



ANEXO 1

FORMATO PARA LA PRESENTACIÓN DE PROYECTOS DE INVESTIGACIÓN CON EL FINANCIAMIENTO DEL FEDU

1. Título del proyecto

Identificación de lenguaje sesgado utilizando procesamiento de lenguaje natural para la lengua castellana

2. Área de Investigación

Área de investigación	Línea de Investigación	Disciplina OCDE
Ingeniería de Sistemas	Sistemas, computación e informática	Inteligencia Artificial

3. Duración del proyecto (meses)

12

4. Tipo de proyecto

Individual	☒
Multidisciplinario	☐
Director de tesis pregrado	☐

4. Datos de los integrantes del proyecto

Apellidos y Nombres	SOSA MAYDANA CARLOS BORIS
Escuela Profesional	INGENIERÍA DE SISTEMAS
Celular	986739074
Correo Electrónico	BORISSOSA@GMAIL.COM

- I. Título (El proyecto de tesis debe llevar un título que exprese en forma sintética su contenido, haciendo referencia en lo posible, al resultado final que se pretende lograr. Máx. palabras 25)

IDENTIFICACIÓN DE LENGUAJE SESGADO UTILIZANDO PROCESAMIENTO DE LENGUAJE NATURAL PARA LA LENGUA CASTELLANA

- II. Resumen del Proyecto de Tesis (Debe ser suficientemente informativo, presentando -igual que un trabajo científico- una descripción de los principales puntos que se abordarán, objetivos, metodología y resultados que se esperan)

El presente proyecto se enmarca dentro del desarrollo del procesamiento de lenguaje natural en lengua castellana, para identificar lenguaje sesgado que generan las personas en nuestra región, utilizando técnicas y el corpus correspondiente que permita de forma eficaz y eficiente identificar el uso ambiguo, tendencioso o imparcial de las palabras. Como se trata de algo muy subjetivo, depende mucho de la intención y la ideología del que habla/escribe y la ideología del que escucha/lee.



III. Palabras claves (Keywords) (Colocadas en orden de importancia. Máx. palabras: cinco)

Lenguaje sesgado, Procesamiento Lenguaje Natural, Inteligencia Artificial, Inteligencia Computacional, Análisis semántico, análisis de textos

IV. Justificación del proyecto (Describa el problema y su relevancia como objeto de investigación. Es importante una clara definición y delimitación del problema que abordará la investigación, ya que temas cuya definición es difusa o amplísima son difíciles de evaluar y desarrollar)

El procesamiento de lenguaje natural es una rama de la inteligencia artificial y el lenguaje, esta estudia la interacción entre el lenguaje humano y la máquina, entre sus empleos en la ciencia se encuentra el reconocimiento del habla, la traducción, resumen, detección de emociones, generación, comprensión y recuperación de lenguaje natural.

El lenguaje ha evolucionado y se ha ido adaptando a las necesidades de la comunicación humana, en ese sentido han variado y se ha definido 2 clases de lenguaje, el natural y el formal. Desde el área de la inteligencia artificial el que nos interesa es el estudio del lenguaje natural que tiene 2 objetivos bien definidos, el facilitar la comunicación con la computadora para que accedan a ella usuarios no especializados y modelar los procesos cognoscitivos que entran en juego en la comprensión del lenguaje para diseñar sistemas que realice tareas lingüísticas complejas.

V. Antecedentes del proyecto (Incluya el estado actual del conocimiento en el ámbito nacional e internacional. La revisión bibliográfica debe incluir en lo posible artículos científicos actuales, para evidenciar el conocimiento existente y el aporte de la Tesis propuesta. Esto es importante para el futuro artículo que resultará como producto de este trabajo)

En cuanto a los antecedentes podemos mencionar que en la actualidad poca información referida al tema, teniendo mayor información en Procesamiento de Lenguaje Natural para el idioma Inglés.

VI. Hipótesis del trabajo (Es el aporte proyectado de la investigación en la solución del problema)

Es posible generar un modelo que permita identificar lenguaje sesgado para el idioma castellano

VII. Objetivo general

Generar un modelo computacional capaz de identificar lenguaje sesgado para el idioma castellano.

VIII. Objetivos específicos

- Recopilar información referida al lenguaje sesgado en lenguaje castellano.
- Recopilar e identificar técnicas de inteligencia artificial para procesamiento de lenguaje natural en idioma castellano.
- Generar un modelo computacional para identificar lenguaje sesgado en idioma castellano

- IX. Metodología de investigación (Describir el(los) método(s) científico(s) que se empleará(n) para alcanzar los objetivos específicos, en forma coherente a la hipótesis de la investigación. Sustentar, con base bibliográfica, la pertinencia del(los) método(s) en términos de la representatividad de la muestra y de los resultados que se esperan alcanzar. Incluir los análisis estadísticos a utilizar)

Básicamente se utilizará la metodología de investigación científica con la implementación de técnicas de inteligencia artificial para el procesamiento de lenguaje natural.

- X. Referencias (Listar las citas bibliográficas con el estilo adecuado a su especialidad)

- Cortez Vasquez, A., Vega Huerta, H., & Pariona Quispe, J. (17 de Diciembre de 2021). UNMSM. Obtenido de [sisbib.unmsm.edu.pe: https://sisbib.unmsm.edu.pe/bibvirtual/publicaciones/risi/2009_n2/v6n2/a06v6n2.pdf](https://sisbib.unmsm.edu.pe/sisbib.unmsm.edu.pe/bibvirtual/publicaciones/risi/2009_n2/v6n2/a06v6n2.pdf)
- Gonzalez Hernandez, M. (17 de Diciembre de 2021). UNIVERSIDAD DE CHILE. Obtenido de [repositorio.uchile.cl: https://repositorio.uchile.cl/bitstream/handle/2250/136443/Tesis_Melisa_Gonzalez_Hernandez.pdf](https://repositorio.uchile.cl/bitstream/handle/2250/136443/Tesis_Melisa_Gonzalez_Hernandez.pdf)
- Greelane, U. d. (17 de Diciembre de 2021). GREELANE. Obtenido de [greelane.com: https://www.greelane.com/es/humanidades/ingles/what-is-biased-language-1689168/](https://www.greelane.com/es/humanidades/ingles/what-is-biased-language-1689168/)
- Ocaña Fernandez, Y., Valenzuela Fernandez, L., & Garro Aburto, L. (17 de Diciembre de 2021). SCIELO. Obtenido de [scielo.org.pe: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2307-79992019000200021](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2307-79992019000200021)
- Soluciones, D. (17 de Diciembre de 2021). DECIDE. Obtenido de [decidesoluciones.es: https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza](https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza)
- Vallez, M., & Pedraza Jimenez, R. (17 de Diciembre de 2021). HIPERTEXT. Obtenido de [upf.edu: https://www.upf.edu/hipertextnet/numero5/pln.html#Conclusion](https://www.upf.edu/hipertextnet/numero5/pln.html#Conclusion)
- Berry M. and Linoff G. "Data Mining Techniques". John Wiley & Sonsa, 1997.
- Bradley Efron, R.J. Tibshirani. "An Introduction to the Bootstrap". Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 1994.
- Bry X. "Analyses factorielles simples", Ed. Economica, Paris, 1995.
- Christopher Gandrud. "Reproducible Research with R and RStudio". Chapman & Hall/CRC the R Series, 2013.
- Hastie, Tibshirani and Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, 2009. John M. Chambers. Programming with R: Software for Data Analysis Springer, Stanford University, Palo Alto, 2008.
- Gareth James, Daniela Witten, Trevor Hastie y Robert Tibshirani. "An Introduction to Statistical Learning: with Applications in R". Springer Texts in Statistics, 2013.
- Giudici Paolo. "Applied Data Mining: Statistical Methods for Business and Industry". Wiley, 2005.
- Graham Williams, Data Mining with R. Rattle and R. Springer, New York, 2011.



- Dunhan M. "Data mining: Introductory and Advanced Topics". Prentice Hall, 2002.
- Han J. and Kamber M. "Data Mining: concepts and techniques", Morgan Kaufman Publishers 2001.

XI. Uso de los resultados y contribuciones del proyecto (Señalar el posible uso de los resultados y la contribución de los mismos)

El uso será para efectos de investigación y estudio de casos en los cursos correspondientes.

XII. Impactos esperados

i. Impactos en Ciencia y Tecnología

Generación de nuevas técnicas y/o algoritmos para el análisis de textos o procesamiento de lenguaje natural de la lengua castellana.

ii. Impactos económicos

iii. Impactos sociales

Identificación de lenguaje sesgado de las personas que se podrían aplicar en la selección de personal.

iv. Impactos ambientales

Al ser un modelo computacional contribuye de alguna medida en ahorrar costos o suprimir operaciones que tienen un impacto negativo en nuestro medio ambiente.

XIII. Recursos necesarios (Infraestructura, equipos y principales tecnologías en uso relacionadas con la temática del proyecto, señale medios y recursos para realizar el proyecto)

01 Computadora Core I7 11ava generación
1 TB de Disco Duro
DVD Writer
Lectora de CD
Tarjeta de Video NVidia GeForce2 MX 400
01 Impresora Laser Jet HP
01 Impresora Desk Jet HP
Conexión a internet dedicada

XIV. Localización del proyecto (indicar donde se llevará a cabo el proyecto)

Puno, Perú



XV. Cronograma de actividades

Actividad	Trimestres											
Analizar y seleccionar métodos o algoritmos de procesamiento de lenguaje natural	X	X	X	X								
Seleccionar los casos para ser analizados				X	X	X						
Desarrollar el modelo de identificación de lenguaje sesgado y probarlo						X	X	X	X			
Validar los resultados						X	X	X	X	X	X	
Redacción del documento final										X	X	X

XVI. Presupuesto

Descripción	Unidad de medida	Costo Unitario (S/.)	Cantidad	Costo total (S/.)
1. Personal	Persona	01	10000	10000
2. Materiales y equipos	Equipo	01	10000	10000
3. Servicios	Servicio	03	500	1500
4. Imprevistos	Otros	01	1500	1500
TOTAL				23000