



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DE LA CIENCIOMETRÍA EN BIOCIENCIAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICAS APLICADAS

PRESENTA:

MINERVA MARÍA ROMERO PÉREZ

DIRECTORA DE TESIS:

DRA. LAYLA MICHÁN AGUIRRE

Ciudad Universitaria, Ciudad de México 2024



Índice general

1	Introducción	1
§1.1	Lenguaje natural y lenguaje máquina	1
§1.2	Los metadatos	1
§1.3	Las bases de datos bibliográficas	2
§1.4	PubMed y Europe PMC	3
§1.5	Las APIs	4
§1.6	Revisiones de literatura	5
§1.7	La cienciometría	5
2	Desarrollo experimental	7
3	Resultados	8
4	Conclusiones	9

1 Introducción

Las biociencias son ...

Los análisis métricos de las publicaciones científicas ha crecido junto con la tecnología informática e informacional en la Web...

Entre métodos más usados para realizar los análisis están: la minería de textos, la ciencia de datos, los análisis de redes, los análisis métricos y los grandes datos...

instituciones e investigadores que utilicen la cienciometría...

Actualmente, existen instituciones como el Centre for Science and Technology Studies (CWTS) traducido al español como Centro de Estudios de Ciencia y Tecnología, en el cual con la bibliometría y la cienciometría estudian la investigación científica (<https://www.cwts.nl/about-cwts>) ...

Silvio Peroni...

Ludo Wltman...

Daniel Torres Salinas ...

La ciencia abierta es un ...

1.1. Lenguaje natural y lenguaje máquina

El lenguaje natural es la forma en la que las personas hablamos y vemos la información, esta es la que entendemos y es como visualizamos comúnmente las cosas en la web. Mientras que el lenguaje máquina es estructurado, interoperable y no es comprensible a simple vista, es leído por computadoras y es como funcionan los programas y la web.

1.2. Los metadatos

Los metadatos son comúnmente conocidos como los datos sobre los datos. Estos constituyen Información estructurada que describe, explica, localiza y hace que sea más fácil de recuperar, utilizar o manejar la información (NISO,XX). Estos pueden ser estructurales, descriptivos o administrativos. Los estructurales son ...

El uso de metadatos adecuados permiten una correcta estructuración y descripción de los datos, puesto que detallan las características de la información que se genera, como la fecha de creación, la licencia de uso, el autor, el tipo de objeto digital (texto, imagen, sonido, video) o la dirección electrónica en la que se encuentra alojado.

Los metadatos bibliográficos son útiles para el fácil procesamiento de la información, ayudan al estudio de las publicaciones científicas, con ellos se pueden generar análisis para ver patrones y responder preguntas científicas sobre las investigaciones. Este tipo de metadatos se originan a partir de la información proporcionada por los autores, revistas y bases de datos bibliográficas.

Cada revista científica tiene una fecha de publicación de sus números, lo que genera los metadatos de fecha de publicación, las bases de datos generan los metadatos de cuando reciben y agregan las publicaciones, entonces se van generando distintos metadatos con fechas, para realizar una revisión de literatura y tomar el conjunto de publicaciones se tienen que tener en consideración estas, ya que hay que tener muy bien definido el intervalo de tiempo del que se van a obtener los textos.

La diferencia entre la lectura natural y la lectura máquina de los metadatos se muestra en la figura 1, en la cual se puede ver que el lenguaje natural es como encontramos comúnmente la información de la Web, mientras que el lenguaje máquina muy pocos lo conocen, pero es la forma en la que se maneja la información para su análisis, manejo y procesamiento.

Formato RIS son las siglas de la empresa que lo desarrollo Research Information Systems, en español Sistemas de información de investigación, es un formato de etiquetado en el cual se separan los metadatos, se agrega la etiqueta espacio, guion, espacio y el metadato asociado a la etiqueta. Se utiliza para el intercambio de información de citas. Por ejemplo, los gestores bibliográficos aceptan este formato para manejar las citas, y las bases de datos bibliográficas manejan este para tener los metadatos y poder exportarlos.

1.3. Las bases de datos bibliográficas

Las bases de datos son fuentes en las cuales permiten procesar, guardar, ordenar, mantener y presentar datos. Son sistematizadas, presentan la información en tablas con un orden establecido, por lo que se tiene mucha información en poco espacio. Las bases de datos permiten realizar búsquedas a partir de distintos criterios, procesar datos de forma cuantitativa y cualitativa, interrelacionar los resultados utilizando distintas variables y actualizar la información fácil y rápidamente.

Las bases de datos se extienden por módulos y se pueden diseñar, almacenar, manejar y analizar en una computadora, son compatibles con otras bases de datos, en estas se pueden hacer consultas ('queries') para realizar cálculos, las bases de datos relacionales son las más

utilizadas, presentan los datos en dos o más tablas (constituidas por registros o renglones y columnas) que están relacionadas entre sí por medio de la repetición de un atributo. Las bases de datos pueden ser estáticas en donde la información no varía en el tiempo o dinámicas las cuales se actualiza, cambia o elimina la información después de algún tiempo.

Las bases de datos biológicas son una herramienta indispensable para el quehacer de los biólogos, estas permiten...

Las bases de datos bibliográficas se encargan de manejar información estructurada, sobre los artículos, libros, capítulos o memorias. Las bases de datos bibliográficas manejan los metadatos dados por los autores y revistas, algunas agrega metadatos como identificadores, ontologías y tesauros.

Existen muchas bases de datos bibliográficas, desde especializadas en áreas como la medicina, que solo contienen artículos o solo libros y hay multidisciplinarias y que contienen cualquier tipo de publicación como artículos, preprints, libros, patentes, etc. Entonces es importante saber qué información contienen y que información necesitamos para elegir la o las más relevantes para trabajar. Entre las características más importantes a considerar están: criterios de inclusión de documentos, contenido (tipo de registros), tipo de acceso, la cobertura (representación) tipológica, temática, geográfica y temporal de la colección. Dentro de las bases de datos biológicas se encuentran las bases de datos bibliográficas sobre biociencia estas resultan útiles para generar nuevo conocimiento. Entonces algo fundamental en la Web son los datos ligados estos nos permiten tener la información conectada desde diferentes enfoques...

Indexar es agregar a un catálogo cualquier tipo de información, por ejemplo revistas, artículos, términos estructurados, etc. Entonces se tiene que las bases de datos bibliográficas indexan revistas de acuerdo a ciertos criterios que implementa la base de datos.

1.4. PubMed y Europe PMC

Una de las bases de datos bibliográficas más importantes para la medicina es PubMed, una de sus características más importantes es que es de acceso abierto, es amigable y tienen una gran cantidad de herramientas, estas características y más pueden consultarse en ... (artículo herramientas de pubmed texto ,XX).

Europe PMC esta es una base de datos bibliográfica que es socio de PubMed Central International y fue construido y desarrollado en colaboración con PubMed Central USA y PMC Canadá. El proyecto se lanzó en 2007 y actualmente es mantenido por EMBL-EBI. Se caracteriza por tener una gran cantidad de herramientas, es amigable, y además de tener las anotaciones SciLite ... (XX, XX)

¿Que se puede obtener de Europe PMC? Para este trabajo es de interés el tesoro MeSH

asociado a PubMed este es un vocabulario controlado ... Por otro lado, Europe PMC contiene las anotaciones SciLite que nos permiten conectar los artículos con ontologías y tesauros <https://www.ebi.ac.uk/training/online/courses/europepmc-quick-tour/searching-and-visualising-scilite-annotations/> ...

1.5. Las APIs

Una API o Interfaz de Programación de Aplicaciones (por sus siglas en inglés, Application Programming Interface), es un conjunto de reglas y protocolos que permiten que diferentes aplicaciones o sistemas informáticos se comuniquen entre sí.

Las APIs son esenciales para la integración y la interoperabilidad entre sistemas de software, y se utilizan en una variedad de contextos, incluyendo aplicaciones web, aplicaciones móviles, sistemas operativos, bases de datos, servicios en la nube y más. Algunos de los conceptos clave relacionados con las APIs son:

- Estándares y protocolos: para garantizar la consistencia y la compatibilidad. Interfaz de usuario vs. interfaz de programación: en donde la interfaz de usuario permite que las personas interactúen con la aplicación, mientras que la API o interfaz de programación permite que los programas y sistemas interactúen entre sí sin necesidad de intervención humana. Datos estructurados: las APIs suelen transmitir datos en un formato estructurado, esto puede ser en formato JSON (JavaScript Object Notation) o XML (eXtensible Markup Language) y facilitan la interpretación y el procesamiento de la información por parte de las aplicaciones. Seguridad: las APIs pueden requerir autenticación y autorización para garantizar que solo las aplicaciones autorizadas puedan acceder a sus servicios. Esto es fundamental para proteger la privacidad y la seguridad de los datos.
- Documentación: Las buenas APIs suelen estar acompañadas de documentación detallada que describe cómo utilizarlas, qué métodos están disponibles y qué datos se pueden esperar en las respuestas.

Las APIs son utilizadas en una amplia gama de aplicaciones, desde la integración de servicios de terceros en aplicaciones móviles hasta la automatización de tareas en sistemas empresariales. Permiten a los desarrolladores aprovechar las funcionalidades de otros sistemas de manera eficiente, lo que acelera el desarrollo de software y la creación de aplicaciones más poderosas y versátiles.

En este trabajo nos interesa explorar las APIs de bases de datos bibliográficas, ya que estas son las que nos van a permitir extraer metadatos de las publicaciones de una manera eficiente y rápida para poder realizar análisis de la información.

Adicionalmente, PubMed se tiene disponible la API (Application Programming Interface) que sea la cual define los métodos y datos que las aplicaciones pueden utilizar para interactuar con la base de datos unas con otras de manera estandarizada y predecible, para lo que permite para exportar registros y para procesarlos por métodos informáticos que implican programación y conocimiento en el funcionamiento de las aplicaciones. Además, hay una buena documentación disponible con los detalles técnicos, lo que promueve que incluso los usuarios puedan crear aplicaciones.

1.6. Revisiones de literatura

Una revisión de literatura es ...

Existen distintos tipos de revisiones de literatura, en capítulo biorevisiones de literatura, X se pueden ver algunas con sus ejemplos y..

1.7. La cienciometría

La cienciometría puede considerarse un tipo de revisión de literatura (capítulo biorevisiones de literatura,X)

Para realizar análisis cienciométricos existen aplicaciones web y software para computadora desde los cuales se puede procesar los datos, algunas de estas herramientas son amigables pero, algunas no ... (XX) La ciencimetría abierta es ..., las ventajas de esta son (XX) Leyes e indicadores cienciométricos Cuando se hace un análisis bibliométrico se utilizan las tres leyes de la bibliometría y los indicadores cienciométricos, las cuales ... Ley de Lotka La proporción de autores que han publicado n artículos dentro de un campo científico es inversamente proporcional al cuadrado de n . De lo cual obtenemos lo siguiente (Urbizagástegui Alvarado, 1999): $A(n)=C1n^2$ De donde: $A(n)$: es la función que indica la fracción de autores que han publicado n artículos $C1$: es el número de publicaciones por cada autor n : el número de publicaciones dentro de un campo científico Ley de branford Ley de Zipf Cienciometría para ciencias de la vida y de la salud La cienciometría es Análisis de redes Una red es una gráfica dirigida en la cual se le asocia una función a los arcos, nodos o ambos ... Objetivos Como los análisis han contribuido en nuevo conocimiento El objetivo de mi investigación fue la bibliografía en Biociencias para entender cómo se ha usado la cienciometría, y otros métodos métricos relacionados, como la bibliometría, la infometría y la almetría, para responder preguntas basadas en la literatura, mi intención fue obtener las aplicaciones cienciométricas en biociencias más interesantes, creativas y singulares, para implementarlas. Además, reuní las herramientas informáticas, los procesos y los protocolos informacionales e informáticos implementados en los análisis métricos para seleccionar los casos más relevantes y la tecnología de vanguardia, como, los grandes datos, el análisis de redes, la semántica, las ontologías, los datos ligados, algoritmos e inteligencia artificial. También sistematicé un conjunto normalizado de datos extraídos de las publicaciones para probar distintas teorías y modelos matemáticos, con la finalidad de generar nuevas hipóte-

sis con base en las variables definidas. Y por último, obtuve visualizaciones, tanto estáticas como dinámicas, que resulten en imágenes, tablas, gráficas, infografías, por ejemplo, que integren, representen y contextualicen mejor los resultados cienciométricos. Este proyecto es multi, intra, trans e interdisciplinario, es original, inédito e integrativo, porque generalmente los análisis bibliométricos estudian un tema científico específico, pero en este caso tengo como objetivo principal estudiar los artículos cienciométricos publicados en biociencias, esto es, haré ciencimetría de la ciencimetría en ciencias de la vida y de la salud para sistematizarla y sintetizarla con los siguientes propósitos: Entender las diferentes formas en las que se ha utilizado la ciencimetría en el área de ciencias biológicas y de covid. Reunir las herramientas digitales y los métodos informáticos más útiles e innovadoras que se han integrado a la ciencimetría en biociencias. Descubrir la mejor manera de integrar, representar y contextualizar la información por medio de la visualización. Reunir todos los análisis realizados para México. Herramientas amigables y de código (?)

2 Desarrollo experimental

De esta manera se tienen los desdoblamientos hiperfinos, los cuales se muestran en la

3 Resultados

De esta manera se tienen los desdoblamientos hiperfinos, los cuales se muestran en la

4 Conclusiones

De esta manera se tienen los desdoblamientos hiperfinos, los cuales se muestran en la