

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PUCP

**IDENTIFICACIÓN DEL NIVEL DE COMPLEJIDAD DE TEXTO PARA
EL ENTRENAMIENTO DE CHATBOTS BASADO EN MACHINE
LEARNING: UNA REVISIÓN DE LITERATURA
TRABAJO DE INVESTIGACIÓN PARA LA OBTENCIÓN DEL GRADO
DE BACHILLER EN CIENCIAS CON MENCIÓN EN INGENIERÍA
INFORMÁTICA**

AUTOR

Hans Matos Rios

ASESOR:

Dr. Cesar Beltrán Castañón

Lima, Diciembre, 2020

Resumen

El nivel de complejidad textual puede ser un inconveniente para algunas personas al momento de usar Chatbots, debido a que estos programas podrían dar respuestas cuyo nivel de complejidad no sea el que entienda el usuario. Entonces, aquellos Chatbots deberían ser entrenados con un conjunto de datos cuya complejidad textual sea la deseada, para evitar confusiones con los usuarios. Para ello, se define una revisión sistemática, en la cual se usan las bases de datos de Google Scholar, ACM Digital Library e IEEE Xplore, de las cuáles se obtiene la información necesaria empleando las palabras claves definidas por el método PICOC, obteniendo un total de treinta y ocho documentos que evidencian la existencia de distintas métricas para analizar la complejidad textual de textos, así como experimentos de entrenamiento con Chatbots y los correspondientes resultados de sus interacciones con los usuarios. Además, analizando documentos de tesis asociadas al tema de investigación, se refuerzan los conceptos de que la complejidad textual puede ser analizado mediante conjunto de métricas. Finalmente, en base a lo desarrollado en la revisión de la literatura y documentos de tesis, se presentan las conclusiones deducidas.



Tabla de Contenido

1	Introducción.....	1
2	Método.....	2
2.1	Revisión Sistemática.....	2
2.1.1	<i>Preguntas de Investigación.....</i>	2
2.1.2	<i>Proceso de Búsqueda.....</i>	2
2.1.2.1	Estrategia de búsqueda.....	2
2.1.2.2	Motores de búsqueda a usar.....	2
2.1.2.3	Cadenas de búsqueda a usar.....	3
2.1.3	<i>Criterios de Inclusión y Exclusión.....</i>	4
2.1.3.1	Criterios de inclusión.....	4
2.1.3.2	Criterios de Exclusión.....	5
2.1.4	<i>Datos Extraídos.....</i>	5
2.1.5	<i>Datos Analizados.....</i>	5
2.2	Resultados.....	6
2.2.1	<i>Resultados de Búsqueda.....</i>	7
2.3	Discusión.....	9
2.3.1	<i>Respuesta a la pregunta “¿Cómo afecta al entrenamiento de un Chatbot el uso de texto complejo?”.....</i>	10
2.3.2	<i>Respuesta a la pregunta “¿Cuáles son las métricas/medidas de análisis de complejidad textual y como se usan para analizar la complejidad de textos?”.....</i>	10
2.3.3	<i>Respuesta a la pregunta “¿Por qué es necesario realizar un análisis de complejidad textual?”.....</i>	10
2.4	Revisión de Tesis.....	11
3	Conclusiones.....	12
4	Referencias.....	13

Índice de figuras



Índice de tablas

Tabla 1. Criterios PICOC para la pregunta 1.....	3
Tabla 2. Criterios PICOC para la pregunta 2.....	3
Tabla 3. Criterios PICOC para la pregunta 3.....	4
Tabla 4. Cadenas de búsqueda.....	4
Tabla 5. Formulario de Extracción.....	5
Tabla 6. Documentos encontrados por motor de búsqueda.....	6
Tabla 7. Tabla de resultado de revisión de literatura para la pregunta P1.....	7
Tabla 8. Tabla de resultado de revisión de literatura para la pregunta P2.....	7
Tabla 9. Tabla de resultado de revisión de literatura para la pregunta P3.....	8

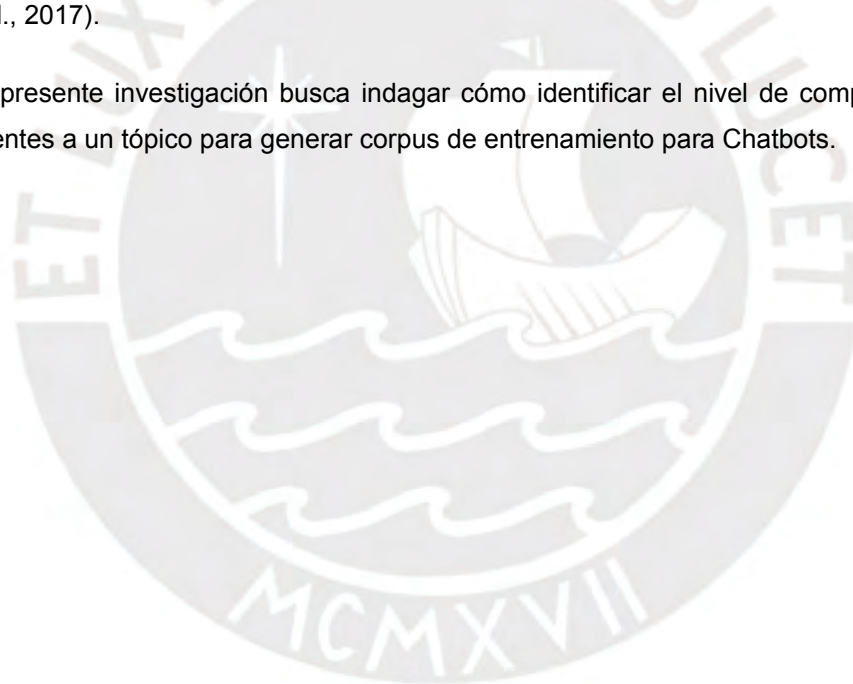


1 Introducción

Actualmente, los Chatbots son entrenados con textos referente a un dominio o tópico sin tomar en cuenta la complejidad que puedan tener sus respuestas (Beaver, 2018a). Dicha situación genera que las respuestas del Chatbot sean muy complejas en algunas situaciones, lo que lleva al usuario a replantear sus preguntas y así se genera confusión por parte del usuario; aquella situación puede ser observada en el estudio realizado por Beaver (2018) en donde el Chatbot que desarrollaron ocasionaba que los usuarios se confundieran cuando las respuestas que daba dicho Chatbot eran muy complejas y, debido a ello, los usuarios tenían que replantear sus preguntas (Beaver, 2018b).

Por otro lado, para analizar la complejidad textual se utilizan distintos tipos de métricas, como lo son las características descriptivas de un texto, como el conteo de caracteres, oraciones, párrafos y sílabas (Stevanoski & Gievska, 2019); las características morfo-sintácticas de los textos, como la cantidad de las clases gramaticales (Solovyev et al., 2018); el significado de las palabras (Fitzgerald et al., 2015, 2016); métricas clásicas de legibilidad, como los índices de Flesch, Flesch-Kincaid, New Dale-Chall y Fernández-Huerta (Crossley et al., 2019; López-Anguita et al., 2018; Ojha et al., 2018; Trozsek et al., 2017).

Por ello, la presente investigación busca indagar cómo identificar el nivel de complejidad textual de textos referentes a un tópico para generar corpus de entrenamiento para Chatbots.



2 Método

Para realizar la revisión del Estado del Arte, se decidió usar la metodología de la revisión sistemática, la cual fue propuesta por Barbara Kitchenham. El uso de esta metodología de investigación permitió que al investigador resumir la evidencia existente con respecto al uso de la tecnología usada para resolver problemas similares al que se investigó en este documento (Keele, 2007).

También, se revisaron tesis del repositorio de tesis PUCP.

2.1 Revisión Sistemática

Las cadenas de búsqueda fueron ejecutadas por última vez el 9 de Mayo del 2020, por lo que los resultados obtenidos reflejan aquel instante de tiempo.

2.1.1 Preguntas de Investigación

En esta revisión sistemática, se busca responder a las siguientes preguntas:

- P1. ¿Cómo afecta al entrenamiento de un *Chatbot* el uso de texto complejo?
- P2. ¿Cuáles son las métricas/medidas de análisis de complejidad textual y como se usan para analizar la complejidad de textos?
- P3. ¿Por qué es necesario realizar un análisis de complejidad textual?

2.1.2 Proceso de Búsqueda

En esta sección describimos primero la estrategia de búsqueda a utilizar, luego los motores de búsqueda a usar, terminando con las cadenas de búsqueda a usar.

2.1.2.1 Estrategia de búsqueda

Para analizar el estado del arte, se buscaron estudios primarios en tres bases de datos académicas mediante 1 cadena de búsqueda por pregunta: Google Scholar, ACM Digital Library e IEEE Explore.

Sin embargo, no todos los documentos encontrados fueron usados debido a que no todos cumplían con los criterios de inclusión y exclusión, algunos eran documentos repetidos o no guardaban relación con el tema a investigar.

2.1.2.2 Motores de búsqueda a usar

Para encontrar información que ayuden en la presente investigación, se ha decidido utilizar algunas bases de datos referenciadas por la Pontificia Universidad Católica del Perú, así como las que fueron sugeridas por el asesor debido a su experiencia. Estas bases de datos a usar son las siguientes:

- Google Scholar.
- ACM Digital Library.
- IEEE Explore.

2.1.2.3 Cadenas de búsqueda a usar

Para generar las cadenas de búsqueda a usar por cada pregunta, se utilizaron los conectores lógicos como AND y OR, además de agregar sinónimos a las palabras clave en caso sea posible. Dichas palabras clave fueron usadas en inglés y no en español. Para esto, se realizó una tabla PICOC por cada pregunta,.

A continuación en la Tabla 1, se muestra la tabla PICOC para la pregunta 1, la cual es “¿Cómo afecta al entrenamiento de un Chatbot el uso de texto complejo?”.

Tabla 1. Criterios PICOC para la pregunta 1

Criterio PICOC	Consideraciones
Population	Chatbot, Text
Intervention	
Comparison	
Outcome	
Context	Text Complexity, Text Difficulty

Para la pregunta 2, la cual es “¿Cuáles son las métricas/medidas de análisis de complejidad textual y como se usan para analizar la complejidad de textos?”, se muestran los criterios PICOC en la Tabla 2.

Tabla 2. Criterios PICOC para la pregunta 2

Criterio PICOC	Consideraciones
Population	Text complexity measures, Text complexity measure, Text complexity metrics
Intervention	Texto Classification, Classification
Comparison	
Outcome	
Context	Machine Learning

Los criterios PICOC de la pregunta 3, la cual es “¿Por qué es necesario realizar un análisis de complejidad textual?”, están en la Tabla 3.

Tabla 3. Criterios PICOC para la pregunta 3

Criterio PICOC	Consideraciones
Population	Text complexity
Intervention	Text Classification, Classification
Comparison	
Outcome	
Context	Coh-Metrix

Gracias al análisis hecho mediante el uso de las tablas con criterios PICOC, se pudo crear las cadenas de búsqueda a usar por pregunta, las cuales se muestran en la Tabla 4.

Tabla 4. Cadenas de búsqueda.

Pregunta	Cadena de búsqueda
P1	("Chatbot") AND ("Training" OR "Train") AND ("Text Complexity" OR "Text Difficulty")
P2	("Text Complexity Measure" OR "Text Complexity Metrics" OR "Text Complexity Measures") AND ("Text Classification" OR "Classification") AND ("Machine Learning")
P3	("Text Complexity" AND "Machine Learning" AND ("Text Classification" OR "Classification") AND "Coh-Metrix" AND ("Reason" OR "Objective")) OR "AzterTest"

2.1.3 Criterios de Inclusión y Exclusión

Para esta investigación, no será posible utilizar todos los documentos encontrados, debido a la gran cantidad de los resultados obtenidos. Por ello, y para evitar caer en la imparcialidad, se han formulado los siguientes criterios de inclusión y exclusión.

2.1.3.1 Criterios de inclusión

- El documento encontrado debe haber usado la herramienta "Coh-Metrix". Este criterio aplica para los documentos encontrados para la pregunta 3.
 - Este criterio ha sido considerado debido a que esta herramienta es la más completa para analizar la complejidad textual ya que implementa cerca de 200 medidas de cohesión, lenguaje y legibilidad (Graesser et al., 2004; McNamara et al., 2014).
- El documento debe haber sido citado al menos 1 vez. Este criterio aplica para los documentos encontrados para todas las preguntas.
 - Este criterio se considera ya que así se podrá saber que tan relevante es en el campo de investigación. Habrá excepciones si es que en el documento se analiza la complejidad textual para textos en español.

- El tema del documento debe guardar relación con el tema a desarrollar en esta investigación. Este criterio aplica para todos los documentos que hayan pasado los filtros de fecha, idioma y cantidad de citas.
- Este criterio se considera debido a que algunos de los documentos encontrados son una revisión del estado del arte; otros documentos están relacionados a otras áreas de estudio como por ejemplo las ciencias sociales, dejando de lado el tema de la Informática.

2.1.3.2 Criterios de Exclusión

- El documento debe ser del año 2014 o más reciente. Este criterio aplica para los documentos encontrados para todas las preguntas.
 - Como se hará una revisión del estado del arte, la antigüedad del documento no debe ser mayor a los 6 años. Este filtro no tiene efecto sobre los documentos para la revisión teórica.
- El documento debe estar en español o inglés. Este criterio aplica para los documentos encontrados para todas las preguntas
 - Se consideró solo utilizar documentos en español o inglés porque el primero es la lengua nativa del investigador y el segundo porque la mayoría de documentos están en inglés.
- El documento no debe ser una patente. Este criterio aplica para todos los documentos que hayan pasado los filtros de fecha, idioma y cantidad de citas.
 - Se consideró este criterio debido a que, al realizar la revisión de patentes de Google encontrada, estas no contenían mucha información y no aportaban a la investigación.

2.1.4 Datos Extraídos

Las partes de interés de la literatura encontrada será, en primera instancia, el resumen o introducción para poder filtrar aquellos documentos de manera inicial utilizando los criterios de inclusión y exclusión. Una vez analizado el resumen o introducción, se utilizará el cuerpo del documento encontrado para entender a mayor detalle el contenido, así como las conclusiones.

2.1.5 Datos Analizados

Para analizar los datos extraídos, se creó un formulario de extracción para analizar los datos, el cual está mostrado en la Tabla 5.

Tabla 5. Formulario de Extracción.

Campo	Descripción	Pregunta General
-------	-------------	------------------

Título	El título del artículo	Criterio general
Autor	El o los autores del artículo	Criterio general
Año de publicación	Año de publicación del artículo	Criterio general
Idioma	Idioma en el que está el artículo	Criterio general
Modelo de aprendizaje de máquina usado	Modelo tradicional o de red neuronal usado en el artículo	P1
Estructura del modelo de aprendizaje de máquina usado	Arquitectura del modelo de red neuronal, si es que fue usada una red	P1
Medidas de rendimiento usado en los modelos	Puede ser "accuracy", "precision", "recall", "F1 score"	P1
Resultados numéricos de las métricas de medidas de rendimiento	Se necesita para determinar qué tan exitoso fue el experimento	P1
Nombre del conjunto de datos usado	El nombre del dataset de entrenamiento	P1, P2 Y P3
Idioma del conjunto de datos usado	El idioma del dataset con el cual fue realizado el experimento	P1, P2 Y P3
Nombre de los conjuntos de métricas de complejidad textual	Las métricas de complejidad textual son agrupadas por grupos distintos.	P2
Nombre individual de cada métrica de complejidad textual	Las métricas de complejidad textual son agrupadas por grupos distintos	P2
Objetivos	Objetivo del experimento realizado	P3
Resultados	Resultados del experimento realizado	P3

2.2 Resultados

En la Tabla 6 se muestran la cantidad de documentos por pregunta, usando las cadenas de búsqueda creadas, detallando la cantidad de documentos totales antes y después de aplicar los criterios de inclusión y exclusión.

Tabla 6. Documentos encontrados por motor de búsqueda

Motor de búsqueda	P1	P2	O3
Google Scholar	13	45	103

ACM Digital Library	2	2	3
IEEE Explore	1	0	0
Total sin criterios	16	47	106
Total con criterios	3	11	24

2.2.1 Resultados de Búsqueda

En esta sección se detallan en una tabla por cada pregunta planteada, los resultados finales de la revisión sistemática realizada.

Tabla 7. Tabla de resultado de revisión de literatura para la pregunta P1

Título	Autores
A machine learning approach to persian text readability assessment using a crowdsourced dataset	Mohammadi, Hamid and Khasteh, Seyed Hossein
Automatic Conversation Review for Intelligent Virtual Assistants	Beaver, Ian R.
Deep Reinforcement Learning for Chatbots Using Clustered Actions and Human-Likeness Rewards	Cuáyahuatl, H. and Lee, D. and Ryu, S. and Choi, S. and Hwang, I. and Kim, J.

Tabla 8. Tabla de resultado de revisión de literatura para la pregunta P2

Título	Autores
A context-aware approach for the identification of complex words in natural language texts	Davoodi, Elnaz and Kosseim, Leila and Mongrain, Matthew
Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications	Balakrishna, Sowmya Vajjala
Assessment of reading difficulty levels in Russian academic texts: Approaches and metrics	Solovyev, Valery and Ivanov, Vladimir and Solnyshkina, Marina
Complexity Measures and POS N-grams for Author Identification in Several Languages	López-Anguita, Rocío and Montejo-Ráez, Arturo and Díaz-Galiano, Manuel C.
Computational lyricology: quantitative approaches to understanding song lyrics and their interpretations	Choi, Kahyun
Has First-Grade Core-Reading-Program Text Complexity Changed across Six Decades?	Fitzgerald, Jill and Elmore, Jeff and Hiebert, Elfrieda H. and Relyea, Jackie Eunjung and Stenner, A. Jackson
Important text characteristics for early-grades text complexity.	Fitzgerald, Jill and Elmore, Jeff and Koons, Heather and Hiebert, Elfrieda H. and Bowen, Kimberly and Sanford-Moore, Eleanor E. and

	Stenner, A. Jackson
Linguistic Metadata Augmented Classifiers at the CLEF 2017 Task for Early Detection of Depression.	Trotzek, Marcel and Koitka, Sven and Friedrich, Christoph M.
Moving beyond classic readability formulas: new methods and new models	Crossley, Scott A. and Skalicky, Stephen and Dascalu, Mihai
Perusal of readability with focus on web content understandability	Ojha, Pawan Kumar and Ismail, Abid and Kuppusamy, K. S.
Team Ned Leeds at SemEval-2019 Task 4: Exploring Language Indicators of Hyperpartisan Reporting	Stevanoski, Bozhidar and Gievska, Sonja

Tabla 9. Tabla de resultado de revisión de literatura para la pregunta P3

Título	Autores
"Is There Choice in Non-Native Voice?" Linguistic Feature Engineering and a Variationist Perspective in Automatic Native Language Identification	Bich, Serhiy
A corpus driven computational intelligence framework for deception detection in financial text	Minhas, Saliha Z.
An analysis of student model portability	Aguirre, Benjamín Valdés and Uresti, Jorge A. Ramírez and Du Boulay, Benedict
Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications	Balakrishna, Sowmya Vajjala
Assessing automatic text classification for interactive language learning	Branco, António and Rodrigues, João and Costa, Francisco and Silva, João and Vaz, Rui
Assessing text and web accessibility for people with autism spectrum disorder	Yaneva, Victoria
Automated assessment of non-native learner essays: Investigating the role of linguistic features	Vajjala, Sowmya
Automated evaluation of text and discourse with Coh-Metrix	McNamara, Danielle S. and Graesser, Arthur C. and McCarthy, Philip M. and Cai, Zhiqiang
Automatic classification of the complexity of nonfiction texts in portuguese for early school years	Hartmann, Nathan and Cucatto, Livia and Brants, Danielle and Aluísio, Sandra
Automatic Essay Scoring Based on Coh-Metrix Feature Selection for Chinese English Learners	Li, Xia and Liu, Jianda
CaMLA Working Papers Predicting Listening Item Difficulty with Language Complexity Measures: A Comparative Data Mining Study	Aryadoust, Vahid and Goh, Christine C.M.

Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing	Kleijn, Suzanne
Coh-Matrix-Esp: A complexity analysis tool for documents written in Spanish	Quispesaravia, Andre and Perez, Walter and Cabezudo, Marco Sobrevilla and Alva-Manchego, Fernando
Cohesion network analysis of CSCL participation	Dascalu, Mihai and McNamara, Danielle S. and Trausan-Matu, Stefan and Allen, Laura K.
Computational lyricology: quantitative approaches to understanding song lyrics and their interpretations	Choi, Kahyun
Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning	Sung, Yao-Ting and Chen, Ju-Ling and Cha, Ji-Her and Tseng, Hou-Chiang and Chang, Tao-Hsing and Chang, Kuo-En
Crawling by readability level	Wagner Filho, Jorge A. and Wilkens, Rodrigo and Zilio, Leonardo and Idiart, Marco and Villavicencio, Aline
Development and validation of an automated essay scoring framework by integrating deep features of English language	Latifi, Syed Muhammad Fahad
From spin to swindle: identifying falsification in financial text	Minhas, Saliha and Hussain, Amir
Important text characteristics for early-grades text complexity.	Fitzgerald, Jill and Elmore, Jeff and Koons, Heather and Hiebert, Elfrieda H. and Bowen, Kimberly and Sanford-Moore, Eleanor E. and Stenner, A. Jackson
Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR	YAO-TING, S. U. N. G. and WEI-CHUN, L. I. N. and Dyson, Scott Benjamin and CHANG, KUO-E. N. and YU-CHIA, C. H. E. N.
OneStopEnglish corpus: A new corpus for automatic readability assessment and text simplification	Vajjala, Sowmya and Lučić, Ivana
Towards context-aware automated writing evaluation systems	Patout, Pierre-André and Cordy, Maxime
Using technology in language assessment	Li, Haiying and Shubeck, K. and Graesser, Arthur C.
AzterTest: Open source linguistic and stylistic analysis tool	Bengoetxea, Kepa, Gonzales-Dios, Itziar, Aguirregoitia, Amaia

2.3 Discusión

Luego de haber recopilado los distintos documentos obtenidos de las distintas bases de datos, se posee ahora la información necesaria para responder a las preguntas planteadas.

2.3.1 Respuesta a la pregunta “¿Cómo afecta al entrenamiento de un Chatbot el uso de texto complejo?”

Para responder esta pregunta se pudo deducir que el nivel de complejidad de una respuesta hecha por un Chatbot puede ocasionar que el usuario reformule su pregunta o no (Beaver, 2018a). Por otro lado, un Chatbot suficientemente entrenado puede identificar el nivel de dificultad de las respuestas de un usuario en una conversación en aplicativos de mensajería (Mohammadi & Khasteh, 2018). Además, un Chatbot que ha sido entrenado con texto de longitud extensa puede predecir de manera más exacta qué diálogos de una conversación son buenos y cuáles malos (Cuáyahuitl et al., 2019). Entonces, se puede responder que el entrenamiento de un Chatbot con textos de distintos niveles de complejidad puede influenciar en el flujo de la conversación.

2.3.2 Respuesta a la pregunta “¿Cuáles son las métricas/medidas de análisis de complejidad textual y como se usan para analizar la complejidad de textos?”

Luego de analizar los datos encontrados en el formulario para esta pregunta, se pudo recopilar que existen varias formas de medir la complejidad de un texto. Por ejemplo, se tienen medidas que analizan las características descriptivas de un texto: Número de caracteres, oraciones, párrafos y sílabas (Stevanoski & Gievska, 2019). También se analizan las características morfo-sintácticas de los textos, por ejemplo: Cantidad de sustantivos, verbos, adjetivos, pronombres, pronombres personales, negaciones y conectores por oración (Solovyev et al., 2018). Además, se puede analizar el significado de las palabras mediante la edad de adquisición, la abstracción y la rareza de la palabra (Fitzgerald et al., 2015, 2016). Otras métricas comúnmente usadas son las fórmulas clásicas de legibilidad, las cuales mezclan los valores calculados en su mayoría por las medidas de características descriptivas. Estas son las siguientes: Flesch-Kincaid, Flesch, New Dale-Chall, Gunning Fog Index, Fernández-Huerta (Crossley et al., 2019; López-Anguita et al., 2018; Ojha et al., 2018; Trozsek et al., 2017). Existen otras métricas que se usan (Balakrishna, 2015; Choi, 2018; Davoodi et al., 2017), pero las anteriores son las que se usaron con más frecuencia en los documentos analizados. Entonces, existen distintas métricas que se usan para analizar distintas características textuales, como se explicó líneas más arriba, para así determinar qué tan complejo es un texto.

2.3.3 Respuesta a la pregunta “¿Por qué es necesario realizar un análisis de complejidad textual?”

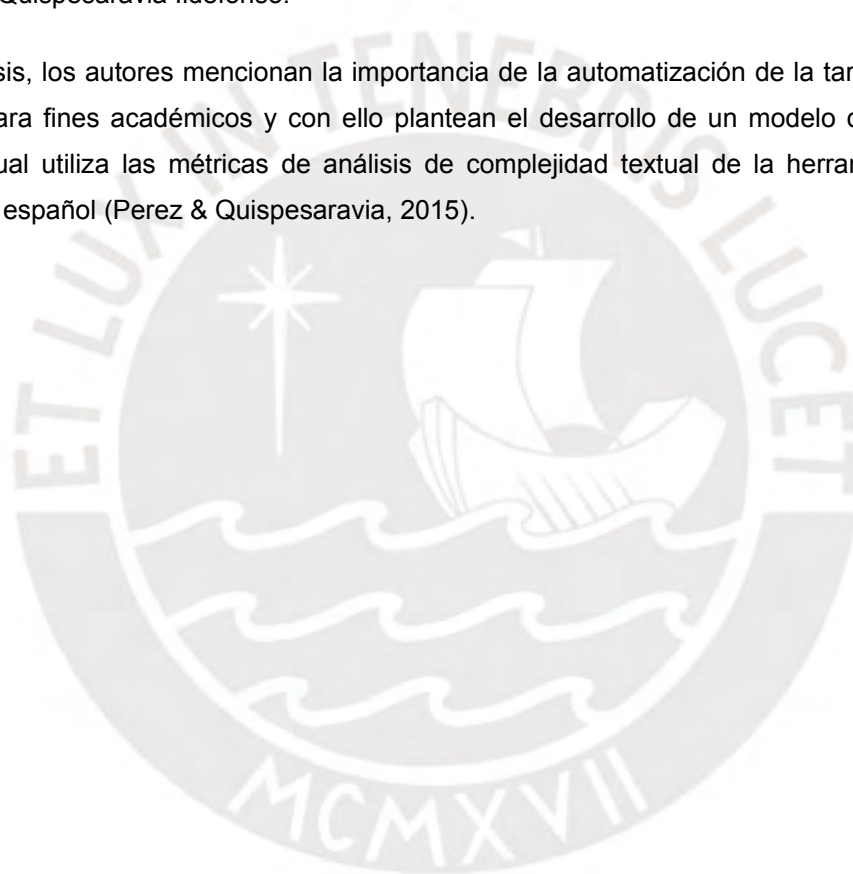
Con los documentos recopilados, se pudo concluir que el análisis de la complejidad textual es necesario para resolver distintos problemas del mundo real, como por ejemplo: Detectar fraude financiero (S. Minhas & Hussain, 2016; S. Z. Minhas, 2016); inferir el lenguaje nativo de un autor (Bich, 2017); clasificar modelos estudiantiles (Aguirre et al., 2016); simplificar texto complejo (Balakrishna, 2015); clasificar textos según su complejidad (Bengoetxea et al., 2020; Branco et al., 2014; Fitzgerald et al., 2016; Hartmann et al., 2016; Kleijn, 2018; Quispesaravia et al., 2016; Sung et al., 2015); evaluar la accesibilidad de los textos (Yaneva, 2016); calificar ensayos automáticamente (Latifi, 2016; Li & Liu, 2016; Patout & Cordy, 2019; Vajjala, 2018) analizar la dificultad de textos auditivos (Aryadoust & Goh,

2014); evaluar el grado de participación de un estudiante en un ambiente educativo virtual (Dascalu et al., 2018); analizar la complejidad de letras de canciones (Choi, 2018); generar automáticamente corpus textuales clasificados según su complejidad textual (Vajjala, 2018; Wagner Filho et al., 2016) y evaluar la legibilidad de los textos (YAO–TING et al., 2015).

2.4 Revisión de Tesis

Para la revisión de Tesis, se usó el repositorio de la PUCP, conocido como Tesis PUCP. Para ello, se utilizó el filtro de búsqueda “Complejidad textual”. Una vez obtenido los textos del repositorio, se procedió a filtrar según los criterios de búsqueda, basándose en un análisis del resumen. Como resultado, luego de haber realizado dicho proceso, se obtuvo una tesis titulada “Herramienta de Análisis y Clasificación de Complejidad de Textos en Español”, cuyos autores son Walter Perez Urcia y André Raúl Quispesaravia Ildelfonso.

En dicha tesis, los autores mencionan la importancia de la automatización de la tarea de la selección de textos para fines académicos y con ello plantean el desarrollo de un modelo de clasificación de textos, el cual utiliza las métricas de análisis de complejidad textual de la herramienta Coh-Metrix adaptada al español (Perez & Quispesaravia, 2015).



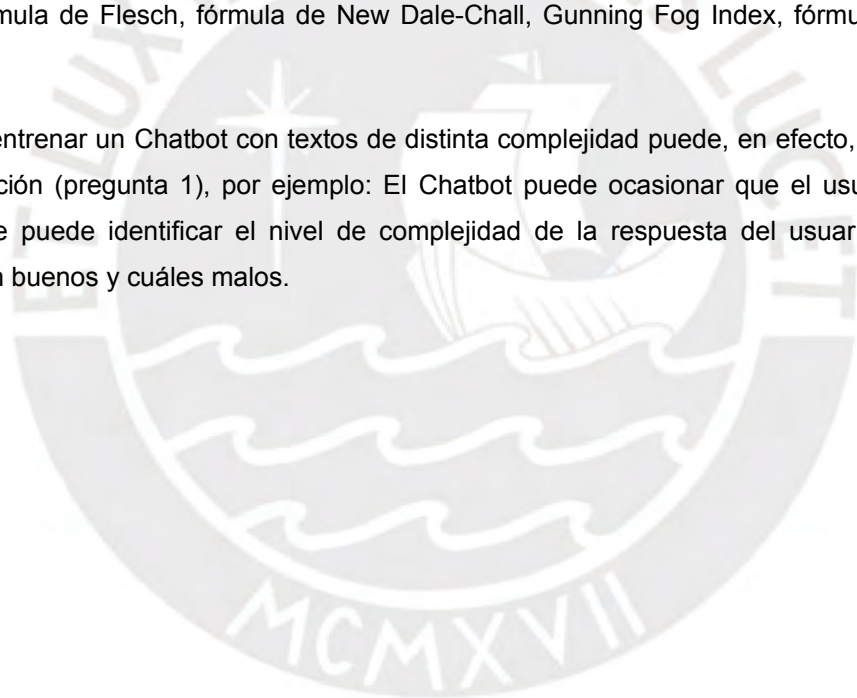
3 Conclusiones

De lo que se obtuvo luego de realizar la revisión sistemática y responder a las tres preguntas planteadas, se puede concluir que:

Primero, la complejidad textual tiene muchas aplicaciones en el mundo real (pregunta 3), las cuales son las siguientes: Detectar fraude financiero; inferir el lenguaje nativo de un autor; clasificar modelos estudiantiles; clasificar textos según su complejidad; evaluar la accesibilidad y legibilidad de los textos; calificar un ensayo automáticamente; analizar la dificultad de textos auditivos; evaluar el grado de participación de un estudiante en un ambiente educativo virtual; analizar la complejidad de letras de canciones y generar automáticamente corpus textuales clasificados según su complejidad textual.

Segundo, para analizar la complejidad de los textos, se disponen de varias métricas y fórmulas que se pueden usar (pregunta 2), como por ejemplo: Número de caracteres, oraciones, párrafos, sílabas, cantidad de sustantivos, verbos, adjetivos, pronombres, pronombres personales, negaciones, conectores por oración, edad de adquisición, la abstracción, la rareza de la palabra, fórmula de Flesch-Kincaid, fórmula de Flesch, fórmula de New Dale-Chall, Gunning Fog Index, fórmula de Fernández-Huerta.

Tercero, el entrenar un Chatbot con textos de distinta complejidad puede, en efecto, afectar el flujo de la conversación (pregunta 1), por ejemplo: El Chatbot puede ocasionar que el usuario replantee su pregunta, se puede identificar el nivel de complejidad de la respuesta del usuario y predecir que diálogos son buenos y cuáles malos.



4 Referencias

- Aguirre, B. V., Uresti, J. A. R., & Du Boulay, B. (2016). An analysis of student model portability. *International Journal of Artificial Intelligence in Education*, 26(3), 932-974.
- Aryadoust, V., & Goh, C. C. M. (2014). *CaMLA Working Papers Predicting Listening Item Difficulty with Language Complexity Measures: A Comparative Data Mining Study*.
- Balakrishna, S. V. (2015). *Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications*.
- Beaver, I. R. (2018a). *Automatic Conversation Review for Intelligent Virtual Assistants*.
- Beaver, I. R. (2018b). *Automatic Conversation Review for Intelligent Virtual Assistants*.
- Bengoetxea, K., Gonzalez-Dios, I., & Aguirregoitia, A. (2020). AzterTest: Open source linguistic and stylistic analysis tool. *Procesamiento de Lenguaje Natural*, 64, 61-68.
<https://doi.org/10.26342/2020-64-7>
- Bich, S. (2017). « *Is There Choice in Non-Native Voice?*» *Linguistic Feature Engineering and a Variationist Perspective in Automatic Native Language Identification*.
- Branco, A., Rodrigues, J., Costa, F., Silva, J., & Vaz, R. (2014). Assessing automatic text classification for interactive language learning. *International Conference on Information Society (i-Society 2014)*, 70-78.
- Choi, K. (2018). *Computational lyricology: Quantitative approaches to understanding song lyrics and their interpretations*. University of Illinois at Urbana-Champaign.
- Crossley, S. A., Skalicky, S., & Dascalu, M. (2019). Moving beyond classic readability formulas: New methods and new models. *Journal of Research in Reading*, 42(3-4), 541-561.
- Cuáyahuitl, H., Lee, D., Ryu, S., Choi, S., Hwang, I., & Kim, J. (2019). Deep Reinforcement Learning for Chatbots Using Clustered Actions and Human-Likeness Rewards. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1-8.
- Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L. K. (2018). Cohesion network analysis of CSCL participation. *Behavior Research Methods*, 50(2), 604-619.

- Davoodi, E., Kosseim, L., & Mongrain, M. (2017). A context-aware approach for the identification of complex words in natural language texts. *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, 97-100.
- Fitzgerald, J., Elmore, J., Koons, H., Hiebert, E. H., Bowen, K., Sanford-Moore, E. E., & Stenner, A. J. (2015). Important text characteristics for early-grades text complexity. *Journal of Educational Psychology*, *107*(1), 4-4.
- Fitzgerald, J., Elmore, J., Relyea, J. E., Hiebert, E. H., & Stenner, A. J. (2016). Has first-grade core reading program text complexity changed across six decades? *Reading Research Quarterly*, *51*(1), 7-28.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, *36*(2), 193-202. <https://doi.org/10.3758/BF03195564>
- Hartmann, N., Cucatto, L., Brants, D., & Aluísio, S. (2016). Automatic classification of the complexity of nonfiction texts in portuguese for early school years. *International Conference on Computational Processing of the Portuguese Language*, 12-24.
- Keele, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical report, Ver. 2.3 EBSE Technical Report. EBSE.*
- Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. LOT.
- Latifi, S. M. F. (2016). *Development and validation of an automated essay scoring framework by integrating deep features of English language*.
- Li, X., & Liu, J. (2016). Automatic Essay Scoring Based on Coh-Metrix Feature Selection for Chinese English Learners. *International Symposium on Emerging Technologies for Education*, 382-393.
- López-Anguita, R., Montejo-Ráez, A., & Díaz-Galiano, M. C. (2018). *Complexity Measures and POS N-grams for Author Identification in Several Languages*.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

- Minhas, S., & Hussain, A. (2016). From spin to swindle: Identifying falsification in financial text. *Cognitive computation*, 8(4), 729-745.
- Minhas, S. Z. (2016). *A corpus driven computational intelligence framework for deception detection in financial text*.
- Mohammadi, H., & Khasteh, S. H. (2018). A machine learning approach to persian text readability assessment using a crowdsourced dataset. *arXiv preprint arXiv:1810.06639*.
- Ojha, P. K., Ismail, A., & Kuppusamy, K. S. (2018). Perusal of readability with focus on web content understandability. *Journal of King Saud University-Computer and Information Sciences*.
- Patout, P.-A., & Cordy, M. (2019). Towards context-aware automated writing evaluation systems. *Proceedings of the 1st ACM SIGSOFT International Workshop on Education through Advanced Software Engineering and Artificial Intelligence*, 17-20.
- Perez, W., & Quispesaravia, A. (2015). *Herramienta de Análisis y Clasificación de Complejidad de Textos en Español* [Tesis de Pregrado, Pontificia Universidad Católica del Perú].
http://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/6042/PEREZ_WALTER_QUISPESARAVIA_ANDR%C3%89_HERRAMIENTA_TEXTO.pdf?sequence=1&isAllowed=y.
- Quispesaravia, A., Perez, W., Cabezudo, M. S., & Alva-Manchego, F. (2016). Coh-Matrix-Esp: A complexity analysis tool for documents written in Spanish. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4694-4698.
- Solovyev, V., Ivanov, V., & Solnyshkina, M. (2018). Assessment of reading difficulty levels in Russian academic texts: Approaches and metrics. *Journal of Intelligent & Fuzzy Systems*, 34(5), 3049-3058.
- Stevanoski, B., & Gievska, S. (2019). Team Ned Leeds at SemEval-2019 Task 4: Exploring Language Indicators of Hyperpartisan Reporting. *Proceedings of the 13th International Workshop on Semantic Evaluation*, 1026-1031.
- Sung, Y.-T., Chen, J.-L., Cha, J.-H., Tseng, H.-C., Chang, T.-H., & Chang, K.-E. (2015). Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning. *Behavior research methods*, 47(2), 340-354.

Trotzek, M., Koitka, S., & Friedrich, C. M. (2017). Linguistic Metadata Augmented Classifiers at the CLEF 2017 Task for Early Detection of Depression. *CLEF (Working Notes)*.

Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education*, 28(1), 79-105.

Wagner Filho, J. A., Wilkens, R., Zilio, L., Idiart, M., & Villavicencio, A. (2016). Crawling by readability level. *International Conference on Computational Processing of the Portuguese Language*, 306-318.

Yaneva, V. (2016). *Assessing text and web accessibility for people with autism spectrum disorder*.

YAO-TING, S. U. N. G., WEI-CHUN, L. I. N., Dyson, S. B., CHANG, K. N., & YU-CHIA, C. H. E. N. (2015). Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR. *The Modern Language Journal*, 99(2), 371-391.

