Aprendizaje automático 2

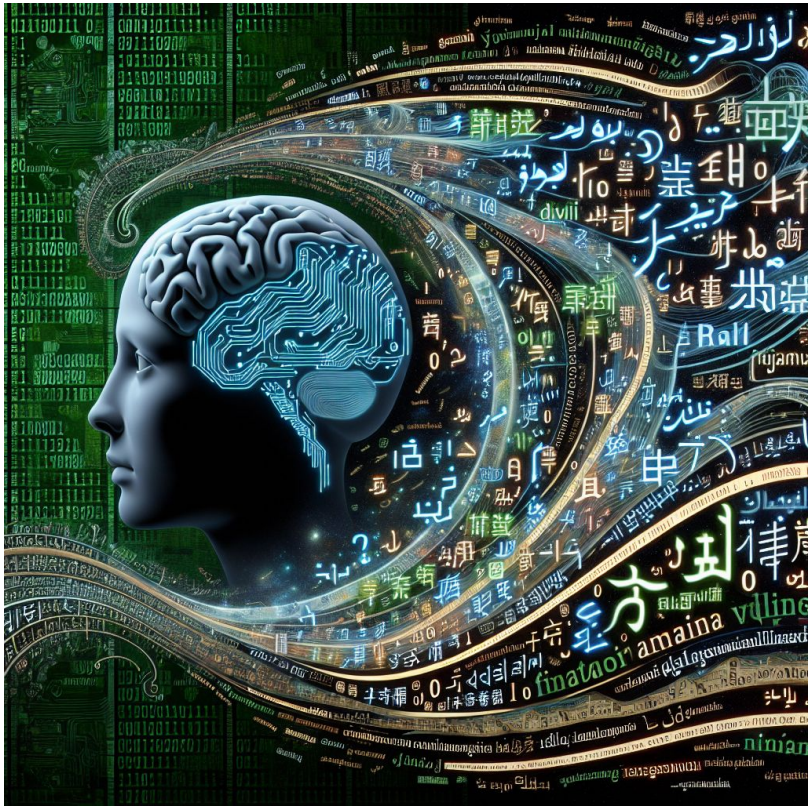
Conceptos de procesamiento del lenguaje natural (PLN)

ISTEA

¿Qué y quiénes somos? ¿Cómo nos comunicamos?

¿Qué y quiénes somos? ¿Cómo nos comunicamos?

¿Qué y quiénes somos? ¿Cómo nos comunicamos?



Lenguaje escrito

Hemos han usado la escritura para guardar y transmitir conocimiento por miles de años.

Los avances en la escritura han fomentado grandes revoluciones sociales y científicas, desde la invención de la escritura, la prensa y hasta los post en internet.

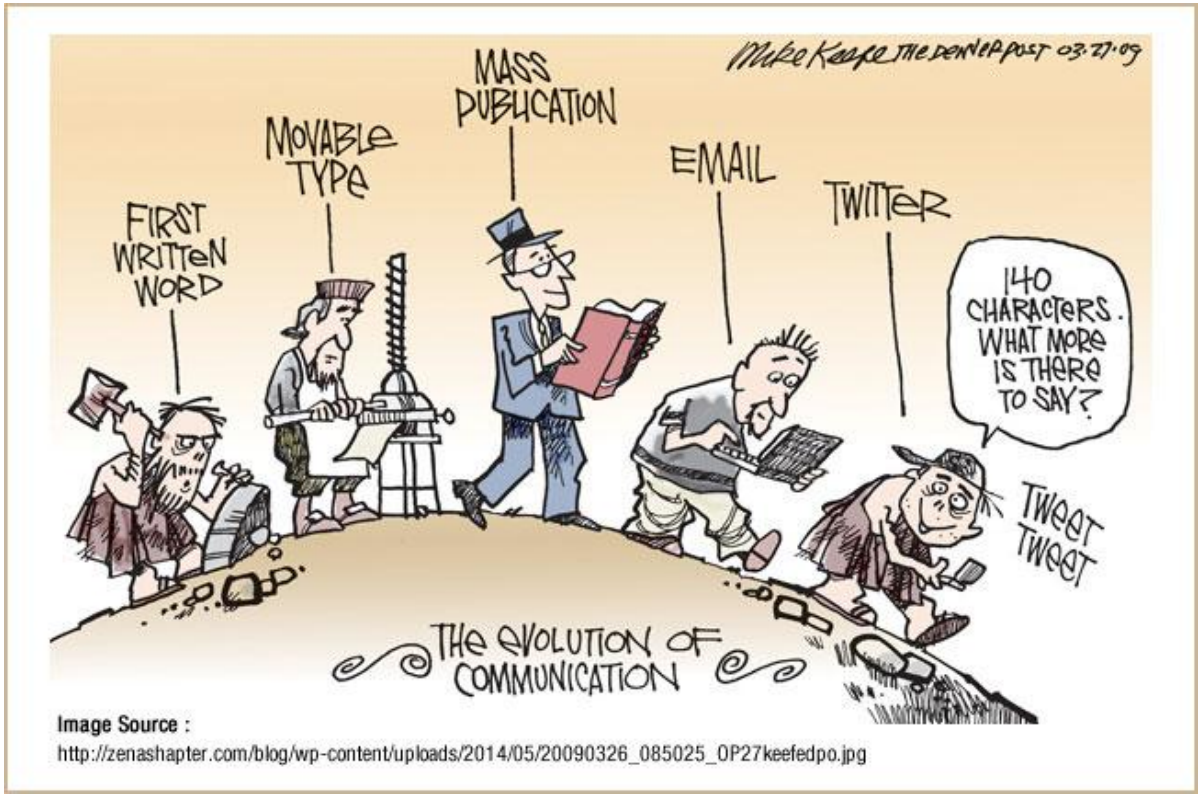


Image Source :
http://zenashapter.com/blog/wp-content/uploads/2014/05/20090326_085025_OP27keefedpo.jpg

Introducción al PLN

- Definición de PLN: Rama de la inteligencia artificial que se centra en la interacción entre computadoras y el lenguaje humano.
- Importancia del PLN: Facilita la comunicación hombre-máquina, análisis de datos textuales y comprensión de información.
- Historia del PLN: Desde el análisis de sintaxis hasta el uso de redes neuronales y modelos de lenguaje avanzados.
- Aplicaciones del PLN: Chatbots, traducción automática, análisis de sentimientos, motores de búsqueda.

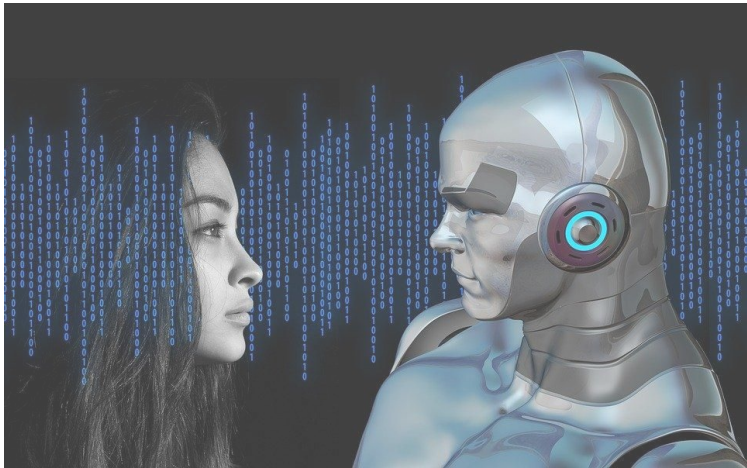
Lenguaje escrito

- Internet y la producción de datos digitales han revolucionado el flujo de la información otra vez
- Existe demasiada información para procesar, por suerte tenemos la IA para buscar, analizar, clasificar, recuperar, sintetizar, y interpretar (entender)
- IA puede imitar las intuiciones humanas para entender significados
  (de qué trata un texto, ej.: corrupción)
- IA puede detectar patrones, que quizás a los humanos les cueste mucho
  (ej.: estilo de un autor particular de un set de textos sin nombres)

"Siempre imaginé que el paraíso sería algún tipo de biblioteca."
"Hay derrotas que tienen más dignidad que una victoria."

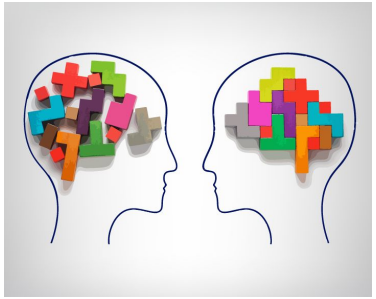
¿Qué es PLN?

- Habilidad de un programa de computadora para interpretar el lenguaje humano, tanto hablado como escrito, también llamado *lenguaje natural*.
- Uno de sus fines principales es mejorar la *interacción hombre-computadora (en inglés, HCI)*.
- De forma similar a los humanos, que tenemos diferentes sensores (los oídos, la voz, los ojos, etc.), las computadoras también (micrófono, altavoz, cámaras, etc.) y el PLN procesa la información por medio de la IA y le da un sentido que la computadora puede interpretar y viceversa.



¿Qué es PLN?

- Pertenece al campo de la Lingüística Computacional
  Natural Language Understanding, (NLU),
  Natural Language Generation (NLG): Text-to-speech y ASR



- La sigla PLN es también utilizada para describir la técnica denominada programación neurolingüística (John Grinder -lingüista- y Richard Bandler -matemático-, 1973).
  Está relacionada con procesos neurológicos, el lenguaje y patrones de comportamiento aprendidos por experiencias.

- Su inicio en 1950, con la publicación "Computing machinery and intelligence" (Turing)

  (https://academic.oup.com/mind/article/LIX/236/433/986238)

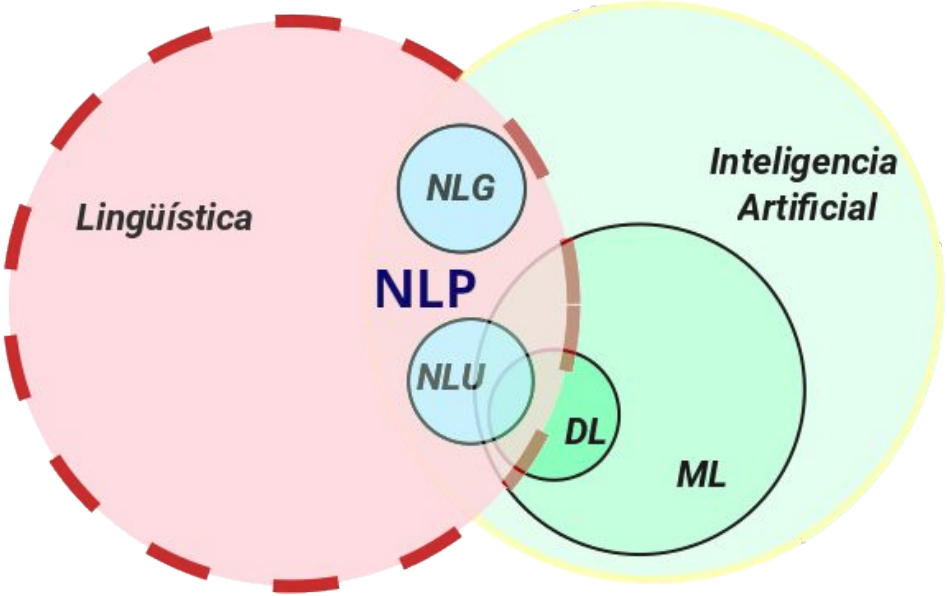
- Es un tema de actualidad porque el problema siempre se puede complejizar más...
  Dificultades: ambigüedad de las lenguas,
  recepción imperfecta de los datos,
  separación entre palabras, entre otros.

I am going to Beijing tomorrow. 我明天要去北京。

I go swimming at school every morning at 7am.
我每天早上7点去学校游泳。

I am going shopping with my friends at Central on Sunday.
我和我的朋友星期天要去中环买东西。

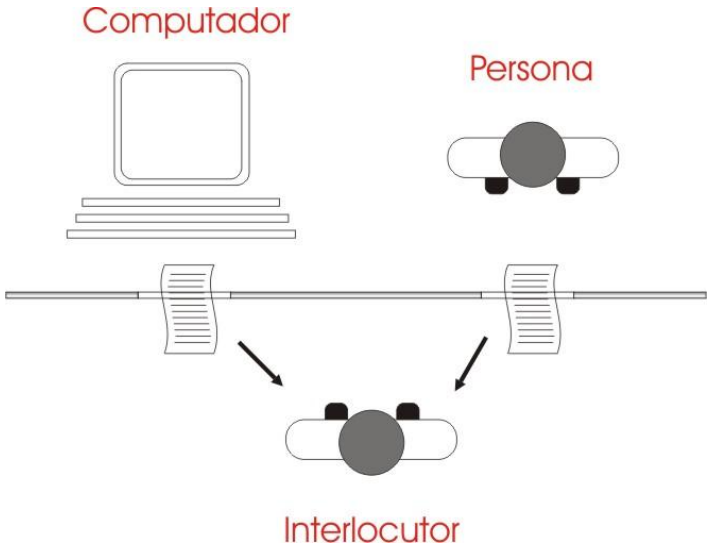
¿Qué es PLN?



- Inicios 1950s. Alan Turing
- 1950s-1990s. PLN ampliamente basado en reglas, desarrolladas por lingüistas para determinar cómo las máquinas podían procesar lenguajes.
- 1990s. Las computadoras se vuelven más rápidas y se usan para desarrollar reglas basadas en estadísticas lingüísticas sin que un lingüista creara todas las reglas. Entran al juego ingenieros y cientistas de datos.
- 2000-. PLN experimentó un crecimiento exponencial en popularidad científica y de consumidores. Hoy en día, implican una combinación de lingüística clásica, métodos estadísticos, inteligencia artificial y un gran poder de cómputo.

El test de Turing

- Un humano evalúa conversaciones en lenguaje natural entre un humano y una máquina diseñada para generar respuestas similares a las de un humano (limitada a un medio únicamente textual)
- Si después de un tiempo el evaluador no puede distinguir entre el humano y la máquina acertadamente, la máquina ha pasado la prueba.
- No se evalúa el conocimiento de la máquina respecto a su capacidad de responder preguntas correctamente, solo se toma en cuenta la capacidad de generar respuestas similares a las que daría un humano.

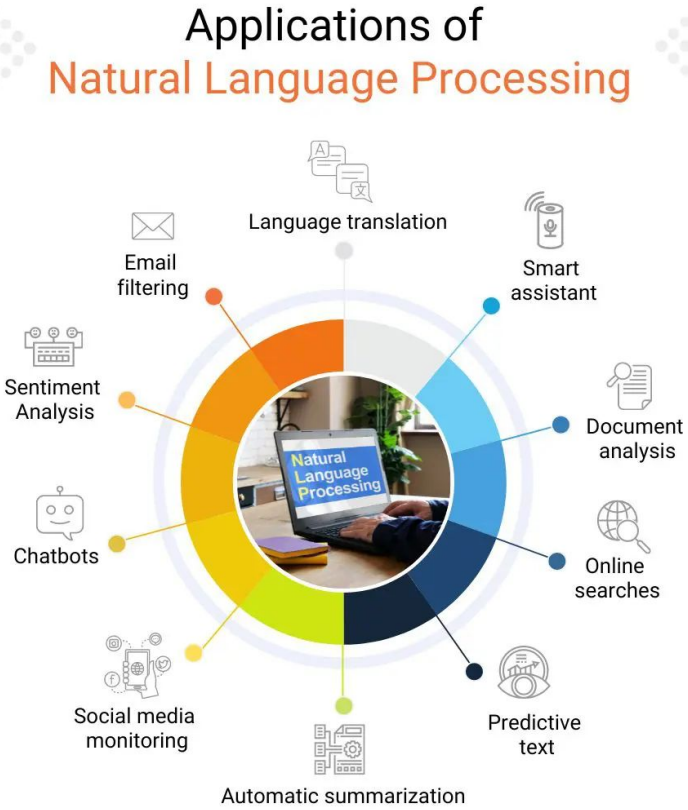


(Turing, 1950)
“¿Pueden pensar las máquinas?”

“¿Existirán computadoras digitales imaginables que tengan un buen desempeño en el juego de imitación?”

→ Mitsuku y otros!

¿Qué hacemos con PLN en texto?

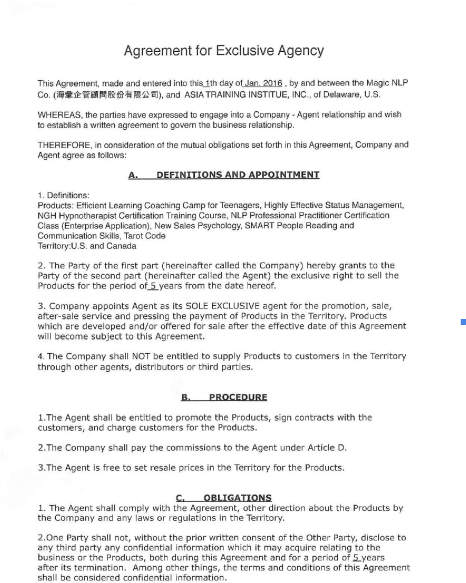


Tareas del PLN

- **Reconocimiento de Entidades Nombradas (NER)**
  Identifica y clasifica entidades importantes en el texto, como nombres de personas, lugares, organizaciones, etc.
  Utilizado en: extracción de información, análisis de noticias, blogs, etc.

Tareas del PLN

- **Extracción de Información**
  Identifica y estructura información específica a partir del texto (posiblemente no estructurado).
  Ejemplo: recuperación de respuestas en motores de búsqueda, análisis de opinión, documentos legales.



Agreement for Exclusive Agency

This Agreement, made and entered into this 1th day of Jan, 2016 , by and between the Magic NLP Co. (海量企管顧問股份有限公司), and ASIA TRAINING INSTITUTE, INC., of Delaware, U.S.

WHEREAS, the parties have expressed to engage into a Company - Agent relationship and wish to establish a written agreement to govern the business relationship.

THEREFORE, in consideration of the mutual obligations set forth in this Agreement, Company and Agent agree as follows:

A. DEFINITIONS AND APPOINTMENT

1. Definitions:
Products: Efficient Learning Coaching Camp for Teenagers, Highly Effective Status Management, NGH Hypnotherapist Certification Training Course, NLP Professional Practitioner Certification Class (Enterprise Application), New Sales Psychology, SMART People Reading and Communication Skills, Tarot Code
Territory:U.S. and Canada

2. The Party of the first part (hereinafter called the Company) hereby grants to the Party of the second part (hereinafter called the Agent) the exclusive right to sell the Products for the period of 5 years from the date hereof.

3. Company appoints Agent as its SOLE EXCLUSIVE agent for the promotion, sale, after-sale service and pressing the payment of Products in the Territory. Products which are developed and/or offered for sale after the effective date of this Agreement will become subject to this Agreement.

4. The Company shall NOT be entitled to supply Products to customers in the Territory through other agents, distributors or third parties.

B. PROCEDURE

1.The Agent shall be entitled to promote the Products, sign contracts with the customers, and charge customers for the Products.

2.The Company shall pay the commissions to the Agent under Article D.

3.The Agent is free to set resale prices in the Territory for the Products.

C. OBLIGATIONS

1. The Agent shall comply with the Agreement, other direction about the Products by the Company and any laws or regulations in the Territory.

2.One Party shall not, without the prior written consent of the Other Party, disclose to any third party any confidential information which it may acquire relating to the business or the Products, both during this Agreement and for a period of 5 years after its termination. Among other things, the terms and conditions of this Agreement shall be considered confidential information.

Partes del Contrato

Objeto y Duración del Contrato

Obligaciones y Compromisos

Cláusulas Importantes

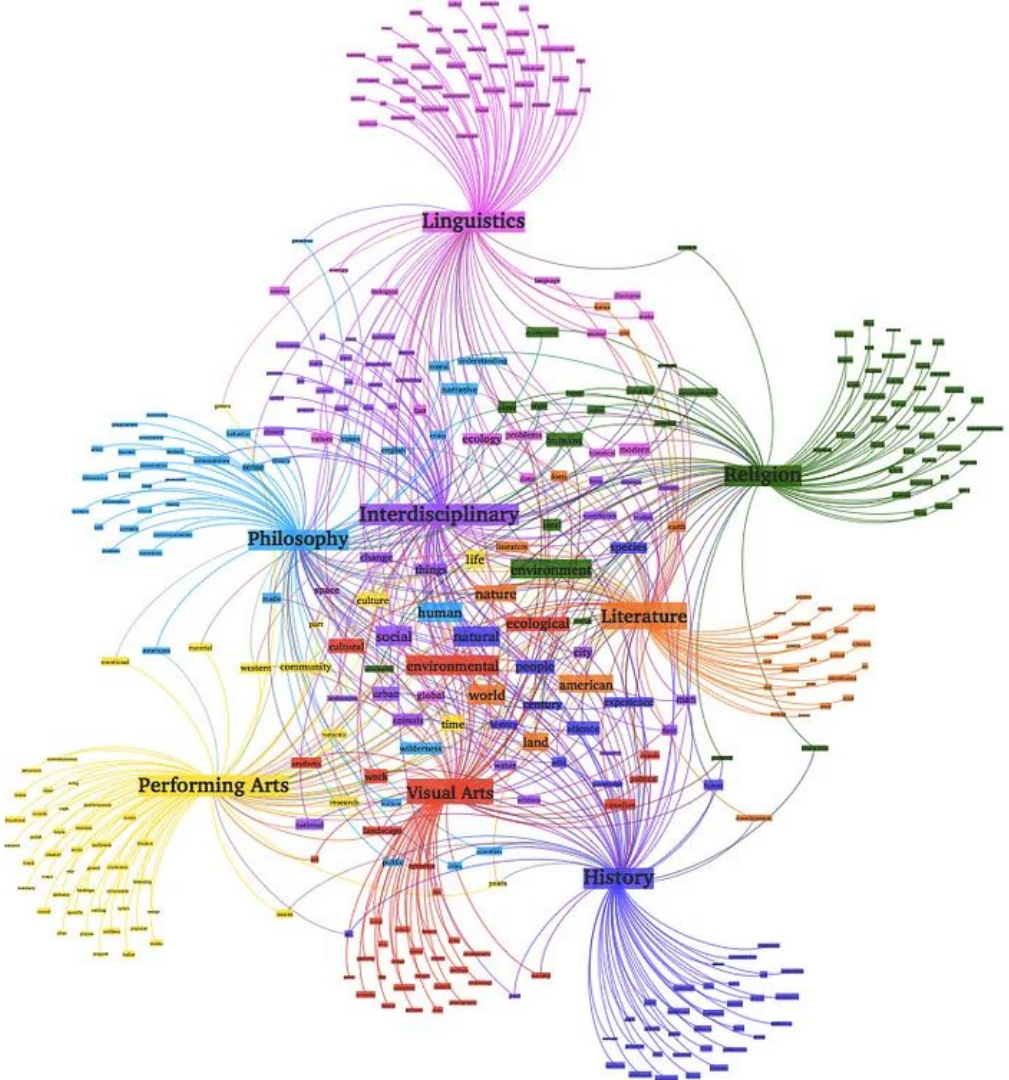
Palabras clave:
https://wordcount.com/es/keyword-extractor

Tareas del PLN

- Modelado de temas. asignar etiquetas a los textos para agruparlos en categorías.

  Descubre temas latentes o dominantes en grandes volúmenes de texto.

  Utilizado en:
  análisis de tendencias,
  agrupación de documentos,
  detección de comportamientos anómalos,
  organización automática en disco,
  envíos automatizados en organizaciones,
  etc...

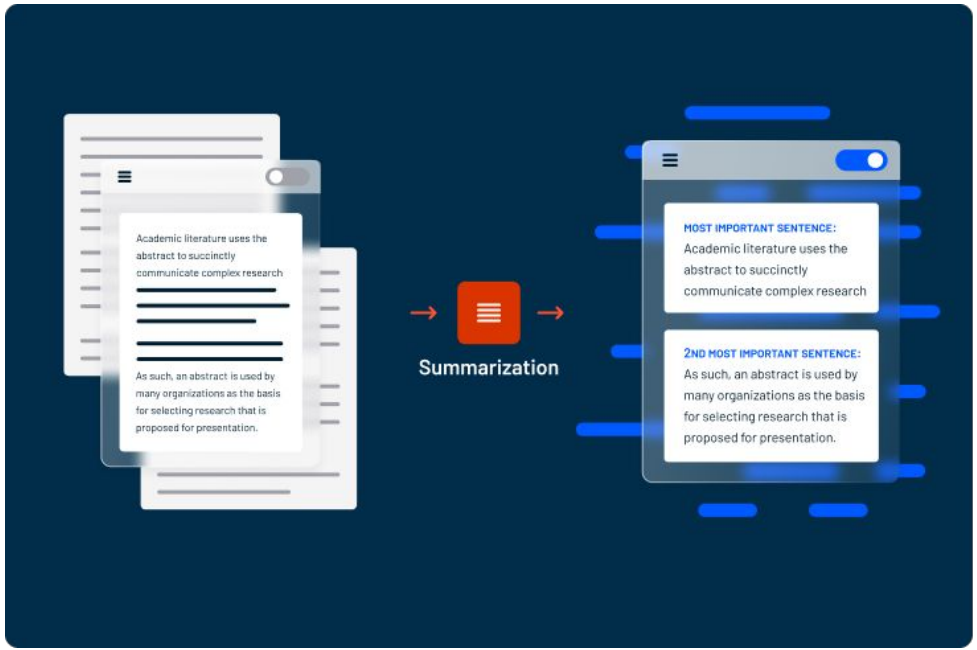


Tareas del PLN

- **Análisis de Sentimiento**
  Determina la actitud emocional detrás de un texto, como positiva, negativa o neutra.
  Aplicaciones: análisis de opiniones en redes sociales, comentarios de clientes, etc.

Tareas del PLN

- **Resumen Automático**
  Crea resúmenes concisos de textos largos, preservando la información esencial.
  Aplicaciones: resumen de artículos científicos, noticias, documentos legales.



Tareas más comunes del PLN

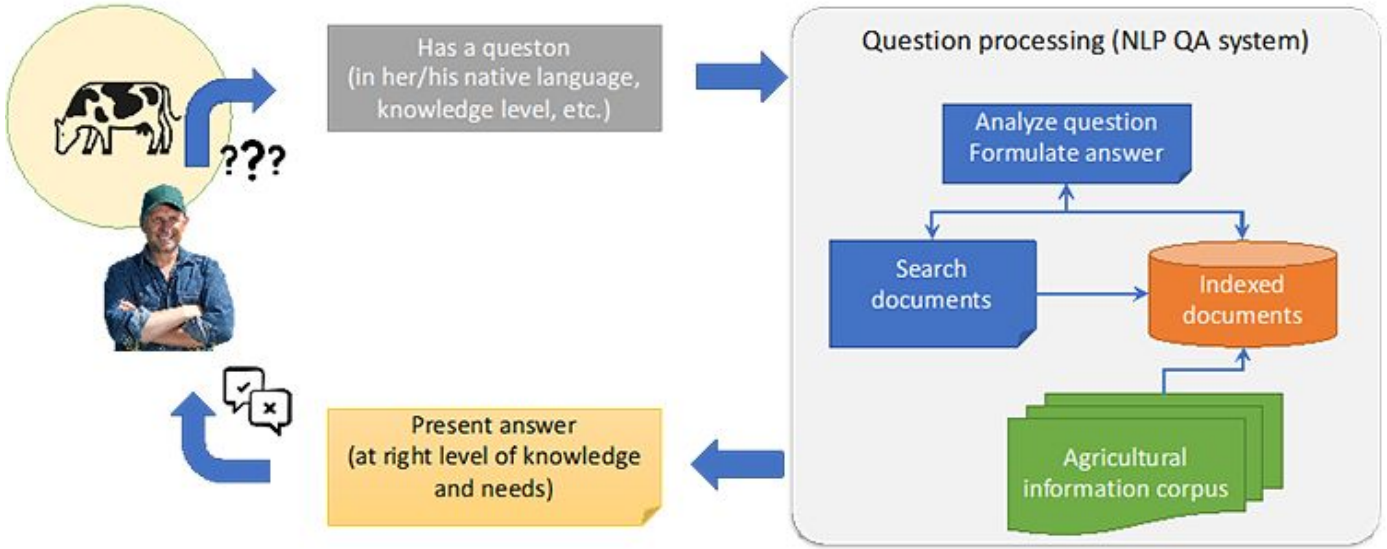
- **Traducción Automática**
  Transforma texto de un idioma a otro de manera automática.
  Ejemplos: Google Translate, DeepL, Microsoft Translator, Amazon Translate, chatGPT...

Tareas más comunes del PLN

- **Generación de Lenguaje Natural**
  Produce texto coherente y gramaticalmente correcto de manera automática.
  Ejemplo: chatbots, generación de contenido automatizado.

Tareas del PLN

- **Preguntas y Respuestas Automatizadas (QA)**
  Entiende preguntas formuladas en lenguaje natural y proporciona respuestas precisas.
  Ejemplo: asistentes virtuales, sistemas de soporte al cliente automatizados.



Aplicaciones del PLN

- Cuidado de la salud (healthcare)
  - análisis y categorización de registros médicos
  - informes de registros automatizado

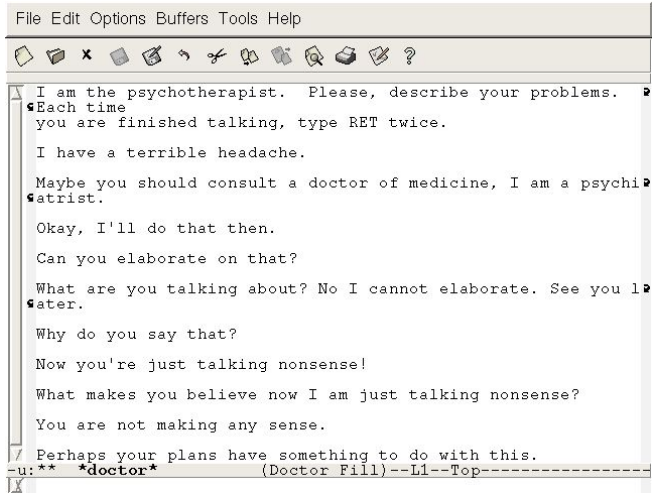
- Marketing inteligente
  - permite saber qué hacen los competidores y estar actualizados través del análisis de millones de blogs, sitios webs, etc.
  - análisis de comentarios de los clientes: analizar de las reseñas de las redes sociales

- Servicio al cliente por asistentes con ASR y TTS
  - comprender lo que dice el cliente y actuar en consecuencia
  - detectar emociones para evitar pérdida de clientes

- Servicios financieros y legales
  - previsión de acciones e información sobre el comercio financiero.
  - automatización de las tareas de litigio de rutina

CHATBOTS

THE HISTORY OF CHATBOTS



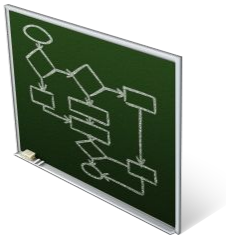
- 1964 - 1966 : ELIZA (http://www.deixilabs.com/eliza.html)
  - programa informático diseñado en el MIT por Joseph Weizenbaum.
  - Parodiaba al psicólogo Carl Rogers

- 1970-2010: Incontables aportes...
  - Parry, RACTER, Dr. Sbaitso, Alice, Watson, ...

- 2010-2016: Asistentes personales
  - SIRI (Apple), Google Assistant, Alexa (Amazon), Cortana (Microsoft)

- Ahora: modelos de lenguaje de gran tamaño
  - 2021: ChatGPT (OpenAI)
  - 2023: Gemini (Google), Copilot (Microsoft), Llama (Facebook, Claude (Anthropic)
  - Un largo etc: LuzIA, POE, YouChat, CharacterAI, Aria, Socratic, GPT4All, ...

Herramientas más populares para el PLN



- NLTK (Natural Language Toolkit) - https://www.nltk.org/
  - disponible para Windows, Mac OS X, and Linux
  - es libre, de código abierto free, proyecto guiado por la comunidad
  - libro “Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit” disponible en https://www.nltk.org/book/

- SpaCy - https://spacy.io/
  - diseñada para aplicaciones en producción
  - basada en modelos de aprendizaje profundo preentrenados, con soporte para múltiples idiomas
  - documentación y guías disponibles en https://spacy.io/usage
  - licencia MIT, con un enfoque en velocidad y eficiencia
    proyecto abierto y colaborativo, ampliamente utilizado en investigación y aplicaciones industriales

- The Apache OpenNLP library - https://opennlp.apache.org/
  - herramienta para NLP en texto, basadas en aprendizaje maquinal
  - manual de uso disponible en https://opennlp.apache.org/docs/
  - licencia Apache versión 2.0, proyecto colaborativo

- Gensim - https://pypi.org/project/gensim/
  - librería de Python para modelado de tópicos, indexación de documentos, recuperación de similaridades en corpus enormes
  - licencia LGPL-2

Fin de la clase