



Universitat
de les Illes Balears

TRABAJO DE FIN DE MÁSTER

ANÁLISIS DE PERSONAJES MITOLÓGICOS

Sergi Fornés Fornés

Máster Universitario

Análisis de Datos Masivos en Economía y Empresa

Centro de Estudios de Postgrado

Año Académico 2021-22

Análisis de Personajes Mitológicos

Sergi Fornés Fornés

Trabajo de Fin de Máster

Centro de Estudios de Postgrado

Universidad de las Illes Balears

Año Académico 2021-22

Palabras clave del trabajo:

Mitología, Web scraping, Análisis de redes sociales, Análisis de texto, Dashboard

Nombre Tutor/Tutora del Trabajo: Ricardo Alberich Martí

Índice

Resumen	2
Abstract	2
Introducción	3
Obtención del conjunto de datos	3
Análisis de grafos y texto	4
Relación entre las mitologías	4
Relación entre los personajes	4
Análisis de texto	5
Visualización de los resultados	5
Exploración de la Base de Datos	5
Relación entre Mitologías	5
Relación entre Personajes	5
Temática de los Personajes	6
Análisis de Mitología	6
Análisis de Personaje	6
Conclusiones	6
Bibliografía	7
Apéndice 1	7
Apéndice 2	9
Apéndice 3	9

Índice de figuras

1.	Distribución de grados	4
2.	Exploración de la Base de Datos	5
3.	Relación entre Mitologías	5
4.	Relación entre Personajes	6
5.	Temática de los Personajes	6
6.	Análisis de Mitología	6
7.	Análisis de Personaje	6
8.	Modelo de datos inicial	9
9.	Modelo de datos con resultados	9

Resumen

Se ha creado una herramienta con el objetivo de analizar dinámicamente los personajes pertenecientes a las diversas mitologías del mundo. Esta herramienta está publicada en internet y es accesible a través del siguiente [enlace](#).

Todos los datos de las mitologías y personajes mitológicos han sido obtenidos de [Wikipedia](#). Mediante diversas librerías de scrapeo de Python, se ha creado la base de datos. Después se ha usado el software R para realizar diferentes análisis de grafos y texto, y con los resultados se ha creado un panel en Power BI.

Abstract

We have created a tool to analyze dynamically the mythological characters belonging to mythologies all around the world. This tool is published online and it is accesible through the following [link](#).

Mythologies and characters data have been obtained from [Wikipedia](#). The data base has been created using various scrapping libraries from Pyhton. We used R software to do the network and text analysis. Finally, with the results we have created a dashboard on Power BI.

Introducción

La mitología tiene una gran influencia en nuestra cultura actual. A pesar de que existen mitos con miles de años de antigüedad, hoy en día no es extraño encontrarse con referencias mitológicas en cualquier ámbito artístico y cultural. Todas las culturas cuentan con sus propios mitos y son un componente indispensable para conocerlas con mayor profundidad.

Aunque no existe una definición rigurosa de mito, se puede decir que son historias tradicionales que responden cuestiones importantes de manera sencilla. En los mitos aparecen personajes sobrenaturales y las relaciones que mantienen con los seres humanos. Existen cientos de mitologías y una misma mitología puede estar compuesta de hasta cientos de estos personajes mitológicos, cada uno con sus peculiaridades.

A continuación, se explica detalladamente el proceso de creación de una herramienta que tiene como objetivo el análisis de mitologías y personajes mitológicos. Primero se obtiene el conjunto de datos que se utilizará, después se analiza y se obtienen diversas métricas, y finalmente se crean las visualizaciones.

Obtención del conjunto de datos

Para empezar, es necesario una base de datos con personajes mitológicos. Se puede definir un personaje mitológico como un personaje que aparece en los mitos de una cultura, ¿pero qué es exactamente un mito? La definición de mito es bastante difusa, está claro que el mito de Prometeo es un mito, ¿pero qué pasa con el mito de Papá Noel o el del Ratoncito Pérez? ¿Pueden ser considerados mitos de nuestra sociedad moderna?

Actualmente se intentan diferenciar los mitos con las leyendas, las ficciones poéticas y las fábulas, aunque parece que la distinción no está claramente definida, por lo que se sigue teniendo el problema de definir un personaje mitológico. Para obtener el conjunto de datos y solucionar este problema de definición, se acude a Wikipedia.

Wikipedia es una de las plataformas de contenido libre más completas de internet, por lo que es seguro que aparecerán una gran cantidad de los personajes mitológicos que estamos buscando. Utilizando Wikipedia no es necesario proponer una definición de personaje mitológico, ya que es la propia comunidad hispanohablante de Wikipedia la que ha diferenciado lo que es un personaje mitológico de lo que no lo es.

Los artículos de Wikipedia están distribuidos en categorías en función del concepto al que hace referencia cada artículo, por lo que es posible acceder a las categorías de personajes mitológicos y obtener los artículos incluidos. Si un artículo pertenece a una categoría de personaje mitológico, lo identificaremos como tal. Además, el propio nombre de la categoría nos ofrece información importante sobre los personajes que contiene, como la mitología a la que pertenece y el rol que asume dentro de la mitología. Por ejemplo, sabemos que los personajes dentro de la categoría *Criaturas de la mitología nórdica* tienen el rol de “criatura”, y pertenecen a la mitología nórdica.

La obtención de los datos se ha realizado en Python. Las librerías usadas para el scrapeo de Wikipedia han sido “BeautifulSoup” y “Selenium”.

La primera estrategia para obtener los artículos de los personajes mitológicos consiste en acceder a todas las categorías específicas de personajes mitológicos.

El conjunto de mitologías se ha obtenido de tres listas diferentes de Wikipedia [1](#) [2](#) [3](#). Por otro lado, tras una exploración de las categorías, hemos clasificado los personajes mitológicos entre siete roles distintos: Personaje, Deidad, Criatura, Demonio, Gigante, Héroe y Dragón. Con las combinaciones de estos dos elementos (mitología y rol) se han encontrado la mayoría de las categorías específicas de personajes mitológicos. Existen personajes mitológicos que pertenecen a categorías que no siguen una combinación de los elementos anteriores, esto ocurre por ejemplo con la categoría *Titanes*, que son deidades de la mitología griega. Por lo tanto, para completar el conjunto de datos, se han añadido estas categorías manualmente (Apéndice 1). Una vez se tienen todas las categorías, se obtienen todos los personajes pertenecientes a ellas.

La segunda estrategia consiste en encontrar personajes que se encuentren en diversas categorías simultáneamente. Por ejemplo, si un artículo que pertenece tanto a la categoría *Gigante* como a la categoría *Mitología mexicana*, sabemos que ese artículo pertenece a un personaje con el rol de “gigante” de la mitología mexicana.

Con la metodología anterior se ha obtenido una gran cantidad de artículos de Wikipedia de personajes mitológicos, la mitología a la que pertenecen, y el rol que tienen. Algunos personajes importantes han sido añadidos manualmente debido a que la gran variabilidad de categorías que no son exclusivas de personajes mitológicos. Se extrae el texto de los artículos de los personajes (ignorándose algunos apartados como “Véase también”, “Bibliografía”,...) para realizar el análisis de texto, y también se obtienen los enlaces a artículos de otros personajes mitológicos para reali-

zar el análisis de grafos. Además se tiene información de que un personaje es femenino debido a que hay categorías que contienen únicamente “Diosas”. Esta variable está claramente sesgada: existen personajes femeninos que únicamente pertenecen a una categoría de género neutro, por lo que a priori no se puede saber si se trata de un personaje femenino. A pesar de ello, puede ser interesante filtrar los resultados para los personajes que sabemos con total seguridad que son femeninos.

La región y el continente de la mitología son atributos que nos pueden interesar en el análisis. Para ello, creamos sus tablas correspondientes a partir de esta [lista](#) en la que se relacionan mitologías, regiones y continentes. Aquí no aparecen todas las mitologías obtenidas anteriormente, por lo que se añaden manualmente las regiones a las mitologías restantes.

Con todo esto, ya tenemos nuestro modelo de datos (Apéndice 2). Está compuesto por la tabla de clases (roles), tabla de personajes, tabla de mitologías, tabla de regiones y tabla de continentes. También hay dos tablas que relacionan la tabla de personajes con las tablas de clases y mitologías debido a que tienen relaciones m:n, es decir, un personaje puede tener más de un rol y pertenecer a más de una mitología (y evidentemente un rol o una mitología contiene más de un personaje).

Análisis de grafos y texto

Para crear la herramienta de análisis de personajes mitológicos, debemos añadir a la base de datos los resultados de los análisis de grafos y texto (Apéndice 3). En total se añaden cuatro nuevas tablas y nuevas variables a la tabla de personajes.

La creación y transformación de estas tablas se ha realizado en R. Se ha usado la librería “igraph” para el análisis de grafos y “tidytext” para el análisis de texto.

Relación entre las mitologías

Un personaje puede pertenecer a diversas mitologías, por esto mismo se ha creado una nueva tabla que contiene la cantidad de personajes que comparte cada par de mitologías. Esto es considerado como una medida de semejanza entre las mitologías, por lo que será usado para la visualización de un grafo.

Relación entre los personajes

Además del modelo de datos comentado anteriormente, se han extraído los enlaces entre los diferentes artículos de Wikipedia de los personajes mitológicos. Si un personaje tienen un enlace al artículo de otro personaje, están relacionados. Con esto se ha creado una nueva tabla con las relaciones entre los personajes para realizar visualizaciones de grafos.

Se ha creado un grafo dirigido utilizando estas relaciones entre los personajes. Este grafo está formado por 3.622 nodos (personajes) y 18.852 aristas (relaciones entre personajes). Es un grafo no conexo con 315 componentes. El orden del mayor componente es de 3277 (90 % de los nodos), por lo que se puede decir que el grafo tiene un componente gigante. Es un grafo poco denso (el 0.14 % de las aristas posibles están definidas), aunque no se puede decir que sea sparse.

Parece que el grafo tiene una estructura “scale-free”, es decir, unos pocos personajes tienen muchas relaciones aunque generalmente se trata de un grafo poco denso. La distribución de grados de los nodos sigue una “power-law”, esto es un indicio de la estructura “scale-free”.

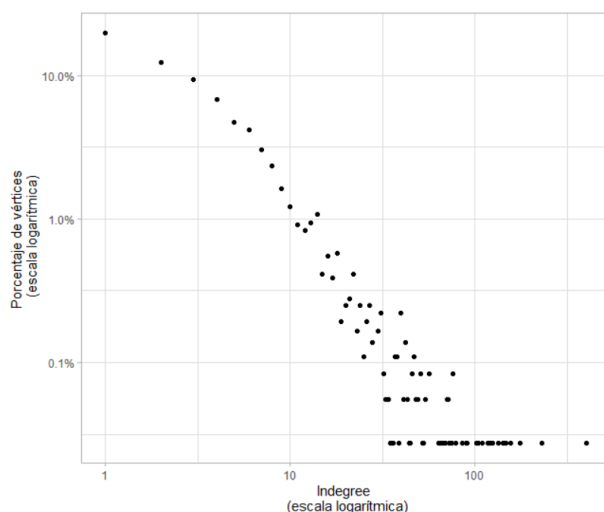


Figura 1: Distribución de grados

Un personaje mitológico está relacionado (indegree), de media, con otros 5 personajes mitológicos.

Se han obtenido diferentes medidas de centralidad de los nodos: el Page Rank y el Betweenness centrality. Estas métricas pueden ser usadas como índices de importancia de los personajes, por lo que se añaden la tabla de personajes de la base de datos.

Análisis de texto

El análisis de texto se ha realizado tanto para analizar los personajes como las mitologías, por lo que se han obtenido dos tablas diferentes.

Para el análisis de los personajes se ha usado el corpus formado por el texto en los artículos de los personajes. Se han eliminado las stopwords típicas obtenidas de diversas fuentes, los nombres de los personajes, los nombres de las mitologías y algunas palabras que aparecen como texto auxiliar en Wikipedia (“cita”, “requerida” y “editar”). Una vez hecho esto, se crea una tabla con la frecuencia y el “term frequency - inverse document frequency” (tf-idf) para cada combinación de personaje-palabra.

Se crea otra tabla similar usando la misma metodología, pero agregando los textos de los artículos por mitologías.

Con estas tablas es posible analizar el uso de los términos más usados en los artículos de Wikipedia. Además, con el uso del tf-idf, podemos descubrir cuales son los términos que caracterizan a un personaje o mitología.

Visualización de los resultados

Se ha creado una herramienta de visualización de los resultados obtenidos utilizando el software Power BI de Microsoft. El resultado está publicado en la web y puede encontrarse en el siguiente [link](#).

El informe de visualización cuenta principalmente con 6 páginas diferentes: (1) Exploración de la Base de Datos, (2) Relación entre Mitologías, (3) Relación entre Personajes, (4) Temática de los Personajes, (5) Análisis de Mitología y (6) Análisis de Personaje.

Exploración de la Base de Datos

Con esta primera página se puede explorar fácilmente el conjunto de los datos. Se puede observar la cantidad mitologías y de personajes totales, la cantidad de personajes por mitología, por región, por continente y por clase. También se pueden ver los índices de importancia de cada personaje (Page Rank y Betweenness Centrality), la cantidad de palabras en el corpus (no se cuentan las stopwords eliminadas anteriormente), y la cantidad de apariciones de cada palabra.

Es posible acceder directamente a un análisis de cualquier mitología, personaje o palabra desde las tablas.

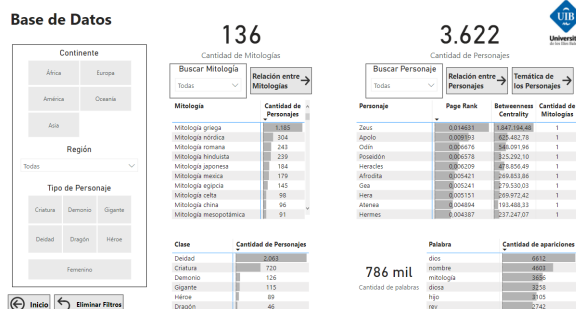


Figura 2: Exploración de la Base de Datos

Para las anteriores visualizaciones, se pueden filtrar los personajes por continente, región, rol del personaje y género femenino.

Relación entre Mitologías

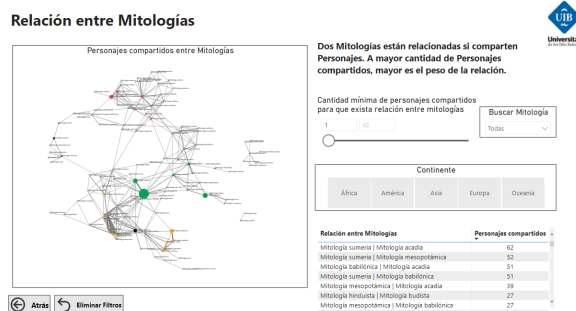


Figura 3: Relación entre Mitologías

La visualización principal de esta página es un grafo con las relaciones entre las distintas mitologías. Dos mitologías están relacionadas si comparten algún personaje mitológico. Se puede cambiar la cantidad mínima de personajes compartidos entre las mitologías para que exista una relación entre ellas. También se puede filtrar por una mitología o continente específico. Al mismo tiempo existe una tabla en la que se observa el peso de las relaciones entre las mitologías.

Relación entre Personajes

En esta página se el grafo con las relaciones entre los personajes analizado en el apartado anterior. La relación entre dos personajes puede ser débil o fuerte dependiendo de si la dirección del enlace es unidireccional o bidireccional. Se puede filtrar el grafo por personaje, mitología, continente y tipo de relación.

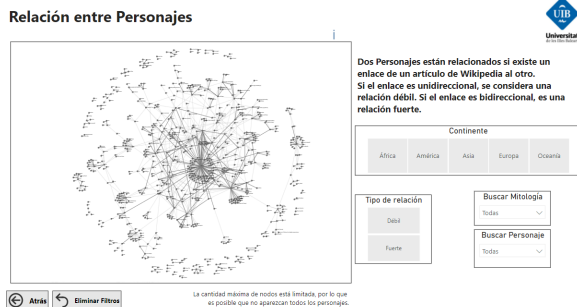


Figura 4: Relación entre Personajes

Si no se aplica ningún filtro, no aparecen todos los personajes dentro de la visualización del grafo debido a la limitación de cantidad de aristas de la herramienta. Cuando se supera el límite máximo de aristas impresas, aparece una señal de aviso en la visualización.

Temática de los Personajes

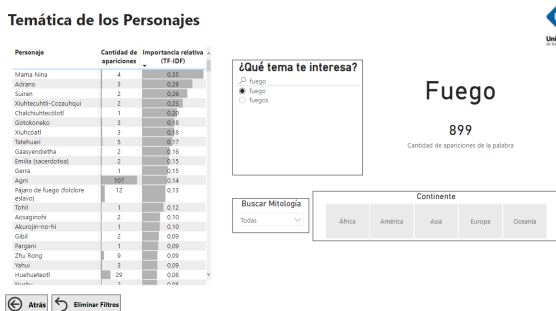


Figura 5: Temática de los Personajes

Página de análisis de las palabras. El objetivo de esta página es encontrar los personajes mitológicos que tengan mayor relación con la palabra de interés. Se inserta una palabra como input y aparecen los personajes con mayor valor tf-idf en relación a esa palabra. También aparece la cantidad de apariciones de la palabra.

Los personajes pueden filtrarse tanto por mitología como por continente.

Análisis de Mitología

Página de análisis de las mitologías. En esta página se observa la cantidad de personajes de la mitología, los personajes de la mitología con sus índices de importancia, las mitologías relacionadas, las palabras más importantes de la mitología (tanto frecuencia como

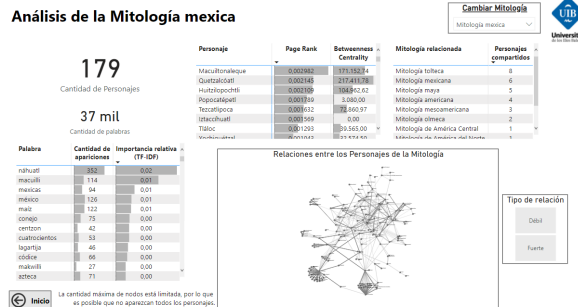


Figura 6: Análisis de Mitología

tf-idf) y el grafo de relaciones entre los personajes de la mitología. Se pueden filtrar este grafo en función del tipo de relación (débil o fuerte) y cambiar de mitología analizada.

Análisis de Personaje

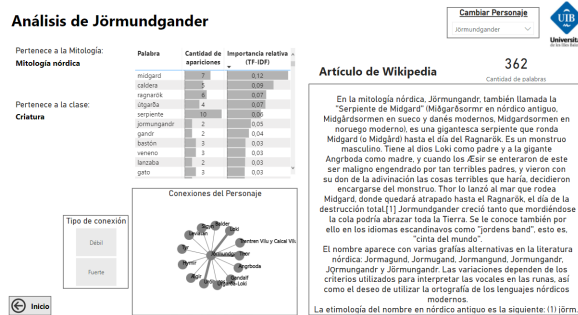


Figura 7: Análisis de Personaje

Página de análisis de los personajes. Se puede observar tanto la mitología a la que pertenece el personaje como su rol. También se observa la frecuencia y el tf-idf de las palabras que aparecen en el artículo del personaje, el texto del artículo y el grafo con las relaciones que tiene con otros personajes. Se puede cambiar el tipo de relación que se observa en el grafo y el personaje analizado.

Conclusiones

En este trabajo se ha creado una herramienta útil para explorar y analizar personajes mitológicos y mitologías utilizando únicamente datos de Wikipedia. Se han desarrollado todas las etapas del proceso de creación de la herramienta: Extracción de los personajes de Wikipedia, modelado, análisis y visualización de los

datos en web. Todo ello con el uso de diferentes tipos de software en función de cada necesidad específica, principalmente Python para el scrapeo de los datos, R para el análisis, y Power BI para la visualización de los resultados.

Se ha visto que casi un tercio de los personajes pertenecen a la mitología griega. Este sesgo seguramente se deba a que es la mitología más popular en la sociedad hispanohablante, por lo que sería interesante obtener datos similares con los artículos de Wikipedia de otros idiomas y analizar sus diferencias.

En este [repositorio de GitHub](#) se pueden encontrar todos los archivos referentes a este trabajo: El scrapeo en Python, los datos, el análisis en R, y la visualización en Power BI.

Bibliografía

- García Gual, C. (1992). Introducción a la mitología griega. Alianza Editorial.
- Kirk, G. S. (1970). El mito. Paidós.
- Silge, J. y Robinson, D. (2017). Text Mining with R: A Tidy Approach. O'Reilly. <https://www.tidytextmining.com/>

Apéndice 1

Mitología griega

Conceptos personificados (deidad)
Grayas (diosa)
Horas (diosa)
Oceánidas (dios)
Ourea (deidad)
Autómatas (criatura)
Centauros (criatura)
Monstruos (criatura)
Harpías (criatura)
Sátiros (criatura)
Diosas del mar (diosa)
Nereidas (diosa)
Oceánides (diosa)
Erinias (diosa)
Musas (diosa)
Ninfas (diosa)
Dríades (diosa)
Náyades (diosa)
Pléyades (diosa)
Dioses ctónicos (deidad)
Dioses del mar (dios)
Dioses olímpicos (deidad)
Dioses olímpicos menores (deidad)
Dioses primordiales (deidad)
Titanes (deidad)
Dioses del viento (deidad)
Deidades (deidad)

Mitología nórdica

Caballos (criatura)
Lobos (criatura)
Ásynjur (diosa)
Nornas (diosa)
Æsir (dios)
Valquirias (diosa)
Vanir (deidad)
Enanos (criatura)
Jotuns (gigante)

Mitología hinduista

Avatares (personaje)
Deidades rigvédicas (deidad)
Razas no humanas (personaje)
Apsará (diosa)
Demonios (demonio)
Ráksasas (demonio)

Mitología japonesa

Divinidades sintoistas (deidad)

Yōkai (criatura)

Myō-ō (deidad)

Mitología ainu

Kamui (deidad)

Mitología budista

Ráksasas (demonio)

Criaturas (criatura)

Deidades femeninas (diosa)

Dakinis (diosa)

Deidades del budismo tibetano (deidad)

Dharmapalas (deidad)

Myō-ō (deidad)

Devas (deidad)

Seres (personaje)

Mitología mexicana

Criaturas (criatura)

Cihuateteo (diosa)

Macuiltonaleque (dios)

Deidades del amor y la fertilidad (deidad)

Centzon Huitznáhuac (deidad)

Centzon Totochtin (deidad)

Cinteteo (dios)

Conceptos personificados (deidad)

Dioses del viento (deidad)

Dioses primordiales (deidad)

Divinidades de la muerte (deidad)

Tezcatlipocas (deidad)

Tlaloque (deidad)

Tonalteteuctin (deidad)

Mitología china

Los ocho inmortales (deidad)

Tres augustos y cinco emperadores (personaje)

Mitología celta

Deidades britanas (deidad)

Fomorianos (deidad)

Tuatha Dé Danann (deidad)

Mitología irlandesa

Fomorianos (deidad)

Tuatha Dé Danann (deidad)

Mitología yoruba

Orisha (deidad)

Mitología del judaísmo

Criaturas (criatura)

Ángeles (criatura)

Mitología persa

Daevas (demonio)

Yazatas (deidad)

Mitología cántabra

Criaturas (criatura)

Mitología asturiana

Criaturas (criatura)

Mitología de América del Norte

Criaturas (criatura)

Mitología de Oriente Medio

Criaturas (criatura)

Dioses (deidad)

Diosas elamitas (diosa)

Dioses elamitas (dios)

Mitología del islam

Criaturas (criatura)

Ángeles (criatura)

Apéndice 2

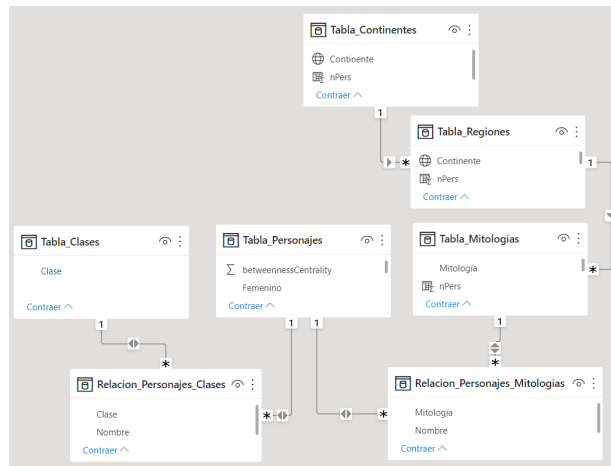


Figura 8: Modelo de datos inicial

Apéndice 3

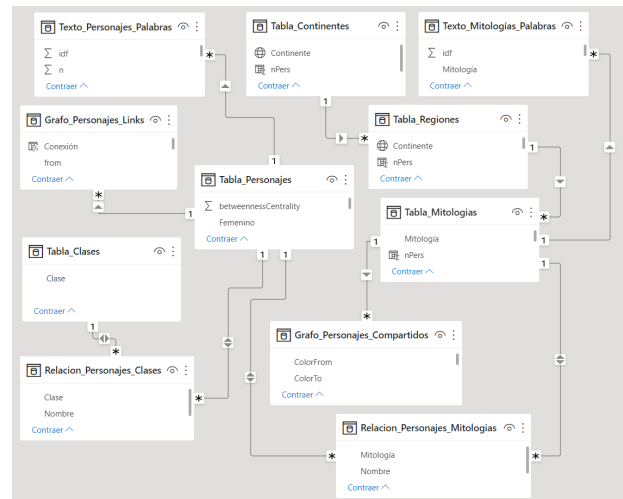


Figura 9: Modelo de datos con resultados