

UNIVERSIDAD CATÓLICA DE PEREIRA

OPTATIVA I PROCESAMIENTO DEL
LANGUAGE NATURAL

Apuntes de clase

Juan David Garcia Acevedo

Docente de la materia
Juan Carlos Blandón ANDRADE

8 de agosto de 2024

Índice general

| | | |
|-----------|--|----------|
| I | Introducción a la Inteligencia Artificial | 3 |
| 1. | Introducción a la asignatura | 4 |
| 1.1. | Clase 1 Agosto | 4 |
| 1.2. | Introducción a la IA | 4 |

| | | |
|-------------|---|-----------|
| II | PLN en Java | 7 |
| III | Introducción a Python | 8 |
| IV | Bases PLN | 9 |
| V | NLP - Pipeline | 10 |
| VI | Text Representation | 11 |
| VII | Text Clasification | 12 |
| VIII | Information Extraction | 13 |
| IX | Spech Recognition and Synthesis | 14 |
| X | Chatbots | 15 |
| XI | Search and Information Retrieval | 16 |
| XII | Topic Modelling | 17 |
| XIII | Recommender Systems for Textual Data | 18 |
| XIV | Metodología de Calificación | 19 |
| 2. | Cortes | 20 |
| 2.1. | Porcentajes evaluativos | 20 |

Parte I

Introducción a la Inteligencia Artificial

Capítulo 1

Introducción a la asignatura

1.1. Clase 1 Agosto

Procesamiento del lenguaje natural

La materia esta enfocada a la inteligencia artificial y se presentará un componente muy fuerte de investigación durante el transcurso de la asignatura. La idea de la asignatura es solucionar problemas y tratar en la medida de lo posible aplicar soluciones que sean de calidad.

Los resultados de aprendizaje de la materia se centrarán en habilidades teórico/prácticas. Permitiendo conceptualizar conceptos; entender y poder dar definiciones ya que todo es basado en habilidades que denoten calidad.

El procesamiento del lenguaje como su nombre indica, procesar el lenguaje. El ingeniero debe y tiene que saber gramática, saber escribir, leer y redactar textos.

Hoy día se usan transformadores, la idea es poder combinar el sistema de reglas con *Machine Learning* con entrenamiento para hacer seguimientos. (Ya hablaremos de reglas más adelante).

Usando reglas y patrones. Árboles, semántica y se trabaja bastante.

1.2. Introducción a la IA

Podemos decir que el tema de la asignatura **Procesamiento del Lenguaje Natural** arranca con un Paper considerado el **Génesis en 1943** del PLN, el Paper es *A logical calculus of the ideas immanent in nervous activity*, donde se comienza a hablar sobre el sistema nervioso, se habla sobre el funcionamiento nervioso. Entran autores como *Warren McCulloch*, *Walter Pitts*, *Alan Turing* y este señor Alan Turing brinda la programación computacional como Teoría de la Computación.

Esta apareció en 1943 cuando Alan Turing hace unas contribuciones con un Paper *Computing machinery and intelligence*, donde este señor ya estaba pensando en aprendizaje automático, algoritmos genéricos y aprendizaje por refuerzo.

Seguido a este aporte de Alan Turing, se nos entrega el **Test de Turing**.

El Test de Turing en resumidas cuentas, tiene a dos humanos, donde uno de ellos envía preguntas las cuales responde un humano y una máquina pero, el test dice que cuando el entrevistador humano no sabe quién contesto la pregunta, se puede decir que la inteligencia artificial llegó a un grado importante engañando al humano entrevistador. Hablamos de 1950.

La IA como término nace en 1956 por *John Macartey* llamando este enfoque como Inteligencia artificial "**siendo la ciencia de crear máquinas inteligentes, programas de computo inteligentes**".

El uso del Test de Turing es una prueba para darse cuenta que tan avanzados están los algoritmos de IA después de un entrenamiento.

En 1967 nace el Perceptrón Mark 1, este es básicamente un conjunto de redes neuronales, con entradas a realizar un proceso y dar una salida. El Perceptrón está centrado a una nuestra naturaleza humana es decir, el perceptrón tiene unas entradas, trabaja como unas neuronas se trata y con Deep Learning. Entre más neuronas hay desventajas y ventajas como mayor precisión más consumo de máquina.

Debido a esto procesar media página vs 10 páginas, es una tarea que puede ser pesada en términos de precisión. En Deep Learning tenemos muchas neuronas conectadas entre sí para poder aprender mediante algoritmos.

Hoy se habla del concepto Capas de Neuronas que se comunican y dan una respuesta 0 o 1.

1980 Nace la Habitación China y es la antítesis del Test de Turing, ya que se refuta su idea de determinar el punto al que ha llegado una IA, porque dice que una máquina que pasa el Test de Turing no quiere decir que tenga inteligencia, sólo la simula. Por tanto, no entendió el contenido semántico la inteligencia artificial, no comprendió lo sintáctico.

■ Eventos importantes

1. IBM 1997 vence al campeón de ajedrez
2. IBM 2011 vence al campeón de jeopardy
3. MINWA 2015 super imágenes
4. 2017 nace la base de los generadores de texto, todos de sobre el principio de generación de texto, que paso atrás y que paso a futuro

Importante para estos generadores de texto del 2017, **Question Answer** competencias de IA a desarrollar problemas.

Hablamos de ChatGPT y generadores de texto, construir sistemas que ayudan a alimentar los algoritmos mediante pruebas que ellas mismas como IA deben de resolver.

El corpus es el base de conocimiento, y siguen aprendiendo.

La programación funcional es muy rápida y usa muchos parentesis ya que generan tiempos muy eficientes de computo.

Inteligencia artificial, como la rama de la ciencia de la computación que se ocupa

de la automatización dejando una entrada, procesos y salida automatizando la conducta inteligente Luger 2005.

La inteligencia artificial tiene como objetivo las capacidades inteligentes, con influencia de la filosofía, matemática, psicología, generar ideas enfocadas en resolución de problemas como áreas específicas, Ejemplo ¿Quién mueve el automóvil autónomamente?, para medir la IA y sus capacidades.

no es solo python, tiene muchas disciplinas involucradas

Enfoques, pero siguen siendo sistemas IA: robots que piensan como humanos, actúan como humanos lógica formal, sistemas expertos sistema que pueda tener un sensor enviar correo electrónico, detectar enfermedades Agentes, asistentes virtuales actúan racionalmente

Parte II

PLN en Java

Parte III

Introducción a Python

Parte IV

Bases PLN

Parte V

NLP - Pipeline

Parte VI

Text Representation

Parte VII

Text Clasification

Parte VIII

Information Extraction

Parte IX

Speech Recognition and Synthesis

Parte X

Chatbots

Parte XI

Search and Information Retrieval

Parte XII

Topic Modelling

Parte XIII

Recommender Systems for
Textual Data

Parte XIV

Metodología de Calificación

Capítulo 2

Cortes

2.1. Porcentajes evaluativos

Cortes 1, 2, 3

- Porcentajes

1. Talleres 20 %
2. Proyecto 20 %
3. Parcial-1 20 %
4. Parcial-2 20 %
5. Parcial-3 20 %