

Procesamient de Lenguaje

¿Qué es e PLN?

El PLN en la actualidad

El lenguaj natural

Temario

Bibliografía

## Procesamiento de Lenguaje Natural:

¿Cómo entienden las computadoras el lenguaje humano?

Mauricio Toledo-Acosta mauricio.toledo@unison.mx

Departamento de Matemáticas Universidad de Sonora



Procesamiento de Lenguaje

#### ¿Qué es e PLN?

El PLN en la actualidad

El lenguaje

Temario

Bibliografí

#### Section 1

¿Qué es el PLN?



## Procesamiento de Lenguaje Natural

El procesamiento de lenguaje natural (PLN) es un campo de las ciencias de la computación que estudia las interacciones entre las computadoras y el lenguaje humano, así como los detalles computacionales de las lenguas naturales.

El PLN combina modelos de lingüística computacional y el aprendizaje automáticos para procesar el lenguaje humano



#### Distintivos Niveles de Abstracción

Procesamient de Lenguaje Natural:

¿Qué es e PLN?

EI PLN en la actualidad

El lenguajo natural

Temario

Bibliogr

El análisis del NLP puede ocurrir a distintos niveles cognitivos:

- Nivel de caracteres.
- Nivel de tokens/palabras.
- Nivel de frases.
- Nivel de oraciones.
- Nivel de documentos.

Además, podemos ver la sintaxis y semántica:

- Incoloras ideas verdes duermen furiosamente.
- Mi niñas duerme tranquilo.



Procesamiento de Lenguaje

#### ¿Qué es el PLN?

El PLN en la actualidad

El lenguaj natural

Temario

Bibliograf

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto



Procesamiento de Lenguaje Natural:

#### ¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliogra

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto



rocesamiento de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguaj natural

Temario

Bibliograf

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto



rocesamiento de Lenguaje

¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliogra

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



rocesamiento de Lenguaje

¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Ribliogr

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



rocesamiento de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliogr

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



rocesamiento de Lenguaje

¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliograf

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto



rocesamiento de Lenguaje

¿Qué es el PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Bibliograf

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto



Las tareas puedes decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis
- Generación de texto.



actualidad

natural

Date:



Las tareas puedes decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



actualidad

El lenguaje natural

Temaric



Las tareas puedes decantarse hacia la lingüística o la computación.

- Análisis de sentimientos.
- Detección de tópicos o modelaje de tópicos.
- Estudios lingüísticos.
- Análisis de dependecia gramatical.
- Extracción de información.
- Reconocimiento de Entidades Nombradas (NER).
- Preguntas y respuestas.
- Traducción.
- Detección de plagio y autoría.
- Paráfrasis.
- Generación de texto.



actualidad

natural

remano

Dibliografia



Procesamiento de Lenguaje

¿Qué es e PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliogra<sup>\*</sup>

### Section 2

El PLN en la actualidad



#### Ejemplos de PLN: Llenado de formas

Procesamient de Lenguaje

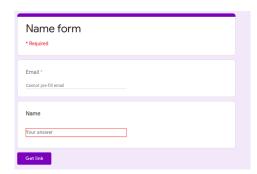
¿Qué es el PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Bibliografía



|          |          |      |     | • |
|----------|----------|------|-----|---|
| =        | Contacts |      | ۹ : |   |
|          | Name     |      |     |   |
|          | 1141110  |      |     |   |
|          | Phone    | Area | •   |   |
| 9        | Address  |      |     |   |
|          | City     |      |     |   |
|          | State ▼  | Zip  |     |   |
| ~        | Email    |      |     |   |
| <u>.</u> | Birthday |      |     |   |



## Ejemplos de PLN: Llenado de formas

Procesamiento de Lenguaje

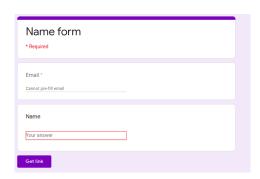
¿Qué es el PLN?

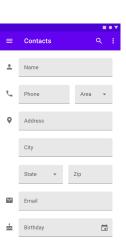
El PLN en la actualidad

El lenguaje natural

Temario

Ribliogra





Retos: Mayúsculas/mínusculas, caracteres invalidos, fórmatos.



## Ejemplos de PLN: Busquedas

Procesamiento de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliografía



Ejemplos de busqueda: Búsqueda 1, Búsqueda 2, Búsqueda 3



### Ejemplos de PLN: Busquedas

Procesamiento de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguaje natural

Temaric

Bibliogra



Ejemplos de busqueda: Búsqueda 1, Búsqueda 2, Búsqueda 3

Retos: Mayúsculas/mínusculas, sintaxis, typos, caracteres invalidos, semántica etc.

BERT en las busquedas



#### Ejemplos de PLN: Large Language Models

¿Qué es el

El PLN en la actualidad

El lenguaj: natural

Temario

Bibliografí

Los LLMs son modelos de lenguaje diseñados para procesar y generar texto de manera autónoma. Es capaz de mantener conversaciones con usuarios en lenguaje natural, responder preguntas, proporcionar información, traducción automática, resumen de texto, todo esto a través de texto. Su versatilidad radica en su capacidad para comprender y generar texto coherente y relevante en una amplia gama de contextos y aplicaciones.

Detección de texto generado por Al



#### Un poco más sobre ChatGPT 3

rocesamiento de Lenguaje Natural:

¿Qué es e PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Ha obtenido los siguientes puntajes:

- Obtuvo 1020 puntos de 1600 posibles en los exámenes SAT.
- Obtuvo un nivel de C+ en los exámenes de cuatro cursos de la Facultad de Derecho de la Universidad de Minnesota.
- Obtuvo un nivel de B a B- en el examen de un curso de gestión empresarial de la Wharton School of Business de la Universidad de Pensilvania.
- Obtuvo una puntuación igual o cercana al aprobado en el Examen de Licencia Médica de EE.UU., que los médicos deben aprobar para obtener la licencia médica.



#### Las alucinaciones

rocesamient de Lenguaje Natural:

¿Qué es e PLN?

El PLN en la actualidad

El lenguaje natural

Temario

D:L::-..

ChatGPT y otros LLMs para generación de textos no son infalibles.

- Acerca de si los dinosaurios construyeron una civilización:
  Some species of dinosaurs even developed primitive forms of art, such as engravings on stones. Corrección
- When did Leonardo da Vinci paint the Mona Lisa?
  Leonardo da Vinci painted the Mona Lisa in 1815.
  Corrección
- Ha inventado referencias y documentos.
- Puede reproducir sesgos.
- Casos concretros Gemini, Claude.



## Tendencias en la investigación

Procesamient de Lenguaje

¿Qué es e PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Bibliografí

Ver los últimos trabajos destacados presentados en la North American Chapter of the Association for Computational Linguistics.

https://2024.naacl.org/

¿Qué temáticas y tendencias observas?



Procesamiento de Lenguaje

¿Qué es e PLN?

El PLN en la actualidad

El lenguajo natural

Temario

Bibliografi

#### Section 3

El lenguaje natural



## Los dificultades del lenguaje natural

Procesamient de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguaje natural

Temario

 Ambigüedad. Una misma palabra o frase puede tener múltiples significados dependiendo del contexto. Ejemplos: Banco, vela, planta.

- Variabilidad lingüística. Existen muchas variaciones en la forma en que las personas utilizan el lenguaje, incluyendo diferencias regionales, culturales y generacionales. Ejemplos: Padre, chido, cool, genial, etc.
- Ironía y sarcasmo. Son difíciles de detectardebido a su naturaleza no literal.



## Los dificultades del lenguaje natural

¿Qué es e

El PLN en la actualidad

El lenguajo natural

Temario

Bibliografía

Resolución de la correferencia. Identificar a qué se refieren pronombres como "él", "ella", "eso" en un texto.
 Se requiere comprensión del discurso y conocimiento del mundo. Ejemplo: María le dio un libro a Ana. Ella lo leyó rápidamente.

- Ambigüedades. ¿Cuál es el significado de las siguientes frases?
  - Vi al niño con los binoculares.
  - I shot my neighbor in my pijamas.



## Los dificultades del lenguaje natural

Procesamient de Lenguaje Natural:

¿Qué es el PLN?

El PLN en la actualidad

El lenguaj natural

Temario

Bibliograf

- Procesamiento de sentimientos y emociones. Capturar el tono, el sentimiento y las emociones detrás del lenguaje es un desafío, ya que requiere comprensión de las sutilezas del discurso humano.
- Tratamiento de errores y ruido. Los errores gramaticales, los errores tipográficos y otros tipos de ruido en el texto pueden dificultar el procesamiento del lenguaje natural.



#### ¿Cuál es la solución?

Procesamient de Lenguaje Natural

¿Qué es e PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Date:

La solución puede ser más o menos sofisticada de acuerdo a la tarea que se quiera resolver.

Puede requerir un algoritmo muy sencillo o un módelo tan avanzado como un Large Language Model (LLM), por ejemplo Chat-GPT, LlaMa, DeepSeek, Qwen, etc.



Procesamiento de Lenguaje

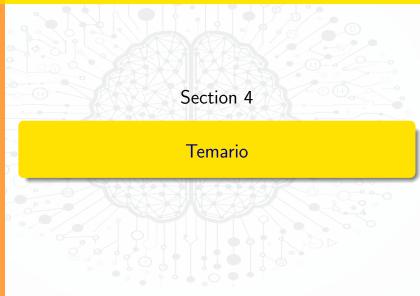
¿Qué es e PLN?

El PLN en la actualidad

El lenguaje

Temario

Bibliografía





#### **Temario**

Qué es e

El PLN en la actualidad

El lenguaje natural

Temario

.....

- Introducción:
  - Tareas del NLP
  - Expresiones regulares
- Métodos de modelado clásico de NLP
  - N-gramas.
  - Módelos BOW (1954), TF-IDF (1970s).
- Algoritmos Clásicos del NLP
  - Naive Bayes
  - Latent Dirchlet Allocation
  - Latent Semantic Analysis
- Métodos de redes neuronales de NLP
  - Embeddings de tokens y documentos.
  - Redes recurrentes y LSTM.
  - Modelos Word2Vec (Google, 2013), Glove (Standford, 2014), FastText (FB, 2016), Doc2Vec (Google, 2014), WordRank (2016).



## **Temario**

Out as

El PLN en la

El lenguaje natural

Temario

Ribliogra

- Transformadores y LLM.
  - Arquitectura Transformers (2017): Attention is all you need.
  - BERT (Google, 2018), RoBERTa (FB, 2019), T5 (2020), GTP-3 (OpenAI, 2020), LLaMA (Meta AI, 2022), GPT-4 (Open AI, 2023).
  - Afinación de modelos pre-entrenados.
  - Prompt Engineering.
  - Retrieval Augmented Generation (RAG).

La evaluación del curso será mediante la elaboración y presentación de proyectos y actividades (teóricos y/o prácticos).



Procesamiento de Lenguaje

¿Qué es e PLN?

El PLN en la

El lenguaj

Temario

Bibliografía

## Section 5

# Bibliografía



# Bibliografía

Procesamient de Lenguaje Natural:

PLN?

El PLN en la actualidad

El lenguaje natural

Temario

Bibliogra<sup>-</sup>

- Goldberg, Y. (2022). Neural network methods for natural language processing. Springer Nature.
- Eisenstein, J. (2018). Natural language processing. Jacob Eisenstein.
- Jurafsky, D., Martin, J. H. (2019). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.