



UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE QUITO

CARRERA DE COMPUTACIÓN

**TEMA:
ESTADO DE ARTE PARA LA ELABORACIÓN DE PRUEBAS
UTILIZANDO EL PROCESAMIENTO DE LENGUAJE NATURAL
EN EL AREA DE FÍSICA Y MATEMÁTICA.**

Trabajo de titulación previo a la obtención del
Título de Ingeniero en Ciencias de la Computación

AUTORES:
ANTHONY DAVID PIZARRO MALACATUS
BRANDON ARIEL LÓPEZ NUÑEZ

TUTOR: GUSTAVO ERNESTO NAVAS RUILOVA

Quito – Ecuador

2023

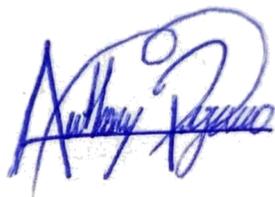
CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Anthony David Pizarro Malacatus con documento de identificación N° 1104541428 y Brandon Ariel López Nuñez con documento de identificación N° 1726729534; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 16 de Agosto del año 2023

Atentamente,



Anthony David Pizarro Malacatus
1104541428



Brandon Ariel López Nuñez
1726729534

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Nosotros, Anthony David Pizarro Malacatus con documento de identificación N° 1104541428 y Brandon Ariel López Nuñez con documento de identificación N° 1726729534, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Artículo Académico: “Estado del Arte para la Elaboración de Pruebas Utilizando el Procesamiento de Lenguaje Natural en el Área de Física y Matemática”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 16 de Agosto del año 2023

Atentamente



Anthony David Pizarro Malacatus
1104541428



Brandon Ariel López Nuñez
1726729534

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Gustavo Ernesto Navas Ruilova con documento de identificación N° 1705675625, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ESTADO DEL ARTE PARA LA ELABORACIÓN DE PRUEBAS UTILIZANDO EL PROCESAMIENTO DE LENGUAJE NATURAL EN EL ÁREA DE FÍSICA Y MATEMÁTICA, realizado por Anthony David Pizarro Malacatus con documento de identificación N° 1104541428 y Brandon Ariel López Nuñez con documento de identificación N° 1726729534, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 16 de Agosto del año 2023

Atentamente,



Ing. Gustavo Ernesto Navas Ruilova, MSc
1705675625

ESTADO DEL ARTE PARA LA ELABORACIÓN DE PRUEBAS UTILIZANDO EL PROCESAMIENTO DE LENGUAJE NATURAL EN EL ÁREA DE FÍSICA Y MATEMÁTICA.

1st Anthony David Pizarro
Malacatus
*Universidad Politécnica
Salesiana*
Quito - Ecuador
apizarrom@est.ups.edu.ec

2nd Brandon Ariel López
Nuñez
*Universidad Politécnica
Salesiana*
Quito - Ecuador
blopez2@est.ups.edu.ec

3rd Gustavo Ernesto Navas
Ruilova
*Universidad Politécnica
Salesiana*
Quito - Ecuador
gnavas@ups.edu.ec

Resumen

El presente estado del arte analiza la investigación actual sobre el uso del procesamiento natural de lenguaje (por sus siglas NLP en inglés) en la creación de pruebas en materias a nivel universitario de física y matemáticas. Los resultados muestran que los investigadores están conscientes de la necesidad de automatizar la creación de pruebas en los sistemas de aprendizaje electrónico y ven el aprendizaje automático como una solución prometedora. Los hallazgos de este análisis del estado del arte brindan información sobre la investigación en el campo de la generación de pruebas en física y matemáticas. Estos descubrimientos complementan la información actual y sirven como base para investigaciones y progreso en el campo de la evaluación educativa.

Palabras Clave- NLP, Área educativa, Evaluación, elaboración de preguntas.

Abstract

These state-of-the-art reviews current research on the use of natural language processing (by its acronym NLP in English) in the creation of tests in college-level subjects of physics and mathematics. The results show that researchers are aware of the need to automate test creation in e-learning systems and see machine learning as a promising solution. The findings of this state-of-the-art analysis provide insights into research in the field of evidence generation in physics and mathematics. These findings complement current information and serve as the basis for research and progress in the field of educational assessment.

Keywords-NLP, Educational area, Assessment, elaboration of questions.

1. Introducción

La evaluación del aprendizaje es un aspecto fundamental en la educación, ya que permite medir el nivel de conocimientos adquiridos por los estudiantes y evaluar la efectividad de los métodos pedagógicos utilizados [1]. Por lo tanto, las pruebas son una herramienta de uso frecuente para evaluar el aprendizaje en varios campos del conocimiento, como las matemáticas y la física. Sin embargo, la elaboración y evaluación de pruebas puede ser un proceso costoso y laborioso para los docentes [2]. Particularmente cuando se trata de grandes grupos de estudiantes. Además, existe el riesgo de que las preguntas no sean lo suficientemente claras o precisas, eso que podría afectar negativamente la calidad de las pruebas. El NLP “se enfoca en la capacidad de las computadoras para entender, interpretar y generar lenguaje humano de manera natural” [3]. En el contexto del desarrollo de pruebas, el uso de NLP se centra en la generación automática de preguntas y respuestas, evaluando la complejidad del texto utilizado en las preguntas e identificando errores y ambigüedades en la respuesta del estudiante.

Abordando esta temática, se empleará la metodología de Systematic Mapping Study (SMS) y la revisión de la literatura sistemática (SLR). El SMS [4] Permitirá un análisis exhaustivo y sistemático sobre las investigaciones existentes en el campo utilizando un enfoque estructurado y riguroso. Por otro lado, la SLR [5] Se centrará en analizar de manera crítica y exhaustiva los estudios previos pertinentes para determinar sus ventajas, desventajas y conclusiones principales.

Con el uso de estas técnicas combinadas, podremos obtener una comprensión profunda del cuerpo de trabajo ya realizado en el dominio de la preparación de exámenes de física y matemáticas utilizando el procesamiento natural.

Al combinar estos enfoques, tiene como objetivo crear un marco de referencia confiable que pueda usarse para sugerir nuevas direcciones de estudio y aumentar nuestra comprensión de

esta área. Esto hará posible no solo reconocer los patrones actuales, sino también detectar el potencial para mejorar y ampliar el uso del procesamiento del lenguaje natural en la creación de pruebas de física y matemáticas.

En las etapas siguientes, nos sumergiremos en una descripción detallada de la metodología que implementaremos, destacando las respectivas etapas de: método PICOC, los criterios de inclusión, las interrogantes de investigación y el enfoque de búsqueda. Cada una de estas etapas ejercen un papel vital en el proceso de investigación.

2. Metodología

Para la realización de esta investigación, se emplearon dos metodologías fundamentales: el mapeo y la revisión sistemáticos de la literatura, en conjunto con la metodología PICOC [6]. El objetivo de aplicar estos enfoques era obtener una comprensión completa y profunda del estado del campo en estudio, particularmente con respecto al uso del procesamiento natural en la creación de pruebas matemáticas y físicas.

Ambos abordajes se realizaron de acuerdo con estrictas normas y procedimientos, lo que aseguró la objetividad y repetibilidad del proceso. Para asegurarse de que los artículos incluidos en el análisis fueran de alto calibre y relevancia, se realizaron búsquedas exhaustivas en bases de datos académicas. También se aplicaron estrictos criterios de selección.

Conseguimos compilar una descripción general exhaustiva y bien investigada del cuerpo de literatura en el área temática al combinar el mapeo sistemático, las revisiones sistemáticas de literatura y el enfoque PICOC. Esto sirvió como un buen punto de partida para realizar un análisis crítico, llegar a conclusiones y determinar posibles direcciones de estudio futuras.

Primero, exploraremos el método PICOC, que nos ayuda a definir claramente los elementos clave de nuestra investigación, como los participantes, intervenciones, comparaciones, resultados y contexto. A continuación, abordaremos los criterios de elegibilidad, que nos permiten establecer los criterios que deben cumplir los estudios incluidos en nuestra revisión. Luego, nos enfocaremos en las preguntas de investigación, que guiarán nuestra indagación y nos ayudarán a obtener respuestas significativas. Por último, nos sumergiremos en la estrategia de búsqueda, que nos permitirá localizar y seleccionar las fuentes de información relevantes para nuestro estudio. Al comprender en detalle cada una de estas etapas, estaremos preparados para llevar a cabo una investigación rigurosa y completa.

Etapa 1.

En esta etapa, el método PICOC se estableció como estrategia de investigación para obtener artículos sobre el uso del NLP en la educación para la elaboración de pruebas, y se establecieron las características de búsqueda. (ver tabla 1).

Tabla 1: Método Picoc.

PICOC	Consideraciones
Población(P): ¿Quién?	Procesamiento de Lenguaje Natural (NLP) en la educación o en las áreas de física y matemática.
Intervención(I): ¿Qué?, ¿Cómo?	Técnicas de NLP usadas en la educación.
Comparación (C): ¿Con qué comparar?	Estudios que presenten las técnicas de implementación de NLP para elaborar evaluaciones.
Resultados (O): ¿Qué se busca conseguir/mejorar?	La elaboración de evaluaciones que usen NLP para hacerlo más rápido.
Contexto (C): ¿En qué tipo de organización y bajo qué circunstancias?	En la educación en el área de física y matemática para mejorar la elaboración de evaluaciones.

Posteriormente, los términos fueron establecidos y utilizados en la forma en que se organizan las cadenas de búsqueda (ver tabla 2).

Las expresiones booleanas frecuentes como "AND" y "OR" se utilizaron para obtener esta cadena de búsqueda.

La cadena de búsqueda que se empleó: (natural language processing OR NLP) AND (physics OR math) AND (evaluation OR test) AND (education) es una combinación de términos clave relacionados con el tema de interés. A continuación, se presenta una justificación para cada uno de los componentes de la cadena:

1. "natural language processing OR NLP": Esta parte de la cadena se refiere al procesamiento de lenguaje natural (NLP) y busca incluir estudios que se centren en técnicas y enfoques de NLP aplicados a la elaboración de pruebas. Al incluir "natural language processing" o "NLP" como términos alternativos, se amplía la cobertura de la búsqueda para incluir estudios que utilicen diferentes terminologías.

2. "physics OR math": Esta sección de la cadena se refiere a los campos de física y matemáticas. Al incluir "physics" o "math" como términos alternativos, se busca abarcar estudios que se centren en la elaboración de pruebas en cualquiera de estos dos campos.

3. "evaluation OR test": Aquí se incluyen los términos "evaluation" y "test", que se refieren a la evaluación y las pruebas en el contexto de la elaboración de pruebas utilizando el procesamiento natural. Se busca encontrar estudios que se centren en la evaluación y el análisis de pruebas generadas mediante técnicas de NLP.

4. "education": Esta última parte de la cadena se refiere al ámbito educativo. Al incluir este término, se busca encontrar estudios que se centren en la aplicación de la elaboración de pruebas con NLP en el contexto educativo, lo que

puede incluir la educación en física y matemáticas.

La combinación de estos términos en la cadena de búsqueda permite enfocar la búsqueda en estudios que aborden la elaboración de pruebas utilizando el procesamiento natural en el ámbito de la física y las matemáticas, con un enfoque en la evaluación y el contexto educativo.

Tabla 2: Términos Para Utilización De Búsqueda.

Términos		Términos Semejantes
Natural processing	language	NLP
Physics		Math
Evaluation		test
Education		

Criterios de elegibilidad

Estos términos sobre inclusión y exclusión en metodologías SMS y SLR se utilizaron para seleccionar los estudios más pertinentes. Estos criterios se detallan a continuación:

Criterios de inclusión: Los tipos de grupos temporales que componen la población de un estudio se conocen como criterios de inclusión [7]. Se incluyeron en el análisis de artículos que cumplan con los siguientes criterios de selección: dispongan un DOI, acceso en formato digital en línea y presentación de evidencia empírica sobre el uso de NLP en educación (ver tabla 3).

Criterios de exclusión: Se excluyeron aquellos tipos de datos que presentaban dificultades para la extracción de información, de acuerdo con los criterios de exclusión establecidos. Esto incluyó artículos que no estaban escritos en inglés o que tenían menos de 10 páginas de extensión. (ver tabla 3).

Etapa 2.

Preguntas de investigación: Se ha realizado una investigación acerca de cómo la utilización del procesamiento natural puede contribuir en la

generación de pruebas en los ámbitos de las matemáticas y la física. Realizando dos tipos de preguntas, en el ámbito de Mapeo Sistemático (SMS) y Revisión de Literatura (SLR), fueron creados para dirigir y definir esta investigación. Seguidamente, se exponen las interrogantes que han surgido durante el desarrollo de la investigación.

- *SMS1:* ¿Cuáles son las técnicas utilizadas en el procesamiento de lenguaje natural para la elaboración de pruebas en el ámbito de la física y las matemáticas?

- *SMS2:* ¿Cuál es la efectividad del análisis de respuestas en la evaluación de los estudiantes mediante el uso de técnicas de procesamiento de lenguaje natural en pruebas de física y matemáticas?

- *SLRS1:* ¿Cuál es el estado actual de la investigación sobre la elaboración de pruebas utilizando el procesamiento de lenguaje natural en el área de física y matemática?

- *SLRS2:* ¿Cuál es la razón por la que se busca automatizar la elaboración de evaluaciones utilizando el procesamiento de lenguaje natural en el área de física y matemática?

Tabla 3: Criterios De Elegibilidad

Tipo	Inclusión	Exclusión
Mapeo Sistemático	Artículos digitales en línea, publicado en los archivos de repositorios ACM, IEEE, Science Direct durante el período de 2020 a 2023, artículos en línea.	Artículos publicados antes al 2020.
	Recopilación exhaustiva de documentos en inglés. dispongan DOI.	Recopilación de investigaciones y análisis en diversos campos del conocimiento, presentados en lenguajes diferentes al inglés.
	Artículos que tienen acceso completo a su información.	Investigaciones secundarias, como informes o libros. Artículos científicos de pago o sin estructura de investigación.
Revisión Sistemática Literatura	Artículos con keywords relacionadas con su título y resumen.	Artículos que no proporcionaban palabras clave o términos relacionados, así como aquellos que tenían menos de 10 páginas.
	Investigaciones sobre el uso del Procesamiento Natural de Lenguaje (NLP) en el área educativo, específicamente en el contexto de evaluaciones o pruebas.	Investigaciones que no se enfocan en la creación de pruebas en física y las matemáticas utilizando el procesamiento natural de lenguaje.
	Artículos que apliquen pruebas elaboradas mediante el procesamiento natural de lenguaje en el ámbito de la física y las matemáticas.	Estudios que no contemplen el uso de pruebas elaboradas mediante el procesamiento natural de lenguaje en el ámbito de la física y las matemáticas.

Etapa 3.

Estrategias de búsqueda: ACM, ScienceDirect, IEEE y Web of Science fueron los cuatro repositorios científicos elegidos previamente que se utilizaron en el desarrollo del método de búsqueda para este estudio. Estos archivos fueron escogidos debido a sus extensos fondos y su aplicabilidad al tema de estudio.

La selección de estos repositorios se basó en la cantidad de artículos disponibles para revisión sistemática (SLR) y mapeo sistemático (SMS). Se consideró crucial reunir un número considerable de publicaciones científicas pertinentes para garantizar la representatividad y exhaustividad de la investigación.

Se utilizaron varios criterios durante el proceso de búsqueda para perfeccionar los resultados y

obtener la información que queríamos. Estas restricciones incluyeron establecer un rango de fechas pertinente, limitar la búsqueda a literatura científica revisada por pares y elegir idiomas particulares para incluir en la búsqueda.

Mediante el uso de estos filtros, fue posible garantizar el calibre y la relevancia de los artículos recopilados y reducir la posibilidad de encontrar material redundante o no relacionado. Se adquirió una colección de información correcta y pertinente para un análisis e interpretación.

- **Filtro 1.** Para el primer filtro se realizó una recopilación exhaustiva de 3.510 publicaciones e investigaciones que cumplieran con la cadena de búsqueda establecida. En este paso inicial, se realizó una búsqueda exhaustiva que abarcó los años 2020 a 2023. El criterio principal fue que los documentos se publicaran en inglés, y

se dio prioridad a los presentados en conferencias.

Además, se definió como criterio de inclusión que las publicaciones e investigaciones sean de libre acceso y que todo su contenido sea accesible. Tras la aplicación de estos criterios, se obtuvieron un total de 1.607 estudios que cumplieran con todas las condiciones definidas. Dentro del abanico científico abordado, estos estudios representan una selección rigurosa y relevante.

• **Filtro 2.** En este punto, Se creó una cadena de búsqueda precisa seleccionando cuidadosamente las palabras clave más relevantes en los títulos y resúmenes de los documentos. Se utilizó el software Mendeley para procesar y filtrar los artículos e investigaciones, este procedimiento condujo a una reducción considerable en el conjunto inicial de artículos, dando como resultado 141 investigaciones en total que coincidían con los criterios predeterminados.

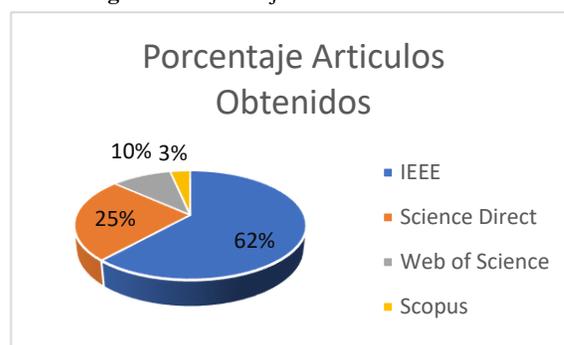
• **Filtro 3.** El tercer filtro utilizado en el procedimiento de selección y análisis de estudios y artículos. Con base en los resúmenes de artículos recopilados, cada estudio y artículo se descargó y revisó individualmente en este punto. El propósito principal de este filtro fue identificar los temas más pertinentes a nuestra investigación acerca de la aplicación del procesamiento del lenguaje natural en las evaluaciones de física y matemáticas.

Esperamos obtener una visión general de cada estudio y establecer su relevancia y utilidad para nuestra investigación leyendo y analizando los resúmenes. Para ello, se aplicaron criterios de inclusión y exclusión. Aquellos artículos cuyos resúmenes demostraron una clara asociación con nuestra área de interés, es decir, la elaboración de pruebas mediante el procesamiento natural de lenguaje en el ámbito de la física y las matemáticas, fueron seleccionados para una

investigación adicional. Por otro lado, se excluyeron de nuestro análisis los artículos que no cumplieran con los criterios de inclusión, como aquellos que no estaban directamente relacionados con el tema de estudio.

Una vez completada la selección, se leyeron detenidamente todas las publicaciones e investigaciones. Los enfoques metodológicos, los resultados producidos y las conclusiones de cada publicación se examinaron minuciosamente durante este proceso para extraer material pertinente a nuestra investigación sobre el uso del procesamiento del lenguaje natural en las evaluaciones de física y matemáticas.

Figura 1: Porcentaje de artículos obtenidos.



La figura 1 muestra los repositorios y el porcentaje de documentos de cada uno. Tras completar exhaustivamente la búsqueda, se lograron recopilar un total de 60 artículos relevantes.

TABLE 4: cadena de búsqueda y filtros.

Repositorio	Cadena de búsqueda	tipo de artículo	Resultados de cadena	Filtro 1	Filtro 2	Filtro 3
IEEE	"Natural Language Processing" OR "NLP" AND "Physics" OR "Math" AND "Evaluation" OR "Test" AND "Education"	Artículos y Conferencias	1238	267	85	37
Science Direct	"Natural Language Processing OR NLP" AND "Physics OR Math" AND "Evaluation OR Test" AND "Education"	Artículos y Conferencias	2215	602	31	15
Web of Science	(Natural Language Processing OR NLP) AND (Physics OR Math) AND (Evaluation OR Test) AND (Education)	Artículos y Conferencias	49	31	12	6
Scopus	(natural language processing OR NLP) AND (physics OR math) AND (evaluation OR test) AND (education)	Artículos	8	7	5	2
TOTAL			3510	1607	141	60

La identificación de palabras clave importantes en los resúmenes se realiza a medida que se leen los artículos. Estas palabras clave se organizan luego en temas pertinentes para que sea más fácil elegir el contenido más pertinente e importante. Vale la pena

mencionar varias palabras clave que se descubrieron a lo largo de esta investigación y que fueron cruciales para el proceso de recopilación de información. Estas palabras han sido reconocidas, con algunos ejemplos:

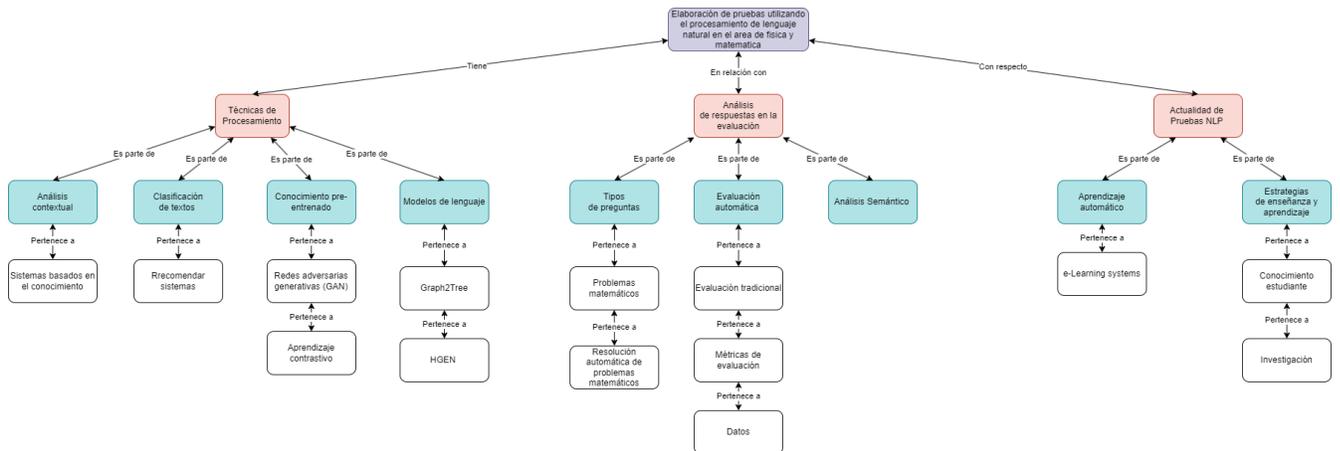
Tabla 5: Artículos recolectados a través de la biblioteca virtual.

N.º	Palabras	Frecuencia	Numero de Artículos	N.º	Palabras	Frecuencia	Numero de Artículos
1	model	3028	58	12	student	1085	42
2	based	2676	55	13	proposed	1070	42
3	learning	2310	53	14	text	1070	42
4	question	2298	53	15	accuracy	986	38
5	data	1933	52	16	methods	985	38
6	answer	1767	49	17	system	961	38
7	knowledge	1604	49	18	different	925	35
8	word	1439	47	19	approach	910	30
9	language	1416	45	20	training	908	28
10	performance	1139	45	21	research	903	25
11	information	1096	43				

El esquema de clasificación se creó a partir de las palabras clave mencionadas anteriormente

y presenta tres categorías, como se visualiza en la Figura 2.

Figura 2. Esquemas de clasificación.



Esta sección muestra los análisis de la información en los documentos de investigación clave. Se utilizó el programa ATLAS.ti para procesar la información, lo que permitió realizar una lista de palabras clave (tabla 5) relacionadas con los artículos elegidos. Estas palabras clave se dividieron en tres categorías principales: técnicas de procesamiento, análisis de respuestas en la evaluación y actualidad de las pruebas de procesamiento natural.

Sistemas basados en el conocimiento:

Según los estudios, “los sistemas de inteligencia artificial utilizan información contextual y conocimiento previo para realizar tareas y tomar decisiones” [7]. Estos sistemas emplean técnicas de procesamiento natural para comprender el significado de las palabras y frases utilizadas en conversaciones y preguntas, entendiendo así su contexto. El análisis contextual implica comprender el contexto en el que se realiza una tarea o se formula una pregunta, lo que puede mejorar la precisión y relevancia de las respuestas [8]. En el artículo "Efficient SPARQL Queries Generator for Question Answering Systems" [9] Propone el uso de sistemas basados en conocimiento con análisis contextual para aumentar la precisión del proceso de clasificación de textos cortos en redes sociales, la creación de consultas SPARQL a partir de preguntas de lenguaje natural y la respuesta a preguntas

complejas relacionadas con una tabla específica [9].

Conocimiento pre-entrenado:

Nuestra investigación ha revelado una relación importante entre el conocimiento pre-entrenado, las redes adversarias generativas y el aprendizaje contrastivo en el campo del aprendizaje automático [10]. Al combinar estos enfoques, se logra mejorar tanto la precisión como la eficiencia de los modelos de aprendizaje automático. Los conocimientos pre-entrenados brindan una base sólida para comprender los datos, mientras que las redes adversarias generativas generan datos sintéticos realistas para ampliar el conjunto de entrenamiento. Por su parte, el aprendizaje contrastivo permite aprender representaciones más discriminativas y robustas [11]. Esta combinación de enfoques potencia el rendimiento del modelo en diversas tareas de aprendizaje automático [12].

Tipos de preguntas:

La investigación ha demostrado que estos conceptos son valiosos en una amplia gama de aplicaciones, aunque su efectividad puede ser aún mayor en problemas que se puedan expresar clara y concisamente en lenguaje natural [13]. Estos enfoques pueden ayudar a comprender y representar mejor los problemas matemáticos, lo que a su vez permite una resolución más eficiente

y precisa mediante técnicas de aprendizaje automático. La combinación de estos campos abre la puerta a soluciones automatizadas más sólidas y eficaces en el ámbito matemático [11].

HGEN: La investigación y posterior integración de HGEN, Graph2Tree y Language Models en el ámbito de la generación de texto automatizada dio como resultado un análisis exhaustivo de las necesidades y desafíos actuales en el procesamiento natural. Inicialmente, se identificó la necesidad de modelos que pudieran capturar y representar de manera efectiva las relaciones complejas entre palabras y conceptos en un texto. A partir de ahí, se exploraron diferentes enfoques y técnicas, y se encontró que HGEN, un modelo basado en grafos heterogéneos era capaz de capturar estas relaciones de manera más precisa[14]. Sin embargo, se observó que HGEN tenía limitaciones en la generación de estructuras sintácticas adecuadas[14]. Para abordar este problema, se investigó Graph2Tree, un modelo que convertía las estructuras de grafos en árboles sintácticos, mejorando así la coherencia y fluidez del texto generado[15]. Finalmente, se reconoció la importancia de los Language Models, especialmente aquellos basados en transformers, para capturar patrones y contextos lingüísticos de manera más precisa. La integración de estos tres enfoques permitió desarrollar un modelo más completo y efectivo para la generación de texto automatizada, abriendo nuevas posibilidades en áreas como la redacción de informes, la traducción automática y la generación de contenido creativo.

E-learning Systems: Los investigadores comprendieron que mejorar la generación automatizada de evaluaciones en los sistemas de e-Learning y, en busca de una solución efectiva, consideraron la aplicación del machine learning. Durante este proceso, se examinaron meticulosamente tanto las ventajas como los desafíos asociados con la integración de técnicas de machine learning en los sistemas de e-Learning, y se identificaron las áreas específicas en las que esta sinergia podría tener una

influencia significativa. posee un gran potencial para mejorar sustancialmente la generación de evaluaciones personalizadas y adaptativas en los entornos virtuales de aprendizaje[11]. Esta sinergia permite una mayor eficiencia y precisión en la creación de evaluaciones adaptadas a las necesidades y cualidades únicas de cada estudiante, optimizando así su proceso de aprendizaje en entornos virtuales[16].

A través de la exhaustiva aplicación de la SLR, se ha identificado una alta frecuencia en la aparición de varios términos clave en el campo de estudio. Estos términos incluyen la "Clasificación de textos", "Modelos de lenguaje", "Sistemas basados en el conocimiento" y las "Redes adversarias generativas (GAN)". Estos hallazgos se presentan y se detallan en la Tabla 6, proporcionando una visión clara y concisa de la frecuencia y relevancia de estos términos dentro del corpus de literatura revisado.

Etapa 4.

Extracción de Datos.

Durante esta etapa de extracción de los artículos recopilados, se dirige la atención hacia distintas secciones clave, como los resúmenes, introducciones y conclusiones. Estas secciones son analizadas en función de los conceptos formados a partir de las palabras clave previamente identificadas. Una vez que tenemos la información necesaria, podemos sintetizar y comparar los datos para llevar a cabo la resolución a las preguntas planteadas de manera precisa y fundamentada.

1. Técnicas de Procesamiento: El término "técnicas de procesamiento de lenguaje natural se refiere a un conjunto de métodos y algoritmos utilizados para analizar, comprender y manipular el lenguaje humano en forma de texto" [17]. Estas técnicas permiten a las computadoras procesar y entender el lenguaje del ser humano de manera similar a como lo hacen los humanos.

1.1 Análisis contextual: Estos sistemas están basados en el conocimiento, aquellos que realizan tareas de análisis y toma de decisiones utilizando una base de conocimientos explícita en el contexto del análisis contextual [11]. Estos sistemas se basan en la noción de que las máquinas pueden representar y usar el conocimiento explícito para efectuar tareas complejas más eficiente y precisa. Los sistemas basados en el conocimiento pueden utilizarse en el análisis contextual para realizar tareas de clasificación y categorización de datos, así como para reconocer patrones y relaciones en grandes cantidades de datos.

1.2 La clasificación de textos: Es un enfoque que “busca comprender el significado de una palabra o frase en función de la situación en la que se usa. En lugar de simplemente analizar la palabra por sí sola, el análisis contextual considera las palabras que la rodean y cómo se utilizan para determinar su significado” [17]. Se tiene como objetivo comprender una palabra o declaración para determinar su significado. El análisis contextual considera las palabras circundantes y su uso para definir una palabra en lugar de simplemente la palabra en sí.

1.3 Conocimiento pre-entrenado: La clasificación de textos relacionada con los sistemas de recomendación se refiere al proceso de etiquetar o clasificar los textos con el objetivo de organizarlos y facilitar la recomendación de contenido relevante a los usuarios [18]. En este contexto, los sistemas de recomendación utilizan métodos de clasificación de textos para analizar y comprender el contenido de los textos, identificar patrones y características relevantes, y luego utilizar esta información para recomendar contenido similar o relacionado a los usuarios.

1.4 Modelos de lenguaje: Los modelos de lenguaje desempeñan un papel fundamental en el procesamiento natural, permitiendo abordar diversas tareas de manera eficiente y precisa. Entre estos modelos, destacan HGEN y Graph2Tree. HGEN se especializa en resolver

ejercicios matemáticos al emplear una codificación jerárquica de gráficos heterogéneos en lenguaje natural [13]. Su enfoque se centra en enfrentar la complejidad de estos problemas y su representación lingüística. Por su parte, Graph2Tree se dedica a generar árboles sintácticos a partir de oraciones en lenguaje natural, empleando técnicas de aprendizaje automático para capturar la estructura y relación sintáctica. Ambos modelos demuestran la versatilidad y eficiencia del procesamiento natural en una amplia gama de contextos y aplicaciones [19]. Su contribución destaca el potencial de estos modelos para mejorar la comprensión y generación de texto en distintos dominios.

Tabla 6: Técnicas de Procesamiento.

N.º	Técnicas de Procesamiento	Referencia
1	Análisis contextual	[8]
2	Clasificación de textos	[10], [17], [20]–[25]
3	Conocimiento pre-entrenado	[11], [26], [27]
4	Modelos de lenguaje	[15], [21], [23], [26]–[36]
5	Sistemas basados en el conocimiento	[10], [12], [24], [25], [37]–[45]
6	Recomendar sistemas	[26], [38], [46], [47]
7	Redes adversarias generativas (GAN)	[10], [48], [49]
8	Aprendizaje contrastivo	[12]
9	Graph2Tree	[14], [15]
10	HGEN	[14]

2. Análisis de Respuesta en la Evaluación: Se refiere al proceso de clasificación y análisis de las respuestas abiertas proporcionadas por los participantes en una evaluación o estudio de investigación. Este análisis puede involucrar el uso de técnicas de procesamiento natural para convertir las respuestas en datos numéricos que se pueden analizar estadísticamente [20].

2.1 Tipos de preguntas: Se refiere a las diferentes categorías de preguntas que se pueden hacer en un examen o evaluación, como preguntas de opción múltiple, preguntas abiertas, preguntas de verdadero o falso [35].

2.3 Evaluación automática: Se indica al uso de la informática para evaluar las respuestas a preguntas en un examen o evaluación

automáticamente. Esto puede incluir la comparación con una base de datos predefinida y el uso de algoritmos para identificar patrones en las respuestas correctas e incorrectas [50].

2.4 Análisis semántico: Se indica al proceso el cual el ser humano comprende y extrae significado de las conversaciones y preguntas hechas en lenguaje natural [25]. El objetivo del análisis semántico es obtener comprensión más profunda del significado de las palabras y frases utilizadas en las conversaciones y las preguntas, lo que puede ayudar a generar respuestas más precisas y pertinentes.

Tabla 7: Análisis de respuestas en la evaluación

N.º	Análisis de respuestas en la evaluación	Referencia
1	Tipos de preguntas	[35]
2	Evaluación automática	[17], [30], [51], [52]
3	Análisis Semántico	[16], [26], [30], [37], [50]
4	Problemas matemáticos	[15], [30], [35], [53]
5	Métricas de evaluación	[30], [33], [46], [51]
6	Resolución automática de problemas matemáticos	[15], [29]–[31], [51]

3. Actualidad de pruebas NLP:

3.1 Aprendizaje automático: se refiere a un enfoque de IA que, sin ser programadas explícitamente, permite a las máquinas aprender y mejorar automáticamente a través de la experiencia [22].

3.2 Planificación de enseñanza y aprendizaje: métodos, técnicas y enfoques utilizados por los educadores para facilitar el proceso de enseñanza y promover el aprendizaje efectivo de los estudiantes. Estas estrategias pueden incluir actividades interactivas, uso de tecnología, colaboración entre pares, retroalimentación formativa, entre otros enfoques pedagógicos [54].

Tabla 8: Área educativa

N.º	Área educativa	Referencia
1	Aprendizaje automático	[26], [33]
2	Sistemas de e-learning	[16]
3	Estrategias de enseñanza y aprendizaje	[32], [34]
4	Conocimiento estudiante	[12], [40], [43]
5	Investigación	[20], [27], [50], [55], [56]

RESULTADOS Y DISCUSIÓN

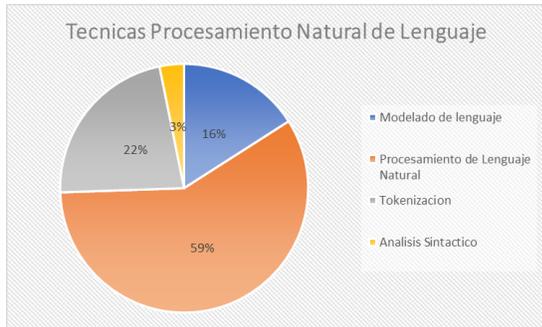
Los resultados revelan una relación sólida entre los documentos analizados en cuanto a las técnicas de procesamiento utilizadas, el análisis de respuestas en la evaluación y la actualidad de las pruebas basadas en NLP. Estos resultados brindan una base sólida para futuras investigaciones y el desarrollo de herramientas y técnicas más avanzadas en el campo de la elaboración de pruebas utilizando el procesamiento natural.

• *SMS1: ¿Cuáles son las técnicas utilizadas en el procesamiento de lenguaje natural para la elaboración de pruebas en el ámbito de la física y las matemáticas?*

Después de investigar sobre las técnicas utilizadas en el procesamiento natural para la elaboración de pruebas en el ámbito de la física y las matemáticas, podemos decir que existen varias técnicas comunes aplicables en este contexto. “Estas técnicas incluyen la tokenización es una forma de dividir un texto en partes más pequeñas como palabras o símbolos matemáticos” [38]. El análisis sintáctico, que facilita la comprensión de la conformación gramatical de una oración o expresión matemática; la extracción de características, que busca palabras clave o conceptos cruciales en una pregunta [57], y “el modelado de lenguaje, que utiliza modelos de aprendizaje automático o estadísticos para predecir la probabilidad de una secuencia de palabras o símbolos” [[19]. Estas técnicas son fundamentales para el procesamiento y comprensión de preguntas en el ámbito de la física y las matemáticas, pero es importante

consultar fuentes adicionales para obtener información más detallada y referencias específicas.

Figura 3. Técnicas Utilizadas en NLP



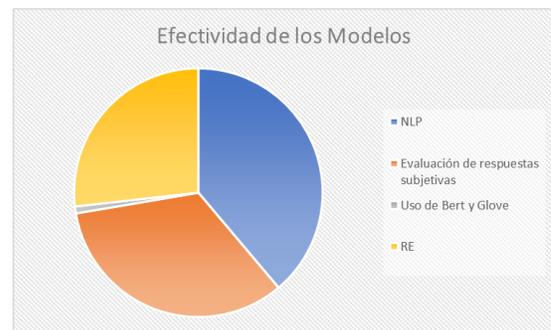
se identificaron diferentes enfoques y técnicas utilizadas (figura 3). Se observó que el análisis sintáctico representa aproximadamente el 3% de los estudios analizados, lo cual destaca la importancia de comprender la estructura gramatical de los textos. La tokenización, por su parte, ocupa un porcentaje significativo del 22%, evidenciando su relevancia en la segmentación de palabras y la generación de unidades de texto más pequeñas. Con un 16% de atención, el modelado de lenguaje es importante en la comprensión y creación de texto coherente. Finalmente, el procesamiento de lenguaje representa el 59% de los estudios analizados, lo que destaca su importancia en la creación y mejora de pruebas mediante técnicas de procesamiento natural de lenguaje. Estos resultados demuestran la diversidad y la importancia de las diversas técnicas utilizadas en el área de estudio, proporcionando una base sólida para el desarrollo de futuras investigaciones.

• *SMS2: ¿Cuál es la efectividad del análisis de respuestas en la evaluación de los estudiantes mediante el uso de técnicas de procesamiento de lenguaje natural en pruebas de física y matemáticas?*

Tras realizar la investigación, se ha concluido que no se han encontrado registros acerca de la efectividad del análisis de respuestas en los campos de física y matemáticas. No obstante, se han identificado métodos potencialmente

adecuados para estos ámbitos. Estas opciones incluyen el procesamiento natural del lenguaje. Es importante destacar que el uso de estas técnicas puede mejorar la eficiencia y la eficacia en las tareas relacionadas con la ingeniería de requisitos (RE) según [19]. Así mismo, [22] menciona una propuesta para evaluar respuestas subjetivas que se compara con el método actual de evaluación manual. Según los resultados experimentales presentados (figura 4), esta nueva metodología demuestra ser aproximadamente tres veces más eficiente en términos de tiempo y muestra una precisión que oscila entre el 75% y el 87.5%, comparado con la versión manual del sistema [18]

Figura 4. Efectividad de Modelos



• *SLRS1: ¿Cuál es el estado actual de la investigación sobre la elaboración de pruebas utilizando el procesamiento de lenguaje natural en el área de física y matemática?*

Se ha realizado un análisis exhaustivo del estado actual de la investigación en relación con la elaboración de pruebas utilizando el procesamiento natural en el área de física y matemática. Realizando una revisión exhaustiva de la literatura y la consulta de una variedad de fuentes académicas. La aplicación del procesamiento de lenguaje natural en la elaboración de pruebas en las disciplinas de física y matemática está ganando popularidad. Se han realizado numerosos estudios que exploran diferentes enfoques y técnicas para aprovechar el poder del procesamiento natural en la generación de preguntas, evaluación automática y personalización de las pruebas. En cuanto a la

aplicación específica del procesamiento natural en el área de física y matemática, se han identificado avances significativos. Se han propuesto modelos y algoritmos que permiten analizar y comprender el lenguaje natural presente en los enunciados de problemas. Sin embargo, a pesar de los avances logrados, aún existen desafíos por superar. Algunos de estos desafíos incluyen la adaptación de los modelos de procesamiento de lenguaje natural a la complejidad y especificidad de los conceptos de física y matemática.

• *SLRS2: ¿Cuál es la razón por la que se busca automatizar la elaboración de evaluaciones utilizando el procesamiento de lenguaje natural en el área de física y matemática?*

Eficiencia y ahorro de tiempo: La automatización de la “elaboración de evaluaciones utilizando el procesamiento de lenguaje natural permite generar rápidamente un gran número de preguntas y ejercicios”[53]. Esto ahorra tiempo a los educadores y facilita la preparación de evaluaciones completas y variadas.

Personalización y adaptabilidad: La generación automatizada de evaluaciones “basada en el procesamiento de lenguaje natural permite adaptar las preguntas y ejercicios según las necesidades individuales de los estudiantes” [27]. Esto facilita la creación de evaluaciones personalizadas que se ajusten a los niveles de habilidad y conocimiento de cada estudiante.

Mejora de la retroalimentación: Al utilizar “el procesamiento de lenguaje natural, se pueden generar respuestas automáticas y retroalimentación inmediata para los estudiantes” [29]. Esto les permite recibir comentarios instantáneos sobre su desempeño y les ayuda a identificar áreas de mejora.

Estandarización y objetividad: La automatización de la elaboración de evaluaciones utilizando “el procesamiento de lenguaje natural garantiza una mayor objetividad y estandarización en la evaluación” [22]. Al

eliminar la influencia subjetiva de los evaluadores humanos, se reduce el sesgo y se obtiene una evaluación más imparcial y consistente.

Escalabilidad y accesibilidad: La automatización permite escalar el proceso de elaboración de evaluaciones, lo que es especialmente beneficioso en entornos educativos con un gran número de estudiantes [36]. Además, el uso del procesamiento natural puede hacer que las evaluaciones sean más accesibles para estudiantes con diferentes estilos de aprendizaje o necesidades especiales.

Estas son algunas de las razones por las que se busca automatizar la elaboración de evaluaciones utilizando el procesamiento natural en el área de física y matemática. Es importante destacar que esta investigación continúa en desarrollo y se exploran nuevas aplicaciones y ventajas de la automatización en la educación.

CONCLUSIONES

Existen diversas técnicas de procesamiento natural de lenguaje aplicadas en la elaboración de pruebas en física y matemáticas. Entre ellas, se destacan el análisis sintáctico, la tokenización, el modelado de lenguaje y el procesamiento de lenguaje en general. Estas técnicas son fundamentales para comprender la estructura del texto, segmentar palabras, generar texto coherente y mejorar la calidad de las pruebas.

La automatización de la elaboración de evaluaciones mediante el uso del procesamiento natural es una tendencia creciente en el ámbito educativo. Los investigadores han reconocido la necesidad de mejorar los sistemas de evaluación en entornos de aprendizaje virtual, y el procesamiento de lenguaje natural se presenta como una solución efectiva para lograrlo.

El estado del arte en el procesamiento natural en el ámbito de la física y las matemáticas muestra un campo de investigación en constante progreso y evolución. La mejora de la generación automatizada de evaluaciones y la

personalización del proceso de evaluación en entornos de aprendizaje virtuales son posibles gracias a las técnicas y enfoques analizados. Estas conclusiones brindan una base sólida para futuras investigaciones que buscan apoyar el avance del entendimiento y la mejora de los sistemas de evaluación educativos.

REFERENCIAS

- [1] Torrico, J. M. F., & Zubieta, C. L. F., “La evaluación en el proceso de aprendizaje,” *redalyc.org*, vol. 19, pp. 15–30, 2007.
- [2] A. Santos Del Real, “Evaluación docente / Teacher evaluation.” [Online]. Available: www.sep.gob.mx/es/sep1/C0570511
- [3] Moreira, D., Cruz, I., Gonzalez, K., Quirumbay, A., Magallan, C., Guarda, T., ... & Castillo, C., “Análisis del Estado Actual de Procesamiento de Lenguaje Natural Analysis of the Current State of Natural Language Processing.”
- [4] N. Paternoster, C. Giardino, M. Unterkalmsteiner, T. Gorschek, and P. Abrahamsson, “Software development in startup companies: A systematic mapping study,” *Inf Softw Technol*, vol. 56, no. 10, pp. 1200–1218, 2014, doi: 10.1016/j.infsof.2014.04.014.
- [5] D. Carrizo and C. Moller, “Estructuras metodológicas de revisiones sistemáticas de literatura en Ingeniería de Software: un estudio de mapeo sistemático Methodological structures of systematic literature review in software engineering: a systematic mapping study.”
- [6] T. F. Frandsen, M. F. Bruun Nielsen, C. L. Lindhardt, and M. B. Eriksen, “Using the full PICO model as a search tool for systematic reviews resulted in lower recall for some PICO elements,” *J Clin Epidemiol*, vol. 127, pp. 69–75, Nov. 2020, doi: 10.1016/J.JCLINEPI.2020.07.005.
- [7] L. M. Connelly, “Inclusion and Exclusion Criteria.”
- [8] S. Al Sulaimani and A. Starkey, “Short Text Classification Using Contextual Analysis,” *IEEE Access*, vol. 9, pp. 149619–149629, 2021, doi: 10.1109/ACCESS.2021.3125768.
- [9] Y. H. Chen, E. J. L. Lu, and Y. Y. Lin, “Efficient SPARQL Queries Generator for Question Answering Systems,” *IEEE Access*, vol. 10, no. September, pp. 99850–99860, 2022, doi: 10.1109/ACCESS.2022.3206794.
- [10] Y. Zhao, K. Wang, G. Guo, and X. Wang, “Learning compact yet accurate Generative Adversarial Networks for recommender systems,” *Knowl Based Syst*, vol. 257, p. 109900, 2022, doi: 10.1016/j.knosys.2022.109900.
- [11] W. Yu, Y. Wen, F. Zheng, and N. Xiao, “Improving Math Word Problems with Pre-trained Knowledge and Hierarchical Reasoning,” *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pp. 3384–3394, 2021, doi: 10.18653/v1/2021.emnlp-main.272.
- [12] X. Song, J. Li, Q. Lei, W. Zhao, Y. Chen, and A. Mian, “Bi-CLKT: Bi-Graph Contrastive Learning based Knowledge Tracing,” *Knowl Based Syst*, vol. 241, p. 108274, 2022, doi: 10.1016/j.knosys.2022.108274.
- [13] D. Liu, G. Wang, and J. Yang, “Help Transformer Improve Performance in Automatic Mathematics Word Problem-Solving,” *IEEE Access*, vol. 10, no. November, pp. 123020–123027, 2022, doi: 10.1109/ACCESS.2022.3220777.
- [14] Y. Zhang, G. Zhou, Z. Xie, and J. X. Huang, “HGEN: Learning Hierarchical Heterogeneous Graph Encoding for Math Word Problem Solving,”

- IEEE/ACM Trans Audio Speech Lang Process*, vol. 30, pp. 816–828, 2022, doi: 10.1109/TASLP.2022.3145314.
- [15] Z. Liang, J. Zhang, and X. Zhang, “Analogical Math Word Problems Solving with Enhanced Problem-Solution Association,” *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, pp. 9454–9464, 2022.
- [16] F. Skopljanac-Macina, I. Zakarija, and B. Blaskovic, “Towards Automated Assessment Generation in e-Learning Systems Using Combinatorial Testing and Formal Concept Analysis,” *IEEE Access*, vol. 9, pp. 52957–52976, 2021, doi: 10.1109/ACCESS.2021.3070510.
- [17] A. Hadifar, S. K. Bitew, J. Deleu, C. Delder, and T. Demeester, “EduQG: A Multi-Format Multiple-Choice Dataset for the Educational Domain,” *IEEE Access*, vol. 11, no. March, pp. 20885–20896, 2023, doi: 10.1109/ACCESS.2023.3248790.
- [18] M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska, and S. S. Band, “Subjective Answers Evaluation Using Machine Learning and Natural Language Processing,” *IEEE Access*, vol. 9, pp. 158972–158983, 2021, doi: 10.1109/ACCESS.2021.3130902.
- [19] R. Sonbol, G. Rebdawi, and N. Ghneim, “The Use of NLP-Based Text Representation Techniques to Support Requirement Engineering Tasks: A Systematic Mapping Review,” *IEEE Access*, vol. 10, pp. 62811–62830, 2022, doi: 10.1109/ACCESS.2022.3182372.
- [20] J. Wilson, B. Pollard, J. M. Aiken, M. D. Caballero, and H. J. Lewandowski, “Classification of open-ended responses to a research-based assessment using natural language processing,” *Phys Rev Phys Educ Res*, vol. 18, no. 1, p. 10141, 2022, doi: 10.1103/PhysRevPhysEducRes.18.010141.
- [21] Shen, J. T., Yamashita, M., Prihar, E., Heffernan, N., Wu, X., Graff, B., & Lee, D., *MathBERT: A Pre-trained Language Model for General NLP Tasks in Mathematics Education*, vol. 1, no. 1. Association for Computing Machinery, 2021. [Online]. Available: <http://arxiv.org/abs/2106.07340>
- [22] G. Sneha, G. Shreya Shree, T. Babu, and R. R. Nair, “Evaluation of Subjective Answers Using Machine Learning,” *2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2022*, pp. 1–6, 2022, doi: 10.1109/ICICT55121.2022.10064615.
- [23] S. Matsumori, K. Okuoka, R. Shibata, M. Inoue, Y. Fukuchi, and M. Imai, “Mask and Cloze: Automatic Open Cloze Question Generation Using a Masked Language Model,” *IEEE Access*, vol. 11, no. February, pp. 9835–9850, 2023, doi: 10.1109/ACCESS.2023.3239005.
- [24] C. Liu and X. Xu, “AMFF: A new attention-based multi-feature fusion method for intention recognition,” *Knowl Based Syst*, vol. 233, p. 107525, 2021, doi: 10.1016/j.knosys.2021.107525.
- [25] V. Udandarao, A. Agarwal, A. Gupta, and T. Chakraborty, “InPHYNet: Leveraging attention-based multitask recurrent networks for multi-label physics text classification,” *Knowl Based Syst*, vol. 211, p. 106487, 2021, doi: 10.1016/j.knosys.2020.106487.
- [26] M. Zulqarnain, R. Ghazali, M. G. Ghouse, N. A. Husaini, A. K. Z. Alsaedi, and W. Sharif, “A comparative analysis on question classification task based on deep learning approaches,” *PeerJ Comput Sci*, vol. 7, pp. 1–27, 2021, doi: 10.7717/PEERJ-CS.570.

- [27] J. Schneider, R. Richner, and M. Riser, “Towards Trustworthy AutoGrading of Short, Multi-lingual, Multi-type Answers,” *Int J Artif Intell Educ*, vol. 33, no. 1, pp. 88–118, 2023, doi: 10.1007/s40593-022-00289-z.
- [28] P. K. Muhuri and P. K. Gupta, “A novel solution approach for multiobjective linguistic optimization problems based on the 2-tuple fuzzy linguistic representation model,” *Applied Soft Computing Journal*, vol. 95, p. 106395, 2020, doi: 10.1016/j.asoc.2020.106395.
- [29] P. Wulff, D. Buschhüter, A. Westphal, L. Mientus, A. Nowak, and A. Borowski, “Bridging the Gap Between Qualitative and Quantitative Assessment in Science Education Research with Machine Learning — A Case for Pretrained Language Models-Based Clustering,” *J Sci Educ Technol*, vol. 31, no. 4, pp. 490–513, 2022, doi: 10.1007/s10956-022-09969-w.
- [30] K. Kim and C. Chun, “Synthetic Data Generator for Solving Korean Arithmetic Word Problem,” *Mathematics*, vol. 10, no. 19, 2022, doi: 10.3390/math10193525.
- [31] P. Wulff, A. Westphal, L. Mientus, A. Nowak, and A. Borowski, “Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers’ written reflections,” *Front Educ (Lausanne)*, vol. 7, no. January, pp. 1–18, 2023, doi: 10.3389/educ.2022.1061461.
- [32] J. Bernard, E. Popescu, and S. Graf, “Improving online education through automatic learning style identification using a multi-step architecture with ant colony system and artificial neural networks,” *Appl Soft Comput*, vol. 131, p. 109779, 2022, doi: 10.1016/j.asoc.2022.109779.
- [33] M. A. Kia, A. Garifullina, M. Kern, J. Chamberlain, and S. Jameel, “Adaptable Closed-Domain Question Answering Using Contextualized CNN-Attention Models and Question Expansion,” *IEEE Access*, vol. 10, pp. 45080–45092, 2022, doi: 10.1109/ACCESS.2022.3170466.
- [34] Q. Duan, J. Huang, and H. Wu, “Contextual and Semantic Fusion Network for Multiple-Choice Reading Comprehension,” *IEEE Access*, vol. 9, pp. 51669–51678, 2021, doi: 10.1109/ACCESS.2021.3068993.
- [35] L. Araujo, F. Lopez-Ostenero, J. Martinez-Romo, and L. Plaza, “Deep-Learning Approach to Educational Text Mining and Application to the Analysis of Topics’ Difficulty,” *IEEE Access*, vol. 8, pp. 218002–218014, 2020, doi: 10.1109/ACCESS.2020.3042099.
- [36] V. Kumar, D. Reforgiato Recupero, R. Helaoui, and D. Riboni, “K-LM: Knowledge Augmenting in Language Models Within the Scholarly Domain,” *IEEE Access*, vol. 10, no. September, pp. 91802–91815, 2022, doi: 10.1109/ACCESS.2022.3201542.
- [37] Y. Wen, X. Zhu, and L. Zhang, “CQACD: A Concept Question-Answering System for Intelligent Tutoring Using a Domain Ontology with Rich Semantics,” *IEEE Access*, vol. 10, no. May, pp. 67247–67261, 2022, doi: 10.1109/ACCESS.2022.3185400.
- [38] Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P., & Galligan, L., “A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis,” *IEEE Access*, vol. 10, pp. 56720–56739, 2022, doi: 10.1109/ACCESS.2022.3177752.

- [39] A. Balali, M. Asadpour, R. Campos, and A. Jatowt, “Joint event extraction along shortest dependency paths using graph convolutional networks,” *Knowl Based Syst*, vol. 210, p. 106492, 2020, doi: 10.1016/j.knosys.2020.106492.
- [40] Yang, H., Qi, T., Li, J., Guo, L., Ren, M., Zhang, L., & Wang, X., “A novel quantitative relationship neural network for explainable cognitive diagnosis model,” *Knowl Based Syst*, vol. 250, p. 109156, 2022, doi: 10.1016/j.knosys.2022.109156.
- [41] K. Chrysafiadi, S. Papadimitriou, and M. Virvou, “Cognitive-based adaptive scenarios in educational games using fuzzy reasoning,” *Knowl Based Syst*, vol. 250, p. 109111, 2022, doi: 10.1016/j.knosys.2022.109111.
- [42] S. Zhu, X. Cheng, and S. Su, “Conversational semantic parsing over tables by decoupling and grouping actions,” *Knowl Based Syst*, vol. 204, p. 106237, 2020, doi: 10.1016/j.knosys.2020.106237.
- [43] Y. Ren, K. Liang, Y. Shang, and Y. Zhang, “MulOER-SAN: 2-layer multi-objective framework for exercise recommendation with self-attention networks,” *Knowl Based Syst*, vol. 260, p. 110117, 2023, doi: 10.1016/j.knosys.2022.110117.
- [44] X. Zhan, Y. Huang, X. Dong, Q. Cao, and X. Liang, “PathReasoner: Explainable reasoning paths for commonsense question answering,” *Knowl Based Syst*, vol. 235, p. 107612, 2022, doi: 10.1016/j.knosys.2021.107612.
- [45] Su, Y., Cheng, Z., Luo, P., Wu, J., Zhang, L., Liu, Q., & Wang, S., “Time-and-Concept Enhanced Deep Multidimensional Item Response Theory for interpretable Knowledge Tracing,” *Knowl Based Syst*, vol. 218, 2021, doi: 10.1016/j.knosys.2021.106819.
- [46] H. A. Ahmed, N. Z. Bawany, and J. A. Shamsi, “Capbug-a framework for automatic bug categorization and prioritization using nlp and machine learning algorithms,” *IEEE Access*, vol. 9, pp. 50496–50512, 2021, doi: 10.1109/ACCESS.2021.3069248.
- [47] L. Q. Cai, M. Wei, S. T. Zhou, and X. Yan, “Intelligent Question Answering in Restricted Domains Using Deep Learning and Question Pair Matching,” *IEEE Access*, vol. 8, pp. 32922–32934, 2020, doi: 10.1109/ACCESS.2020.2973728.
- [48] K. Palasundram, N. Mohd Sharef, K. A. Kasmiran, and A. Azman, “Enhancements to the Sequence-to-Sequence-Based Natural Answer Generation Models,” *IEEE Access*, vol. 8, pp. 45738–45752, 2020, doi: 10.1109/ACCESS.2020.2978551.
- [49] H. Wen, Y. Wu, J. Li, C. Yang, H. Duan, and Y. Yang, “Transferring Inter-Class Correlation for Teacher–Student frameworks with flexible models,” *Knowl Based Syst*, vol. 242, p. 108316, 2022, doi: 10.1016/j.knosys.2022.108316.
- [50] H. A. Abdeljaber, “Automatic Arabic Short Answers Scoring Using Longest Common Subsequence and Arabic WordNet,” *IEEE Access*, vol. 9, no. 1, pp. 76433–76445, 2021, doi: 10.1109/ACCESS.2021.3082408.
- [51] Q. Wu, Q. Zhang, and X. Huang, “Automatic Math Word Problem Generation with Topic-Expression Co-Attention Mechanism and Reinforcement Learning,” *IEEE/ACM Trans Audio Speech Lang Process*, vol. 30, pp. 1061–1072, 2022, doi: 10.1109/TASLP.2022.3155284.
- [52] A. Praveen Kumar, A. Nayak, K. Manjula Shenoy, R. J. Manoj, and A. Priyadarshi, “Pattern-Based Syntactic Simplification of Compound and Complex Sentences,” *IEEE Access*, vol.

- 10, pp. 53290–53306, 2022, doi: 10.1109/ACCESS.2022.3174846.
- [53] Y. Xu, R. Smeets, and R. Bidarra, “Procedural generation of problems for elementary math education,” *International Journal of Serious Games*, vol. 8, no. 2, pp. 49–66, 2021, doi: 10.17083/ijsg.v8i2.396.
- [54] T. Steuer, A. Filighera, and T. Tregel, “Investigating Educational and Noneducational Answer Selection for Educational Question Generation,” *IEEE Access*, vol. 10, pp. 63522–63531, 2022, doi: 10.1109/ACCESS.2022.3180838.
- [55] N. B. Chaudhuri and D. Dhar, “Designing deep-network based novelty assessment model in Design education,” *Appl Soft Comput*, vol. 134, p. 109966, 2023, doi: 10.1016/j.asoc.2022.109966.
- [56] S. Shang, J. Liu, and Y. Yang, “Multi-Layer Transformer Aggregation Encoder for Answer Generation,” *IEEE Access*, vol. 8, pp. 90410–90419, 2020, doi: 10.1109/ACCESS.2020.2993875.
- [57] M. R. A. H. Rony, U. Kumar, R. Teucher, L. Kovriguina, and J. Lehmann, “SGPT: A Generative Approach for SPARQL Query Generation from Natural Language Questions,” *IEEE Access*, vol. 10, no. April, pp. 70712–70723, 2022, doi: 10.1109/ACCESS.2022.3188714.