



Big data - Parte 2

Oscar Bustos

September 28, 2023



Pontificia Universidad
JAVERIANA
Bogotá

VIGILADA MINEDUCACIÓN



This template is based on [SINTEF Presentation](#) from [Federico Zenith](#) and its derivation [Beamer-LaTeX-Themes](#) from Liu Qilong and Andrea Gasparini



Table of Contents

1 Introducción

▶ **Introducción**

▶ Índice Invertido

▶ Embeddings de Palabras

▶ Conclusiones

¿Qué es el Procesamiento de Lenguaje Natural?

1 Introducción

- El Procesamiento de Lenguaje Natural (PLN) es una subdisciplina de la inteligencia artificial que se enfoca en la interacción entre las computadoras y los lenguajes humanos.
- Implica el desarrollo de algoritmos y modelos para comprender, interpretar y generar lenguaje humano.



Table of Contents

2 Índice Invertido

▶ Introducción

▶ Índice Invertido

▶ Embeddings de Palabras

▶ Conclusiones

Índice Invertido

2 Índice Invertido

- Un índice invertido es una estructura de datos utilizada en sistemas de recuperación de información.
- Mapea términos (palabras) a sus ubicaciones en un conjunto de documentos o textos.
- Esto permite búsquedas de texto completo eficientes y es un componente fundamental de los motores de búsqueda.



Table of Contents

3 Embeddings de Palabras

- ▶ Introducción
- ▶ Índice Invertido
- ▶ **Embeddings de Palabras**
- ▶ Conclusiones

Embeddings de Palabras

3 Embeddings de Palabras

- Los embeddings de palabras son representaciones vectoriales de palabras en un espacio vectorial continuo.
- Capturan relaciones semánticas entre palabras y se utilizan en diversas tareas de PLN como el análisis de sentimientos, la traducción automática y más.

Verbatim Code Example

3 Embeddings de Palabras

```
#include <stdio.h>

int main() {
    printf("Hello, world!\n");
    return 0;
}
```



Table of Contents

4 Conclusiones

▶ Introducción

▶ Índice Invertido

▶ Embeddings de Palabras

▶ Conclusiones

Conclusiones

4 Conclusiones

- El Procesamiento de Lenguaje Natural es un campo diverso y emocionante con muchas aplicaciones.
- Los índices invertidos permiten una recuperación de texto eficiente.
- Los embeddings de palabras capturan el significado semántico y son cruciales para varias tareas de PLN.

¿Preguntas?

4 Conclusiones

¡Gracias por su atención!
¿Alguna pregunta?