



Análisis Cualitativo de Textos en Español

Clave de registro del proyecto ante la Secretaría de Ciencia 7052/2024CIB

Integrantes del proyecto de investigación: Alma Delia Cuevas Rasgado (Responsable técnico), Yedid Erandini Niño Membrillo (corresponsable), Rafael Hernández Espinosa, Rosa María Rodríguez Aguilar, Esther Figueroa Hernández, Josué Padilla Cuevas, Gabriela Alejandra García Robledo, Gerardo Hernández Jiménez, Leonel, Jorge Luis, Eduardo, Pablo, Sergio Robledo Villagomez

Espacio Académico Universidad Autónoma Del Estado de México Centro universitario Texcoco

Introducción.

El Procesamiento de Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial (IA) que proviene de las ciencias de datos, consiste en procesar de manera automática los datos textuales para analizarlos, entenderlos y obtener información relevante de una manera eficiente e inteligente. Usando PLN, se puede analizar grandes cantidades de textos de manera automática que para un humano sería tedioso, consumiría tiempo, costoso o casi imposible. Por tal razón se requiere de un software de tipo Front-End para computadora personal, portátil y dispositivo móvil para el Análisis Cualitativo de Textos llamado AnaCual que procese información obtenida de encuestas, entrevistas sobre los quehaceres diarios, opiniones sobre productos, eventos para analizar sus opiniones, sentimientos, clasificar gustos, afinidades, extraer información útil a problemas sociales, económicos, contables, en fin. Este sistema AnaCual usará la metodología de ingeniería de software más apropiada, así como el ciclo de vida con métodos ágiles y herramientas de Procesamiento de Lenguaje Natural PLN como: creador de resúmenes, identificación de entidades nombradas, análisis de sentimientos muy útiles en el análisis de texto. Será una herramienta para la comunidad académica del área de Ciencias Sociales. Actualmente hay aplicaciones como MAXQDA que proporcionan estos beneficios, pero son productos con venta de licencias y el costo de estas es muy alto y está fuera del alcance económico de los estudiantes. Con la creación de AnaCual el proceso de enseñanza y aprendizaje de estos estudiantes se verá beneficiado. La aplicación será probada, por lo pronto, por los estudiantes del centro universitario UAEM Texcoco.

Método.

Descripción de las técnicas y procedimientos. Para el desarrollo del software es necesario establecer la directriz que guiaría el proceso a seguir en las actividades por realizar, a fin de cumplir con el objetivo ya planteado.

A continuación, se describen los pasos o fases planeados, así como la manera en que se llevarían a cabo:

1. Recolección de información. Se planeó el uso de dos técnicas de recolección de datos, la primera serán entrevistas al experto en análisis cualitativo para obtener las especificaciones que deberán considerarse para el análisis de textos cualitativos y que servirán para la creación del software. La segunda técnica de recolección de datos a utilizar sería la observación del manejo, por el experto, de las herramientas MAXQDA y AtlasTi para visualizar la funcionalidad de cada una de las herramientas con respecto a cómo se realiza el análisis de información para la investigación cualitativa. En este mismo sentido, la observación de las herramientas antes mencionadas se realizaría sin el experto para examinar las funcionalidades que ofrecen para comprender mejor la manera en que se realiza el análisis cualitativo de textos.

2. Análisis y validación de requerimientos. Considerando la recolección de información descrita en el punto anterior, se llevaría a cabo un análisis exhaustivo para identificar los requerimientos del software y se clasificarían en dos tipos, funcionales y no funcionales. Además, se identificarán los módulos y componentes por desarrollar y sus alcances, mismos que conformarían el software. Ya con los requerimientos clasificados, los módulos y el alcance, se realizarían la validación de los requerimientos



Reporte de Informe final de Proyectos de Investigación

Secretaría de Ciencia
Dirección de Investigación

con el experto, mediante sesiones denominadas Desarrollo Conjunto de Aplicaciones (JAD) en donde se reuniría el equipo de desarrollo con el experto para determinar los requerimientos, con la finalidad de asegurar que estos sean correctos, completos y claros con base en las especificaciones solicitadas.

3. Diseño del software. Se traducirían los requerimientos, módulos y componentes en diagramas que serían la base para la creación del software. Se realizarían los diseños de los datos, de la interfaz y de los módulos y componentes, que en su conjunto permitirían visualizar la organización y comprensión de la funcionalidad del software por desarrollar. Cabe destacar, que durante esta fase se deberían verificar los diagramas a fin se encontraran estandarizados y elaborados en la herramienta de diseño, la cual no se definió.

4. Implementación. Considerando los diferentes diagramas realizados en el diseño del software, se ejecutarían cada uno de ellos para poder codificar los módulos y componentes en el lenguaje de programación Python y utilizando técnicas y herramientas de IA para el PLN, como las bibliotecas. Asimismo, se crearían las interfaces gráficas de usuario dándole la funcionalidad a cada elemento que la conforman, se crearía la base de datos, tentativamente se utilizará MySQL.

Però nada de lo anterior se pudo culminar en su totalidad, solo unas partes que se exponen a continuación y fueron desarrollados con los recursos particulares de los estudiantes y profesores.

Resultados

"Análisis Cualitativo de documentos de texto en español", enlace en repositorio UAEMex: http://hdl.handle.net/20.500.11799/143755	Artículo
Ponencia Usos y Cuidados de la Inteligencia Artificial en la vida cotidiana	Ponencia
Usos y Cuidados de las Redes Sociales	Ponencia
MédicoBERT: A medical language model for Spanish NLP tasks with an application in question-answering using hyperparameter optimization.	Ponencia
Análisis cualitativo de textos en español, ICOFEST 2024	Ponencia
Análisis cualitativo de textos en español, Semana de la Ciencia	Ponencia
Métodos de IA para la informática forense, jornada académica de la Licenciatura en Informática Administrativa	Ponencia
Exposición de Proyectos de Inteligencia Artificial	Ponencia



Reporte de Informe final de Proyectos de Investigación

Secretaría de Ciencia
Dirección de Investigación

Retribución social

Robótica divertida para niñas y niños en el CIC IPN	taller
Usos y cuidados de las redes sociales, en escuela SECUNARIA OFIC. NO. 0468 ISIDRO FABELA	Comunicación social
Inteligencia artificial para niños, en Primaria de Colegio El Cedral	Resultados ante grupos

Corresponsable:

Dra. en I. S. Yedid Erandini Niño Membrillo (Lic. en Informática Administrativa), yeninom@uaemex.mx

Colaboradores Profesores:

1. Dr. en A. Rafael Hernández Espinosa (Lic. en Administración), rhernandez@uaemex.mx
2. Dra. en D. Rosa María Rodríguez Aguilar (Ing. En Diseño, UAE Nezahualcóyotl), dmrodriguez@uaemex.mx

Estudiantes de licenciatura:

AREA: Lic. en informática administrativa 8º semestre

1. Galván Peralta Jorge Luis, jlgalvanp@alumno.uaemex.mx
2. Hernández Vallarta Luis Eduardo, lehernandezv@uaemex.mx
3. Escalante Catorce Leonel, lescalantec@uaemex.mx
4. Hernández Jiménez Gerardo, ghernandezj@uaemex.mx
5. Sergio Robledo Villagómez, sroble dov@uaemex.mx

AREA: Lic. ingeniería en computación 8º semestre

1. Pablo Rafael Bravo Salazar

Estudiantes de Doctorado en ciencias de la computación:

- 1.- Josué Padilla Cuevas
2. Gabriela García Robledo

Conclusión:

- 1.- Módulo para el Análisis Cualitativo de Textos en español, desarrollado por Hernández Jiménez Gerardo, en registro de autor ante abogado general y artículo enviado a revista RIIIT, aceptado y publicado.
- 2.- Módulo para el identificador de entidades nombradas en español, desarrollado por Escalante Catorce Leonel, en preparación de artículo.
- 3.- Módulo para el análisis de sentimientos en español, desarrollado por Pablo Rafael Bravo Salazar, en preparación de artículo.
- 4.- Participación como ponente en Congreso internacional de Inteligencia Artificial y Machine Learning realizado en Tokio Japón, derivado de un módulo de la tesis doctoral Algoritmo de pregunta-respuesta a partir de textos clínicos de Josué Padilla Cuevas ya graduado de Doctor, se solicitó apoyo al Programa de Internacionalización pero la solicitud fue rechazada.
- 5.- Se desarrolló parte de un módulo de la tesis de Gabriela García Robledo el del reconocedor de entidades nombradas, ya graduada de Doctora.



Reporte de Informe final de Proyectos de Investigación

Secretaría de Ciencia
Dirección de Investigación

ANEXO

Acepto que el informe final del proyecto de investigación y los datos de contacto, sean publicados en la página de la Secretaría de Investigación y Estudios Avanzados; si es necesario autorizo se realicen las modificaciones de forma para su difusión.

Nombre de los integrantes

Firma de los Integrantes

Dra. En C. de la Comp. Alma Delia Cuevas Rasgado

Dra. En I.S. Yedid Erandini Niño Membrillo

Dra. en D. Rosa María Rodríguez Aguilar

Dra. en E. Esther Figueroa Hernández

Dr. en C. Comp. Josué Padilla Cuevas

Dra. en C. Comp. Gabriela García Robledo

L. I. Galván Peralta Jorge Luis

L. I. Hernández Vallarta Luis Eduardo

Escalante Catorce Leonel

Hernández Jiménez Gerardo

Pablo Rafael Bravo Salazar



Reporte de Informe final de Proyectos de Investigación

Secretaría de Ciencia
Dirección de Investigación

Rafael Hernández Espinosa