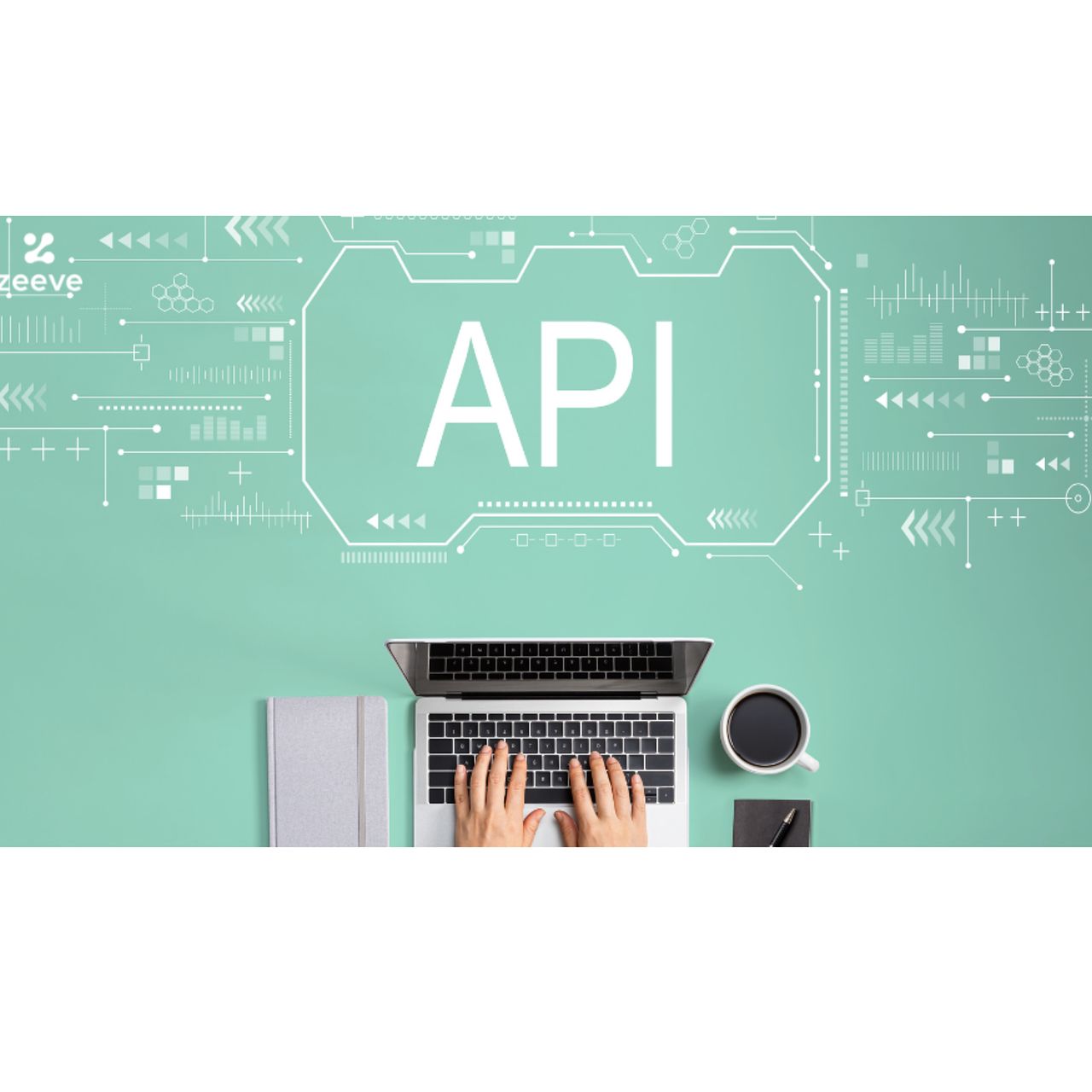


Explotación de datos web

Año: 2023

Ensayo

“API en bioinformática”



Salim Taleb, Nasim A

Docentes: Alejandro J. Hadad; Walter Elías; Maximiliano Fernandez

Carrera: Tecnicatura universitaria en procesamiento y explotación de datos

# ¿Qué es una API?

Una API (Application Programming Interfaces) es una interfaz de programación de aplicaciones. Es un conjunto de reglas que permiten la comunicación entre dos aplicaciones.

Generalmente, se utilizan para comunicar servicios web, dónde se estructura una petición para enviar al servidor, el cual procesa esta petición y devuelve los resultados pertinentes.

Facilita el desarrollo de aplicaciones, ya que permite con una simple petición procesar datos y devolver resultados sin necesidad de crear el código o utilizar procesamiento local. Algunas API tiene un acceso limitado a la cantidad de peticiones permitidas y también pueden cobrar dinero para permitir una mayor cantidad de accesos o tareas disponibles.

# Reactome

Reactome es una base de datos de rutas biológicas de acceso libre, de código abierto, curada y revisada por pares. Su objetivo es proporcionar herramientas de bioinformática intuitivas para la visualización, interpretación y análisis del conocimiento de las rutas para apoyar la investigación básica y clínica, el análisis del genoma, la modelización, la biología de sistemas y la educación. Fundado en 2003, el proyecto Reactome es dirigido por Lincoln Stein de OICR, Peter D'Eustachio de NYU Langone Health, Henning Hermjakob de EMBL-EBI y Guanming Wu de OHSU.

El pilar de Reactome es una base de datos relacional de libre acceso y de código abierto de moléculas de señalización y metabólicas y sus relaciones organizadas en rutas y procesos biológicos. La unidad básica del modelo de datos de Reactome es la reacción. Las entidades (ácidos nucleicos, proteínas, complejos, vacunas, terapéuticos anticancerígenos y pequeñas moléculas) que participan en las reacciones forman una red de interacciones biológicas y se agrupan en rutas. Ejemplos de rutas biológicas en Reactome incluyen el metabolismo intermediario clásico, la señalización, la regulación transcripcional, la apoptosis y las enfermedades. El proceso de curación de Reactome para una ruta es similar a la edición de una revisión científica. Un experto externo en el dominio aporta su experiencia, un curador lo formaliza en la estructura de la base de datos y un experto externo en el dominio revisa la representación. Un sistema de seguimiento de evidencia asegura que todas las afirmaciones estén respaldadas por la literatura primaria.

Todos los datos y software están disponibles para su descarga gratuita. Los datos de interacciones, reacciones y rutas se proporcionan como archivos planos descargables, Neo4j GraphDB, MySQL, BioPAX, SBML y PSI-MITAB y también son accesibles a través de nuestras API de servicios web. El software e instrucciones para la instalación local de la base de datos de Reactome, el sitio web y las herramientas de entrada de datos también están disponibles para permitir curación independiente de las vías. [What is Reactome?](https://reactome.org/what-is-reactome)

Reactome contiene dos servicios de API diferentes, uno que permite realizar servicios de [análisis](https://reactome.org/dev/analysis) y otro servicio de [contenido](https://reactome.org/dev/content-service), que permite acceder a los datos curados. A su vez, existe una librería llamada “[Reactome2py](https://github.com/reactome/reactome2py)” que facilita esta comunicación con las API. Finalmente, lo más útil de reactome, sus [gráficos de vías](https://reactome.org/dev/pathways-overview/js#API), que pueden utilizarse mediante una API de JavaScript en páginas web.

Caso de uso

Se adjuntan dos casos de utilización con códigos de Python:

reactome\_req.py: En este caso se realiza un pedido directamente a la API de Reactome mediante la librería requests, una vez devueltos los resultados se cuentan la cantidad de vías afectadas por cada gen y se muestran en un gráfico utilizando la librería matplotlib.

reactome\_reactome2py.py: Este caso es un poco más complejo, se busca recrear la salida de la herramienta web de Reactome, pero en una aplicación de ventana de PyQt5, en este caso para comunicarse con la API de Reactome se utiliza la librería reactome2py, enviando una petición de análisis al servidor de Reactome, con los genes a analizar, y luego creando una ventana que contenga un explorador web que accede a los resultados de dicho análisis.

# g:Profiler

g:Profiler es un servidor web público para caracterizar y manipular listas de genes. g:Profiler tiene una interfaz de usuario amigable y fácil de usar, con visualizaciones potentes y está actualmente disponible para más de 400 especies, incluyendo mamíferos, plantas, hongos, insectos de Ensembl y Ensembl Genomes. g:Profiler se actualiza aproximadamente cada tres meses y sigue los lanzamientos trimestrales de las bases de datos de Ensembl.

El conjunto de herramientas g:Profiler incluye:

g:GOSt, el núcleo de g:Profiler, realiza análisis de enriquecimiento estadístico para proporcionar interpretación a las listas de genes proporcionadas por el usuario. Las listas de genes pueden ser planas u ordenadas. Aceptamos la mayoría de los tipos de identificadores, regiones cromosómicas y ID de términos de función. Proporcionamos datos de múltiples fuentes de evidencia funcional, incluyendo términos de la Gene Ontology, rutas biológicas, motifs regulatorios de factores de transcripción y microARN, anotaciones de enfermedades humanas y interacciones proteína-proteína.

g:Convert es una herramienta de conversión de identificadores de genes. Utiliza información en las bases de datos de Ensembl para manejar cientos de tipos de ID para genes, proteínas, transcritos, probesets de microarray, etc, para muchas especies, plataformas experimentales y bases de datos biológicas. g:Convert es flexible: acepta una lista mixta de ID y reconoce automáticamente sus tipos. También puede servir como un servicio para obtener todos los genes pertenecientes a una categoría funcional particular.

g:Orth es una herramienta para mapear genes homólogos entre organismos relacionados según los datos de Ensembl. Dado un organismo objetivo seleccionado, g:Orth recupera los genes del organismo objetivo que son similares en secuencia a los genes iniciales en la entrada.

g:SNPense es una herramienta para mapear polimorfismos de nucleótido único en humanos (SNP) a nombres de genes, ubicaciones cromosómicas y términos de consecuencias variantes del Proyecto de Ontología de Secuencia.

g:Profiler es desarrollado y mantenido en Estonia, en la Universidad de Tartu, Instituto de Informática, Grupo de Bioinformática, Algoritmos y Minería de Datos BIT. Actualmente, g:Profiler es desarrollado y mantenido por un equipo de desarrolladores de software profesionales, estadísticos e investigadores - Uku Raudvere, Ivan Kuzmin, Liis Kolberg, Priit Adler, Hedi Peterson y Jaak Vilo. Anteriormente, los principales contribuyentes han sido Jüri Reimand y Tambet Arak. A lo largo del tiempo, g:Profiler también ha recibido contribuciones valiosas de miembros del BIT, notablemente Jaanus Hansen, Raivo Kolde, Meelis Kull y Sulev Reisberg. La primera versión de g:Profiler se conocía como GOSt (Gene Ontology Statistics) y se hizo disponible a principios de 2005. g:Profiler es una herramienta de ELIXIR y su desarrollo es apoyado a través de proyectos de la Fundación del Desarrollo Científico Europea "Estonian Life Science Infrastructure for Biological Information". [g:Profiler](https://biit.cs.ut.ee/gprofiler/page/docs)

Caso de uso

Se planteó analizar de una lista de genes que términos GO asociados a proceso biológico (BP) tiene dicha lista, y también teniendo en cuenta cada gen en particular, mediante dos alternativas, similar al caso anterior, mediante requests y con una librería de Python específica de g:Profiler.

gprofiler\_grupal\_lib: Se analizó de manera grupal a la lista de genes, utilizando la librería de g:Profiler y se mostraron los nombres de los términos GO y su respectiva significancia en forma de p-value.

gprofiler\_grupal\_req: Similar al anterior pero utilizando un pedido POST directamente a la API de g:Profiler con la librería requests.

gprofiler\_individual\_lib: En este caso se utilizó también la librería, pero se realizó un análisis para cada gen particular de la lista y se mostraron los resultados.

gprofiler\_individual\_req: Finalmente se realizó un análisis individual de cada gen de la lista utilizando pedidos POST a la API.