



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

“Generador de resumen automático”

Alumna(s): Buendía Valdivia María Angélica
Espinosa Curiel Oscar
Isabel Gómez Yareli Elizabeth
Zagoya Mellado Roberto Uriel

Grupo 1

Semestre 2020-1



Fecha de entrega: 12 de junio 2020

Índice

I. Introducción3

Motivación (interna o Externa)5

Justificación (que es lo que justifica la existencia de ese trabajo)5

Problema/Pregunta (¿Qué es lo que pretende responder con este trabajo?).6

Solución Propuesta6

II. Marco teórico11

Trabajos relacionados. (Comparación con los trabajos y el nuestro)12

III. Método experimental15

III.I. Datos ¿Cómo se obtuvieron?15

III.II. Descripción de su método.15

Descripción del experimento (Responder a la pregunta).¡Error! Marcador no definido.

Presentación de resultados.30

IV. Conclusión39

Discusión de resultados39

Descripción de observaciones51

Trabajo futuro52

V. Referencias53

I. Introducción

Durante la existencia de la humanidad el conocimiento se ha convertido esencial para su desarrollo y crecimiento, por lo cual la lectura de información ha sido un punto primordial para desempeñar y llegar a cumplir cualquiera de sus objetivos.

Para comprender la mayoría de la información se tiene que leer un texto sumamente largo, lleno de conceptos y con una letra muy pequeña que en algunas ocasiones no se alcanza a ver. Para algunos, la lectura no es su hobby, además consume mucho tiempo y energía para comprender una idea, así que surge la necesidad de acortar la información más relevante. Dado esto, nace el resumen, este es un escrito que logra captar todas las ideas principales de un texto y sintetizarlas en un escrito mucho más pequeño. La extensión de un resumen varía, ya que no todos los textos son de la misma extensión, pero la mayoría tiene el 25% de la extensión original.

El objetivo del resumen es lograr reducir información a la más importante de cualquier tema, demostrar que se entiende lo resumido del texto original y que sea de utilidad para las personas. Los resúmenes han sido de gran ayuda ya que han servido para la realización de presentaciones, para estudiar algún tema, para realizar alguna exposición o para un trabajo académico.

Existen tipos de resúmenes como el indicativo y el informativo, el indicativo es aquel donde se incluyen los temas del documento resumido y la metodología y no supera la extensión de las 100 o 200 palabras. El resumen informativo es aquel en donde se profundiza sobre el texto original siendo un método científico (objetivos, metodología, resultados y conclusiones), en el proyecto se implementa el resumen indicativo.

En décadas anteriores aún se veía que se elaboraban con más frecuencia los resúmenes a mano, es decir, la persona leía todo el texto y al final o durante la lectura hacía anotaciones y subrayaba los puntos que a esa persona se le hacían relevantes e importantes, en la actualidad aún se sigue utilizando pero conforme pasa el tiempo se ha hecho de suma importancia el elaborar resúmenes en gran volumen, así que gracias a que la tecnología ha ido evolucionando hoy en día existen diferentes

softwares que nos permiten realizar estos resúmenes de una manera más práctica, automática. Estas herramientas ahora permiten ingresar el texto y resumirlo en cuestión de milisegundos, por lo que así se puede acortar mucho más el tiempo que invierte el lector en realizar esta actividad y obtener el mismo resultado de haber captado la misma información.

Algunas tecnologías actuales pueden hacer un resumen coherente de un texto al tener en cuenta variables como la longitud, la escritura y la sintaxis.

El resumen automático es parte del área del aprendizaje automático y de la minería de datos. Existen dos enfoques del resumen automático: la extracción y la abstracción, la extracción funciona seleccionando un subconjunto de palabras, frases u oraciones existentes en el texto original para formar el resumen, en la abstracción es un método que construye una representación semántica interna, a la que le aplican técnicas de generación de lenguaje natural para crear un resumen más similar al que realizaría un humano que más adelante se explicarán con más detalle.

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 los resúmenes, muchas veces se junta la carga de trabajo y no da tiempo de leer todos los documentos completos porque algunos tienen gran extensión de contenido y la mayoría de los profesores dejan muchos artículos y algunos libros para leer, pero para eso también sirven los resúmenes para reducir el tiempo y poder captar las ideas esenciales del documento sin tener que haber leído todo. Al elaborar este proyecto lo que se pretende es que los alumnos tengan una herramienta útil para facilitar y analizar su contenido desde artículos hasta textos de internet, de igual manera el ayudar a tener una mejor comprensión lectora.

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, lo que nos llamó la atención en el análisis de sentimientos es que este es muy utilizado en redes sociales para verificar la opinión que se tiene de las comunidades sobre ciertos temas, también es utilizada en organizaciones mundiales, un ejemplo es Obama, él lo utilizó para sondear la opinión en las elecciones. Lo que encontramos de este proyecto es que el entender los sentimientos no es una tarea fácil ya que se requiere de conocer culturas, interpretar palabras dentro de un contexto, asumir faltas de ortografía y a partir de ello determinar diferentes reglas que nos indiquen que es a lo que el usuario se quiere referir, etc.. En pocas palabras es un proyecto que requería de mucho tiempo y de algunos conceptos extra-clase. Analizando el proyecto de chatbot, éste nos pareció interesante ya que este es un programa informático con el que es posible mantener una conversación, tanto si queremos pedirle algún tipo de información o que lleve a cabo una acción. Para la elaboración de este proyecto el profesor dijo que teníamos poco tiempo para su realización ya que en un semestre no lo íbamos a poder terminar a menos que obtuviéramos ciertos programas ya elaborados, pero ese no era el objetivo, ya que el profesor

quería que nosotros empezáramos el proyecto desde cero, considerando la opinión del profesor opinamos que la dificultad era muy alta para elaborar este proyecto en un semestre teniendo más materias por las cuales se debe de poner atención al igual que en esta materia, así que se llegó a la conclusión que los tres proyectos desde nuestro punto de vista valían la pena realizarlos, los consideramos muy interesantes para implementarlos en la materia.

Dado que nuestras opciones fueron aprobadas por el profesor, al haber tomado en cuenta el factor de tiempo, el profesor nos sugirió realizar el óptimo para realizarlo durante el semestre y el que se eligió fue el de "Resumen de textos".

Problema/Pregunta (¿Qué es lo que se pretende responder con este trabajo?).

Hoy en día existen muchas actividades que hacer, la sociedad cada vez trata de realizar todo de una forma más práctica. Conforme pasa el tiempo el internet de las cosas (IoT) y todo lo que se refiere a las TIC ha ido entrando más en nuestros hogares, esto ha provocado que la mayoría de las personas quieran realizar sus actividades de una manera más fácil y automática, por esta razón ha surgido este proyecto el "Generador de resumen". Lo que se pretende es precisamente eso, es permitirles a las personas realizar sus resúmenes de una manera más fácil, práctica y en poco tiempo, sin tener que escribir a mano y sin invertir mucho tiempo.

Solución Propuesta.

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

En lo 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 eligió para la solución y así llegar al resumen del texto.

1.-Elegir el texto

Este paso consiste en brindar al usuario dos opciones para obtener su resumen, la primera es pegando un texto que tenga escrito en computadora y la segunda es a través de la URL que se encuentre en una página web la cual contenga el texto que quiere resumir.

Para la explicación de la solución propuesta pondremos un ejemplo de un párrafo de un discurso del famoso Denzel Washington en los premios de Imagen NNAACP.

"Entonces, sigue trabajando. Sigue esforzándote. Nunca te rindas. Cáete siete veces, levántate ocho. La facilidad es una mayor amenaza para el progreso que las dificultades. Entonces, sigue moviéndote, sigue creciendo, sigue aprendiendo. Nos vemos en el trabajo".

En el párrafo anterior básicamente lo que está diciendo el famoso es que está motivando a otros para que trabajen duro y nunca se rindan. Para resumir el párrafo anterior se utilizan las técnicas basadas en NLP mencionadas anteriormente, por las que para hacer el resumen se deben de realizar los siguientes pasos.

2.-Convertir párrafos a oraciones

Primero necesitamos convertir todo el párrafo en oraciones, es decir, dividir el párrafo cada vez que se encuentra un período. Entonces, si dividimos el párrafo en oraciones, obtenemos lo siguiente:

- Así que sigue trabajando
- Sigue esforzándote
- Nunca te rindas
- Cáete siete veces, levántate ocho
- La facilidad es una mayor amenaza para el progreso que las dificultades
- Así que sigue moviéndote, sigue creciendo, sigue aprendiendo.
- Nos vemos en el trabajo

3.-Preprocesamiento de texto

Después de convertir el párrafo en oraciones, debemos eliminar todos los caracteres especiales, detener palabras y números de todas las oraciones. Después del preprocesamiento, obtenemos las siguientes frases:

sigue trabajando

sigue esforzándote

nunca te rindas

caer siete veces conseguir ocho

aliviar la amenaza mayor progreso dificultades

sigue moviendo sigue creciendo sigue aprendiendo

ver trabajo

4.-Tokenizado las oraciones

Necesitamos tokenizar todas las oraciones para obtener todas las palabras que existen en estas. Después de tokenizar las oraciones, obtenemos una lista de las siguientes palabras:

```
['mantener',  
'trabajando',  
'mantener',  
'esforzarse',  
'Nunca',  
'dar',  
'otoño',  
'Siete',  
'hora',  
'obtener',  
'ocho',  
'facilitar',  
'mayor',  
'amenaza',  
'Progreso',  
'privación',  
'facilitar',  
'mayor',  
'amenaza',  
'Progreso',  
'privación',  
'mantener',
```



```
'Moviente',
'mantener',
'creciente',
'mantener',
'aprendizaje',
'ver',
'trabajo']
```

5.-Encontrar la frecuencia ponderada de ocurrencia.

A continuación, necesitamos encontrar la frecuencia ponderada de las ocurrencias de todas las palabras. La frecuencia ponderada de cada palabra se obtiene al dividir su frecuencia por la frecuencia de la palabra que aparece con mayor frecuencia. La siguiente tabla contiene las frecuencias ponderadas para cada palabra:

Palabra	Frecuencia	Frecuencia ponderada
facilitar	2	0.40
Ocho	1	0,20
otoño	1	0,20
obtener	1	0,20
dar	1	0,20
mayor	2	0.40
creciente	1	0,20
privación	2	0.40
mantener	5	1.00
aprendizaje	1	0,20
emocionante	1	0,20
Nunca	1	0,20
Progreso	2	0.40
ver	1	0,20

Palabra	Frecuencia	Frecuencia ponderada
Siete	1	0,20
esforzarse	1	0,20
amenaza	2	0.40
hora	1	0,20
trabajo	1	0,20
trabajando	1	0,20

Dado que la palabra "mantener" tiene la frecuencia más alta de 5, por lo tanto, la frecuencia ponderada de todas las palabras se ha calculado al dividir su número de ocurrencias entre 5.

6.-Reemplazar palabras por frecuencia ponderada en oraciones originales

El paso final es conectar la frecuencia ponderada en lugar de las palabras correspondientes en las oraciones originales y encontrar su suma. Es importante mencionar que la frecuencia ponderada de las palabras eliminadas durante el preprocesamiento (palabras de detención, puntuación, dígitos, etc.) será cero y por lo tanto, no es necesario agregarlas, como se menciona a continuación:

Frase	Suma de frecuencias ponderadas
Así que sigue trabajando	$1 + 0.20 = 1.20$
Sigue esforzándote	$1 + 0.20 = 1.20$
Nunca te rindas	$0.20 + 0.20 = 0.40$
Cáete siete veces, levántate ocho	$0.20 + 0.20 + 0.20 + 0.20 + 0.20 = 1.0$
La facilidad es una mayor amenaza para el progreso que las dificultades	$0.40 + 0.40 + 0.40 + 0.40 + 0.40 = 2.0$

Frase	Suma de frecuencias ponderadas
La facilidad es una mayor amenaza para el progreso que las dificultades	$0.40 + 0.40 + 0.40 + 0.40 + 0.40 = 2.0$
Así que sigue moviéndote, sigue creciendo, sigue aprendiendo.	$1 + 0.20 + 1 + 0.20 + 1 + 0.20 = 3.60$

Nos vemos en el trabajo $0.20 + 0.20 = 0.40$

7.-Ordenar oraciones en orden descendente de la suma.

El paso final es ordenar las oraciones en orden inverso de su suma. Las oraciones con las frecuencias más altas resumen el texto. Por ejemplo, mire la oración con la suma más alta de frecuencias ponderadas:

Así que sigue moviéndote, sigue creciendo, sigue aprendiendo.

Puedes juzgar fácilmente de qué se trata el párrafo. De manera similar, puede agregar la oración con la segunda suma más alta de frecuencias ponderadas para tener un resumen más informativo. Echa un vistazo a las siguientes oraciones:

Entonces, sigue moviéndote, sigue creciendo, sigue aprendiendo. La facilidad es una mayor amenaza para el progreso que las dificultades.

Estas dos oraciones dan un resumen bastante bueno de lo que se dijo en el párrafo.

II. Marco teórico.

Investigando acerca de los resúmenes automáticos y de sus inicios, han tenido una gran evolución desde que se comenzó la idea.

En los años finales de 1950 y 1960 se sugería que generar resúmenes de texto por computadora era viable, aunque no sencillo.

En el año de 1960 aparecen algunas bibliotecas en EUA, donde se quería agregar digitalmente el contenido de la biblioteca para facilitar su búsqueda, al existir límites de volumen de almacenamiento fue imposible almacenar todos los documentos, para ello se creó la idea de realizar resúmenes de estos, para que quien los consultará pudiera conocer más acerca de ellos.

Algunos documentos ya proporcionaban este pequeño resumen, pero otros no, esto hizo que se despertará el interés por técnicas que permitieran realizar resúmenes automáticos.

En los 30 años siguientes se continuó usando esa técnica de resúmenes en otros ámbitos, pero hasta los años 90's es con el surgimiento y la extensión de WWW (World Wide Web) y la aparición de nuevas fuentes de información, es cuando los resúmenes automáticos cobran más importancia.

Actualmente, es muy diferente a la situación de los años 60's, ya que ahora contamos con más capacidad de almacenamiento más barata y seguramente irá incrementando durante los años siguientes. El problema que surgió hoy en día es filtrar la información abundante en pequeños resúmenes para que esta puede ser entendible. De esto surge la importancia de inventar nuevas técnicas para poder solucionar este problema. Tras algunas décadas, los avances en el PLN, texto en línea y especialmente en la web, han renovado el interés y las técnicas en la generación automática de resúmenes de texto.

Trabajos relacionados. (Comparación con los trabajos y el nuestro)

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, esperamos que con lo realizado los resultados obtenidos contribuyan y aporten a dicho tema. En el trabajo que se 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. Josué Antonio Careaga Moya publicada en el año 2012 por la UNAM, la cual nos permite entender que como característica principal un resumidor debe generar un resumen en el cual las ideas principales estén presentes 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

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 presentan 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 resultó útil e interesante ya que consideramos que nos permite evaluar si el resumen creado está 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 considerado importantes y necesarios en nuestro trabajo.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.

Algunos de los trabajos relacionados son los mencionados anteriormente, 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

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>

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>

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 ordenándolas 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 los mismos 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 presente 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

III.I. Datos ¿Cómo se obtuvieron?

Los datos para el desarrollo de nuestro proyecto se obtuvieron de diversos trabajos ya realizados y consultados para basarnos de ahí y realizar nuestro proyecto. Dado que Python cuenta con la biblioteca nltk, nosotros decidimos usar este lenguaje por su utilidad y facilidad para obtener más rápido y de forma eficiente los resultados.

La mayor parte de los datos aquí presentados fueron obtenidos de las fuentes ya consultadas y mencionadas en la sección del Marco Teórico del proyecto y de la solución propuesta.

III.II. Descripción de su método.

Para la descripción de nuestro método es importante conocer lo que realizan cada una de las bibliotecas que utilizamos del lenguaje de programación Python. A continuación, se detallará que es lo que hacen estas librerías.

III.III. Librería NLTK: Por sus siglas en inglés (Natural Language Toolkit), es una plataforma que se utiliza para realizar programas para análisis de texto. Esta plataforma fue liberada por Steven Bird y Edward Loper, gracias a un concurso de lingüística computacional en la Universidad de Pensilvania en el año 2001.

Es un conjunto de herramientas para el procesamiento del lenguaje natural, simbólico y estadístico para el lenguaje de programación Python. También incluye demostraciones gráficas, pero en este proyecto no se verá su utilidad.

Este está destinado para apoyar a la enseñanza como una herramienta de estudio individual y como plataforma en los sistemas de investigación de PLN y en áreas muy relacionadas como lingüística empírica, las ciencias cognitivas, inteligencia artificial y aprendizaje automático.

Para su instalación en Windows 10 se debe de realizar desde el cmd, escribiendo

```
pip install nltk
```

Para su instalación en Linux distribución Ubuntu se ejecuta el comando desde la terminal.

```
sudo apt-get update  
sudo apt-get install python-nltk
```

III.IV. ¿Por qué utilizar NLTK?

Primero debemos entender que el lenguaje natural es el que usamos los humanos para comunicarnos diariamente, a diferencia de los lenguajes de programación. El análisis de lenguaje natural puede ir desde solo contar frecuencias o bien desde analizar expresiones completas y darles utilidad.

III.IV.I. ¿Qué nos proporciona NLTK?

NLTK nos proporciona herramientas para el procesamiento del lenguaje natural, es decir, el lenguaje que utilizamos diariamente para la comunicación humana. El procesamiento del lenguaje natural puede ser sencillo como el caso de contar frecuencias como es el ejemplo del proyecto presentado en este trabajo, que muestra como a partir del conteo de frecuencias y valorización de oraciones se genera un resumen o bien el hecho de tomar los resultados para destacar el contenido de una oración, su relevancia, entre otros factores.

Por lo que para este trabajo se recalca nuevamente que se hará uso de Natural Language Toolkit (NLTK) la cual es una biblioteca de código abierto de Python, con el fin de manipular los datos para obtener un resultado útil.

III.IV.II. NLTK con 4 objetivos principales.

Sencillez.	Brindar a los usuarios un uso práctico sin agobiarlos con la depuración asociada al procesamiento del lenguaje natural.
Consistencia.	Proporcionar interfaces y estructuras fáciles de deducir.
Extensibilidad.	Proporcionar estructura en la que los nuevos módulos de software puedan acomodarse fácilmente, tomando en cuenta enfoques competitivos.

Modularidad.	Proporcionar componentes que puedan usarse de manera independiente sin necesidad de comprender el resto del kit de herramientas.
---------------------	--

III.IV. III. ¿Cómo vemos que se cumplen los objetivos de NLTK en nuestro proyecto?

Para el proyecto implementado del cual se habla en el documento presente identificamos que cumple al menos con tres objetivos como lo son la sencillez, la cual la identificamos porque los paquetes como stopwords o punkt son fáciles de asociar a la actividad que proveen, la consistencia nos resultó fácil de utilizar y, en cuanto a la modularidad podemos considerarla como verdadera ya que de NLTK solo usamos dos herramientas que son stopwords y punkt.

III.V. ¿Cómo funciona NLTK en nuestro proyecto?

Nos ayuda a realizar un conteo de las veces que aparece una palabra en un texto, el ver cuántas veces se encuentra en un texto podría no resultarnos aún del todo útil. Ya que podría arrojarnos un resultado sin sentido de palabras y números. Para poderle dar sentido a las palabras es necesario ver su importancia en el texto completo y por oración. La manera que esto es posible es a partir de las herramientas que nos proporciona nltk para el desarrollo de nuestro proyecto.

III.V.I. Token.

En el programa realizado se tokenizará todo el texto en un inicio, para lo cual haremos uso del módulo punkt y en específico de word_tokenizer, el cual nos dará subcadenas de una cadena, estos tokens son cada una de las formas que conforman al texto en general como lo pueden ser palabras, puntuación incluso números o caracteres numéricos, word_tokenize recibe lo que queremos tokenizar y el idioma.

Para el proyecto se utilizó `word_tokenize` el cual nos arroja cada una de las formas que se encuentren en un texto como se muestra en la imagen 1, donde apreciamos una lista que contiene cada una de las subcadenas que se pueden generar a partir del texto.

```
[ 'En', 'informática', ',', 'la', 'inteligencia', 'artificial', '(', 'IA', ')',
  ',', 'a', 'veces', 'llamada', 'inteligencia', 'de', 'máquina', ',', 'es', 'inte',
  'ligencia', 'demostrada', 'por', 'máquinas', ',', 'en', 'contraste', 'con', 'la',
  ',', 'inteligencia', 'natural', 'que', 'muestran', 'los', 'humanos', 'y', 'los', 'animales',
  ',', 'Los', 'principales', 'libros', 'de', 'texto', 'de', 'IA', 'def',
  'inen', 'el', 'campo', 'como', 'el', 'estudio', 'de', 'los', 'agentes', 'i',
  'nteligentes', ':", 'cualquier', 'dispositivo', 'que', 'perciba', 'su', 'e',
  'ntorno', 'y', 'tome', 'medidas', 'que', 'maximicen', 'sus', 'posibilidades', 'd',
  'e', 'lograr', 'con', 'éxito', 'sus', 'objetivos', ',', '[', '1', ']', 'Coloquia',
  'lmente', 'el', 'término', 'inteligencia', 'artificial', '"', 'se',
  'usa', 'a', 'menudo', 'para', 'describir', 'máquinas', '(', 'o', 'computadoras',
  ',', 'que', 'imitan', 'funciones', 'cognitivas', '"', 'que', 'los', 'h',
  'umanos', 'asocian', 'con', 'la', 'mente', 'humana', 'como', 'aprendi',
  'zaje', 'y', 'resolución', 'de', 'problemas', '"', 'A', 'medid',
  'a', 'que', 'las', 'máquinas', 'se', 'vuelven', 'cada', 'vez', 'más', 'capaces',
  ',', 'las', 'tareas', 'que', 'se', 'consideran', 'que', 'requieren', 'in',
  'teligencia', '"', 'a', 'menudo', 'se', 'eliminan', 'de', 'la', 'definición',
  'de', 'IA', 'un', 'fenómeno', 'conocido', 'como', 'efecto', 'de', 'IA',
  ',', '[', '3', ']', 'Una', 'broma', 'en', 'el', 'Teorema', 'de', 'Tesler', 'dice',
  'que', 'AI', 'es', 'lo', 'que', 'no', 'se', 'ha', 'hecho', 'todavía', '"',
  ',', '[', '4', ']', 'Por', 'ejemplo', 'el', 'reconocimiento', 'óptico',
  'de', 'caracteres', 'a', 'menudo', 'se', 'excluye', 'de', 'las', 'cosas', 'co',
  'nsideradas', 'como', 'AI', '[', '5', ']', 'convirtiéndose', 'en', 'una', ']
```

Imagen 1. Subcadenas

III.V.II. Frecuencias.

Es necesario indicar que para el proyecto en curso se hará uso de la frecuencia ponderada, que es el valor obtenido entre la frecuencia de una palabra dividida entre el número mayor de frecuencia de una palabra dentro del texto.

III.V.III. ¿Qué hay con la tabla de frecuencias?

La tabla de frecuencias es un diccionario, el cual se usa con el fin de poder almacenar una clave con su respectivo valor en este caso las claves son cada una de las formas o tokens que están presentes en el texto y los valores son los números correspondientes a las frecuencias, como se muestra en la imagen 2.

```
{'informática': 1, ',': 22, 'inteligencia': 7, 'artificial': 3, '(': 7, 'ia': 6, ')': 8, 'veces': 1, 'llamada': 1, 'máquina': 2, 'demostrada': 1, 'máquinas': 3, 'contraste': 1, 'natural': 1, 'muestran': 1, 'humanos': 2, 'animales': 1, ':': 10, 'principales': 1, 'libros': 1, 'texto': 1, 'definen': 1, 'campo': 1, 'estudio': 1, '``': 15, 'agentes': 1, 'inteligentes': 1, '': 1, 'cualquier': 1, 'dispositivo': 1, 'perciba': 1, 'entorno': 1, 'tome': 1, 'medidas': 1, 'maximice n': 1, 'posibilidades': 1, 'lograr': 1, 'éxito': 2, 'objetivos': 2, '[': 18, ']' : 1, ']: 18, 'coloquialmente': 1, 'término': 1, '"': 7, 'usa': 1, 'menudo': 4, 'describir': 1, 'computadoras': 1, 'imitan': 1, 'funciones': 1, 'cognitivas': 1, 'asocian': 1, 'mente': 1, 'humana': 2, 'aprendizaje': 2, 'resolución': 1, 'problemas': 1, 'medida': 1, 'vuelven': 1, 'cada': 1, 'vez': 1, 'capaces': 1, 'tareas': 1, 'consideran': 1, 'requieren': 1, 'eliminan': 1, 'definición': 1, 'fenómeno': 1, 'conocido': 2, 'efecto': 1, '3': 1, 'broma': 1, 'teorema': 1, 'tesler': 1, 'dice': 1, 'ai': 3, 'hecho': 1, 'todavía': 1, '4': 1, 'ejemplo': 2, 'reconocimiento': 1, 'óptico': 1, 'caracteres': 1, 'excluye': 1, 'cosas': 1, 'consideradas': 1, '5': 1, 'convirtiéndose': 1, 'tecnología': 1, 'rutina': 1, '6': 1, 'capacidades': 1, 'modernas': 1, 'generalmente': 1, 'clasificadas': 1, 'incluyen': 1, 'comprensión': 1, 'exitosa': 1, 'habla': 1, '7': 1, 'competir': 1, 'alto': 1, 'nivel': 1, 'sistemas': 1, 'juego': 1, 'estratégico': 1, 'ajedrez': 1, 'go': 1, '8': 1, 'vehículos': 1, 'autónomos': 1, 'enrutamiento': 1, 'inteligente': 1, 'redes': 2, 'entrega': 1, 'contenido': 1, 'simulaciones': 1, 'militares': 1, 'fundada': 1, 'disciplina': 1, 'académica': 1, '1955': 1, 'años': 1, 'transcurridos': 1, 'entonces': 1, 'experimentado': 1, 'varias': 1, 'oleadas': 1, 'optimismo': 1, '9': 1, '10': 2, 'seguido': 2, 'decepción': 1, 'pérdida': 1, 'fondos': 1, 'invierno': 1, '11': 1, '12': 1, 'nuevos': 1, 'enfoques': 1, 'financiación': 1, 'renovada': 1, '13': 1, 'mayor': 1, 'parte': 1, 'historia': 1, 'i
```

Imagen 2. Tabla de frecuencia absoluta

Como podemos observar se está ocupando un diccionario esto debido a que tenemos lo que se llama **lista simple**, es decir un token acompañado de su frecuencia, podemos encontrar otro tipo de listas como las que describen en la tabla siguiente.

Tipo de lista.	Descripción.	Ejemplo.
Simple.	Token acompañado de su frecuencia.	Clasificadas: 1
Canónicas.	Lista ordenada alfabéticamente y a cada línea le suceden las palabras que corresponden a dicha forma canónica.	El trabajo de Juan López Chávez y Marina Arjona de <i>Lexicometría y fonometría del Primer Sueño de Sor Juana Inés de la Cruz</i> .
Lemas.	Ordenadas por lemas o raíces, sin importar la parte de la oración de la que trate.	Se puede agrupar el lema nación en: Naciones, nacionalizar, nacionalmente y nacionalización.
Dos o más palabras.	No son palabras simples sino pares, terceras de palabras que ocurren contiguamente. Cada	Informática:1 sin: 2 amor:4

	una se acompaña de su frecuencia.	
Partes de la oración.	Lista que solo contenga partes de la oración.	Se usan solo referencias de categorías para fines estadísticos.

Una vez obtenida la frecuencia absoluta de nuestras palabras o ya tokens, obtenemos la frecuencia ponderada de cada una de ellas que como ya se mencionó corresponde al número total de apariciones en el texto entre el mayor número correspondiente a una palabra, dichas frecuencias ponderadas se muestran en la siguiente imagen.

```

Valor mayor de frecuencia de una palabra: 22
{'informática': 0.0454545454545414, ',': 1.0000000000000018, 'inteligencia': 0.31818181818181657, 'artificial': 0.136363636363624, '(': 0.31818181818181657, 'ia': 0.272727272727216, ')': 0.36363636363636154, 'veces': 0.0454545454545414, 'llamada': 0.0454545454545414, 'máquina': 0.090909090909083, 'demostrada': 0.0454545454545414, 'máquinas': 0.136363636363624, 'contraste': 0.0454545454545414, 'natural': 0.0454545454545414, 'muestran': 0.0454545454545414, 'humanos': 0.090909090909083, 'animales': 0.0454545454545414, '.': 0.4545454545454515, 'principales': 0.0454545454545414, 'libros': 0.0454545454545414, 'texto': 0.0454545454545414, 'definen': 0.0454545454545414, 'campo': 0.0454545454545414, 'estudio': 0.0454545454545414, '': 0.6818181818181763, 'agentes': 0.0454545454545414, 'inteligentes': 0.0454545454545414, ':': 0.0454545454545414, 'cualquier': 0.0454545454545414, 'dispositivo': 0.0454545454545414, 'perciba': 0.0454545454545414, 'entorno': 0.0454545454545414, 'tome': 0.0454545454545414, 'medidas': 0.0454545454545414, 'maximicen': 0.0454545454545414, 'posibilidades': 0.0454545454545414, 'lograr': 0.0454545454545414, 'éxito': 0.090909090909083, 'objetivos': 0.090909090909083, '[': 0.8181818181818148, '1': 0.0454545454545414, ']': 0.8181818181818148, 'co loquialmente': 0.0454545454545414, 'término': 0.0454545454545414, '"': 0.31818181818181657, 'usa': 0.0454545454545414, 'menudo': 0.181818181818166, 'describir': 0.0454545454545414, 'computadoras': 0.0454545454545414, 'imitan ': 0.0454545454545414, 'funciones': 0.0454545454545414, 'cognitivas': 0.0454545454545414, 'asocian': 0.0454545454545414, 'mente': 0.0454545454545414, 'humana': 0.090909090909083, 'aprendizaje': 0.090909090909083, 'resolución': 0.0454545454545414, 'problemas': 0.0454545454545414, 'medida': 0.045454

```

Imagen 3.Frecuencias ponderadas.

III.V.III.I. ¿De qué nos sirve la forma en que se crea la tabla de frecuencias?

Además, la tabla que podemos apreciar en este proyecto tiene un **orden por aparición**, ya que si nos percatamos no veremos los tokens con mayor frecuencia primero o en orden inverso, tampoco se muestran de manera alfabética o por la longitud de la palabra. Lo que nos indica que la tabla se va construyendo según vayan apareciendo en el texto.

Entonces el hecho de que se vayan acomodando las palabras por aparición y por lo tanto llenando de esta manera la tabla de frecuencias, nos beneficia en el hecho de que las palabras solo aparecerán una vez de acuerdo con el texto y de ahí el número

de veces que estén el texto, lo que se asemeja a la manera en que lo haríamos humanamente ya que primero contabilizaríamos y finalmente solo escribiríamos la palabra con su respectivo valor total. Adicionalmente nos permite conocer la distribución de las palabras y ver si esta indica un cambio en el sentido de la oración por el hecho de aparecer en un lugar distinto. Independientemente del tipo de lista con el que contemos nos permitirá en sentido general del proyecto conocer cuál es el sentido de la palabra, sin embargo, los tipos de orden que les demos nos harán más fácil su análisis o lo complicarán en menor medida.

III.VI. Problemas al realizar los diccionarios o listas de palabras.

Podemos enfrentarnos a la pérdida de información relevante o no, es decir, podemos obtener listas de palabras que contengan guiones que son esenciales en palabras compuestas, signos de puntuación eliminados que quiten sentido a las oraciones e incluso eliminación de mayúsculas que correspondan a determinadas siglas. Además de la eliminación podemos tener confusión en el sentido de números que estén escritos con letras y no con números, lo que quitaría el sentido a dichas cifras.

Entonces la pregunta relevante es, ¿Qué debemos eliminar y de acuerdo con qué? Para lo cual se han planteado soluciones como:

- Pre- procesamiento del texto: en donde se realcen todas las palabras incluyendo fechas más importantes, así como el uso de expresiones regulares.
- Ajustar el programa: indica que el programa usado o método empleado deberá ajustarse a lo que uno decida conservar o eliminar, dando uno el propio sentido.

III.VI.I. ¿Cómo saber que se ignora y a qué se le da importancia?

En un proyecto que desea generar un resumen como es el caso presentado, no es útil darle importancia a todas las palabras del texto que un usuario nos proporcione. Muchas de las veces los humanos al hacer un resumen creemos que es útil toda la información que nos encontramos al ir leyendo, pero no es así, al realizar un resumen debemos tener en cuenta el hecho de que solo queremos dar a conocer la idea principal, lo más importante o relevante de todo el texto.

Los resúmenes de los seres humanos contienen palabras conectoras que repetimos con frecuencia las cuales variarán de acuerdo con el idioma que se use. Para poder dar énfasis a las palabras más significativas de un texto en este proyecto se emplea la librería de Python stopwords, lo anterior no nos quiere indicar que no sean necesarias, sino que se trata en un análisis de este tipo de encontrar aquellas que dan un aporte sustancial.

Los stopwords son un conjunto de preposiciones, conjunciones, artículos, adverbios, pronombres y algunos verbos y forman alrededor del 30% al 40% de los textos en español, este tipo de palabras se les caracteriza como palabras vacías que si bien no se eliminan, se omiten para poder obtener lo más relevante.

Entonces un stopwords funcionará de la siguiente forma.

En informática, la inteligencia artificial (IA).
--

Al ejecutar el stopwords, lo que obtenemos es:

informática , inteligencia artificial (ia)
--

Lo que nos indica que hemos descartado el siguiente conjunto:
{En, la}

Que corresponde a artículos y preposiciones propiamente.

El uso de stopwords se emplea de la siguiente manera para python, primero importamos el corpus de stopwords, es decir el

conjunto de palabras donde se encuentran contenidas todas aquellas preposiciones, conjunciones, artículos, etc.

```
import nltk
# stopwords = palabras reservadas.
from nltk.corpus import stopwords|
```

Imagen 4. Stopwords en Python

Una vez importado se configura el idioma, para los casos implementados en este proyecto son español e inglés, como se muestra en las siguientes imágenes.

```
# arreglo 1) - SW - configura el método de entrada o
# idioma en el que se
# van a trabajar las palabras en el metodo words de
# stopwords, en este caso, Español.
SW = set(stopwords.words("spanish"))
```

Imagen 5. Idioma en que se puede cambiar el texto a ingresar (español).

```
SW = set(stopwords.words("english"))
```

Imagen 6. Idioma en que se puede cambiar el texto a ingresar (inglés).

Posteriormente se crea un diccionario de frecuencias llamado tabla de frecuencias que se ha mencionado en la sección anterior en este se guardarán las palabras y su valor asociado a estas, de acuerdo con la condición de encontrarse o no en SW que corresponde al corpus de las stopwords, de la manera siguiente.

```
# Con un for se recorre el texto y se almacena en la tabla.
for word in words:
    word = word.lower() # setea las palabras en minúscula y las almacena en
    word.
    if word in SW:
        continue # Si la palabra se encuentra en SW, continua con el ciclo.
    if word in freqTable: # Si la palabra ya se encuentra en la tabla
    frecuencia,
        freqTable[word] += 1 # Suma 1 a la posición donde se encuentra la
    palabra.
    else:
        freqTable[word] = 1 # Sino, la palabra en la TF va a ser igual a 1.
```

Imagen 7. Creación de tabla de frecuencias.

III.VI.II. ¿Cómo generar las oraciones?

Para el hecho de generar las oraciones o crear una lista donde estas estén albergadas se usa `sent_tokenize`, la cual permite al igual que `Word_tokenize` separar, aunque en este caso serán oraciones. Lo que nos dará un resultado de una lista que contenga las oraciones agrupadas de acuerdo con los signos de puntuación del texto original de dónde son tomadas. Por ejemplo, para el caso de un texto como el de la imagen obtendremos al utilizar `sent_tokenize` una lista como la que se muestra en la imagen inferior. Las oraciones al igual que las palabras por sí solas tendrán un valor ponderado que es igual a su frecuencia absoluta más su frecuencia ponderada obtenida en pasos previos ya descritos.

```
# text para definir el texto a resumir.
text = ""En informática , la inteligencia artificial ( IA ), a veces llamada
inteligencia de máquina , es inteligencia demostrada por máquinas , en
contraste con la inteligencia natural que muestran los humanos y los animales .
Los principales libros de texto de IA definen el campo como el estudio de los "
agentes inteligentes ": cualquier dispositivo que perciba su entorno y tome
medidas que maximicen sus posibilidades de lograr con éxito sus objetivos.
[1]Coloquialmente, el término "inteligencia artificial" se usa a menudo para
describir máquinas (o computadoras) que imitan funciones "cognitivas" que los
humanos asocian con la mente humana , como "aprendizaje" y "resolución de
problemas". A medida que las máquinas se vuelven cada vez más capaces, las
tareas que se consideran que requieren "inteligencia" a menudo se eliminan de
la definición de IA, un fenómeno conocido como efecto de IA . [3] Una broma en
el Teorema de Tesler dice que "AI es lo que no se ha hecho todavía". [4] Por
ejemplo, el reconocimiento óptico de caracteres a menudo se excluye de las
cosas consideradas como AI [5] , convirtiéndose en una tecnología de rutina.
```

Imagen 8. Definir el texto a resumir.

Se obtienen las siguientes oraciones.

```
SENTENCES
['En informática , la inteligencia artificial ( IA ), a veces llamada intelligen
cia de máquina , es inteligencia demostrada por máquinas , en contraste con la
inteligencia natural que muestran los humanos y los animales .', 'Los principal
es libros de texto de IA definen el campo como el estudio de los " agentes inte
ligentes ": cualquier dispositivo que perciba su entorno y tome medidas que max
imicen sus posibilidades de lograr con éxito sus objetivos.', '[1]Coloquialment
e, el término "inteligencia artificial" se usa a menudo para describir máquinas
(o computadoras) que imitan funciones "cognitivas" que los humanos asocian con
la mente humana , como "aprendizaje" y "resolución de problemas".', 'A medida
que las máquinas se vuelven cada vez más capaces, las tareas que se consideran
que requieren "inteligencia" a menudo se eliminan de la definición de IA, un fe
nómeno conocido como efecto de IA .', '[3] Una broma en el Teorema de Tesler di
ce que "AI es lo que no se ha hecho todavía".', '[4] Por ejemplo, el reconocimi
ento óptico de caracteres a menudo se excluye de las cosas consideradas como AI
[5] , convirtiéndose en una tecnología de rutina.', '[6] Las capacidades moder
```

Imagen 9. Oraciones obtenidas.

Además, permite separar las oraciones de acuerdo con los saltos de línea o tabulaciones que se indican en un texto, como en el caso del texto siguiente.

```
# text para definir el texto a resumir.
text = """Información Taxonómica

Las diez especies que componen actualmente el género Sus se localizan
principalmente en Asia, aunque el jabalí euroasiático, Sus scrofa, la especie
más abundante, muestra un área de distribución histórica más amplia, que
incluye Europa y el norte de África. El estudio taxonómico de la especie es
complejo por la variación de caracteres debido fundamentalmente a la
domesticación.

Descripción

Muestra la máxima altura en la región de las extremidades anteriores; el cuello
es poco aparente, las orejas son pequeñas y se mantienen erguidas.

Las dimensiones son muy variables; en Europa la corpulencia aumenta hacia el
este, observándose los menores tamaños en el sur de la Península Ibérica y los
mayores en los Cárpatos.

Medidas corporales de ejemplares adultos del Montseny (Cataluña), CC:
133,0-148,0 cm (machos), 118,0-137,0 cm (hembras); C: 17,5-24,0 cm (machos),
13,0-20,0 cm (hembras); CR: 72,0-85,5 cm (machos), 71,0-73,0 cm (hembras); Ps:
60-118 kg (machos), 40-65 kg (hembras).

Fórmula dentaria: 3.1.4.3/3.1.4.3. Caninos de puntas agudas y ángulos cortantes.
```

Imagen 10. Separar los textos de acuerdo a las tabulaciones.

Obteniendo.

```
['Información Taxonómica\n\nLas diez especies que componen actualmente el género Sus se localizan principalmente en Asia, aunque el jabalí euroasiático, Sus scrofa, la especie más abundante, muestra un área de distribución histórica más amplia, que incluye Europa y el norte de África.', 'El estudio taxonómico de la especie es complejo por la variación de caracteres debido fundamentalmente a la domesticación.', 'Descripción\n\nMuestra la máxima altura en la región de las extremidades anteriores; el cuello es poco aparente, las orejas son pequeñas y se mantienen erguidas.', 'Las dimensiones son muy variables; en Europa la corpulencia aumenta hacia el este, observándose los menores tamaños en el sur de la Península Ibérica y los mayores en los Cárpatos.', 'Medidas corporales de ejemplares adultos del Montseny (Cataluña), CC: 133,0-148,0 cm (machos), 118,0-137,0 cm (hembras); C: 17,5-24,0 cm (machos), 13,0-20,0 cm (hembras); CR: 72,0-85,5 cm (machos), 71,0-73,0 cm (hembras); Ps: 60-118 kg (machos), 40-65 kg (hembras).', 'Fórmula dentaria: 3.1.4.3/3.1.4.3.', 'Caninos de puntas agudas y ángulos cortantes.', 'Posee un marcado dimorfismo sexual, los machos son más corpulento
```

Imagen 11. Conjunto de oraciones.

Una vez que tenemos un conjunto de oraciones que aparecen en nuestro texto, es de importancia saber cuál es el proceso que determina como se valorizarán cada una de las oraciones.

Para esto se recorren las oraciones que se encuentran en el texto, posteriormente se recorren las partes que componen a la tabla de frecuencias, entonces si la palabra está en las

oraciones en minúscula y además en diccionario de oraciones a valorizar se suma uno a su frecuencia en la posición, en caso contrario el valor asignado para su posición corresponde al de su frecuencia absoluta, de esta forma se obtienen las oraciones con valor, como se muestra en la imagen siguiente.

```
# Variable sentences para almacenar las oraciones a valorizar del texto.
sentences = sent_tokenize(text)

print(sentences)

# Diccionario sentenceValue para almacenar los valores de las oraciones.
sentenceValue = dict()

#Se crea un ciclo for para recorrer las oraciones que se encuentran en el texto
for sentence in sentences:
    # Se crea un segundo for para recorrer los items que se encuentran en la TF
    for word, freq in freqTable.items():
        if word in sentence.lower(): # Si la palabra se encuentra en las
oraciones (en minúscula)
            if sentence in sentenceValue: # Y si la oración está en el
diccionario de las oraciones a valorizar.
                sentenceValue[sentence] += freq # Entonces suma 1 al número de
frecuencia en la posición de la oración del sv.
            else:
                sentenceValue[sentence] = freq # Sino, que el valor de la
posición de la oración sea igual a la frecuencia.

#Muestra las oraciones ya valorizadas con su respectivo puntaje.
#sentenceValue

print('\t\t\t*****ORACIONES VALORIZADAS*****')
```

Imagen 12. Código para obtener oraciones con valor.

A continuación, se muestran las oraciones ya valorizadas, el ejemplo mostrado es el del texto de la informática, se muestran las oraciones de acuerdo con los valores de frecuencia ponderada obtenidos con anterioridad.

```
{'En informática , la inteligencia artificial ( IA ), a veces llamada inteligen  
cia de máquina , es inteligencia demostrada por máquinas , en contraste con la  
inteligencia natural que muestran los humanos y los animales .': 3.545454545454  
537, 'Los principales libros de texto de IA definen el campo como el estudio de  
los " agentes inteligentes ": cualquier dispositivo que perciba su entorno y t  
ome medidas que maximicen sus posibilidades de lograr con éxito sus objetivos.'  
: 1.8181818181813, '[1]Coloquialmente, el término "inteligencia artificial" s  
e usa a menudo para describir máquinas (o computadoras) que imitan funciones "c  
ognitivas" que los humanos asocian con la mente humana , como "aprendizaje" y "  
resolución de problemas".' : 5.772727272727252, 'A medida que las máquinas se vu  
elven cada vez más capaces, las tareas que se consideran que requieren "intelig  
encia" a menudo se eliminan de la definición de IA, un fenómeno conocido como e  
fecto de IA .': 3.090909090909086, '[3] Una broma en el Teorema de Tesler dice  
que "AI es lo que no se ha hecho todavía".' : 2.545454545454535, '[4] Por ejempl  
o, el reconocimiento óptico de caracteres a menudo se excluye de las cosas cons  
ideradas como AI [5] , convirtiéndose en una tecnología de rutina.' : 3.99999999  
9999991, '[6] Las capacidades modernas de la máquina generalmente clasificadas  
como IA incluyen la comprensión exitosa del habla humana , [7] competir al más  
alto nivel en los sistemas de juego estratégico (como el ajedrez y el Go)), [8]  
vehículos autónomos , enrutamiento inteligente en redes de entrega de contenid  
o y simulaciones militares.La inteligencia artificial fue fundada como una disc  
iplina académica en 1955, y en los años transcurridos desde entonces ha experim  
entado varias oleadas de optimismo, [9] [10] seguido por la decepción y la pérd  
ida de fondos (conocido como un " invierno AI "), [11] [12] seguido de nuevos e  
nfosques, éxito y financiación renovada.' : 7.681818181818143, '[10] [13] Durante
```

Imagen 13. Oraciones ya valorizadas.

III.VII. ¿Cómo determinar que oraciones se añaden al resumen final?

Para llegar a crear el resumen una vez que hemos dado valor a cada oración será necesario hacer la suma total de los valores que se le han dado a estas, para lo cual se recorrerán una por una y en caso de que la oración pertenezca a las oraciones valorizadas sumaremos el valor a una variable que contenga el valor total.

Una vez que hemos obtenido la suma total lo que queda es obtener el promedio dividiendo el valor total entre el número de oraciones valorizadas. Una vez hecho esto podemos proseguir a añadir las oraciones al resumen.

Las oraciones se añadirán siempre y cuando el valor ponderado de la oración sea mayor a 1.2 por el valor obtenido de tal manera que al multiplicar el promedio por dicho valor priorizamos el que solo las oraciones con mayor valor se añadan.

El proceso y resultado se muestran en las siguientes imágenes.


```

# Valor promedio de una oración desde un texto original
average = int(sumValues/ len(sentenceValue)) # Divide la suma de valores en la
total de oraciones valorizadas.

# Se crea una variable para almacenar el resumen a imprimir.
summary = ''

# Se crea un for para recorrer las oraciones almacenadas
for sentence in sentences:

    #Donde si, la oración está las oraciones Valorizadas y la posición de la
    oración es mayor que 1.2 veces el promedio:
    if (sentence in sentenceValue) and (sentenceValue[sentence] > (1.2 *
    average)):

        # El resumen va a agregar un espacio más la oración que aprobó la
        condición.
        summary += " " + sentence

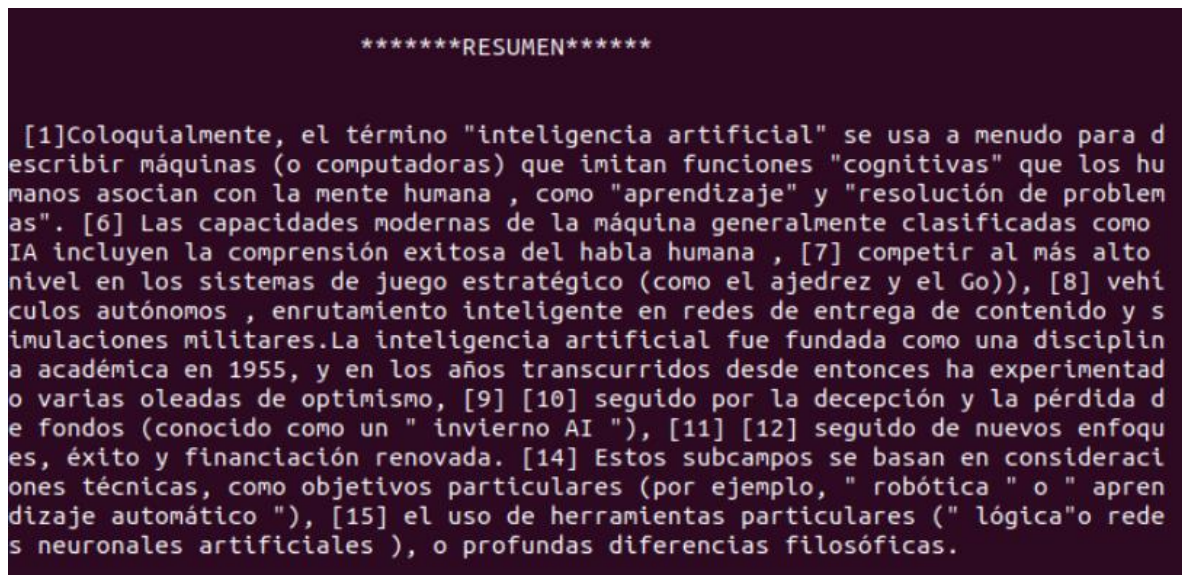
# Se imprime el resumen.
print('\t\t\t*****RESUMEN*****')
print("\n"+ "\n"+summary)

print("\n"+ "\n"+summary)

```

Imagen 14. Obtención de oraciones finales.

El resultado es:



```

*****RESUMEN*****

[1]Coloquialmente, el término "inteligencia artificial" se usa a menudo para d
escribir máquinas (o computadoras) que imitan funciones "cognitivas" que los hu
manos asocian con la mente humana , como "aprendizaje" y "resolución de problem
as". [6] Las capacidades modernas de la máquina generalmente clasificadas como
IA incluyen la comprensión exitosa del habla humana , [7] competir al más alto
nivel en los sistemas de juego estratégico (como el ajedrez y el Go)), [8] vehí
culos autónomos , enrutamiento inteligente en redes de entrega de contenido y s
imulaciones militares.La inteligencia artificial fue fundada como una disciplin
a académica en 1955, y en los años transcurridos desde entonces ha experimentad
o varias oleadas de optimismo, [9] [10] seguido por la decepción y la pérdida d
e fondos (conocido como un " invierno AI "), [11] [12] seguido de nuevos enfoqu
es, éxito y financiación renovada. [14] Estos subcampos se basan en consideraci
ones técnicas, como objetivos particulares (por ejemplo, " robótica " o " apren
dizaje automático "), [15] el uso de herramientas particulares (" lógica"o rede
s neuronales artificiales ), o profundas diferencias filosóficas.

```

Imagen 15. Resumen final.

Y es de esta forma que podemos obtener un resumen.

III.VIII. ¿Qué hay de la interfaz gráfica?

Librería PYQT5: Es una biblioteca gráfica Qt para el lenguaje de programación Python, está desarrollada por la firma británica "Riverbank Computing". Esta es muy utilizada para la creación de interfaces gráficas en Python, ya que es muy fácil de utilizar y entender en poco tiempo.

Para su instalación en Windows 10 se debe de realizar desde el cmd, escribiendo:

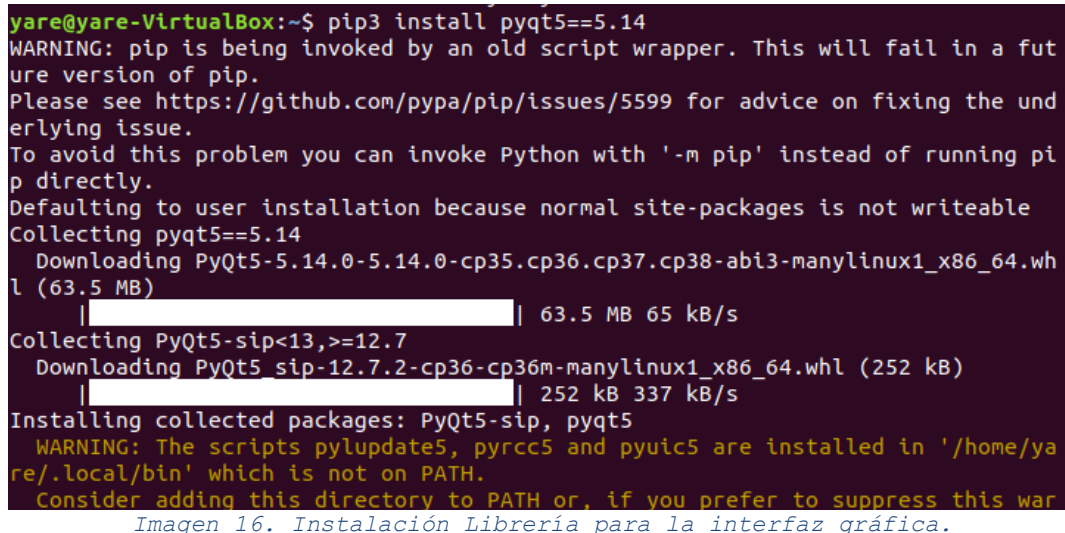
```
$pip install PyQt5
```

Para su instalación en Linux distribución Ubuntu se ejecuta el comando desde la terminal.

```
$pip3 install PyQt5-sip
Descomprimiéndolo
$python3 configure.py
$make
$make install
```

¿Qué hacer para tener una interfaz gráfica?

Primero para poder tener una interfaz gráfica e importar en nuestro código Qt Core, QtGui, QtWidgets es necesario instalar PYQT5 librería que nos permitirá importar lo necesario para poder mostrar al usuario una interfaz.



```
yare@yare-VirtualBox:~$ pip3 install pyqt5==5.14
WARNING: pip is being invoked by an old script wrapper. This will fail in a future
version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the und
erlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running pi
p directly.
Defaulting to user installation because normal site-packages is not writeable
Collecting pyqt5==5.14
  Downloading PyQt5-5.14.0-5.14.0-cp35.cp36.cp37.cp38-abi3-manylinux1_x86_64.wh
l (63.5 MB)
    |████████████████████████████████████████| 63.5 MB 65 kB/s
Collecting PyQt5-sip<13,>=12.7
  Downloading PyQt5_sip-12.7.2-cp36-cp36m-manylinux1_x86_64.whl (252 kB)
    |████████████████████████████████████████| 252 kB 337 kB/s
Installing collected packages: PyQt5-sip, pyqt5
  WARNING: The scripts pylupdate5, pyrcc5 and pyuic5 are installed in '/home/ya
re/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this war
n
```

Imagen 16. Instalación Librería para la interfaz gráfica.

III.IX. Panorama general de la implementación.

Para su implementación lo que se realizó, fue seguir las siguientes tareas de manera sintetizada.

1. Importar los métodos para trabajar con NLTK.
2. Importar stopwords para poder checar posteriormente las palabras reservadas.
3. Descargar los métodos para palabras reservadas y tokenizar.
4. Configurar el método de entrada para el idioma requerido.
5. Definir el texto que queremos resumir.
6. Tener un arreglo para almacenar los tokens del texto.
7. Crear una tabla para las frecuencias de las palabras.
8. Recorrer el texto y añadir las palabras con su frecuencia a la tabla de frecuencia y actualizarla si es necesario.
9. Mostrar la tabla de frecuencias.
10. Crear un diccionario para almacenar los valores de las oraciones.
11. Recorrer las oraciones en el texto y verificar si está en el diccionario y tabla de frecuencias y sumar puntos a la palabra.
12. Mostrar el valor por oración.
13. Evaluar las oraciones de acuerdo con el diccionario.
14. Obtener el valor promedio de la palabra de acuerdo con el texto original.
15. Añadir el texto a la variable de resumen.
16. Recorrer las oraciones almacenadas y si la oración tiene valor y su posición es mayor que 1.2 veces el promedio añadir al resumen.
17. Imprimir resumen.

III.X. Presentación de resultados.

Los primeros resultados aquí mostrados corresponden a los resultados de un resumen obtenido desde la terminal, como podemos observar aún no se muestra una interfaz gráfica para el usuario, pero estos resultados fueron los primeros que obtuvimos.

•Versión en español.

-Resumen.

```
*****RESUMEN*****

[1]Coloquialmente, el término "inteligencia artificial" se usa a menudo para d
escribir máquinas (o computadoras) que imitan funciones "cognitivas" que los hu
manos asocian con la mente humana , como "aprendizaje" y "resolución de problem
as". [6] Las capacidades modernas de la máquina generalmente clasificadas como
IA incluyen la comprensión exitosa del habla humana , [7] competir al más alto
nivel en los sistemas de juego estratégico (como el ajedrez y el Go)), [8] vehí
culos autónomos , enrutamiento inteligente en redes de entrega de contenido y s
imulaciones militares.La inteligencia artificial fue fundada como una disciplin
a académica en 1955, y en los años transcurridos desde entonces ha experimentad
o varias oleadas de optimismo, [9] [10] seguido por la decepción y la pérdida d
e fondos (conocido como un " invierno AI "), [11] [12] seguido de nuevos enfoqu
es, éxito y financiación renovada. [14] Estos subcampos se basan en consideraci
ones técnicas, como objetivos particulares (por ejemplo, " robótica " o " apren
dizaje automático "), [15] el uso de herramientas particulares (" lógica"o rede
s neuronales artificiales ), o profundas diferencias filosóficas.
```

Imagen 17. Resumen en versión español.

•Ingles.

Podemos implementar distintos lenguajes siempre y cuando estos se encuentren dentro del stopwords.

-Resumen.

```
*****RESUMEN*****

[1] Colloquially, the term "artificial intelligence" is often used to describe
machines (or computers) that mimic "cognitive" functions that humans associate
with the human mind, such as "learning" and "problem solving." [6] Modern mach
ine capabilities generally classified as AI include successful understanding of
human speech, [7] competing at the highest level in strategic game systems (su
ch as chess and Go)), [8] autonomous vehicles , intelligent routing in content
delivery networks and military simulations. Artificial intelligence was founded
as an academic discipline in 1955, and in the years since then has experienced
several waves of optimism, [9] [10] followed by disappointment and loss of fun
ds (known as an "AI winter"), [11] [12] followed by new approaches, success, an
d renewed funding. [14] These subfields are based on technical considerations,
such as particular objectives (eg "robotics" or "machine learning"), [15] the u
se of particular tools ("logic" or artificial neural networks), or profound phi
losophical differences .
```

Imagen 18. . Resumen (inglés).

Resultados de la versión 2.

•Menú.

Para probar el proyecto, lo que se realiza es direccionarse dentro de la terminal en donde se encuentra guardado el proyecto y ejecutarlo para que nos habrá un menú en donde se podrá realizar el resumen.

```
yare@yare-VirtualBox:~/Documentos/AnálisisYProcesamientoDeTextosProyectoFinal-master/ProjectPython/version2/Interfaz/VentanaPrincipal$ python3 VentanaMenu.py
```

Imagen 19. Ejecución del programa.

Que nos genera la siguiente ventana.



Imagen 20. Menú principal

Al seleccionar el resumir texto podemos llegar a una nueva ventana donde podremos introducir nuestro texto a resumir. Como se muestra a continuación.

Texto desde el ordenador.

Una vez instalada nos debemos mover a la ruta donde se encuentra nuestra interfaz y ejecutar la ventana principal cuya extensión es .py. Como se muestra en la imagen siguiente. Esta es una manera de poder visualizarla o bien nos podemos dirigir a este punto desde la ejecución del menú mostrada en el punto anterior.

```
yare@yare-VirtualBox:~/Documentos/AnálisisYProcesamientoDeTextosProyectoFinal-master/ProjectPython/version2/Interfaz/VentanaPrincipal$ python3 VentanaPrincipal.py
```

Imagen 21. Compilando y ejecutando el programa.

Con lo que obtendremos un resultado como el que se muestra a continuación, en la ventana introduciremos el texto a resumir y al presionar el botón de resumir obtendremos una tabla de frecuencias, oraciones valorizadas y el resumen. Así mismo veremos una ventana emergente que nos indicará que nuestro resumen está listo.

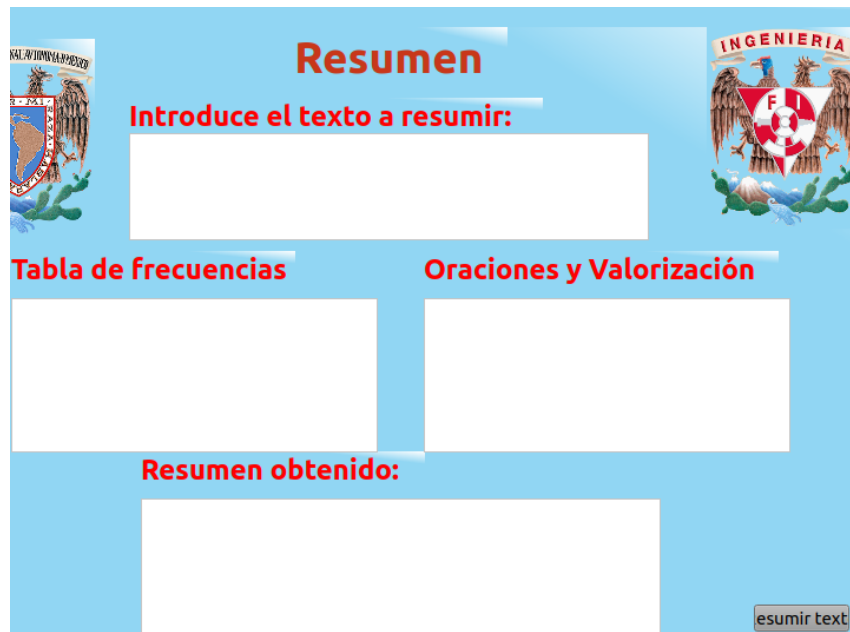


Imagen 22. Ventana Menú

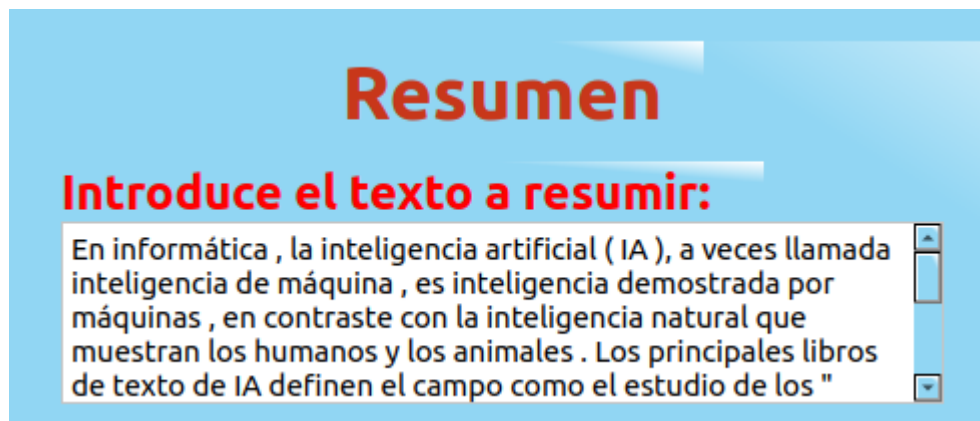


Imagen 23. Square para introducir el texto.

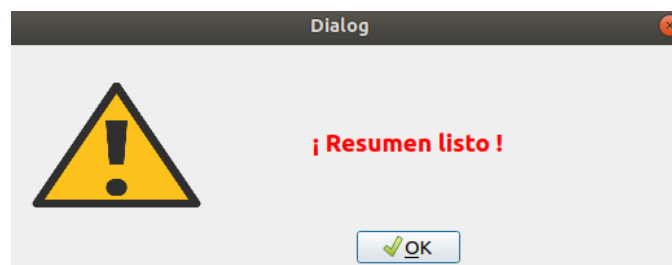


Imagen 24. Dialogo que indica que el resumen se ha realizado.

Al presionar ok podremos apreciar los campos ya mencionados.



informática:	1
.: 22	
inteligencia:	7
artificial:	3
(:	7
ia:	6
):	8

basado:	1
factores:	1
sociales:	1
instituciones:	1
trabajo:	1
investigadores:	1

Imagen 25. Tabla de frecuencias (Interfaz).

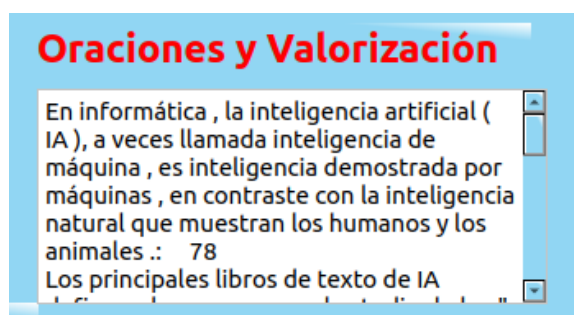


Imagen 26. Oraciones y valorización (Interfaz).

Resumen obtenido.

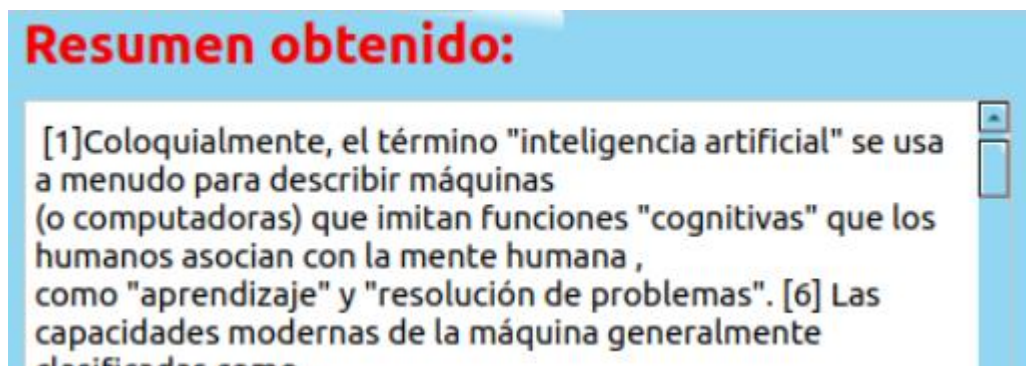


Imagen 27. Resumen Obtenido (Interfaz).

Para un resumen en inglés los resultados son los siguientes.

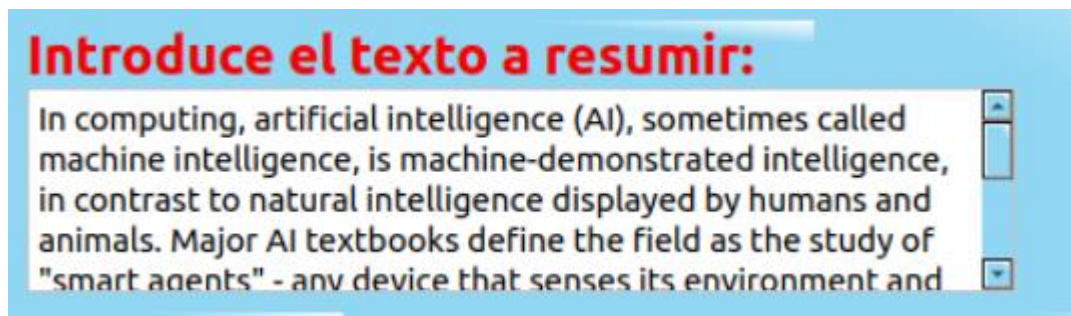


Imagen 28. Cuadro para ingresar el texto en inglés (Interfaz).

Del cual obtenemos su tabla de frecuencias, oraciones valorizadas y el resumen en general.

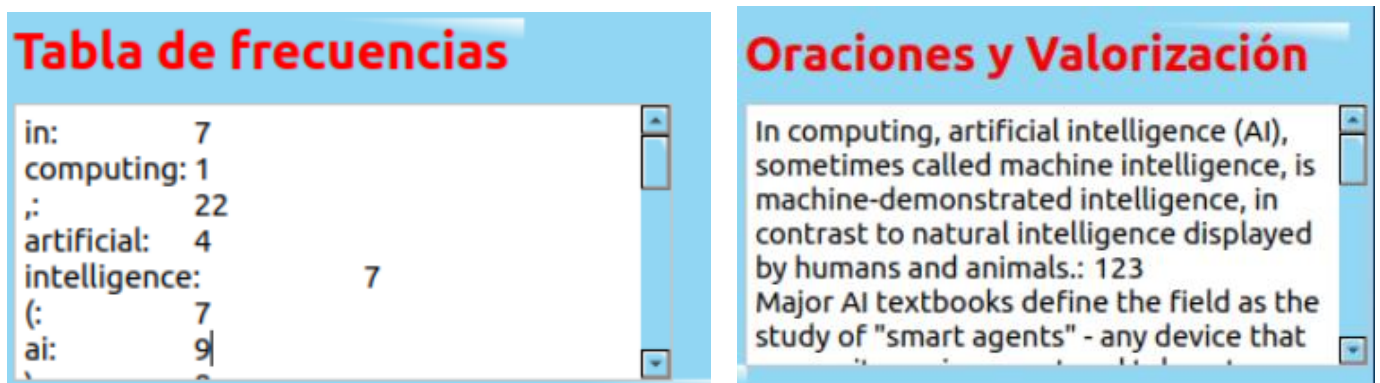


Imagen 29. Tabla de frecuencias y oraciones con valorización (Interfaz).

Obtención del resumen.

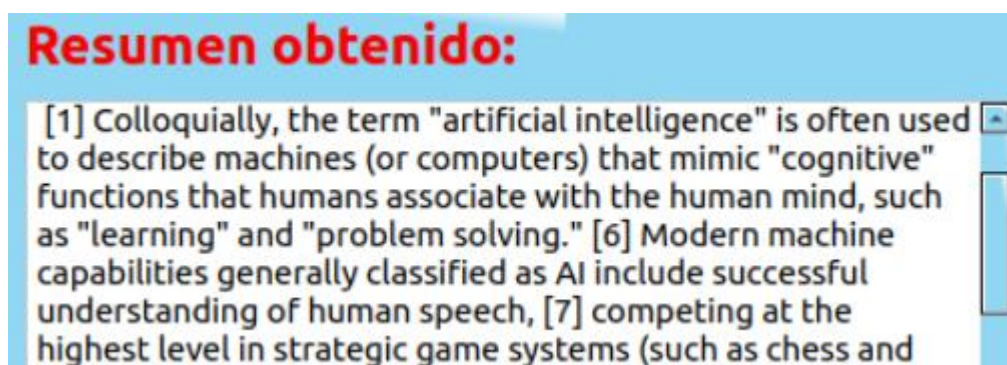


Imagen 30. Resumen obtenido en inglés (Interfaz).

Resultados de la versión 3.

Una vez avanzado el proyecto se visualizaron las áreas de mejora de este y por tal razón para esta versión del proyecto

se implementaron botones adicionales como el de limpiar cuadro ubicado en la parte inferior derecha, el cual permite eliminar el contenido de forma automática.

Una visión general de esta versión es la que se muestra en la imagen siguiente. Sigue contando con las ventanas de la versión 2 además de la ventana emergente que indica cuando se ha resumido un texto.

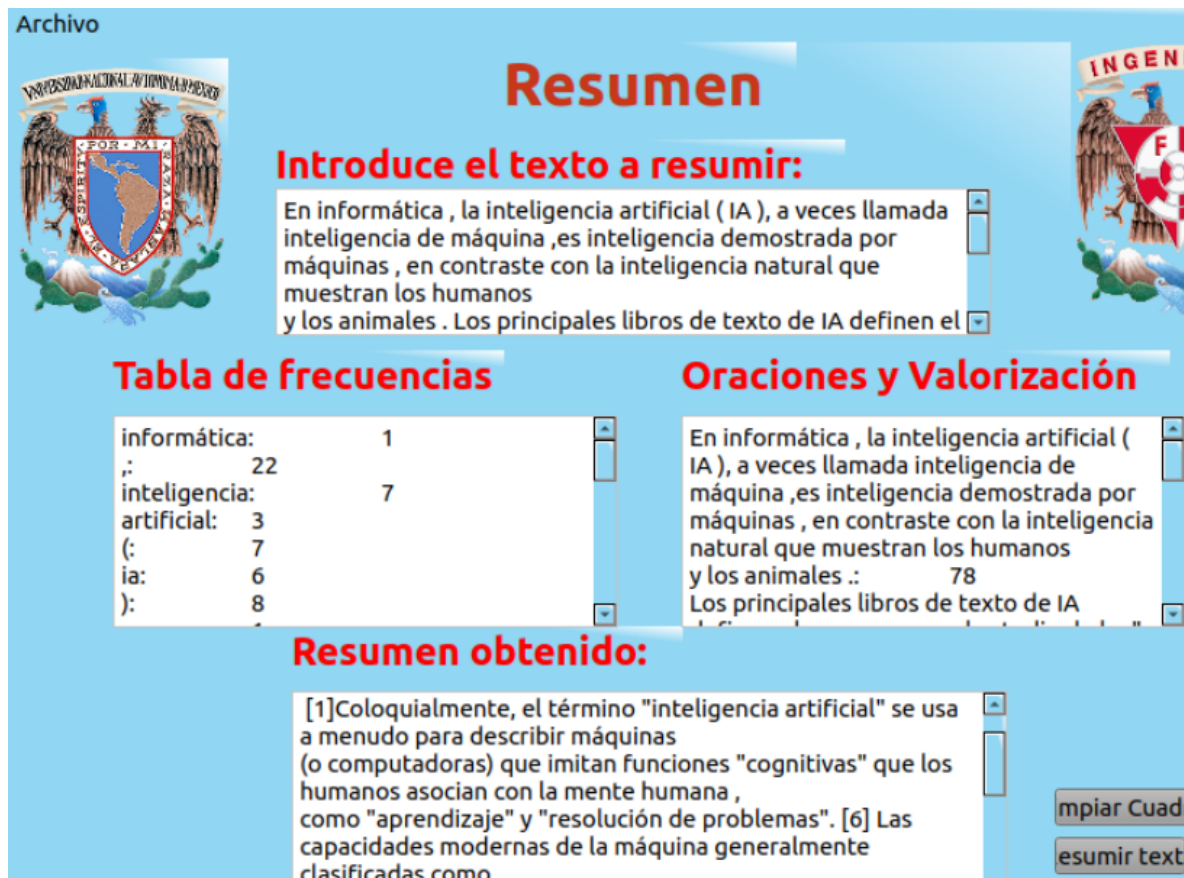


Imagen 31. Menú con botón para limpiar e introducir otra vez (Interfaz)

Mejoras.

Además, nos brinda la posibilidad de poder guardar el resumen en archivo .txt, para lo cual nos debemos colocar en la parte superior izquierda presionar Archivo y Guardar en archivo, lo que nos generara en nuestra carpeta local tener el resumen.

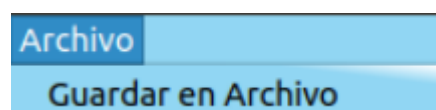


Imagen 32. Guardar el resumen (Interfaz).

Al ser guardado nos aparecerá una ventana como la siguiente.



Imagen 33. Resumen guardado (Interfaz).

Que nos indicará que se ha guardado satisfactoriamente, cada vez que demos clic en Guardar Archivo se creará un archivo que contendrá nuestro resumen, por lo que hay que tener cuidado de no guardarlo muchas veces.

Además, nos ofrece la implementación de poder obtener un texto resumido a través de la URL que introduzcamos. En una nueva ventana se puede introducir la URL a partir de la cual se desea hacer el resumen.



Imagen 34. Ventana para insertar URL.

De la URL deseada, se escribe en el recuadro correspondiente y posterior a dar clic en "Resumir texto", se obtiene una salida similar a la imagen 31.

Resumen

Introduce un URL con texto:

Tabla de frecuencias

procesamiento:	15
lenguaje:	34
natural:	17
language:	2
processing:	2
sas:	12
industria banca comunicaciones educación	

Oraciones y Valorización

1 Programación 2 SAS Enterprise Guide
 1 Programación de SAS para usuarios de R
 Habilitación de SAS Viya Todos los
 cursos Formatos de aprendizaje e-
 Learning Web en vivo Ubicación de las
 aulas Tutoriales en vídeo Todos los
 formatos de

Resumen obtenido:

Procesamiento del lenguaje natural
 Qué es y por qué es importante

El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) es una rama de la que ayuda a las computadoras a entender, interpretar y manipular el lenguaje humano. Cómo se utiliza

Métodos

Cargar Cuadr
 Resumir text
 Salir

Imagen 35. Resumen a partir de una URL.

Es importante mencionar que se ha omitido todo el texto que hace referencia a otras páginas web con el fin de evitar palabras que redirigen a otras páginas web, como redes sociales, o algunos recursos dentro del mismo sitio. Considerando esto, se han perdido algunas palabras que podrían aportar al resumen. Sin embargo, todo esto depende de la forma en que se ha programado dicha página web.

IV. Conclusión

Discusión de resultados

Para darle una valoración a los resúmenes generados en nuestro proyecto la forma más común sería compararlo con los resúmenes realizados por un humano. Para ello existen varias técnicas para conocer si el resumen que se obtuvo es coherente o lo más coherente posible, una de ellas es la evaluación intrínseca y extrínseca. La intrínseca prueba el sistema de resumen en sí mismo, y la extrínseca prueba el resumen basado en la forma que se afectaría la realización de alguna otra tarea.

Las evaluaciones intrínsecas evalúan la coherencia y la informatividad de los resúmenes.

Las evaluaciones extrínsecas ponen a prueba el impacto del resumen de tareas como la evaluación de la pertinencia y la comprensión de lectura.

Existen otros métodos como los son los intertextuales, estos evalúan la salida de un resumen en específico. Los intratextuales centran el análisis comparativo con varios sistemas de integración.

Hay infinidad de métodos por los cuales podemos saber si nuestro proyecto es correcto, pero elaborar un resumen en muy subjetivo, ya que todas las personas realizamos los resúmenes de diferentes maneras, por ejemplo, no todos le damos la misma importancia a las cosas, algunas personas pueden considerar más valiosa cierta información que otra. Por lo tanto, si le damos la valoración comparando el resumen automático con el de un humano, seguro existirán semejanzas, pero de igual manera diferencias.

Sabiendo esto consideramos que nuestro proyecto lo que realiza es una valuación intrínseca ya que sólo se evalúan la coherencia y la informatividad de los resúmenes.

Coherencia: Dado que nuestro método para generar el resumen se enfoca en la extracción de la información, donde se extrae la información sin modificarla y se utilizan frases clave, donde el objeto es seleccionar frases enteras, para crear un breve resumen de párrafo. Puede que al tomar una frase se saque del contexto del texto y es posible que este deje de tener sentido en el lugar donde se introduzca, por lo tanto, se debe de verificar que exista coherencia.

Informativo: Calcular el nivel de retención para dar un resultado de cuanta información de la fuente mantiene el resumen.

Similitud del contenido: Comparar la similitud del resumen generado con el resumen de referencia.

Tomando en cuenta un texto llamado "Texto original -Resumen Proceso Administrativo", que se dejó como tarea realizar a un alumno del equipo, se consideró que el resumen generado fue correcto, ya que hubo coherencia en el resultado comparándolo con el texto original.



MainWindow

Archivo

Resumen

Introduce el texto a resumir:

Unidad 8. Proceso administrativo 8.1. El proceso administrativo 8.1.1. Concepto y características La administración, como ciencia, exige en su desarrollo teórico y práctico un estilo de pensamiento sistematizado, con bases sólidas en la investigación y en la elaboración de sus teorías, que rompa con sistemas rutinarios e improvisaciones por parte de quien la practica. Éstas son algunas de las características de la administración que da como resultado el uso de la herramienta más importante que pudo haber adoptado el administrador en su práctica: el método científico.

Tabla de frecuencias

unidad:	4
proceso:	17
administrativo:	14
concepto:	2
características:	3
administración:	6
ciencia:	1
exige:	3
desarrollo:	2
teórico:	1

Oraciones y Valorización

Unidad 8.: 6
Proceso administrativo 8.1.: 33
El proceso administrativo 8.1.1.: 33
Concepto y características La administración, como ciencia, exige en su desarrollo teórico y práctico un estilo de pensamiento sistematizado, con bases sólidas en la investigación y en la elaboración de sus teorías, que rompa con sistemas rutinarios e improvisaciones por parte de quien la practica.: 43
Éstas son algunas de las características de la...

Resumen obtenido:

Éstas son algunas de las características de la administración que da como resultado el uso de la herramienta más importante que pudo haber adoptado el administrador en su práctica: el método científico, y con ello, la creación de una verdadera estructura de trabajo adecuada a su particular forma de ser: el proceso administrativo. El proceso administrativo ha sido un modelo a seguir durante generaciones, el cual fue desarrollado y utilizado a finales del siglo XIX y principios del XX por Henry Fayol, y a partir de ese momento se le ha identificado como la estructura básica de la práctica administrativa, otorgándole a ésta una capacidad de abstracción más amplia y la posibilidad de generar conceptos teóricos cada vez más particularizados a las

Limpiar Cuadros
Resumir texto
Salir

Imagen 36. Texto en español- Proceso administrativo (Interfaz)

Unidad 8. Proceso administrativo

8.1. El proceso administrativo

8.1.1. Concepto y características

La administración, como ciencia, exige en su desarrollo teórico y práctico un estilo de pensamiento sistematizado, con bases sólidas en la investigación y en la elaboración de sus teorías, que rompa con sistemas rutinarios e improvisaciones por parte de quien la practica. Éstas son algunas de las características de la administración que da como resultado el uso de la herramienta más importante que pudo haber adoptado el administrador en su práctica: el método científico, y con ello, la creación de una verdadera estructura de trabajo adecuada a su particular forma de ser: el proceso administrativo.

El proceso administrativo ha sido un modelo a seguir durante generaciones, el cual fue desarrollado y utilizado a finales del siglo XIX y principios del XX por Henry Fayol, y a partir de ese momento se le ha identificado como la estructura básica de la práctica administrativa, otorgándole a ésta una capacidad de abstracción más amplia y la posibilidad de generar conceptos teóricos cada vez más particularizados a las necesidades de las organizaciones, siendo también un modelo con el cual se estandariza la función del administrador.

El proceso administrativo se define como una consecución de fases o etapas a través de las cuales se lleva a cabo la práctica administrativa. Algunos autores que estudian dicho proceso lo han dividido de acuerdo con su criterio en tres, cuatro, cinco o seis etapas, pero sólo se refieren al grado de análisis del proceso, ya que el

9 x 279.4 mm

lo es el mismo que manejan todos. Por eso, en la actualidad, la división

Imagen 37. Texto Original- "Proceso administrativo"

Resumen Automático “Proceso Administrativo”.

Éstas son algunas de las características de la administración que da como resultado el uso de la herramienta más importante que pudo haber adoptado el administrador en su práctica: el método científico, y con ello, la creación de una verdadera estructura de trabajo adecuada a su particular forma de ser: el proceso administrativo.

El proceso administrativo ha sido un modelo a seguir durante generaciones, el cual fue desarrollado y utilizado a finales del siglo XIX y principios del XX por Henry Fayol, y a partir de ese momento se le ha identificado como la estructura básica de la práctica administrativa, otorgándole a ésta una capacidad de abstracción más amplia y la posibilidad de generar conceptos teóricos cada vez más particularizados a las necesidades de las organizaciones, siendo también un modelo con el cual se estandariza la función del administrador.

El proceso administrativo se define como una consecución de fases o etapas a través de las cuales se lleva a cabo la práctica administrativa. Considerando la importancia de este instrumento, mencionaremos algunas características del proceso administrativo:

a) El proceso administrativo está formado por un conjunto de fases o pasos de manera cíclica, es decir, el fin se convierte de nuevo en el principio de la tarea administrativa.

b) Una de las principales observaciones es que en cualquier modelo que analicemos del proceso administrativo, las fases que invariablemente encontraremos son las de planeación y control.

Naturaleza

La naturaleza del proceso administrativo se encuentra en su razón de ser, es decir, el proceso nace como una necesidad de explicar y estructurar toda una función administrativa que tiene como objetivo coordinar las demás funciones de la organización.

La disciplina que exige se encuentra en los principios que lo regulan; el orden se manifiesta en la consecución de las fases, nos organiza, nos enseña a pensar y trabajar en un estilo lógico y sistemático en el proceso de decisiones; la flexibilidad, siendo una característica natural de la administración, crea un modelo manejable de trabajo, para amoldarse a las necesidades, cultura organizacional y estilos gerenciales de la organización de que se trate.

La eficiencia del proceso administrativo no está proyectada exclusivamente para los aspectos de sistemas, procedimientos y maquinaria de una organización. Principios Terry y Franklin, en su libro Principios de administración nos definen el concepto de principio: “Se puede definir un principio

Imagen 38. Resumen- “Proceso administrativo”- Por humano.

Probando nuestro proyecto con el método de comparación, lo realizamos con la de un humano. Desde nuestro punto de vista consideramos que esté método no es el óptimo ya que existen muchas diferencias ya que todos pensamos de forma diferente.

Comparando nuestro resultado del resumen automático que generamos con el resumen elaborado por un humano, observamos ciertas diferencias. Cabe destacar que también esta fue una tarea de la vida real, que se dejó a un miembro del equipo en la materia de Administración de Redes.

A continuación, se mostrarán. Los textos se encuentran en la carpeta del Proyecto con el nombre “Resumen-Proceso administrativo-Prueba”, ahí está el texto original, el resumen automático y el resumen elaborado por el humano.

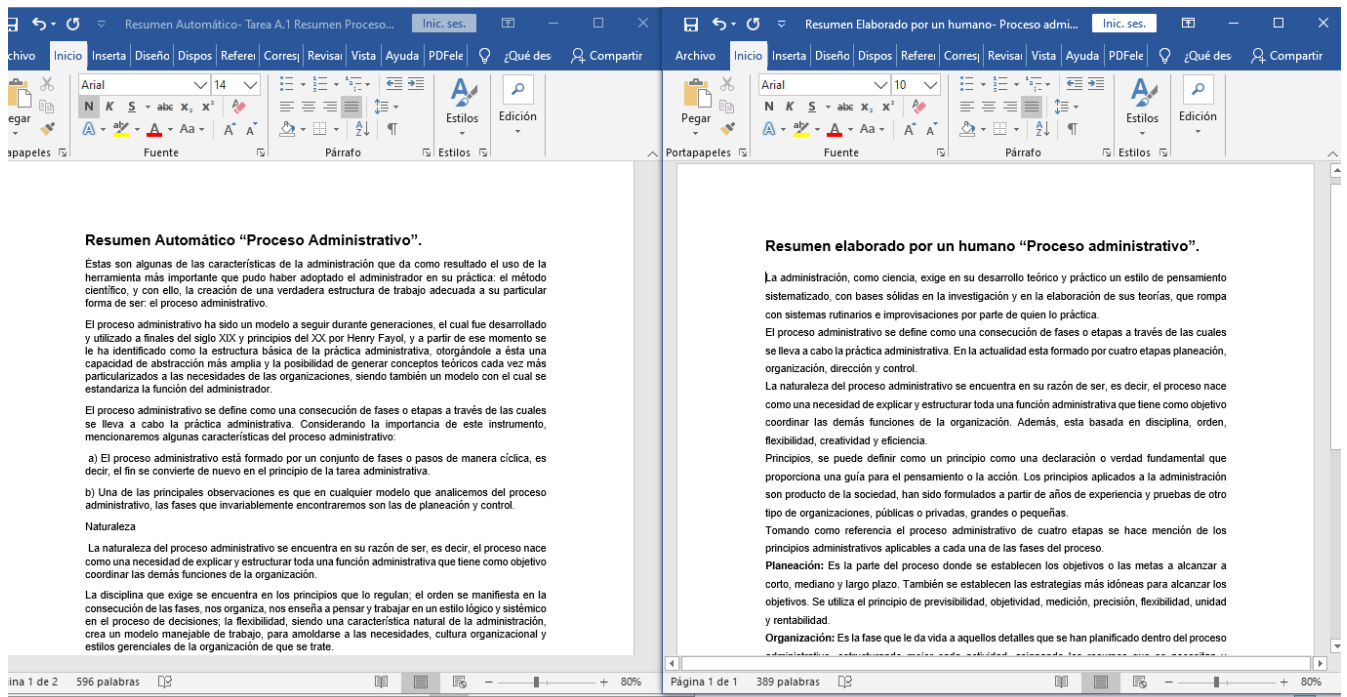


Imagen 39. Comparación de Textos Automático VS Elaborado por humano.

En la imagen que se encuentra en la parte de arriba se puede observar algunas de las diferencias y similitudes.

Una similitud que se encontró es que algunos de los conceptos mencionados en la lectura tienen coherencia y se parece mucho a lo que el humano también realizó.

Resumen automático

El proceso administrativo se define como una consecución de fases o etapas a través de las cuales se lleva a cabo la práctica administrativa. Considerando la importancia de este instrumento, mencionaremos algunas características del proceso administrativo:

Imagen 40. Resumen automático.

La naturaleza del proceso administrativo se encuentra en su razón de ser, es decir, el proceso nace como una necesidad de explicar y estructurar toda una función administrativa que tiene como objetivo coordinar las demás funciones de la organización.

Imagen 41. Continuación.

Resumen elaborado por un humano

El proceso administrativo se define como una consecución de fases o etapas a través de las cuales se lleva a cabo la práctica administrativa. En la actualidad esta formado por cuatro etapas planeación, organización, dirección y control.

Imagen 42. Resumen elaborado por humano.

La naturaleza del proceso administrativo se encuentra en su razón de ser, es decir, el proceso nace como una necesidad de explicar y estructurar toda una función administrativa que tiene como objetivo coordinar las demás funciones de la organización. Además, esta basada en disciplina, orden, flexibilidad, creatividad y eficiencia.

Imagen 43. Continuación resumen elaborado por humano.

De las diferencias que notamos es que no toma en cuenta los subtítulos o palabras remarcadas.

¿Qué resultados se obtuvieron de aplicar el mismo texto a varias personas?

Los resultados aquí presentados fueron obtenidos a través del hecho de solo pedir a compañeros realizar un resumen de un texto dado, sin que ellos conocieran la funcionalidad o posible resultado de nuestro generador de resumen.

-Resumen automático.

Como bien podemos recordar trabajamos a lo largo de nuestro trabajo con un texto que trata de inteligencia interfacial el cual se muestra en la siguiente imagen.

```
# text para definir el texto a resumir.
text = ""En informática , la inteligencia artificial ( IA ), a veces llamada
inteligencia de máquina , es inteligencia demostrada por máquinas , en
contraste con la inteligencia natural que muestran los humanos y los animales .
Los principales libros de texto de IA definen el campo como el estudio de los "
agentes inteligentes ": cualquier dispositivo que perciba su entorno y tome
medidas que maximicen sus posibilidades de lograr con éxito sus objetivos.
[1]Coloquialmente, el término "inteligencia artificial" se usa a menudo para
describir máquinas (o computadoras) que imitan funciones "cognitivas" que los
humanos asocian con la mente humana , como "aprendizaje" y "resolución de
problemas". A medida que las máquinas se vuelven cada vez más capaces, las
tareas que se consideran que requieren "inteligencia" a menudo se eliminan de
la definición de IA, un fenómeno conocido como efecto de IA . [3] Una broma en
el Teorema de Tesler dice que "AI es lo que no se ha hecho todavía". [4] Por
ejemplo, el reconocimiento óptico de caracteres a menudo se excluye de las
cosas consideradas como AI [5] , convirtiéndose en una tecnología de rutina.
```

Imagen 44. Texto para resumir.

El hecho interesante aquí es que contamos con un texto relativamente corto del cual obtendremos un resumen de forma automática con nuestro generador de resúmenes y varios resúmenes que hemos obtenido de diversos estudiantes universitarios que son nuestros amigos de diferentes carreras (no solo ingeniería) para comparar las similitudes y diferencias de los resultados finales.

-Resumen automático.

El resumen obtenido por el generador de textos es el siguiente.

```
*****RESUMEN*****

[1]Coloquialmente, el término "inteligencia artificial" se usa a menudo para d
escribir máquinas (o computadoras) que imitan funciones "cognitivas" que los hu
manos asocian con la mente humana , como "aprendizaje" y "resolución de problem
as". [6] Las capacidades modernas de la máquina generalmente clasificadas como
IA incluyen la comprensión exitosa del habla humana , [7] competir al más alto
nivel en los sistemas de juego estratégico (como el ajedrez y el Go)), [8] vehí
culos autónomos , enrutamiento inteligente en redes de entrega de contenido y s
imulaciones militares.La inteligencia artificial fue fundada como una disciplin
a académica en 1955, y en los años transcurridos desde entonces ha experimentad
o varias oleadas de optimismo, [9] [10] seguido por la decepción y la pérdida d
e fondos (conocido como un " invierno AI "), [11] [12] seguido de nuevos enfoqu
es, éxito y financiación renovada. [14] Estos subcampos se basan en consideraci
ones técnicas, como objetivos particulares (por ejemplo, " robótica " o " apren
dizaje automático "), [15] el uso de herramientas particulares (" lógica"o rede
s neuronales artificiales ), o profundas diferencias filosóficas.
```

Imagen 45. Resumen automático(Generador de resumen).

Posteriormente procedimos a revisar uno a uno los resultados que nuestros compañeros nos enviaron, a continuación, mostramos algunos de ellos, que son los más parecidos a nuestro resumen generado de forma automática.

La inteligencia artificial es un campo de la informática, fundada como disciplina académica en 1955, la cual ha ido variando con el tiempo, de tal forma que se han dividido subcampos que se relacionan entre sí, basados en técnicas o factores sociales.

De esta manera, la inteligencia artificial (IA) se puede definir como la inteligencia demostrada por maquinas.

Aunque en diversos libros de texto, es definida como el estudio de los agentes inteligentes, es decir, a partir de la percepción del entorno que realice un dispositivo se tomaran decisiones que aumentaran la posibilidad de lograr un objetivo.

Ahora bien, coloquialmente es definida para describir computadoras que imitan funciones cognitivas asociadas al aprendizaje y resolución de problemas.

Con el paso del tiempo, la capacidad de las computadoras ha ido aumentando, lo que ha provocado que algunas tareas ya no requieran ser resueltas con IA, lo anterior es conocido como efecto IA.

Dicho lo anterior, se asocia a la broma en el Teorema de Tesler, la cual dice que la IA es lo que no se ha hecho todavía.

Actualmente, las capacidad de las computadoras clasificadas como IA incluye comprensión del habla humana, competición de alto nivel en juegos estratégicos, vehículos autónomos, enrutamiento inteligente en redes para entrega de contenido y simulaciones militares.

Imagen 46.Resumen participante 1.

A partir del resumen mostrado en la imagen 44 podemos situar a nuestro generador de resumen en una buena escala, ya que, si bien este participante usa palabras propias para la generación de dicho resumen, sigue resaltando las mismas ideas que son extraídas de nuestro generador de resumen, como el año en que creo la IA, los retos, avances y repercusiones.

Con lo que nos damos cuenta de que este participante ha tomado en cuenta las ideas principales del texto de prueba, además el equipo considera a este participante como alguien que sabe hacer resúmenes, dado que las ideas son muy similares podríamos puntuar a nuestro generador de resúmenes con una buena puntuación.

En informática la inteligencia artificial es una inteligencia demostrada por máquinas a diferencia de la natural. La inteligencia artificial se usa para describir máquinas que imitan las funciones “cognitivas” que se asocian con la mente humana como “aprendizaje” y “resolución de problemas”. Cada vez que las máquinas se vuelven más capaces se eliminan de la definición de IA, fenómeno conocido como efecto de la IA.

Las capacidades modernas clasificadas como IA incluyen la comprensión exitosa del habla humana, vehículos autónomos, entre otras. La IA fue fundada en 1955 desde entonces ha experimentado optimismo, decepción y pérdida de fondos (invierno AI), nuevos enfoques, éxito y financiación renovada.

Los subcampos en que se divide la IA se basan en objetos particulares como la robótica, redes neuronales o en factores sociales como instituciones particulares o trabajo de investigadores particulares. |

Imagen 47. Resumen participante 2.

De igual manera para este caso podemos notar que incluye palabras similares a las del resumen realizado por medio del generador de textos. Lo que nos da una buena puntuación aceptable en nuestro trabajo, ya que consideramos que el participante 2 es bueno realizando resúmenes, a pesar de que las ideas están estructuradas de diferente forma podemos concluir que esto se debe a que la persona ha priorizado de distinta forma las ideas, con las que ha generado el resumen.

La inteligencia artificial es inteligencia demostrada por máquinas, comparada con la inteligencia que mostramos los humanos y los animales. Libros de IA han definido el campo como el estudio de los "agentes inteligentes" de esta manera cualquier dispositivo que perciba su entorno y tome medidas que maximicen sus posibilidades de lograr con éxito sus objetivos. El término "inteligencia artificial" se usa para describir máquinas que imitan funciones que los humanos asocian con la mente humana, como ejemplo el aprendizaje y la resolución de problemas. A su vez las máquinas se han vuelto más veloces tanto que las tareas que se consideran difíciles y que necesite inteligencia se han eliminado de la IA. Por esto mismo en el Teorema de Tesler se dice que la IA es lo que no se ha hecho todavía. Otro ejemplo sería el reconocimiento óptico de caracteres a menudo se excluye de las cosas consideradas como IA, convirtiéndose en una tecnología de rutina.

De este modo las capacidades modernas de la máquina de IA incluyen la comprensión del habla humana, competir al más alto nivel en los sistemas de juego estratégico, vehículos autónomos, enrutamiento inteligente en redes de entrega de contenido y simulaciones militares, etc. La inteligencia artificial fue fundada 1955, y en los años transcurridos desde entonces ha experimentado distintas fases, así como el optimismo, decepción, pérdida de fondos, seguido de nuevos enfoques, éxito y financiación renovada.

Durante su historia, la investigación de IA se ha dividido en subcampos.

Estos subcampos se basan en consideraciones técnicas, como objetivos particulares como la robótica, el uso de herramientas particulares como la lógica, o profundas diferencias filosóficas.

Imagen 48. Resumen participante 3.

Para este caso podemos notar que el participante ha generado un resumen mucho más parecido al del generador de resumen automático, debido a que incluye más conceptos, es decir, que no ha priorizado tanto unos conceptos sobre otros. Este resumen contribuiría a justificar que el generador realizado se aproxima al que los humanos llegamos a hacer.

En informática la inteligencia artificial es aquella demostrada por máquinas a diferencia de la natural, algunos libros la definen como: cualquier dispositivo que perciba su entorno y tome medidas que maximicen sus posibilidades de lograr con éxito sus objetivos. La inteligencia artificial se usa para describir máquinas que imitan las funciones "cognitivas" que se asocian con la mente humana como "aprendizaje" y "resolución de problemas". Cada vez que las máquinas se vuelven más capaces en la resolución y comprensión de problemas, se eliminan de la definición de IA, a éste fenómeno se le conoce como efecto de la IA.

Las capacidades modernas clasificadas como IA incluyen la comprensión exitosa del habla humana, estrategia, simulaciones militares, vehículos autónomos, entre otros. La IA fue fundada en 1955.

Los subcampos en que se encuentra dividida la IA se basan en objetos particulares como la robótica, redes neuronales o en factores sociales como instituciones particulares o trabajo de investigadores particulares.

Imagen 49. Resumen participante 4.

El resumen anterior es bastante parecido de igual forma al generado por el generador de resúmenes, si bien es más corto incluye la mayor parte de las ideas principales que se muestran en el resultado final del resumen automático.

La inteligencia artificial (IA), es inteligencia demostrada por máquinas , en contraste con la inteligencia natural que muestran los humanos y los animales . Los principales libros de texto de IA definen el campo como el estudio de los " agentes inteligentes ": cualquier dispositivo que perciba su entorno y tome medidas que maximicen sus posibilidades de lograr con éxito sus objetivos. El término "inteligencia artificial" se usa a menudo para describir máquinas que imitan funciones "cognitivas" que los humanos asocian con la mente humana , como "aprendizaje" y "resolución de problemas". A medida que las máquinas se vuelven cada vez más capaces, las tareas que se consideran que requieren "inteligencia" a menudo se eliminan de la definición de IA, un fenómeno conocido como efecto de IA . Una broma en el Teorema de Tesler dice que "AI es lo que no se ha hecho todavía". Por ejemplo, el reconocimiento óptico de caracteres a menudo se excluye de las cosas consideradas como AI , convirtiéndose en una tecnología de rutina. Las capacidades modernas de la máquina generalmente clasificadas como IA incluyen la comprensión exitosa del habla humana , competencias en juegos de estrategia (como el ajedrez y el Go)), vehículos autónomos , enrutamiento inteligente en redes de entrega de contenido y simulaciones militares. La inteligencia artificial fue fundada como una disciplina académica en 1955, y en los años transcurridos ha experimentado varias oleadas de optimismo, seguido por la decepción y la pérdida de fondos (conocido como un " invierno AI "), seguido de nuevos enfoques, éxito y financiación renovada. La investigación de IA se ha dividido en subcampos, estos subcampos se basan en consideraciones técnicas, como objetivos particulares (por ejemplo, " robótica " o " aprendizaje automático "), el uso de herramientas particulares (" lógica " o redes neuronales artificiales), o profundas diferencias filosóficas. Los subcampos también se han basado en factores sociales (instituciones particulares o el trabajo de investigadores particulares).

Imagen 50. Resumen participante 5.

Como bien lo podemos notar este es uno de los resúmenes generado por humanos que más se parece al realizado por nuestro programa, lo que nos daría un indicativo que esta persona ha seguido una serie de pasos para poder obtener la información más relevante del texto original, lo cual nos lleva a determinar que nuestro generador resulta útil y da resultados muy parecidos a los humanos.

La inteligencia artificial, también conocida como inteligencia de máquina se define como el campo de estudio de los “agentes inteligentes”, es decir, cualquier dispositivo que perciba su entorno y tome medidas que maximicen sus posibilidades de lograr con éxito sus objetivos. De manera coloquial se le hace referencia al término de “inteligencia artificial” para describir máquinas o computadores que son capaces de imitar funciones “cognitivas” que los humanos asocian con la mente humana, como “aprendizaje” y “resolución de problemas”. Un término importante de englobar es el concepto de “efecto de IA”, en el cual se menciona que mientras más capaces sean las máquinas de realizar tareas que requieren “inteligencia” suelen eliminarse de la definición de IA. Continuando con el concepto de máquina, estas incluyen características como: la comprensión exitosa del habla humana, competir al más alto nivel en los sistemas de juego estratégico (el ajedrez y el Go)), vehículos autónomos, enrutamiento inteligente en redes de entrega de contenido y simulaciones militares,

Fundada como una disciplina académica en 1955 hasta el día de hoy, la inteligencia artificial ha experimentado varias oleadas de optimismo, seguido por la decepción y la pérdida de fondos (invierno AI), seguido de enfoques, éxito y financiación renovada. A lo largo de la historia, la investigación de IA, se ha dividido en subcampos que a menudo se comunican entre sí. Dichos subcampos se basan en consideraciones técnicas como lo son los objetivos y uso de herramientas particulares o las profundas diferencias filosóficas, considerando su base en factores sociales.

Imagen 51. Resumen participante 6.

En informática se le denomina Inteligencia Artificial (IA) a la inteligencia demostrada por máquinas, siendo el contraste de la inteligencia natural que generan los humanos y animales.

Cualquier dispositivo que perciba su entorno y maximice sus posibilidades de lograr el éxito, se considera que tiene IA.

La capacidad moderna de las máquinas, imita funciones cognitivas que los humanos asocian como aprendizaje y resolución de problemas.

Con el paso del tiempo la IA se ha dividido en subcampos. Estos subcampos se basan en: objetivos particulares, el uso de herramientas o diferentes filosofías.

Imagen 52. Resumen participante 7.

En informática, la inteligencia artificial es aquella que demuestran poseer las máquinas, diferente a la que poseen los humanos y animales. A menudo el término de “inteligencia artificial” se usa para describir máquinas que imitan funciones que los humanos asocian con la mente humana, como resolver problemas o aprender. Debido a que las máquinas cada vez se vuelven más capaces, las tareas que se considera que requieren inteligencia se reducen. Las capacidades modernas clasificadas como inteligencia artificial incluyen la comprensión exitosa del habla humana, competir al mas alto nivel de juegos estratégicos, etc. La inteligencia artificial fue fundada como una disciplina académica en 1955, en los años transcurridos ha pasado por diferentes fases, como el conocido “invierno AI”. En la mayor parte de su historia la IA se ha dividido en subcampos, que se basan en objetivos particulares, como robótica, o también se han basado en factores sociales, como instituciones particulares o trabajo de investigadores particulares.

Imagen 53. Resumen participante 8.

En estas tres últimas imágenes podemos encontrar las mismas observaciones que ya se han comentado para los resúmenes anteriores, es decir, mismo empleo de palabras u oraciones más relevantes del texto e incluso misma secuencia en el acomodo de las ideas para generar un resultado final, lo que nos lleva a pensar que la mayoría de nuestros compañeros han priorizado las ideas de la misma manera que lo hace el generador de resúmenes de textos, sin embargo y a pesar de que usen en algunas ocasiones palabras propias o distintos conectores mantiene la misma estructura que hemos observado en el resultado del generador de resúmenes automático.

Podríamos puntuar de manera general de acuerdo con los resultados por compañeros a nuestro resumidor de textos como un resumidor de textos bueno, ya que nos ofrece un resultado muy o bastante parecido al que podría obtener una persona. Algunos de los resúmenes son más o menos parecidos en cuanto a la forma que se estructuran las ideas, en la extensión e incluso en palabras más usadas, lo que nos lleva a la siguiente pregunta ¿nuestros compañeros qué algoritmo utilizan para resolverlo?, como respuesta a ello nos comentaron que se enfocaron en el título para que, a partir de este saber que ideas deben obtener, en características que señalen del tema, objeto o personaje, fechas relacionadas con el tema, causas o consecuencias si es que lo amerita, cabe resaltar que nosotros los humanos poseemos la capacidad de poder hacer los resúmenes de acuerdo a nuestra experiencia y no por el análisis completo de como generarlo.

Con lo anterior podemos obtener el resultado siguiente, las librerías de Python nos permiten hacer una diferencia entre las palabras más o menos importantes como lo hace la mente de una persona, posteriormente el darle valor a las palabras y

oraciones será otro proceso normal de nuestro cerebro. Y la estructura final depende de su puntaje, esta vemos que difiere de manera no tan alta en los resúmenes obtenido por personas, lo cual podemos entender que difiere de lo que una persona priorice.

Descripción de observaciones

Lo que pudimos observar en la realización de los resúmenes automáticos es que se deben tener en cuenta ciertas variables, pero obviamente para realizarlo lo más completo, se necesita de más tiempo. Algunas de estas variables que consideramos son las siguientes: como la persona a la que va dirigido, que dispositivo lo realizará, el medio de donde se va a extraer (tomar el cuanto el tipo de datos, el idioma, etc.). Por lo que un resumen muchas veces se debe de acoplar a diferentes necesidades de acuerdo al tipo de persona y a la situación.

Fuente: Se debe de tomar en cuenta si la información se está tomando de una sola fuente o de más, ya que si se trata de más el grado de dificultad se complica al programarlo ya que se vuelve más compleja la búsqueda de la importancia de la información.

Idioma: El idioma de la fuente es importante ya que, en la elaboración del resumen, no siempre es posible encontrar una buena técnica para cualquier idioma.

Género: Es importante tomar en cuenta que se quiere resumir, si es una revista, un texto de una página web, un artículo o un libro, ya que si es un libro se debe de tomar en cuenta el título, el índice e incluso el propio resumen de la portada. Si estamos hablando de una página web tiene varios aspectos a considerar como lo son las etiquetas.

Tipo de resumen: Se debe de considerar que tipos de resumen podrá lograr realizar el programa de resúmenes automáticos. Ya que existen los informativos, descriptivos, abstractos, y de síntesis.

Tipos de dispositivos: Se debe de tomar en cuenta para que tipos de dispositivos estará disponible y para qué tipo de plataformas.

La parte en que se obtiene el resumen a partir de una página web resultó ser un punto sensible y difícil de manipular. Este resumen depende mucho de la forma en que se ha programado dicha página web, pues se omiten del texto a analizar palabras en las que se tienen enlaces a otros sitios web, como a publicidad, redes sociales entre otros, así como palabras en el menú del propio sitio.

Se tomó la decisión de omitir dichas palabras para evitar que palabras fuera del contexto del texto a analizar fueran consideradas dentro del resumen. Sin embargo, en algunas ocasiones se omiten palabras con valor dentro del resumen porque en ellas se encuentra un enlace a otro sitio donde se puede encontrar su definición. A pesar de esto, el resumen final no se vio alterado (en sentido negativo) de manera significativa por estas omisiones.

Trabajo futuro.

Uno de los problemas al que nos enfrentamos fue el realizar un resumen a partir de un texto en una página web, lo cual se logró de la mejor manera, sin embargo, consideramos que es necesario depurar la página aún más para que no nos lleguen a parecer etiquetas de la misma página, con el fin de dar un resumen más limpio al usuario.

El trabajo podría ser mejorado si se implementará una opción para poder jalar incluso documentos completos en versión PDF o Word o incluso otra extensión y que, a partir de ellos, una vez que el usuario elija el idioma de su texto original le pueda ser generado un documento totalmente resumido, lo que permitirá solo obtener un resumen completo de un texto con mayor extensión.

Otra oportunidad de mejora que vemos en el resumen desde una página web, es que se pudiera diferenciar las palabras que sí aportan valor al texto a resumir de las que no. Nosotros pudimos notar que, generalmente, las palabras que sí aportan valor se encuentran dentro de la misma etiqueta HTML que el párrafo que la contiene. Las que no aportan valor se encuentran en etiquetas distintas que generalmente contienen bastantes referencias a otros sitios web y poco texto “visible” en el sitio web.

Si bien logramos producir textos muy parecidos a los realizados por personas, en un futuro nos podríamos enfocar más en tratar de dejar al usuario elegir sobre las palabras que él considera

claves, con el fin de que se obtengan resultados más personalizados.

El análisis de textos inteligentes ha crecido a través de los años y con el avance que el ser humano ha tenido en diferentes ámbitos, sin embargo, hace falta mucho más desarrollo para poder alcanzar resultados lo más parecidos a los de un humano. Con lo cual concluimos que nuestra extensión del trabajo aquí presente correspondería a generar mejores resúmenes a partir de contenidos más extensos, brindar más opciones al usuario para que este pueda exportar e importar documentos sin importar su tipo, de tal forma que facilite totalmente la actividad de resumir por completo.

V. Referencias

<https://redhistoria.com/historia-y-evolucion-del-resumen-de-textos/>

<https://www.instintoprogramador.com.mx/2019/07/resumen-de-texto-con-nltk-en-python.html?m=1>

<https://code.tutsplus.com/es/tutorials/introducing-the-natural-language-toolkit-nltk--cms-28620>

<http://www.nltk.org/book/>

<https://likegeeks.com/es/tutorial-de-pyqt5/>

https://es.qwe.wiki/wiki/Automatic_summarization#Applications_and_systems_for_summarization

Herramientas y técnicas. Consultado el 25 de mayo en:

http://www.corpus.unam.mx/cursocorpus/4_1_Conteo_de_palabras.html