

082057 – Procesamiento del Lenguaje Natural

Trabajo Práctico Individual

CODIGO: 03 – Detección <https://normas-apa.org/formato/> de Emociones y Recuperación de la Información (Information Retrieval)

1 LINEAMIENTOS

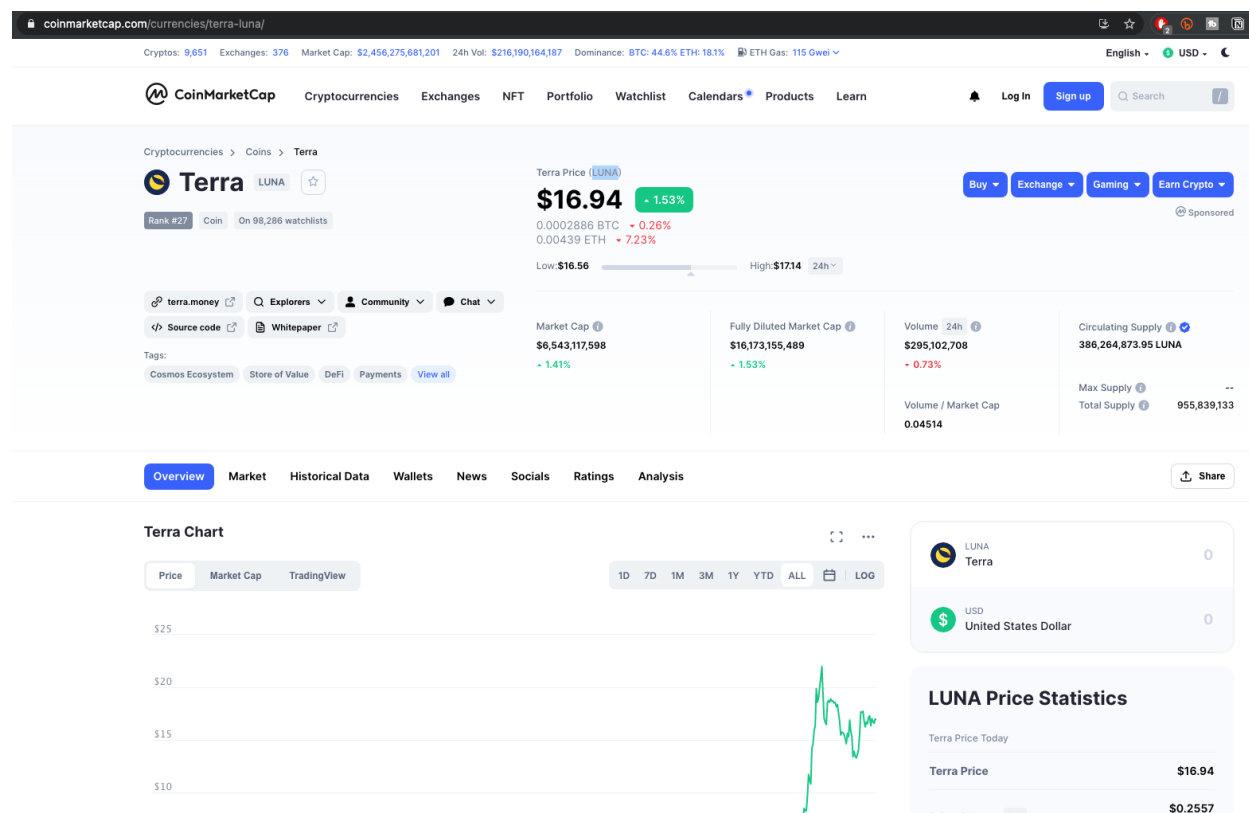
- El trabajo práctico es 100% individual.
- Todo el código fuente desarrollado (si hubiera) así como el documento de texto que lo acompaña con las justificaciones, técnicas, bibliografía, autores y demás correctamente citados, deben ser subidos a la cuenta de Github <https://github.com/> **personal y privada** de cada estudiante y compartida con la cuenta de github del profesor (<https://github.com/hernanborre>). **Sólo así se considera el trabajo entregado.**
- El/La estudiante, deberá realizar una **defensa oral** de su código, técnicas utilizadas, citar autores utilizados si es necesario y poder expresar claramente tanto su desarrollo cognitivo a la solución, así como las conclusiones obtenidas.
- Este trabajo práctico constituye la única evaluación en primera instancia de esta materia, por lo cual la producción de la solución del mismo se espera que esté a la altura o supere el tiempo dedicado a un parcial.
- Fecha de entrega: **24 de Junio y 8 de Julio de 2022**

2 CONSIGNA

El análisis y detección de sentimientos (positivos, negativos o neutros) es uno de los campos más importantes del Procesamiento del Lenguaje Natural. Diversas técnicas y estado del arte (SOTA) son aportadas día a día por la comunidad científica, lo cuál luego lleva a la implementación de los mismos por parte de las organizaciones.

En este Trabajo Práctico, además de aplicar análisis de sentimientos, se deberá usar accesos a información de diversas APIs (las que se nombren en este enunciado son solamente a modo de ejemplo, sin limitar de ninguna forma las que puedan ser implementadas). Además, técnicas de Recuperación de Información (Information Retrieval) y Web Scrapping deberán ser usadas para capturar y recuperar la información analizada sobre los tópicos o palabras claves (keywords) pedidas.

En este año las keywords serán nombres de criptomonedas y su correspondiente símbolo con el cuál operan en los mercados (Ejemplo, para Bitcon será Bitcon y \$BTC, o para Ethereum, \$ETH, o en el screenshot adjunto se muestra la moneda TERRA LUNA y su símbolo es \$LUNA, y así sucesivamente). **El TP se considera complejo con solamente elegir y realizar el análisis de una sola criptomenda a tu elección.**



Dado que es un proyecto que requiere de Information Retrieval, se requiere implementar una base de datos para almacenar la información analizada (queda a criterio del/a estudiante qué base de datos utilizar) – se sugiere una de estilo no relacional Documental – mongoDB – o bien la más utilizada en el campo es ElasticSearc. (sin embargo con la primera opción es suficiente para este Trabajo Práctico)

Se deben utilizar varias fuentes de información – **al menos 3** - (por ejemplo: google.com, bing.com. twitter, reddit, cointelegraph) desde dónde se tomará información de las 20 criptomonedas elegidas previamente, se guardarán sus X principales palabras (frequency), su timestamp, sus NERs (Named Entity Recognition) y su Sentimiento asociado al pedazo de información analizado. Se pueden combinar varias fuentes de información para tomar una decisión final.

Cómo opcional se puede guardar un timestamp del momento dónde se evaluá y predice el sentimiento así como el valor en el mercado de esa criptomoneda (por ejemplo se puede utilizar la API de coinmarketcap o de coingecko, entre otras)

2 ENTREGA, CODIGO Y EJEMPLOS

El código deberá ser escrito en Python 3 y/o Javascript / Typescript. Se pueden usar todas las librerías que se crean necesarias.

Además del código desarrollado, se debe entregar un documento de texto (puede ser un .doc o docx, o bien recomendamos usar la herramienta de generación de textos científicos LaTeX), en la cual se explica la solución y se le da crédito a los autores consultados, las técnicas usadas, el código fuente o ejemplos tomados de otros blogs, videos, etc.

Las citas deben ser en formato **APA**