



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

FUNDAMENTO DE BASE DE DATOS



NOMBRE:

- *Julio César Luna Díaz*

FECHA:

09/02/2022

DOCENTE:

Nelson Oswaldo Piedra Pullaguari, Ph.D

Informe de Avances del desarrollo del Proyecto Integrador

02, enero de 2022

Tema: Interacción con el archivo .csv (archivo con las bases de datos)

Introducción

En este inicio de actividades se comenzó obteniendo el archivo que contiene las bases de datos con la que se va a trabajar en el presente proyecto, además de realizar un breve análisis a la información contenida en dicho archivo.

Desarrollo

1. Se realizó la descarga desde un repositorio de github, perteneciente al usuario “*rashida048*”, del archivo .csv denominado **movie_dataset.csv**. Este archivo contiene una amplia data de aproximadamente 5000 películas, de donde se desglosa su nombre, su productora, su director, su personal, fecha de lanzamiento, etc.
2. Posterior mediante el uso de las herramientas para el manejo de datos del programa Excel se procedió a organizar la data en sus respectivos campos.
3. Luego, teniendo ya organizada de una mejor manera la data se procedió a establecer los diversos campos que el archivo *movie_dataset.csv* nos proporcionó, obteniendo los campos:
 - index
 - budget
 - genres
 - homepage
 - id
 - keywords
 - original_language
 - original_title
 - overview
 - popularity
 - production_companies
 - production_countries
 - release_date
 - revenue
 - runtime
 - spoken_languages

- | | |
|----------------|--------------|
| ▪ status | ▪ vote_count |
| ▪ tagline | ▪ cast |
| ▪ title | ▪ crew |
| ▪ vote_average | ▪ director |

4. Después de identificado los diversos campos, se intentó realizar el planteamiento de entidades para toda esta data. Observándose que el establecimiento de estas entidades no iba a cumplir el criterio de que el número de entidades se iban a establecer de acuerdo con el número de campos en los que esta organizada la data, debido a que, en ciertos campos la data no es atómica y esto implica que no se dé el cumplimiento de uno de los principios de los atributos en una entidad. Por ello, luego de una revisión a cada uno de los campos se estableció las siguientes entidades con sus respectivos atributos:

❖ MOVIE

- Index
- IdMovie
- Budget
- Homepage
- Keywords
- Original_title
- Overview
- Popularity
- Release_date
- Revenue
- Runtime
- Tagline

➤ TitleMovie

- Vote_average
- Votecount

❖ GENRE

- IdGenre
- GenreName

❖ ORIGINALLANGUAGE

- Id_OriginalLanguage
- e
- OrigLanguageName

❖ PRODUCTIONCOMPAN

Y

- Id_ProductionComp
- any
- ProdCompanyName

❖ PRODUCTIONCOUNTRY

- Id_ProductionCount
- ry
- ISO_3166_1
- ProdCountryName

❖ SPOKENLANGUAGE

- Id_SpokenLanguage
- ISO_639_1
- SpokenLangName

❖ STATUS

- IdStatus
- StatusName

❖ CAST

- IdCast
- CastName

❖ CREW

- IdCrew
- GenderCrew
- Credit_idCrew
- CrewName

❖ DEPARTMENTCREW

- IdDepartment
- DepartcrewName

❖ JOBCREW

- IdJob
- JobcrewName

❖ DIRECTOR

- IdDirector
- DirectorName

5. Luego se procedió a establecer las diversas dependencias, entre las que se encontró:

IdMovie → {Budget, Homepage, Keywords, Index, Original_Title, Overview, Popularity, Release_Date, Revenue, Runtime, Status, Tagline, TitleMovie, Vote_Average, VoteCount}

IdGenre → {GenreName}

IdOriginalLanguage → {OrigLanguageName}

IdProductionCompany → {ProdCompanyName}

IdProductionCountry → {ISO_3166_1, ProdCountryName}

IdSpokenLanguage → {ISO_639_1, SpokenLangName}

Status → {StatusName}

Cast → {CastName}

Crew → {GenderCrew, Credit_idCrew, CrewName}

IdDepartment → {DepartcrewName}

IdJob → {JobcrewName}

IdDirector → {DirectorName}

6. Posterior a ello se procedió a aplicar a las tablas obtenidas las 3 reglas de la normalización (Primera Forma Normal, Segunda Forma Normal, Tercera Forma Normal):

Normalización

1FN – Primera Forma Normal:

Para la primera forma normal se analizo si en caso los atributos cumplen con ser atómicos, es decir, son indivisibles.

- Se llego a la conclusión de que las tablas cumplen con que los atributos *si* son atómicos.

Multivaluados		
Genres	Keywords	Cast

Multivaluados Compuestos			
ProductionCompanies	ProductionCountry	SpokenLanguage	Crew

2FN – Segunda Forma Normal:

Para el cumplimiento de la segunda forma normal se tiene que cada uno de los atributos deben tener únicamente una dependencia funcional completa, y no dependencias funcionales parciales.

- Finalizado el análisis se concluyó que las tablas propuestas si cumplen con el cumplimiento de una dependencia funcional completa

Entidades en Resumen

ENTIDADES			
MOVIE	GENRE	ORIGINALLANGUAGE	PRODUCTIONCOMPANY
<u>IdMovie</u>	<u>IdGenre</u>	<u>IdOriginalLanguage</u>	<u>IdProductionCompany</u>
Budget			
Homepage	GenreName	OrigLanguageName	ProdCompanyName
Keywords			
Index			
Original_title			
Overview			
Popularity			
Release_date			
Revenue			
Runtime			
Tagline			
TitleMovie			
Vote_average			
Votecount			

ENTIDADES			
PRODUCTIONCOUNTRY	SPOKENLANGUAGE	STATUS	CAST
<u>IdProductionCountry</u>	<u>IdSpokenLanguage</u>	<u>IdStatus</u>	<u>IdCast</u>

ISO_3166_1	ISO_639_1	StatusName	CastName
ProdCountryName	SpokenLangName		

ENTIDADES			
CREW	DEPARTMENTCREW	JOBCREW	DIRECTOR
<u>IdCrew</u>	<u>IdDepartment</u>	<u>IdJob</u>	<u>IdDirector</u>
GenderCrew Credit_idCrew CrewName	DepartcrewName	JobcrewName	DirectorName

3FN – Tercera Forma Normal:

Para poder cumplir con los parámetros de la tercera forma normal se busca evitar que se presenten dependencias transitivas entre los atributos de una tabla.

- Al observar las tablas propuestas, se observa que si se cumple que los atributos no tienen una dependencia transitiva.

Entidades en Resumen

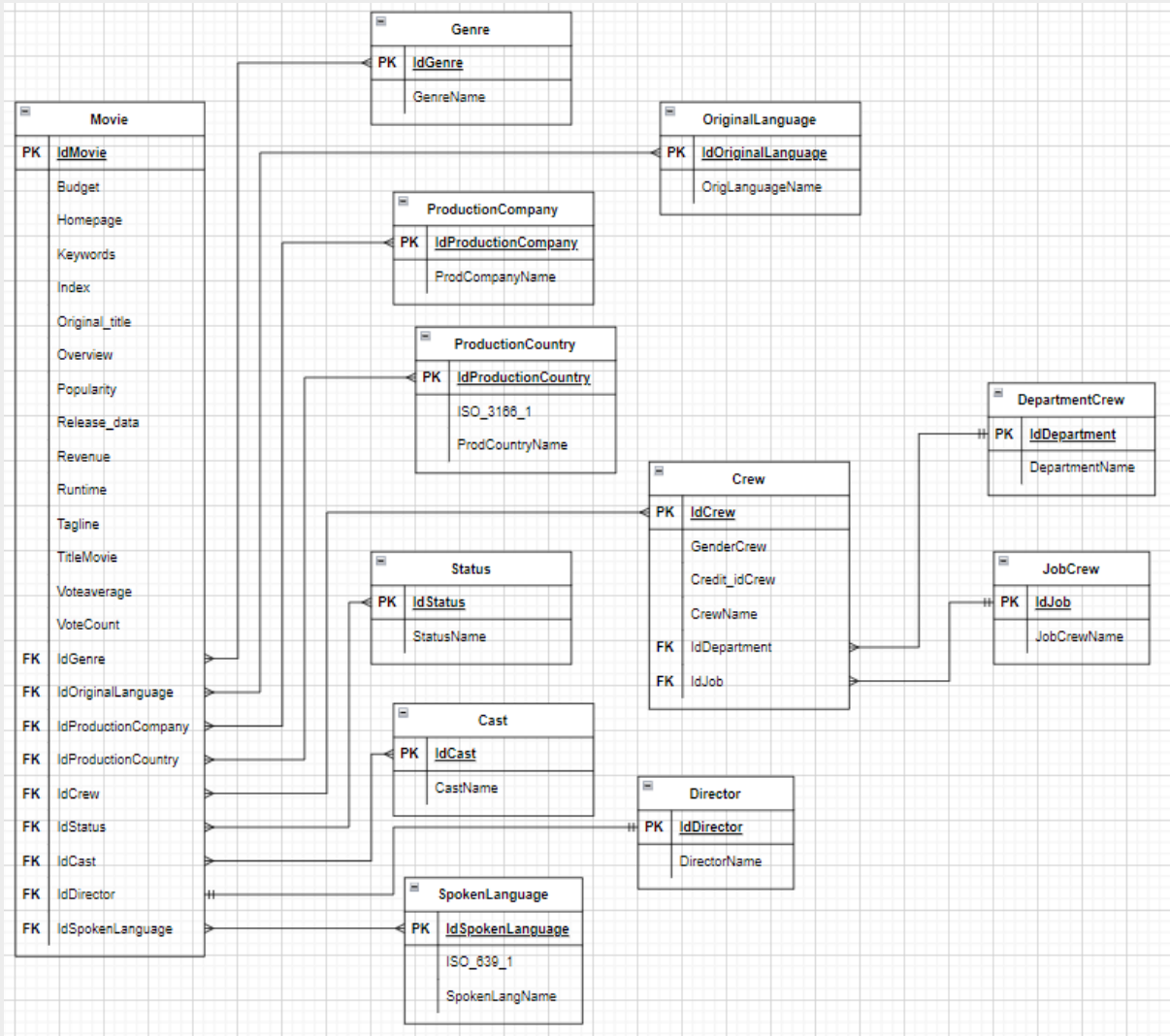
ENTIDADES			
MOVIE	GENRE	ORIGINALLANGUAGE	PRODUCTIONCOMPANY
<u>IdMovie</u>	<u>IdGenre</u>	<u>IdOriginalLanguage</u>	<u>IdProductionCompany</u>
Budget Homepage Keywords Index Original_title Overview Popularity Release_date Revenue Runtime Tagline TitleMovie Vote_average Votecount	GenreName	OrigLanguageName	ProdCompanyName

ENTIDADES			
PRODUCTIONCOUNTRY	SPOKENLANGUAGE	STATUS	CAST
<u>IdProductionCountry</u>	<u>IdSpokenLanguage</u>	<u>IdStatus</u>	<u>IdCast</u>
ISO_3166_1	ISO_639_1	StatusName	CastName

ProdCountryName	SpokenLangName
-----------------	----------------

ENTIDADES			
CREW	DEPARTMENTCREW	JOBCREW	DIRECTOR
<u>IdCrew</u>	<u>IdDepartment</u>	<u>IdJob</u>	<u>IdDirector</u>
GenderCrew Credit_idCrew CrewName	DepartcrewName	JobcrewName	DirectorName

7. Luego de ello se procedió a realizar el modelo relacional, basado en las entidades y atributos previamente recopilados.



8. A continuación, se procedió a realizar un desglose de cada uno de los atributos de las entidades (Dominio, Opcionalidad, Multivaluado, Tipo, Comentario)

Definición de la Entidad *MOVIE*

Property	Domain	Optional	MultiValuated	Type	Comment
Index	Int(5000)	Not Null	No	Primary Key	Número en el catalogo
IdMovie	Varchar(16)	Not Null	No	Primary Key	Id asignado a cada película
Budget	Int(5000)		Yes		Presupuesto
Homepage	Varchar(200)		Yes		Página Web
Keywords	Varchar(300)		Yes		Palabras identificativas
Original_title	Varchar(100)		Yes		Titulo original (Idioma de Origen)
Overview	Varchar(500)		Yes		Descripcion
Popularity	Int		Yes		Popularidad
Release_date	Date		Yes		Fecha de Lanzamiento
Revenue	Int		Yes		Ingresos
Runtime	Int		Yes		Tiempo de duración
Tagline	Varchar(200)		Yes		Lema o frase
TitleMovie	Varchar(100)		Yes		Titulo Película (inglés)
Vote_average	Int		Yes		Promedio de Votación del publico
Votecount	Int		Yes		# Personas que votaron

Definición de la Entidad *GENRE*

Property	Domain	Optional	MultiValuated	Type	Comment
IdGenre	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Género
GenreName	Varchar(10)	Not Null	Yes		Nombre del Género

Definición de la Entidad *ORIGINALLANGUAGE*

Property	Domain	Optional	MultiValuated	Type	Comment
IdOriginalLanguage	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Idioma Original
OrigLanguageName	Varchar(5)	Not Null	Yes		Nombre del Idioma Original

Definición de la Entidad *PRODUCTIONCOMPANY*

Property	Domain	Optional	MultiValuated	Type	Comment
IdProductionCompany	Varchar(20)	Not Null	No	Primary Key	Id Asignado a la Compañía Productora
ProdCompanyName	Varchar(10)	Not Null	Yes		Nombre de la Compañía Productora

Definición de la Entidad *PRODUCTIONCOUNTRY*

Property	Domain	Optional	MultiValuated	Type	Comment
IdProductionCountry	Varchar(20)	Not Null	No	Primary Key	Id Asignado al País de la Productora
ISO_3466_1	Varchar(5)	Not Null	Yes		Abreviatura del País
ProdCountryName	Varchar(50)	Not Null	Yes		Nombre del País

Definición de la Entidad *SPOKENLANGUAGE*

Property	Domain	Optional	MultiValuated	Type	Comment
IdSpokenLanguage	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Idioma Hablado
ISO_639_1	Varchar(5)	Not Null	Yes		Abreviatura del Idioma
SpokenLangName	Varchar(20)	Not Null	Yes		Nombre del Idioma

Definición de la Entidad *STATUS*

Property	Domain	Optional	MultiValuated	Type	Comment
IdStatus	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Estatus
StatusName	Varchar(15)	Not Null	Yes		Nombre del Estatus

Definición de la Entidad *CAST*

Property	Domain	Optional	MultiValuated	Type	Comment
IdCast	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Elenco
CastName	Varchar(50)	Not Null	Yes		Nombre del Elenco

Definición de la Entidad *CREW*

Property	Domain	Optional	MultiValuated	Type	Comment
IdCrew	Varchar(20)	Not Null	No	Primary Key	Id Asignado al Personal
GenderCrew	Int	Not Null	Yes		Genero del Personal
Credit_idCrew	Varchar(30)	Not Null	Yes		Id de crédito del Personal
CrewName	Varchar(100)	Not Null	Yes		Nombre del Personal

Definición de la Entidad *DEPARTMENTCREW*

Property	Domain	Optional	MultiValuated	Type	Comment
IdDepartment	Varchar(20)	Not Null	No	Primary Key	Id Asignado al tipo Departamento del Elenco
DepartcrewName	Varchar(15)	Not Null	Yes		Nombre del tipo de Departamento del Elenco

Definición de la Entidad *JOBCREW*

Property	Domain	Optional	MultiValuated	Type	Comment
IdJob	Varchar(20)	Not Null	No	Primary Key	Id Asignado Trabajo del Personal
JobcrewName	Varchar(30)	Not Null	Yes		Nombre del Trabajo del Personal

Definición de la Entidad *DIRECTOR*

Property	Domain	Optional	MultiValuated	Type	Comment
IdDirector	Varchar(20)	Not Null	No	Primary Key	Id Asignado al director
DirectorName	Varchar(30)	Not Null	Yes		Nombre del director

Anexos

Link de uso para la descarga de archivo .csv:

https://github.com/rashida048/Datasets/blob/master/movie_dataset.csv Enlaces a un sitio externo.