

Test Tema 2: PLN

1. ¿Qué es el Procesamiento del Lenguaje Natural (PLN)?
 - a) Un campo de la física que estudia las ondas de sonido
 - b) Un campo de la medicina que se enfoca en el tratamiento de trastornos del habla
 - c) Un campo de la informática y la inteligencia artificial que estudia cómo interactúan los ordenadores con las lenguas humanas
 - d) Un campo de la ingeniería que se enfoca en la construcción de máquinas que hablan
2. ¿Cuál es el objetivo principal del PLN?
 - a) Construir robots humanoides
 - b) Crear videojuegos con personajes que hablan
 - c) Desarrollar nuevas teorías lingüísticas
 - d) Permitir a las máquinas entender, analizar y generar lenguaje humano de forma efectiva
3. ¿Cuál de las siguientes NO es una tarea común del PLN?
 - a) Clasificación de textos
 - b) Generación de resúmenes
 - c) Traducción de textos
 - d) Diseño de hardware para computadoras
4. ¿Qué es la clasificación de textos?
 - a) Un proceso para convertir texto en audio
 - b) Una tarea para categorizar o predecir la clase de un documento de texto
 - c) Un método para comprimir archivos de texto
 - d) Un algoritmo para traducir texto a diferentes idiomas
5. ¿Cuáles son los dos tipos principales de sistemas de clasificación de textos?
 - a) Sistemas binarios y multiclase
 - b) Sistemas basados en reglas y sistemas basados en aprendizaje automático
 - c) Sistemas supervisados y no supervisados
 - d) Sistemas de código abierto y sistemas propietarios
6. ¿Qué caracteriza a los sistemas de clasificación de textos basados en reglas?
 - a) Utilizan un conjunto de reglas lingüísticas construidas manualmente
 - b) Aprenden a clasificar el texto automáticamente a partir de datos
 - c) Son más fáciles de mantener y ampliar que los sistemas basados en aprendizaje automático
 - d) No requieren de conocimientos profundos en la materia
7. ¿Cuál es una desventaja de los sistemas de clasificación de textos basados en reglas?
 - a) Son más precisos que los sistemas basados en aprendizaje automático
 - b) Son fáciles de implementar
 - c) Requieren profundos conocimientos en la materia y son difíciles de mantener y ampliar
 - d) No consumen mucho tiempo
8. ¿Cómo funcionan los sistemas de clasificación de textos basados en aprendizaje automático?
 - a) Aprenden el mapeo de los datos de entrada con las etiquetas utilizando algoritmos de aprendizaje automático supervisado
 - b) Utilizan reglas lingüísticas predefinidas
 - c) Se basan en la comprensión intuitiva del lenguaje
 - d) No requieren datos etiquetados para entrenar el modelo

9. ¿Cuáles son las dos fases principales de los sistemas de clasificación de textos basados en aprendizaje automático?
- a) Análisis y síntesis
 - b) Entrenamiento y predicción
 - c) Compilación y ejecución
 - d) Codificación y decodificación
10. ¿Qué es el preprocesamiento de datos de texto?
- a) Un proceso para convertir texto en código binario
 - b) Un paso importante para limpiar y preparar los datos de texto para su posterior tratamiento
 - c) Un método para cifrar mensajes de texto
 - d) Un proceso para traducir texto a diferentes idiomas
11. ¿Cuál es el objetivo principal del preprocesamiento de texto?
- a) Hacer que los datos de texto sean más difíciles de analizar
 - b) Introducir ruido en los datos de texto
 - c) Eliminar el ruido y preparar los datos para su análisis
 - d) Aumentar la cantidad de datos de texto
12. ¿Cuál de las siguientes NO es una tarea común en el preprocesamiento de texto?
- a) Tokenización
 - b) Eliminación de palabras vacías (stop words)
 - c) Lematización y stemming
 - d) Traducción de texto a diferentes idiomas
13. ¿Qué es la tokenización?
- a) Un proceso para eliminar palabras vacías del texto
 - b) Un proceso para dividir un texto en unidades más pequeñas llamadas tokens
 - c) Un método para reducir palabras a su forma base
 - d) Un algoritmo para traducir texto
14. ¿Qué son las stop words?
- a) Palabras que son muy importantes en el texto
 - b) Palabras muy frecuentes en un idioma que no aportan significado relevante al análisis del texto
 - c) Palabras que son difíciles de entender
 - d) Palabras que están escritas con errores ortográficos
15. ¿Qué son la lematización y el stemming?
- a) Técnicas para reducir las palabras a sus formas base o raíz
 - b) Métodos para dividir el texto en unidades más pequeñas
 - c) Algoritmos para traducir texto a diferentes idiomas
 - d) Procesos para aumentar el tamaño de los datos de texto
16. ¿Cuál es la diferencia principal entre la lematización y el stemming?
- a) La lematización considera el contexto y las reglas gramaticales, mientras que el stemming simplemente corta los sufijos de las palabras
 - b) El stemming es más preciso que la lematización
 - c) La lematización es más rápida que el stemming
 - d) No hay diferencia entre ellas

17. ¿Cuál de los siguientes es un método para extraer características del texto?
- a) Bolsa de palabras (Bag of Words)
 - b) Análisis de sentimiento
 - c) Reconocimiento de entidades nombradas
 - d) Traducción automática
18. ¿Cómo funciona el modelo de Bolsa de Palabras (Bag of Words)?
- a) Representa el texto como una colección de palabras, donde cada documento se convierte en un vector que contiene la frecuencia de aparición de cada palabra del vocabulario
 - b) Pondera la importancia de cada palabra en un documento
 - c) Asigna un vector denso a cada palabra
 - d) Representa el texto como una secuencia de n-gramas
19. ¿Qué significa TF-IDF?
- a) Transformación de frecuencia inversa de palabras
 - b) Texto de formato idéntico
 - c) Frecuencia de término-frecuencia inversa de documento
 - d) Tasa de filtrado de datos de texto
20. ¿Cuál es la función principal de TF-IDF?
- a) Convertir texto a audio
 - b) Ponderar la importancia de cada palabra en un documento
 - c) Eliminar palabras vacías del texto
 - d) Crear un vocabulario de palabras
21. ¿Qué son los word embeddings?
- a) Representaciones de texto basadas en frecuencias de palabras
 - b) Representaciones vectoriales densas de palabras que capturan el significado semántico
 - c) Métodos para comprimir archivos de texto
 - d) Algoritmos para traducir texto
22. ¿Qué capturan los word embeddings?
- a) La frecuencia de las palabras en el texto
 - b) La longitud de las palabras
 - c) El significado semántico de las palabras
 - d) Las propiedades sintácticas de las palabras
23. ¿Cuál de los siguientes NO es un modelo popular de word embeddings?
- a) Word2Vec
 - b) GloVe
 - c) FastText
 - d) TF-IDF
24. ¿Qué son los N-gramas?
- a) Representaciones de texto basadas en frecuencias de palabras
 - b) Secuencias contiguas de 'n' elementos de un texto o discurso
 - c) Métodos para comprimir archivos de texto
 - d) Algoritmos para traducir texto

25. ¿Para qué se utilizan los N-gramas en PLN?
- a) Para comprimir datos de texto
 - b) Para modelar el contexto y la estructura del texto
 - c) Para traducir textos de un idioma a otro
 - d) Para eliminar palabras vacías del texto
26. ¿Qué es el Reconocimiento de Entidades Nombradas (NER)?
- a) Un método para clasificar textos
 - b) Una técnica para identificar y clasificar entidades mencionadas en un texto
 - c) Un algoritmo para traducir textos
 - d) Un proceso para resumir textos
27. ¿Qué tipo de entidades puede identificar NER?
- a) Solo nombres de personas
 - b) Solo nombres de lugares
 - c) Solo fechas
 - d) Nombres de personas, organizaciones, lugares, fechas, cantidades, etc.
28. ¿Qué es el análisis de sentimientos?
- a) Una técnica para determinar las emociones, opiniones y actitudes expresadas en un texto
 - b) Un método para clasificar textos
 - c) Un algoritmo para traducir textos
 - d) Un proceso para extraer información de textos
29. ¿Cuál es el objetivo del análisis de sentimientos?
- a) Clasificar las expresiones en categorías como positivas, negativas o neutras
 - b) Traducir el texto a diferentes idiomas
 - c) Comprimir el texto
 - d) Aumentar la cantidad de datos de texto
30. ¿Cuál es un ejemplo de aplicación del PLN?
- a) Diseño de nuevos procesadores
 - b) Desarrollo de software de contabilidad
 - c) Filtro de spam
 - d) Creación de modelos 3D
31. ¿Qué es la minería de texto?
- a) El proceso de transformar texto no estructurado en un formato estructurado para identificar patrones significativos
 - b) Un método para traducir texto a diferentes idiomas
 - c) Un algoritmo para comprimir texto
 - d) Una técnica para generar resúmenes de texto
32. ¿Cuál es otro nombre para la minería de texto?
- a) Minería de datos de texto
 - b) Análisis de sentimientos
 - c) Procesamiento del habla
 - d) Minería de opiniones
33. ¿Qué tipo de datos utiliza la minería de texto?
- a) Datos numéricos
 - b) Datos no estructurados, como texto
 - c) Datos categóricos
 - d) Datos de imagen
34. ¿Qué porcentaje de los datos en el mundo están en formato no estructurado como texto?

- a) 50%
 - b) 75%
 - c) 90%
 - d) 25%
- 35.¿Qué técnica se utiliza para dividir el texto en unidades básicas como palabras o frases?
- a) Tokenización
 - b) Lematización
 - c) Stemming
 - d) TF-IDF
- 36.¿Qué técnica reduce las palabras a su forma base según el contexto?
- a) Lematización
 - b) Stemming
 - c) Tokenización
 - d) TF-IDF
- 37.¿Qué técnica reduce las palabras a su raíz eliminando sufijos?
- a) Lematización
 - b) Stemming
 - c) Tokenización
 - d) TF-IDF
- 38.¿Qué técnica ayuda a eliminar palabras comunes que no aportan significado relevante?
- a) Lematización
 - b) Stemming
 - c) Eliminación de stop words
 - d) TF-IDF
- 39.¿Qué técnica evalúa la importancia de una palabra en un documento en relación con un conjunto de documentos?
- a) Lematización
 - b) Stemming
 - c) Eliminación de stop words
 - d) TF-IDF
- 40.¿Qué técnica representa palabras en un espacio vectorial continuo capturando su significado semántico?
- a) TF-IDF
 - b) N-gramas
 - c) Word embeddings
 - d) Eliminación de stop words
- 41.¿Qué son los n-gramas?
- a) Secuencias contiguas de n elementos de un texto
 - b) Representaciones vectoriales de palabras
 - c) Métodos para eliminar palabras vacías
 - d) Técnicas para reducir palabras a su forma base

42. ¿Qué técnica identifica y clasifica entidades como personas, organizaciones o lugares en un texto?
- a) Análisis de sentimientos
 - b) Reconocimiento de entidades nombradas (NER)
 - c) Clasificación de texto
 - d) TF-IDF
43. ¿Qué técnica determina las emociones expresadas en un texto?
- a) Análisis de sentimientos
 - b) Reconocimiento de entidades nombradas (NER)
 - c) Clasificación de texto
 - d) TF-IDF
44. ¿Qué son los embeddings en el contexto de la PNL?
- a) Representaciones vectoriales de palabras o frases que capturan su significado y relaciones semánticas
 - b) Técnicas para eliminar ruido de los datos de texto
 - c) Algoritmos para traducir texto de un idioma a otro
 - d) Métodos para aumentar el tamaño de los datos de texto
45. ¿Qué se utiliza para ponderar la importancia de las palabras en cada documento?
- a) Bolsa de palabras
 - b) TF-IDF
 - c) Word2Vec
 - d) N-gramas
46. ¿Qué tipo de tarea puede ser clasificar canciones por género musical?
- a) Clasificación de texto
 - b) Análisis de sentimientos
 - c) Reconocimiento de entidades nombradas
 - d) Traducción automática
47. ¿Qué técnica se usa para convertir las palabras en números para que un ordenador pueda entender el lenguaje natural?
- a) Word embeddings
 - b) Análisis de sentimientos
 - c) Reconocimiento de entidades nombradas
 - d) Tokenización
48. ¿Cuál es el propósito del entrenamiento de un modelo de word embeddings?
- a) Aprender a comprimir y ordenar las palabras de manera óptima para resolver un problema específico
 - b) Traducir texto a diferentes idiomas
 - c) Aumentar la cantidad de datos de texto
 - d) Eliminar palabras vacías del texto
49. ¿Qué se puede hacer para tratar el desbalance en un conjunto de datos?
- a) Sobremuestreo o submuestreo
 - b) Eliminar palabras vacías del texto
 - c) Realizar la tokenización
 - d) Utilizar TF-IDF

50. ¿Qué modelo fue uno de los primeros en despegar en 2013 y es de los más utilizados hasta la fecha?

- a) Word2Vec
- b) GloVe
- c) FastText
- d) TF-IDF

1. ¿En qué década se originó el Procesamiento del Lenguaje Natural (PLN)?

- a) 1950
- b) 1960
- c) 1970
- d) 1980

2. ¿Con qué disciplinas está íntimamente relacionado el PLN?

- a) Física y Química
- b) Biología y Medicina
- c) Inteligencia Artificial y Computación Lingüística
- d) Matemáticas y Estadística

3. ¿Cuál de las siguientes NO es una fase del análisis del lenguaje en PLN?

- a) Preprocesamiento
- b) Análisis léxico
- c) Análisis financiero
- d) Análisis semántico

4. ¿Qué implica el preprocesamiento de textos?

- a) Análisis sintáctico
- b) Preparación del texto para el análisis
- c) Análisis semántico
- d) Análisis pragmático

5. La segmentación también se conoce como:

- a) Lematización
- b) Stemming
- c) Tokenización
- d) Normalización

6. ¿Qué hace el análisis léxico?

- a) Analiza la estructura de las oraciones
- b) Identifica la función de cada término en una frase
- c) Comprende el significado de las palabras
- d) Analiza la intención de las frases

7. ¿Qué son las etiquetas POS?

- a) Partes de una oración
- b) Etiquetas de partes del discurso
- c) Partes del preprocesamiento
- d) Patrones sintácticos

8. ¿Qué herramienta se utiliza para analizar la estructura de las oraciones?
- a) Lematizador
 - b) Stemmer
 - c) Parser
 - d) Tokenizador
9. ¿Qué busca el "chunking"?
- a) Patrones o subfrases con una estructura específica
 - b) Raíces de las palabras
 - c) Palabras vacías
 - d) La intención de las frases
10. ¿Qué se analiza en el nivel semántico?
- a) La estructura de las oraciones
 - b) El significado de las palabras y las oraciones
 - c) La función de cada palabra en la frase
 - d) La intención de las frases
11. ¿Cuál de estos NO es un recurso utilizado en el análisis semántico?
- a) Diccionarios
 - b) Tesoros
 - c) Analizadores sintácticos
 - d) Ontologías
12. ¿Qué analiza el nivel pragmático?
- a) La estructura de las oraciones
 - b) El significado de las palabras
 - c) La intencionalidad de las frases
 - d) La función de las palabras
13. ¿Cuál es el primer paso en el preprocesamiento de textos?
- a) Decodificación del texto
 - b) Tokenización
 - c) Lematización
 - d) Stemming
14. ¿Qué hace la identificación del idioma en el preprocesamiento?
- a) Traduce el texto
 - b) Determina el idioma del texto
 - c) Elimina elementos no relevantes
 - d) Corrige errores
15. ¿Qué tipo de elementos se eliminan en el preprocesamiento?
- a) Palabras vacías
 - b) Raíces de las palabras
 - c) Secciones no textuales
 - d) Lematizaciones
16. ¿Qué tarea realiza la limpieza y normalización de términos?
- a) Elimina palabras vacías
 - b) Reduce palabras a su lema
 - c) Elimina acentos, guiones y normaliza palabras
 - d) Corrige errores ortográficos

17. ¿Qué son los "spell-checkers"?
- a) Herramientas de corrección ortográfica
 - b) Herramientas de segmentación
 - c) Herramientas de lematización
 - d) Herramientas de stemming
18. ¿Qué son las "stop words"?
- a) Raíces de las palabras
 - b) Palabras vacías
 - c) Errores ortográficos
 - d) Palabras normalizadas
19. ¿Qué hace la lematización?
- a) Reduce palabras a su raíz
 - b) Reduce palabras a su forma base
 - c) Elimina palabras vacías
 - d) Corrige errores
20. ¿Qué hace el stemming?
- a) Reduce palabras a su raíz común
 - b) Reduce palabras a su forma base
 - c) Elimina palabras vacías
 - d) Corrige errores
21. ¿Qué es la tokenización?
- a) Reducción de palabras a su lema
 - b) División del texto en unidades básicas
 - c) Eliminación de palabras vacías
 - d) Corrección de errores
22. ¿Qué es NLTK?
- a) Un idioma
 - b) Una biblioteca de Python para PLN
 - c) Un parser
 - d) Un stemmer
23. ¿Qué es spaCy?
- a) Un corpus lingüístico
 - b) Una biblioteca de Python para PLN
 - c) Un diccionario
 - d) Una ontología
24. ¿Qué es UDPipe?
- a) Una herramienta multilingüe para PLN
 - b) Un corpus de textos en inglés
 - c) Una técnica de stemming
 - d) Una herramienta de corrección ortográfica
25. ¿Qué es NLP-Cube?
- a) Un analizador sintáctico
 - b) Una herramienta para analizar grandes volúmenes de texto en varios idiomas
 - c) Un tipo de tokenización
 - d) Un diccionario de sinónimos

26. ¿Qué es el British National Corpus (BNC)?
- a) Un corpus de referencia del inglés británico
 - b) Un corpus de alemán
 - c) Un corpus de italiano
 - d) Un corpus de japonés
27. ¿Qué es el Corpus of Contemporary German (DTA)?
- a) Un corpus de inglés
 - b) Un corpus de alemán contemporáneo
 - c) Un corpus de italiano
 - d) Un corpus de japonés
28. ¿Qué es el Corpus PAISÀ?
- a) Un corpus de inglés
 - b) Un corpus de alemán
 - c) Un corpus de textos italianos contemporáneos de la web
 - d) Un corpus de japonés
29. ¿Qué es el NINJAL-LWP?
- a) Un corpus de inglés
 - b) Un corpus de alemán
 - c) Un corpus de italiano
 - d) Un corpus de referencia para el estudio del japonés
30. ¿Qué herramienta de etiquetado gramatical soporta varios idiomas y realiza tareas como tokenización y lematización?
- a) Stanza
 - b) GermaNER
 - c) Tint
 - d) MeCab
31. ¿Qué herramienta destaca por su rapidez y eficiencia en el procesamiento del lenguaje natural y tiene modelos preentrenados para inglés, alemán e italiano?
- a) Stanza
 - b) spaCy
 - c) UDPipe
 - d) NLP-Cube
32. ¿Qué herramienta está basada en el proyecto Universal Dependencies y es capaz de analizar textos en una amplia variedad de idiomas?
- a) Stanza
 - b) spaCy
 - c) UDPipe
 - d) NLP-Cube
33. ¿Qué herramienta de procesamiento del lenguaje natural se basa en redes neuronales para analizar grandes volúmenes de texto?
- a) Stanza
 - b) spaCy
 - c) UDPipe
 - d) NLP-Cube

34. ¿Cuál de las siguientes herramientas está diseñada específicamente para el alemán?
- a) NLTK
 - b) GermaNER
 - c) Tint
 - d) MeCab
35. ¿Cuál de las siguientes herramientas está diseñada específicamente para el italiano?
- a) NLTK
 - b) GermaNER
 - c) Tint
 - d) MeCab
36. ¿Cuál de las siguientes herramientas está diseñada específicamente para el japonés?
- a) NLTK
 - b) GermaNER
 - c) Tint
 - d) MeCab
37. Según los textos ¿Cuál es el objetivo principal del British National Corpus (BNC)?
- a) Servir como una herramienta para la traducción automática.
 - b) Ofrecer una muestra de textos antiguos del inglés británico.
 - c) Servir como un recurso representativo y equilibrado del uso del inglés moderno.
 - d) Permitir la creación de modelos de lenguaje específicos para el inglés.
38. ¿Qué características tiene el Corpus of Contemporary German (DTA)?
- a) Se enfoca en el alemán antiguo y medieval.
 - b) Contiene textos de ficción, literatura y periodismo del alemán contemporáneo.
 - c) Es un corpus limitado con pocos textos.
 - d) Solo contiene textos técnicos.
39. ¿Cuál es el enfoque del Corpus PAISÀ?
- a) Textos académicos y formales.
 - b) Textos auténticos contemporáneos extraídos de la web.
 - c) Textos traducidos a varios idiomas.
 - d) Textos antiguos y manuscritos.
40. ¿Qué particularidades tiene el NINJAL-LWP?
- a) Contiene principalmente textos escritos.
 - b) Se centra en el japonés antiguo.
 - c) Incluye textos tanto escritos como hablados del japonés contemporáneo.
 - d) Se utiliza principalmente para la traducción automática.
41. ¿Qué es un token en el contexto del PLN?
- a) Una unidad básica de un texto, como una palabra o frase.
 - b) Una función gramatical de una palabra.
 - c) Un tipo de análisis semántico.
 - d) Una forma de lematización.
42. ¿Qué tipo de información puede ser representada con el sistema de codificación de caracteres?
- a) Solo caracteres latinos.
 - b) Caracteres árabes, chinos, japoneses, europeos y símbolos de internet.
 - c) Solo símbolos matemáticos.
 - d) Solo imágenes.

43. ¿Qué se entiende por "forma canónica" de un token?
- a) Una forma única que representa varias grafías de un mismo concepto.
 - b) Una forma gramaticalmente correcta.
 - c) La raíz de una palabra.
 - d) El significado de una palabra.
44. ¿Cuál es el objetivo de las expresiones regulares en el PLN?
- a) Normalizar textos.
 - b) Detectar palabras con características especiales o patrones.
 - c) Corregir errores ortográficos.
 - d) Eliminar stop words.
45. ¿Qué se entiende por "overstemming" en el proceso de stemming?
- a) El hecho de recortar demasiado una raíz, dando lugar a un resultado no representativo.
 - b) El hecho de no recortar lo suficiente una raíz.
 - c) El proceso de encontrar el lema de una palabra.
 - d) La eliminación de stop words.
46. ¿Qué es un tesoro en el contexto del PLN?
- a) Una lista estructurada de términos que representa el contenido de los documentos.
 - b) Una herramienta de análisis sintáctico.
 - c) Un algoritmo de lematización.
 - d) Un tipo de ontología.
47. ¿Qué es una ontología en el PLN?
- a) Un diccionario de sinónimos.
 - b) Una herramienta para etiquetar partes del discurso.
 - c) La descripción explícita de una conceptualización de un dominio de trabajo.
 - d) Un método de corrección de errores.
48. ¿Qué es la desambiguación de palabras en el PLN?
- a) La eliminación de palabras vacías.
 - b) La reducción de palabras a su raíz.
 - c) La identificación del significado correcto de una palabra en su contexto.
 - d) La normalización de un texto.
49. ¿Qué son las collocations?
- a) Raíces de las palabras.
 - b) Palabras con significado similar.
 - c) Conjuntos de palabras que tienen sentido cuando van juntas.
 - d) Palabras de categorías gramaticales específicas.
50. ¿Qué es el análisis pragmático en el PLN?
- a) El estudio de la estructura de las oraciones.
 - b) El estudio de los significados de las palabras.
 - c) El estudio de la intencionalidad de las frases y cómo el contexto afecta su interpretación.
 - d) El estudio de la función de las palabras.

Soluciones

1.c	11.c	21.b	31.a	41.a
2.d	12.d	22.c	32.a	42.b
3.d	13.b	23.d	33.b	43.a
4.b	14.b	24.b	34.c	44.a
5.b	15.a	25.b	35.a	45.b
6.a	16.a	26.b	36.a	46.a
7.c	17.a	27.d	37.b	47.a
8.a	18.a	28.a	38.c	48.a
9.b	19.c	29.a	39.d	49.a
10.b	20.b	30.c	40.c	50.a

1.b	11.c	21.b	31.b	41.a
2.c	12.c	22.b	32.c	42.b
3.c	13.a	23.b	33.d	43.a
4.b	14.b	24.a	34.b	44.b
5.c	15.c	25.b	35.c	45.a
6.b	16.c	26.a	36.d	46.a
7.b	17.a	27.b	37.c	47.c
8.c	18.b	28.c	38.b	48.c
9.a	19.b	29.d	39.b	49.c
10.b	20.a	30.a	40.c	50.c