

|  |
| --- |
| **Universidad Internacional de La Rioja**  **Escuela Superior de Ingeniería y Tecnología**  **Máster Universitario en Análisis y Visualización de Datos Masivos** |
|  |
| Análisis de la Cronología y relaciones entre personajes bíblicos y eventos a través del tiempo |

**Trabajo Fin de Máster**

**Tipo de trabajo:** Piloto Experimental

**Presentado por:** González Cano, David

**Director/a:** Fernández García, Diego

**Resumen**

A través del uso de diferentes técnicas de Big Data y Visualización del Procesamiento de Lenguaje Natural (NPL), usando técnicas dirigidas al análisis de modernos análisis de texto, usando Python y herramientas de software libre como Google BERT.

La propuesta de trabajo de Fin de Master (TFM) tiene el propósito central de utilizar estas técnicas para analizar la cronología de los personajes bíblicos, los eventos y las relaciones que estos personajes tienen en los libros de la Biblia.

Para ello se utilizan diferentes datasets públicos que contienen variables como “fecha, “pasaje de la biblia”, mediante técnicas de NPL se analizan las variables y finalmente se crea una plataforma de visualización que sea accesible a cualquier usuario que lo requiera.

Para la recopilación, limpieza, depuración, análisis, clasificación y visualización de datos se utilizará Python con sus respectivas librerías.

Para analizar y trabajar con textos relacionados con la cronología bíblica se va a utilizar **NPL** (Procesamiento de Lenguaje Natural)

Para la visualización se está pensando en herramientas colaborativas como Google sheet que permita obtener tanto visualizaciones estáticas como dinámicas.

Lo que se espera tener de esta investigación es poder dejar un prototipo que utilizando el texto de la biblia permita usar las herramientas para hacer preguntas sobre la cronología tanto de eventos como de personajes y presente visualizaciones claras de lo requerido.

**Abstract**

The Master's Thesis proposal (TFM) has the central purpose of using different Big Data and Data Visualization techniques to analyze the chronology of biblical characters, the events and the relationships that these characters have in the books of the Bible.

Using the different datasets that are public, with different variables and different formats.

Python with its respective libraries will be used to collect, clean, debug, analyze, classify and visualize data.

To analyze and work with texts related to biblical chronology, NPL (Natural Language Processing) will be used.

For visualization, collaborative tools such as Google Sheet are being considered that allow obtaining both static and dynamic visualizations.

What is expected from this research is to be able to leave a prototype that, using the text of the Bible, allows the use of tools to ask questions about the chronology of both events and characters and presents clear visualizations of what is required.

Índice de Contenido

[1. Introducción 7](#_Toc152654510)

[1.1 Justificación 7](#_Toc152654511)

[1.2 Planteamiento del trabajo 8](#_Toc152654512)

[1.3 Estructura de la memoria 8](#_Toc152654513)

[2. Contexto y estado del arte 9](#_Toc152654514)

[2.1 Introducción al tema 9](#_Toc152654515)

[2.2. Desarrollo (base teórica): antecedentes, estudios actuales, autores de referencia 9](#_Toc152654516)

[2.2. 1 NLP (Natural Lenguaje Processing) 9](#_Toc152654517)

[2.2.2 Google BERT (Bidirectional Encoder Representations from Transformers) to 10](#_Toc152654518)

[2.2.3. Herramientas Colaborativas 11](#_Toc152654519)

[2.2.3.1 GitHub / GitLab / Bitbucket / Gitkraten / SourceGorge 11](#_Toc152654520)

[*2.2.3.2* Google *Colab* 11](#_Toc152654521)

[2.2.4. Bibliotecas Python 11](#_Toc152654522)

[2.2.4. La Biblia 12](#_Toc152654523)

[2.3 Conclusiones (nexo de unión de lo investigado con el trabajo a realizar) 16](#_Toc152654524)

[3. Objetivos concretos y metodología de trabajo 17](#_Toc152654525)

[3.1. Objetivo general 17](#_Toc152654526)

[3.2. Objetivos específicos 17](#_Toc152654527)

[6. Bibliografía 19](#_Toc152654528)

[Bibliografía 19](#_Toc152654529)

**Índice de figuras**

[Ilustración 1 Cronologia Biblica www.biblechronological.net Rick Aschmann 15](#_Toc148477273)

[Tabla 1 Libros de la Biblia Hebrea o Antiguo testamento 12](#_Toc148988966)

# 1. Introducción

## 1.1 Justificación

Meditando sobre los proyectos que dan gasolina en el día a día y que se vuelven retos ultraístas, el leer la Biblia es uno de ellos, desde su inicio hasta el final contando con más de cuarenta escritores y aproximadamente 773.746 palabras en un libro que nos lleva desde el inicio del mundo, va recorriendo con eventos, personajes los diferentes tiempos de la humanidad y ahí, en el mismo texto lleno de aventuras nos sumergimos en cual será el final de la humanidad con señales que se conectan en los 66 libros divididos en 1189 capítulos y 31.102 versículos y 3.5 35.666.480 letras, Biblia Reina Valero 1960. (Concepto y definicion Net, 2023)

La secuencia de los tiempos, los personajes, los eventos en los escritos no son tan fáciles de visualizar debido a que las fechas no están implícitas en el texto y se necesitan estudios teológicos y contextos históricos para poder entender la cronología de los eventos con sus respectivos personajes.

Para llevar a cabo esta investigación se van a utilizar diferentes técnicas, métodos y herramientas para encontrar una forma de analizar y presentar cronológicamente eventos, personajes utilizando textos que están ya colocados en la web y también si es posible desarrollar un prototipo de base de datos que nos permita ir actualizando las fechas en una línea de tiempo actualizable por los investigadores de este tema.

La necesidad de esta investigación surge de la diversidad de autores en sus escritos, las figuras literarias y estilísticas que se utilizan como metáforas, símiles, personificaciones, parábolas, hipérboles, ironías, y muchas más; la identificación de fechas en cada evento, la no existencia hasta el momento de una base de datos estructurada que coloque cada evento o personaje en una línea de tiempo.

Por todo ello se propone una herramienta a la población interesada en poder dar un contesto en el tiempo de los escritos ya sea por estudio u otro interés dando una visualización en una forma clara y utilizando herramientas de datos masivos.

En la población mundial aproximadamente de 8300 millones y con un cálculo de 2500 millones de cristianos en el mundo (EL orden Mundial EOM, 2019), tienen como base de su fe la Biblia, la cual se divide en Antiguo y nuevo testamento. Así a esta población se puede sumar el pueblo judío aproximadamente 13 millones (Embajada de Israel en España, s.f.), los cuales su base de lectura en su creencia es el Antiguo Testamento. ya que tienen una división diferente y se conoce como la biblia hebrea o Tanaj la cual está dividida en tres partes: la Torá (ley) , los Nevi’im (profetas) y los Ketuvim (escritos), además los judíos no consideran la parte de los libros del nuevo testamento como parte de las escrituras.

## 1.2 Planteamiento del trabajo

Utilizando técnicas de procesamiento de lenguaje natural (NLP) y modelos de machine Learning que permitan extraer información y procesarla cronológicamente. Creando una base de datos estructurada que manaje y permita trabajar con fechas de los personajes bíblicos y utilizar herramientas de visualización que permitan de una forma fácil presentar los eventos y personajes en forma dinámica.

## 1.3 Estructura de la memoria

La primera parte es describir las técnicas de manejo de Texto

Lo segunda parte está dirigida al análisis de la biblia como un dataset y como se logra organizar para obtener información cronológica

El tercer es el prototipo planteado de la unión de las técnicas y la estructura bíblica con herramientas de Visualización.

# 2. Contexto y estado del arte

## 2.1 Introducción al tema

Es interesante pensar que cuando hablamos de tiempo, aunque no lo sentimos material, tiene estados de pasado, presente y futuro; y cuando lo graficamos con personajes y eventos se nos convierte en una línea de tiempo, una herramienta visual para presentar cualquier información que mezcla datos y eventos formando una función de sucesión temporal.

La cronología bíblica se refiere al estudio y ordenamiento de los eventos históricos descritos en la biblia en secuencia temporal y busca determinar fechas aproximadas de eventos, así como la duración de periodos y la secuencia en la que ocurrieron.

## 2.2. Desarrollo (base teórica): antecedentes, estudios actuales, autores de referencia

Las técnicas a describir son las mas conocidas en el momento para el procesamiento de texto en el mundo la Big data.

### 2.2. 1 NLP (Natural Lenguaje Processing)

“Un idioma no son solo palabras. Es una cultura, una tradición, una unificación de una comunidad, una historia completa que cruza lo que es una comunidad. Todo está incorporado en un lenguaje".

-Noam Chomsky

Es un campo entre la intersección de ciencias de computación e inteligencia artificial y logística. Es un tema de construir sistemas que procesan y entienden el lenguaje humano.

(Sowmya Vajjala, 2020)

El lenguaje natural es el primer medio de comunicación entre humanos desde los comienzos de la humanidad. Los computadores se comunican en una codificación de 0s y 1s. La forma en que las maquinas puede entender el lenguaje es lo que se denomina Procesamiento de lenguaje natural.

Esta es un área de la ciencia de datos que trabaja con métodos de análisis, modelos y lenguaje humano.

La base practica del NPL se puede dividir en aprender primero

Word Embedding

Tokeniztion

Modelos secuenciales red Neuronal recurrentes RRNNs y LSTMs … bag words

Transformers (codificador-decodificador)

Je suis etudiant - input – embeddings – positional encoding.

Modelado de lenguaje

Aplicación con FastAPI:NPL en Producción

### 2.2.2 Google BERT (Bidirectional Encoder Representations from Transformers) to

Es un modelo de procesamiento de lenguaje natural desarrollado por Google. Utiliza la arquitectura de Transformers para entender el contexto de las palabras en una oración, teniendo en cuenta tanto palabras anteriores como siguientes. Esto mejora la capacidad del modelo para comprender el significado de una palabra en función de su contexto, lo que resulta en una mejor comprensión del lenguaje natural y una mejora de la calidad de las respuestas generadas por los modelos de procesamiento de lenguaje natural.

(Ravichandiran, January 2021) compartir sus proyectos de software de manera centralizada.

### 2.2.3. Herramientas Colaborativas

El

#### 2.2.3.1 GitHub / GitLab / Bitbucket / Gitkraten / SourceGorge

Usar una plataforma que permita la colaboración en este proyecto puede ser la base para que analistas de datos con interés en el procesamiento natural del lenguaje, sea experimentado que puedan aportar en ideas o en el propio desarrollo pueden ser un aspecto importante para tener éxito en el proyecto que se está desarrollando.

Git es una tecnología que empezó en el 2005 …

Comandos Linux

GitHub es una plataforma de desarrollo colaborativo que utiliza el sistema de control de versiones Git. Permite a los desarrolladores trabajar juntos en proyectos de software, facilitando el seguimiento de cambios en el código fuente, la colaboración entre equipos y la gestión eficiente de proyectos. Los usuarios pueden alojar sus repositorios de código en GitHub, lo que significa que pueden almacenar.

Algunos aspectos útiles para desarrollar el proyecto:

**Control de versiones:** Con lo que facilita un contexto de trabajo colaborativo. Lleva el control de versiones muy importante en el mundo donde una o varias personas están desarrollando un proyecto,

**Repositorios**: Es el espacio donde se almacena todo el código fuente, archivos de configuración, documentos otros recursos relacionados con un proyecto. Puede ser publico (accesible para todos) o privado (accesible solo para personas autorizadas)

Colaboración de desarrolladores.

Seguimiento a problemas(bugs) Los usuarios pueden informar bugs, solicitar nuevas características o discutir ideas a través del seguimiento de problemas GitHub.

Despliegue: Utiliza un despliegue en sitios de web y aplicaciones a través de GitHub Pages, una función que permite alojar sitios web directamente desde los repositorios de GitHub

GitHub crea un ambiente que permite almacenar código en un servidor remoto, dando la habilidad de compartir el código con otras personas y hacer más fácil que una persona agregue, modifique o borre código del mismo archivo y proyecto, mientras que guarda una fuente de verdad para este archivo. (Sarah Guthals, 2019)

#### *2.2.3.2* Google *Colab*

### 2.2.4. Bibliotecas Python

Hay varias bibliotecas en Python que son de Python son elementos centrales del lenguaje de programación y como se puede escribir, reusar y compartir código en Python.

Para el NLP procesamiento de lenguaje natural hay varias bibliotecas poderosas y populares como son:

NLTK: Es una biblioteca extensa para trabajar con datos de texto. Proporciona herramientas de tokenizacion, stemming, lematización, análisis sintético y mas. (NLTK, Consultada en 2023)

spaCy: Es una biblioteca abierta, diseñada para ser rápida y eficiente. Proporciona modelo preentrenados. Tiene muy buen rendimiento en grandes cantidades de texto.

TextBlob: Es una biblioteca fácil de usar para el procesamiento de texto basado en NLTK. Ofrece una API para tareas sencillas, como análisis de sentimiento, extracción de frases clave y clasificación de texto. (Textblob: Simplified Text Processing, Consultado en 2023)

Gensim: Esta biblioteca se centra en el modelado de temas y similitud de documentos. Es útil para la construcción y entrenamiento de modelo de vectores de palabras (Word embedddings) y modelos de tópicos. (GENSIM topic modeling for humans, Consultada en Nov 22 2023)

Transformer: Proporciona implementaciones de modelos de lenguaje preentrenados de vanguardia, como BERT, GPT-2 . (Tranformers Huggin Face, Consultado en nov 22 2023)

### 2.2.5. La Biblia

La base teórica para desarrollar este punto es extensa y a veces muy controversial, pero se planteará el ser lo más objetivos posibles, estudiando libros de historiadores, teólogos, pastores de iglesias, y revisando documentos que nos den la base en cuanto a tener un texto que sea aceptado en la mayoría.

La palabra “Biblia” tiene raíz del latín medieval y deriva del griego byblos, que significa “libros”, pero si vamos hacia su origen etimológico del termino llegamos a que byblos era una antigua ciudad Fenicia situada sobre las costas del actual Líbano. Los fenicios inventaron el alfabeto que todavía utilizamos y les enseñaron a escribir a los griegos. Desde Biblos los fenicios exportaban los papiros en los que fueron escritos los primeros libros (El papiro una planta semejante al junco era abierta en tiras que se humedecían y entretejían. Una vez secas, constituían un excelente papel en que escribir.) Aunque byblos originalmente significa “papiro” en griego, con el tiempo paso a significar “libros” Los libros \*tal como lo conocemos deben su denominación a la antigua ciudad. (Davis, 1998).

Ahora bien, el termino no es solo un libro sino la recopilación de muchos libros que fueron escritos en un tiempo aproximado de 4000 años por diferentes autores que trataron temas de leyes, poesía, filosofía e historia sabiduría. La reunión y organización de la biblia también se vuelve en un tema de estudio interesante, ya que dependiendo del grupo religioso que se esté basado su organización cambia, por ejemplo, la biblia de un judío, no es la misma de un católico y la biblia de un católico no es la de un protestante.

La Biblia se divide en Antiguo Testamento y Nuevo Testamento. Es interesante preguntarse a que se refiere la palabra testamento, que tiene varios significados. Uno es algo que mucho no quisiéramos pensar y es la la última voluntad que es un documento legal para el uso de los bienes terrenales de un difunto. Otro significado es el dejar evidencia de algo, por ejemplo …….. y la utilizada para aludir a nuestro tema en cuestión es la manera antigua de decir “pacto” que aludía a un acuerdo o contrato. O sea, el antiguo testamento era el pacto entre Dios y su pueblo y con el nuevo testamento es un nuevo pacto a través de la vida, la muerte y la resurrección de Jesús.

Para los judíos que no tienen sino el antiguo testamento ya que no creen en el nuevo testamento escrita hebreo antiguo se puede decir que es el equivalente al antiguo testamento de la biblia de los cristianos. Se divide en tres: Tora, Profetas y escrituras

|  |  |  |
| --- | --- | --- |
| **Tora** | **Profetas** | **Escrituras** |
| Génesis  Éxodo  Levítico  Números  Deuteronomio | Jesúa  Jueces  Samuel (I)  Samuel (II)  Reyes (I)  Reyes (II)  Isaías  Jeremías  Ezequiel  Oseas  Joel  Amos  Abdías  Jonás  Miqueas  Nahum  Habacuc  Sofonías  Hageo  Zacarias  Malaquías | Salmos  Proverbios  Job  Cantar de Los cantares  Ruth  Lamentaciones  Eclesiastés  Ester  Daniel  Esdras  Nehemías  Crónicas (I)  Crónicas (II) |

Tabla 1 Libros de la Biblia Hebrea o Antiguo testamento

La biblia para la mayoría de los cristianos antiguo testamento

|  |
| --- |
| Libros Antiguo Testamento en la mayoría de las biblias cristianas |
| Génesis  Éxodo  Levítico  Números  Deuteronomio  Josué  Jueces  Ruth  Samuel (I)  Samuel (II)  Reyes (I)  Reyes (II)  Crónicas (I)  Crónicas (II)  Esdras  Nehemías  Ester  Job  Salmos  Proverbios  Eclesiastés  Cantares  Isaías  Jeremías  Lamentaciones  Ezequiel  Daniel  Oseas  Joel  Amos  Abdías  Jonás  Miqueas  Nahum  Habacuc  Sofonías  Hageo  Zacarias  Malaquías |

Tabla 2 Libros Antiguo testamento mayoría biblias cristianas

Otro punto que de los autores hay controversias pero que en los mismos escritos se pueden extractar quien es el que los escribe aunque si hay libros que siguen con diferentes teorías de quien es el autor, Un ejemplo es Deuteronomio que se le considera a Moisés el autor, pero en el ultimo capitulo data su muerte, por lo que probablemente este fue escrito por Josué. (Jesúa en el Tora)

* La Biblia como texto base (Ya digitados o en sitios web)
  + Antiguo testamento
  + Nuevo Testamento
  + Libros que ayudan en la cronología libros judíos y históricos.
* Procesamiento de Lenguaje Natural, por sus siglas en ingles Natural Language Processing.
* El método de visualización se esta investigando con varias herramientas, Google Sheet, Google Bert, Power Bi

Un primer sitio consultado es un sitio web dedicado a la cronología bíblica realizada por Rick Aschman la cual tiene un estudio que presenta datos y una visualización que nos da una orientación hacia el uso de datos para presentar eventos, personajes y fechas con una visualización en una línea de tiempo. (Aschmann, 2022)

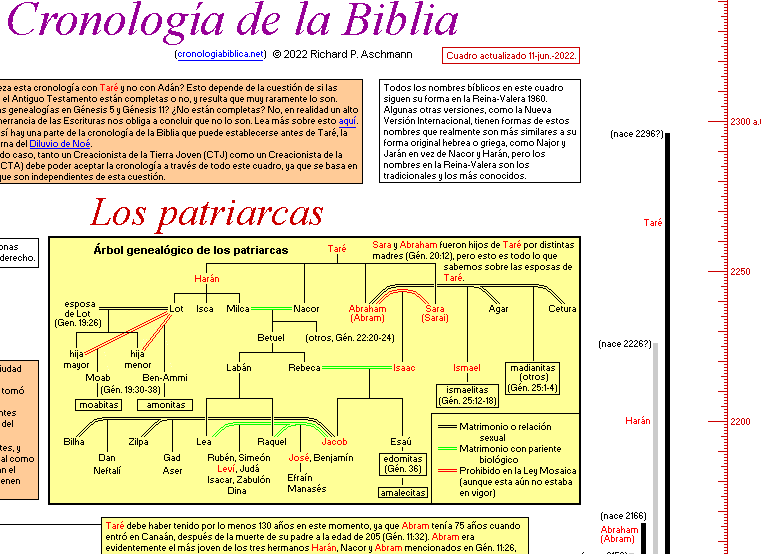


Ilustración 1 Parte de la visualización de cronología de la Biblia (Aschmann, 2022)

<https://aschmann.net/BibleChronology/castellano.html>

## 2.3 Conclusiones (nexo de unión de lo investigado con el trabajo a realizar)

* Se esta trabajando un grupo de pastores en Colombia que serán parte del proceso de analizar las dificultades y serán parte de la investigación, por medio de entrevistas y soporte en el prototipo de parte del uso.

# 3. Objetivos concretos y metodología de trabajo

## 3.1. Objetivo general

El objetivo general de este piloto experimental es desarrollar un sistema de análisis de datos masivos y visualización que facilite la comprensión de la narrativa bíblica en los contextos históricos de tiempo y cada momento de la historia.

## 3.2. Objetivos específicos

* Desarrollar un prototipo de modelo NLP que permita extraer información cronológica de los textos bíblicos.
* Crear una base estructurada de las fechas de los personajes y eventos bíblicos.
* Utilizar herramientas de visualización para presentar los eventos y personajes en forma dinámica.

# 6. Bibliografía

# Bibliografía

Aschmann, R. (21 de 05 de 2022). *Chronology of the Bible*. Obtenido de https://aschmann.net/BibleChronology/castellano.html

*Concepto y definicion Net*. (28 de 7 de 2023). Obtenido de https://conceptodefinicion.net/descubre-cuantas-palabras-hay-en-la-biblia/

Davis, K. C. (1998). *Que se yo de la Biblia.* (T. Arijon, Trad.) New York: Sudamericana.

*EL orden Mundial EOM*. (20 de 12 de 2019). Obtenido de El cristianismo en el mundo: https://elordenmundial.com/mapas-y-graficos/cristianismo-en-el-mundo/

*Embajada de Israel en España*. (s.f.). Obtenido de https://embassies.gov.il/madrid/AboutIsrael/AmongtheNations/Pages/ENTRE-NACIONES-Judeidad.aspx

*GENSIM topic modeling for humans*. (Consultada en Nov 22 2023). Obtenido de https://radimrehurek.com/gensim/

*NLTK*. (22 de 11 de Consultada en 2023). Obtenido de https://www.nltk.org/

Ravichandiran, S. (January 2021). *Getting Started with Goggle BERT.* Birmingham, UK: Packt Publishing Ltda.

Sarah Guthals, P. P. (2019). *GitHub dummies A wiley brand.* Hooboken, NJ, USA: John Wiley & Son, inc.

Sowmya Vajjala, B. M. (2020). *Practical Natural language Processing.* Sebastopol, CA, USA: O'Relly Media, inc.

*Textblob: Simplified Text Processing*. (Noviembre de Consultado en 2023). Obtenido de https://textblob.readthedocs.io/en/dev/

*Tranformers Huggin Face*. (Consultado en nov 22 2023). Obtenido de https://huggingface.co/docs/transformers/index