

**Universidad Nacional Autónoma de México**





**Facultad de Ingeniería**



División de Ingeniería Eléctrica

Asignatura: Análisis y Procesamiento Inteligente de Textos

Profesor (a):M.P. Octavio Augusto Sánchez Velázquez

PROYECTO

**“RESUMEN DE TEXTOS AUTOMÁTICO”**

Alumna(s): Buendía Valdivia María Angélica

Isabel Gómez Yareli Elizabeth

Zagoya Mellado Roberto Uriel

Grupo 1

Semestre 2020-1

Fecha de entrega: 2020

Índice

[**I.Introducción**](#_30j0zll) **3**

[Motivación](#_ecilo6iy6jw8) 3

[Justificación](#_14l0qks1afp1) 3

[Problema](#_vegg65aaymt2) 3

[Solución Propuesta](#_vegg65aaymt2) 3

[**II.Marco teórico**](#_myrxt2j8zwhc) **3**

[Antecedentes](#_8zxiqsrw9dn7) 3

[Trabajos relacionados](#_foktee8t9uic) 3

[Descripción](#_lvtpvn3jxxos) 3

[**III.Método experimental**](#_17tn2qdhaq0x) **3**

[Datos ¿Cómo se obtuvieron?](#_cjcl4r9sr9fz) 3

[Descripción de su método](#_myohg9kd7osg) 3

[Descripción del experimento](#_flv771hsngry) 3

[Presentación de resultados](#_1ttobnke38lp) 3

[**IV.Conclusión**](#_tajh1x6g08d) **3**

[Discusión de resultados](#_jjcxji5veuf) 3

[Descripción de observaciones](#_g91yfzfw2xgv) 3

[Trabajo futuro](#_l1d4xpkv055f) 3

[**V.Referencias**](#_m4xtkaq771sx) **4**

# **I.Introducción**

Durante la existencia de la humanidad el conocimiento se ha convertido esencial para su desarrollo y crecimiento, por lo cual la lectura ha sido un punto principal para desempeñar y llegar a cumplir cualquier objetivo. Pero muchas veces existen lecturas con un escritos de gran extensión, en las cuales las ideas principales se pueden dar a conocer en un escrito mucho más corto y conciso.

Para comprender la mayoría de la información se tiene que leer un texto sumamente largo, lleno de conceptos y con una letra muy chica que en veces no alcanzas a ver, para algunos, la lectura, no es su fuerte, además de consumir mucho tiempo y energía para comprender una idea, pero ahora ya existen herramientas que te permiten ingresar el texto y resumirlo en cuestión de milisegundos. Así se puede acortar mucho más el tiempo que inviertes en realizar esta actividad lector y obtener el mismo resultado de haber captado la misma información.

Para este proyecto realizaremos una aplicación la cual podrá resumir un texto conservando las ideas más importantes que hay en este, se podrá ingresar (copiar y pegar) el texto, directamente en esta aplicación o, con la URL de la página que contiene un texto del cual tú quieres obtener un resumen. Este no será un sistema sin fallas y puede ser que en los resúmenes, contenga equivocaciones porque funcionará con base en los algoritmos propios de las librerías de python, las cuales se describirán en secciones posteriores, y no usa un pensamiento lógico y crítico del ser humano, por lo tanto, deberá estar atento al resumen obtenido y se deberá revisar los párrafos finales, para verificar la coherencia y verificar si es el resultado.

### Motivación (interna o Externa)

La motivación para realizar este proyecto fue ver la necesidad principalmente de los estudiantes para realizar resúmenes y así facilitar y analizar su contenido desde artículos a textos de internet.

### Justificación (que es lo que justifica la existencia de ese trabajo)

El proyecto fue elegido entre tres diferentes proyectos, las opciones que existían eran “El análisis de sentimientos”, “Chatbot” y “El resumen de textos”. Estos tres fueron elegidos por diferentes motivos pero desde nuestro punto de vista los consideramos muy interesantes para implementarlo en la materia. Nuestras opciones fueron aprobadas por el profesor, pero dado a el factor de tiempo, el profesor nos sugirió realizar el más óptimo para realizarlo durante el semestre y el que se eligió fue el de “Resumen de textos”.

### Problema/Pregunta (¿Qué es lo que pretende responder con este trabajo)

Hoy en día existen muchas actividades que realizar, la sociedad cada vez trata de realizar todo de una manera más automática.

### Solución Propuesta

Para la solución a nuestro proyecto estuvimos investigando en varias fuentes por lo que encontramos varias opciones para la solución. La opción que nos pareció como la mejor es la siguiente.

En lo que primero que se pensó fue en el lenguaje de programación, estuvimos investigando y “python” fue el elegido ya que cuenta con Bibliotecas útiles como NLTK, esta sirve para el análisis de procesamiento de lenguaje natural (PLN) por lo cual fue de suma importancia para realizar el proyecto.

Otro punto que se tomó en cuenta fue el proceso para la elaboración del resumen paso a paso, el siguiente procedimiento es el que se tomó en cuenta para la solución y llegar al resumen del texto.

1.-Elegir el texto

2.-Convertir párrafos a oraciones

3.-Preprocesamiento de texto

4.-Tokenizado las oraciones

5.-Encontrar la frecuencia ponderada de ocurrencia

6.-Reemplazar palabras por frecuencia ponderada en oraciones originales

7.-Ordenar oraciones en orden descendente de la suma

# **II.Marco teórico**

### Antecedentes

Para cualquier problema planteado con el análisis y procesamiento de textos existe un campo amplio de información en el cual nos basamos para realizar nuestro trabajo y esperamos que con lo realizado los resultados obtenidos contribuyan y aporten a dicho tema. En el trabajo presenta se mostrarán diversos trabajos que nos permitirán o aportarán en mayor o menor medida recursos necesarios para el desarrollo de nuestra aplicación.

Uno de los principales documentos que consultamos fue la tesis titulada *Sistema Resumidor-Traductor Automático*  del Ing. Jouse Antonio Careaga Moya publicada en el año 2012 por la UNAM, la cual nos permite reforzar la parte que el resumidor que construyamos debe contener las ideas principales y dar una idea completa del texto. Nos guiamos en este trabajo ya que emplea la librería NLTK la cual brinda la posibilidad de segmentar el texto en enunciados, separados por comas, puntos o saltos de línea, por lo tanto podemos indicar que el trabajo citado nos orienta en cómo llevar a cabo el proceso de elaboración ya mencionado. [[1]](#footnote-1)

A lo largo de que el lector conozca más sobre el trabajo presentado se dará cuenta que el trabajo citado en el párrafo anterior es de gran ayuda para la solución de nuestro planteamiento, sin embargo, se analizaron otros trabajos como el de Pedro Luis Mateo, José Carlos González, Julio Villena y José Luis Martínez titulado *Un sistema para resumen automático de textos en castellano* en el cual se presenta las técnicas que deben ser aplicadas para detectar palabras clave, dar ponderaciones con el fin de no solo crear extracciones simples y sin coherencia ya que el fin de un resumen es mostrar la idea del texto basado en el original y no sólo copiar las frases idénticamente. El trabajo nos resulta útil e interesante ya que consideramos que nos permite evaluar si el resumen creado esta cumpliendo con el concepto, si este está siendo analizado correctamente para crear finalmente un resumen de la mejor calidad, los cuales son puntos que hemos considerados importantes y necesarios en nuestro trabajo. [[2]](#footnote-2)

Consideramos que los dos trabajos citados son los más relevantes ya que contienen una amplia descripción y brindan una o más soluciones al problema planteado, además que coinciden con la solución propuesta.

### Trabajos relacionados.

Algunos de los trabajos relacionados son los mencionados en la parte de antecedentes, para el primero de ellos se implementa Python y algunas de las librerías que son útiles para la creación de un resumidor de textos, el segundo es un trabajo internacional que de igual manera nos permite tener una guía para dar solución a nuestro problema. En general como ya se mencionó en un inicio contamos con una amplia variedad de artículos y trabajos que nos proporcionan ideas y soluciones que debemos implementar en nuestro propio proyecto.

Además tenemos como guía las ventajas que son definidas en todas las aplicaciones disponibles online o en aplicaciones, las cuales nos permiten a través de una comparativa identificar cuáles son los principales puntos con los que debe contar nuestro programa.

**Descripción.**

**-Resumen de textos automático.**

Un resumen de textos automático es aquel que se ejecutará por el usuario pero que no requerirá la intervención de este en la generación del documento final, dicho documento debe contener una idea concisa del documento original el cual será procesado por diferentes funciones del programa clasificando las ideas más relevantes y ordenandolas con el fin de generar un documento que sea lo más coherente posible. El hecho de generar un resumen de textos es una tarea que corresponde a la disciplina del procesamiento de lenguaje natural.

**-Procesamiento de Lenguaje Natural.**

Corresponde a una disciplina que emplea diversas técnicas con el fin de que el lenguaje natural, es decir el lenguaje hablado por los humanos sea procesado por las computadoras o una máquina, dicho problema no tiene una solución fácil ya que se requiere que el programador conozca tanto el lenguaje natural como el de la máquina con el fin de hacer que una computadora genere los procesos necesarios con el fin de obtener resultados que los humanos realizamos como el comparar textos, resumir textos o encontrar una idea principal.

**-Recuperación de la información.**

Indica el problema que se tiene con poder recuperar la información extraída por el documento en un inicio, con proporcionar la información solicitada y mostrarla de manera exitosa.

**-Técnicas para generar un resumen de textos.**

Podemos encontrarnos con diversas técnicas que van desde generarlo a partir de temas específicos, otro ejemplo es introducir un aprendizaje para que compare posteriormente los textos solicitados o bien los que se basan en análisis discursivos a partir de árboles por ejemplo, en el trabajo presenta se pretende generar un análisis a partir de los tokens identificados a través de NLTK de Python. En general en una aplicación de resumen de textos veremos como punto clave el uso de la lingüística y sus conceptos la cual nos permitirá identificar los elementos necesarios en un texto y a partir de esto generar un resumen lo más adecuado y completo posible.

# **III.Método experimental**

### Datos ¿Cómo se obtuvieron?

### Descripción de su método

### Descripción del experimento

### Presentación de resultados

# **IV.Conclusión**

### Discusión de resultados

### Descripción de observaciones

### Trabajo futuro

# **V.Referencias**

**LIBRERIAS A INSTALAR CHICOS ;)**

* nltk
* beautifulsoup
* lxml
* heapq

python 3.7

# 

**Buendía Valdivia María Angélica**

**Isabel Gómez Yareli Elizabeth**

**Zagoya Medallo Roberto Uriel**

1. Careaga, J.2012. *Sistema Resumidor-Traductor Automático. México, CDMX. UNAM. Consultado 27/03/2020 en:* [*http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/2640/Tesis.pdf?sequence=1*](http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/2640/Tesis.pdf?sequence=1) [↑](#footnote-ref-1)
2. Mateo, P. González, J.Villena, J. Martínez, J. *Un sistema para resumen automático de textos en castellano. Centro de Empresas la Arboleda. Madrid. Consultado el 20/03/2020 en:* [*http://rua.ua.es/dspace/handle/10045/1486*](http://rua.ua.es/dspace/handle/10045/1486) [↑](#footnote-ref-2)