Recuperación de Información y Recomendaciones en la Web

Proyecto: Token Analyzer

Integrantes	**CI**
Aymara Nohely Melo Cuello	4.944.732-2
Nicolás Roberto Comerci Zunino	5.132.630-8
Santiago Nicolás Martinez Fernandez	5.085.397-8

Introducción:

El objetivo del proyecto se centra en el mundo de las criptomonedas y cómo los usuarios interactúan haciendo uso de estas.

Una criptomoneda o token es la representación virtual o digital de un valor económico. En este proyecto se va a trabajar sobre uno de sus usos, la adquisición de servicios o productos financiados a través del correspondiente valor económico digital de los tokens.

Para ello existe un registro de transacciones llamado Blockchain, donde se almacena un bloque con información asociada a todas las transacciones digitales realizadas. Además, cada usuario puede crear su wallet desde la cual podrá interactuar con la blockchain, guardando activos, enviándolos a otras wallets, pagando comisiones de la red, etc.

Dentro del ecosistema existen diferentes redes o blockchains, entre las que se encuentran: Bitcoin, Ethereum, Cardano, Solana, Binance Smart Chain (BSC), etc. Cada una tiene diferentes características que la distinguen como velocidad en las transacciones, el costo de gas o fees que se necesitan pagar para efectuar una transacción, las criptomonedas que soporta la red, y las diferentes dapps asociadas a la misma.

En este contexto es de interés enfocarse en la BSC, dado que a día de hoy es la red que alberga una gran cantidad de proyectos populares entre usuarios ajenos al ecosistema. Esto se debe a las bajas comisiones de la red y a su similitud con la red de Ethereum la más popular para contratos inteligentes, sin embargo, esta presenta comisiones enormes que hacen inaccesible para los usuarios menos adinerados el uso de la red.

En el proyecto se va a utilizar la información que guarda la blockchain de la BSC, la cual va a ser accedida mediante distintas formas, con el objetivo de recabar información, dado que para que los usuarios puedan realizar una transacción de forma segura, es importante que tengan consciencia de cuáles son las direcciones de plataformas, individuos o sitios inseguros y maliciosos con los cuales se debe evitar interactuar de cualquier forma.

Por este motivo el proyecto se centra en obtener la información de seguridad asociada a un contrato o dirección (los cuales identifican a las criptomonedas), estos son representados por cadenas hexadecimales como por ejemplo “0xd3d865e400e1686ce35d83c642a96f13d46946d5”. La información se obtendrá mediante los sitios web de BsCheck y TokenSniffer. Estos servicios permiten analizar Tokens de la red Binance Smart Chain, respaldada por Binance (plataforma exchange para la compra, venta e intercambio de criptomonedas). De esta forma se obtendrá la información que le permitirá al usuario saber si está realizando una transacción poco segura, o a un sitio sospechoso

Problema:

Dado que para una transacción realizada por un usuario la información registrada puede no ser fácil de interpretar, y a su vez los usuarios con poco conocimiento del funcionamiento o que iniciaron recientemente en el ecosistema de las criptomonedas pueden no detectar que están realizando una transacción insegura, es de vital importancia tener conocimiento de las criptomonedas que se poseen y su grado de seguridad.

En los tiempos recientes la principal estrategia de fraude utilizada por los usuarios maliciosos consiste en enviar en forma de 'regalo' un token cuyo valor en dólares es muy alto. Por tal motivo los usuarios que reciban este token se van a ver impresionados por encontrar que el valor de sus wallet asciende a cifras gigantes de miles de dólares, emocionados por dichas ganancias los usuarios intentan cambiar esos tokens por el equivalente al dólar en las criptomonedas que son las monedas estables o 'stablecoins'. Para efectuar este intercambio el usuario debe de firmar el contrato del token malicioso (mecanismo que es común a todas las criptomonedas y, por lo tanto, no genera sospecha alguna, es similar al 'términos y condiciones' comúnmente observado en productos de software que nadie lee), al firmar el contrato el usuario da acceso total a su wallet, y aprovechándose de esto el atacante vacía la cuenta del usuario, quedando la wallet comprometida y sin fondos.

Enfoque de la solución:

Con base en el problema planteado se propone la creación de una app web que permita a los usuarios escanear su wallet personal y la aplicación genere un reporte de los tokens que posee el usuario en dicha wallet. Además, para cada criptoactivo se mostrará el balance del mismo en dólares, de esta forma el usuario puede fácilmente tener un control sobre sus activos y el valor de los mismos. Además, la aplicación basándose en ciertas evidencias en los contratos de los tokens y la reputación de los mismos en internet, generará un informe del grado de riesgo que conlleva utilizar el token y los contratos relacionados con el mismo.

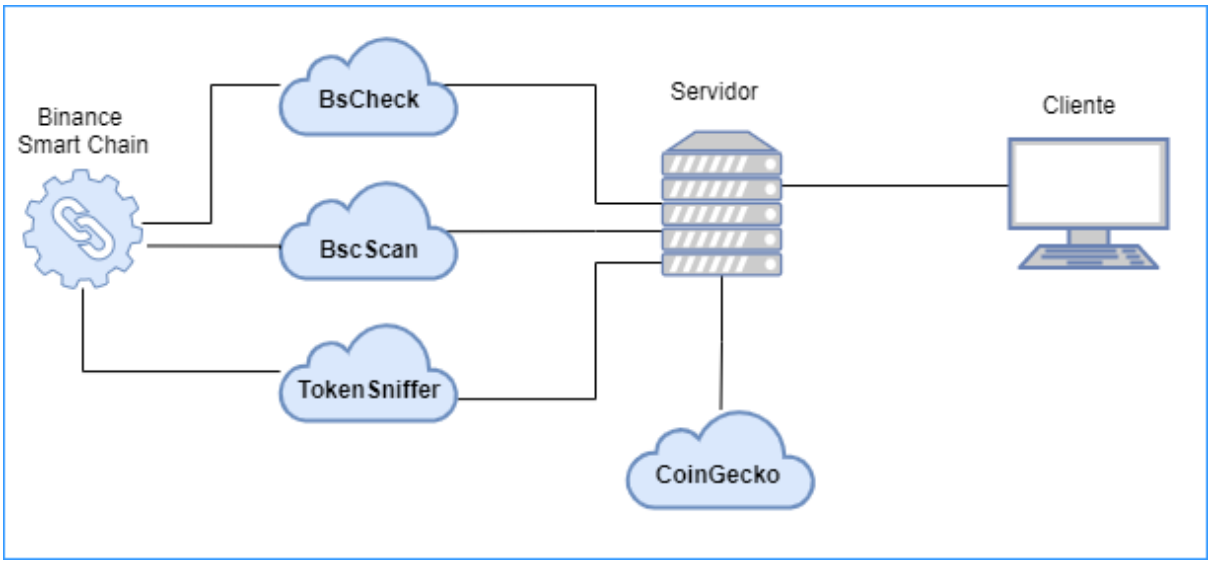
Diseño:

Como se mencionó anteriormente la solución propuesta consiste en una aplicación web a través de la cual los usuarios sean capaces de escanear su wallet para ver los activos que poseen y el valor de los mismos en dólares.

Para esto la aplicación utilizará la api proporcionada por BscScan y obtendrá los tokens que posea dicha wallet en la blockchain de BSC. Luego haciendo uso de la api proporcionada por CoinGecko (portal web con información de criptoactivos de relevancia), se obtendrá el valor de los activos expresados en dólares, además de información variada de los mismos como el histórico de precios, el nombre del token y la imagen que lo identifica popularmente, entre otros.

Posteriormente, aquellos tokens que le generen sospecha al usuario (ya sea porque los análisis anteriores no brindaron información acerca del mismo o porque el usuario no recuerda haberlos adquirido por voluntad propia), podrán ser sometidos a un análisis de seguridad. Para ejecutar los análisis mencionados se hará uso de los sitios 'BSCheck' y 'TokenSniffer', estos proveen ciertas medidas de seguridad con las cuales se puede cuantificar el nivel de seguridad de una criptomoneda. Dado que dichos sitios no ofrecen una API se realizará scrapping sobre los mismos.

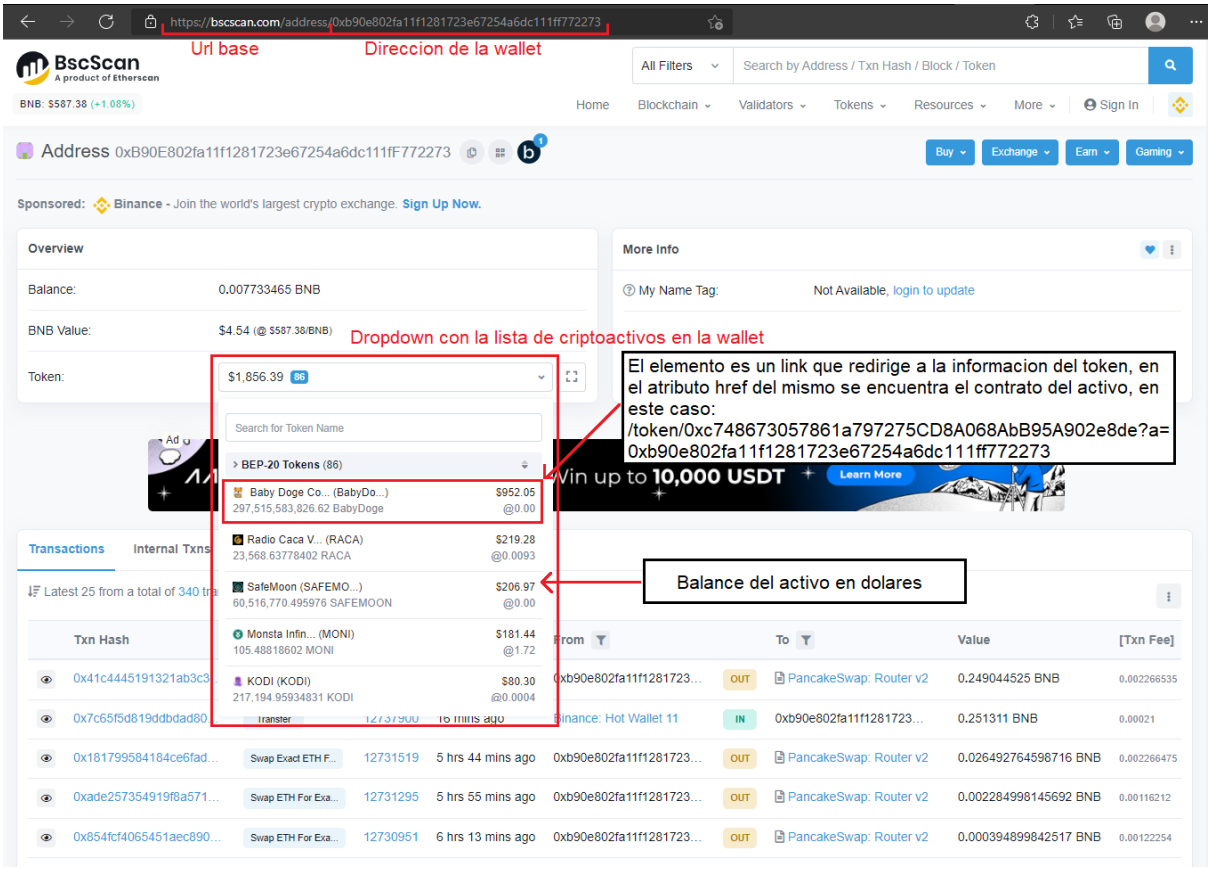
El siguiente diagrama ilustra la arquitectura de la aplicación, los usuarios correrán un Frontend web, el cual a su vez consumirá los servicios provistos por un servidor backend que hará uso de las herramientas mencionadas anteriormente. Además, la aplicación web requerirá que los usuarios se registren e inicien sesión cuando deseen utilizar la web, esto como medida de seguridad antispam, y permitirá fácilmente en el futuro guardar datos sobre los usuarios, o cachear información sobre los mismos.



Implementación:

Para la construcción del Frontend se utilizó la librería de JavaScript React.js y para el backend se utilizó el lenguaje Python con el framework Flask para desarrollo de APIs web. A su vez se tiene un registro donde se guarda la información de los usuarios registrados en la plataforma, como se mencionó en la sección de diseño esto se realizó principalmente con el objetivo de agregar una capa de seguridad antispam, pero en un futuro el mismo podría ser usado para cachear información de las consultas realizadas por los usuarios y de esa forma acortar el tiempo de ejecución de las mismas.

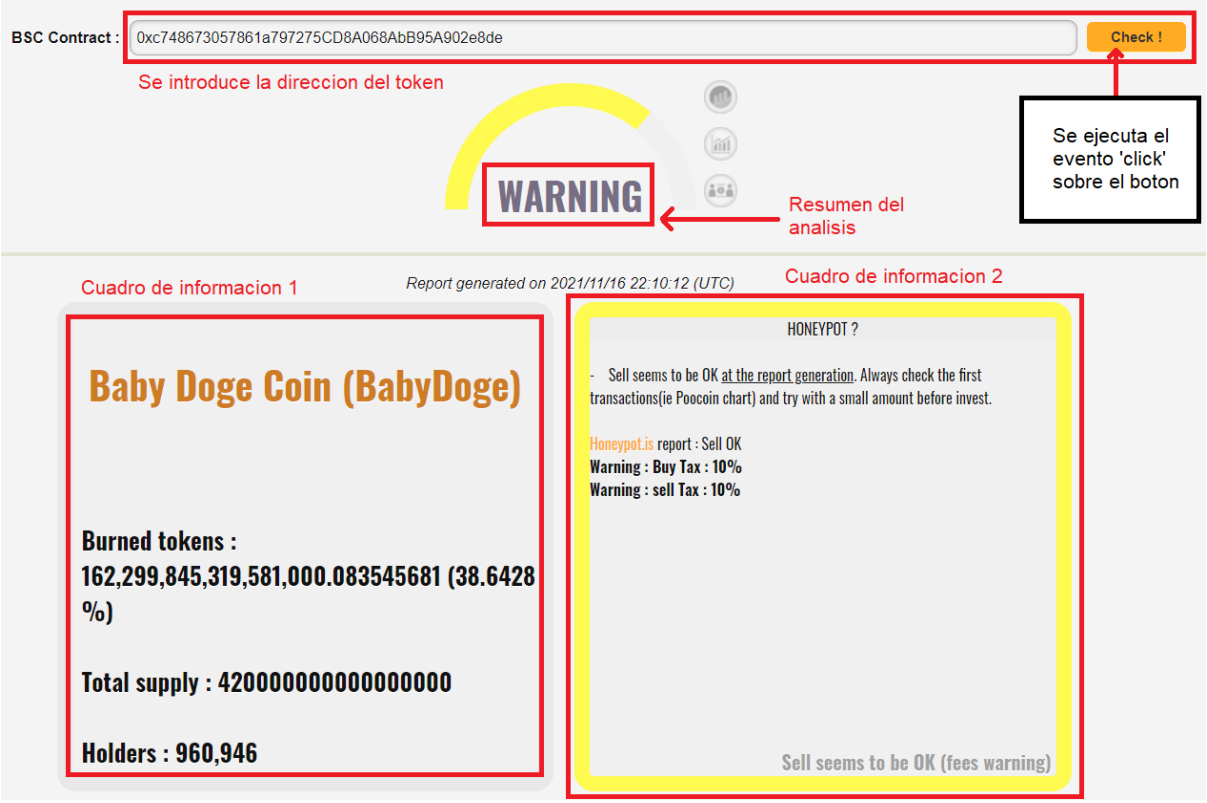
Durante la implementación surgió el problema de que los principales servicios a consumir expuestos por la API del sitio web BSCScan, necesarios para la construcción de la plataforma eran privados y la herramienta solicitaba el pago de una suscripción para poder hacer uso de dichos servicios. Para subsanar el problema el equipo decidió hacer scrapping sobre la web misma de la herramienta, ya que esta lista toda la información de la wallet que se analiza a través de su buscador integrado.



En la imagen anterior se puede apreciar el esquema que presenta la web de BSCScan, para acceder a la vista anterior es necesario obtener el HTML de la URL base (como se muestra en la imagen) más la dirección de la wallet al final de la misma. Una vez obtenido el HTML basta con obtener la lista de ítems del Dropdown y de estos extraer la dirección del contrato con una regex de la etiqueta <a /> en el atributo href. Para finalizar se obtuvo además el balance del criptoactivo en dólares desde el atributo señalado en la imagen.

Una vez obtenidos los tokens y su balance en dólares, solo resta obtener información a partir de sus contratos para mostrársela al cliente. Con dicho fin se utilizó la API provista por CoinGecko, la misma recibe el contrato del token y retorna un JSON con toda la información relevante del mismo: su histórico de precios, el nombre de la criptomoneda, la imagen que popularmente la identifica, entre otra información que no es relevante para la solución.

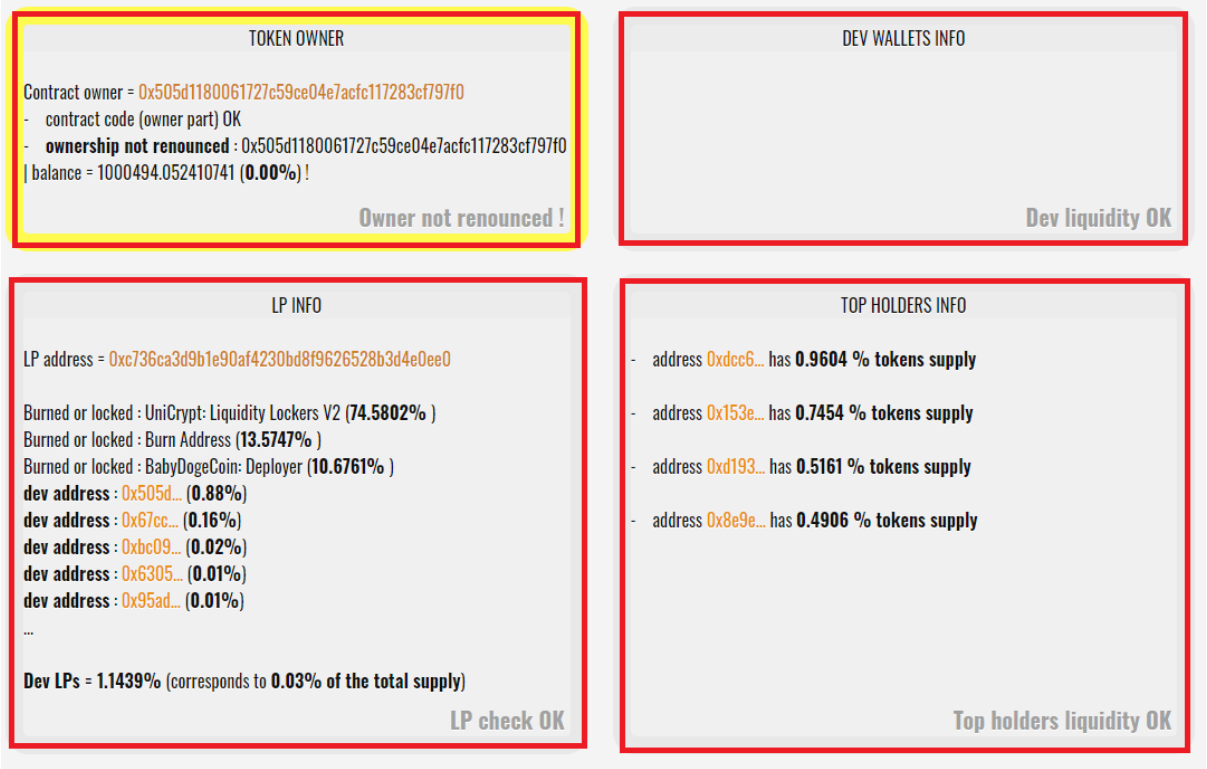
Para la segunda funcionalidad de analizar las criptomonedas/tokens en busca de posibles estafas o inseguridades es que se procede a realizar el scrapping sobre la web de BSCheck.



A diferencia del scrapping anterior en este caso para extraer la información fue necesario interactuar con la web ejecutando JavaScript. Primero se introduce la dirección del token en la barra de búsqueda, luego se ejecuta el evento click del botón para que comience el análisis. Una vez finalizado este la página muestra un resumen global del análisis con el grado de riesgo de la criptomoneda.

En el cuadro de información 1, primero se muestra el nombre del token. Luego se muestra la cantidad de tokens quemados, en el ecosistema de criptomonedas ‘quemar un token’ significa que cierta cantidad del supply total es enviada a una wallet de la cual no se pueden extraer nunca más, de esta forma se reduce la oferta del mismo, acción que normalmente tiene como resultado el aumento del valor del mismo. Luego en el cuadro aparece el supply total de la criptomoneda y la cantidad de holders, es decir cuánta gente utiliza/confía en dicho activo.

En el cuadro de información 2, se obtiene si la criptomoneda se trata de la clásica estafa conocida como 'Honeypot' o tarro de miel, esta consiste en que la criptomoneda en cuestión no tiene implementada una función de venta, es decir el activo solo se puede comprar, de esta forma los usuarios al ver que las gráficas de precio solo aumentan comienzan a comprar con mayor esmero hasta que se dan cuenta de que no son capaces de vender y de esta forma se va llenando la liquidez o el 'tarro de miel' hasta que el desarrollador de la criptomoneda decide vender (el sí puede hacerlo) todo lo que él posee llevándose toda la liquidez en dólares del activo.



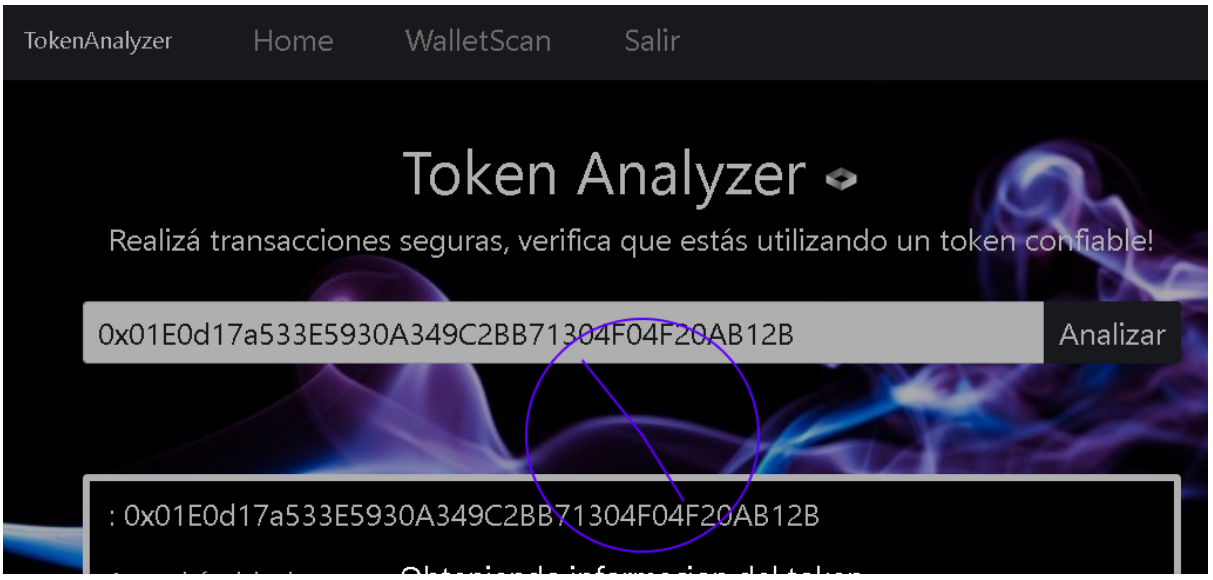
El análisis del token owner es bastante irrelevante por el motivo de que a día de hoy los proyectos que han llegado a la descentralización completa son muy pocos. La información sobre las wallets de los devs que se obtiene (en el ejemplo no aparece) hace referencia a que porcentaje del supply total les pertenece a estos y en caso de que el valor sea muy elevado podría ser un peligro, dado que los devs serían capaces de manipular el precio de mercado del activo. El checkeo de LP (Liquidity Pool) verifica que nadie pueda manipular la liquidez del activo, esto es sumamente importante para evitar estafas conocidas como RugPulls. Por último el análisis muestra información sobre los holders con mayor porcentaje del supply total, esto es de interés para conocer cuál es la influencia que puede llegar a tener en el mercado una sola wallet.

Con respecto a la herramienta del Token Sniffer esta no pudo ser utilizada debido a las protecciones de captcha que hacen inviable scrappear la web, por dicho motivo se colocó un botón en la solución que redirige al sitio web en cuestión con el token que se desea analizar.

Funcionalidades y uso:

El sistema presenta dos grandes funcionalidades, donde la primera es el análisis de Tokens de la blockchain de Binance. Estos Tokens son criptomonedas que los usuarios pueden tener en sus wallets y realizar transferencias con ellas. Antes de invertir o por cualquier otra razón obtener un Token, es una buena idea realizar una investigación sobre el mismo, ya que existen Tokens maliciosos que son usados para realizar fraudes financieros, de los que uno no quiere ser víctima.

Nuestro sistema permite realizar un análisis del Token para obtener información de seguridad sobre él:

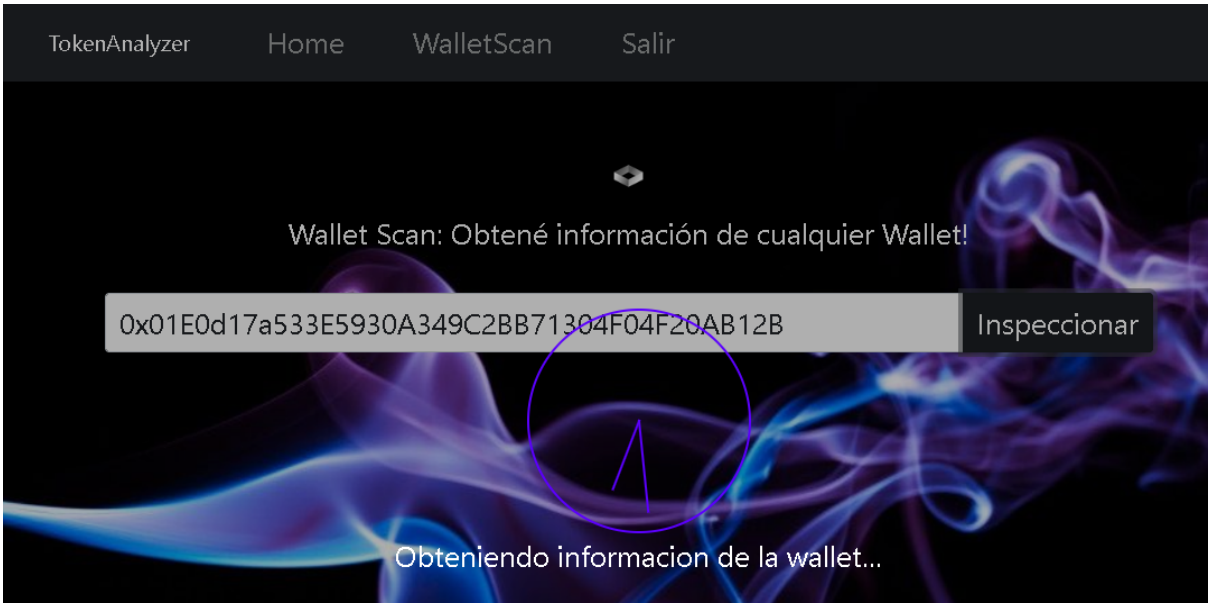


(Escaneo de seguridad de un Token en proceso)

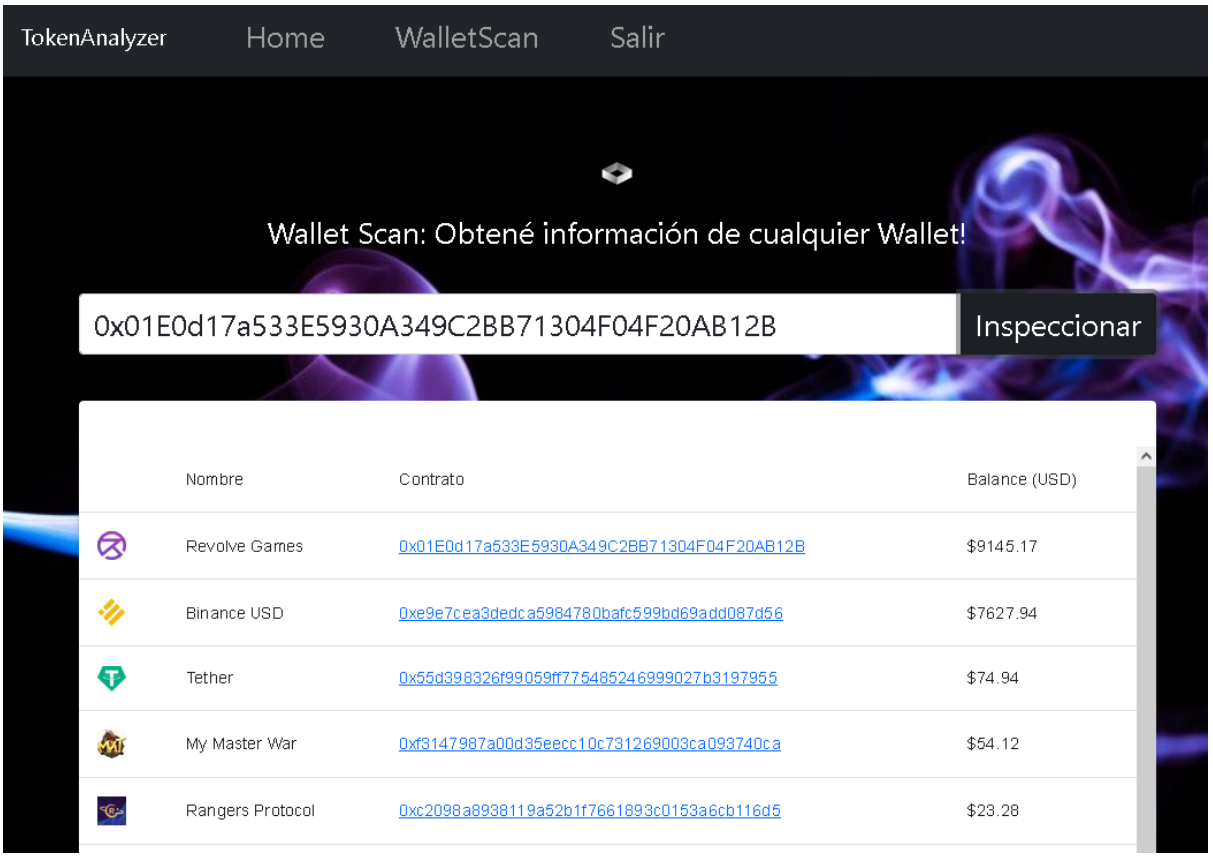
(Escaneo terminado, muestra de información de seguridad del Token)

La información presentada por esta funcionalidad permite a los usuarios tomar decisiones informadas y responsables a la hora de decidir si se desea comprar, invertir o conservar cierto balance de este Token.

La segunda funcionalidad permite analizar una Wallet con el fin de determinar que Tokens tiene y el balance de dichos Tokens en dólares:



(Escaneo de una Wallet en proceso)



(Resultados de una Wallet y sus Tokens)

Con esta información, el usuario puede revisar los Tokens contenidos en la Wallet escaneada. Además, el sistema también permite hacer click en cualquiera de los Tokens contenidos en la Wallet, iniciando un escaneo de seguridad del mismo. De esta forma se relaciona la segunda funcionalidad con la primera, ya que pudiendo analizar los Tokens de la Wallet se puede tener una mejor idea del estado de esta.

Conclusiones:

Dado que para una transacción realizada por un usuario la información registrada puede no ser fácil de interpretar, y a su vez la información se encuentra dispersa en varios sitios, contar con un sitio web en el cual pueda analizarse tanto una Wallet como un Token es de gran utilidad. Con ello se logra mantener los datos centralizados en un solo lugar para poder cotejarlos o analizarlos de forma sencilla.

La posibilidad de analizar una Wallet y a su vez cada uno de los Token asociados a la misma le brinda al usuario mayor seguridad y sentimiento de confianza respecto a sus transacciones o movimientos. Se contará con información al alcance de la mano, en cualquier momento y con una interfaz amigable que permita una interpretación de los datos sin mayor esfuerzo.

Una de las dificultades enfrentadas en la recuperación de información fue que para obtener acceso a los datos de algunas APIs era necesario una suscripción paga. Motivo por el cual se tuvo que recuperar información de diversos sitios para que en conjunto se llegase al producto completo que se quería ofrecer.

Otra dificultad fue que al momento de acceder a algunos sitios, si bien los mismos no requerían una suscripción paga, tenían control de bots y eso requirió buscar la manera de recuperar la información sin ser bloqueados por dichos controles.

En conclusión se considera que la aplicación Web desarrollada cumple con las expectativas respecto a sus funcionalidades. Lo que se podría mejorar en cuanto a experiencia de usuario son los tiempos de respuesta, ya que al reunir la información de varios sitios se demora la respuesta a la request enviada a la API de la aplicación al analizar (ya sea Wallet o Token).

Trabajo a futuro:

Si bien el producto actual cumple su función y es más que suficiente como prueba de concepto, ciertos aspectos de su implementación podrían no ser apropiados para un sistema de producción, o al menos, podrían ser poco rendidores o escalables como para proporcionar una experiencia de usuario de primer nivel.

En primer lugar, toda la información de los tokens y wallets se obtiene, sea mediante queries o scrapping, desde recursos externos. Esto limita el rendimiento, velocidad y volumen de datos que se puede procesar en un momento dado.

La solución es clara: cachear y precomputar sobre los datos obtenidos. De esta forma el costo de obtener datos externos y procesar los mismos deberá ser pagado solo la primera vez que se piden los datos de un token o wallet específico. Al ingresar estos datos en nuestro propio corpus se aceleran las subsecuentes consultas. Además, los datos sensibles al tiempo pueden ser actualizados desde las fuentes externas si los datos de nuestro corpus son suficientemente viejos (por ej, una vez por día).

Este corpus podría ser implementado, por ejemplo, usando ElasticSearch, que permite guardar los datos en una base NoSql donde se puede buscar por cualquier campo que se desee usando la técnica de indexación reversa. Esto haría que nuestra recuperación de información fuera órdenes de magnitud más rápida.

Otra opción aún más drástica sería omitir recursos externos (por ejemplo en el caso de obtener información de las wallets) y obtener los datos directamente desde la blockchain. Esto puede sonar extremadamente complejo y computacionalmente caro, pero no lo es. La blockchain de Bitcoin es de 360GB aprox.[1], mientras la de Binance es de 1.07TB[2]. En términos de almacenamiento para hoy en día, esto no es para nada excesivo. Por ejemplo, en Amazon Web Services, a razón de $0.125/GB/Mes[3], costaría $45/Mes alojar la blockchain de Bitcoin, $133.76/Mes alojar la blockchain de Binance, $178.75/Mes alojar ambas. Menos de $200 como costo operativo por mes, lo que es muy razonable.

Alojar la blockchain como nodo participante en ella tiene la ventaja de que automáticamente recibimos actualizaciones de todas las transacciones, tokens y wallets desde la misma. Esto significa que podemos reaccionar a la nueva información y precomputar analíticas en el momento que recibimos constancia de una nueva transacción.

De esta forma, sería posible expandir el proyecto actual a uno que no simplemente recupera información desde fuentes externas, sino que podría realizar esta misma función pero con un rendimiento extremadamente superior, consultando a la fuente primaria, la blockchain. Requeriría un esfuerzo de desarrollo substancial, pero perfeccionaría el producto.

---

[1] https://www.statista.com/statistics/647523/worldwide-bitcoin-blockchain-size/
[2] https://ycharts.com/indicators/ethereum_chain_full_sync_data_size
[3] https://aws.amazon.com/ebs/pricing/

Referencias:

- *Binance Smart Chain Whitepaper*. (s. f.). Binance.org. Recuperado 18 de noviembre de 2021, de https://www.binance.org/en/smartChain
- *Bscheck*. (s. f.). Bscheck.eu. Recuperado 18 de noviembre de 2021, de https://www.bscheck.eu/
- *CoinGecko: Precios de criptomonedas y capitalización del mercado*. (s. f.). CoinGecko. Recuperado 18 de noviembre de 2021, de https://www.coingecko.com/es
- *¿Cómo funciona Bitcoin? - Bitcoin*. (s. f.). Bitcoin.org. Recuperado 18 de noviembre de 2021, de https://bitcoin.org/es/como-funciona
- *¿Qué es la tecnología de blockchain? - IBM Blockchain | IBM*. (s. f.). IBM Corp. Recuperado 18 de noviembre de 2021, de https://www.ibm.com/mx-es/topics/what-is-blockchain
- *¿Qué es un contrato inteligente?* (s. f.). Coinbase. Recuperado 18 de noviembre de 2021, de https://www.coinbase.com/es-LA/learn/crypto-basics/what-is-a-smart-contract
- *¿Qué es una wallet o monedero de criptomonedas?* (s. f.). Bit2Me Academy. Recuperado 18 de noviembre de 2021, de https://academy.bit2me.com/wallet-monederos-criptomonedas/
- *¿Qué son las criptomonedas y cómo funcionan?* (s. f.). Guias Santander. Recuperado 18 de noviembre de 2021, de https://www.santander.com/es/stories/guia-para-saber-que-son-las-criptomonedas
- *Token Sniffer*. (s. f.). tokensniffer.com. Recuperado 18 de noviembre de 2021, de https://tokensniffer.com/
- *What Is a Rug Pull?* (s. f.). CoinMarketCap. Recuperado 18 de noviembre de 2021, de https://coinmarketcap.com/alexandria/glossary/rug-pull