (arredondada às unidades) de visualizações).*

Interrogação #3:

- *Qual o criador do vídeo com upload este ano (2021) que tem mais do que 10 visualizações?*

Interrogação #4:

- *Quais são os utilizadores que não escreveram nenhum comentário?*

Interrogação #5:

- *Quais os vídeos que estão nas tendências em Portugal?*

Interrogação #6:

- *Quais são os utilizadores que assistiram, por completo, ao vídeo mais visualizado da plataforma?*

Interrogação #7:

- *Quais os comentários do vídeo com upload mais recente?*

Interrogação #8:

- *Quais são os vídeos que foram adicionados a listas de reprodução com o estado "Publico"?*

Interrogação #9:

- *Apresentar as estatísticas de todos os utilizadores. (ID, Nome, País, N° de subscritores, N° de vídeos carregados, N° de comentários escritos, N° de reações "Like" e N° de reações "Dislike").*

Interrogação #10:

- *Qual é o top3 de vídeos assistidos por país.*

H. Adição de gatilhos à base de dados

Foram adicionados os seguintes gatilhos à base de dados:

1) Quando da adição de um novo tuplo na tabela de utilizadores identificados verifica-se se já existe na tabela pessoa um tuplo com o *Loguser_id* correspondente, pois trata-se de uma chave estrangeira, uma vez que é uma generalização. Não existindo é adicionado à tabela pessoa um tuplo para depois se adicionar o utilizador identificado pretendido.

2) Ao adicionar um vídeo à lista de reprodução é feita a verificação se esse vídeo já foi publicado, se ainda não tiver sido publicado a adição não é executada e uma mensagem de erro é apresentada ao utilizador. Para fazer esta verificação é comparada a data de publicação na tabela *Criador* a data de adição na tabela de informação da lista de reprodução.

3) Ao atualizar a reação a um comentário (tabela classificação de comentários), verificamos se o tipo do comentário que se está a atualizar é oposto ao novo Tipo, uma vez que os tipos de reações são 'like' ou 'dislike', ao satisfazer esta condição atualizamos também o número de dislikes e número de likes do comentário em questão.