

# Inspira Crea Transforma

## Proyecto Final Almacenamiento y Recuperación de Información

**Inspira Crea Transforma**

UNIVERSIDAD  
**EAFIT**<sup>®</sup>

## **Integrantes**

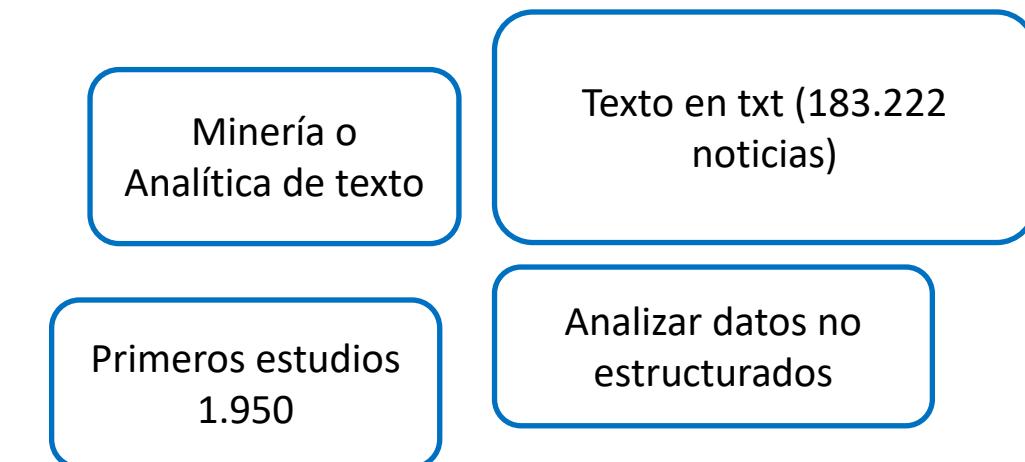
Juan Diego Estrada Pérez

Liceth Mosquera Galvis

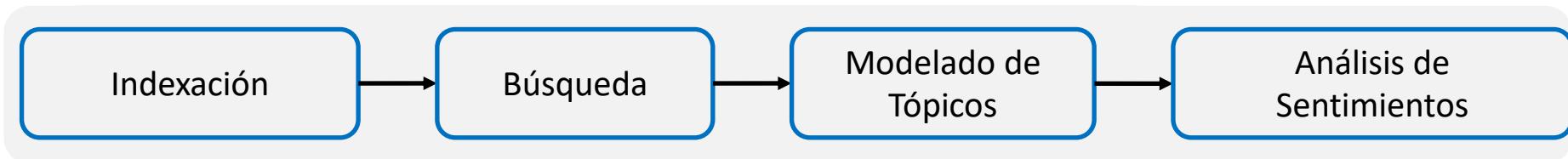
Johan Steward Rios

Juan Mauricio Cuscagua

# Contexto



- Objetivo General
1. Recuperación de la información
  2. Procesamiento de Lenguajes Naturales (NLP)
  3. Extracción de la información relevante
  4. Minería de texto



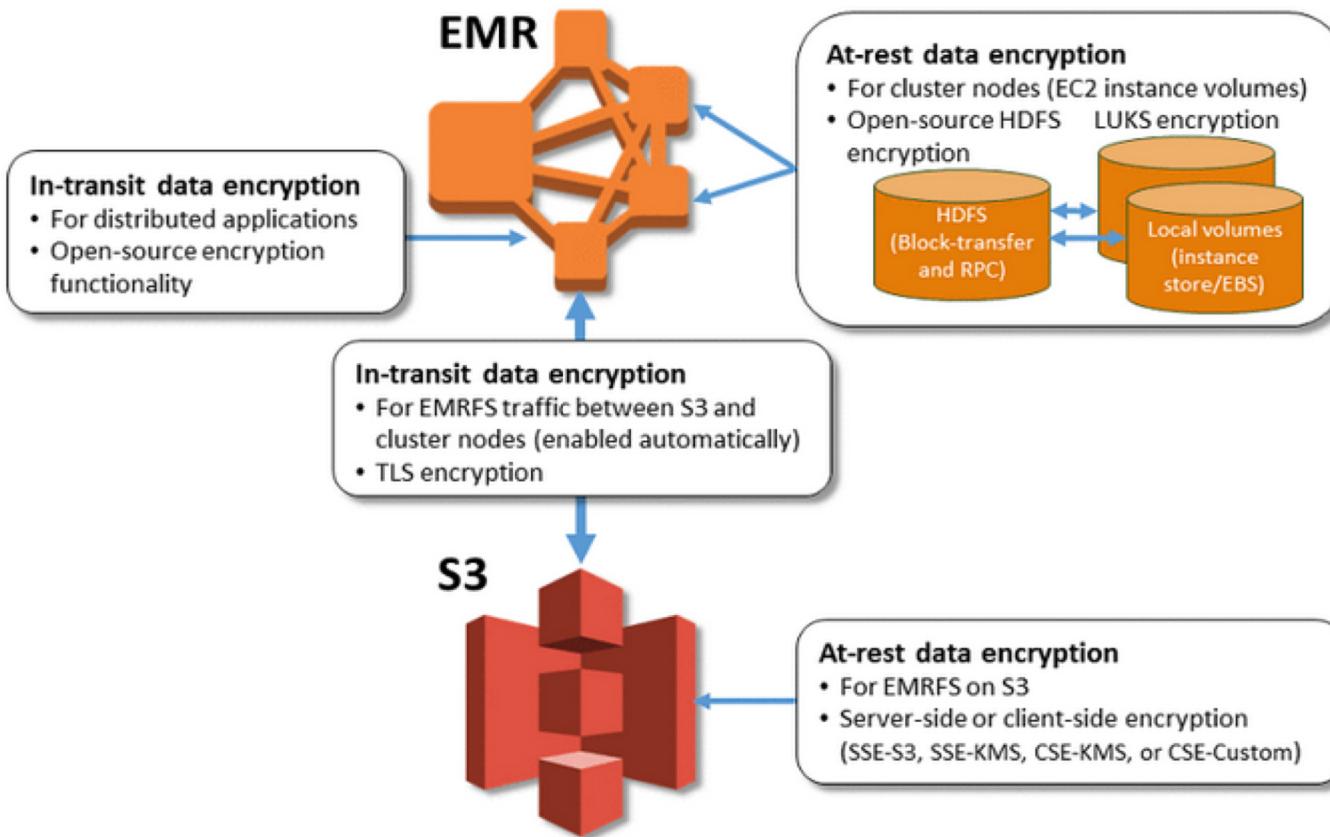


Imagen tomada de:

<https://aws.amazon.com/blogs/big-data/secure-amazon-emr-with-encryption/>

# Meta

1. ¿Qué es?

2. Carga de Información

3. Paquetes

4. Toml

5. Recuperación de Docs

6. Índice Invertido

7. Ranking

	Indri IR	Lucene IR	MALLET ML/NLP	LIBLEAR ML	SVM <sup>MULT</sup> ML	scikit ML/NLP	CoreNLP ML/NLP	META all
Feature generation	✓	✓	✓			✓	✓	✓
Search	✓	✓						✓
Classification			✓	✓	✓	✓	✓	✓
Regression			✓	✓	✓	✓	✓	✓
POS tagging			✓					✓
Parsing							✓	✓
Topic models			✓			✓		✓
<i>n</i> -gram LM								✓
Word embeddings			✓				✓	✓
Graph algorithms								✓
Multithreading		✓	✓			✓	✓	✓

Table 1: Toolkit feature comparison. Citations for all toolkits may be found in their respective comparison sections.

Tabla sacada del artículo META: A Unified Toolkit for Text Retrieval and Analysis

# Modelado de Tópicos

Documents



LDA

Creation of topics

	weight	words
Topic 1	3%	flower
	2%	rose
	1%	plant
...		
Topic 2	2%	company
	1%	wage
	1%	employee

Topics allocation to documents

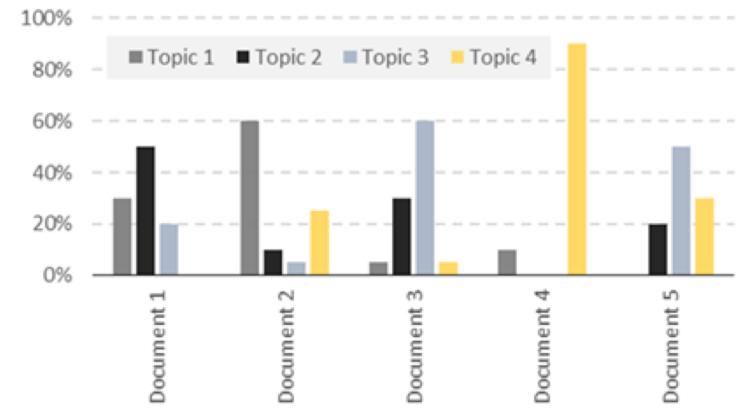


Imagen creada por:

<https://towardsdatascience.com/the-complete-guide-for-topics-extraction-in-python-a6aaa6cedbbc>

# Análisis de sentimientos



Imagen bajada de [www.freepng.es](http://www.freepng.es) y editada por el autor de este blog:

<https://medium.com/@tomyuz/a-sentiment-analysis-approach-to-predicting-stock-returns-d5ca8b75a42>

# Referencias

Una ampliación del trabajo así como las referencias utilizadas para lograr el mismo pueden ser encontradas en el documento Text mining process.pdf. Esta presentación, el documento y la demo se encuentran disponibles en el siguiente github:

[https://github.com/jdestradap/text\\_mining](https://github.com/jdestradap/text_mining)