

Universidad Politécnica de Madrid

MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL

SISTEMA DE PREGUNTA-RESPUESTA

Asignatura: Ingeniería Lingüística

Autores:

Joel Pardo Ferrera

Daniel Carmona Pedrajas

Correos:

joel.pardof@alumnos.upm.es

daniel.carmonap@alumnos.upm.es

15 Enero 2023

Índice

1. Introducción	2
2. Repositorios	2
3. Definición del dominio	2
3.1. Dimensiones del dominio	3
3.2. Valores de dimensiones de nuestros repositorios	4
4. Definición de preguntas tipo y parafraseo	5
5. Definición de reglas de identificación de dimensión	12
5.1. Pronombres interrogativos	12
5.2. Verbos	13
5.3. Palabras clave	13
6. Experimentación	13
6.1. Implementación	14
6.2. Validación	15
7. Conclusiones	16
A. Repositorios	16
A.1. Repositorio de Steve Jobs	16
A.2. Repositorio de Elon Musk	18
B. Pesos	19
B.1. Pesos de los pronombres interrogativos	19
B.2. Listas de verbos para cada dimensión	19
B.3. Listas de palabras clave para cada dimensión	22

1. Introducción

En esta memoria se recoge el proceso de diseño e implementación de un sistema de pregunta-respuesta. Este tipo de sistema se encuentra dentro de los llamados sistemas de extracción de información, que tratan de encontrar datos que satisfagan una necesidad de información dentro de una colección extensa de documentos de una naturaleza desestructurada, es decir, que a priori no puede ser procesada por un ordenador [1].

En nuestro caso, contaremos con un repositorio de información estructurado y el objetivo será diseñar un sistema que sea capaz de entender por qué dimensión de nuestro repositorio nos están preguntando para así responder con el dato que el usuario está buscando.

Seguiremos los siguientes pasos:

- En la **Sección 2** comentamos la temática de nuestros repositorios y explicamos por qué la hemos elegido.
- En la **Sección 3** analizamos nuestros repositorios para definir el dominio sobre el que trabajaremos y obtener las dimensiones del mismo. Una vez hecho esto, definiremos los valores de cada dominio para nuestros repositorios de información por los que el usuario preguntará.
- En la **Sección 4** definimos las preguntas tipo para cada una de las dimensiones y las parafraseamos para entender mejor qué caracteriza a cada dimensión.
- En la **Sección 5** extraemos las reglas que nos permitan identificar a qué dimensión pertenece una pregunta formulada por el usuario en base a las preguntas tipo y parafraseo de la Sección 4.
- En la **Sección 6** explicamos algunos detalles de implementación y probamos nuestro sistema.

2. Repositorios

Nuestro sistema de pregunta respuesta se centra en el campo de las biografías. Hemos seleccionado este campo porque incluye una extensa colección de tipos de datos como pueden ser datos numéricos (edad, número de hijos), datos temporales (fecha de nacimiento, fecha de fallecimiento), datos espaciales (lugar de nacimiento) y más. En resumen, una biografía contiene tanto datos cualitativos como cuantitativos lo que nos dará un mayor rango de posibles preguntas posibilitando así el diseño de un sistema de pregunta respuesta que sea fácilmente escalable, robusto y que pueda abstraerse a otros dominios.

Lo primero que necesitamos para realizar el sistema de pregunta-respuesta es un repositorio de información. El repositorio de información es el lugar en el que se encuentra la información que queremos buscar, se puede encontrar de manera estructurada (en una base de datos relacional) o no estructurada (textos). Nosotros hemos decidido usar dos textos con información biográfica de personalidades del mundo tecnológico: [Steve Jobs](#) y [Elon Musk](#) cuyas biografías hemos obtenido en [Wikipedia](#), una enciclopedia colaborativa online creada y mantenida por usuarios.

Se pueden consultar los repositorios completos en el apéndice [A](#).

3. Definición del dominio

Hasta ahora hemos estado hablando de **dominios** y **dimensiones** pero no nos hemos parado a explicar estos conceptos, pues bien:

- **Dominio:** Es un subespacio del espacio de información que contiene ciertas dimensiones que lo caracterizan y distinguen de los demás.
- **Dimensión:** Es el componente por el que está compuesto un dominio. Una dimensión define un tipo de dato que contiene una unidad de información y que está presente en una gran parte de las instancias o repositorios de un dominio.

Para determinar si un repositorio pertenece a un dominio ya creado habría que comprobar qué número de dimensiones del dominio tienen un valor no nulo para ese mismo repositorio, si obtenemos una gran cantidad de dimensiones con valor no nulo quiere decir que podríamos asignar el repositorio al dominio en cuestión. En esta práctica no hemos realizado ese proceso porque en este momento no tenemos definido nuestro dominio sino que lo vamos a crear, extrayendo de nuestro repositorio las dimensiones del dominio de biografía.

3.1. Dimensiones del dominio

Para obtener las dimensiones de nuestro dominio hemos comparado ambas biografías y extraído dimensiones comunes a ambos repositorios, creando así la tabla 1 donde la columna *Dimensión* representa el nombre de una dimensión detectada y la columna *Observación* detalla qué información contiene y qué valor asignar a esa dimensión en casos atípicos.

Dimensión	Observación
Primer Apellido	El primer apellido del padre de la persona.
Segundo Apellido	El primer apellido de la madre de la persona.
Lugar de nacimiento	El país de nacimiento. Si no se conoce asignar el valor "Desconocido"
Día de nacimiento	Si no se conoce asignar el día 1.
Mes de nacimiento	Si no se conoce asignar el mes de enero.
Año de nacimiento	Si no se conoce se debe inferir de la edad.
Fecha de nacimiento	Dimensión compuesta por día, mes y año de nacimiento en el formato dd-mm-yyyy.
Edad	Si no se conoce se debe inferir del año de nacimiento, si ha fallecido asignar la edad con la que falleció.
Día de fallecimiento	Si la persona no ha fallecido asignar el día 0.
Mes de fallecimiento	Si la persona no ha fallecido asignar el mes "Ninguno".
Año de fallecimiento	Si la persona no ha fallecido asignar el año 0.
Fecha de fallecimiento	Dimensión compuesta por día, mes y año de fallecimiento en el formato dd-mm-yyyy. Si la persona no ha fallecido hasta el momento actual asignar el valor 00-00-0000.
Marido	Si la persona ha tenido más de un marido, asignar la lista de sus nombres. Si la persona no ha tenido marido, asignar el valor "No".
Esposa	Si la persona ha tenido más de una esposa, asignar la lista de sus nombres. Si la persona no ha tenido esposa nunca, asignar el valor "No".
Pareja actual	Asignar el nombre de la pareja actual de la persona. Si la persona no tiene pareja actual asignar el valor de "Soltero".
Número de hijos	El número de hijos que ha tenido la persona.
Nombre del padre	Si se desconoce el nombre del padre, asignar el valor "Desconocido"
Nombre de la madre	Si se desconoce el nombre de la madre, asignar el valor "Desconocido"
Universidad	Si la persona ha accedido a más de una universidad, asignar la lista de nombres de la universidad. Si la persona no ha accedido a la universidad, asignar el valor "Ninguna".

Carrera universitaria	Si la persona ha obtenido más de una carrera, asignar la lista de carreras universitarias completadas. Si la persona no ha conseguido un título universitario asignar el valor "Ninguna".
Puestos empresariales	Una lista de posiciones en las empresas que haya trabajado la persona. Si la persona no ha trabajado nunca, asignar el valor "Ninguno".

Tabla 1: Dimensiones y observaciones del dominio Biografía

3.2. Valores de dimensiones de nuestros repositorios

En la tabla 4 podemos ver el valor de las dimensiones para cada uno de nuestros repositorios.

Dimensión	Elon Musk	Steve Jobs
Nombre	Elon	Steve
Primer Apellido	Musk	Jobs
Segundo Apellido	Reeve	Hagopian
Lugar de nacimiento	Sudáfrica	Estados Unidos
Día de nacimiento	28	24
Mes de nacimiento	Junio	Febrero
Año de nacimiento	1971	1955
Fecha de nacimiento	28-06-1971	24-02-1995
Edad	51	56
Día de fallecimiento	0	5
Mes de fallecimiento	Ninguno	Octubre
Año de fallecimiento	0	2011
Fecha de fallecimiento	00-00-0000	05-11-2011
Marido	No	No
Esposa	Justine Wilson, Talulah Riley	Laurene Powell, Chrisann Brennan
Pareja actual	Grimes	No
Número de hijos	10	4
Nombre del padre	Errol	Paul
Nombre de la madre	Maye	Clara
Universidad	Pretoria, Queen's, Pennsylvania	Reed, Portland, Oregón
Carrera universitaria	Física, Economía	Ninguna
Puestos empresariales	"Fundador, CEO, jefe ingeniero de Space", "CEO y arquitecto de producto en Tesla", "CEO de Twitter", "Presidente la Fundación Musk", "Founder of The Boring Company", "Co-fundador de Neurallink", "Co-fundador of OpenAI", "Cofundador de Zip2"	"Cofundador de Apple", "Cofundador de Pixar", "Empleador Apple, Pixar, NeXT y Atari", "Creador de Apple I, Apple Lisa, iMac, iPod, iPhone, iPad, Mac OS X v10.0, NEXTSTEP, iTunes y App Store"

Tabla 2: Valores de dimensiones de nuestros repositorios

4. Definición de preguntas tipo y parafraseo

En este punto vamos a definir la pregunta tipo para cada dimensión y posteriormente, buscaremos parafraseos de las mismas que nos faciliten encontrar patrones con las que clasificar las preguntas del usuario en la siguiente sección. Por cada pregunta escribiremos de 8 a 10 parafraseos.

Pregunta tipo para la dimensión Nombre: ¿Cuál es su nombre?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se llama?
- ¿Cómo le bautizaron sus padres?
- ¿Con qué nombre le llaman sus padres?
- ¿Cuál es su apodo?
- ¿Cuál es el nombre que le dieron sus padres?
- ¿Por qué nombre le llaman sus padres?
- ¿Qué nombre tiene?
- ¿Qué nombre le han dado sus padres?
- ¿Qué nombre le han puesto sus padres?
- ¿Qué nombre aparece en su documento de identidad?
- ¿Quién es?

Pregunta tipo para la dimensión Primer Apellido: ¿Cuál es su primer apellido?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se apellida?
- ¿Cómo se llama de segundo nombre?
- ¿Cómo se conoce a su familia?
- ¿Cómo se conoce a la familia de su padre?
- ¿Cuál es el apellido de su padre?
- ¿Cuál es el nombre de su familia?
- ¿Cuál es el nombre de la familia de su padre?
- ¿Cuál es su segundo nombre?
- ¿Qué apellido tiene como primer apellido?
- ¿Qué apellido tiene la familia de su padre?

Pregunta tipo para la dimensión Segundo Apellido: ¿Cuál es su segundo apellido?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se llama de tercer nombre?
- ¿Cómo se conoce a la familia de su madre?
- ¿Cuál es su último apellido?
- ¿Cuál es el apellido de su madre?
- ¿Cuál es el apellido que va después del primer apellido?
- ¿Cuál es el nombre de la familia de su madre?
- ¿Cuál es su tercer nombre?
- ¿Qué apellido tiene como segundo apellido?
- ¿Qué apellido tiene la familia de su madre?

Pregunta tipo para la dimensión Lugar de Nacimiento: ¿Dónde nació?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su lugar de nacimiento?
- ¿Cuál es el lugar en el que nació?
- ¿Cuál es el lugar en el que ha nacido?
- ¿Cuál es el país en el que nació?
- ¿Cuál es el país en el que ha nacido?
- ¿Cuál es su nacionalidad?
- ¿Dónde se encuentra su lugar de nacimiento?
- ¿En qué país nació?
- ¿En qué lugar nació?
- ¿En qué país le dió a luz su madre?
- ¿Qué país es su país de nacimiento?

Pregunta tipo para la dimensión Día de nacimiento: ¿Qué día nació?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su día de nacimiento?
- ¿Cuál es el día en el que nació?
- ¿Cuál es el número del día en el que nació?
- ¿En qué día nació?

Pregunta tipo para la dimensión Mes de nacimiento: ¿Qué mes nació?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su mes de nacimiento?
- ¿Cuál es el mes en el que nació?
- ¿En qué mes nació?

Pregunta tipo para la dimensión Año de nacimiento: ¿Qué año nació?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su año de nacimiento?
- ¿Cuál es el año en el que nació?
- ¿De qué año es?
- ¿En qué año nació?

Estas tres dimensiones no se suelen preguntar de forma separada, es por esto que vamos a definir aquí una **dimensión compuesta** llamada **Fecha de nacimiento** que concatena las dimensiones de Día de nacimiento, Mes de nacimiento y Año de nacimiento en el formato dd-mm-yyyy.

Pregunta tipo para la dimensión Fecha de nacimiento: ¿Cuándo nació?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuándo fue su nacimiento?
- ¿Cuál es su fecha de nacimiento?
- ¿Cuál es la fecha en la que nació?
- ¿Cuál es la fecha en la que ha nacido?
- ¿Cuándo fue su nacimiento?
- ¿En qué fecha fue dado a luz?
- ¿En qué fecha nacio?
- ¿En qué fecha ha nacido?
- ¿Qué día nació? (Habrá ambigüedad con Día de nacimiento)
- ¿Qué día fue dado a luz?

Pregunta tipo para la dimensión Edad: ¿Qué edad tiene?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cómo de mayor es?
- ¿Cuál es su edad?
- ¿Cuán mayor es?
- ¿Cuántos años tiene?
- ¿Cuántos años han pasado desde su nacimiento?
- ¿Cuántos años se ha mantenido con vida?
- ¿Cuántos años ha vivido?
- ¿Hace cuántos años nació?
- ¿Qué tiempo total de vida tiene?
- ¿Qué periodo de tiempo ha estado vivo?
- ¿Qué número de años tiene?

Pregunta tipo para la dimensión Día de fallecimiento: ¿Qué día falleció?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su día de fallecimiento?
- ¿Cuál es el día en el que falleció?
- ¿Cuál es el día en el que murió?
- ¿Cuál es el número del día en el que falleció?
- ¿Cuál es el número del día en el que murió?
- ¿En qué día falleció?

Pregunta tipo para la dimensión Mes de fallecimiento: ¿Qué mes falleció?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su mes de fallecimiento?
- ¿Cuál es el mes en el que falleció?
- ¿En qué mes falleció?
- ¿En qué mes murió?

Pregunta tipo para la dimensión Año de fallecimiento: ¿Qué año falleció?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es su año de fallecimiento?
- ¿Cuál es el año en el que falleció?
- ¿Cuál es el año en el que murió?
- ¿En qué año falleció?
- ¿En qué año murió?

Análogamente a la dimensión de Fecha de nacimiento, creamos una **dimensión compuesta** para el fallecimiento; **Fecha de fallecimiento** que concatena las dimensiones de Día de fallecimiento, Mes de fallecimiento y Año de fallecimiento en el formato dd-mm-yyyy.

Pregunta tipo para la dimensión Fecha de fallecimiento: ¿Cuándo falleció?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuándo fue su fallecimiento?
- ¿Cuándo fue su muerte?
- ¿Cuál es la fecha de su muerte?
- ¿Cuál es su fecha de fallecimiento?
- ¿Cuál es la fecha en la que falleció?
- ¿Cuál es la fecha en la que murió?
- ¿Cuál es la fecha en la que ha muerto?
- ¿Cuándo fue su fallecimiento?
- ¿Cuándo fue su muerte?
- ¿En qué fecha falleció?

- ¿En qué fecha murió?
- ¿En qué fecha ha muerto?
- ¿Qué día murió? (Habrá ambigüedad con Día de fallecimiento)
- ¿Qué día falleció? (Habrá ambigüedad con Día de fallecimiento)

Pregunta tipo para la dimensión Marido: ¿Quién es su marido?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Con qué hombre está casada/o?
- ¿Con qué hombre ha contraído matrimonio?
- ¿Cómo se llama su marido?
- ¿Cómo se llama el hombre con el que está casado?
- ¿Cómo se llama el hombre con el que ha contraído matrimonio?
- ¿Cuál es el nombre de su marido?
- ¿Cuál es su marido?
- ¿Qué nombre tiene su marido?
- ¿Quién es el hombre con el que se ha casado?

Pregunta tipo para la dimensión Esposa: ¿Quién es su esposa?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Con qué mujer está casada/o?
- ¿Con qué mujer ha contraído matrimonio?
- ¿Cómo se llama su mujer?
- ¿Cómo se llama la mujer con la que está casado?
- ¿Cómo se llama la mujer con la que ha contraído matrimonio?
- ¿Cuál es el nombre de su mujer?
- ¿Cuál es su mujer?
- ¿Qué nombre tiene su mujer?
- ¿Quién es la mujer con la que se ha casado?

Pregunta tipo para la dimensión Pareja actual: ¿Quién es su pareja actual?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se llama su pareja actual?
- ¿Con quién mantiene una relación?
- ¿Con quién mantiene una relación sentimental?
- ¿Con qué persona mantiene una relación sentimental?
- ¿Cuál es el nombre de su pareja actual?
- ¿Cuál es su pareja actual?
- ¿Qué persona es su pareja actual?
- ¿Qué nombre tiene su pareja actual?
- ¿Quién es la persona con la que tiene una relación actualmente?

Pregunta tipo para la dimensión Número de hijos: ¿Cuántos hijos tiene?

Esta dimensión es cuantitativa. Los parafraseos que hemos detectado son:

- ¿Cuál es el número de hijos que tiene?
- ¿Cuántos retoños tiene?
- ¿De cuántas personas es padre/madre?
- ¿Qué número de hijos tiene?
- ¿Qué número de retoños tiene?
- ¿Qué cantidad de hijos ha tenido?

Pregunta tipo para la dimensión Nombre del padre: ¿Quién es su padre?

Esta dimensión es cualitativa, podemos formular preguntas con ¿qué?, ¿cómo?, ¿cuál?. Los parafraseos que hemos detectado son:

- ¿Cómo se llama su padre?
- ¿Cómo se llama el hombre que le crió?
- ¿Cómo se llama el hombre del que descende?
- ¿Cómo se llama su progenitor?
- ¿Cuál es el nombre de su padre?
- ¿Cuál es el nombre de su progenitor?
- ¿Cuál es el nombre del hombre del que descende?
- ¿Qué nombre tiene su padre?
- ¿Quién es su progenitor?
- ¿Quién es el hombre del que descende?

Pregunta tipo para la dimensión Nombre de la madre: ¿Quién es su madre?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se llama su madre?
- ¿Cómo se llama la mujer que le crió?
- ¿Cómo se llama la mujer del que descende?
- ¿Cómo se llama la mujer de la que nació?
- ¿Cómo se llama su progenitora?
- ¿Cuál es el nombre de su madre?
- ¿Cuál es el nombre de su progenitora?
- ¿Cuál es el nombre de la mujer del que descende?
- ¿Cuál es el nombre de la mujer de la que nació?
- ¿Qué nombre tiene su madre?
- ¿Quién es su progenitora?
- ¿Quién es la mujer de la que descende?
- ¿Quién es la mujer de la que nació?

Pregunta tipo para la dimensión Universidad: ¿En qué universidad estudió?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿A qué universidad asistió para completar sus estudios?
- ¿Cómo se llama la universidad en la que estudió?
- ¿Cuál es el nombre de la universidad en la que estudió?
- ¿Cuál es la universidad en la que estudió?
- ¿Dónde estudió?
- ¿Dónde ha estudiado?
- ¿Dónde obtuvo sus títulos universitarios?
- ¿Dónde realizó sus estudios universitarios?
- ¿Dónde ha cursado su carrera?
- ¿En qué universidad realizó sus estudios universitarios?
- ¿En qué universidad obtuvo sus títulos universitarios?

Pregunta tipo para la dimensión Carrera Universitaria: ¿Qué carrera tiene?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cómo se llama la carrera que tiene?
- ¿Cuál es la carrera que ha estudiado?
- ¿Cuál es su título universitario?
- ¿Cuál es el nombre de su título universitario?
- ¿Cuál es el nombre de su carrera universitaria?
- ¿Cuáles son sus estudios?
- ¿Qué título universitario tiene?
- ¿Qué carrera ha estudiado?
- ¿Qué estudios tiene?
- ¿Qué estudios ha realizado?
- ¿Qué ha estudiado?

Pregunta tipo para la dimensión Puestos empresariales: ¿En qué puestos ha trabajado?

Esta dimensión es cualitativa. Los parafraseos que hemos detectado son:

- ¿Cuál ha sido su puesto empresarial?
- ¿Cuál es su ocupación?
- ¿Cuál es su trabajo?
- ¿Cuál es su experiencia laboral?
- ¿Cuál es su carrera profesional?
- ¿De qué empresas ha sido parte?
- ¿Dónde ha trabajado?

- ¿En qué empresas ha trabajado?
- ¿Qué puestos ha ocupado?
- ¿Qué experiencia laboral tiene?
- ¿Qué experiencia laboral tiene?
- ¿Qué tipo de trabajo ha realizado?

5. Definición de reglas de identificación de dimensión

En esta sección vamos a definir unas reglas que nos permitan identificar a qué dimensión pertenece una pregunta. Algunas ideas que se nos pasan por la cabeza son, dada una pregunta:

- Utilizar algún tipo de red neuronal con mecanismos de atención para clasificar nuestra pregunta como alguna de las dimensiones de nuestro dominio, pero esta técnica requiere muchos datos de entrenamiento y tiempo, cosas de las que no disponemos dentro del marco temporal de la práctica.
- Utilizar algún modelo sintáctico que nos permita detectar los verbos y nombres de la pregunta para asignarle la dimensión correcta. El problema de esta opción es que construir este tipo de modelo será costoso y no tenemos los conocimientos ni la experiencia que tiene un lingüista que pueda crear una serie de estructuras comunes de pregunta según el dominio.

Lo que se nos ocurre entonces es crear un diccionario de palabras en base a nuestros parafraseos de la Sección 4 y para una pregunta calcular cierto *score* o puntuación que nos permita clasificarla en la dimensión correcta. La idea con este método es crear una tabla dimensión-palabra que en cada celda guarde cierto valor. Usaremos la tabla de la siguiente manera, para una pregunta recibida:

1. Crearemos un vector de puntuaciones de tamaño igual al número de dimensiones.
2. Para cada palabra de la actualizaremos nuestro vector de puntuaciones sumando el vector de la palabra que viene dado por su columna en la tabla dimensión-palabra. Si la palabra no se encuentra en la tabla su vector será el vector cero.
3. Al final, asumiremos que la pregunta pertenece a la dimensión para la que mayor puntuación haya obtenido.

Este método es algo rudimentario y tiene semejanzas con el modelo de espacio de vectores de la práctica de clasificación de textos con glosarios. Aún con todo, creemos que será capaz de cubrir con gran precisión los parafraseos definidos anteriormente y que podrá detectar correctamente otras estructuras de preguntas y parafraseos que no hemos tenido en cuenta.

A continuación, explicamos el procedimiento mediante el cual crearemos nuestra tabla dimensión-palabra, poniendo como ejemplo la dimensión de **Carrera universitaria**.

5.1. Pronombres interrogativos

Los pronombres interrogativos pueden darnos algo de información sobre la pregunta aunque se repitan en todas. Fijándonos en las preguntas tipo y su parafraseo, en las dimensiones cuantitativas los pronombres cuánto y cuándo aparecen con más frecuencia que las dimensiones cualitativas donde usamos pronombre como quién, dónde o qué. Para asignarle un valor o peso a un pronombre interrogativo p dada una dimensión d usaremos la siguiente fórmula:

$$v_p = n_p / n_d \quad (1)$$

Donde n_p es el número de preguntas (contando pregunta tipo y parafraseos) y n_d el número total de preguntas que hemos definido para la dimensión d . De esta forma, para la dimensión Carrera Universitaria, los pesos nos quedarían como se muestra en la tabla 3.

Cabe destacar que hemos decidido condensar los posibles géneros y números del pronombre en su forma masculina singular, por ejemplo, *cuántas* se contará como una repetición más de *cuántos*. Se pueden comprobar todos los pesos obtenidos con este método en el apéndice B.1 (tabla 5).

Cómo	Cuál	Cuán	Cuándo	Cuánto	Dónde	Qué	Quién
0.083	0.416	0	0	0	0	0.5	0

Tabla 3: Pesos de los pronombres interrogativos para la dimensión Carrera universitaria

5.2. Verbos

Nos fijamos que otro elemento que contiene información sobre la dimensión es el verbo que acompaña al nombre de la dimensión, por ejemplo, carrera suele estar acompañado del verbo estudiar. Además creemos que estos verbos dan más información que los pronombres interrogativos acerca de la dimensión a la que pertenece la pregunta formulada por el usuario por lo que decidimos darles un peso mayor. Decidimos darle a cada verbo el mismo peso, 2 puntos por verbo si éste aparece en alguno de los parafraseos. Así sólo necesitaremos una lista de verbos para cada dimensión. Usaremos el infinitivo para representar un verbo, pero comprobaremos si la frase contiene ese verbo en tercera persona y los tiempos verbales de presente, pretérito perfecto y pretérito perfecto compuesto. Como ejemplo mostraremos la lista de verbos para la dimensión de Carrera Universitaria que obtenemos de los parafraseos de las preguntas:

- **Estudiar**, comprobando “estudia”, “estudió” y “ha estudiado”.
- **Cursar**, comprobando “cursa”, “cursó” y “ha cursado”.
- **Tener**, comprobando “tiene”, “tuvo” y “ha tenido”.

De esta manera, la pregunta *¿Qué estudios tiene?* tendrá 2 puntos en la dimensión de Carrera Universitaria. Cabe destacar que hemos descartado el verbo “ser” porque está presente en la inmensa mayoría de parafraseos y no aporta demasiada información sobre la dimensión a la que la pregunta se refiere. Las listas de verbos (y sus formas) para todas las dimensiones se pueden encontrar en el apéndice B.2.

5.3. Palabras clave

Por último, detectamos que hay palabras clave que están fuertemente relacionadas con la dimensión. El nombre de la dimensión suele aparecer en la pregunta, también encontramos sinónimos y adjetivos en el parafraseo. Es por esto que a este tipo de palabras clave les vamos a dar el mayor peso, 5 puntos si aparecen en el nombre de la dimensión o en alguna de sus preguntas. Este proceso está hecho a mano, alguien con mayor conocimiento del dominio podría obtener palabras más representativas que las que encontraremos nosotros, incluso, si dispusiéramos de más repositorios podríamos extraer términos importantes en base a métricas como el tf-idf. Como no disponemos de lingüistas ni más repositorios, el resultado que obtendremos será posiblemente mejorable, pero nos encargaremos de que nuestro sistema sea escalable y modificable en la siguiente sección.

El resultado de este análisis para la dimensión de Carrera universitaria es el siguiente:

- **Carrera**, incluyendo “carreras”.
- **Universitario**, incluyendo “universitarios”, “universitaria” y “ha estudiado”.
- **Título**, incluyendo “títulos”.
- **Estudio**, incluyendo “estudios”.

De nuevo, asumimos que cada palabra clave engloba sus número singular y plural y su género masculino y femenino con algunas excepciones en las que el género o el número permita distinguir entre dimensiones como puede ser el caso de progenitor/padre y progenitora/madre para las dimensiones de Nombre del padre y Nombre de la madre. Para ver la lista completa de palabras clave por dimensión se puede consultar el apéndice B.3.

6. Experimentación

En esta sección explicaremos la implementación de nuestro sistema a nivel alto y lo validaremos con una serie de preguntas.

6.1. Implementación

Podemos dividir la implementación de este sistema en dos fases:

- **Cálculo de pesos**, en la que obtenemos nuestra tabla de pesos para nuestras palabras definidas en la sección 5
- **Procesamiento del input del usuario**, en el que procesamos las preguntas formuladas e intentamos dar la respuesta correcta.

Cálculo de pesos: Como hemos mencionado en secciones anteriores, queremos un sistema que sea flexible y escalable, es por esto que nuestro cálculo de pesos está diseñado de tal manera que si es necesario añadir preguntas, palabras o dimensiones se puedan recalcular dándole a un botón. El script [create_weights.py](#) se encarga de esto, dando como output el archivo [weights.csv](#) donde se guarda nuestra tabla de pesos.

La forma en la que podemos modificar las palabras o preguntas es la siguiente:

- Modificar o añadir **Parafraseos**: Accedemos a la carpeta [Preguntas/](#), seleccionamos el archivo de preguntas de la dimensión a modificar (o crear), por ejemplo, *preguntas_universidad.txt* y añadir una nueva pregunta que se tendrá en cuenta para ajustar los pesos.
- Modificar o añadir **Pronombres**: Accedemos a la carpeta [Pronombres/](#) y abrimos el archivo *pronombres.txt* añadiendo una nueva línea con formato `{representante}:{representante},{forma1},{forma2}....`. De esta forma se tendrán en cuenta estos pronombres a la hora de calcular los pesos como se especifica en la Sección 5.1.
- Modificar o añadir **Verbos**: Accedemos a la carpeta [Verbos/](#) y abrimos el archivo *formas_verbos.txt* añadiendo una nueva línea con formato `{representante}:{representante},{forma1},{forma2}....`. De esta forma se tendrán en cuenta estos verbos a la hora de calcular los pesos como se especifica en la Sección 5.2.
- Modificar o añadir **Palabras clave**: Accedemos a la carpeta [PalabrasClave/](#) y abrimos el archivo *formas_palabras.clave.txt* añadiendo una nueva línea con formato `{representante}:{representante},{forma1},{forma2}....`. De esta forma se tendrán en cuenta estas palabras a la hora de calcular los pesos como se especifica en la Sección 5.3.

Una vez realizados los cambios y si estamos satisfechos simplemente habría que ejecutar el archivo [create_weights.py](#) y automáticamente se actualizarán los pesos en nuestra tabla [weights.csv](#).

Procesamiento de preguntas: Para preguntar a nuestro sistemas debemos ejecutar el script [qa_system.py](#) que esperará una pregunta o consulta que el usuario introduzca por la consola. Una vez introducida, el script irá buscando en nuestros archivos de palabras su representante (el representante de universitarias es universitario por ejemplo), si no lo encuentra pasará la siguiente palabra, si la encuentra consultará la tabla de pesos y actualizará la puntuación asignada a cada dimensión con el peso correspondiente. Cuando haya procesado todas las palabras, el sistema responderá con las tres dimensiones que más puntuación tengan ordenadas de mayor a menor, esto es así porque queremos comprobar cómo de grande es un fallo en caso de que ocurra, no es lo mismo que al preguntar por el nombre del padre nos reponda con el de la madre a que nos responda con el número de hijos. Se da por sentado que el usuario preguntará por una de las personalidades de las que tenemos guardado su repositorio, en caso contrario, se mostrará un mensaje de error. En la Figura 1 se muestra un ejemplo de pregunta-respuesta, como vemos responde dimensiones relacionadas con su nacimiento y el apellido de su madre.

```
Por favor, introduce una pregunta acerca de Elon Musk o Steve Jobs:
Dime en qué lugar del mundo dió a luz a elon musk su madre
Respuestas de mayor a menor probabilidad (sólo se muestran 3):
  Dimensión lugar_de_nacimiento: Sudáfrica
  Dimensión fecha_de_nacimiento: 28-06-1971
  Dimensión segundo_apellido: Reeve
```

Figura 1: Ejemplo de pregunta-respuesta de nuestro sistema

6.2. Validación

Para la validación empleamos el script [qa_system.py](#) explicado anteriormente. Hacemos 30 preguntas, por lo tanto obtenemos 90 soluciones, 3 por cada una de ellas. El objetivo de esta validación es combinar diferentes dimensiones en una misma pregunta con el propósito de confundir al sistema. Para ello, haremos pruebas con preguntas que contengan de 1 a 3 dimensiones. Tanto las soluciones como las respuestas que ofrece el sistema, se muestran en la siguiente tabla:

Pregunta	Dimension 1	Dimension 2	Dimension 3
¿El primer apellido de Steve Jobs es de la madre?	segundo apellido	primer apellido	nombre de la madre
¿La mujer de Elon Musk está muerta?	nombre de la madre	esposa	puestos empresariales
¿Qué estudió Steve Jobs en el año 1972?	edad	año de nacimiento	año de fallecimiento
¿Elon Musk tuvo un hijo a los 30 años?	numero de hijos	edad	año de fallecimiento
¿El padre de Steve Jobs tiene marido?	marido	primer apellido	nombre de la madre
¿Elon Musk estudió en Estados Unidos??	universidad	carrera universitaria	puestos empresariales
¿Steve Jobs tenía novia?	puestos empresariales	carrera universitaria	primer apellido
¿Elon musk tiene descendencia en Estados Unidos?	puestos empresariales	esposa	primer apellido
¿Steve Jobs estaba con su mujer cuando murió?	nombre de la madre	esposa	mes de fallecimiento
¿Elon Musk estudió filología?	universidad	carrera universitaria	puestos empresariales
¿Steve Jobs tiene descendencia?	puestos empresariales	esposa	primer apellido
¿Elon Musk no tiene padre?	primer apellido	nombre de la madre	nombre del padre
¿Cuando murió Steve Jobs?	mes de fallecimiento	día de fallecimiento	fecha de fallecimiento
¿En qué universidad estudió Elon Musk?	universidad	carrera universitaria	mes de fallecimiento
¿Steve Jobs estudió una carrera?	universidad	carrera universitaria	puestos empresariales
¿Los hijos de Elon Musk tienen apellidos diferentes por que los ha tenido con diferentes mujeres?	primer apellido	nombre de la madre	segundo apellido
¿Steve Jobs fundó alguna empresa con su madre?	puestos empresariales	nombre de la madre	segundo apellido
¿El abuelo de los hijos de Elon Musk es su padre?	primer apellido	nombre de la madre	nombre del padre
¿Steve Jobs no estudió en Columbia?	universidad	carrera universitaria	puestos empresariales
¿Elon Musk murió a una edad joven?	mes de fallecimiento	día de fallecimiento	fecha de fallecimiento
¿Steve Jobs es mayor de 50 años?	edad	año de nacimiento	año de fallecimiento
¿En que año se casó Elon Musk?	edad	año de nacimiento	año de fallecimiento
¿La madre de Steve Jobs estudió en su misma universidad?	universidad	nombre de la madre	segundo apellido

¿El padre de Elon Musk es empresario?	primer apellido	nombre de la madre	nombre del padre
¿Dónde vive Steve Jobs?	edad	universidad	lugar de nacimiento
¿Elon Musk ha dejado de estudiar?	universidad	carrera universitaria	puestos empresariales
¿A que edad dejó de estudiar Steve Jobs?	universidad	carrera universitaria	puestos empresariales
¿La mujer de Elon Musk tiene madre?	nombre de la madre	segundo apellido	esposa
¿El hijo de Steve Jobs está muerto?	numero de hijos	mes de fallecimiento	fecha de fallecimiento
¿La mujer de Elon Musk está casada?	esposa	nombre de la madre	marido

Tabla 4: Tabla de dimensiones respondidas a nuestras preguntas ordenadas por probabilidad

7. Conclusiones

El objetivo de este proyecto era el de diseñar e implementar un sistema interactivo de pregunta-respuesta sobre bibliografías.

Para la elaboración del mismo, hemos tenido que documentarnos acerca de la vida de dos personas con una vida de un calibre significativo, diseñado las preguntas tipo, además de unas reglas para poder identificar la dimensión a la que se refieren dichas preguntas. Por último, lo hemos automatizado para que un usuario sea capaz de obtener una respuestas con cierto sentido de él.

A la vista de los resultados obtenidos en la evaluación, cuando aparecen varias dimensiones dentro de una misma pregunta, el sistema tiende a confundirse. Aparecen dimensiones en la respuesta que no tienen nada que ver con las dimensiones de la pregunta.

Hay preguntas que tienen una complejidad estructural, por lo tanto, la semántica que con la que el sistema genera la respuesta no es la correcta.

Hay que tener en cuenta que intervienen diversos factores a la hora de estructurar una pregunta, por lo tanto, los rendimientos pueden ser variables. Para un buen funcionamiento, deberíamos de utilizar, como máximo, dos dimensiones en una misma pregunta.

Como posibles mejoras, en primer lugar podríamos pensar en ampliar las dimensiones para que tuviera más capacidad de respuesta. Otra de las mejoras que se plantea es ampliar tanto los verbos como las palabras clave para cada dimensión, utilizando así n-gramas para dotarle de más comprensión.

En definitiva, hemos cumplido con el objetivo del proyecto. Tenemos un mecanismo de pregunta-respuesta básico pese a que todo sistema sea siempre susceptible a mejora.

Referencias

- [1] Christopher D. Manning, Prabhakar Raghavan e Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. DOI: [10.1017/CB09780511809071](https://doi.org/10.1017/CB09780511809071).

A. Repositorios

En este apéndice se encuentran los textos completos de nuestros repositorios de biografías, los valores de las dimensiones en cada repositorio están remarcados en negrita.

A.1. Repositorio de Steve Jobs

Steve Jobs nació en San Francisco, Estados Unidos, 24 de febrero de 1955 fruto de la relación entre Abdulfattah Jandali, un inmigrante sirio musulmán, luego doctorado en ciencias políticas, y Joanne Carole

Schieble, una estadounidense de ascendencia alemana y suiza, por entonces dos jóvenes estudiantes universitarios que lo entregaron en **adopción a una pareja de clase media, Paul Jobs y Clara Hagopian**, ella de ascendencia armenia. Sus padres biológicos se casaron luego y tuvieron otra hija, la novelista Mona Simpson, a quien Steve no conocería hasta la edad adulta. En esa nueva familia Steve creció junto a su otra hermana, Patricia. **Su padre adoptivo, Paul Jobs**, era maquinista en la empresa estatal de transporte ferroviario y su madre ama de casa.

De la mano de Steve Wozniak comienza a asistir a las reuniones del Homebrew Computer Club, donde Wozniak le contó que estaba intentando construir un pequeño computador casero. Steve Jobs se encarga de las ventas y negociaciones y Steve Wozniak, en secreto, de construir la máquina electrónica. Según afirma Nolan Bushnell, luego de su regreso de la India, a donde fue acompañado por un antiguo compañero de la escuela secundaria y más tarde primer empleado de Apple, Daniel Kottke, decidió renunciar a Atari y fundar Apple Computer. Steve ofreció a Bushnell un porcentaje de Apple, 50 000 dólares, el cual no aceptó.

En 1972 entra en la universidad Reed College de Portland (Oregón). Asiste a ella tan solo 6 meses antes de abandonarla, debido al alto coste de sus estudios. En lugar de regresar a casa, continúa asistiendo a clases como oyente unos 18 meses más, viviendo a base de trabajos con ingresos ínfimos. Curiosamente, sus estudios en caligrafía, enseñada por Robert Palladino, le serían de utilidad cuando diseñara las tipografías del primer Mac.

Inicios de Apple Computer. Fue así como en 1976 nació Apple Computer Company. Tras la consecución del primer computador personal, **bautizado como Apple I**, Jobs se dedicó a su promoción entre otros aficionados a la informática, tiendas y ferias de electrónica digital, llegando a vender unos 200 ejemplares.

Tras abandonar Apple en 1986, **Steve Jobs compra por 5 millones de dólares la empresa The Graphics Group destinando otros 5 millones adicionales como inversión, conocida en lo sucesivo como Pixar**, una subsidiaria de Lucasfilm especializada en la producción de gráficos por computador. Steve Jobs empezó a firmar varios acuerdos para producir películas animadas para la compañía Walt Disney. En 1995 se estrenó en los cines Toy Story, el primer largometraje generado completamente por computadora, conseguido con su propio software de renderización, RenderMan. Toy Story fue el mayor éxito de taquilla de 1995 y la primera película del binomio Walt Disney-Pixar en ganar un premio óscar.

Estuvo casado desde 1991 con Laurene Powell, a quien conoció en la Universidad de Stanford. Vivieron en Palo Alto, California, con sus tres hijos. Steve tuvo además otra hija llamada Lisa, fruto de una relación de juventud con Chris Ann Brennan. Su paternidad no la reconoció sino hasta 1991. Sufrió varios problemas graves de salud. En 2004 se le diagnosticó un cáncer de páncreas, enfermedad que superó tras un tratamiento en una clínica oncológica californiana. A principios del 2009 anunció que padecía un desequilibrio hormonal y que debía apartarse necesariamente de la compañía, y delegó la mayor parte de sus responsabilidades en Timothy Cook, por entonces jefe de operaciones. En abril de 2009 se sometió a un trasplante de hígado, y en septiembre de ese mismo año volvió al trabajo. El 17 de enero de 2011, Jobs dejó nuevamente Apple por problemas de salud, a poco más de un mes de la **presentación del iPad 2**. Mientras tanto, la compañía quedó a cargo de Tim Cook. Jobs presentó públicamente el iPad 2 el miércoles 2 de marzo. No obstante, declaró que desde su residencia seguiría ocupándose de las decisiones más relevantes de la compañía, como en efecto ocurrió.

Tras dejar Apple, a los 30 años de edad, decidió continuar su carrera empresarial en la industria de la computación y **fundó la empresa NeXT Computer Inc.** En el plan de negocios se estableció que, al igual que se hacía en Apple, la compañía vendiese al cliente no solo el hardware, sino también el sistema operativo y parte del software de usuario. Como consecuencia, Jobs en 1993 centró la estrategia de su compañía en la producción de software, cambiando el nombre de la empresa por el de Next Software Inc. Apple Computer anunció el 20 de diciembre de 1996 la adquisición de NeXT Software por 400 millones de dólares con el fin de actualizar el sistema operativo de las computadoras Macintosh, después del fracaso de la compañía con Copland, un proyecto que nunca llegó a terminarse.

Así, Steve Jobs volvió a formar parte de la compañía Apple. El regreso a Apple. La vuelta de Steve Jobs a la empresa Apple se produjo cuando la empresa se encontraba en declive, así que se decidió a recuperar el control de ésta, se ganó la confianza de la dirección de la compañía en detrimento del entonces director ejecutivo, Gil Amelio, logrando que se lo nombrara director interino el 16 de septiembre de 1997. Steve Jobs tenía unos tumores neuroendocrinos de, relativamente, poca gravedad, comparados con el muy agresivo adenocarcinoma que tienen el 95 por ciento de los pacientes de cáncer de páncreas. Jobs no recibió radioterapia o quimioterapia. Durante la ausencia de Jobs, Tim Cook, que era director de ventas y operaciones de Apple, dirigió la compañía.

A principios de agosto de 2006 Jobs realizó una presentación en la Conferencia Anual de Desarrolladores de Apple. Sin embargo, de acuerdo con un artículo de Ars Technica los asistentes a la Conferencia que vieron a Jobs dijeron que tenía buen aspecto. Después de la presentación un portavoz de Apple dijo que Steve tenía

una salud de hierro. Los responsables de Apple afirmaron que Jobs padecía una afección corriente y que estaba tomando antibióticos, mientras que otros achacaban su estado demacrado al procedimiento Whipple aplicado en su cirugía.

Durante una conferencia en julio en la que se discutía sobre los beneficios de Apple, los participantes respondieron a las preguntas sobre la salud de Jobs que era un "sunto privado". Otros opinaron que los accionistas tenían el derecho a saber más dado el estilo personal de Jobs al dirigir la compañía. El 28 de agosto de 2008 el servicio de noticias de empresa Bloomberg publicó por error un obituario de Jobs de 2500 palabras que contenía espacios para su edad y la causa de la muerte. Aunque el error se rectificó inmediatamente, muchas agencias de noticias escribieron sobre él, intensificando los rumores sobre la salud de Jobs.

En un acto posterior Jobs terminó su presentación con una diapositiva en la que se leía "110/70", en referencia a su presión sanguínea, y dijo que no respondería más preguntas sobre su salud. El 16 de diciembre de 2008 Apple anunció que el vicepresidente de marketing Philip W. Esto reactivó las preguntas sobre la salud de Jobs. En una declaración del 5 de enero de 2009 en Apple, Jobs dijo que había estado sufriendo un desequilibrio hormonal durante varios meses.

El 14 de enero de 2009 en un memorando interno de Apple Jobs escribió que la semana anterior supo que los asuntos relacionados con su salud eran más complejos de lo que pensaba, y anunció una excedencia de seis meses hasta el final de junio de 2009 para poder centrarse en su salud.

Steve Jobs falleció en su casa de California a las 2 de la tarde del 5 de octubre de 2011, a los 56 años, a consecuencia de un paro respiratorio derivado de las metástasis del cáncer neuroendocrino de páncreas que le fue descubierto en 2004, por el que en 2009 había recibido un trasplante de hígado. El día anterior había perdido la conciencia y murió estando su esposa, hijos y hermana a su lado.

Su vida quedó plasmada en creaciones como Apple I, Apple Lisa, iMac, iPod, iPhone, iPad, Mac OS X v10.0, NEXTSTEP, iTunes y App Store.

A.2. Repositorio de Elon Musk

Elon Reeve Musk (Pretoria, Sudáfrica, 28 de junio de 1971), conocido como Elon Musk, es un empresario, inversor y magnate sudafricano que también posee las nacionalidades canadiense y estadounidense.

Musk nació de madre canadiense (Maye) y padre sudafricano blanco (Errol), y se crió en Pretoria (Sudáfrica). Estudió brevemente en la Universidad de Pretoria antes de trasladarse a Canadá a los 17 años. Se matriculó en la Universidad de Queen y se trasladó a la Universidad de Pensilvania dos años después, donde se graduó en Economía y Física.

En diciembre de 2000 se tomó unas vacaciones en Brasil y en una reserva de caza en Sudáfrica. En enero de 2001 volvió a California y a los pocos días empezó a sentirse mal. En un principio lo diagnosticaron con meningitis viral, enfermedad que trataron y le dieron el alta. Unos días después empeoró y lo trasladaron al Sequoia Hospital en Redwood City. Le diagnosticaron y trataron por malaria. Estando en la unidad de cuidados intensivos se encontró que la cepa correspondía a plasmodium falciparum, que es el más letal. Cambiaron el tratamiento a otro más agresivo que podía tener efectos secundarios peligrosos como arritmias y fallo orgánico. Estuvo a punto de morir, perdió 20 kilos y tardó 6 meses en recuperarse totalmente.

Musk vive en Bel-Air, California. En 2000 Elon se casó con su primera **esposa, la autora canadiense Justine Wilson**, a la que conoció cuando ambos eran estudiantes en la Universidad de Queen. En 2002 tuvieron **un hijo** llamado Nevada Alexander Musk, quien a las diez semanas falleció por el síndrome de muerte súbita. Musk y Wilson tienen **cinco hijos**: un par de gemelos y unos trillizos. Se separaron en el año 2008.

Ese mismo año conoció a la actriz británica **Talulah Riley. Se casaron** en 2010 y se divorciaron en 2012. Se volvieron a casar en 2013 y terminaron por separarse legalmente. En noviembre de 2017 terminó su relación con la actriz Amber Heard debido a sus conflictos de agendas.

El 7 de mayo de 2018, Musk **comenzó a salir con la cantante canadiense Grimes** y en enero de 2020 anunciaron que estaban esperando su primer **hijo, el sexto de Musk**. El 4 de mayo de 2020 nació su hijo, X AE A-XII Musk. La pareja anunció su separación en septiembre de 2021. La pareja regresó en diciembre del 2021. En marzo de 2022 se hizo público que había nacido su **segunda hija biológica con Grimes**, a quien llamaron Exa Dark Sideræl, por gestación subrogada en diciembre de 2021; su octavo hijo en total.

En julio de 2022, Insider hizo público que había tenido **gemelos con Shivon Zilis**, directora de operaciones y proyectos especiales en Neuralink, en noviembre de 2021.

Elon Musk es un empresario, inversor y magnate sudafricano que también posee las nacionalidades canadiense y estadounidense. Es el fundador, **consejero delegado e ingeniero jefe de SpaceX; inversor ángel, CEO y arquitecto de productos de Tesla, Inc.; fundador de The Boring Company; cofundador de Neuralink y OpenAI, aunque ya no tiene mas participación en OpenAI por desacuerdos en el**

rumbo de la empresa, además de ser el actual dueño de Twitter. Con un patrimonio neto estimado en unos 140 mil millones de dólares en enero de 2023, Musk es la segunda persona más rica del mundo según el índice de multimillonarios de Bloomberg y la lista de multimillonarios en tiempo real de Forbes.

B. Pesos

En este apéndice se pueden encontrar las tablas de los pesos de las palabras de nuestro diccionario para cada Dimensión.

B.1. Pesos de los pronombres interrogativos

En esta sección se muestra la tabla 5 donde se muestran los pesos obtenidos para los pronombres interrogativos siguiendo el método explicado en la Sección 5.1.

	cómo	cuál	cuán	cuándo	cuánto	dónde	qué	quién
nombre	0.167	0.250	0.000	0.000	0.000	0.000	0.500	0.083
primer_apellido	0.364	0.455	0.000	0.000	0.000	0.000	0.182	0.000
segundo_apellido	0.200	0.600	0.000	0.000	0.000	0.000	0.200	0.000
lugar_de_nacimiento	0.000	0.500	0.000	0.000	0.000	0.167	0.333	0.000
dia_de_nacimiento	0.000	0.600	0.000	0.000	0.000	0.000	0.400	0.000
mes_de_nacimiento	0.000	0.500	0.000	0.000	0.000	0.000	0.500	0.000
anno_de_nacimiento	0.000	0.400	0.000	0.000	0.000	0.000	0.600	0.000
fecha_de_nacimiento	0.000	0.273	0.000	0.273	0.000	0.000	0.455	0.000
edad	0.083	0.083	0.083	0.000	0.417	0.000	0.333	0.000
dia_de_fallecimiento	0.000	0.714	0.000	0.000	0.000	0.000	0.286	0.000
mes_de_fallecimiento	0.000	0.400	0.000	0.000	0.000	0.000	0.600	0.000
anno_de_fallecimiento	0.000	0.500	0.000	0.000	0.000	0.000	0.500	0.000
fecha_de_fallecimiento	0.000	0.312	0.000	0.312	0.000	0.000	0.312	0.000
marido	0.300	0.200	0.000	0.000	0.000	0.000	0.300	0.200
esposa	0.300	0.200	0.000	0.000	0.000	0.000	0.300	0.200
pareja_actual	0.100	0.200	0.000	0.000	0.000	0.000	0.300	0.400
numero_de_hijos	0.000	0.143	0.000	0.000	0.429	0.000	0.429	0.000
nombre_del_padre	0.364	0.273	0.000	0.000	0.000	0.000	0.091	0.273
nombre_de_la_madre	0.357	0.286	0.000	0.000	0.000	0.000	0.071	0.286
universidad	0.083	0.167	0.000	0.000	0.000	0.417	0.333	0.000
carrera_universitaria	0.083	0.417	0.000	0.000	0.000	0.000	0.500	0.000
puestos_empresales	0.000	0.385	0.000	0.000	0.000	0.077	0.538	0.000

Tabla 5: Pesos de los pronombres interrogativos para la dimensión Carrera universitaria

B.2. Listas de verbos para cada dimensión

Como hemos especificado en la sección 5.2, obtenemos los verbos de los parafraseos de cada dimensión, de tal forma que si aparecen al menos una vez, tendrán un peso de 2 para la misma.

Los verbos para la dimensión **nombre** son:

- **aparecer**, comprobando “aparecer”, “aparece”, “apareció”, “aparecido”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **bautizar**, comprobando “bautizar”, “bautizó”, “bautizaron”, “bautizado”.
- **dar**, comprobando “dar”, “da”, “dió”, “dieron”, “dado”.
- **poner**, comprobando “poner”, “pone”, “puso”, “pusieron”, “puesto”.

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.

Los verbos para la dimensión **primer_apellido** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **apellidar**, comprobando “apellidar”, “apellida”, “apedilló”, “apedillaba”, “apedillado”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **conocer**, comprobando “conocer”, “conoce”, “conoció”, “conocido”.

Los verbos para la dimensión **segundo_apellido** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **conocer**, comprobando “conocer”, “conoce”, “conoció”, “conocido”.
- **ir**, comprobando “ir”, “va”, “fue”, “ido”.

Los verbos para la dimensión **lugar_de_nacimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.
- **dar**, comprobando “dar”, “da”, “dió”, “dieron”, “dado”.
- **encontrar**, comprobando “encontrar”, “encuentra”, “encontró”, “encontrado”.

Los verbos para la dimensión **dia_de_nacimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.

Los verbos para la dimensión **mes_de_nacimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.

Los verbos para la dimensión **anno_de_nacimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.

Los verbos para la dimensión **fecha_de_nacimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.
- **dar**, comprobando “dar”, “da”, “dió”, “dieron”, “dado”.
- **ir**, comprobando “ir”, “va”, “fue”, “ido”.

Los verbos para la dimensión **edad** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.
- **vivir**, comprobando “vivir”, “vive”, “vivió”, “vivido”.
- **pasar**, comprobando “pasar”, “pasa”, “pasó”, “pasaron”, “pasado”.
- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **mantener**, comprobando “mantener”, “mantiene”, “mantuvo”, “mantenido”.

Los verbos para la dimensión **dia_de_fallecimiento** son:

- **fallecer**, comprobando “fallecer”, “fallece”, “falleció”, “fallecido”.
- **morir**, comprobando “morir”, “muere”, “murió”, “muerto”.

Los verbos para la dimensión **mes_de_fallecimiento** son:

- **fallecer**, comprobando “fallecer”, “fallece”, “falleció”, “fallecido”.
- **morir**, comprobando “morir”, “muere”, “murió”, “muerto”.

Los verbos para la dimensión **anno_de_fallecimiento** son:

- **fallecer**, comprobando “fallecer”, “fallece”, “falleció”, “fallecido”.
- **morir**, comprobando “morir”, “muere”, “murió”, “muerto”.

Los verbos para la dimensión **fecha_de_fallecimiento** son:

- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.
- **fallecer**, comprobando “fallecer”, “fallece”, “falleció”, “fallecido”.
- **morir**, comprobando “morir”, “muere”, “murió”, “muerto”.
- **ir**, comprobando “ir”, “va”, “fue”, “ido”.

Los verbos para la dimensión **marido** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **casar**, comprobando “casar”, “casa”, “casó”, “casado”.
- **contraer**, comprobando “contraer”, “contrae”, “contrajo”, “contraído”.

Los verbos para la dimensión **esposa** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **casar**, comprobando “casar”, “casa”, “casó”, “casado”.
- **contraer**, comprobando “contraer”, “contrae”, “contrajo”, “contraído”.

Los verbos para la dimensión **pareja_actual** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **mantener**, comprobando “mantener”, “mantiene”, “mantuvo”, “mantenido”.

Los verbos para la dimensión **numero_de_hijos** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.

Los verbos para la dimensión **nombre_del_padre** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **criar**, comprobando “criar”, “cria”, “crió”, “criado”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.

Los verbos para la dimensión **nombre_de_la_madre** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **nacer**, comprobando “nacer”, “nace”, “nació”, “nacido”.
- **criar**, comprobando “criar”, “cria”, “crió”, “criado”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.

Los verbos para la dimensión **universidad** son:

- **realizar**, comprobando “realizar”, “realiza”, “realizó”, “realizado”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **cursar**, comprobando “cursar”, “cursa”, “cursó”, “cursado”.
- **asistir**, comprobando “asistir”, “asiste”, “asistió”, “asistido”.
- **estudiar**, comprobando “estudiar”, “estudia”, “estudió”, “estudiado”.

Los verbos para la dimensión **carrera_universitaria** son:

- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **estudiar**, comprobando “estudiar”, “estudia”, “estudió”, “estudiado”.
- **llamar**, comprobando “llamar”, “llama”, “llamó”, “llamaron”, “llamado”.
- **realizar**, comprobando “realizar”, “realiza”, “realizó”, “realizado”.

Los verbos para la dimensión **puestos_empresariales** son:

- **trabajar**, comprobando “trabajar”, “trabaja”, “trabajó”, “trabajado”.
- **realizar**, comprobando “realizar”, “realiza”, “realizó”, “realizado”.
- **poner**, comprobando “poner”, “pone”, “puso”, “pusieron”, “puesto”.
- **tener**, comprobando “tener”, “tiene”, “tuvo”, “tenido”.
- **ocupar**, comprobando “ocupar”, “ocupa”, “ocupó”, “ocupado”.

B.3. Listas de palabras clave para cada dimensión

Las palabras clave para la dimensión **nombre** son:

- **apodo**, incluyendo “apodo”, “apodos”.
- **padres**, incluyendo “padres”.
- **identidad**, incluyendo “identidad”, “identidades”.
- **documento**, incluyendo “documento”, “documentos”.
- **nombre**, incluyendo “nombre”, “nombres”.
- **puesto**, incluyendo “puesto”, “puestos”.

Las palabras clave para la dimensión **primer_apellido** son:

- **primer**, incluyendo “primer”, “primero”, “primeros”, “primera”, “primeras”.
- **padre**, incluyendo “padre”.
- **familia**, incluyendo “familia”, “familias”.
- **apellido**, incluyendo “apellido”, “apellidos”.
- **segundo**, incluyendo “segundo”, “segunda”, “segundos”, “segundas”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **segundo_apellido** son:

- **madre**, incluyendo “madre”, “madres”.
- **primer**, incluyendo “primer”, “primero”, “primeros”, “primera”, “primeras”.

- **tercer**, incluyendo “tercer”, “tercero”, “tercera”, “terceros”.
- **familia**, incluyendo “familia”, “familias”.
- **apellido**, incluyendo “apellido”, “apellidos”.
- **segundo**, incluyendo “segundo”, “segunda”, “segundos”, “segundas”.
- **último**, incluyendo “último”, “última”, “últimos”, “últimas”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **lugar_de_nacimiento** son:

- **país**, incluyendo “país”, “países”.
- **lugar**, incluyendo “lugar”, “lugares”.
- **luz**, incluyendo “luz”, “luces”.
- **madre**, incluyendo “madre”, “madres”.
- **nacionalidad**, incluyendo “nacionalidad”, “nacionalidades”.
- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.

Las palabras clave para la dimensión **día_de_nacimiento** son:

- **número**, incluyendo “número”, “números”.
- **día**, incluyendo “día”, “días”.
- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.

Las palabras clave para la dimensión **mes_de_nacimiento** son:

- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.
- **mes**, incluyendo “mes”, “meses”.

Las palabras clave para la dimensión **anno_de_nacimiento** son:

- **año**, incluyendo “año”, “años”.
- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.

Las palabras clave para la dimensión **fecha_de_nacimiento** son:

- **fecha**, incluyendo “fecha”, “fechas”.
- **día**, incluyendo “día”, “días”.
- **luz**, incluyendo “luz”, “luces”.
- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.

Las palabras clave para la dimensión **edad** son:

- **mayor**, incluyendo “mayor”, “mayores”.
- **nacimiento**, incluyendo “nacimiento”, “nacimientos”.
- **número**, incluyendo “número”, “números”.
- **año**, incluyendo “año”, “años”.
- **vida**, incluyendo “vida”, “vidas”.

Las palabras clave para la dimensión **día_de_fallecimiento** son:

- **número**, incluyendo “número”, “números”.
- **día**, incluyendo “día”, “días”.
- **fallecimiento**, incluyendo “fallecimiento”, “fallecimiento”.

Las palabras clave para la dimensión **mes_de_fallecimiento** son:

- **fallecimiento**, incluyendo “fallecimiento”, “fallecimiento”.
- **mes**, incluyendo “mes”, “meses”.

Las palabras clave para la dimensión **anno_de_fallecimiento** son:

- **año**, incluyendo “año”, “años”.
- **fallecimiento**, incluyendo “fallecimiento”, “fallecimiento”.

Las palabras clave para la dimensión **fecha_de_fallecimiento** son:

- **muerte**, incluyendo “muerte”, “muertes”.
- **fecha**, incluyendo “fecha”, “fechas”.
- **día**, incluyendo “día”, “días”.
- **fallecimiento**, incluyendo “fallecimiento”, “fallecimiento”.

Las palabras clave para la dimensión **marido** son:

- **marido**, incluyendo “marido”, “maridos”.
- **casado**, incluyendo “casado”, “casada”, “casados”, “casadas”.
- **hombre**, incluyendo “hombre”, “hombres”.
- **matrimonio**, incluyendo “matrimonio”, “matrimonios”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **esposa** son:

- **mujer**, incluyendo “mujer”, “mujeres”.
- **casado**, incluyendo “casado”, “casada”, “casados”, “casadas”.
- **matrimonio**, incluyendo “matrimonio”, “matrimonios”.
- **nombre**, incluyendo “nombre”, “nombres”.
- **esposa**, incluyendo “esposa”, “esposas”.

Las palabras clave para la dimensión **pareja_actual** son:

- **actual**, incluyendo “actual”, “actuales”.
- **relación**, incluyendo “relación”, “relaciones”.
- **pareja**, incluyendo “pareja”, “parejas”.
- **persona**, incluyendo “persona”, “personas”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **numero_de_hijos** son:

- **número**, incluyendo “número”, “números”.
- **retoño**, incluyendo “retoño”, “retoños”.

- **hijo**, incluyendo “hijo”, “hija”, “hijos”, “hijas”.
- **persona**, incluyendo “persona”, “personas”.

Las palabras clave para la dimensión **nombre_del_padre** son:

- **nombre**, incluyendo “nombre”, “nombres”.
- **padre**, incluyendo “padre”.
- **hombre**, incluyendo “hombre”, “hombres”.
- **progenitor**, incluyendo “progenitor”, “progenitores”.

Las palabras clave para la dimensión **nombre_de_la_madre** son:

- **mujer**, incluyendo “mujer”, “mujeres”.
- **madre**, incluyendo “madre”, “madres”.
- **progenitora**, incluyendo “progenitora”, “progenitoras”.
- **padre**, incluyendo “padre”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **universidad** son:

- **título**, incluyendo “título”, “títulos”.
- **universitaria**, incluyendo “universitaria”, “universitario”, “universitarias”, “universitarios”.
- **universidad**, incluyendo “universidad”, “universidades”.
- **estudio**, incluyendo “estudio”, “estudios”.
- **carrera**, incluyendo “carrera”, “carreras”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **carrera_universitaria** son:

- **título**, incluyendo “título”, “títulos”.
- **universitaria**, incluyendo “universitaria”, “universitario”, “universitarias”, “universitarios”.
- **estudio**, incluyendo “estudio”, “estudios”.
- **carrera**, incluyendo “carrera”, “carreras”.
- **nombre**, incluyendo “nombre”, “nombres”.

Las palabras clave para la dimensión **puestos_empresariales** son:

- **laboral**, incluyendo “laboral”, “laborales”.
- **empresa**, incluyendo “empresa”, “empresas”.
- **empresarial**, incluyendo “empresarial”, “empresariales”.
- **profesional**, incluyendo “profesional”, “profesionales”.
- **trabajo**, incluyendo “trabajo”, “trabajos”.
- **ocupación**, incluyendo “ocupación”, “ocupaciones”.
- **experiencia**, incluyendo “experiencia”, “experiencias”.
- **carrera**, incluyendo “carrera”, “carreras”.
- **puesto**, incluyendo “puesto”, “puestos”.