



SEP

SECRETARÍA DE
EDUCACIÓN PÚBLICA

SECRETARÍA DE EDUCACIÓN PÚBLICA



**ESCUELA NACIONAL DE
BIBLIOTECONOMÍA Y ARCHIVONOMÍA**

**APLICACIÓN DE LA BIBLIOMINERÍA
METODOLÓGICA EN LA ELABORACIÓN DE
UNA ONTOLOGÍA COMO SISTEMA DE
REPRESENTACIÓN DEL CONOCIMIENTO DE LA
ENFERMEDAD DEL TIFO EN MÉXICO, 1904-1977**

TESIS

**QUE PARA OBTENER EL TÍTULO DE:
LICENCIADO EN BIBLIOTECONOMÍA**

PRESENTA:

**MARTÍNEZ ACEVEDO KAREN VANESSA
POLO BAUTISTA LUIS ROBERTO**

**ASESORES: Dr. Francisco Collazo Reyes
Xochitl Flores Vargas**

CIUDAD DE MÉXICO

2021

Dedicatoria

Dedicado especialmente a nuestras familias, quienes cada día nos impulsaron a seguir adelante, por su apoyo incondicional y por aportar grandes cosas a nuestras vidas.

Agradecimientos

Agradecimientos especiales al Dr. Francisco Collazo Reyes, XOchitl Flores Vargas, a todo elequipo del Atlas Histórico de la Ciencia Mexicana y al Dr. Luis Manuel Vilches Blázquez quienes confiaron en nosotros y nos compartieron conocimiento que sirvió como base para la realización de este trabajo.

Agradecemos a la ENBA, a nuestros profesores, compañeros y amigos por formar parte de nuestra formación académica.

Luis Roberto Polo Bautista

Karen Vanessa Martínez Acevedo

Tabla de contenido

Prefacio	iv
Introducción	v
Capítulo 1. Antecedentes de la enfermedad del tifo a nivel internacional y nacional	1
1.1 Enfoque Internacional	1
1.1.1 Antecedentes	1
1.1.2 Desarrollo del tifo durante la Primera Guerra Mundial	3
1.1.3 Desarrollo del tifo durante la Segunda Guerra Mundial	6
1.2 Enfoque Nacional.....	10
1.2.1 Antecedentes	10
1.2.2 Tifo en la Nueva España (siglo XVI-XVIII)	11
1.2.3 Tifo en el siglo XIX	12
1.2.4 Tifo en el siglo XX	13
1.3 Agente transmisor	16
1.4 Etiología	18
1.5 Desarrollo de vacunas	21
1.6 Prevención	23
1.7 Erradicación	25
Referencias	27
Capítulo 2. Herramientas tecnológicas para análisis de textos y sistemas de organización del conocimiento	30
2.1 Descubrimiento de conocimiento en bases de datos	30
2.1.1 Definición del KDD	30
2.1.2 Etapas del KDD	31
2.1.2.1 Minería de datos	32
2.1.2.1.1 Definición de minería de datos	32
2.1.2.1.2 Algoritmos / técnicas de minería de datos.....	34
2.1.2.1.3 Aplicaciones de minería de datos.....	35
2.2 Modelos matemáticos de la bibliometría	37
2.2.1 Ley de la productividad científica de los autores	38

2.2.2 Ley de dispersión de la bibliografía científica	38
2.2.3 Ley de crecimiento exponencial de la información científica	39
2.2.4 Ley de Zipf	41
2.3 Bibliominería	41
2.3.1 Definición de bibliominería	42
2.3.2 Bibliominería metodológica	43
2.3.3 Aplicaciones de bibliominería en bibliotecas	44
2.4 Análisis de Redes Sociales (ARS)	45
2.5 Teoría de Grafos	46
2.6 Procesamiento de lenguaje natural (NLP)	49
2.6.1 Definición del NLP	49
2.6.2 Estructura del NLP	50
2.6.3 Aplicaciones del NLP	51
2.7 Reconocimiento de entidades nombradas	52
2.8 Extracción de relaciones	53
2.9 Frecuencia de Término-Frecuencia Inversa de Documento TF-IDF (Term Frequency- Inverse Document Frequency)	54
2.10 Softwares para el procesamiento de lenguaje natural y análisis bibliométrico	55
2.10.1 Software de análisis de contenido y minería de texto	56
2.10.2 Lenguajes de programación de alto nivel	57
2.10.2.1 Herramientas de Python para hacer Procesamiento de Lenguaje Natural	57
2.10.3 Bibliometrix (Biblioshiny)	58
2.11 Organización del conocimiento	60
2.12 Sistemas de organización del conocimiento	60
2.12.1 Ontologías	62
2.12.1.1 Tipología de ontologías	63
2.12.1.2 Estructura de las ontologías	65
2.12.1.3 Metodologías para crear ontologías	66
2.12.1.3.1 Metodología methontology	66
2.12.1.4 Lenguajes de representación de ontologías	68
2.12.1.5 Herramientas para el desarrollo de ontologías	70

2.12.1.5.1 Protégé.....	71
Referencias	72
Capítulo 3. Bibliominería y metodología Methontology para la creación de una ontología sobre el tifo en México 1904-1977	80
3.1 Aplicación de la bibliominería.....	80
3.1.1 Identificación del área de aplicación	80
3.1.2 Identificación de fuentes de datos internas y externas	81
3.1.3 Recolección y limpieza de los datos en un almacén de datos	81
3.1.4 Selección de herramientas de análisis apropiadas	87
3.1.5 Identificación de patrones a través de la minería de datos	88
3.1.6 Análisis de los resultados mediante la metodología metonhtology	90
3.2 Codificación en Protégé	97
Referencias	103
Capítulo 4. Sistema de representación del conocimiento de la enfermedad del tifo en México	104
4.1 Análisis bibliométrico del corpus textual.....	104
4.2 Resultados de la bibliominería	108
4.2.1 Objeto de estudio.....	110
4.2.2 Procedimientos	111
4.2.3 Instrumentos	112
4.2.4 Lugares.....	115
4.2.5 Instituciones.....	117
4.3 Resultados en Protégé.....	119
Conclusiones	127
Referencias	129
Anexos	129

Prefacio

Los factores que intervinieron en la elección del tema de este trabajo de titulación fueron específicamente dos, el primero está orientada al tema central de esta investigación, el cual es la enfermedad del tifo en México, y el segundo está dirigido al tratamiento y enfoque de la propia información, la cual es la creación de una ontología y las herramientas necesarias para desarrollarla.

Respecto al primer punto, fue un factor importante el haber realizado nuestro servicio social en el *Centro de Investigación y de Estudios Avanzados del IPN* (CINVESTAV), en un proyecto denominado *Atlas Histórico de la Ciencia Mexicana (AHCM)*, desarrollado en el área de bibliometría de la biblioteca de física, matemáticas y matemática educativa “Jerzy Plebanski”.

Este acercamiento se debió a una invitación por parte del Dr. Francisco Collazo Reyes profesor de bibliometría en la Escuela Nacional de Biblioteconomía y Archivonomía, cuando cursábamos el sexto semestre de la licenciatura.

Un tema de investigación dentro del área de bibliometría es la enfermedad del tifo, el cual nos pareció interesante para desarrollar un estudio de ese dominio del conocimiento, considerando un punto de vista bibliométrico.

Por otro lado, lo que influyó en el tratamiento y enfoque de esta investigación, fue un artículo titulado *Construcción de una red de ontologías sobre eventos meteorológicos a partir de periódicos históricos*, el cual uno de sus autores, concretamente el Dr. Luis Manuel Vilches-Blázquez, nos compartió información acerca de las herramientas de análisis de texto que sirvieron como base para la elaboración de la ontología sobre el tifo.

Introducción

La elaboración de una ontología en torno al tema del tifo en México, 1904-1977; surge con la necesidad de sistemas de representación del conocimiento que permitan organizar la información tomando en cuenta datos del contenido del documento y, por ende, permitan su recuperación considerando estos datos como puntos de acceso. Los actuales sistemas de organización de la información que se implementan dentro de las bibliotecas tradicionales son muy generales y están adecuados con la estructura de una publicación científica actual, tomando en cuenta los datos más comunes, como *título, autor (mención de responsabilidad), fecha de publicación, etc.*

Sin embargo, los sistemas de información históricos requieren de un tratamiento más especializado de la información y a falta de una estructura generalizada del texto científico, requieren considerar otros datos para conformar una genealogía de la formación del discurso científico, como es el caso del Atlas Histórico de la Ciencia Mexicana.

Esta investigación pretende dar otro enfoque respecto al desarrollo de la ciencia médica en México, mediante la representación del conocimiento sobre la enfermedad del tifo por medio de una ontología, tomando en cuenta categorías de análisis que reflejen la conformación genealógica del texto científico como *objetos de estudio, procedimientos, instrumentos, lugares e instituciones.*

Este trabajo de investigación se estructura de la siguiente manera:

En el primer capítulo se abordan los antecedentes de la enfermedad del tifo, tomando en cuenta la zona geográfica internacional (principalmente el continente europeo) y la zona geográfica nacional; así como algunos casos importantes del impacto epidemiológico de la enfermedad del tifo en México, la etiología, el desarrollo de vacunas, la prevención y finalmente su erradicación.

En el segundo capítulo se describen de manera teórica algunas herramientas para realizar análisis de textos, como la minería de datos, la bibliometría y la bibliominería

metodológica; así como el análisis de redes sociales, el procesamiento de lenguaje natural y los sistemas de organización del conocimiento.

En el tercer capítulo se describe la implementación metodológica de la bibliominería, a través de diferentes procesos como, la conformación del corpus textual, el preprocesamiento de datos (limpieza de texto), la utilización del procesamiento de lenguaje natural y la metodología methontology para el desarrollo de la ontología de la enfermedad epidemiológica del tifo en México.

En el cuarto capítulo se muestran los resultados obtenidos en el análisis bibliométrico del corpus de información, el procesamiento de lenguaje natural y aspectos específicos de la utilización y enfoque de la ontología como sistema de representación y organización del conocimiento sobre el tema del tifo en México. Finalmente se muestran las conclusiones y los anexos.

Capítulo 1. Antecedentes de la enfermedad del tifo a nivel internacional y nacional

En este capítulo se realizó una investigación documental acerca de los antecedentes de la enfermedad del tifo considerando la zona geográfica internacional, principalmente el continente europeo y la zona geográfica nacional, algunos casos importantes del impacto del tifo en México, la etiología, el desarrollo de vacunas, la prevención y finalmente su erradicación.

1.1 Enfoque Internacional

Para poder establecer un origen del surgimiento de la enfermedad del tifo a nivel mundial, es necesario destacar los primeros casos registrados de tifo en Europa, así como también el impacto que tuvo a nivel epidemiológico.

1.1.1 Antecedentes

Como una enfermedad del ser humano, el tifo epidémico ha existido durante siglos, aunque se ha especulado que ciertas epidemias antiguas probablemente eran tifo, los primeros relatos contemporáneos aparecieron cerca del final del siglo XV. De 1489 a 1490, durante las guerras civiles de Granada, los médicos españoles describieron una enfermedad similar al tifo que mató a 17,000 soldados españoles, seis veces el número de muertos en combate con los moros (Kiple, 1993).

A principio del siglo XVI, se originó una enfermedad similar en Italia. Durante el ataque francés de Nápoles en 1528, una aparente epidemia de tifo pudo haber alterado el curso posterior de la historia europea. Los franceses estaban en el punto decisivo de la victoria sobre las fuerzas de Carlos V, cuando la enfermedad se propagó y mató a 30,000 soldados franceses, lo que obligó a los sobrevivientes del ejército a retirarse (Kiple, 1993).

En 1546, Girolamo Fracastoro, quien había observado las epidemias en Italia, publicó la primera descripción clara de lo que denominó *fiebre lenticular, punteada o petequial*, también caracterizada por dolor de cabeza y malestar en general. En las regiones de los Balcanes, donde las tropas alemanas, italianas y francesas se

reunieron para combatir a los turcos, la enfermedad del tifo golpeó a muchos soldados, incluso antes de llegar al campo de batalla (Kiple, 1993).

Como la enfermedad se difundió en toda Europa por las fuerzas que regresaban de Hungría, al tifo también se le conoció como *morbus hungaricus* [...] El tifo aumentó dramáticamente a principios del siglo XIX. Entre 1816 y 1819, una gran epidemia atacó a 700,000 personas en Irlanda, cuya población era de solo 6 millones (Kiple, 1993).

Sin embargo, durante varias décadas, la confusión caracterizó la comprensión médica de la enfermedad. A finales del siglo XVIII, el médico Boissier de Sauvages había comenzado a usar el término tifo para describir los síntomas neurológicos de aquella enfermedad (Kiple, 1993).

Se realizaron pocas investigaciones para distinguir patológicamente entre el tifo y la fiebre tifoidea. Incluso en el siglo XX, la confusión entre la fiebre tifoidea y el tifo se caracterizaba en la nomenclatura. En muchos países europeos, la fiebre tifoidea se le conocía como *typhus abdominalis* y el tifo como *typhus exanthematicus* (Kiple, 1993).

Tiempo después, la investigación patológica y epidemiológica adicional que realizaron William Jenner y Austin Flint convenció a la mayoría de los médicos estadounidenses de que el tifo y la fiebre tifoidea eran entidades de enfermedades distintas. Durante una epidemia de tifo particularmente grave en la Alta Silesia, el médico y político alemán Rudolph Virchow publicó una evaluación radical de la epidemia, al observar que la enfermedad afectaba a los pobres, a los no educados y los inmundos (Kiple, 1993).

Un ataque de tifo confiere inmunidad prolongada. Los niños varones en regiones donde el tifo es frecuente pueden contraer casos leves que los protegen de una infección posterior más grave. Debido a que el organismo infeccioso persiste en los tejidos de sus víctimas incluso después de recuperarse, esta puede reaparecer años más tarde, especialmente en condiciones de estrés o cuando el sistema inmunitario de la víctima es bajo (Kiple, 1993).

Este fenómeno se observó, pero no se reconoció en 1898, cuando el médico de Nueva York, Nathan Brill describió una enfermedad frecuentemente diagnosticada como tifoidea, pero que tenía síntomas más estrechamente relacionados con el tifo. En 1910, Brill publicó un estudio de 221 casos y su minuciosidad condujo a la designación de la *enfermedad de Brill* como una trampa para síntomas desconocidos, similares al tifo (Kiple, 1993).

1.1.2 Desarrollo del tifo durante la Primera Guerra Mundial

La enfermedad del tifo era conocida por varios nombres, como: *fiebre de la cárcel*, *fiebre maligna*, *fiebre manchada*, etc., pero seguramente el nombre más adecuado, al menos durante algún tiempo, fue *fiebre de guerra*. La palabra alemana para tifus es Fleckfieber, que literalmente significa "fiebre manchada" (Allen, 2014).

El tifo desempeñó un papel importante en muchas campañas militares, entre ellas la invasión catastrófica de Rusia por parte de Napoleón. Su gran ejército marchó a Moscú con 500,000 hombres y regresó con solo 3,000. Es probable que el 20% de las víctimas fueran a causa del tifo. Los soldados de Napoleón no tenían idea de qué los golpeó, pero los científicos determinaron la primacía del piojo en el tifo justo a tiempo para la Primera Guerra Mundial (Allen, 2014).

El científico francés Charles Nicolle, director del Instituto Pasteur en Túnez, África del Norte (Kiple, 1993) fue el primero en probar la transmisión del tifo por piojos. En 1903, Nicolle notó que, en el hospital de la ciudad, los trabajadores de admisión a menudo enfermaban de tifo, pero las enfermeras de la sala no. Después de enterarse de que los pacientes que ingresaban al hospital fueron despiojados y lavados, Nicolle razonó que el agente que propagaba la enfermedad debía estar en su ropa (Allen, 2014).

Más tarde realizó algunos experimentos decisivos, utilizando piojos para transmitir la enfermedad humana al chimpancé, del chimpancé a un macaco, y del macaco a otro chimpancé. Su estudio publicado en 1909, fue enormemente influyente y por lo tanto, lo hizo ganador del Premio Nobel de Medicina en 1928 (Allen, 2014).

Durante la Primera Guerra Mundial, las fuerzas militares de ambos lados suavizaron el conflicto y actuaron a partir del descubrimiento de Nicolle, desarrollando procedimientos para la eliminación del tifo (Kiple, 1993). Cada país tenía médicos entrenados en temas relacionados al tifo, y muchos de ellos intentaron y posteriormente fallaron, en la creación y desarrollo de vacunas contra esta enfermedad. En 1914, los institutos médicos alemanes enviaron a los mejores científicos al campo para estudiar y combatir enfermedades infecciosas (Allen, 2014). Entre las poblaciones civiles pobres en la Europa del Este, donde no se aplicaron tales medidas preventivas, la enfermedad continuó teniendo una alta tasa de mortalidad (Kiple, 1993).

Los médicos militares alemanes se dieron cuenta por primera vez de la amenaza del tifo para sus tropas después de la Batalla de Tannenberg de 1914, en la cual, una fuerza alemana no tripulada destruyó dos ejércitos zaristas y capturó a 92,000 rusos. Durante ese proceso, un oficial ruso le menciona a S. Ansky un socorrista y escritor que, estaba luchando contra tres enemigos mortales: los alemanes, los piojos y sus propios piojos; el último enemigo es considerado el más peligroso (Allen, 2014).

El tifo siempre acechaba en partes de Rusia; viajó por donde quiera que iban los ejércitos reclutas del zar, en los piojos que se aferraban a su ropa sin lavar. Un médico alemán encontró 6,000 piojos en un solo prisionero ruso. Después de ese hecho, el tifo se diversificó rápidamente en los campos de prisioneros de guerra (Allen, 2014).

Los alemanes tomaron medidas extremas contra estos artrópodos. El ejército instaló estaciones de limpieza donde los soldados y prisioneros se desnudaron y esperaron durante horas en carpas con corrientes de aire o edificios de concreto mientras su ropa estaba expuesta al vapor de alta presión y baños de creosota diseñados para matar a los piojos (Allen, 2014).

Los alemanes desinfectaron a 3.5 millones de prisioneros durante la guerra, y mientras el sistema fue brutal, trajo resultados positivos. De las 33,000 muertes militares alemanas por enfermedades infecciosas durante la guerra, menos de

1,500 fueron causadas por el tifo. El control de la enfermedad en Alemania probablemente evitó que el tifo se extendiera al frente occidental (Allen, 2014).

Aunque no se informó de tifo en el Occidente, las tropas sí enfermaron de *fiebre de trinchera*, una enfermedad que era transmitida por piojos, que los alemanes llamaron *fiebre Wolhynian* y los polacos la denominaron *Quintana* (el agente causal finalmente se le denominó *Rickettsia Quintana*). Pocas veces mataba, pero era muy desagradable y extremadamente común, representando más de un tercio de las víctimas británicas; la fiebre surgía cada cinco días y, a veces, reaparecía años después. Cientos de miles de casos ocurrieron entre ejércitos aliados y alemanes en el frente occidental (Allen, 2014).

A finales de 1917, los bolcheviques declararon un conflicto armado con Alemania, y Lenin retiró a los ejércitos de Rusia del frente. Regresaron a casa hambrientos, harapientos y llenos de tifo; la enfermedad se sembró en la población justo cuando el país se sumió en una guerra civil que atrajo a los países cercanos y enemigos de Rusia. El régimen bolchevique carecía de los recursos para rastrearlo y tenía poco interés en revelar el inestable estado de su salud pública a un mundo hostil (Allen, 2014).

La guerra civil duró desde 1917 hasta 1921, fue un choque caótico de ejércitos cuyas tropas a menudo cambiaron de bando: los rojos bolcheviques, por ejemplo, lucharon contra los bolcheviques blancos, los nacionalistas ucranianos y los polacos liderados por Józef Piłsudski; los ejércitos blancos bajo varios generales lucharon contra comunistas, polacos, ucranianos y las fuerzas del anarquista Néstor Makhno. Los ejércitos devastadores, la poca higiene, las columnas de refugiados hambrientos y enfermos y los seguidores del campamento extendieron el tifo por toda la región (Allen, 2014).

Al final de la Primera Guerra Mundial, la peor epidemia de la historia se extendió por toda Rusia desde Siberia hasta Polonia, causando entre 30 y 40 millones de casos de enfermedades y matando a aproximadamente tres millones de personas. Fue en la antesala de esta gran catástrofe donde el Dr. Rudolf Weigl y su ayudante Ludwik Fleck fueron reconocidos como investigadores del tifo (Allen, 2014).

1.1.3 Desarrollo del tifo durante la Segunda Guerra Mundial

Durante las décadas de 1920 y 1930, la investigación empírica sobre la vacuna contra el tifo se vio obstaculizada porque los organismos no podían cultivarse en cantidades necesarias fuera de las células vivas. Al mostrar ingenio frente a esta problematización, los investigadores prepararon vacunas de intestinos de piojos infectados, de heces de piojos y de los pulmones de ratas infectados con tifo (Kiple, 1993).

La Segunda Guerra Mundial comenzó el 1 de septiembre de 1939, cuando Alemania invadió Polonia con el pretexto de un falso ataque fronterizo polaco. Durante ese conflicto, la ideología nazi había identificado el tifo como una enfermedad característica de los judíos. A la medicina nazi le causaba terror el tifo, por lo cual, se tomaron medidas drásticas aparentemente para combatirlo. Estas incluyeron el cierre de grupos judíos en ciudades como Varsovia, Cracovia y Lviv, asegurando que la enfermedad se propagaría, pero solo entre los judíos. Los médicos alemanes capturados se convencieron de que era mejor matar a los judíos que permitirles contaminar a otros (Allen, 2014).

Dos meses después del inicio de la guerra, Hermann Eyer, quien entonces tenía 33 años, llegó a Cracovia y comenzó a establecer una instalación de producción en masa para que la vacuna contra el tifo se produjera mediante el método Weigl. Para esto, ocupó el Instituto de Microbiología de la Universidad Jagiellonian y confiscó los departamentos de la facultad judía en la calle Mickiewicz para alojar a su personal alemán (Allen, 2014).

Tiempo después, Hermann Eyer comenzó a buscar biólogos polacos para trabajar en el laboratorio; la mayoría de ellos habían sido detenidos en la redada en el Jagiellonian. Robert Kudicke, un microbiólogo alemán de 63 años alertó a Eyer de uno de los hombres arrestados, Zdzisław Przybykiewicz, un profesor de 31 años en el Instituto de Microbiología, el cual, fue ubicado en el cuartel de la Wehrmacht y llevado a Eyer. Przybykiewicz pasó de ser un hombre condenado en una prisión de la Gestapo a ser subdirector del Instituto de Investigación de Virus y Tifo de las

fuerzas armadas de Wehrmacht, así mismo, Eyer lo asignó para dirigir la producción de vacunas en el laboratorio de Cracovia (Allen, 2014).

En los días que precedieron a la invasión polaca, a Eyer le habían comentado que las tropas de la Wehrmacht estarían ocupando Lwów, y que su misión era controlar el laboratorio Weigl y expandir su producción de vacunas para los soldados alemanes, sin embargo, en lugar de enviarlo a Cracovia, se vio obligado a luchar por su cuenta para instalar un laboratorio de vacunas contra el tifo (Allen, 2014).

Eyer sustrajo la vacuna de las clínicas del ejército polaco y la usó para inmunizar a 50 empleados alemanes. Quince de estos hombres, incluido Eyer, enfermaron de tifo durante los siguientes seis meses, pero el curso de la enfermedad fue leve. Ninguno de los empleados murió y solo unos pocos sufrieron algunos síntomas como erupciones o alucinaciones (Allen, 2014).

Con ese hecho, se logró demostrar la efectividad de la vacuna de Weigl. Eyer, en colaboración con Friedrich Weigl, hermano mayor de Rudolf y un destacado abogado polaco en Cracovia que lo ayudaba a encontrar empleados; logró poner en marcha la producción de vacunas en poco tiempo y logró entregar su primer lote de vacuna contra el tifo a la inspección de salud de la Wehrmacht en abril de 1940 (Allen, 2014).

Durante ese tiempo, había 50 personas alimentando piojos al menos una vez al día en el laboratorio. Varios eran miembros del personal y otros fueron contratados. Para poder seleccionar con cuidado a las personas que formaban parte de ese trabajo, un asociado describió que los alimentadores de piojos recomendados eran personas de sexo masculino, maduros, con temperamento tranquilo, de piel suave y sana, deberían tener gran paciencia y autocontrol para resistir el deseo de rascarse (Allen, 2014).

Adam Finkel, un hematólogo que trabajó en el laboratorio de Weigl, señaló en una publicación de 1933 que 100 piojos, típicamente succionan hasta 40 miligramos de sangre en 15 minutos; tuvieron que alimentarse dos veces al día durante ocho días antes de estar listos para cosechar. Algunos alimentadores de piojos distintos

alimentaron 30,000 piojos a la vez, lo que significa que la sangre de una persona que alimenta piojos durante un poco más de una semana podría proporcionar vacunas para aproximadamente 300 personas (Allen, 2014).

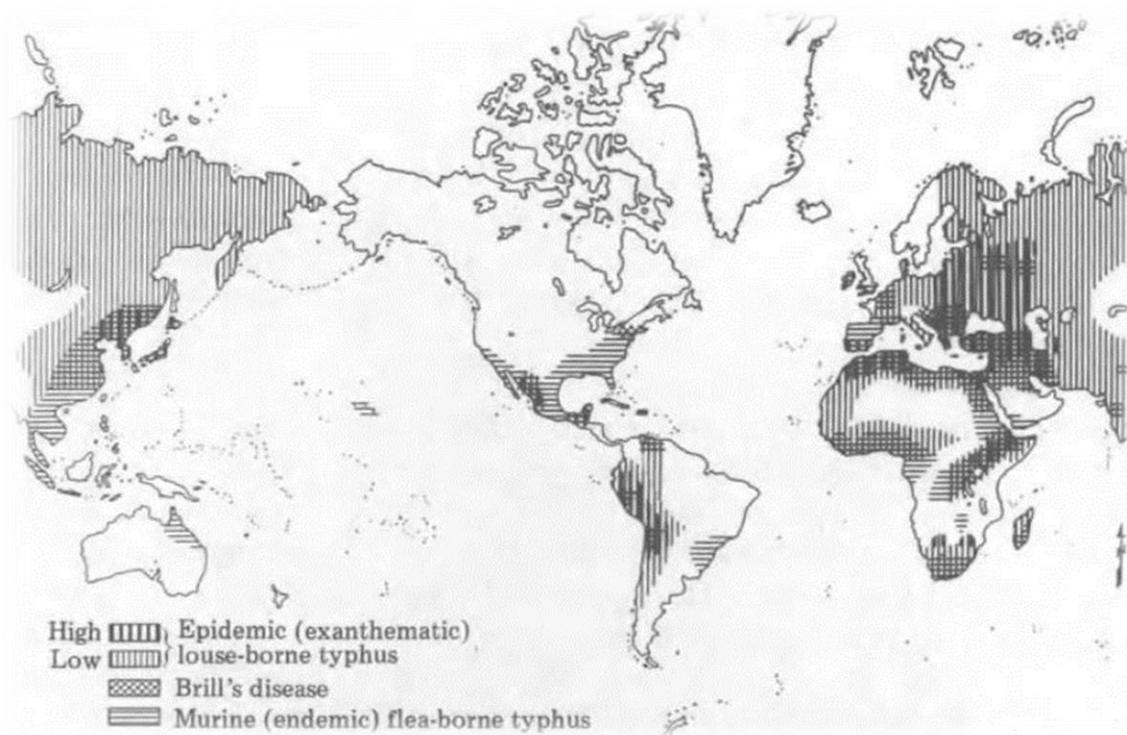
En 1937, el investigador del Servicio de Salud Pública de Estados Unidos, Herald R. Cox descubrió que las rickettsias crecían exorbitantemente en las yemas de los huevos de gallina fértil. Estos métodos simplificaron la producción de vacunas y la hicieron comercialmente factible [...] La amenaza del tifo fue un factor clave en los planes militares aliados durante la Segunda Guerra Mundial y en 1942 Franklin D. Roosevelt creó la Comisión de Tifo de Estados Unidos, con la finalidad de combatir el tifo donde sea que amenace los esfuerzos militares de Estados Unidos (Kiple, 1993).

La llamada vacuna Cox se suministró a todo el personal militar aliado, y aunque la experiencia durante la guerra indicó que en realidad no previno la enfermedad, claramente mejoró su curso. La investigación intensiva sobre agentes de seguridad por parte de grupos gubernamentales y privados demostró que los piojos podrían controlarse eficazmente con diclorodifenil-tricloroetano, más comúnmente llamado DDT (Kiple, 1993).

Este polvo podría aplicarse con una "máquina de soplado" para administrarlo debajo de la ropa sin que sus dueños tengan que quitársela. El método no solo fue más rápido, sino que fue aceptado incluso por los civiles más modernos. El DDT también demostró ser altamente efectivo en la epidemia de tifo que ocurrió durante el invierno de 1943 a 1944 en Nápoles, Italia (Kiple, 1993).

A finales de 1943, Ludwik Fleck se unió a un grupo de prisioneros que intentaban cultivar gérmenes de tifo en los pulmones de conejos vivos. Esta era una tarea de gran importancia, ya que, si pudieran cultivar los gérmenes, los cultivos se utilizarían para hacer una vacuna, lo que sería inmensamente valioso para el ejército alemán. Mientras el laboratorio contribuyera a la defensa inmunológica de sus soldados, el régimen nazi mantendría el laboratorio abierto y se abstendría de asesinar a su personal (Allen, 2014).

La profilaxis y el control del tifo con la vacuna Cox y con la utilización del DDT durante la Segunda Guerra Mundial redujeron el tifo de una amenaza mayor a una mera molestia entre las tropas americanas. Solo se produjeron 104 casos entre el personal militar de Estados Unidos, sin muertes. En cambio, se originaron epidemias graves entre civiles en el norte de África, Yugoslavia, los campos de concentración alemanes, Japón y Corea (Kiple, 1993). En el mapa 1 se muestra la diversificación del tifo epidémico, murino y la enfermedad de Brill durante la Segunda Guerra Mundial.



Mapa 1. Distribución geográfica del tifo en la Segunda Guerra Mundial

Fuente: Stanhope Bayne-Jones citado por (Kiple, 1993)

Años antes y durante la Segunda Guerra Mundial, la Alemania nazi llevó a cabo una serie de medidas para que la sociedad alemana tuviera un nulo contacto con las personas que percibían como amenazas biológicas a la salud de la nación. Los nazis obtuvieron ayuda de médicos, genetistas, psiquiatras y antropólogos para desarrollar políticas de salud radical. Estas políticas iniciaron con la esterilización en masa de muchas personas en hospitales y otras instituciones, y terminaron por

casi exterminar a los judíos de Europa (Museum United States Holocaust Memorial, 2018).

Algunos experimentos se llevaban a cabo dentro de los campos de concentración alemanes de Sachsenhausen, Dachau, Natzweiler, Buchenwald y Neuengamme, donde los científicos usaron a los internos de los campos para probar compuestos de inmunización y anticuerpos para la prevención y tratamiento de enfermedades contagiosas como la malaria, el tifo, la tuberculosis, la fiebre tifoidea, la fiebre amarilla y la hepatitis infecciosa (Museum United States Holocaust Memorial, 2018).

1.2 Enfoque Nacional

Como el objetivo principal de esta investigación está orientado a la enfermedad del tifo en México, a continuación, se abordarán de una forma más extendida el contexto social y científico que tuvo el tifo en la república mexicana, desde el origen de la enfermedad en el continente hasta su erradicación.

1.2.1 Antecedentes

Los antecedentes del tifo se pueden considerar a partir de dos enfoques, el primero señala que la enfermedad ya existía en el nuevo continente antes de que los conquistadores llegaran; algunos argumentos que apoyan esta afirmación son los siguientes:

La existencia del *matlazahuatl* o *tabardete* en nuestro territorio nacional data desde la protohistoria y sus terribles estragos se encuentran reflejados en las pinturas y relaciones toltecas. Esta enfermedad ocasionó la destrucción del reino de Tollán en el año 1116 y es de sospechar que de igual manera causó las emigraciones de Huehuetlapallan a México en el año 583 (León, 1982).

Elsa Malvido infiere la existencia de la enfermedad en América por la aparición del vector (piojo humano) en una momia y por el hecho de que en el libro sagrado de los mayas (Pop-Wuj) se hablaba sobre estos artrópodos (Guevara Flores, 2011).

El segundo enfoque sostiene que los orígenes del tifo se remontan hasta la época Colonial; donde la rata negra (*Mus rattus*), reservorio del tifo, pudo ser transportada

por barcos europeos a América, llegando a Cuba, para después pasar por las costas de Yucatán y México por los viajes de Hernán Cortés, donde se pudo diversificar entre la población por la presencia de piojos (Guevara Flores, 2011).

1.2.2 Tifo en la Nueva España (siglo XVI-XVIII)

Guerra Francisco (1988) menciona que la primera epidemia de tabardillo o tifo exantemático en México fue en 1526 en Veracruz, lugar denominado Medellín, donde las personas procedentes de Sevilla, entre ellos Luis Ponce de León, juez de residencia de Cortés, murieron a causa de esta enfermedad en un tiempo relativamente corto (Bertrán, 2006) citado por (Canales Guerrero, 2017).

Algunos aspectos importantes relacionados con la presencia del tifo en la Nueva España fueron los siguientes:

Según Fray Bernardino de Sahagún entre 1541 y 1545 hubo un brote epidémico, lo que causó grandes pérdidas humanas, algunos denominaban ese brote como *pestilencia o cocolixtle* (Castillo, 1956; Hurtado Ochotereta & Matías Juan, 2005).

En 1570 Pedro Ocharte imprimió en México el tratado de medicina más antiguo publicado en América, titulado *Opera Medicinales*, por el doctor Francisco Bravo, el cual se estructuraba en cuatro volúmenes. En uno de sus primeros capítulos el autor realizó un estudio sobre el *tabardillo* o *tabardete* (Castillo, 1956).

En 1576 una epidemia denominada *Matlazahuatl* o *tabardete* invadió gran parte de la Nueva España ocasionando la muerte de aproximadamente dos millones de personas. Durante esa epidemia se realizaron por primera vez en América investigaciones anatómo-patológicas en cadáveres humanos (Castillo, 1956).

Nicolás León menciona que en 1579 el doctor García de Farfán publica el *Tratado breve de Medicina* y debido a su aceptación fue reimpresso en los años 1592, 1604 y 1610. También publicó el *Tratado breve de Cirugía*, el cual describe en alguno de sus capítulos al tifo como *tavardete* (Castillo, 1956).

Fernández del Castillo señala que en los años 1588 y entre 1595 y 1596, hubo brotes de *Matlazahuatl* no tan graves como lo sucedido en 1576, pero de igual manera produjeron grandes estragos (Márquez Morfín, 1994).

En 1608 el doctor Juan de Barrios publicó el libro titulado *Verdadera medicina, cirugía y astrología*, el cual habla sobre el *tabardete*. En 1618 se publica el *Tratado de higiene y climatología* por el doctor Diego Cisneros, que describe al tifo como una enfermedad frecuente (Castillo, 1956).

Elsa Malvido menciona que entre 1641 y 1643 hubo una epidemia denominada *Matlazahuatl*, la cual sólo afectó al Estado de Puebla. Otros brotes epidémicos se registraron en 1667 y 1686 (Márquez Morfín, 1994).

Nuevamente en 1735 y 1736 la epidemia de *Matlazahuatl* ocasionó aproximadamente 40,150 muertos en el territorio considerado como el Arzobispado de México. Por motivo de esta epidemia se construyeron los hospitales provinciales de San Rafael, Santa Catarina de Mártir, de la Teja y Nuestra Señora de los Milagros (Castillo, 1956; Hurtado Ochotereta & Matías Juan, 2005).

Cooper describe que la última aparición de la epidemia *Matlazahuatl* del siglo XVIII ocurrió entre 1761 y 1762, simultáneamente con un brote de viruela (Márquez Morfín, 1994).

1.2.3 Tifo en el siglo XIX

En 1812 se registró una epidemia con características similares al tifo epidémico, la cual denominaron *epidemia de fiebres misteriosas*. Este padecimiento se propagó durante los conflictos bélicos de la guerra de independencia. En febrero de ese año, en Cuautla, la situación de los insurgentes se deterioró debido a la falta de alimentos y agua; estas circunstancias desencadenaron una fuerte epidemia (Molina del Villar, 2015).

Durante 1813, hubo un brote epidémico de tifo originado en Puebla, de ahí, se diversificó por los pueblos cercanos a la Ciudad de México (Márquez Morfín, 1994). En los primeros meses de ese año, el Diario de México comenzó a publicar noticias

respecto a la aparición de fiebres, así como otras noticias acerca de la salubridad (“Diario de México”, 1813) citado por (Márquez Morfín, 1994).

En 1846 después de numerosas necropsias efectuadas en enfermos del hospital de San Juan de Dios, el doctor Miguel Francisco Jiménez establece que el tifo de México (*tabardillo*), es diferente al *typhus* o dotienteria de Europa, llamada después tifoidea (Castillo, 1956).

Entre diciembre de 1874 y octubre de 1877 el tifo cobró un total de 3,458 vidas y seguramente alcanzó una morbilidad por encima de los 17,000 enfermos. La epidemia vivió su año más crítico en 1876, en el que se dieron 1,482 casos de muerte y tocó su punto máximo durante el mes de abril de 1877, en el cual fallecieron 227 personas (Herrera Rangel, 2011).

Como se menciona en el expediente 1009 “Sobre el desarrollo del tifo en la cárcel municipal” de 1891 del Archivo Histórico de la Ciudad de México, en ese año la enfermedad del tifo se propagó por la cárcel de Belem, la cual, causó grandes pérdidas humanas por su agresividad y recurrencia. Debido al alto índice de morbilidad que tenía la enfermedad, se tuvo que dar parte al presidente Porfirio Díaz para que, a través de la Secretaría de Gobernación, se implementarían medidas drásticas para detener el avance de la epidemia (Flores Flores, 2008).

Algunas de las causas por la cual se transmitía el tifo eran la carencia de todos los requisitos exigidos por la higiene, como aire, luz o que haya suficiente ventilación [...] las medidas que se tomaron para disminuir los casos de tifo consistieron en la separación de los reos en galeras, alejándolos del aire contaminado, bañar los edificios de la prisión con lechado de cal y regar los pisos con una solución de bicloruro de mercurio, además el establecimiento debería de asearse diariamente, y cualquier persona que presente fiebre sería remitida al hospital (Flores Flores, 2008).

1.2.4 Tifo en el siglo XX

En 1900 se contabilizaron aproximadamente 1,836 enfermos de tifo en algunos estados de México de los trescientos mil habitantes. Seis años más tarde, la

Secretaría de Instrucción Pública propone un acuerdo, en donde se concede una bonificación económica de \$20,000 a quien descubra el germen patógeno del tifo, otra bonificación de \$20,000 a quien descubra el medio de inmunizar ese germen, o el modo de trasmisión y una tercera bonificación de \$10,000 que se distribuirán entre todos los que realicen los trabajos más útiles para los descubrimientos relacionados (Castillo, 1956).

En 1907 el Instituto Patológico investiga el tifo exantemático; la sección a cargo de la investigación se encontraba dirigida por el doctor Genaro Escalona. Durante ese mismo año el doctor Miguel Otero experimentaba con sangre de personas infectadas de tifo para inyectarla a individuos de prueba, por tal motivo, adquirió experiencia en un caso de tifo típico (Castillo, 1956).

Durante 1909 y 1910 el tifo mexicano había sido objeto de interesantes estudios experimentales realizados por los doctores Gaviño, Girard, Goldtdern y Anderson en el Instituto Bacteriológico Nacional y clínicos por el doctor Genaro Escalona en el Hospital General. Los primeros logran la inoculación del tifo humano al mono (Castillo, 1956).

Howard Taylor Ricketts de las Universidades de Pensylvania y Chicago, vino a México a sustituir a Goldberger, el cual, se infectó de tifo y tuvo que abandonar el país. Ricketts estudió el tifo exantemático y su modo de propagación, por ese motivo se contagió durante sus investigaciones y murió en el Hospital Americano. El doctor Ricketts en colaboración con el doctor Russell M. Wilder escribieron algunos trabajos sobre el tifo, en los cuales realizaron importantes aportaciones (Castillo, 1956; Ruiz-Castañeda, 1972).

En 1910, en homenaje a la memoria del doctor Howard Taylor Ricketts se da el nombre *Howard T. Ricketts* al laboratorio del Instituto Bacteriológico Nacional donde el doctor realizó sus investigaciones (Castillo, 1956).

A partir de 1911 las condiciones sanitarias empeoraron en la capital del país. Este hecho provocó un incremento de enfermedades infecciosas, como el tifo, la viruela y la escarlatina. El organismo encargado de hacer frente a esta problematización

fue el Consejo Superior de Salubridad, el cual dependía de la Secretaría de Gobernación y estaba integrado por sólo seis miembros y cerca de 2,000 empleados, especialistas en diferentes disciplinas (Molina del Villar, 2015).

En 1914 estalló la guerra civil; el hambre, la movilización de tropas y la crisis en los servicios sanitarios propiciaron la aparición de diversas epidemias, como el tifo de 1915 y 1916, el número de personas infectadas en esos años, tan solo en el Hospital General, fue de 600. Durante esos años, el doctor Alfonso Pruneda es nombrado director de la campaña contra el tifo, empleando métodos modernos basados en las nociones recientes acerca de la transmisión de esa enfermedad por medio del piojo (Castillo, 1956; Molina del Villar, 2015).

Cuatro años más tarde, nuevamente una epidemia de tifo azotó la Ciudad de México. En 1919 se verifica en Toluca el Congreso de Tabardillo. En ese mismo año se establece La Comisión Central para el estudio del tabardillo en el Hospital General (Castillo, 1956).

En 1920 el doctor Tomas G. Perrín introduce en la práctica clínica la reacción del método Weil-Félix para el diagnóstico del *tabardillo*. Los cultivos de *Proteus X 19* fueron enviados a México por el doctor Rodolfo Cruz D., del Instituto Bacteriológico de Buenos Aires (Castillo, 1956).

En 1927 se descubre la existencia del tifo murino, cuyos agentes de propagación fueron la pulga y la rata. Un año más tarde, el doctor Hermann Mooser confirma en México la observación de Neill, describiendo que el tifo experimental del cual produce lesiones en la túnica vaginal con abundantes *rickettsias* (Castillo, 1956).

Esta peculiaridad del tifo mexicano no lo produce el tifo europeo, esto se justifica con las observaciones de una gran multitud de médicos mexicanos, quienes consideraban al tifo mexicano o *tabardillo*, como entidad nosológica independiente (Castillo, 1956).

A finales de febrero y principios de marzo de 1934, los doctores Miguel Bustamante, Gerardo Valera y Bosque Pichardo hicieron una visita a Totomihuacán, en la ciudad de Puebla de Zaragoza, en la cual, hubo una epidemia de tifo con 33 muertes en

los últimos dos meses. Esta situación fue favorable para realizar investigaciones acerca del valor preventivo del suero contra el tifo, que fue proporcionado por los doctores H. Zinsser y M. Ruiz Castañeda, de la universidad de Harvard (Bustamante, Varela, & Bosque Pichardo, 1934).

La investigación implicó la utilización del lote número dos de suero, que se preparó el 15 de septiembre de 1933, el cual, protegía al cuy contra el tifo mexicano. Posteriormente inyectaron el suero a 27 personas, cada una con 5 c.c si se trataba de un adulto y 3 c.c si era niño; las personas seleccionadas tenían que estar en contacto con tifosos y no debían de haber sufrido tifo con anterioridad. Los doctores seleccionaban a una o dos personas por choza para inmunizarlas y a las demás no se les suministró suero (Bustamante et al., 1934)

En una segunda visita a Totomihuacán (15 días después) se mostró que ninguna de las personas inmunizadas contrajo tifo, a diferencia de las que no se suministró el suero, ya que algunas si enfermaron. Un mes después, la epidemia había disminuido y ninguno de los inmunizados lograron contagiarse de tifo. Los doctores detallaron que al principio de su investigación tuvieron dificultades, ya que, los nativos no se prestaban para ser inmunizados; tiempo después y con la efectividad que se logró sucedió lo contrario (Bustamante et al., 1934). Durante 1932 y 1933, se realizaron varias investigaciones dirigidas por la Comisión del Departamento de Salubridad acerca de la inmunidad del suero contra el tifo, concretamente los casos estudiados corresponden a 23 enfermos del Hospital General y tres enfermos en el Manicomio General, en Mixcoac. Los resultados obtenidos de los estudios fueron favorables para las personas inmunizadas (Bustamante et al., 1934).

Durante esos años, Mooser, H., Ruiz Castañeda, M., & Zinsser, H. (1931) realizaron investigaciones sobre la transmisión del virus del tifo de rata en rata y el mecanismo por el cual la enfermedad se mantiene en esos organismos, en donde, las ratas que se encontraban en la prisión de Belem sirvieron para la realización el estudio.

1.3 Agente transmisor

Durante el siglo XIX en Zacatecas, se creía que el tifo era transmitido mediante el aire por emanaciones desprendidas de sustancias animales en descomposición y

una vez generada la enfermedad, su propagación era favorecida por el contacto con los enfermos, la falta de higiene, la ventilación insuficiente, y el contacto con objetos insalubres (Miño-Grijalva, 2019).

Aunque no se conocía como tal el vector que fungía como reservorio de la enfermedad del tifo, las creencias y prevenciones que se tomaban en aquella época eran adecuadas, ya que, el organismo infeccioso se encontraba diversificado entre contextos poco higiénicos, por medio de ratas y ratones.

El virus del tifo del nuevo mundo se encontraba en la pulga de la rata y era transmitido principalmente por medio de esta hacia el hombre. Algunos estudios indican que el virus típico del tifo que se propagó en México y Estados Unidos, así como en algunas partes de Europa, principalmente por la presencia de estos roedores (Zinsser, 1935).

El tifo es causado principalmente por dos tipos de bacterias, la *Rickettsia prowazekii* causa el tifo epidémico o exantemático; el principal reservorio natural de esta bacteria son los seres humanos y se transmiten por el vector piojo humano. La *Rickettsia typhi* causa el tifo murino o endémico; los principales reservorios de esta bacteria son los roedores *Mus rattus* (rata negra) y *Rattus norvegicus* (rata café), de igual manera los ratones, musarañas, perros, gatos y zarigüeyas, como vector las pulgas *Xenopsylla cheopis* y algunos ácaros y piojos (García-Acosta, Aguilar-García, & Aguilar-Arce, 2017; Peniche Lara, Dzul-Rosado, Zavala Velázquez, & Zavala-Castro, 2012; Petri, 2017; Sorvillo et al., 1993; Zinsser, 1935).

Cabe mencionar que la Nueva España era vulnerable a estas dos bacterias que causaban el tifo *murino* y el tifo *exantemático* o epidémico también conocido como tifo humano, el primero teniendo una letalidad del 2% y el segundo alcanzaba una letalidad del 70% en los adultos (Molina del Villar, Pardo Hernández, & Mora, s.f.; Vaughan, R. James McKay, & Waldo E. Nelson, 1990).

Sin embargo, las epidemias catastróficas en la Nueva España eran causadas por el tifo humano, ya que, el tifo *murino* no se contagia de hombre a hombre, siempre se requiere a la pulga de la rata como transmisor del microorganismo *Rickettsia typhi*

[...] el *tifo murino* dada su baja letalidad, no podía provocar una epidemia (Canales Guerrero, 2017).

1.4 Etiología

Desde los primeros años de vida colonial, el tifo exantemático ha sido una de las epidemias más renombradas y mortíferas de la Nueva España, por lo que ha sido de gran interés para los médicos el conocer e intentar erradicar esta enfermedad. En 1909 Nicolle presentó pruebas concluyentes sobre la transmisión interhumana y ofreció métodos para el estudio experimental del tifo (Ruiz-Castañeda, 1972). Estas pruebas sirvieron de base para controlar la infección en los ejércitos que combatieron durante la primera guerra mundial (Ruiz Castañeda, 1956).

Los estudios de Nicolle fueron confirmados por Gaviño, Girard, Anderson y Golberger. Durante sus investigaciones, Golberger se infectó de tifo y tuvo que abandonar el país siendo sustituido por el investigador Howard Taylor Ricketts, el cual había conseguido prestigio gracias a sus estudios sobre la fiebre de las Montañas Rocallosas (Ruiz-Castañeda, 1972). Los trabajos de Ricketts en colaboración con Wilder se enfocaron en la búsqueda del agente causal del tifo con resultados que describieron la presencia de numerosos corpúsculos bacteriformes en el intestino del piojo (Ruiz Castañeda, 1956).

Lamentablemente Ricketts también fue víctima del tifo dejando sus investigaciones en incertidumbre. Cinco años después; Da Rocha Lima, un investigador brasileño quien dirigía el Departamento de Patología del Instituto de Medicina Tropical de Hamburgo, confirmó que en el piojo infectado hay un parasitismo intracelular abundante y de posición definida (Ruiz-Castañeda, 1972; Ruiz Castañeda, 1956).

En 1916 Da Rocha Lima demostró que la diferencia entre los corpúsculos tifosos y los observados en piojos normales se base en el hecho de que en el tifo esos microorganismos ocupan el citoplasma de las celdillas intestinales del piojo (Ruiz-Castañeda, 1972). En ese mismo año, se demostró que otros ectoparásitos, además del piojo, pueden infectarse con rickettsias, confirmando la dualidad o pluralidad de transmisión del tifo (Ruiz Castañeda, 1956).

Da Rocha Lima honro la memoria de Ricketts dando su nombre al nuevo género microbiano y el de su colaborador Prowazek para designar la especie causante del tifo. Sin embargo, el ambiente científico era poco receptivo a esta rectificación, quedando la etiología del tifo en la incertidumbre durante los siguientes quince años (Ruiz Castañeda, 1956).

El escepticismo con que se recibió la comunicación de Ricketts siguió imperando a pesar de los trabajos realizados por Da Rocha Lima. Nicolle, Blanc y Conseil insistían en la naturaleza viral del tifo, por otra parte, causaron gran impacto los trabajos de Plotz y los de Weil y Félix. En América, sin embargo, después del fracaso de Plotz, Wolbach y algunos colaboradores aceptaron el papel etiológico de *Rickettsia* (Ruiz-Castañeda, 1972).

El doctor Hermann Mooser, vino a México invitado por el Hospital Americano para hacerse cargo del laboratorio clínico de ese mismo nosocomio. Mooser era graduado de la Universidad de Zürich y mostró notable capacidad en las diversas especialidades de laboratorio [...] en la segunda mitad de ese siglo, el tifo exantemático aún era frecuente en la ciudad de México, lo que estimuló su interés por investigar su causa y mecanismos epidemiológicos (Ruiz-Castañeda & Varela, 1971). Hermann Mooser acuñó el término *tifo americano* que por diversas circunstancias cambió la designación a *tifo mexicano*, ambas basadas en ciertas diferencias notadas en la evaluación del *tifo mexicano* en animales de laboratorio, con lo observado por Nicolle y la mayoría quien estudió el *tifo europeo* (Ruiz Castañeda, 1956).

En el Instituto de Higiene comenzó una nueva era para el estudio del tifo, en el, Mooser encontró en la lesión de la túnica vaginal del cobayo, inoculado con tifo mexicano, un sitio de predilección para la multiplicación de rickettsias; Monteiro nombró a ese germen *rickettsia mooseri*. Gracias al descubrimiento de Mooser fue posible estudiar las rickettsias haciendo investigaciones sobre sus características biológicas (sus reacciones), serológicas e inmunológicas (Varela, 1956).

En 1931, Zinsser visitó México y en colaboración con Mooser y Castañeda logró demostrar la existencia de un agente infeccioso en el cerebro de ratas capturadas

en una prisión de la ciudad, agente que resultó idéntico al que se había encontrado en casos humanos de *tabardillo*. Zinsser con ayuda de Mooser y Castañeda consiguió observar la evolución del tifo en pulgas y demostrar sus transmisiones de rata a rata, por medio del piojo de rata (Ruiz-Castañeda & Varela, 1971; Ruiz Castañeda, 1956).

Bustamante M. (1956) describe que en la segunda mitad del siglo XX las rickettsias estaban divididas en cinco grupos según sus manifestaciones clínicas, sus características serológicas e inmunológicas y sus aspectos epidemiológicos, estos grupos son los siguientes:

Tabla 1. Grupo de Rickettsias

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Tifo clásico, producido por <i>Rickettsia prowazeki</i> y el tifo murino producido por la <i>Rickettsia mooseri</i> .	Fiebre manchada, producido por <i>Rickettsia Rickettsi</i>	Fiebre tsutsugamuchi	Fiebre Q, producida por <i>Rickettsia Burnetti</i> .	Fiebre de las trincheras, causada por la <i>Rickettsia Quintana</i>

Fuente: Bustamante M. (1956)

Por otra parte, Hidalgo, Faccini-Martínez, & Valbuena (2013) describen que actualmente y con base en análisis filogenéticos, se proponen dos clasificaciones para el género *Rickettsia*, mismas que se muestran en la Figura 1.

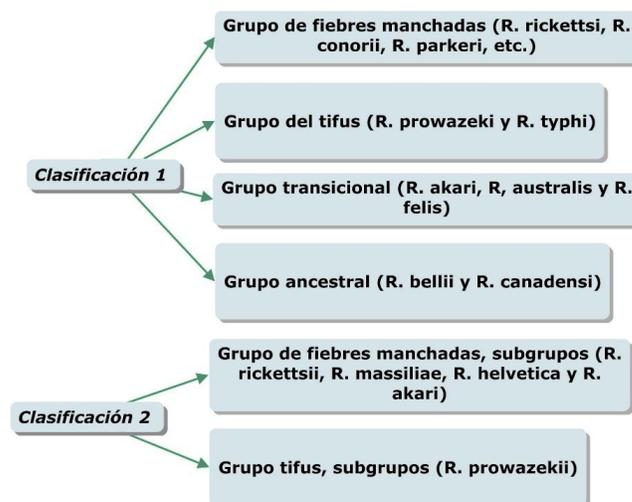


Figura 1. Clasificación de rickettsias

Fuente: Hidalgo, Faccini-Martínez, & Valbuena (2013)

1.5 Desarrollo de vacunas

Entre 1932 y 1934, en el Instituto de Higiene; Zinsser y Castañeda cultivaron cantidades relativamente considerables de rickettsias murinas en el peritoneo de ratas, y con ello lograron producir sueros superinmunes mediante los cuales demostraron la posibilidad de provocar inmunidad pasiva y preparar suero concentrado para su uso en la terapia del tifo (Bustamante, 1956; Ruiz Castañeda, 1956; Varela, 1956).

Durante 1934 y 1935, con lotes de suero proporcionados por Zinsser y Castañeda, Bustamante, Varela, Bosque Picharlo, entre otros; lograron resultados favorables en cuanto a inmunización pasiva en los Estados de Puebla y Oaxaca, así como también en enfermos del Hospital General (Bustamante, 1956; Varela, 1956).

Otra aplicación importante del cultivo de rickettsias, fue el desarrollado por Castañeda sobre métodos serológicos para investigar la presencia de precipitinas, opsoninas y de anticuerpos capaces de fijar en complementos con rickettsias; métodos perfeccionados más tarde por Bengston y Plotz, los cuales sirvieron para diferenciar el tipo antigénico del tifo (Ruiz Castañeda, 1956).

La comisión Mexicana del Tifo, integrada por los doctores Landa, Miranda, Cervera, Varela, entre otros; estudió cuidadosamente el suero ya no como profiláctico, sino como terapéutico de los tifos clásicos y murino y en 1936 dio a conocer los favorables resultados obtenidos (Bustamante, 1956).

Tiempo después, Sánchez Casco inmunizó un grupo de paralíticos con vacuna del tifo y después los inoculó con virus del tifo conservados en cobayos; se logró la protección de la mayoría de los vacunados y en los casos en los que se desarrolló el tifo, éste fue en forma benigna. Este experimento fue repetido por los alemanes durante la Segunda Guerra Mundial en un campo de concentración con resultados similares (Varela, 1956).

Durante los años que precedieron la Segunda Guerra Mundial se tuvo la creencia de que la vacuna murina sería suficiente para afrontar una situación de emergencia, ya que, diversos experimentos desarrollados por los doctores Félix y Veintemillas,

de Bolivia, habían aceptado la utilidad práctica de la vacuna disponible. Sin embargo, Zinsser recomendó que se realizarán más esfuerzos para producir una vacuna homologa que sería de mayor utilidad en caso de una guerra (Ruiz Castañeda, 1956).

Las dificultades técnicas que había que resolver eran enormes y de seguir utilizando la vacuna murina habría que centuplicar su rendimiento. Dos investigadores abordaron el problema simultáneamente. Cox, en Estados Unidos inoculó con éxito huevos embrionados que produjeron vacunas clásicas de poder suficiente para suministrar, si no inmunidad completa, suficiente protección para disminuir el peligro de muerte [...] por otro lado, Ruiz Castañeda, en México, desarrolló un método nuevo de cultivo de rickettsias mediante la inoculación masiva al pulmón de ratas y ratones. La vacuna murina llegó a tener un contenido de rickettsias cuyo límite era lo más que pudiera tolerar el hombre (Ruiz Castañeda, 1956).

De esa manera, para 1938 se había resuelto técnicamente el cultivo por alguno de los dos métodos antes descritos, sin embargo, el método de Castañeda dio resultados poco favorables con rickettsias clásicas, por lo que fue dejado de lado; hasta que Durand, Giroud, Sparrow y otros autores europeos, emplearon el método de Castañeda y lograron preparar vacunas de ratones, conejos y otros animales, con las que afrontaron con éxito la necesidad de vacunas suficientemente activas para proteger al hombre (Ruiz Castañeda, 1956).

En México se adoptó desde 1939 la vacuna clásica, preparada en pulmones de ratones, hasta 1942 cuando se logró obtener preparados de pureza aceptable para ser empleados en el hombre. Sin embargo, fue necesario incorporar esta vacuna murina en una fórmula bivalente de mayor poder experimental pero más tóxica, por lo que hubo que reducir las dosis hasta hacerlas tolerables al hombre.

Con este proceso, se ganó especificidad, pero se perdió potencia, como lo demostraron trabajos de aplicación realizados, durante algunos brotes epidémicos del tifo, por un grupo de epidemiólogos encabezados por el doctor Ortiz Mariotte (Ruiz Castañeda, 1956).

La vacuna bivalente fue perfeccionada aumentando su contenido en rickettsias clásicas hasta alcanzar un tercio del contenido total. La vacuna resultante, aplicada en dosis suficientes, sirvió para prevenir contra el riesgo de infección (Ruiz Castañeda, 1956).

Durante los estudios de las cepas murinas de tifo se contagiaron en el Instituto de Higiene los doctores Mooser, Castañeda, Lozano y León y los ayudantes Ramos, Pozo, Piltz, Quintanar, López Reyes, Matamoros y Serrano. Con esa misma cepa se contagió y posteriormente falleció el doctor Durán del Instituto Bacteriológico de Chile. En el Instituto Pasteur se contagió con la misma cepa el ayudante del doctor Plotz y en la Universidad de Harvard el doctor Zinsser y el doctor Velasco (Varela, 1956).

1.6 Prevención

El tifo era una enfermedad que se asociaba con la pobreza, desnutrición y la insalubridad, aunque el tifo afectó a todos los sectores sociales, el padecimiento estuvo asociado a los cuarteles, los navíos, prisiones, asilos, lazaretos o en lugares donde las personas vivían aglomeradas y con poca higiene (Molina del Villar, 2015). Es por ello por lo que se implementaron estrategias para disminuir los contextos insalubres en la ciudad.

En octubre de 1915, al terminar la guerra de facciones, las autoridades presentaron mayor interés en las emergencias sanitarias, es por ello que José María Rodríguez emprendió una campaña, que sería dirigida por el médico Alfonso Pruneda, empleando métodos actualizados y con base en conocimientos recientes acerca de la transmisión del tifo por medio del piojo (Rodríguez, 2016).

El 9 de diciembre de ese mismo año, la Secretaría de Gobernación expidió un decreto promulgado por el Poder Ejecutivo Federal en el diario *El constitucionalista* con las normas que debían aceptarse para luchar contra el tifo en la capital y sus alrededores. El documento describía que se prohibía la venta de bebidas alcohólicas, en especial el pulque; los centros públicos de reunión deberán clausurarse a las 11 p.m.; quedan prohibidos los bailes, kermeses, veladas y reuniones; se prohíben las reuniones llamadas velorios; se prohíbe que en las casas

haya palomas, gallinas, perros y cualquier clase de animales; se prohíbe el acceso a los lugares públicos a las personas de cualquier clase social que vistan ropa creada directamente de animales que puedan llevar parásitos que sean transmisibles (Castillo, F., 1960) citado por (Rodríguez, 2016).

De igual manera se declaró obligatorio el baño para todos los empleados del gobierno y para ello se implementaron baños gratuitos en: Niágara, Ferrocarril Central, Paraíso, Peralvillo, Victoria, Pajaritos, Plazuela de las Vizcaínas, Hidalgo, Santa María la Ribera, etc. De igual manera, para evitar la transmisión de tifo fue conveniente proporcionar gratuitamente baños y aseo de ropas a la clase proletaria. (Molina del Villar et al., s.f.).

En 1951 se registraron 57 brotes en los Estados de México, Puebla, Guanajuato, Tlaxcala, Oaxaca, Hidalgo, Querétaro, Chiapas, Aguascalientes, Coahuila, Durango y Guerrero. Generalmente cada uno de estos brotes ataca a varias localidades y se ignora el número exacto de víctimas. Por esta razón, en ese año se implementó nuevamente una campaña contra el tifo (Ortiz Mariotte, 1952).

La campaña inició por un acuerdo del Secretario de Salubridad y Asistencia, la cual, tenía como meta la erradicación del microorganismo infeccioso del tifo. Sin embargo, un problema sanitario de tipo económico social de esta magnitud requiere la cooperación de todos [...] la erradicación del piojo de la ropa en las comunidades indígenas no solo requiere agua y jabón, sino también el cambio de hábitos arraigados que, a pesar de esos elementos, lo siguen manteniendo (Ortiz Mariotte, 1952).

La campaña requería de enfermeras en las zonas urbanas y centros de higiene, así como también en zonas rurales; el principal material de la campaña lo constituía el DDT, que se aplicaba para evitar el desarrollo de cepas resistentes; de igual manera se incentivaba la compra de polvo para lavar ropa y jabón [...] a pesar de la poca epidemiología que tiene el piojo de la cabeza, se creyó que en una buena campaña contra la pediculosis el despiojamiento debe ser integral, por ese motivo se preparó una pomada para aplicar al pelo en la misma forma que se hace para un peinado

ordinario, esa vaselina contenía 2% insecticida, teniendo 100% de efectividad (Ortiz Mariotte, 1952).

En términos generales, se inició la distribución de DDT al 10%, en sobrecitos individuales fáciles de adquirir por su bajo costo; se elaboró jabón con DDT para lavado de ropa, habiéndolo perfeccionado después de muchas pruebas industriales, químicas y biológicas. El jabón, con ayuda del polvo, resolvió el problema del tifo y la desparasitación con piojos en la ropa. La vaselina con DDT es un auxiliar en la desparasitación de la cabeza (Ortiz Mariotte, 1952).

A partir de 1942, también existían métodos para exterminar el virus del tifo, como el suero para conferir inmunidad pasiva, la vacuna de Zinsser y Castañeda para prevenirlo, los antibióticos para tratar las formas clínicas y, como se mencionó anteriormente, la aplicación de DDT para destruir los piojos, cuyo uso fue iniciado por Mooser y empleado en gran escala durante la última guerra mundial (Varela, 1956).

1.7 Erradicación

La historia del tifo tuvo un desenlace con la aparición de insecticidas de extraña selectividad para los ectoparásitos que más daño han ocasionado al hombre y más tarde con el surgimiento de una nueva era en la quimioterapia de las infecciones. Estos hechos ocurrieron durante un momento crítico de la historia de las desgracias provocadas por el hombre civilizado, neutralizando para siempre la posibilidad de que el tifo vuelva a constituir el tremendo azote que aun en el siglo pasado era aliado de la guerra. Pero los grandes descubrimientos son a veces accidentes fortuitos que tal vez no hubieran ocurrido sin la conjugación de múltiples factores en un momento oportuno (Ruiz Castañeda, 1956).

Actualmente y por fortuna, el tifo ha sido relegado al grupo de enfermedades vencidas por el ingenio humano [...] el tifo pasó a la historia, pero su trágico recuerdo estará para siempre asociado al nombre de quienes dedicaron su esfuerzo a conquistarlo (Ruiz Castañeda, 1956). En términos generales, podemos mencionar que la enfermedad del tifo tuvo un fuerte impacto en la sociedad a nivel epidemiológico causando muchas muertes en diversas zonas de Europa y

posteriormente en el territorio nacional, las personas más vulnerables a esta enfermedad fueron aquellas que no tenían una higiene óptima y que vivían en zonas insalubres.

Para erradicar esta enfermedad se tuvieron que realizar diversas investigaciones médicas, para más tarde experimentarlas en diversas zonas y probar su efectividad. Es por lo que, el objetivo de esta investigación está orientado al análisis de aquella producción científica concretamente de México, con la finalidad de visibilizar aquellos objetos de estudio, instrumentos, procedimientos, lugares e instituciones, que fueron utilizados para esos fines. Por lo cual, se procederá a describir las herramientas tecnológicas que permitieron analizar esos textos científicos.

Referencias

- Allen, A. (2014). *The fantastic laboratory of DR. Weigl: how two brave scientists battled typhus and sabotaged the Nazis*. New York: W.W. Norton & Company.
- Bustamante, M. E. (1956). Aspectos epidemiológicos de las rickettsias en México. 1934 a 1954. *Gaceta médica de México*, (3).
- Bustamante, M. E., Varela, G., & Bosque Pichardo, G. (1934). El suero contra el tifo como inmunizante. *Boletín de la Oficina Sanitaria Panamericana*, 13(7), 604–607. Recuperado de <http://iris.paho.org/xmlui/bitstream/handle/123456789/10165/v13n7p604.pdf?sequence=1&isAllowed=y>
- Canales Guerrero, P. (2017). Historia natural del tifo epidémico: comprender la alta incidencia y rapidez en la transmisión de la *Rickettsia prowazekii*. En *Epidemias de matlazahuatl, tarbadillo y tifo en Nueva España y México* (Primera ed, p. 228). México: Universidad Autónoma de Coahuila: Escuela de Ciencias Sociales.
- Castillo, F. F. del. (1956). El tifus en México antes de Zinsser. *Gaceta médica de México*, 86(3), 10.
- Flores Flores, G. (2008). A la sombra penitenciaria: la cárcel de Belem de la ciudad de México, sus necesidades, prácticas y condiciones sanitarias, 1863-1900. *Revista Cultura y Religión*, 2(3), 69–73. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2785575>
- García-Acosta, J., Aguilar-García, C., & Aguilar-Arce, I. (2017). Tifus. *Medicina Interna de México*, 33(3), 351–362. <https://doi.org/0186-4866>
- Guerra, F. (1988). Origen de las epidemias en la conquista de América. *Quinto centenario*, 14, 43–51.
- Guevara Flores, S. E. (2011). *Estudio paleoepidemiológico aplicado a la aproximación de matlazahuatl como tifo exantemático*. Escuela Nacional de Antropología e Historia.
- Herrera Rangel, D. (2011). Las pintas de la sirvienta el tifo y el temor a los pobres en la Ciudad de México, 1874-1877. *Estudios de Historia Moderna Contemporánea de México*, 41, 53–77. <https://doi.org/10.22201/iih.24485004e.2011.41.26585>
- Hidalgo, M., Faccini-Martínez, Á. A., & Valbuena, G. (2013). Rickettsiosis transmitidas por garrapatas en las Américas: avances clínicos y epidemiológicos, y retos en el diagnóstico. *Revista Biomédica*, 33. <https://doi.org/http://dx.doi.org/10.7705/biomedica.v33i0.1466>
- Hurtado Ochotereta, C. A., & Matías Juan, N. A. (2005). Historia de la vacunación en México. *Revista Mexicana de Puericultura y Pediatría*, 13(74), 47–52.
- Kiple, K. F. (Ed.). (1993). *The Cambridge World History of Human Disease*.

<https://doi.org/10.2307/2517311>

- León, N. (1982). ¿Qué era el Matlazahuatl y que el Cocoliztli en los tiempos precolombinos y en la época hispana? En *Ensayos sobre la historia de las epidemias en México* (pp. 383–397). México: Instituto Mexicano del Seguro Social.
- Márquez Morfín, L. (1994). *La desigualdad ante la muerte en la Ciudad de México : el tifo y el cólera (1813 y 1833)*. México: Siglo XXI.
- Museum United States Holocaust Memorial. (2018). Experimentos médicos de los nazis. En *The Holocaust Encyclopedia*. Recuperado de <https://encyclopedia.ushmm.org/content/es/article/nazi-medical-experiments>
- Miño-Grijalva, M. (2019). Zacatecas: las epidemias de tifo y viruela a finales del siglo XIX. Una aproximación cuantitativa, 1892-1893. *Historia y Sociedad*, (37), 201–242. <https://doi.org/10.15446/hys>
- Molina del Villar, A. (2015). El tifo en la ciudad de México en tiempos de la revolución mexicana, 1913-1916. *Historia Mexicana*, 64(3), 1163. <https://doi.org/10.24201/hm.v64i3.3001>
- Molina del Villar, A., Pardo Hernández, C. P., & Mora, J. M. L. (s/f). *Las epidemias en la ciudad de México: una revisión de larga duración*. Recuperado de <http://www5.diputados.gob.mx/index.php/camara/Centros-de-Estudio/CESOP/Eventos-y-Presentaciones/Presentaciones.-III-Coloquio-internacional-Las-paradojas-de-la-megalopolis/MESA-REDONDA-1.-Volver-al-pasado-para-resignificar-futuros/Las-epidemias-en-la-ciuda>
- Mooser, H., Ruiz Castañeda, M., & Zinsser, H. (1931). The transmission of the virus of mexican typhus from rat to rat by polyplax spinulosus. *The Journal of experimental medicine*, 54(4), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Ortiz Mariotte, C. (1952). Campaña Nacional Contra el Tifo en México. *Boletín de la oficina sanitaria panamericana*, 236–244.
- Peniche Lara, G., Dzul-Rosado, K. R., Zavala Velázquez, J. E., & Zavala-Castro, J. (2012). Murine Typhus: Clinical and epidemiological aspects. *Colombia médica*, 43(2), 175–180. Recuperado de <http://www.ncbi.nlm.nih.gov/pubmed/24893060><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4001937>
- Petri, W. A. (2017). Tifus epidémico. Recuperado de <https://www.msmanuals.com/es/professional/enfermedades-infecciosas/rickettsias-y-microorganismos-relacionados/tifus-epidémico>
- Rodríguez, M. E. (2016). El tifo en la Ciudad de México en 1915. *Gaceta médica de México*, 152(2), 253–258.
- Ruiz-Castañeda, M. (1972). A la memoria de Howard Taylor Ricketts. *Gaceta Medica de México*, 104(4), 257–263.

- Ruiz-Castañeda, M., & Varela, G. (1971). Hermann Mooser: homenaje a su memoria. *Gaceta médica de México*, 102, 667–668. Recuperado de http://www.anmm.org.mx/bgmm/1864_2007/1971-102-6-667-668.pdf
- Ruiz Castañeda, M. (1956). Hans Zinsser en el XXX aniversario de la vacuna contra el tifo. *Gaceta médica de México*, 189–197.
- Sorvillo, F. J., Gondo, B., Emmons, R., Ryan, P., Barr, A. R., Waterman, S. H., ... Murray, R. A. (1993). A suburban focus of endemic typhus in Los Angeles County: association with seropositive domestic cats and opossums. *The American Journal of Tropical Medicine and Hygiene*, 48(2), 269–273. <https://doi.org/10.4269/ajtmh.1993.48.269>
- Varela, G. (1956). Contribución del Instituto de Higiene al estudio del tifo exantemático. *Gaceta médica de México*, (3).
- Vaughan, V. C., R. James McKay, & Waldo E. Nelson. (1990). *Tratado de pediatría*. México: Salvat.
- Zinsser, H. (1935). Much about rats a little about mice. En *Rats, lice, and history* (pp. 189–211). London: George Routledge and Sons.

Capítulo 2. Herramientas tecnológicas para análisis de textos y sistemas de organización del conocimiento

En este capítulo se describirán de manera teórica algunas herramientas para realizar análisis de textos, como la minería de datos, la bibliometría que solo está enfocada a analizar datos relacionados con la producción científica, la bibliominería, que es entendida como la combinación de la minería de datos y la bibliometría, así como el análisis de redes sociales, el procesamiento de lenguaje natural y los sistemas de organización del conocimiento.

2.1 Descubrimiento de conocimiento en bases de datos

La minería de datos no es una herramienta aislada, pertenece a un proceso mayor, al Descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases en adelante KDD). Antes de describir la minería de datos, se procederá a mencionar aspectos importantes del KDD, como su definición y las etapas del proceso.

2.1.1 Definición del KDD

Se trata de un análisis y modelado automático exploratorio de grandes repositorios de datos. El KDD es el proceso organizado de identificación de patrones válidos, novedosos, útiles y comprensibles a partir de grandes y complejos conjuntos de datos, el modelo se utiliza para comprender los fenómenos a partir de los datos, el análisis y la predicción (Maimon & Rokach, 2010).

Por otro lado, Fayyad, Grinstein, & Wierse (2002) consideran al KDD como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de los datos”. El KDD involucra una secuencia de etapas claramente definidas, cada una fundamental para la transición de los datos en conocimiento (Quiroz Gil, 2012).

Para Piatetsky-Shapiro & Frawley (1991) el KDD reúne investigaciones actuales para descubrir conocimientos útiles e interesantes en bases de datos. Abarca muchos enfoques diferentes para el descubrimiento, incluido el aprendizaje

inductivo, las estadísticas bayesianas, la optimización de la consulta semántica, la adquisición de conocimiento para sistemas expertos, la teoría de la información, etc.

Tomando como base las definiciones antes mencionadas, podemos definir al KDD como un proceso estructurado que engloba varias etapas que en conjunto permiten generar nuevo conocimiento a partir del análisis de grandes cantidades de datos. El núcleo del proceso del KDD es la minería de datos, la cual implica la ejecución de algoritmos que exploran los datos, desarrollan el modelo y descubren patrones previamente desconocidos (Maimon & Rokach, 2010).

2.1.2 Etapas del KDD

Las etapas del proceso del KDD son iterativas e interactivas, lo que significa que se puede regresar a la etapa anterior en cualquier circunstancia, de igual manera, en algunos casos se involucra al usuario en la toma de decisiones (Maimon & Rokach, 2010; Timaran Pereira, Hernandez Arteaga, Caicedo Zambrano, Hidalgo Troya, & Alvarado Perez, 2016). Para Timaran Pereira et al. (2016), define las etapas del proceso del KDD como:

1. *Seleccion*: Una vez identificado el conocimiento relevante y prioritario, y definidas las metas del proceso desde el punto de vista del usuario final, se crea un conjunto de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La seleccion de los datos vara de acuerdo con los objetivos del proyecto.

2. *Procesamiento*: En esta etapa se analiza la calidad de los datos, se aplican operaciones basicas como la remocion de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos, datos nulos, datos duplicados y tecnicas estadsticas para su reemplazo. Para esta etapa es de suma importancia la interaccion con el usuario o analista.

3. *Transformacion*: En esta etapa se mejora la calidad de los datos con transformaciones que involucran la reduccion de dimensiones y transformacion de atributos; este paso puede ser crucial para el xito de todo el proyecto KDD y generalmente es muy especfico (Landa, 2016; Maimon & Rokach, 2010).

4. *Minería de datos*: En esta etapa se decidirá el algoritmo del proceso de minería de datos, el cual puede ser clasificación, regresión o agrupamiento; esta decisión depende principalmente de los objetivos del KDD y de las etapas anteriores. Dentro de la minería de datos hay dos objetivos principales, la predicción (supervisada) y la descripción (no supervisada y visualización) (Maimon & Rokach, 2010).

5. *Interpretación*: En este punto se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores interacciones. También se incluye la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario. De igual manera se consolida el conocimiento descubierto para incorporarlo en otro sistema para facilitar la toma de decisiones (Timarán Pereira et al., 2016).

2.1.2.1 Minería de datos

Actualmente la minería de datos es una herramienta importante para poder analizar grandes cantidades de datos o información, esto es de suma importancia, ya que permite cuantificar mejores resultados en cuanto al descubrimiento de patrones dentro de los datos, permitiendo facilitar la toma de decisiones. Es por ello por lo que es un factor importante dentro de esta investigación, su implementación puede revelar aspectos desconocidos o poco conocidos en la dinámica de las estructuras de comunicación científica referentes a la producción científica del tifo en México.

Para tener un contexto general de lo que es la minería de datos, a continuación, se describe su definición considerando el enfoque de diversos autores, así como también sus algoritmos y las diversas aplicaciones que esta herramienta puede tener en la actualidad.

2.1.2.1.1 Definición de minería de datos

Anteriormente se describió a la minería de datos como una etapa en el proceso de descubrimiento de conocimiento (KDD). Sin embargo, en la industria, en los medios y en el contexto científico, el término minería de datos se usa a menudo para referirse a todo el proceso de descubrimiento de conocimiento, por lo cual, se podría definir como el proceso de descubrir patrones interesantes y el conocimiento de

grandes cantidades de datos; las fuentes de datos pueden incluir bases de datos, almacenes de datos (data warehouses), la web, otros repositorios de información o datos que se transmiten dinámicamente al sistema (Han, Kember, & Pei, 2012).

Otra definición es la propuesta por Beltrán Martínez (s/f), la cual considera a la minería de datos como el descubrimiento de relaciones, tendencias, desviaciones, comportamientos atípicos, patrones y trayectorias ocultas, con el propósito de soportar los procesos de toma de decisiones con mayor conocimiento; la minería de datos se puede ubicar en el nivel más alto de la evolución de los procesos tecnológicos de análisis de datos.

Por otro lado, Bharati & Ramageri (2010) la define como un proceso lógico que se utiliza para encontrar datos útiles a través de una gran cantidad de datos. El objetivo de esta técnica es encontrar patrones que antes eran desconocidos, para utilizarlos en la toma de decisiones dentro de un negocio. De igual manera, Bharati & Ramageri (2010) mencionan que la minería de datos involucra tres etapas importantes, las cuales son:

- *Exploración:* En este paso, los datos se limpian y se transforman en otra forma, también se determinan las variables importantes y luego la naturaleza de los datos en función del problema.
- *Identificación del patrón:* Este paso trata de identificar el patrón o algoritmo más apropiado para el proyecto.
- *Implementación:* En este punto se analizan los patrones para inferir los resultados.

Con base en lo anterior, podemos definir a la minería de datos como un proceso que involucra la utilización de algoritmos para la extracción de patrones, relaciones, anomalías, tendencias, etc., de grandes volúmenes de información o conjuntos de datos. Hay dos tipos principales de minería de datos: el primero está orientado a la verificación y el segundo está al descubrimiento. Los métodos de descubrimiento son aquellos que identifican automáticamente patrones en los datos, por medio de la predicción y la descripción (Maimon & Rokach, 2010).

Los métodos descriptivos están orientados a la interpretación de datos, que se enfoca en comprender la forma en que los datos se relacionan entre sí. Los métodos orientados a la predicción se enfocan principalmente en la utilización de algoritmos que permitan estructurar los datos de una forma específica para poder predecir comportamientos, tendencias, etc. (Maimon & Rokach, 2010).

2.1.2.1.2 Algoritmos / técnicas de minería de datos

La minería de datos implementa varios algoritmos o técnicas como clasificación, agrupación, regresión, redes neuronales, reglas de asociación, árboles de decisión, algoritmos genéticos, etc., para el descubrimiento de conocimiento a partir del análisis a bases de datos (Bharati & Ramageri, 2010). A continuación, se describen algunas técnicas o algoritmos que utiliza la minería de datos:

A) Clasificación: Es la técnica de minería de datos más utilizada y permite clasificar un conjunto de datos de acuerdo con etiquetas o patrones asignados dentro de los datos, mediante árboles de decisión, redes neuronales, redes bayesianas, etc. (Bharati & Ramageri, 2010; Unidad de Información y Análisis Financiero, 2014).

- *Redes neuronales:* Las redes neuronales también se denominan aprendizaje conexionista, ya que crea conexiones entre los datos [...] el algoritmo de red neuronal más popular es la retro-propagación (Gera & Goel, 2015).
“Las redes neuronales son mejores para identificar patrones o tendencias en los datos y son muy adecuadas para las necesidades de predicción o pronóstico” (Bharati & Ramageri, 2010).
- *Árboles de decisión:* Los árboles de decisión proveen de una herramienta de clasificación muy potente, que permite representar y analizar los datos para generar una predicción con base en el comportamiento del conjunto de datos (Bouza & Santiago, 2012).
- *Redes Bayesianas:* “Las redes bayesianas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento basados en teorías probabilísticas (teorema de Bayes) capaces de predecir el valor de variables no observadas y explicar las observadas” (Bonilla Gordillo, Ojeda Schuldt, & Echeverría Briones, 2006).

B) Agrupamiento (Clustering): Es un algoritmo que agrupa datos en categorías considerando sus similitudes y minimizando sus diferencias. Con el uso de esta técnica se pueden identificar zonas dispersas entre datos u objetos y patrones de distribución, así mismo las correlaciones entre los atributos de los datos (Bharati & Ramageri, 2010; Unidad de Información y Análisis Financiero, 2014).

C) Regresión: Para Sansgiry S.S, Bhosle M, Sail K., la técnica de regresión se basa en el aprendizaje supervisado y se utiliza para predecir un dato continuo o numérico, de igual manera el algoritmo puede predecir ventas, ganancias, tasas hipotecarias, pies cuadrados, temperaturas, etc. (Gera & Goel, 2015).

Tabla 2. Algoritmos de minería de datos

Minería de datos			
Verificación	Descubrimiento		
(Lenguaje Estructurado de Consultas) SQL	Descripción	Predicción	
Procesamiento Analítico en Línea (OLAP)	Visualización	<i>Clasificación</i>	<i>Regresión</i>
Análisis estadístico	Agrupamiento	Árboles de decisión	Árboles de regresión
	Reglas de asociación	AANNs	Árboles de modelos
	Descubrimiento de subgrupos (Subgroup Discovery)	Redes bayesianas	AANNs
	Redes de creencia	IBL	

Fuente: (Escobar Terán, Alcivar, & Puris, 2016)

2.1.2.1.3 Aplicaciones de minería de datos

La minería de datos es utilizada en varios sectores de la sociedad, algunas de estas organizaciones incluyen las tiendas minoristas, hospitales, bancos, compañías de seguros, etc. Varias de estas organizaciones están combinando la minería de datos con estadísticas, reconocimiento de patrones y otras herramientas importantes (Bharati & Ramageri, 2010).

La minería de datos se puede utilizar para buscar patrones, tendencias y conexiones que de otra forma serían complicadas de encontrar. La minería de datos es popular en muchas empresas porque les permite aprender más sobre sus clientes o usuarios y tomar decisiones inteligentes con base en esa información, también es de suma importancia para predecir o agrupar las preferencias de los clientes para realizar proyectos de marketing (Bharati & Ramageri, 2010).

El Grupo Logicalis (2014) describe algunos otros sectores donde se podría aplicar la minería de datos:

- *Análisis de datos financieros*: Se emplea tanto en el sector bancario como en el de las finanzas. Se busca proveer datos que aseguren la posibilidad de practicar análisis sistemáticos en condiciones avanzadas y con garantías de fiabilidad.
- *Industria de las telecomunicaciones*: En este sector, los datos son especialmente importantes para alcanzar una buena comprensión del negocio. La minería de datos y aplicaciones específicamente diseñadas para esta área, ayudan en la identificación de los patrones de telecomunicaciones, facilitando la detección de actividades fraudulentas y haciendo posible un mejor uso de los recursos, mejorando la calidad del servicio.
- *Análisis de datos biológicos*: El campo de la biología es uno de los más beneficiados por los avances de la tecnología. La minería de datos podría ayudar a la identificación de patrones de proteínas estructurales, análisis de redes genéticas o integración sistemática de bases de datos.

Para Mata Raúl (2017) la minería de datos también se podría aplicar a las siguientes actividades:

- *Medicina y farmacia*: Diagnóstico de enfermedades y la efectividad de los tratamientos.
- *Seguridad y detección de fraude*: Reconocimiento facial, identificaciones biométricas, accesos a redes no permitidas, etc.
- *Astronomía*: Identificación de nuevas estrellas y galaxias.

- *Geología, minería, agricultura y pesca*: Identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites.
- *Ciencias ambientales*: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales para mejorar su observación, gestión y/o control.
- *Ciencias sociales*: Estudio de los flujos de la opinión pública. Planificación de ciudades: identificar barrios con conflicto en función de valores sociodemográficos, etc.
- *Recuperación de información no numérica*: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.

Otro sector muy importante en donde se podría aplicar la minería de datos es la industria de la información, específicamente en las unidades de información, como las bibliotecas, museos, archivos, centros de documentación, etc. Principalmente enfocado al análisis de datos bibliográficos y datos generados por estas unidades, con la finalidad de evaluar, analizar y mejorar la toma de decisiones referente al crecimiento de los recursos de información, predicción del comportamiento de los usuarios, mejorar la calidad de los servicios, distribuir mejor los ingresos económicos, etc.

2.2 Modelos matemáticos de la bibliometría

La bibliometría tiene como base teórica los modelos matemáticos de la ley de Lotka (ley de la productividad científica de los autores), ley de Bradford (ley de dispersión de la bibliografía científica), Ley de Zipf, etc. Estos modelos se crearon con la finalidad de facilitar el análisis de datos generados dentro de las unidades de información, principalmente en las bibliotecas. Es importante describir aquellos modelos porque tienen una relación directa con la bibliominería, la cual se describe posteriormente.

2.2.1 Ley de la productividad científica de los autores

Como menciona Urbizagástegui Alvarado (2005), la ley de la productividad científica de los autores o Ley de Lotka, es una distribución de probabilidades discretas que describe la productividad científica de un autor. El modelo matemático fue propuesto originalmente por Lotka en 1926, como un modelo del cuadrado inverso, el cual se expresa en la siguiente fórmula:

$$Y_x = C x^{-n}$$

Donde,

Y_x es la probabilidad de que un autor haga x contribuciones sobre un asunto.

C y n son los dos parámetros que deben ser estimados de los datos observados.

Según Ardanuy Jordi (2009) en la actualidad, el modelo matemático se expresa de

la siguiente manera:
$$Y_x = \frac{C}{x^n}$$

En otras palabras, la Ley de Lotka es una técnica cuantitativa que analiza la cantidad de investigaciones o aportaciones científicas de un cierto autor en un área del conocimiento específica. Con la finalidad de medir los avances científicos y el crecimiento de la literatura en esa área en particular.

2.2.2 Ley de dispersión de la bibliografía científica

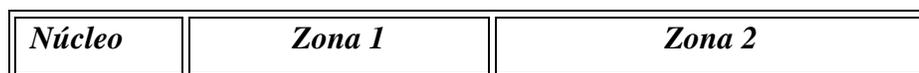
La ley de dispersión de la bibliografía científica o Ley de Bradford fue propuesta en 1934 por Samuel C. Bradford, un matemático, bibliotecario y documentalista británico después de estudiar una bibliografía de geofísica, cubriendo 326 revistas en el campo (Ardanuy, 2009; Jeremy Norman's, 2020; Paul E., 2004).

Ardanuy Jordi (2009) menciona que la ley de Bradford describe lo siguiente: si las revistas científicas se ordenan en una secuencia decreciente de productividad de artículos sobre una disciplina específica, estas pueden dividirse en un núcleo de revistas que comprenden el tema en particular y varios grupos o zonas que

contienen aproximadamente el mismo número de artículos que el núcleo, donde el número de revistas en el núcleo y en las zonas sucesivas se encuentran en la siguiente relación:

$$1 : n : n^2$$

Otra forma de representación de la relación anterior es la propuesta por Falcató Pedro (1989) quien la organiza de la siguiente manera:



Donde la primera zona posterior al núcleo tendrá n veces más revistas que él; la segunda zona tendrá n revistas más que el anterior, y por lo consiguiente $n \times n = n^2$ más revistas que el núcleo.

Con base en lo anterior, podemos mencionar que la ley de Bradford se encarga de analizar un corpus de información sobre un área determinada, con la finalidad de categorizar los documentos en tres grupos principales, donde la primer categoría o núcleo contiene la información más consultada sobre ese tema, la segunda categoría contiene más información que el primer núcleo y es consultada en menor proporción y la tercera categoría contiene más información que el anterior y es consultada en menos proporción.

La ley de Bradford se puede utilizar en situaciones donde se requiera un análisis de los documentos o información más consultados dentro de una disciplina específica.

2.2.3 Ley de crecimiento exponencial de la información científica

Según Centeno M.E., en 1976 Price, J.S notó que el crecimiento de la información científica crecía a un ritmo superior que al de otros procesos sociales, pero es muy similar a procesos que se observan en contextos naturales; el crecimiento es tal, que la información podría duplicarse, dependiendo en gran medida del área del conocimiento. Por lo que existe un crecimiento exponencial que alcanza un límite y una progresión geométrica llamada explosión de la información (Escorcía Otalora, 2008).

Escorcía Otalora (2008) describe el modelo matemático aplicado a la ley del crecimiento exponencial, la cual es la siguiente:

$$N = N_0 x e^{bxt}$$

Donde,

N es la magnitud de la medida relacionada con el tamaño de la ciencia.

N_0 es la magnitud de medida en el tiempo ($t = 0$).

t es el tiempo y

b es la constante que relaciona la velocidad del crecimiento con el tamaño ya adquirido de la ciencia.

Dentro de esta ley, Price propuso varias etapas, dos de ellas son las de crecimiento exponencial y crecimiento lineal. En la primera describe que la tasa de crecimiento es proporcional al tamaño de la muestra; y la segunda describe que la tasa de crecimiento es constante o independiente del tamaño del sistema (Escorcía Otalora, 2008).

De igual manera, Escorcía Otalora (2008) menciona las fases que contempla la Ley de Price, las cuales se muestran a continuación:

- Fase 1: Precursores.
- Fase 2: Crecimiento exponencial.
- Fase 3: Crecimiento línea.
- Fase 4: Colapso del campo científico.

En otras palabras, podemos describir que la Ley de Price calcula el crecimiento de la literatura científica dentro de una disciplina particular, con relación al paso del tiempo.

2.2.4 Ley de Zipf

La Ley de Zipf fue propuesta originalmente por el lingüista de la Universidad de Harvard George Kingsley Zipf a finales de la década de los 40. La Ley se basa en contar el número de veces que se usa cada palabra en un texto, para después realizar un ordenamiento de las palabras más frecuentes a las menos frecuentes por rangos (Carrasco, 2003; Urbizagástegui Alvarado & Restrepo Arango, 2011).

Según Carrasco (2003) la Ley de Zipf establece que el número de apariciones de una palabra es inversamente proporcional a su número de orden, el cual se expresa en la siguiente fórmula:

$$f(n) \simeq \frac{C}{n}$$

Donde,

$f(n)$ es la frecuencia normalizada de aparición de una palabra, es decir, el número de veces que aparece en el texto dividido por el número total de palabras (Montemurro, 2001).

C es una constante que se fija experimentalmente (Carrasco, 2003). Tomando como base lo anterior, podemos mencionar que la Ley de Zipf calcula la frecuencia de aparición de una palabra dentro de un texto, tomando en cuenta el número de veces que aparece y el total de palabras que hay en el texto.

Algunas aplicaciones de la Ley de Zipf pueden estar enfocadas a realizar análisis del discurso científico, mediante un conteo de palabras y redes semánticas, hacer análisis cualitativo de términos técnicos de una disciplina y como estos cambian con el paso del tiempo, en fin, las aplicaciones que esta ley podría tener son muy diversas y está relacionada con el objetivo de la investigación.

2.3 Bibliominería

El término de bibliominería es la combinación de la minería de datos y la bibliometría que está enfocada a analizar datos bibliográficos y datos generados dentro de las

unidades de información. A continuación, se describirá la definición de bibliominería, el proceso que implementa, y las diversas aplicaciones que puede tener dentro de las bibliotecas.

2.3.1 Definición de bibliominería

El uso de la minería de datos para examinar los registros de datos de una biblioteca podría denominarse bibliomining o bibliominería. Con la adopción de los catálogos en línea y los sistemas de recuperación de la información durante el último cuarto de siglo, los especialistas de bibliotecas e información regularmente han utilizado la bibliometría para explorar patrones en información bibliográfica (Nicholson & Stanton, 2003).

Durante el mismo período, varios investigadores han desarrollado y probado técnicas de minería de datos, tales como: métodos estadísticos y de visualización avanzados para localizar patrones no triviales en grandes conjuntos de datos. Por lo cual, la bibliominería se refiere al uso de estas técnicas para analizar las enormes cantidades de datos generados por una biblioteca (Nicholson & Stanton, 2003).

El término bibliominería fue utilizado por primera vez por Nicholson & Stanton (2003) al aplicar las técnicas de la minería de datos, para sondear las enormes cantidades de datos generados por bibliotecas. El término bibliominería es la unión de la bibliometría, la cual se encarga de descubrir patrones en la comunicación científica y el uso de técnicas de minería de datos (Nicholson, 2003).

Para Puente Marcelo (2010) la bibliominería puede describirse como “una disciplina que combina técnicas de la estadística, la bibliometría y de la minería de datos para la elaboración de informes que permitan extraer nueva información o conocimiento sobre los patrones de los sistemas bibliotecarios”.

Como menciona Banerjee, K, la conceptualización del término bibliominería es reciente, pero es una actividad que se ha realizado desde finales de la década de 1990 (Candás Romero, 2006).

Con base en las definiciones antes mencionadas, podemos denominar a la bibliominería como una disciplina que engloba técnicas de minería de datos y bibliometría enfocadas a analizar grandes conjuntos de datos generados por una unidad de información, con la finalidad de extraer nueva información y/o conocimiento que facilite la toma de decisiones.

2.3.2 Bibliominería metodológica

Para Nicholson (2003) la bibliominería metodológica consiste en las siguientes etapas:

1. *Determinar áreas de enfoque*: “El primer paso en la bibliominería metodológica es determinar el área de enfoque. Esta área puede provenir de un problema específico en la biblioteca o puede ser un área general que requiere exploración y toma de decisiones” (Nicholson, 2003). En otras palabras, se trata de especificar el área de aplicación de la bibliominería.

2. *Identificar fuentes de datos internas y externas*: En esta etapa se seleccionan las fuentes de información donde se extraerán los datos para el análisis. Según Nicholson (2003), los datos preferentes deben ser transaccionales, no agregados y de bajo nivel, con la finalidad de proteger los datos de los usuarios.

3. *Recolectar, limpiar y anonimizar los datos en un almacén de datos*: En este punto se recolectan todos los datos para el análisis y se procederá a realizar una limpieza o en su caso depuración de aquellos datos faltantes, nulos, etc., con el propósito de tener un corpus de datos uniforme. Cuando el corpus de datos esté depurado, se procederá a crear un almacén de datos, el cual según Nicholson (2003) “es una base de datos que está separada de los sistemas operativos y contiene una versión limpia y anónima de los datos operativos reformateados para su análisis”.

4. *Seleccionar herramientas de análisis apropiadas*: “Una vez que se ha desarrollado el almacén de datos, se prepara el escenario para el análisis” (Nicholson, 2003). Esto se refiere a seleccionar las herramientas necesarias para poder interpretar los datos generados, como informes tradicionales o softwares especializados.

5. *Descubrir patrones a través de la minería de datos y la creación de informes con herramientas analíticas tradicionales:* Este punto hace referencia a seleccionar aquellas técnicas pertenecientes a la minería de datos para poder analizar el corpus de información.

Como se mencionó anteriormente, “hay dos categorías principales de tareas de minería de datos, la descripción y la predicción. En la descripción, el objetivo es comprender los datos del pasado y el presente [...] Por otro lado, la predicción se utiliza para hacer una declaración sobre lo desconocido en función de lo que se sabe” (Nicholson, 2003).

6. *Analizar e implementar los resultados:* En esta etapa se tienen que probar los datos obtenidos del proceso creando un modelo con la finalidad de probar la solidez de los resultados (Nicholson, 2003).

2.3.3 Aplicaciones de bibliominería en bibliotecas

Para Candás Romero (2006) la bibliominería podría ser de apoyo en las siguientes actividades dentro de la biblioteca:

- Evaluación bibliotecaria.
- Apoyo en la gestión económica / selección y adquisición de recursos.
- Organización de la información.
- Sistemas de clasificación.
- Presentación de resultados y sistema de recomendaciones en el OPAC.
- Estudios de usuarios.

La bibliominería también se podría aplicar para analizar los patrones de error de un software para poder tomar decisiones informadas sobre las mejoras del sistema. De igual manera, se podría aplicar para predecir las necesidades futuras de un usuario; al buscar patrones en documentos de alto uso, los bibliotecarios pueden predecir mejor la demanda de nuevos documentos, para determinar cuántos ejemplares de un trabajo solicitar (Nicholson & Stanton, 2003).

2.4 Análisis de Redes Sociales (ARS)

El análisis de redes sociales ARS (Análisis de Redes Sociales), también conocido como análisis estructural, se ha desarrollado como herramienta de medición y análisis de las estructuras sociales que emergen de las relaciones entre diversos actores como *individuos, organizaciones, naciones, etc.* El ARS es un conjunto de técnicas de análisis aplicadas al estudio formal de las relaciones entre actores y para analizar las estructuras sociales que surgen de la recurrencia de esas relaciones o de la ocurrencia de determinados eventos (Sanz Menéndez, 2003).

El análisis de redes comienza prestando mayor interés al estudio de las conformaciones sociales, insistiendo en comprender las condiciones estructurales de las acciones de los actores. [...] El análisis de redes sociales generalmente estudia la conducta de los individuos a nivel micro, los patrones de relaciones (la estructura de la red) a nivel macro, y las interacciones entre los dos niveles (Sanz Menéndez, 2003).

Por otro lado, Pérez Salazar (2014) considera que el Análisis de Redes Sociales “proporciona un acercamiento metodológico que permite la identificación objetiva de las relaciones que ocurren en todo sistema social, así como la medición de un conjunto de parámetros, que describen tanto la importancia relativa de cada elemento al interior de la red, como su configuración general”.

Según Serrat Olivier (2017) el Análisis de Redes Sociales busca comprender las redes y sus participantes, teniendo dos enfoques principales: los actores y las relaciones entre ellos en un contexto social específico [...] Las redes sociales son nodos que representan individuos, grupos, organizaciones y sistemas relacionados que se unen en uno o más tipos de interdependencias: estos incluyen valores, visiones e ideas compartidas, contactos sociales, parentesco, conflicto, intercambios financieros, comercio, membresía conjunta en organizaciones y la participación grupal en eventos, etc.

Tomando como base las definiciones antes mencionadas, podemos considerar al ARS como un conjunto de herramientas de medición que permiten analizar las condiciones estructurales de un sistema social, que emergen de las relaciones entre

diversos actores sociales como personas, grupos, organizaciones, etc., en contextos determinados. Los actores sociales son representados mediante nodos y las conexiones e interacciones mediante aristas o conectores, y en su conjunto visualizan los diversos patrones de relaciones.

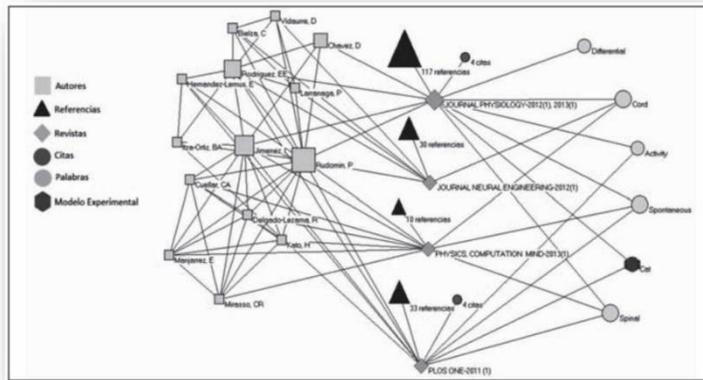


Figura 2. Ejemplo de una Red Social

Fuente: (Collazo Reyes, Luna Morales, & Luna Morales, 2016)

2.5 Teoría de Grafos

La teoría de grafos o teoría de las gráficas es una rama de las matemáticas discretas, que en los últimos años ha tenido un gran desarrollo. Se considera que tiene su inicio en 1736 cuando Leonhard Euler publicó un artículo donde aparece la solución al famoso problema de los puentes de Königsberg (González-Moreno, 2017). Combariza (2003) describe las siguientes definiciones como básicas de la teoría de grafos.

- **Grafo:** Un grafo G es una estructura matemática que consta de un par ordenado de conjuntos (V, E) , siendo $V \neq \emptyset$. Los elementos de V se llaman vértices y los elementos de E se llaman aristas.

Ejemplo: $G = (V, E)$ $V = (1, 2, 3, 4)$ $E = ((1, 2), (1, 3), (2, 3), (2, 4), (3, 4))$

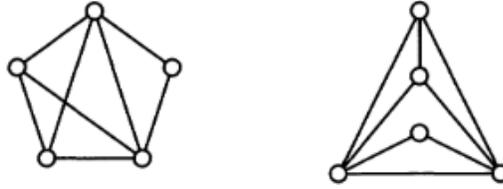


Figura 3. Dos grafos con 5 vértices

Fuente: (Godsil & Royle, 2001)

- **Multígrafo:** Un multígrafo $G = (V, E)$ está formado por un conjunto V de vértices, un conjunto E de aristas y una función f de E a $\{uv / u, v \in V, u \neq v\}$.
- **Pseudografo:** Un pseudografo $G = (V, E)$ está formado por un conjunto V de vértices, un conjunto E de aristas y una función f de E a $\{uv / u, v \in V\}$.

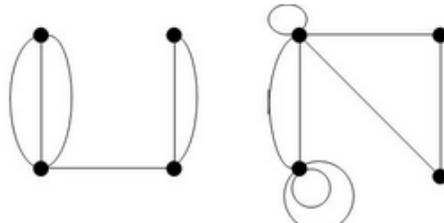


Figura 4. Multigrafo y pseudografo

Fuente: (Rapozo, s/f)

- **Grafo ponderado:** Un grafo ponderado $G = (V, E, w)$ está formado por un grafo $G = (V, E)$ y una aplicación $w: E \rightarrow R^+$

$$e = w(e)$$

que asocia un número, llamado *peso*, con cada arista del grafo.

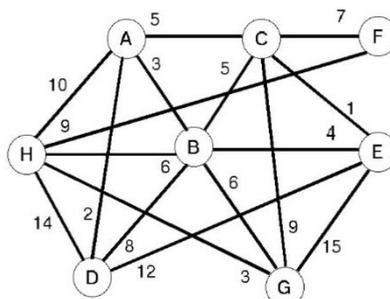


Figura 5. Grafo ponderado

Fuente: (Martínez, 2011)

- *Grafos dirigidos*: Un grafo dirigido o digrafo $D = (V, A)$ consta de un conjunto de vértices V y un conjunto de arcos A que son pares ordenados de elementos de V . En un grafo dirigido, los arcos están representados por flechas, las cuales indican la dirección de estos.

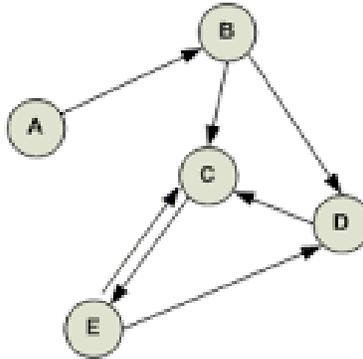


Figura 6. Grafo dirigido

Fuente: (Scla, 2016)

En términos generales, un grafo es una estructura matemática que permite modelar situaciones y contextos específicos, mediante una representación gráfica formada por nodos o vértices y aristas que representan los lazos o relaciones entre los nodos (Rochina, 2017).

Sanz Menéndez (2003) considera que la teoría de grafos es una parte importante del Análisis de Redes Sociales por las siguientes razones:

- Tiene vocabulario que puede ser utilizado para analizar muchas propiedades de las estructuras sociales.
- Ofrece operaciones matemáticas para medir y analizar esas propiedades.
- Permite probar teoremas sobre los grafos y, por lo tanto, deducir y someter a test determinados enunciados.

Con base en lo anterior, podemos mencionar que la teoría de grafos es parte de las matemáticas discretas y se encarga de estudiar las propiedades de conjuntos de datos, por lo cual es la base para analizar y modelar datos abstractos mediante nodos y vértices, así como las relaciones que puedan tener.

2. 6 Procesamiento de lenguaje natural (NLP)

El Procesamiento de Lenguaje Natural (Natural Language Processing en adelante NLP), es aquella área que estudia las interacciones entre la computadora y el lenguaje humano. Este término fue mencionado por primera vez en 1950, en el ensayo de Alan Turing, nombrado *Computing machinery and intelligence*, en el cual, se propuso que un humano evaluará conversaciones en lenguaje natural entre un humano y una máquina (computadora) diseñada para generar respuestas semejantes a las del humano.

Para dar un contexto general del NLP, a continuación, se describirán algunas definiciones, su estructura; aplicaciones como: minería de datos, traducción automática, recuperación de la información, extracción de información que tiene como tareas fundamentales el reconocimiento de entidades nombradas y extracción de relaciones; así como también algunos softwares que pueden utilizarse para el procesamiento de lenguaje natural y minería de texto.

2.6.1 Definición del NLP

El NLP es un área temática de la informática en la que el aprendizaje automático y la lingüística computacional son fundamentales. Esta herramienta computacional se ocupa principalmente de hacer que la interacción entre humanos y computadoras sea fácil y eficiente. La máquina aprende la sintaxis y el significado del lenguaje humano, lo procesa y entrega el resultado al usuario final. El NLP implica hacer que los sistemas informáticos realicen tareas significativas con el lenguaje natural y humano comprensible (Jain, Kulkarin, & Shah, 2018).

Para Bretonnel Cohen & Demner-Fushmanin (2014) El NLP es un tópico que combina la lingüística y la informática que se ocupa de las aplicaciones computacionales cuya entrada es el lenguaje natural. La investigación del procesamiento de lenguaje natural principalmente se ocupa de crear sistemas que procesan el lenguaje en un determinado nivel de restricción lingüística. Por ejemplo, se podrían crear sistemas de procesamiento de lenguaje natural para etiquetar las partes del discurso por medio de las palabras.

Por otro lado, Bird, Klein, & Loper (2009) consideran que el lenguaje natural es un lenguaje que se utiliza para la comunicación cotidiana entre los humanos, en idiomas como inglés, hindi o portugués [...] por lo cual, el procesamiento de lenguaje natural cubre, en un sentido amplio cualquier tipo de manipulación informática del lenguaje natural.

En un extremo, podría ser tan simple como contar frecuencias de palabras para comparar diferentes estilos de escritura. En el otro extremo, el NLP implica "la comprensión de las expresiones humanas completas, al menos hasta el punto de poder darles respuestas útiles (Bird et al., 2009).

Para Garbade M. (2018) el NLP es una disciplina de la inteligencia artificial que se ocupa de la interacción entre computadoras y humanos usando el lenguaje natural teniendo como finalidad leer, descifrar, comprender y dar sentido a los idiomas humanos de una manera que sea valiosa.

Por lo tanto, el NLP se puede considerar como una disciplina de la inteligencia artificial y la lingüística computacional que se ocupa principalmente del análisis y procesamiento de datos textuales, teniendo como objetivo facilitar la interacción entre humanos y máquinas por medio de la creación de sistemas inteligentes que procesan el lenguaje natural.

2.6.2 Estructura del NLP

Cortez Vásquez, Vega Huerta, & Pariona Quispe (2009) mencionan que la estructura o arquitectura de un sistema de procesamiento de lenguaje natural se sustenta en los siguientes niveles:

- *Nivel fonológico*: Se refiere a la relación de las palabras con los sonidos que representan.
- *Nivel morfológico*: Se enfoca a cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas.
- *Nivel sintáctico*: Trata de cómo pueden unirse las palabras para formar oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros.

- *Nivel semántico*: Se refiere al significado de las palabras y de cómo los significados se unen para darle sentido a una oración, también se refiere al significado independiente del contexto, dicho de otra forma, la oración aislada.
- *Nivel pragmático*: Trata de cómo las oraciones se usan en diversos contextos y como su uso impacta en el significado de las oraciones.

2.6.3 Aplicaciones del NLP

Las aplicaciones que puede tener el procesamiento de lenguaje natural son muy diversas, algunas están orientadas a establecer interacciones entre el humano y una computadora por medio del lenguaje, minería de textos, reconocimiento de voz, etc. A continuación, Hernández & Gómez (2013) describe algunas otras aplicaciones del NLP.

- *Minería de datos*: La minería de datos usa técnicas y metodologías de la recuperación de la información (RI), extracción de la información (EI) de corpus de información procesados con técnicas de lingüística computacional.
- *Traducción automática*: Es un área de la lingüística computacional dedicada a la investigación del uso de software para traducir un texto o habla en un lenguaje natural.
- *Recuperación de la información*: La recuperación de la información es definida como “el proceso de encontrar en un repositorio grande de datos, material (usualmente documentos) de naturaleza no estructurada (usualmente texto) o semiestructurada (páginas Web, por ejemplo) que satisfaga una necesidad de información”.
- *Extracción de la información*: La extracción de la información se enfoca en utilizar métodos y lenguajes de consulta a bases de datos como SQL para recuperar datos estructurados de cantidades de información no estructurada (Bird et al., 2009).

Jiang (2013) menciona que existen dos tareas fundamentales de la extracción de información, las cuales son el *reconocimiento de entidad* y *extracción de relación*.

El primero se refiere a la búsqueda de nombres de entidades como personas, organizaciones y ubicaciones; el último está enfocado a encontrar las relaciones semánticas [...] entre entidades.

2.7 Reconocimiento de entidades nombradas

El término *entidad nombrada* surgió en las Conferencias de Entendimiento de Mensajes (MUC) donde se buscaba promover y evaluar la investigación en el área de extracción de la información. En ellas, se definió entidad nombrada como “una palabra o secuencia de palabras que se identifican como nombre de personas, organizaciones, lugares, fechas, tiempo, porcentajes o cantidades” (Babych & Hartley, 2003).

Para Bird et al. (2009) las entidades nombradas son “frases nominales definidas que se refieren a tipos específicos de individuos, como organizaciones, personas, fechas, etc.”

El reconocimiento de entidades nombradas (Named Entity Recognition en adelante NER) “es la tarea de identificar tramos de texto que mencionan entidades con nombre y clasificarlas en categorías predefinidas, como persona, organización, ubicación o cualquier otra clase de interés” (Souza, Nogueira, & Lotufo, 2019). De acuerdo con Sánchez Pérez (2008) el NER es una tarea de la lingüística computacional, en la cual se busca agrupar cada palabra de un documento en una lista de categorías definidas como: persona, lugar, organización o fecha.

NE Type	Examples
ORGANIZATION	<i>Georgia-Pacific Corp., WHO</i>
PERSON	<i>Eddy Bonte, President Obama</i>
LOCATION	<i>Murray River, Mount Everest</i>
DATE	<i>June, 2008-06-29</i>
TIME	<i>two fifty a m, 1:30 p.m.</i>
MONEY	<i>175 million Canadian Dollars, GBP 10.40</i>
PERCENT	<i>twenty pct, 18.75 %</i>
FACILITY	<i>Washington Monument, Stonehenge</i>
GPE	<i>South East Asia, Midlothian</i>

Figura 7. Entidades Nombradas

Fuente: Bird et al.

Con base en lo anterior, podemos describir que el reconocimiento de entidades nombradas es una tarea de la lingüística computacional, que busca identificar y clasificar palabras o secuencias de palabras de un corpus de información en categorías predefinidas, como lugares, personas, organizaciones, fechas, porcentajes, etc. Sin embargo, el NER se puede aplicar para la identificación de palabras específicas con la finalidad de resolver un problema particular.

En el caso de esta investigación, el NER se aplicó para la identificación y extracción de entidades categorizadas en objetos de estudio, instrumentos, procedimientos, lugares e instituciones, todas ellas relacionadas al dominio de conocimiento de la enfermedad del tifo. A continuación, se muestra un ejemplo del reconocimiento de entidades nombradas, aplicado a un párrafo del primer capítulo.

El científico francés Charles Nicolle PERSON, director del Instituto Pasteur ORG en Túnez LOCATION, África del Norte LOCATION (Kiple PERSON, 1993 DATE) fue el primero en probar la transmisión del tifo por piojos. En 1903 DATE, Nicolle PERSON notó que, en el hospital ORG de la ciudad, los trabajadores de admisión a menudo enfermaban de tifo, pero las enfermeras de la sala no.

Figura 8. Ejemplo de Reconocimiento de Entidades Nombradas

Fuente: Elaborada por los autores

2.8 Extracción de relaciones

La extracción de relaciones está enfocada a extraer las relaciones que existen entre tipos específicos de entidad nombrada (Bird et al., 2009). Tolpygo Alexander (2016) define la extracción de relaciones como la vinculación de “dos entidades, como el propietario de una empresa, o la posición de la empresa de alguien y una persona en fuentes de texto no estructuradas”.

Por lo cual, podemos mencionar que la extracción de relaciones está orientada a la recuperación de las relaciones que existen entre tipos específicos de entidades nombradas.

2.9 Frecuencia de Término-Frecuencia Inversa de Documento TF-IDF (Term Frequency- Inverse Document Frequency)

La Frecuencia de Término por Frecuencia Inversa de Documento (Term Frequency-Inverse Document Frequency en adelante TF-IDF). Es un algoritmo desarrollado en la década de los 70 dentro del campo de Recuperación de la Información (Information Retrieval en adelante IR). En términos de recuperación de información bibliográfica tiene que ver con la pertinencia de encontrar el documento más relevante para cierto término en una colección de documentos (González Villa, 2019).

El algoritmo funciona midiendo la frecuencia de aparición de un término o frase dentro de un documento determinado, y lo compara con el número de documentos que mencionan ese término dentro de una colección completa de documentos (González Villa, 2019). La fórmula del TF-IDF se divide en dos partes, las cuales se describirán a continuación.

La TF es la suma de todas las veces que aparece un término en un documento, lo que permite determinar su capacidad de representación (Blázquez Ochando, 2012).

Blázquez Ochando (2012) menciona que para calcular la frecuencia de un término dentro de un documento determinado se efectúa la siguiente operación:

$$tf(n) = \sum_{(n)}^{D^1}$$

Donde, la frecuencia de aparición tf de un término n en un documento D^1 es la suma de las ocurrencias de dicho término.

El algoritmo IDF es inversamente proporcional al número de documentos en los que aparece dicho término. Lo que significa, que cuanto menor sea la cantidad de documentos, así como la frecuencia absoluta de aparición del término, mayor será su factor IDF y por el contrario, cuanto mayor sea la frecuencia absoluta relativa a una alta presencia en todos los documentos, menor será su factor discriminatorio (Blázquez Ochando, 2012).

El mismo autor nos describe la operación que se necesita efectuar para el cálculo del factor IDF.

$$IDF_{(n)} = \log_{10} \frac{N}{DF_{(n)}} + 1$$

Donde, el IDF de un término dado n se realiza aplicando el logaritmo en base 10 de N (Número total de documentos de la colección) dividido entre la frecuencia de documentos DF para un término n en la colección. Al valor resultante se le suma 1 para corregir los valores para los términos con IDF muy bajos (Blázquez Ochando, 2012).

La representación matemática final del cálculo TF-IDF es la siguiente:

$$TF - IDF_{(n,d)} = TF_{(n,d)} \times IDF_{(n)}$$

Donde, el peso de un término en un documento es el producto de su TF en dicho documento y su IDF (Blázquez Ochando, 2012).

Se consigue un TF-IDF alto, cuanto mayor es la frecuencia de un término en una página y menor es el número de documentos que lo mencionan. Conforme aumente el número de documentos que incluyan ese término, baja el valor TF-IDF, hasta el extremo de que puede llegar a 0, si la mayoría de los documentos de una colección lo mencionan (González Villa, 2019).

2.10 Softwares para el procesamiento de lenguaje natural y análisis bibliométrico

Para realizar procesamiento de lenguaje natural existen diversos softwares, tanto de código abierto como propietario o privados. El uso de estos softwares está determinado por el objetivo principal de esta investigación y la visión de los resultados esperados. A continuación, se describirán dos softwares que permiten la realización del procesamiento de lenguaje natural.

2.10.1 Software de análisis de contenido y minería de texto

Software de estadísticas de palabras

Wordstat es un software “de análisis de texto flexible y fácil de usar, ya sea que necesite herramientas de minería de texto para una extracción rápida de temas y tendencias, o una medición cuidadosa y precisa con herramientas de análisis de contenido cuantitativo de última generación” (Provalis Research, s/f).

WordStat 8 se utiliza principalmente para:

- Inteligencia empresarial y análisis de sitios web competitivos.
- Extracción de información y descubrimiento de conocimiento a partir de informes de incidentes, quejas de clientes, etc.
- Análisis de contenido de cobertura de noticias o literatura científica.
- Etiquetado automático y clasificación de documentos.
- Detección de fraude, atribución de autoría, análisis de patentes.
- Desarrollo y validación de taxonomías (Provalis Research, s/f).

Las características principales de WordStat 8 son:

- Permite la importación y exportación de diversos formatos de archivos, como pdf, csv, html, txt, etc.
- Permite analizar grandes cantidades de información.
- Cuenta con diversas herramientas de visualización y exploración de datos.
- Permite la clasificación de datos con base al contenido de la información.
- Cuenta con diversas herramientas de representación de jerarquías y enlaces de datos.
- Cuenta con diversos diccionarios en diversos idiomas.
- Tiene acceso a herramientas de recuperación de palabras clave en contexto.
- Permite generar mapas interactivos (Provalis Research, s/f).

2.10.2 Lenguajes de programación de alto nivel

Python

Python es un lenguaje de programación de alto nivel ampliamente utilizado para la programación de propósito general, fue desarrollado por Guido van Rossum y lanzado por primera vez en 1991.

Python presenta un sistema dinámico y administración automática de memoria, también admite múltiples paradigmas de programación, incluidos los orientados a objetos, imperativo, programación funcional y estilos de procedimiento (Stack Overflow, s/f).

Por otro lado, Rossum & Python development team (2018) consideran que Python es un lenguaje de programación poderoso y fácil de aprender; cuenta con estructuras de datos eficientes y de alto nivel y un enfoque simple y efectivo a la programación orientada a objetos. La sintaxis y el estilo dinámico de Python, lo convierte en un lenguaje ideal para la creación de scripts y el desarrollo de aplicaciones en diversas plataformas.

Para Bird et al. (2009) Python es “un lenguaje de programación simple pero potente con excelente funcionalidad para procesar datos lingüísticos”.

Con base en lo anterior, podemos definir a Python como un lenguaje de programación de alto nivel, enfocado a la programación de propósito general; su estructura es dinámica permitiéndole la administración automática de memoria; también está orientado a objetos, así mismo su sintaxis es simple y potente con funcionalidades de procesamientos de datos lingüísticos.

2.10.2.1 Herramientas de Python para hacer Procesamiento de Lenguaje

Natural

spaCy

“spaCy es una biblioteca de código abierto para el procesamiento avanzado del lenguaje natural (NLP) en Python” (spaCy, 2020a). SpaCy está diseñado específicamente para uso en producción, por lo cual es útil en la creación de

aplicaciones que procesen y comprendan grandes volúmenes de texto o información.

Se puede utilizar para crear sistemas de extracción de información o comprensión del lenguaje natural, de igual manera para procesar texto para el aprendizaje profundo (spaCy, 2020a). Algunas de las características más importantes de spaCy se muestran en la Figura 9.

NOMBRE	DESCRIPCIÓN
Tokenización	Segmentar el texto en palabras, signos de puntuación, etc.
Etiquetado de parte de voz (POS)	Asignar tipos de palabras a tokens, como verbo o sustantivo.
Análisis de dependencias	Asignar etiquetas de dependencia sintáctica, describiendo las relaciones entre tokens individuales, como sujeto u objeto.
Lematización	Asignar las formas base de las palabras. Por ejemplo, el lema de "era" es "ser" y el lema de "ratas" es "rata".
Detección de límites de oración (SBD)	Encontrar y segmentar frases individuales.
Reconocimiento de entidad nombrada (NER)	Etiquetado de objetos denominados del "mundo real", como personas, empresas o ubicaciones.
Entity Linking (EL)	Desambiguando entidades textuales a identificadores únicos en una base de conocimiento.
Semejanza	Comparar palabras, tramos de texto y documentos y qué tan similares son entre sí.
Clasificación de texto	Asignar categorías o etiquetas a un documento completo o partes de un documento.
Coincidencia basada en reglas	Encontrar secuencias de tokens basadas en sus textos y anotaciones lingüísticas, similares a las expresiones regulares.
Formación	Actualizar y mejorar las predicciones de un modelo estadístico.
Publicación por entregas	Guardar objetos en archivos o cadenas de bytes.

Figura 9. Características de spaCy

Fuente: (spaCy, 2020a).

2.10.3 Bibliometrix (Biblioshiny)

Bibliometrix es una herramienta de código abierto desarrollado por Massimo Aria y Corrado Cuccurullo enfocado para la investigación cuantitativa en cienciometría y

bibliometría que incluye todos los principales métodos de análisis bibliométricos [...] el paquete bibliometrix proporciona varios métodos de importación de datos bibliográficos de los índices multidisciplinarios de citas Scopus, Clarivate Analytics, Web of Science, Digital Science Dimensions y otras bases especializadas como *PubMed* y *Cochrane*. El análisis bibliométrico incluye la construcción de matrices para citación, acoplamiento, análisis de colaboración científica y análisis de co-palabras (Aria & Cuccurullo, 2016a).

Bibliometrix es una herramienta desarrollada en computación estadística y lenguaje gráfico R, de acuerdo con un flujo de trabajo lógico bibliométrico. El lenguaje R es altamente extensible porque es un lenguaje de programación orientado a objetos y funcional. Como tiene una naturaleza de software libre, también es fácil obtener ayuda de la comunidad de usuarios, compuesta principalmente por destacados estadísticos. Por lo tanto, bibliometrix es flexible y puede actualizarse rápidamente, de igual manera puede integrarse con otros paquetes de R estadísticos. Por eso, es útil en una disciplina en constante cambio, como la bibliometría (Aria & Cuccurullo, 2016a).

Aria & Cuccurullo (2016b) mencionan que Biblioshiny es una aplicación que proporciona una interfaz web para bibliometrix. Es compatible con los académicos en el uso fácil de las características principales de bibliometrix, como:

- Importación y conversión de datos a la recopilación de marcos de datos.
- Recopilación de datos mediante la colección Dimensions, PubMed y Scopus API.
- Filtrado de datos.
- Análisis y gráficos para tres métricas de nivel diferentes (fuentes, autores, documentos).
- Análisis de tres estructuras de conocimiento (estructura conceptual, estructura intelectual, estructura social).

En términos generales, el software WordStat8 y la biblioteca spaCy de Python son dos herramientas para realizar procesamiento de lenguaje natural, si bien, en el

segundo caso es necesario adentrarse en la programación. Ambos disponen de una flexibilidad y facilidad de trabajar y analizar datos lingüísticos.

Por el otro lado, el software bibliometrix no es especializado para el procesamiento de lenguaje natural, más bien para ámbitos relacionados al análisis bibliométrico. Para motivos de este trabajo de investigación, se utilizará la biblioteca spaCy para el procesamiento y el software bibliometrix para la generación de datos bibliométricos.

2.11 Organización del conocimiento

La organización del conocimiento es “la ciencia de estructurar sistemáticamente grupos de unidades de conocimiento (conceptos) de acuerdo con sus características inherentes” (Dahlberg, 2006).

Luna González (2015) señala que la organización del conocimiento es una disciplina que pertenece a la ciencia de la información que se encarga de estudiar el tratamiento y la recuperación del conocimiento, así como la construcción y control de los lenguajes e instrumentos utilizados en los procesos de representación de los documentos, permitiendo su posterior recuperación por parte de los usuarios.

Entendiendo esto, se puede definir como una disciplina de la ciencia de la información que se encarga de estructurar sistemáticamente grupos de unidades de conocimiento, así como de estudiar el tratamiento y la recuperación del conocimiento.

El objetivo principal de esta ciencia es aplicar operaciones como la clasificación y la ordenación para crear sistemas de distribución física e intelectual que respeten la secuencialidad, la jerarquía, la asociatividad, etc., con la finalidad de garantizar un acceso adecuado a la información y, por ende, al conocimiento (Amador, Pérez, Pérez, & Peñas, 2016).

2.12 Sistemas de organización del conocimiento

Los sistemas para la Organización del Conocimiento (SOC) son herramientas indispensables en las actividades ordenadoras que se realizan en las bibliotecas. En las áreas encargadas del arreglo de los recursos de información se emplean

varios de estos sistemas, como: clasificaciones, encabezamientos de materias, glosarios, tesauros, etc. (Suárez Sánchez, 2017).

Los SOC abarcan todos los tipos de esquemas para su organización, como son *clasificaciones, taxonomías, encabezamientos de materia, tesauros, ontologías, etc.* Conforman una serie de herramientas que ayudan a crear, gestionar y visualizar modelos que muestren una perspectiva simplificada de los conceptos que forman un determinado dominio del conocimiento y de su estructura semántica (Sánchez, 2009).

El objetivo principal de los SOC es ayudar a la comprensión, gestión y recuperación de los conceptos contenidos en cada dominio del conocimiento. Dado que cualquier disciplina tiene necesidad de describirse de forma comprensible y de facilitar mecanismos de gestión y de recuperación conceptual, los sistemas de organización del conocimiento se encuentran en diversos estudios, desde la pedagogía hasta la inteligencia artificial, pasando por la documentación y la ingeniería de software (Sánchez, 2009). Según Codina, Lluís (2018), estos sistemas se encuentran ordenados de acuerdo con su nivel de complejidad, tal y como se muestra en la Figura 10.

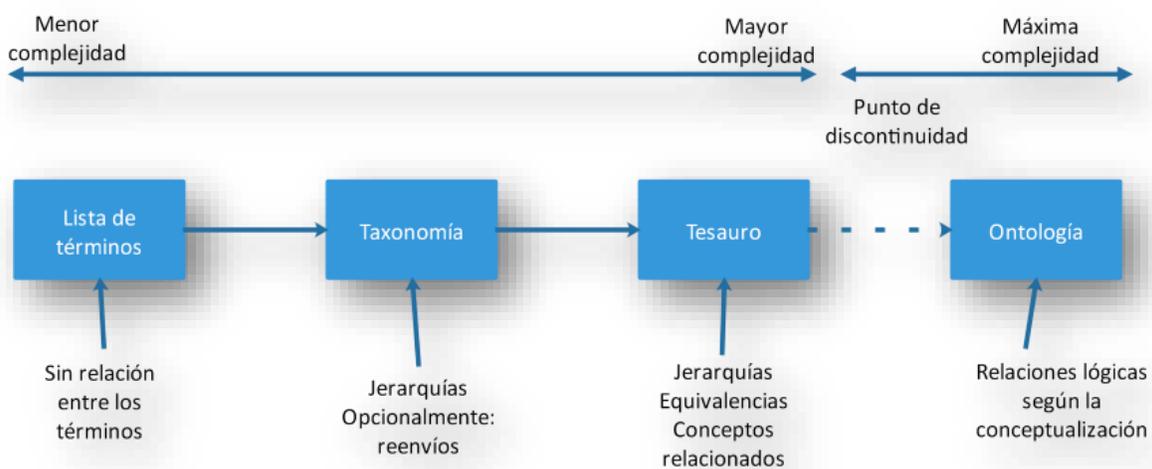


Figura 10. Nivel de complejidad de los Sistemas de organización

Fuente: (Codina, 2018)

2.12.1 Ontologías

Las ontologías son herramientas adecuadas para las definiciones de un vocabulario de representación de un área del conocimiento; coinciden con los tesauros y con las clasificaciones en la capacidad de representación del contenido de un documento a través de la abstracción y de las relaciones entre conceptos (Qin, 2002).

Por otro lado, Codina (2017) menciona que las ontologías se estructuran mediante listas de términos y de las relaciones entre ellos, por medio de clases y subclases [...] una ontología debe estar decodificada en un formato que puede procesar un software, de igual manera, debe estar basada en lógica formal.

Las ontologías se encargan de “definir los términos utilizados para describir y representar un área del conocimiento. Sirven para definir vocabularios que facilitan la búsqueda mediante una herramienta (buscador) que pueda entender, ya que son especificados con precisión como para permitir diferenciar términos y referenciarlos de manera precisa” (Díaz Piraquive, Joyanes Aguilar, & Medina García, 2009).

Considerando lo anterior, podemos mencionar que una ontología representa un vocabulario específico de un área del conocimiento en particular, mediante términos o conceptos abstractos que se relacionan entre sí, basándose en una lógica formal y su estructura es comprendida por un software, con la finalidad de facilitar la recuperación de la información.

Las ontologías se organizan en forma jerárquica, estructuradas en grafos constituidos por nodos y aristas. Cuentan con un nodo “raíz” conectado a todos los otros nodos mediante un único recorrido, de manera que todos los nodos por debajo del nodo inicial tienen solo un nodo padre. Es decir, que la base de una ontología se halla constituida por una generalización o especialización jerárquica de conceptos, como una taxonomía (Barber et al., 2019; Guarino & Oberle, 2009).

Las principales aplicaciones del trabajo con ontologías han sido clasificadas por el Grupo de Trabajo en Ontologías del consorcio del W3C en seis áreas básicas, las cuales son: a) portales web; b) colecciones multimedia; c) administración de sitios

corporativos; d) documentación de diseño; e) agentes inteligentes y f) computación ubicua (Sánchez Trigo, 2012) mejor conocida como la web 4.0.

Así mismo, las ontologías son usadas ampliamente en ingeniería del conocimiento, inteligencia artificial y ciencias de la computación; en aplicaciones orientadas a la gestión del conocimiento, procesamiento de lenguaje natural, comercio electrónico, información de integración inteligente, recuperación de la información, integración de bases de datos, bioinformática y educación (Gómez-Pérez, Fernández-López, & Corcho, 2006).

2.12.1.1 Tipología de ontologías

Van Heijst, Schreiber, & Wielinga (1997) señalan que, las ontologías se pueden clasificar según dos dimensiones: la primera se refiere a la cantidad, el tipo de estructura de la conceptualización y el tema de la conceptualización. Respecto a la primera dimensión, se distinguen tres categorías:

- *Ontologías terminológicas*: Especifican los términos que se utilizan para representar el conocimiento en el dominio del discurso.
- *Ontologías de información*: Especifican la estructura de registros de las bases de datos. Los esquemas de bases de datos son un ejemplo de esta clase de ontologías.
- *Ontologías de modelado del conocimiento*: Especifican conceptualizaciones del conocimiento. En comparación con las ontologías de información, las ontologías de modelado del conocimiento generalmente tienen una estructura interna más rica.

Respecto a la segunda dimensión, se distinguen las siguientes categorías:

- *Ontologías de aplicación*: Contienen todas las definiciones necesarias para modelar el conocimiento requerido para una aplicación en particular. Por lo general, las ontologías de aplicaciones son una combinación de conceptos que se toman de las ontologías de dominio y de las ontologías genéricas.
- *Ontologías de dominio*: Expresan conceptualizaciones que son específicas para dominios particulares.

- *Ontologías genéricas*: Son similares a las ontologías de dominio, pero los conceptos que definen se consideran genéricos en muchos campos. Típicamente, las ontologías genéricas definen conceptos como estado, evento, proceso, acción, componente, etc.
- *Ontologías de representación*: Proporcionan un marco de representación sin hacer afirmaciones sobre el mundo.

Por otro lado, Guarino (1998) las tipifican en las siguientes categorías:

- *Ontologías de nivel superior*: Describen conceptos muy generales como espacio, tiempo, materia, objeto, evento, acción, etc., que son independientes de un problema o dominio particular.
- *Ontologías de dominio y las ontologías de tareas*: Describen respectivamente, el vocabulario relacionado con un dominio genérico (como medicina o automóviles) o una tarea o actividad genérica (como diagnóstico o venta), especializando los términos introducidos en la ontología de nivel superior.
- *Ontologías de aplicación*: Describen conceptos que dependen tanto de un dominio particular como de una tarea, que a menudo son especializaciones de ambas ontologías relacionadas.

Lamarca Lapuente (2013) menciona que de igual manera se pueden establecer distintos tipos de ontologías orientadas a diversos aspectos, como los siguientes:

Según el ámbito de conocimiento al que se apliquen

- *Ontologías generales*: Son las ontologías de nivel más alto ya que describen conceptos generales (espacio, tiempo, materia, objeto, etc.).
- *Ontologías de dominio*: Describen el vocabulario de un dominio concreto del conocimiento.
- *Ontologías específicas*: Son ontologías especializadas que describen los conceptos para un campo limitado del conocimiento o una aplicación concreta.

Según al tipo de agente al que van destinadas

- *Ontologías lingüísticas:* Se vinculan a aspectos lingüísticos, esto es, a aspectos gramaticales, semánticos y sintácticos destinados a su utilización por los seres humanos.
- *Ontologías no lingüísticas:* Destinadas a ser utilizadas por robots y agentes inteligentes.
- *Ontologías mixtas:* Combinan las características de las anteriores.

Según el grado de nivel de abstracción y razonamiento lógico que permitan

- *Ontologías descriptivas:* Incluyen descripciones, taxonomías de conceptos, relaciones entre los conceptos y propiedades, pero no permiten inferencias lógicas.
- *Ontologías lógicas:* Permiten inferencias lógicas mediante la utilización de una serie de componentes como la inclusión de axiomas, etc.

2.12.1.2 Estructura de las ontologías

Los elementos que conforman la estructura de una ontología son los siguientes:

Conceptos: Son la base de una ontología y se presenta en tres modos *simples*, *compuestos* y *por medio de siglas*. Su alcance conceptual se encuentra definido por un dominio de conocimiento (Suárez Sánchez, 2018).

Clases: Describen los conceptos del dominio y se organizan en taxonomías y por lo general, se les aplican mecanismos de herencia (Ramos & Nuñez, 2007).

Instancias: Son los niveles más bajos de una ontología y son un aspecto esencial de proyectos [...] el nivel de instancias es determinado por el dominio, en algunos dominios es prioritario determinar instancias, tal es el caso de especies de animales, tipos de tumores, medicamentos, etc. (Suárez Sánchez, 2018).

Atributos: Definen que es una entidad y emplea sintaxis de conectores. Sirven para generar la estructura del dominio a partir de conceptos y las clases (Suárez Sánchez, 2018).

Relaciones: Representan las interacciones entre los conceptos del dominio, por lo general, contienen relaciones binarias (Ramos & Nuñez, 2007).

Funciones: Son un tipo concreto de relación, donde se identifica un elemento mediante el cálculo a partir de otros elementos que se encuentren en la ontología (Ramos & Nuñez, 2007).

Axiomas: Son teoremas que se declaran sobre relaciones que deban cumplir los elementos de la ontología, usando lógica de primer orden. Por ejemplo: “*Si X y Y son de la clase Z, entonces X no es sub-clase de Y*” o *Para todo X que cumpla con la condición Coun1, A es B*”, etc. (Sánchez López, 2007; Gómez-Pérez et. al., 2006).

2.12.1.3 Metodologías para crear ontologías

Para la creación de cualquier ontología es necesario contar con una metodología específica. En la actualidad, existen diversas metodologías para este fin, como: metodología CYC, metodología de USCHOLD Y KING, metodología GRÜNINGER Y FOX y metodología methontology, la cual es una de las propuestas más completas, ya que toma la creación de ontologías como un proyecto informático (Guzmán Luna, López Bonilla, & Durley Torres, 2012).

2.12.1.3.1 Metodología methontology

La metodología methontology fue desarrollada por el Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid. Esta metodología permite construir ontologías en el nivel de conocimientos, y tiene sus raíces en las actividades identificadas por el proceso de desarrollo de software propuesto por la organización IEEE y en otras metodologías de ingeniería de conocimientos (Corcho, Fernández-López, Gómez-Pérez, & López-Cima, 2005).

Como menciona Corcho et al. (2005) la metodología methontology proporciona una guía sobre el desarrollo de la ontología a través de las actividades de especificación, conceptualización, formalización, implementación y mantenimiento, las cuales se describen a continuación:

- *Especificación:* Permite determinar porque se construye la ontología, cuál será su uso y quienes serán sus usuarios finales.

- *Conceptualización*: Se encarga de organizar y convertir una percepción informal del dominio en una especificación semiformal, basadas en notaciones tabulares y gráficas.
- *Formalización*: Se encarga de la transformación de dicho modelo conceptual a un modelo formal.
- *Implementación*: Se construyen modelos computables en un lenguaje ontológico, como RDF, OWL, etc.
- *Mantenimiento*: Se encarga de la actualización y/o corrección de la ontología, en caso necesario.

Methontology propone una serie de tareas en un orden determinado para la actividad de conceptualización, haciendo énfasis en los componentes (conceptos, atributos, relaciones, constantes, axiomas formales, reglas e instancias) que se construyen en cada una de las tareas, las cuales se muestran a continuación.

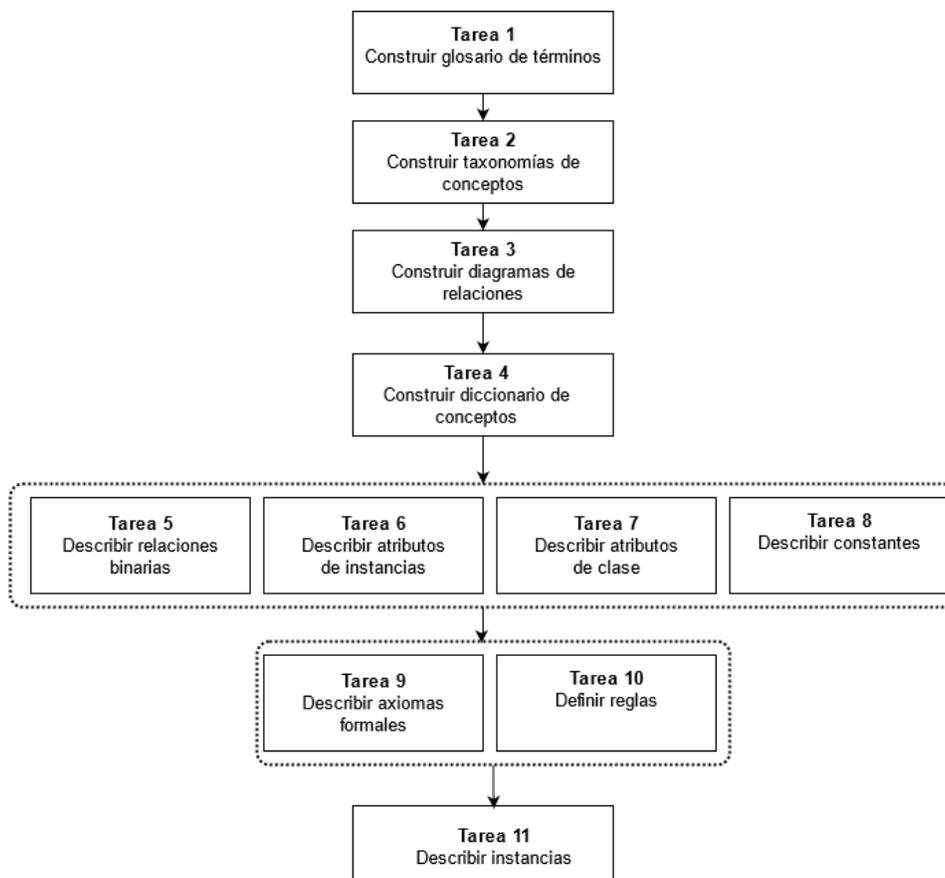


Figura 11. Tareas incluidas en la actividad de conceptualización de methontology

Fuente: (Corcho et al., 2005)

2.12.1.4 Lenguajes de representación de ontologías

Extensible Markup Language (XML)

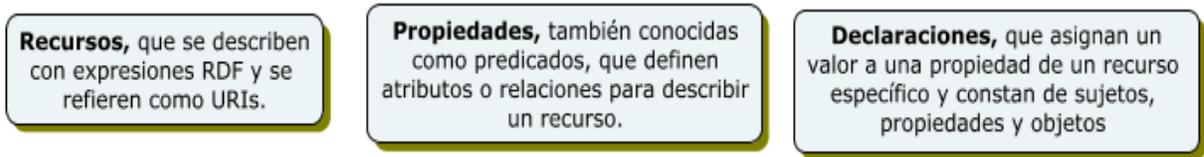
Es definido como un estándar de lenguaje de marcado y una versión simplificada de SGML (Standard Generalized Markup Language), XML fue diseñado especialmente para documentos web. Este lenguaje permite a los diseñadores crear sus propias etiquetas personalizadas, lo cual permite la definición, transmisión e interpretación de datos entre aplicaciones y también entre organizaciones, lo que genera una mejor interoperabilidad (Khosrowpour, 2012).

El lenguaje XML está estrechamente relacionado con la representación de las ontologías web, ya que es la base para el desarrollo del Marco de Descripción de Recursos, en adelante RDF.

Marco de descripción de recursos (RDF)

El RDF es un lenguaje de etiquetado, creado mediante sintaxis XML, que define un modelo de datos para describir recursos en la web (cualquier objeto identificable por un URI) mediante enunciados en forma de tripletas sujeto-predicado-objeto (Cormenzana López, 2017).

El RDF es equivalente a una red semántica, que es un gráfico que incluye un grupo de nodos y aristas. Los nodos representan conceptos, instancias de conceptos y valores de propiedades. Las aristas representan las propiedades de los conceptos o relaciones entre estos (Gómez-Pérez et. al., 2006). El mismo autor nos señala que el lenguaje contiene tres componentes principales, los cuales son:



RDF es un modelo abstracto, de forma que lo especificado mediante este modelo puede serializarse recurriendo a distintos formatos de codificación, como Turtle o JSON-LD. Del mismo modo, otras tecnologías se estructuran mediante RDF y proporcionan una infraestructura para la definición de ontologías basadas en la

Web, que permite una mayor integración e interoperabilidad entre datos ofrecidos y descritos por distintas comunidades, como OWL o SKOS (Cormenzana López, 2017).

Mediante esta combinación del modelo se consigue dotar de significado a las páginas web, lo que la convierte en una de las tecnologías esenciales de la Web Semántica [...] habitualmente, la consulta de datos almacenados en formato RDF, al igual que la manipulación de estos, se realiza a través del lenguaje SPARQL (SPARQL Protocol and RDF Query Language). Los puntos de acceso SPARQL son servicios que aceptan sentencias SPARQL sobre datos RDF a los que se tiene acceso, devolviendo los resultados buscados (Cormenzana López, 2017).

Extensiones de Ontología HTML Simples

Son “un pequeño conjunto de extensiones HTML diseñadas para dar a las páginas web un significado semántico al permitir información como las relaciones de clase, subclase y propiedad” (Heflin, Hendler, & Luke, 1999). De igual manera, Heflin et al. (1999) menciona que tiene como características principales:

A) Es considerado uno de los primeros lenguajes de etiquetado para la estructuración de ontologías de uso web.

B) Las ontologías y registros se manifiestan por medio de etiquetas asignadas a los archivos HTML o XML.

C) Este lenguaje funciona para especificar las clases y relaciones entre clases.

D) Carece de elementos para hacer negaciones.

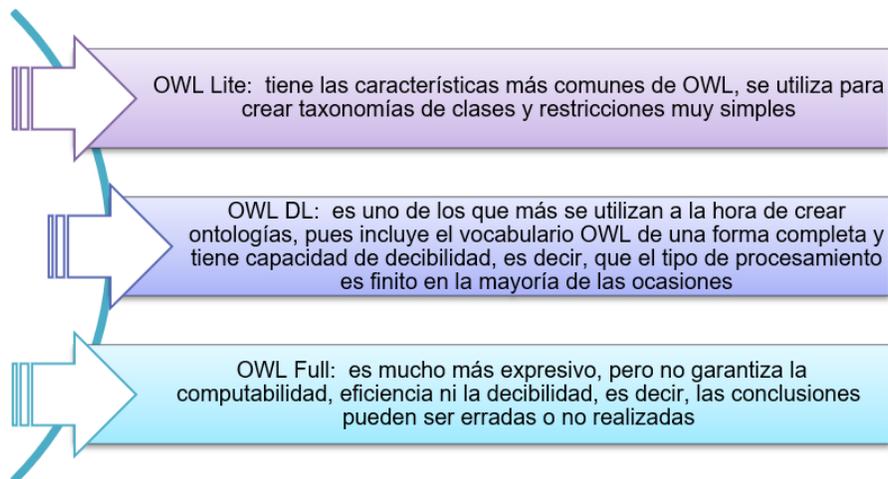
Idioma de Intercambio de Ontologías XML

Es un lenguaje que está diseñado para proporcionar un formato de intercambio de definiciones de ontologías entre un conjunto de partes interesadas. Las definiciones que XOL está diseñado para codificar, incluyen tanto información de esquema (metadatos), como definiciones de clase de bases de datos de objetos, como información no esquemática (hechos básicos), definiciones de objetos de bases de datos de objetos (Karp, Chaudhri, & Thomere, 2002).

Lenguaje de Ontologías Web (OWL)

En el lenguaje OWL se “definen clases, propiedades, instancias, relaciones inversas entre propiedades, clases disyuntas y un conjunto de axiomas, los cuales se utilizan para definir restricciones, que pueden ser tanto de cardinalidad de propiedades como combinaciones de clases” (Staab & Studer, 2009).

OWL es un lenguaje empleado para publicar y compartir datos mediante ontologías. Junto al entorno XML y RDF posibilitan la comprensión de estructuras representadoras y organizadoras tanto para los humanos como las máquinas. Es el lenguaje de representación de ontologías más utilizado y disponible en la mayoría de los editores ontológicos, tanto de código abierto como propietario (Suárez Sánchez, 2018). Según Gómez-Pérez et al. (2006) OWL se compone de los siguientes tres niveles:



2.12.1.5 Herramientas para el desarrollo de ontologías

Los editores ontológicos son softwares que permiten la creación de ontologías. Su función es ser herramientas en las cuales se almacenan los componentes terminológicos, propiedades, desagregaciones y relaciones del dominio de información. Algunos de estos programas son OntoStudio, Ontosaurus, Protégé, WebOnto, IODE, etc. (Suárez Sánchez, 2018).

2.12.1.5.1 Protégé

Protégé es una herramienta de código abierto para el desarrollo de ontologías y sistemas basados en conocimiento. Se desarrolló en Java [...] las aplicaciones que se construyen en Protégé se emplean para la resolución de problemas en dominios específicos. Esta herramienta emplea una interfaz gráfica de usuario integral e intuitiva porque permite la creación de una estructura de frames con clases, slots e instancias (Staab & Svátek, 2006).

Protégé es el software libre de mayor uso para la creación de ontologías. Fue desarrollado por Stanford University. Provee la estructura ontológica y agregaciones de visualización con Graphviz y otros visualizadores ontológicos. Se fundamenta en metas de representación del conocimiento. Tiene alto potencial para el establecimiento de clases, propiedades, desagregación y relaciones. Mantiene relación con estándares con RDF y OWL (Suárez Sánchez, 2018).

La parte teórica de este capítulo representa las herramientas tecnológicas necesarias para un análisis completo de un corpus de información y posteriormente el desarrollo de un sistema de representación y organización del conocimiento.

En términos generales, la información antes mencionada se estructura tomando como base una metodología que facilitará la creación y desarrollo de una ontología sobre el tifo en México.

En el siguiente capítulo, se describe el proceso que se llevó a cabo para el desarrollo del sistema de representación, englobando principalmente el proceso de bibliominería y la metodología menthontology.

Referencias

- Amador, S. R., Pérez, M. D., Pérez, M. J. L. H., & Peñas, D. A. (2016). Patrones para la organización del conocimiento en los Sistemas de Información Curricular. Un caso de estudio. *Investigación Bibliotecológica*, 30(68), 91–107. <https://doi.org/10.1016/j.ibbai.2016.02.005>
- Ardanuy, J. (2009). Breve introducción a la bibliometría. <https://doi.org/10.1038/nmat3485>
- Aria, M., & Cuccurullo, C. (2016a). Bibliometrix. Recuperado el 22 de abril de 2020, de <https://www.bibliometrix.org/>
- Aria, M., & Cuccurullo, C. (2016b). Biblioshiny. Recuperado el 22 de abril de 2020, de <https://bibliometrix.org/Biblioshiny.html>
- Babych, B., & Hartley, A. (2003). *Improving machine translation quality with automatic named entity recognition*. Trabajo presentado en 7th International EAMT workshop on MT and other Language Technology Tools, Improving MT through other Language Technology Tools: Resources and Tools for Building M. Recuperado de <https://dl.acm.org/doi/10.5555/1609822.1609823>
- Barber, E. E., Pisano, S., Romagnoli, S., De Pedro, G., Gregui, C., & Blanco, N. (2019). Fundamentos lógicos de las ontologías Web. *Información, cultura y sociedad*, (41), 81–100. <https://doi.org/10.34096/ics.i41.6196>
- Beltrán Martínez, B. (s/f). *Minería de datos* (Trabajo de grado, Benemerita Universidad Autónoma de Puebla), México.
- Bharati, M., & Ramageri, L. (2010). Data mining: Techniques and applications. *Indian Journal of Computer Science and Engineering*, 1(4), 301–305. <https://doi.org/10.1201/b21638>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol: O'Reilly.
- Blázquez Ochando, M. (2012). Técnicas avanzadas de recuperación de información: Frecuencias y pesos de los términos de un documento. Recuperado de <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com/2012/11/frecuencias-y-pesos-de-los-terminos-de.html>
- Bonilla Gordillo, A. F., Ojeda Schuldt, M. A., & Echeverría Briones, F. (2006). *Implementación de minería de datos basada en redes bayesianas para la toma de decisiones en los registros académicos*. Escuela Superior Politécnica del Litoral.
- Bouza, C. N., & Santiago, A. (2012). La minería de datos: árboles de decisión y su aplicación en estudios médicos. *Modelación Matemática de Fenómenos del Medio Ambiente y la Salud*, (2). Recuperado de https://s3.amazonaws.com/academia.edu.documents/43713947/MINERIA_DE_DATOS_MEDICOS.pdf?response-content-disposition=inline%3B

filename%3DLA_MINERIA_DE_DATOS_ARBOLES_DE_DECISION.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20

- Bretonnel Cohen, K., & Demner-Fushman, D. (2014). Introduction to natural language processing. En John Benjamins Publishing Company (Ed.), *Biomedical Natural Language Processing*. Recuperado de <https://books.google.es/books?hl=es&lr=&id=vXvPAgAAQBAJ&oi=fnd&pg=PR1&dq=Verspoor+y+Kevin+Brettonne+Cel&ots=ZGclNaErKH&sig=EuKdrBeRG2DePeihclX8bjLhY60#v=onepage&q&f=false>
- Candás Romero, J. (2006). Minería de datos en bibliotecas: bibliominería. *BiD: Textos universitaris de biblioteconomia i documentació*, (17), 4. Recuperado de <http://bid.ub.edu/17canda2.htm>
- Carrasco, R. C. (2003). *La ley de Zipf en la Biblioteca Miguel de Cervantes*. España: Universidad de Alicante.
- Collazo Reyes, F., Luna Morales, M. E., & Luna Morales, E. (2016). Impronta genética institucional y desarrollo de identidad internacional: estudio bibliométrico sobre médula-espinal-gato en Cinvestav. *Investigación Bibliotecológica*, 30(70), 41–60. <https://doi.org/10.1016/j.ibbai.2016.10.003>
- Codina, L. (2017). *Taxonomías y Ontologías: Qué son y cómo se aplican en Medios de Comunicación*. Recuperado el 16 de marzo de 2020, de <https://www.lluiscodina.com/taxonomias-ontologias/>
- Codina, L. (2018). *Qué es una taxonomía y cómo se aplica a un sitio web: Área de Formatos Digitales* (F. de C. U. P. Fabra, Ed.). Recuperado de <https://www.lluiscodina.com/taxonomia-sitio-web/>
- Combaranza, G. (2003). Una introducción a la teoría de grafos. Trabajo presentado en *XIV Encuentro de Geometría y II de Aritmética*. Recuperado de <http://funes.uniandes.edu.co/6102/1/CombaranzaUnaIntroducciónGeometría2003.pdf>
- Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Construcción de ontologías legales con la metodología methontology y la herramienta WebODE. *Web Semántica y ontologías jurídicas. Aplicaciones para el derecho en la nueva generación de la red*, 1–17. Recuperado de <http://oa.upm.es/5289/>
- Cormenzana López, R. (2017). *RDA y BIBFRAME: hacia un nuevo paradigma de catalogación*. 1–40. Recuperado de <https://orcid.org/0000-0003-1240-4697>
- Cortez Vásquez, A., Vega Huerta, H., & Pariona Quispe, J. (2009). Procesamiento de lenguaje natural. *Revista de Ingeniería de Sistemas de Informática*, 6(2), 10. Recuperado de <https://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923/5121>

- Dahlberg, I. (2006). Knowledge organization: A new science? *Knowledge Organization*, 33(1), 11–19.
- Díaz Piraquive, F. N., Joyanes Aguilar, L., & Medina García, V. H. (2009). Taxonomía, ontología y folksonomía, ¿qué son y qué beneficios u oportunidades presentan para los usuarios de la web? *Universidad & Empresa*, 8(16), 242–261. Recuperado de <https://www.redalyc.org/pdf/1872/187214803010.pdf>
- Escorcía Otalora, T. A. (2008). *El análisis bibliométrico como herramienta para el seguimiento de publicaciones científicas, tesis y trabajos de grado* (Trabajo de grado, Pontificia Universidad Javeriana.). Recuperado de <https://www.javeriana.edu.co/biblos/tesis/ciencias/tesis209.pdf>
- Escobar Terán, H., Alcivar, M., & Puris, A. (2016). Aplicaciones de Minería de Datos en Marketing. *Revista Publicando*, 3(8), 503–512.
- Fayyad, U. M., Grinstein, G. G., & Wierse, A. (2002). *Information visualization in data mining and knowledge discovery*. San Francisco, California: Morgan Kaufmann Publishers.
- Falcato, P. (1989). *La ley de Bradford y sus aplicaciones: una introducción*. Recuperado de <http://eprints.rclis.org/39384/>
- Garbade, M. J. (2018). *A Simple Introduction to Natural Language Processing*. Recuperado el 3 de abril de 2020, de <https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32>
- Gera, M., & Goel, S. (2015). Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity. *International Journal of Computer Applications*, 113(18), 22–29. <https://doi.org/10.5120/19926-2042>
- Godsil, C., & Royle, G. (2001). Graphs. En *Algebraic Graph Theory* (pp. 1–18), New York: Springer. https://doi.org/10.1007/978-1-4613-0163-9_1
- Gómez-Pérez, A., Fernandez-Lopez, M., & Corcho, O. (2006). *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. España: Springer Science & Business Media. Recuperado de https://books.google.com.mx/books/about/Ontological_Engineering.html?id=qR_3BwAAQBAJ&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false
- González-Moreno, D. A. (2017). *Introducción a la Teoría de las Gráficas*. Recuperado de http://www.cua.uam.mx/pdfs/conoce/libroselec/24Libro_Introduccion_a_la_teor%C3%ADa_de_las_gra.pdf
- González Villa, J. (2019). TF IDF: herramientas para mejorar la relevancia de tus contenidos. Recuperado el 21 de junio de 2020, de <https://useo.es/tf-idf-relevancia/>

- Grupo Logicalis. (2014). *Minería de datos: aplicaciones más populares a día de hoy*. Recuperado el 8 de febrero de 2020, de <https://blog.es.logicalis.com/analytics/mineria-de-datos-aplicaciones-que-ya-son-una-realidad>
- Guarino, N., & Oberle, D. (2009). What is an ontology? En *Handbook on Ontologies* (pp. 1–17). <https://doi.org/10.1007/978-3-540-92673-3>
- Guarino, N. (1998). Formal Ontology and Information Systems. En N. Guarino (Ed.), *Formal Ontology in Information Systems* (pp. 3-15), Trento, Italia: IOS Press.
- Guzmán Luna, J. A., López Bonilla, M., & Durley Torres, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica, Año XVII(50)*, 133–140.
- Han, J., Kember, M., & Pei, J. (2012). What Is Data Mining? En *Data mining: Data mining concepts and techniques* (3a ed., pp. 203–207). United States of America: Elsevier; Morgan Kaufmann Publishers.
- Heflin, J., Hendler, J., & Luke, S. (1999). *SHOE: A Knowledge Representation Language for Internet Applications*. Recuperado de https://www.researchgate.net/publication/2620999_SHOE_A_Knowledge_Representation_Language_for_Internet_Applications
- Hernández, M., & Gómez, J. (2013). Aplicaciones de Procesamiento de Lenguaje Natural. *Revista Politécnica*, 32(1), 87–96. Recuperado de https://rua.ua.es/dspace/bitstream/10045/33514/1/2013_Hernandez_Gomez_RevPolitec.pdf
- Jain, A., Kulkarni, G., & Shah, V. (2018). Natural Language Processing. *International Journal of Computer Sciences and Engineering*, 6(1). <https://doi.org/10.26438/ijcse/v6i1.161167>
- Jeremy Norman's. (2020). *Bradford's Law: Exponentially Diminishing Returns of Extending a Library Search : History of Information*. Recuperado el 9 de febrero de 2020, de <http://www.historyofinformation.com/detail.php?id=614>
- Jiang, J. (2013). Information extraction from text. En *Mining Text Data* (pp. 11–41). https://doi.org/10.1007/978-1-4614-3223-4_2
- Karp, P. D., Chaudhri, V. K., & Thomere, J. (2002). *XOL - XML-Based Ontology Exchange Language*. Recuperado el 17 de marzo de 2020, de Technology Reports website: <http://xml.coverpages.org/xol.html>
- Khosrowpour, M. (2012). Extensible Markup Language. En *Dictionary of Information Science and Technology* (2° ed.). <https://doi.org/10.4018/978-1-4666-2624-9>
- Lamarca Lapuente, M. J. (2013). *Hipertexto, el nuevo concepto de documento en la cultura de la imagen* (Tesis doctoral, Universidad Complutense de Madrid). Recuperado de <http://www.hipertexto.info/documentos/ontologias.htm>
- Landa, J. (2016). *¿Qué es KDD y Minería de Datos?* Recuperado el 8 de febrero de

2020, de <http://fcojlanda.me/es/ciencia-de-los-datos/kdd-y-mineria-de-datos-espanol/>

- Luna González, M. E. (2015). Organización del conocimiento en la red digital. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*, 29(67), 77–89. <https://doi.org/10.1016/j.ibbai.2016.02.037>
- Maimon, O., & Rokach, L. (2010). Introduction to Knowledge Discovery in Databases. En *Data mining and knowledge discovery handbook*. (pp. 1–17). Berlin: Springer.
- Mata, R. (2017). *Minería de datos: qué es, cómo es el proceso y a qué áreas se puede aplicar*. Recuperado el 8 de febrero de 2020, de <https://www.icemd.com/digital-knowledge/articulos/mineria-datos-proceso-areas-se-puede-aplica/>
- Martínez, R. (2011). *Grafos ponderados*. Recuperado el 14 de marzo de 2020, de <https://blogs.ua.es/matematicadiscrecion/2011/01/06/tema-7-grafos-ponderados-sesion-21122010/>
- Montemurro, M. A. (2001). Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Physica A: Statistical Mechanics and its Applications*, 300(3–4), 567–578. [https://doi.org/10.1016/S0378-4371\(01\)00355-7](https://doi.org/10.1016/S0378-4371(01)00355-7)
- Nicholson, S. (2003). The Bibliomining Process: Data Warehousing and Data Mining for Library Decision Making. *Information Technology and Libraries*, 22(4), 146–151. Recuperado de <http://scottnicholson.com/pubs/biblioprocess.pdf>
- Nicholson, S., & Stanton, J. (2003). Gaining Strategic Advantage Through Bibliomining: Data Mining for Management Decisions in Corporate, Special, Digital, and Traditional Libraries. En *Organizational Data Mining: Leveraging Enterprise Data Resources* (pp. 247–262). <https://doi.org/10.4018/978-1-59140-134-6.ch017>
- Paul E., B. (2004). *Bradford's law*. Recuperado el 9 de febrero de 2020, de Dictionary of Algorithms and Data Structures website: <https://xlinux.nist.gov/dads/HTML/bradfordslaw.html>
- Pérez Salazar, G. (2014). Análisis de Redes Sociales: una introducción conceptual. En Universidad Pedagógica Nacional (Ed.), *Ensayos de comunicación, educación y tecnología: vertientes entrelazadas* (pp. 7–20). Recuperado de https://www.researchgate.net/publication/262485338_Analisis_de_redes_sociales_Una_introduccion_conceptual
- Piatetsky-Shapiro, G., & Frawley, W. (1991). *Knowledge discovery in databases*. Recuperado de <https://mitpress.mit.edu/books/knowledge-discovery-databases>
- Provalis Research. (s/f). *Content analysis and text mining software*. Recuperado de <https://provalisresearch.com/products/content-analysis-software/>
- Puente, M. de la. (2010). Bibliominería: bibliometría y minería de datos. *Consultora de Ciencias de la Información*, 9(1), 76–99. Recuperado de

<http://eprints.rclis.org/14459/>

Qin, J. (2002). *Evolving Paradigms of Knowledge Representation and Organization : A Comparative Study of Classification , XML / DTD , and Ontology*.

Quiroz Gil, N. L. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim. *Revista Interamericana de Bibliotecología*, 35(1), 97–108. Recuperado de <http://aprendeenlinea.udea.edu.co/revistas/index.php/RIB/article/view/13341/1937>

Ramos, E., & Nuñez, H. (2007). Ontologías: componentes, metodologías, lenguajes, herramientas y aplicaciones. *Lecturas en Ciencias de la Computación*. Recuperado de <https://docplayer.es/54553669-Ontologias-componentes-metodologias-lenguajes-herramientas-y-aplicaciones.html>

Rapozo, E. (s/f). *Tópicos especiais em Tecnologia de Computação*. Recuperado el 14 de marzo de 2020, de <https://slideplayer.com.br/slide/375110/>

Rochina, P. (2017). *El análisis de redes sociales mediante la teoría de grafos*. Recuperado de <https://revistadigital.inesem.es/informatica-y-tics/teoria-grafos/>

Rossum, G. van, & Python development team. (2018). *Python Tutorial: release 3.7.0*. Recuperado de https://bugs.python.org/file47781/Tutorial_EDIT.pdf

Sánchez López, S. E. (2007). *Modelo de indexación de formas en sistemas VIR basado en ontologías*. Escuela de Ingeniería y Ciencias, Universidad de las Américas Puebla.

Sánchez Trigo, E. (2012). Recursos para la traducción del discurso especializado: ontologías y terminología. En *Traducción e interpretación: estudios, perspectivas y enseñanzas*. Recuperado de <https://books.google.com.mx/books?id=GG0A68pCACMC&printsec=copyright#v=onepage&q&f=false>

Sánchez Pérez, C. R. (2008). *Clasificación de Entidades Nombradas utilizando Información Global* (Instituto Nacional de Astrofísica, Óptica y Electrónica). Recuperado de <http://cseg.inaoep.mx/~mmontesg/tesis-estudiantes/TesisMaestria-CarolinaSanchez.pdf>

Sánchez, S. (2009). *Organización del conocimiento, Sistema de || KOS*. Recuperado el 16 de marzo de 2020, de <http://glossarium.bitrum.unileon.es/Home/organizacion-del-conocimiento-sistema-de>

Sanz Menéndez, L. (2003). Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes. *Apuntes de Ciencia y Tecnología*, (7), 21–28.

Scla, M. (2016). Camino mínimo en grafos. Recuperado de <http://www.oia.unsam.edu.ar/wp-content/uploads/2017/11/dijkstra-prim.pdf>

- Serrat, O. (2017). Social Network Analysis. En *Knowledge Solutions* (pp. 39–43). https://doi.org/10.1007/978-981-10-0983-9_9
- Souza, F., Nogueira, R., & Lotufo, R. (2019). *Portuguese Named Entity Recognition using BERT-CRF*. Recuperado de <http://arxiv.org/abs/1909.10649>
- SpaCy. (2020a). spaCy 101: *Everything you need to know*. Recuperado el 16 de noviembre de 2020, de https://spacy.io/usage/spacy-101#_title
- Staab, S., & Studer, R. (2009). *Handbook on Ontologies* (S. Staab & R. Studer, Eds.). <https://doi.org/10.1007/978-3-540-92673-3>
- Staab, S., & Svátek, V. (Eds.). (2006). *Managing Knowledge in a World of Networks*. <https://doi.org/10.1007/11891451>
- Stack Overflow. (s/f). *Python: notes for professionals*. 790. Recuperado de <https://goalkicker.com/PythonBook>
- Suárez Sánchez, A. (2017). Sistemas para la organización del conocimiento: definición y evolución histórica. *e-Ciencias de la Información*, 7(2), 1. <https://doi.org/10.15517/eci.v7i2.26878>
- Suárez Sánchez, A. (2018). *Ontologías: fundamentos y aplicaciones , una aproximación desde la perspectiva bibliotecológica*. Universidad Autónoma de México.
- Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, S. J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 63–86). <https://doi.org/10.16925/9789587600490>
- Tolpygo, A. (2016). *Natural Language Processing: Relation Extraction I*. Recuperado el 21 de abril de 2020, de <https://sflscientific.com/data-science-blog/2015/12/11/natural-language-processing-information-extraction>
- Unidad de Información y Análisis Financiero. (2014). *Técnicas de minería de datos para la detección y prevención del lavado de activos y la financiación del terrorismo (LA/FT)*. República de Colombia.
- Urbizagástegui Alvarado, R. (2005). La productividad científica de los autores: un modelo de aplicación de la ley de Lotka por el método del poder inverso generalizado. *Información, cultura y sociedad*, 0(12), 51–73. <https://doi.org/10.34096/ICS.112.904>
- Urbizagástegui Alvarado, R., & Restrepo Arango, C. (2011). La ley de Zipf y el punto de transición de Goffman en la indización automática. *Investigación Bibliotecológica*, 25(54), 71–92. <https://doi.org/10.22201/iibi.0187358xp.2011.54.27482>
- Van Heijst, G., Schreiber, A. T., & Wielinga, B. J. (1997). Using explicit ontologies in

KBS development. *International Journal of Human Computer Studies*, 46(2–3), 183–292. <https://doi.org/10.1006/ijhc.1996.0090>

Capítulo 3. Bibliominería y metodología Methontology para la creación de una ontología sobre el tifo en México 1904-1977

En este capítulo se describe cómo se implementó la bibliominería metodológica con la utilización del procesamiento de lenguaje natural y la metodología methontology para la construcción de la ontología del tifo.

3.1 Aplicación de la bibliominería

Con base en una red de ontologías sobre eventos meteorológicos elaborada por Vilches-Blázquez, Comesaña, & Arrieta Moreno (2020), se realizó un esquema (Figura 12) donde se muestran las actividades que se efectuaron por medio de la bibliominería metodológica para la creación de una ontología sobre el tema del tifo en México, mismas que se especifican en los siguientes párrafos.

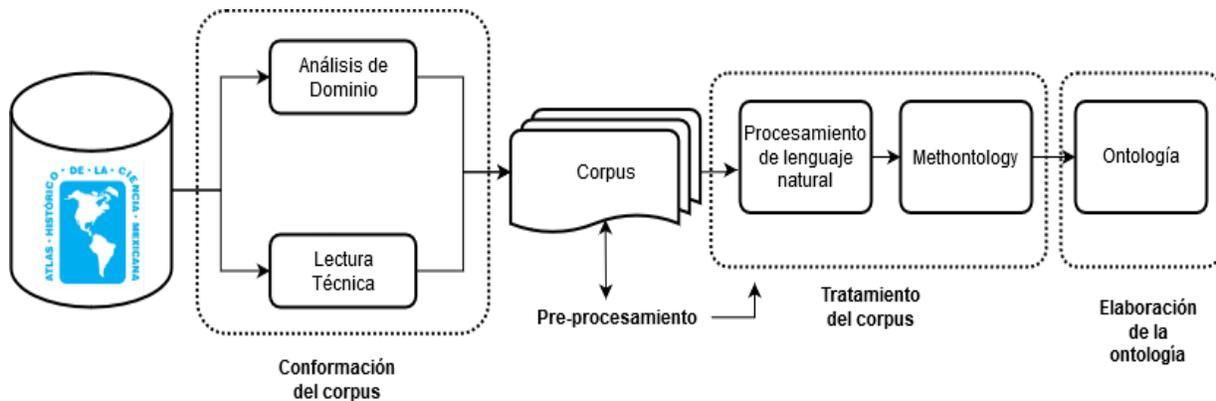


Figura 12. Actividades para la construcción de una ontología.

Fuente: Elaborado por los autores

3.1.1 Identificación del área de aplicación

Como se mencionó anteriormente, el área de aplicación es un dominio de conocimiento específico de la medicina, referente a la enfermedad epidémica del tifo en México, en el periodo 1904-1977. Lo anterior se debe principalmente a que en ese periodo se desarrollaron las investigaciones etiológicas, el desarrollo de vacunas preventivas y la erradicación de la enfermedad.

3.1.2 Identificación de fuentes de datos internas y externas

La fuente de datos seleccionada para la recuperación de la información fue la base de datos del Atlas Histórico de la Ciencia Mexicana (AHCM), un proyecto desarrollado en el área de bibliometría del Centro de Investigaciones y Estudios Avanzados del IPN (CINVESTAV).

Fue iniciado en el año 2004 como un trabajo colaborativo de los departamentos de física, matemáticas y matemática educativa. Actualmente cuenta con 31,537 registros bibliográficos que corresponden a los trabajos sobre México producidos dentro y fuera del país durante el período 1795-1979. (Atlas Histórico de la Ciencia Mexicana [AHCM], 2017).

Se puede acceder directamente a la base de datos del AHCM escaneando el siguiente código QR o entrando al siguiente enlace:
<http://bibliometria.bfm.cinvestav.mx/presentacion.html>



Figura 13. Código QR para acceder a la base de datos AHCM

Fuente: Elaborado por los autores

3.1.3 Recolección y limpieza de los datos en un almacén de datos

Esta etapa consistió en dos partes fundamentales. La primera está orientada a la recolección de los datos (*conformación del corpus*), y la segunda a su limpieza y almacenamiento (*preprocesamiento*).

➤ **Recolección de datos**

Una vez identificada la fuente de datos se procedió a la recuperación de la producción científica en el tema del tifo en México durante el periodo antes mencionado. En este punto se consideran dos pasos importantes, a) *Análisis de dominio* y b) *Lectura técnica*. El primer paso hace referencia a que “la mejor manera de comprender la información es estudiar los dominios del conocimiento como comunidades de pensamiento o discurso” (Hjørland & Albrechtsen, 1995).

En otras palabras, el análisis de dominio se refiere a conocer los términos técnicos de una ciencia, disciplina particular, o un cuerpo de conocimientos temático específico, como este caso, con la finalidad de comprender mejor la información que ahí se trate. Con la terminología obtenida se procedió a identificar los términos dentro de los artículos científicos mediante la lectura técnica. Implementando el análisis de dominio y la lectura técnica se obtuvieron 239 registros de artículos en la base de datos del AHCM, cerca del 1% del total de registros en la base de datos.

Para tener un control sobre los datos bibliográficos de los artículos, se organizaron en una hoja de cálculo en Excel, considerando los siguientes campos de control bibliográfico: *autores, adscripción, título, revista, título abreviado, rev-iso, ISSN, año, vol., número, pag-1, pag-2, citas, URL, DOI, tipo documento, idioma, institución, resumen, área temática, palabras clave, etc.* En la Figura 14 se muestran los registros con la información de algunos de los campos mencionados.

Clave trabajo	CLAVE-ACM_Autores	ADSCRIPCION	TITULO	REVISTA	TITULO ABRE REV-ISO	ISSN	AÑO	VOL	Numero	Pag-1	Pag-2	No-Pag	No REF
50	AHCM:0033-:ANONIMO	EXT	TYPHUS FEVE PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1906	21	16	393	393	1	0
51	AHCM:0033-:ANONIMO	EXT	MEXICO PRE PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1906	21	6	131	131	1	0
52	AHCM:0033-:ANONIMO	EXT	MEXICO SUB PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1906	21	29	853	853	1	0
53	AHCM:0033-:ANONIMO	EXT	REPORT FRO PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1906	21	16	393	393	1	0
54	AHCM:0033-:FRICK, EBER	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1906	21	11	259	260	2	0
55	AHCM:0016-:GAVINO-IGU	LOC	JUICIOS Y CR GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	6	210	0	1	0
56	AHCM:0016-:LICEAGA, EDI	LOC	ALGUNOS D GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	6	287	0	1	0
57	AHCM:0016-:MENDIZABA	LOC	PROFILAXIS GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	7	315	342	1	0
58	AHCM:0016-:PRIETO, IGNU	LOC	NOTA PREVE GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	6	233	0	1	0
59	AHCM:0016-:SORIANO, M	LOC	ALGUNOS F GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	9	409	0	1	0
60	AHCM:0016-:SORIANO, M	LOC	DICTAMEN D GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1906	1	6	254	0	1	0
61	AHCM:0033-:GOTTSCHALH	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1907	22	35	1222	1222	1	0
62	AHCM:0033-:GOTTSCHALH	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1907	22	39	1373	1373	1	0
63	AHCM:0033-:GOTTSCHALH	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1907	22	21	691	691	1	0
64	AHCM:0016-:LUGO-HIDAL	LOC	LA ANURIA E GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1907	2	24	325	0	1	0
65	AHCM:0016-:TERRES, JOSE	LOC	PRONOSTICC GACETA MEE	GAC MED ME	Gac. Med. M	0016-3813	1907	2	13	167	0	1	0
66	AHCM:0033-:ANONIMO	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1908	23	46	1660	1661	2	0
67	AHCM:0033-:ANONIMO	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1908	23	47	1702	1702	1	0
68	AHCM:0033-:ANONIMO	EXT	MEXICO REPI PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1908	23	48	1740	1740	1	0
69	AHCM:0033-:ANONIMO	EXT	REPORT FRO PUBLIC HEAL	PUBLIC HEAL	Public Health	0033-3549	1908	23	45	1625	1625	1	0
70	AHCM:0096-:CARBAJAL, A	LOC	COMPLICACI MEMORIAS E	MEM SOC CII	Mem. Soc. CI	0396-1861	1908	27	1	5	17	15	0

Figura 14. Registros de la producción científica del tifo

Fuente: (Atlas Histórico de la Ciencia Mexicana)

Posteriormente, se procedió a la recuperación de los textos disponibles en línea por medio de diversos sitios web pertenecientes a las revistas donde se encuentran publicados los documentos y principalmente a través de la base de datos *Web of Science* (www.webofknowledge.com).

Los artículos recuperados se exportaron en formato PDF (Portable Document Format); los que no se encontraron en línea, se localizaron físicamente en la biblioteca especializada en historia de la medicina “Nicolas León” de la UNAM. Los documentos recuperados directamente de la biblioteca se digitalizaron con la herramienta CamScanner y se guardaron en formato PDF.

El siguiente paso consistió en realizar un análisis basado en la tecnología de Reconocimiento Óptico de Caracteres (OCR, Optical Character Recognition) mediante el software ABBYY FineReader 12, que permite convertir documentos escaneados, imágenes, PDF y fotografías digitales en archivos electrónicos compatibles con la búsqueda y edición posterior como Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Rich Text.

Este procedimiento nos permitió extraer el texto de los artículos y nos facilitó su tratamiento automático. En la Figura 15 y 16, se muestran ejemplos del análisis OCR efectuado en los artículos científicos del tifo.

LESIONES CEREBRALES

Una búsqueda minuciosa de lesiones cerebrales tíficas en cuyes inoculados con ambas cepas demostró que la orquítica "L" tiene poca tendencia a producirlas en comparación con las no orquíticas. Los animales fueron sacrificados en el séptimo, décimo primero y décimo sexto día, de preferencia en el décimo primero. Ocho cepas no orquíticas presentaron lesiones cerebrales tíficas siendo particularmente numerosas en la cepa número 42, por lo cual la consideramos representativa de las cepas no orquíticas mexicanas. Se ha iniciado un estudio sistemático similar al publicado por Lillie, Dyer y Armstrong (8) con el objeto de adquirir mayores informes sobre las lesiones cerebrales.

PASES DE CEPAS NO ORQUÍTICAS MEXICANAS A TRAVES DE RATON

Nicolle y Laigret (9) y Laigret y Jadin (10) encontraron que el tifo del Viejo Mundo se perdía por pases sucesivos en ratas y ratones, contrariamente a lo que ocurre con las cepas murinas en las que al parecer se conserva dicho virus indefinidamente. Consideramos interesante probar nuestras cepas no orquíticas en el ratón. Las cepas números 27, 42, 44 y 45 fueron inoculadas y pasadas en este animal por tres o cuatro generaciones. Los cuyes que fueron inoculados en el mismo orden para demostrar la presencia de la infección tífica, revelaron que estas cepas fueron perdidas del segundo al tercer pase de ratón a ratón. Estos experimentos agregan una prueba más a la similitud entre nuestras cepas no orquíticas y las cepas del Viejo Mundo.

EXPERIMENTOS DE VACUNACION CONTRA CEPAS ORQUÍTICAS Y NO ORQUÍTICAS MEXICANAS

CASTAÑEDA - SILVA - VARIEDADES DEL TIFO

529

LESIONAS CEREBRALES

Una búsqueda minuciosa de lesiones cerebrales tíficas en cuyes inoculados con ambas cepas demostró que la orquítica "L" tiene poca tendencia a producirlas en comparación con las no orquíticas. Los animales fueron sacrificados en el séptimo, décimo primero y décimo sexto día, de preferencia en el décimo primero. Ocho cepas no orquíticas presentaron lesiones cerebrales tíficas siendo particularmente numerosas en la cepa número 42, por lo cual la consideramos representativa de las cepas no orquíticas mexicanas. Se ha iniciado un estudio sistemático similar al publicado por Lillie, Dyer y Armstrong (8) con el objeto de adquirir mayores informes sobre las lesiones cerebrales.

PASES DE CEPAS NO ORQUÍTICAS MEXICANAS A TRAVES DE RATON

Nicolle y Laigret (9) y Laigret y Jadin (10) encontraron que el tifo del Viejo Mundo se perdía por pases sucesivos en ratas y ratones, contrariamente a lo que ocurre con las cepas murinas en las que al parecer se conserva dicho virus indefinidamente. Consideramos interesante probar nuestras cepas no orquíticas en el ratón. Las cepas números 27, 42, 44 y 45 fueron inoculadas y pasadas en este animal por tres o cuatro generaciones. Los cuyes que fueron inoculados en el mismo orden para demostrar la presencia de la infección tífica, revelaron que estas cepas fueron perdidas del segundo al tercer pase de ratón a ratón. Estos experimentos agregan una prueba más a la similitud entre nuestras cepas no orquíticas y las cepas del Viejo Mundo.

EXPERIMENTOS DE VACUNACION CONTRA CEPAS ORQUÍTICAS Y NO ORQUÍTICAS MEXICANAS

Figura 15. Artículos en español utilizando OCR

Fuente: Elaborada por los autores

A NOTE ON IMPROVEMENT IN THE METHOD OF VACCINE PRODUCTION WITH RICKETTSIAE OF MEXICAN TYPHUS FEVER

HANS ZINSSER AND M. RUIZ CASTANEDA¹

From the Department of Bacteriology and Immunology of the Harvard University Medical School

Received for publication April 23, 1931

In previous publications (1, 2, 3) the writers have reported the successful protection of guinea pigs against infection with both Mexican and European typhus virus by intraperitoneal administrations of Rickettsia suspensions killed with 0.2 per cent formalin in salt solution. Since that time, experiments have been carried out with monkeys which cannot, however, be regarded as anything more than encouraging; for although two vaccinated monkeys appeared to be immune to massive injections of European typhus virus, and two control monkeys exhibited characteristic temperature curves, a third control monkey failed to react. The guinea pig experiments, however, have given such definite results that we consider it important to submit our present method of vaccine production for further investigation by others.

A NOTE ON IMPROVEMENT IN THE METHOD OF VACCINE PRODUCTION WITH RICKETTSIAE OF MEXICAN TYPHUS FEVER

HANS ZINSSER AND M. RUIZ CASTANEDA¹

From the Department of Bacteriology and Immunology of the Harvard University Medical School

Received for publication April 23, 1931

In previous publications (1, 2, 3) the writers have reported the successful protection of guinea pigs against infection with both Mexican and European typhus virus by intraperitoneal administrations of Rickettsia suspensions killed with 0.2 per cent formalin in salt solution. Since that time, experiments have been carried out with monkeys which cannot, however, be regarded as anything more than encouraging; for although two vaccinated monkeys appeared to be immune to massive injections of European typhus virus, and two control monkeys exhibited characteristic temperature curves, a third control monkey failed to react. The guinea pig experiments, however, have given such definite results that we consider it important to submit our present method of vaccine production for further investigation by others.

Figura 16. Artículos en inglés utilizando OCR

Fuente: Elaborada por los autores

Dentro de las limitaciones presentadas en el OCR, fue que algunos artículos estaban mal digitalizados, por lo cual, los párrafos estaban desalineados, el texto estaba mal enfocado, etc., esto dificultó el reconocimiento y, por ende, la recuperación total del contenido de los documentos. Para esto se realizaron las correcciones necesarias por medio del preprocesamiento.

➤ Limpieza de datos

Esta parte pertenece a la actividad de preprocesamiento del corpus, en la cual se corrigieron manualmente los errores ortográficos, gramaticales y de continuidad léxica generada por el OCR/ABBYY FineReader 12. De esta manera se facilitó el tratamiento de estos textos. La Figura 17 muestra un ejemplo de la limpieza de un texto.

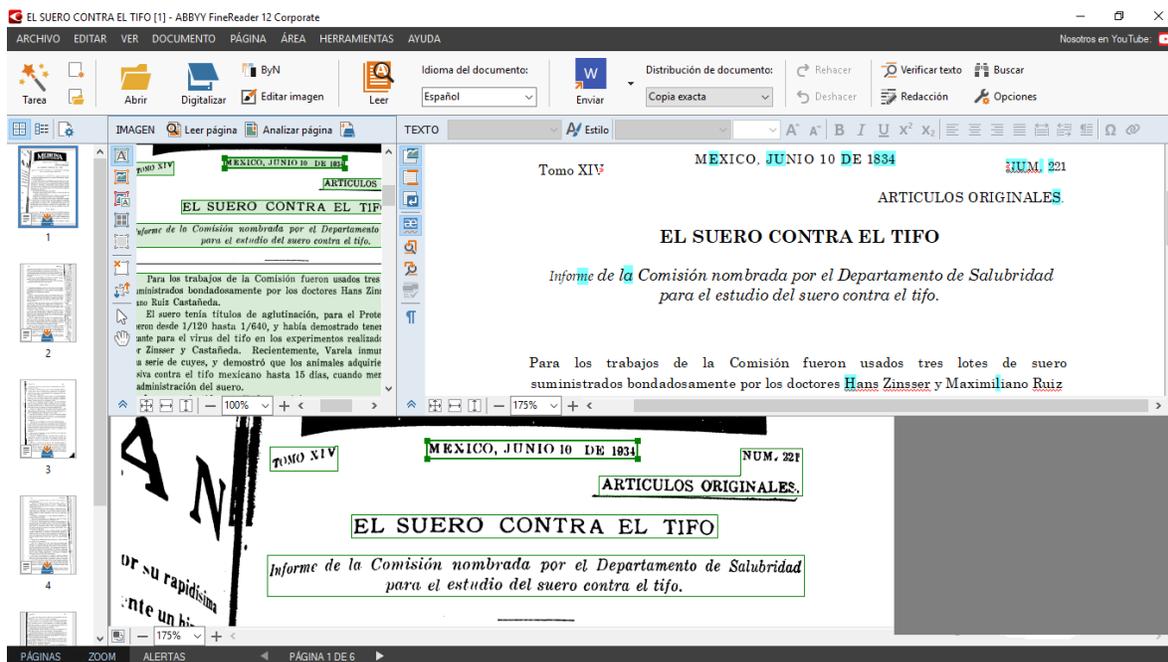


Figura 17. Limpieza de los textos

Fuente: Elaborada por los autores

Después del procedimiento de limpieza de los textos, se procedió a guardar cada uno en formato *txt*, para poder agregarlo a una nueva columna del archivo en Excel, donde el texto correspondiera al artículo en cuestión, así obtuvimos un archivo más completo, como se muestra en la Figura 18.

Autores	TITULO	REVISTA	TITULO ABRE REV-ISO	ISSN	ANO	VOL	Numer	Pag-1	Pag-2	No-P	IDIOMA	CONTENIDO	
ANONIMO	MEXICO SUM	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1906	21	29	853	853	1	Inglés	MEXICO.Summary of typhus fever in City of	
ANONIMO	PRECAUTION	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1906	21	6	131	131	1	Inglés	MEXICO.Precautions to be taken against sp	
FRICK; EBERH	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1906	21	11	259	260	2	Inglés	MEXICO.Report from City of Mexico—Mea:	
ANONIMO	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1906	21	16	393	393	1	Inglés	Report from Veracruz—Inspection of vessel	
GOTTSCHALK	MEXICO REP	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1907	22	35	1222	1222	1	Inglés	MEXICO.Report from City of Mexico—Stati:	
GOTTSCHALK	MEXICO REP	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1907	22	39	1373	1373	1	Inglés	MEXICO.Report from City of Mexico—Stati:	
GOTTSCHALK	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1907	22	21	691	691	1	Inglés	MEXICO.Report from City of Mexico— Com	
GOTTSCHALK	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1908	23	14	429	429	1	Inglés	Report from City of Mexico—Comparative :	
ANONIMO	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1908	23	46	1660	1661	2	Inglés	MEXICO.Report from Mexico City—Further	
ANONIMO	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1908	23	47	1702	1702	1	Inglés	MEXICO.Report from Mexico City—Outbre:	
ANONIMO	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1908	23	48	1740	1740	1	Inglés	MEXICO.Report from Mexico City—Status	
VARELA, GEF	TYPHUS FEVE	AMERICAN J	AM J TROP IV	Am. J. Trop. I	0096-6746	1908	29	1	63	69	7	Inglés	TYPHUS FEVER IN MEXICO: A STUDY OF EPI
ANDERSON J	A NOTE ON T	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1909	24	52	1941	1942	2	Inglés	Association of Schools of Public HealthA Not	
GOLDBERGE	ON THE RELA	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1909	24	50	1861	1862	2	Inglés	On the Relation of Rocky Mountain Spotted	
CARTER	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1909	24	28	998	998	1	Inglés	Report from Veracruz—Inspection and fum	
CARTER	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1909	24	19	615	615	1	Inglés	Report from Veracruz—Smallpox along rail	
CARTER	REPORT FRO	PUBLIC HEAL	PUBLIC HEAL	Public Health 0033-3549	1909	24	20	659	659	1	Inglés	Report from Veracruz—Smallpox—Antimo:	
WILDER, RUS	FURTHER INV	JOURNAL OF	J AMER MED	J. Am. Med. J	0002-9955	1910	55	V	309	311	3	Inglés	FURTHER INVESTIGATIONS REGARDING TH
MC CAMPBEI	OBSERVATIO	JOURNAL OF	J MED RES	J. Med. Res.	0097-3599	1910	23	1	71	83	13	Inglés	OBSERVATIONS ON TYPHUS EXANTHEMATI
ANDERSON, ,	ON THE ETIO	JOURNAL OF	J MED RES	J. Med. Res.	0097-3599	1910	22	3	469	481	13	Inglés	ON THE ETIOLOGY OF TABARDILLO OR ME:
RICKETTS-TA	THE ETIOLOC	JOURNAL OF	J AMER MED	J. Am. Med. J	0002-9955	1910	54	17	1373	1375	3	Inglés	THE ETIOLOGY OF THE TYPHUS FEVER (TAE
WILDER, RUS	THE RELATIC	ARCHIVES OF	ARCH INTERI	Arch. Intern.	0003-9926	1910	5	4	361	370	10	Inglés	THE RELATION OF TYPHUS FEVER (TABARD
WILDER, RUS	THE TYPHUS	JOURNAL OF	J AMER MED	J. Am. Med. J	0002-9955	1910	54	V	463	467	5	Inglés	THE TYPHUS FEVER OF MEXICO (TABARDIL

Figura 18. Registros bibliográficos correspondientes al Corpus tifo

Fuente: Elaborada por los autores

En la Figura 18 se presenta una parte del archivo que se generó a partir de la agregación del texto completo en la columna CONTENIDO, de cada uno de los artículos científicos. Posteriormente, mediante Python se procedió a clasificar la columna *CONTENIDO* considerando el idioma (español e inglés), con la finalidad de generar dos archivos en formato txt que sirvieron como datos de entrada para realizar el procesamiento del lenguaje natural.

De los 239 registros bibliográficos que se obtuvieron originalmente, se recuperaron 128 en texto completo, de los cuales, 90 artículos se encuentran en idioma inglés y 38 en español. Esto corresponde a poco más de la mitad (53%) del total de registros identificados originalmente.

3.1.4 Selección de herramientas de análisis apropiadas

Las herramientas específicas del NLP que se requirieron para el tratamiento del corpus fueron: el reconocimiento de entidades nombradas, la frecuencia de palabras, la implementación del algoritmo TF-IDF y la generación de redes semánticas o grafos, también conocida como Análisis de Redes Sociales, realizados principalmente por la biblioteca spaCy en Python.

Este análisis identificó las palabras más relevantes del corpus de información del tifo y las categorizó automáticamente dependiendo de su similitud y contexto, tal como se muestra en el paso siguiente.

3.1.5 Identificación de patrones a través de la minería de datos

Las categorías que se tomaron en cuenta para el análisis del discurso por medio del procesamiento de lenguaje natural se presentan a continuación:

Para Valencia Martínez (2018) las categorías analizadas se refieren a:

- *Objetos de estudio:* Se incluyen todos aquellos objetos que son sometidos a un proceso de análisis e investigación, con la finalidad de arrojar resultados de un cierto proceso científico.
- *Procedimientos:* Se identificarán los métodos, técnicas y prácticas aplicadas al conocimiento de los objetos de estudio.
- *Instrumentos:* Se comprenden las herramientas y utensilios que son utilizados como recursos de apoyo en el análisis de los objetos de estudio sometidos a un procedimiento físico, químico o metalúrgico.
- *Lugares:* Se identificarán aquellas zonas geográficas en donde se realizaron procedimientos científicos con el objetivo de permitir el desarrollo de indicadores de espacialidad en el tema del tifo en México y la influencia de la geografía local en el desarrollo de conocimientos y resultados orientados a la erradicación de la enfermedad.
- *Instituciones:* Se comprenden aquellas instituciones, organizaciones, centros de investigación que realicen actividades académicas, de investigación o cualquier otra actividad con o sin fines de lucro. Para motivos de esta investigación, se tomaron principalmente instituciones de adscripción, formación e investigación.

A través del NLP se implementó el reconocimiento de entidades nombradas para la identificación y extracción de las palabras pertenecientes a las categorías anteriores, la frecuencia de palabras y el algoritmo TF-IDF para la medición de la representatividad de los términos en ese dominio de conocimiento y las redes semánticas para la modelación léxica de la relación entre tipos distintos de entidades.

Para la identificación de las entidades nombradas que correspondieron a las categorías antes mencionadas, se generó un modelo por medio de la biblioteca spaCy en el lenguaje de programación Python, que permitió incorporar datos de entrenamiento que generen estadísticas de predicción para el etiquetado del discurso y su clasificación como entidad nombrada.

Como se muestra en la Figura 19, los datos de entrenamiento corresponden a la entrada de texto y la asignación de etiquetas en el discurso que identifiquen a las entidades nombradas. Con base en esa información, el modelo generó una “predicción en forma de un gradiente de error de la función de pérdida, que calcula la diferencia entre el ejemplo de entrenamiento y el resultado esperado” (Spacy, 2020b).

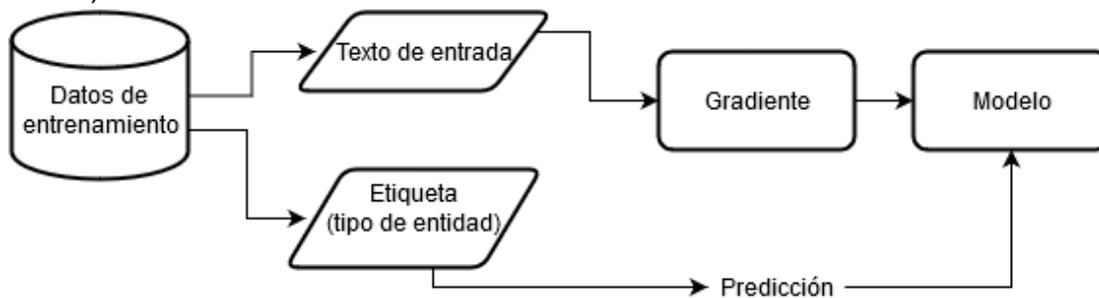


Figura 19. Modelo de reconocimiento de entidades nombradas

Fuente: (Spacy, 2020b)

Con base al modelo anterior y después de un filtrado se extrajeron las palabras en forma de entidades nombradas que correspondieron a cada una de las categorías antes mencionadas, las cuales se analizaron en el punto siguiente.

3.1.6 Análisis de los resultados mediante la metodología methontology

En esta parte se aplicó la metodología methontology adoptando dos enfoques principales i) *top-down* (definición de los conceptos más generales del dominio) y ii) *bottom-up* (definición de las clases más específicas del dominio) (Marciszack, Pérez Cota, Antonelli, & Cárdenas, 2009) para poder analizar los resultados y adecuarlos a las tareas que propone para la creación de la ontología. La terminología que contemplan las tareas de la metodología corresponde al idioma español e inglés.

➤ Tarea 1: Construcción del glosario

Se construyó un glosario de términos, que incluye todas las palabras relevantes del dominio del tifo en México (conceptos, instancias, atributos y relaciones entre conceptos), sus descripciones en lenguaje natural, y sus sinónimos (Corcho et al., 2005). Para la construcción del glosario se tomaron en cuenta las cinco categorías de análisis: *objeto de estudio* (Tabla 3), *procedimientos* (Tabla 4), *instrumentos* (Tabla 5), *lugares* (Tabla 6), *instituciones de adscripción, formación e investigación* (Tabla 7). Cada tabla está constituida por los campos de nombre, variante, descripción y tipología. Las descripciones de las palabras fueron tomadas de la base de datos *Word Net* mediante Python. Las palabras que se encuentran en la sección de *Nombre* se eligieron con base al término frecuente y con el TF-IDF. A continuación, se muestra una parte del glosario de cada una de las categorías.

Tabla 3. Glosario de objetos de estudios

Objetos de estudio			
Nombre	Variante	Descripción	Tipología
rickettsias	rickettsia, ricketssias, rickettsiae	Grupo de bacterias muy pequeñas...	concepto
cuyes	cuy, cuyo, cuya, cuyas, cobaya, cobayo, cobayos	Roedor experimental para pruebas sobre la inmunidad del tifo	concepto
ratas	rata, rats, rat	Roedores de cola larga similares a, pero más grandes que un ratón	conceptos
piojos	piojo, lice, louse	Insecto chupa sangre sin alas...	conceptos

Fuente: Elaborada por los autores

Tabla 4. Glosario de procedimientos

Procedimientos			
Nombre	Variante	Descripción	Tipología
felix reaction	reacción de Weil-Félix, método de Weigl, Weigl method	Método aglutinación de rickettsias	concepto
método de Ruiz Castañeda	método de Castañeda, Castaneda's method, técnica Ruiz Castañeda	Método de cultivo de rickettsias	concepto
método de Giemsa	Giemsa solution	Método para el descubrimiento de presencia de rickettsias	concepto

Fuente: Elaborada por los autores

Tabla 5. Glosario de instrumentos

Instrumentos			
Nombre	Variante	Descripción	Tipología
microscopio	microscope, microscopic	De o relacionado con o utilizado en microscopía	concepto
papel lustre	--	Es un papel que viene en una gran variedad de tonos vivos...	concepto
serum	sueros, serums	Líquido amarillento y acuoso, rico en proteínas...	concepto

Fuente: Elaborada por los autores

Tabla 6. Glosario de lugares

Lugares			
Nombre	Variante	Descripción	Tipología
México	Mexico...	Una República en el sur de América del Norte...	instancia
Oaxaca	--	Es uno de los treinta y un estados que, junto con Ciudad de México, forman los Estados Unidos Mexicanos	concepto
La Laguna	--	Es una región metropolitana perteneciente a los Estados de Durango y Coahuila	atributo
Tlalpujahuá	--	Es un municipio del Estado de Michoacán	atributo

Fuente: Elaborada por los autores

Tabla 7. Glosario de instituciones

Instituciones			
Nombre	Variante	Descripción	Tipología
Instituto de Enfermedades Tropicales	Institute of Health and Tropical Diseases...	Es uno de los principales institutos locales de capacitación e investigación en medicina...	concepto
Manicomio General	--	Fue un hospital psiquiátrico donde se catalogaba a los pacientes enfermos según su sexo...	concepto
Hospital General	General Hospital	Es una institución de asistencia pública perteneciente a la Secretaría de Salud de México...	concepto

Fuente: Elaborada por los autores

➤ **Tarea 2: Construcción de taxonomías de conceptos**

Corcho et al. (2005) considera que, para la construcción de una taxonomía, se deben seleccionar del glosario aquellos términos que son conceptos, también menciona que methontology propone la utilización de las siguientes relaciones taxonómicas:

1. *Subclase-de*: Un concepto C1 es *subclase-de* otro concepto C2 si y solo si todas las instancias de C1 son también instancias de C2.
2. *Descomposición-disjunta*: Una *descomposición-disjunta* de un concepto C es un conjunto de subconceptos de C que no tienen instancias comunes y que no cubren C, es decir, puede haber instancias del concepto C que no son instancias de ninguno de los conceptos que forman la descomposición.
3. *Descomposición-Exhaustiva*: Una *Descomposición-Exhaustiva* de un concepto C es un conjunto de subconceptos de C que lo cubren, es decir, tal que no existe ninguna instancia de C que no sea instancia de al menos uno de los conceptos de la descomposición. Los conceptos que pertenecen a este conjunto pueden tener instancias y subconceptos comunes. En la Figura 20 se muestra un ejemplo de la taxonomía perteneciente a la categoría objeto de estudio.

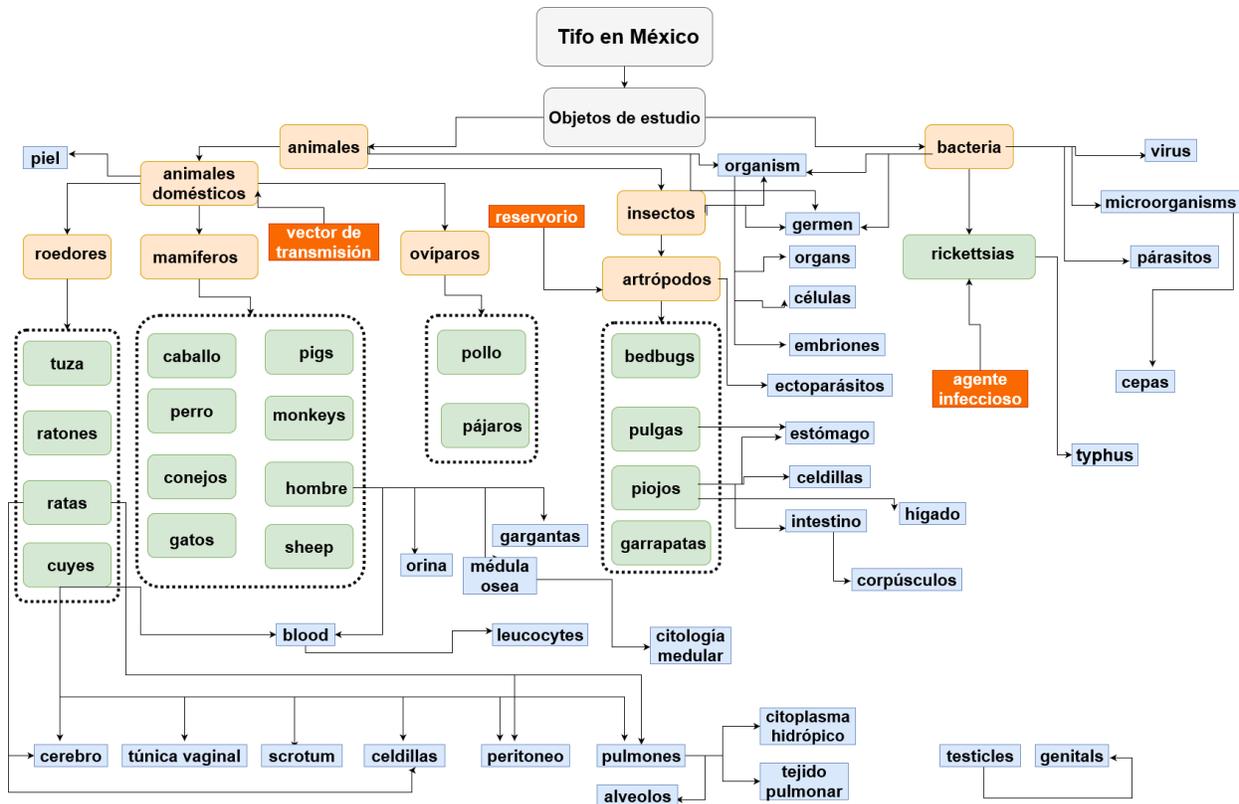


Figura 20. Taxonomía de conceptos

Fuente: Elaborada por los autores

➤ **Tarea 3: Construcción de diagramas de relaciones binarias**

“El objetivo de este diagrama es establecer las relaciones existentes entre conceptos de esta o de distintas taxonomías de conceptos” (Corcho et al., 2005). Cabe mencionar, que la tarea 3 y la tarea 5 son similares, por lo cual ambas tareas se realizaron conjuntamente. En la Figura 21 se presenta un ejemplo de la tarea 3 y la tabla 8 hace referencia a la tarea 5.

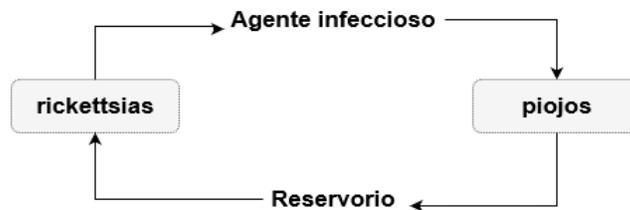


Figura 21. Relaciones binarias

Fuente: Elaborada por los autores

Tabla 8. Descripción de las relaciones binarias

Nombre de la relación	Concepto de origen	Concepto destino	Relación inversa
agente infeccioso	rickettsias	piojos	reservorio
reservorio	piojos	rickettsias	agente infeccioso

Fuente: Elaborada por los autores

➤ **Tarea 4: Construcción del diccionario de conceptos**

El diccionario contiene todos los conceptos del dominio, sus relaciones, sus instancias, y sus atributos de clase y de instancia. Las relaciones especificadas para cada concepto son aquéllas en las que el concepto es el origen de esta (Corcho et al., 2005). En la tabla 9, se muestra un fragmento del diccionario de conceptos.

Tabla 9. Diccionario de conceptos

Nombre del concepto	Instancias	Atributos de clase	Atributos de instancias	Relaciones
cuyes	roedores	peritoneo, cerebro...	--	vector de transmisión
piojos	artrópodos	intestino, celdillas...	ectoparásito	reservorio
rickettsias	bacterias	typhus	virus, microorganismo..	agente infeccioso

Fuente: Elaborada por los autores

➤ **Tarea 6: Describir atributos de instancia**

“El objetivo de esta tarea es describir en detalle todos los atributos de instancia incluidos en el diccionario de conceptos” (Corcho et al., 2005). En la tabla 10, se muestra una parte de la descripción de instancias.

Tabla 10. Descripción de atributos de instancias

Nombre del atributo	Instancias	Tipo de valor	Rango	cardinalidad
ectoparásitos	artrópodos	string	ilimitado	1
virus	bacteria	string	ilimitado	1

Fuente: Elaborada por los autores

La sección *tipo de valor* hace referencia al tipo de dato del atributo; el cual puede ser *string* (cadena de caracteres), *int* (números enteros), *float* (números decimales), etc. El rango se refiere a la cantidad de información que va a corresponder cada atributo, y la cardinalidad es el número de elementos que contiene un conjunto determinado, en este caso, cuantas palabras o tokens conforman el *string*.

➤ **Tarea 7: Describir atributos de clase (concepto)**

El objetivo de esta tarea es describir en detalle todos los atributos de clase incluidos en el diccionario de conceptos. Un ejemplo de esta tarea se muestra en la tabla 11.

Tabla 11. Descripción de atributos de clase

Nombre del atributo	Concepto	Tipo de valor	Rango	cardinalidad
celdillas	ratas, cuyes...	string	ilimitado	1
cerebro	ratas, cuyes	string	ilimitado	1

Fuente: Elaborada por los autores

➤ **Tarea 8: Describir constantes**

“El objetivo de esta tarea es describir en detalle cada una de las constantes identificadas en el glosario de términos” (Corcho et al., 2005). Las constantes especifican información relacionada con el dominio del conocimiento, siempre toman el mismo valor y normalmente se usan en forma numérica (Gómez-Pérez et. al., 2006), por lo cual, esta tarea no aplica para el dominio del tifo en México.

➤ **Tarea 9: Definir axiomas formales**

Los axiomas formales son expresiones lógicas que siempre son verdaderas y se usan como restricciones del modelo, así como las características que debe cumplir algún concepto, atributo, relación o instancias de relación (Cruz Colohua, Reyes Hernández, Chan Hernández, & Sánchez-Cervantes, 2016; Gómez-Pérez et. al., 2006). En la tabla 12, se muestra una parte de la definición de los axiomas formales.

Tabla 12. Definición de axiomas formales

Nombre del axioma	Descripción	Expresión	Concepto	Relaciones	Variab es
Los artrópodos son objetos de estudio	Los artrópodos se sometían a un proceso de ...	$\forall w (w \rightarrow o)$	garrapatas piojos pulgas bedbugs	reservorio reservorio reservorio reservorio	w w w w
Los ovíparos son objetos de estudio	Los ovíparos se sometían a un proceso de ...	$\forall y (y \rightarrow o)$	pollo pájaros	vector de... vector de...	y y

Fuente: Elaborada por los autores

Constantes: objeto de estudio= o, procedimientos= p, instrumentos= i, lugares= l, instituciones=i2

➤ **Tarea 10: Definir reglas**

De manera similar a la tarea previa, en esta parte se deben identificar qué reglas se necesitan en la ontología, y posteriormente describirlas. Para cada regla, methontology propone incluir la siguiente información: nombre, descripción en lenguaje natural, expresión que describe formalmente la regla, y conceptos (Corcho et al., 2005). Sin embargo, para cuestiones de esta investigación y con base a las variables de los axiomas formales solo se tomarán en cuenta: nombre, descripción, expresión y variables, con la finalidad de que ambos tengan una relación.

La metodología propone especificar las expresiones de las reglas utilizando el formato *si <condiciones> entonces <consecuente>*. (Corcho et al., 2005). En la tabla 13, se muestra una parte de la definición de reglas.

Tabla 13. Definición de reglas

Nombre de la regla	Descripción	Expresión	Variable
Un objeto de estudio no es un procedimiento	Si para todo objeto de estudio, entonces no debe ser	$\forall o \rightarrow \neg p, i, l, i2$	w x

Un objeto de estudio no es un instrumento	procedimiento, instrumento, lugar o institución		y z u
Un procedimiento no es un objeto de estudio	Si para todo procedimiento, entonces no debe ser objeto de estudio, instrumento, lugar o institución	$\forall p \rightarrow \neg o, i, l, i2$	w x y
Un procedimiento no es un instrumento			

Fuente: Elaborada por los autores

Constantes: objeto de estudio= o, procedimientos= p, instrumentos= i, lugares= l, instituciones=i2

➤ Tarea 11: Definir instancias

Una vez que el modelo conceptual de la ontología ha sido desarrollado, se pueden definir las instancias que aparecen en el diccionario de conceptos. Para cada instancia se define: su nombre, el nombre del concepto al que pertenece y los valores de sus atributos de instancia, si se conocen (Corcho et al., 2005). En la tabla 14, se mostrará un fragmento de la tabla de la definición de las instancias.

Tabla 14. Definición de instancias

Nombre de la instancia	Concepto	Atributo	Valor
animales	--	--	string
bacterias	rickettsias	organism, germen...	string
animales domésticos	--	--	string

Fuente: Elaborada por los autores

3.2 Codificación en Protégé

En esta parte se procedió a codificar las tareas que conforman la metodología methontology con la utilización del software Protégé, con la finalidad de establecer relaciones semánticas entre los términos y elaborar una ontología como sistema de representación del conocimiento de la enfermedad del tifo en México. En la Figura

22 se muestran las cinco categorías antes mencionadas, las cuales, por medio de estas se desglosan todos los términos analizados de este dominio de conocimiento.

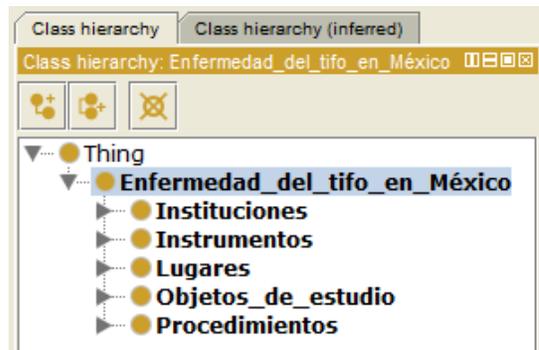


Figura 22. Categorías principales en Protégé

Fuente: Elaborada por los autores

La tarea 1, corresponde a la construcción del glosario de términos. Se elaboró en la pestaña *Entities*, en el apartado *Anotation* se agregó la descripción del término (*isDefinedBy*) así como sus variantes (*seeAlso*), tal y como se muestra en la Figura 23.

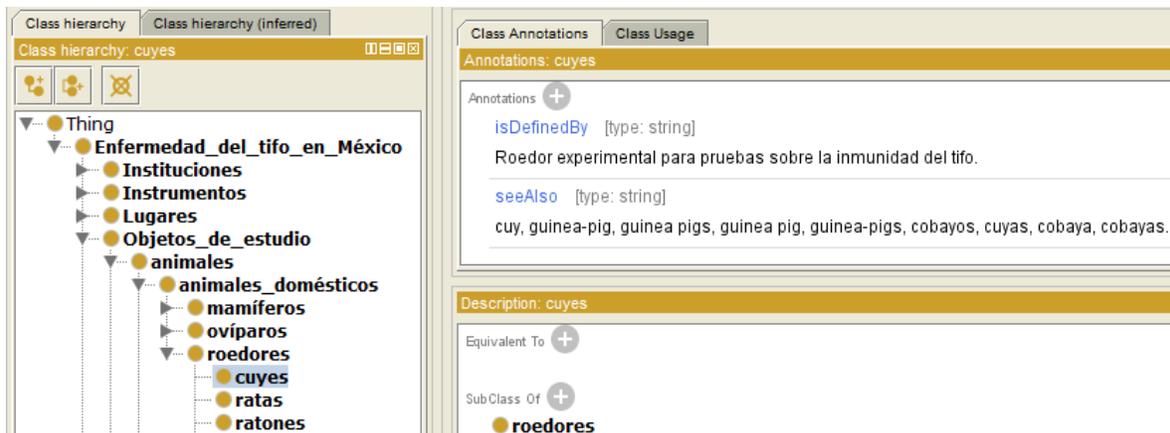


Figura 23. Glosario de términos en Protégé

Fuente: Elaborada por los autores

La tarea 3 se visualiza en la Figura 24. Se creó en el panel *Object properties*, la opción *inverse of* hace referencia a la relación inversa entre las clases, *Domains (intersection)* establece el concepto principal donde se estableció la relación y *Ranges (intersection)* es entendida como el concepto receptivo de tal relación. La

casilla *Symmetric* significa que el concepto “x” está relacionado con el concepto “y” y viceversa.

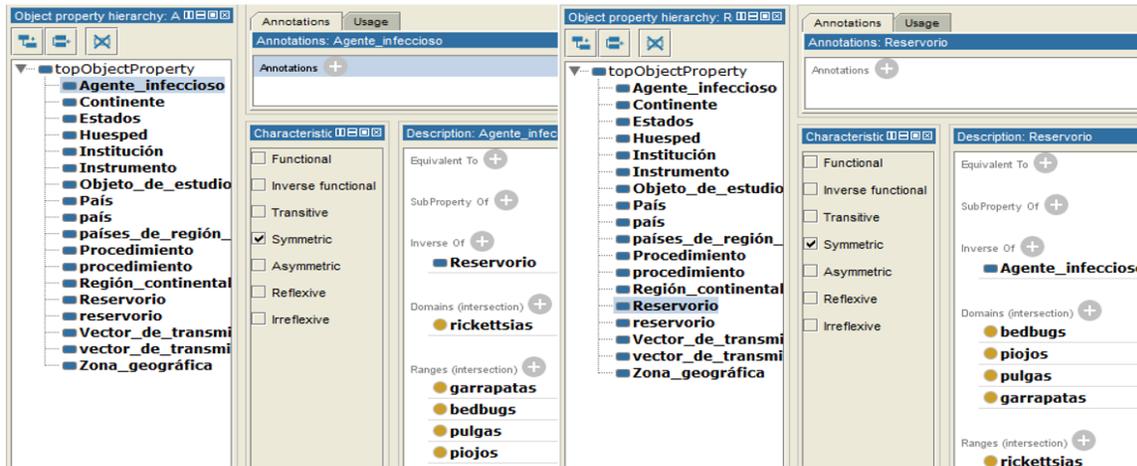


Figura 24. Relaciones binarias en Protégé

Fuente: Elaborada por los autores

La tarea 4, 6, 7 y 11, mismas que se muestran en las Figuras 25 y 26 se desarrollaron en la sección *Data properties* donde se codificaron los atributos de clases e instancias, de igual manera, en la sección *Individuals* se describieron las instancias estableciendo las relaciones con los atributos. Como en el caso de la tarea 1, se utilizó el apartado *Anotation* para agregar su descripción y sinónimos.

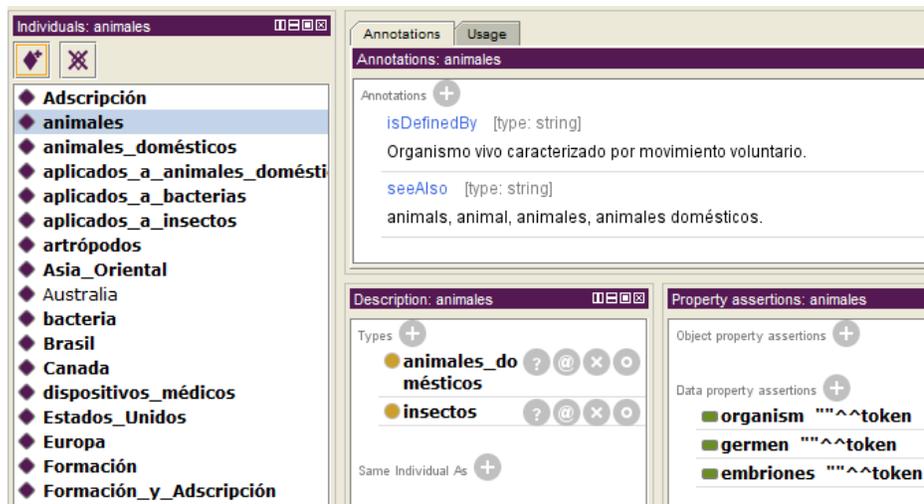


Figura 25. Descripción de instancias en Protégé

Fuente: Elaborada por los autores

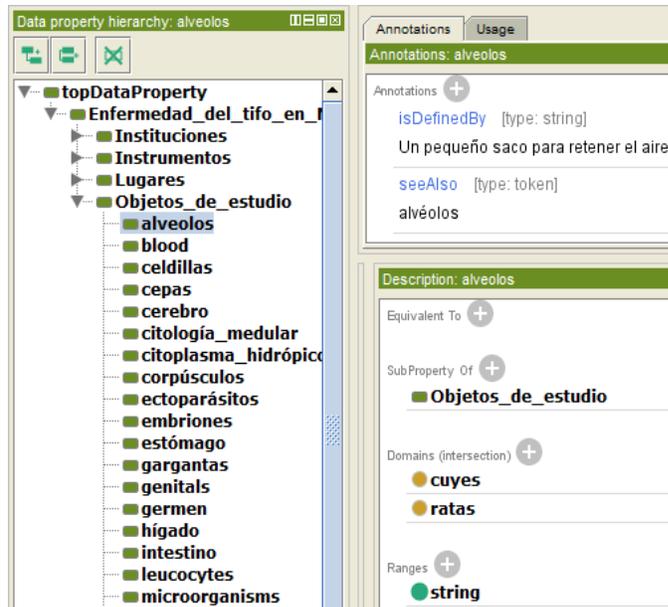


Figura 26. Atributos de conceptos en Protégé

Fuente: Elaborada por los autores

Los axiomas descritos en la tarea 9 están implícitos en la jerarquía de la construcción del diccionario de conceptos, al igual que la tarea 10, la cual está orientada a la conformación de reglas o restricciones de las clases.

La codificación de las reglas se visualiza en la Figura 27 y se implementó la palabra clave *not* que hace referencia a la delimitación de las propiedades de las clases correspondientes.



Figura 27. Codificación de reglas en Protégé

Fuente: Elaborada por los autores

A partir de los datos anteriores, se generó un grafo a través de Protégé, mismo que se muestra en la Figura 28, donde se visualiza el vocabulario empleado organizado en clases y subclases correspondientes a las categorías *Lugares*, *Procedimientos*, *Instituciones*, *Instrumentos* y *Objetos de estudio*.

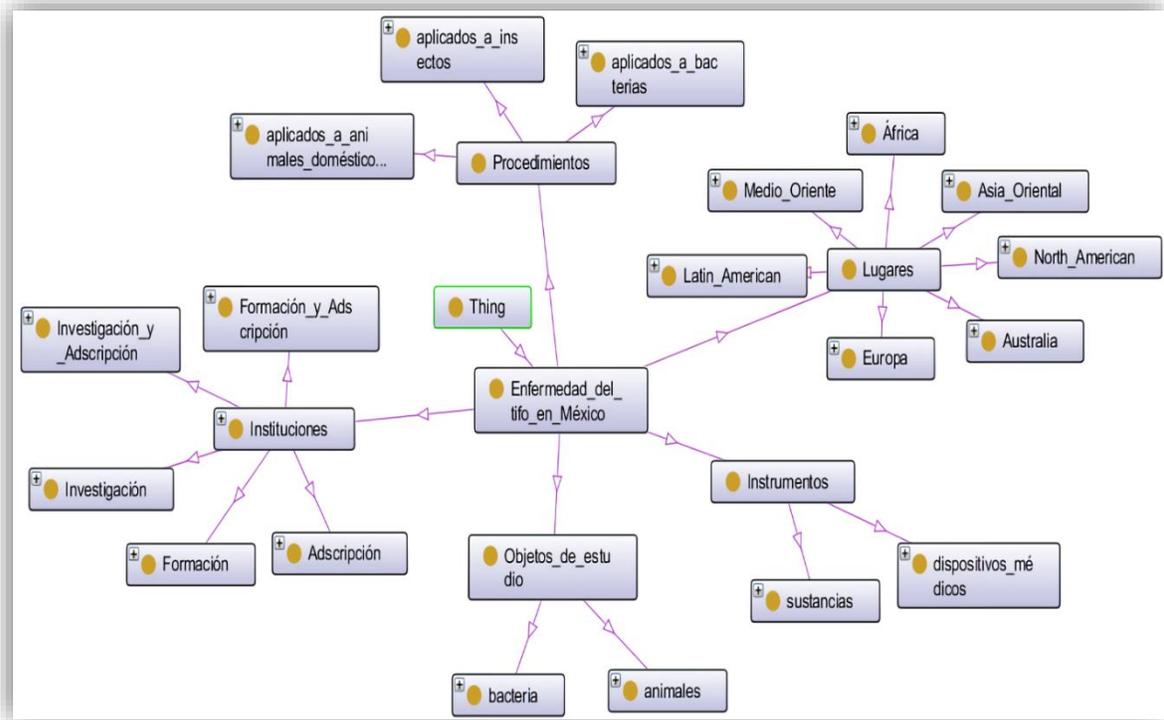


Figura 28. Grafo en Protégé

Fuente: Elaborada por los autores

Para visualizar amplia y detalladamente las tareas pertenecientes a la metodología methontology y descargar la ontología completa, favor de escanear el código QR de la Figura 29 o entrar en el siguiente enlace: https://github.com/LuisPoloBautista/Ontologia_Tifo_Mexico



Figura 29. Código QR para visualizar las tareas de methontology

Fuente: Elaborado por los autores

En la metodología de este trabajo se utilizó la bibliominería para implementar varios procesos computacionales: el procesamiento de lenguaje natural para el análisis del corpus de información y la aplicación de la metodología methontology para la elaboración de la ontología de la enfermedad epidémica del tifo.

Estas actividades fueron modeladas en el diagrama que se encuentra en la Figura 12, mismas que se especifican a lo largo de este capítulo 3.

En el siguiente capítulo se presentarán los resultados obtenidos del análisis bibliométrico del corpus de información, el procesamiento de lenguaje natural aplicado a las categorías *Lugares*, *Procedimientos*, *Instituciones*, *Instrumentos* y *Objetos de estudio* y la utilización de la ontología del tifo como sistema de representación del conocimiento.

Referencias

- Atlas Histórico de la Ciencia Mexicana. (2017). Recuperado el 27 de febrero de 2020, de <http://bibliometria.bfm.cinvestav.mx/inicio/index.html>
- Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Construcción de ontologías legales con la metodología methontology y la herramienta WebODE. *Web Semántica y ontologías jurídicas. Aplicaciones para el derecho en la nueva generación de la red*, 1–17. Recuperado de <http://oa.upm.es/5289/>
- Cruz Colohua, I., Reyes Hernández, L. ., Chan Hernández, G., & Sánchez-Cervantes, J. . (2016). Diseño de una ontología para el proceso de evaluación de las asignaturas técnico-científicas del Instituto Tecnológico de Orizaba. *Research in Computing Science*, 128, 77–89. Recuperado de [https://www.rcs.cic.ipn.mx/2016_128/Diseno de una ontologia para el proceso de evaluacion de las asignaturas tecnico-cientificas.pdf](https://www.rcs.cic.ipn.mx/2016_128/Diseno%20de%20una%20ontologia%20para%20el%20proceso%20de%20evaluacion%20de%20las%20asignaturas%20tecnico-cientificas.pdf)
- Hjørland, B., & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*, 46(6), 400–425. Recuperado de https://www.academia.edu/1304539/Toward_a_new_horizon_in_information_science_domain-analysis
- Marciszack, M., Pérez Cota, M., Antonelli, L., & Cardenas, M. (2009). *Construcción de una ontología utilizando Protégé para la elicitación de requerimientos*. 8. Recuperado de [http://www.institucional.frc.utn.edu.ar/sistemas/Noticias/INV/Seminarios/ontologias_cnit2009\].pdf?fbclid=IwAR05tcS3uBzGm5tsSaFIFSrQ8Aaq-2M4uP0jkip2XABc_7cdB0UabLwpVs](http://www.institucional.frc.utn.edu.ar/sistemas/Noticias/INV/Seminarios/ontologias_cnit2009].pdf?fbclid=IwAR05tcS3uBzGm5tsSaFIFSrQ8Aaq-2M4uP0jkip2XABc_7cdB0UabLwpVs)
- SpaCy. (2020b). *Training spaCy's Statistical Models*. Recuperado el 8 de septiembre de 2020, de <https://spacy.io/usage/training>
- Valencia Martínez, A. (2018). *Modelo ontológico de información para estudiar las primeras formas de comunicación científica moderna en el tema de mineralogía en México, 1795-1849*. (Tesis de licenciatura, Escuela Nacional de Biblioteconomía y Archivonomía), Ciudad de México.
- Vilches-Blázquez, L. M., Comesaña, D., & Arrieta Moreno, L. De J. (2020). Construcción de una red de ontologías sobre eventos meteorológicos a partir de periódicos históricos. *Transinformação*, 32. <https://doi.org/10.1590/1678-9865202032e180077>

Capítulo 4. Sistema de representación del conocimiento de la enfermedad del tifo en México

En este capítulo se muestran los resultados obtenidos referentes al análisis bibliométrico del corpus de información, el procesamiento de lenguaje natural y aspectos específicos de la utilización y enfoque de la ontología como sistema de representación y organización del conocimiento sobre el tifo en México en el periodo 1904-1977.

4.1 Análisis bibliométrico del corpus textual

El análisis bibliométrico del corpus de información se realizó mediante el software bibliometrix. Esta herramienta genera estadísticas con base a los modelos matemáticos mencionados en el capítulo 2. Los datos generales de la producción científica del tifo se muestran en la tabla 16.

Tabla 3. Estadística general del corpus textual

Descripción	Resultados
Documentos	239
Fuentes de publicación (revistas, libros, etc.)	23
Periodo	1904-1977
Autores	127
<i>Idioma</i>	
Español	129
Inglés	108
Frances	2
<i>Tipos de documentos</i>	
Artículos	230
Cartas	9

Fuente: Elaborada por los autores

La Figura 30 muestra los 20 autores más relevantes relacionados con la producción de conocimiento en torno al tema del tifo. Como se puede observar, más de la mitad de los autores escribieron 5 artículos, identificando casos con menor proporción.

Los autores con más producción científica son Ruiz-Castaneda, Varela G, Zinsser H, Silva-Goytia, Mooser H, López-Vallejo y Parada-Gaym. Cabe mencionar que hubo una cantidad considerable de artículos que no se logró identificar a un autor, por lo cual se consideraron como publicaciones anónimas.

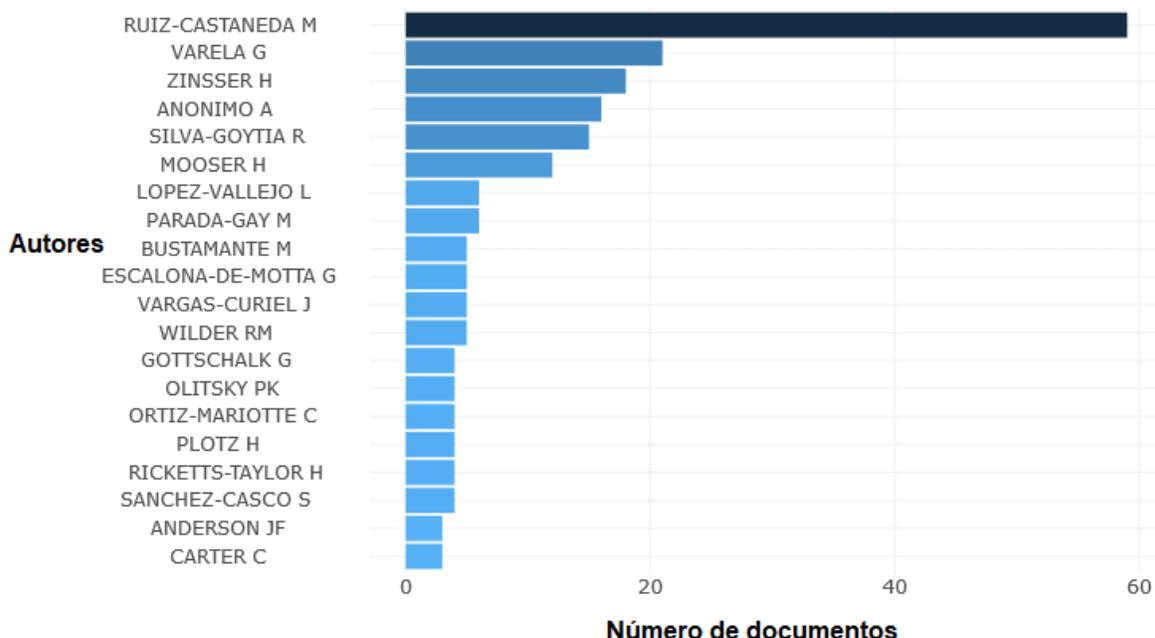


Figura 30. Autores más productivos en el tema del tifo 1904-1977

Fuente: Elaborada por los autores

En la Figura 31 se presentan las 20 revistas más productivas respecto a la cantidad de trabajos publicados. Como se puede observar, solamente tres revistas publicaron más de 20 artículos cada una, siendo la *Revista Mexicana de Medicina* la que encabeza la lista con 75 artículos publicados, y aproximadamente más de la mitad de las revistas solamente publicaron 5 artículos cada una.

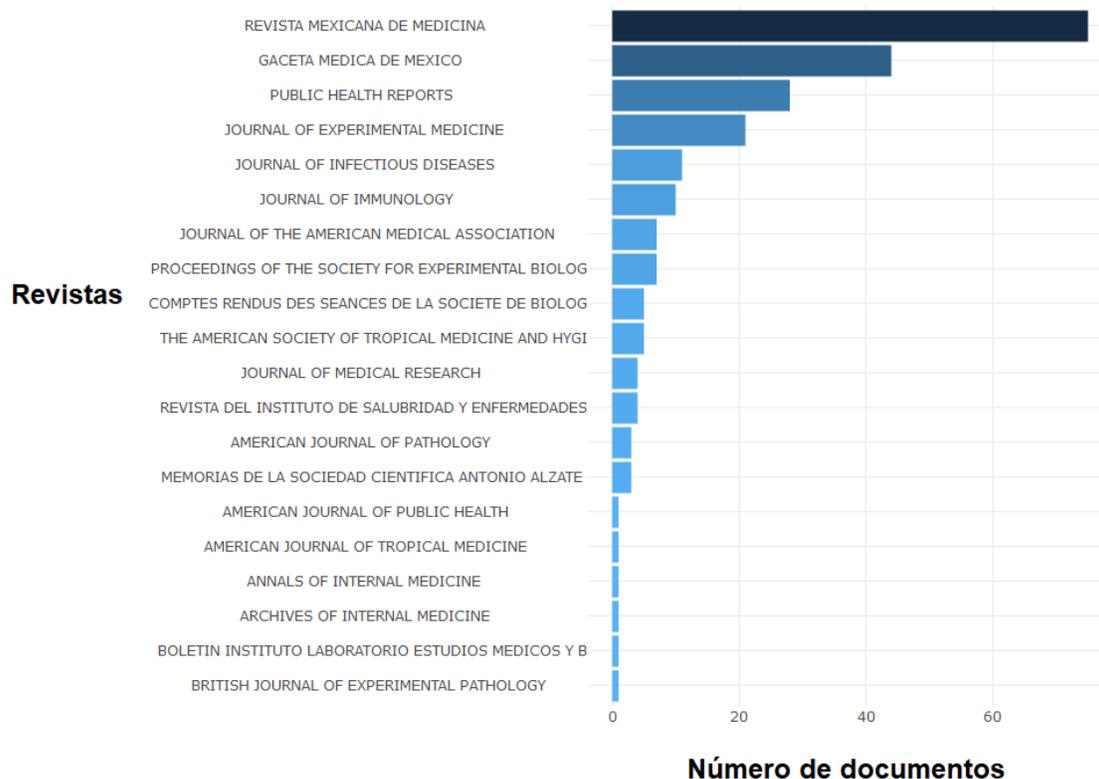


Figura 31. Revistas más productivas en el tema del tifo 1904-1977

Fuente: Elaborada por los autores

Con base en esta información, podemos mencionar que los autores que publicaron entre 20 y 60 artículos lo hicieron mediante las tres revistas más productivas, de las cuales, las dos primeras son editadas en México.

Con la finalidad de mostrar en contexto la colaboración entre los autores y el número de publicaciones, se tomó como base la teoría de grafos y el análisis de redes sociales que representara de una forma visual esa información.

La red que se muestra en la Figura 32, tomó en consideración a los 127 autores que forman parte de la colaboración científica en ese periodo, la cual puede inferir una conformación de comunidades que contribuyeron al desarrollo del conocimiento en torno al tema del tifo.

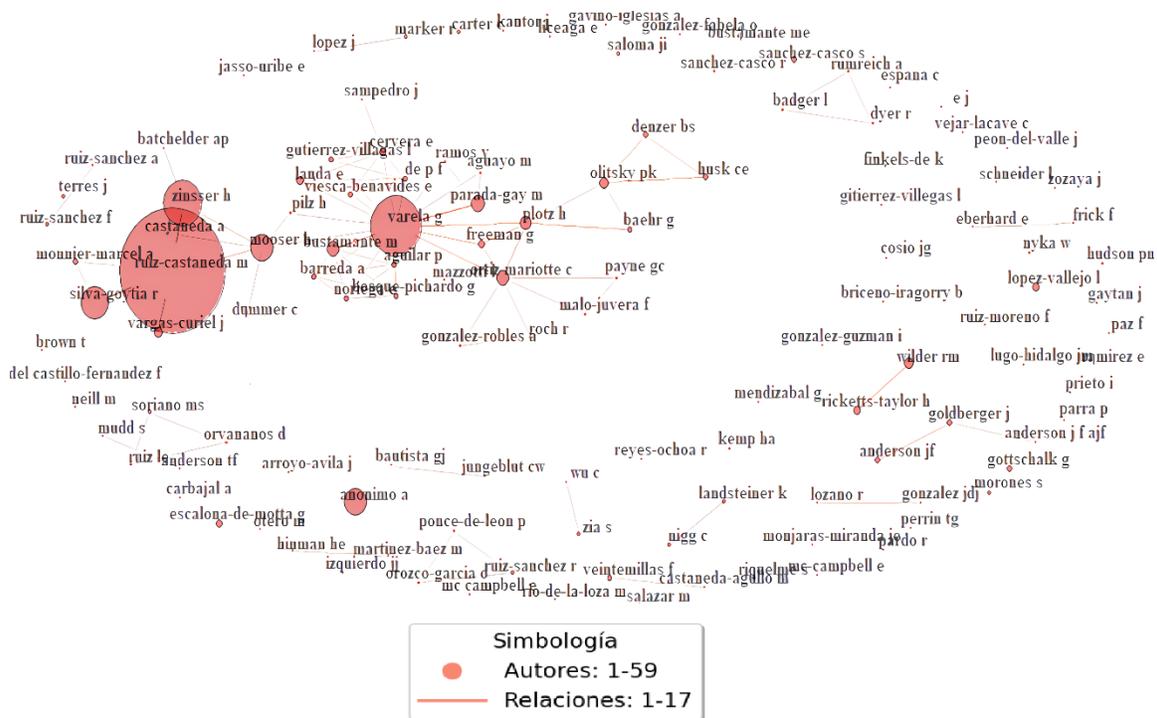


Figura 32. Red bibliométrica de relaciones de coautoría en el tema del tifo 1904-1977

Fuente: Elaborada por los autores

El tamaño del nodo corresponde a la cantidad de publicaciones de cada autor, mientras que el grosor de las líneas muestra la densidad de la colaboración entre autores. De este modo, tenemos que Ruiz Castañeda fue el autor que más publicó y por ende más prestigio tuvo dentro de la comunidad, al igual que Zinsser, Mooser, Varela y Silvia-Goytia.

Así mismo, podemos observar que una gran parte de los autores no colaboraron con nadie y tuvieron poca productividad científica, y se encuentran distribuidos en la periferia de la red, de igual forma, podemos mencionar que estos autores sólo publicaron un artículo científico en algunas de las revistas anteriormente mencionadas.

En términos generales, el análisis bibliométrico refleja de una manera cuantitativa los patrones que tuvo la actividad científica relacionada al tema del tifo en México, 1904 -1977, que influyeron en la erradicación de la enfermedad.

De igual manera se da a conocer a las revistas que fueron relevantes en ese dominio de conocimiento, y por medio de una red bibliométrica de coautoría se refleja aquellos autores que sobresalieron gracias a su aportación de conocimiento y colaboración científica.

4.2 Resultados de la bibliominería

Procesamiento de lenguaje natural

En este apartado se presentan los resultados obtenidos del procesamiento de lenguaje natural a través de la implementación de la bibliominería metodológica. El procesamiento de lenguaje natural fue una parte fundamental en el análisis del discurso de la literatura especializada en el tema del tifo en México y la elaboración de la ontología que es una forma de representar el dominio de conocimiento.

Como se mencionó en el capítulo anterior, la terminología especializada fue analizada por las categorías de análisis: *objetos de estudio, procedimientos, instrumentos, lugares e instituciones*.

Cada categoría se analizó por separado con el objetivo de identificar con mayor precisión las palabras relacionadas a ese contexto. Para esto se tomó como base la frecuencia de palabras, redes semánticas o grafos, reconocimiento de entidades nombradas y el algoritmo TF-IDF.

Cabe mencionar que todo el análisis que incluye el procesamiento de lenguaje natural fue desarrollado en Python utilizando principalmente la biblioteca spaCy, así como bibliotecas complementarias de análisis y visualización de datos como matplotlib, networkx, scikit-learn, etc. El número total de palabras identificadas en cada categoría se muestra en la Figura 33.

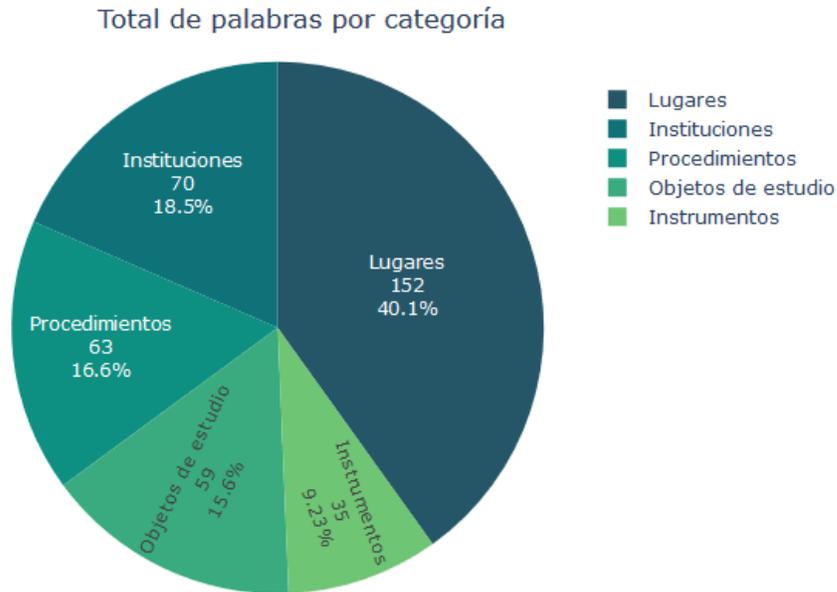


Figura 33. Total de palabras identificadas por categoría

Fuente: Elaborada por los autores

Como se puede observar, la categoría que contiene más palabras es *lugares* con 152, que corresponde a un 40.1% del total de términos identificados. Cabe destacar, que algunas palabras en inglés y en plural no fueron contabilizadas en este proceso, ya que en su mayoría se incluyeron en la lista de variantes de términos en las tareas que conforman la metodología methontology.

El porcentaje de palabras en idioma español tendrá un valor relevante en la elaboración de la ontología. De una forma general, se inició este análisis textual generando nubes de palabras como herramientas de visualización de datos, las cuales consisten en mostrar el tamaño y el énfasis visual de las palabras ponderadas por su frecuencia de aparición en el corpus de información y son útiles para representar gráficamente las palabras prominentes en un texto (Saif, Fernández, He & Alani, 2014).

Con base en esto, se presentan dos nubes de palabras elaboradas a través de Python, las cuales corresponden al corpus textual del tifo en español e inglés. Cabe

Figura 35 se presenta una tabla con la frecuencia de aparición de las primeras 20 palabras pertenecientes a la categoría *objetos de estudio*.

La frecuencia de términos sirve como una medida de ponderación de las palabras más relevantes referentes a los objetos que fueron utilizados en un proceso de análisis e investigación que aporta resultados orientados a entender la enfermedad del tifo.

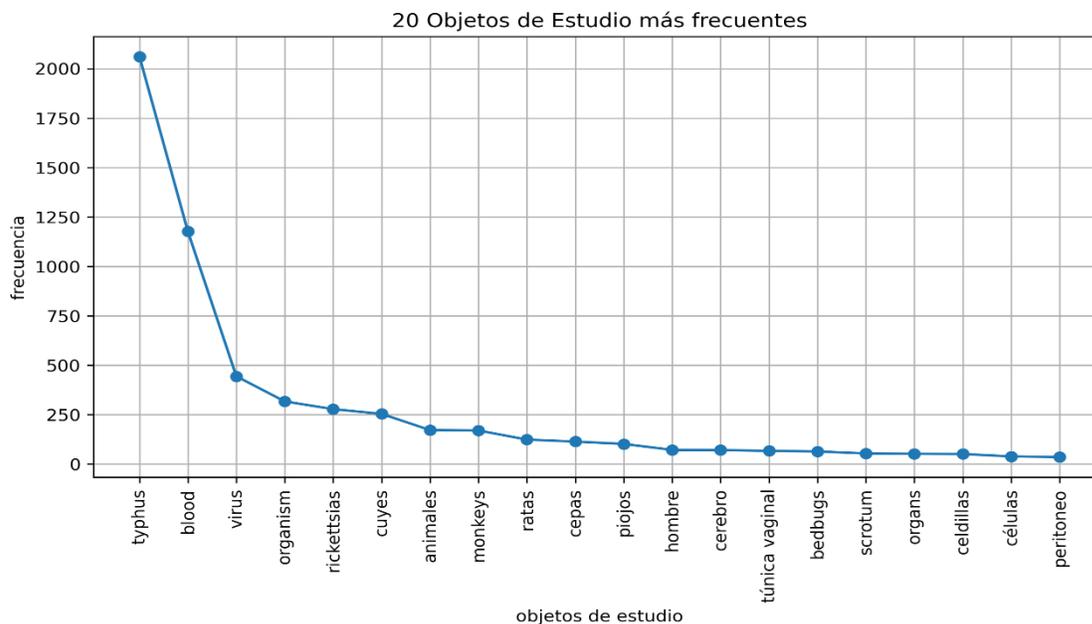


Figura 35. Frecuencia de palabras (objetos de estudio)

Fuente: Elaborada por los autores

Como se puede observar en la Figura 35, las palabras corresponden a distintos modelos de experimentación, entre los que se destacan a roedores (*cuyes* y *ratas*), mamíferos de gran tamaño (*monkeys-monos* y *hombres*) y bacterias (*typhus*), así como los distintos atributos que se estudiaron de cada uno de estos modelos, como cerebros, blood-sangre y organs-órganos. Estos modelos de experimentación fueron utilizados para el desarrollo y pruebas de la efectividad de las vacunas.

4.2.2 Procedimientos

Esta categoría está conformada por 63 palabras (Anexo 1), lo que corresponde a 16.6% del total de términos. La Figura 36 muestra las palabras más relevantes

relacionadas a los métodos, técnicas y prácticas utilizadas en el estudio de los objetos presentados anteriormente. Cabe mencionar, que algunos procedimientos se aplicaron específicamente al estudio de animales (modelos de experimentación) como mamíferos (*cuyes, ratas, etc.*), artrópodos (*piojos, pulgas, garrapatas, etc.*) y bacterias que producen la enfermedad epidémica del tifo (*rickettsias*).

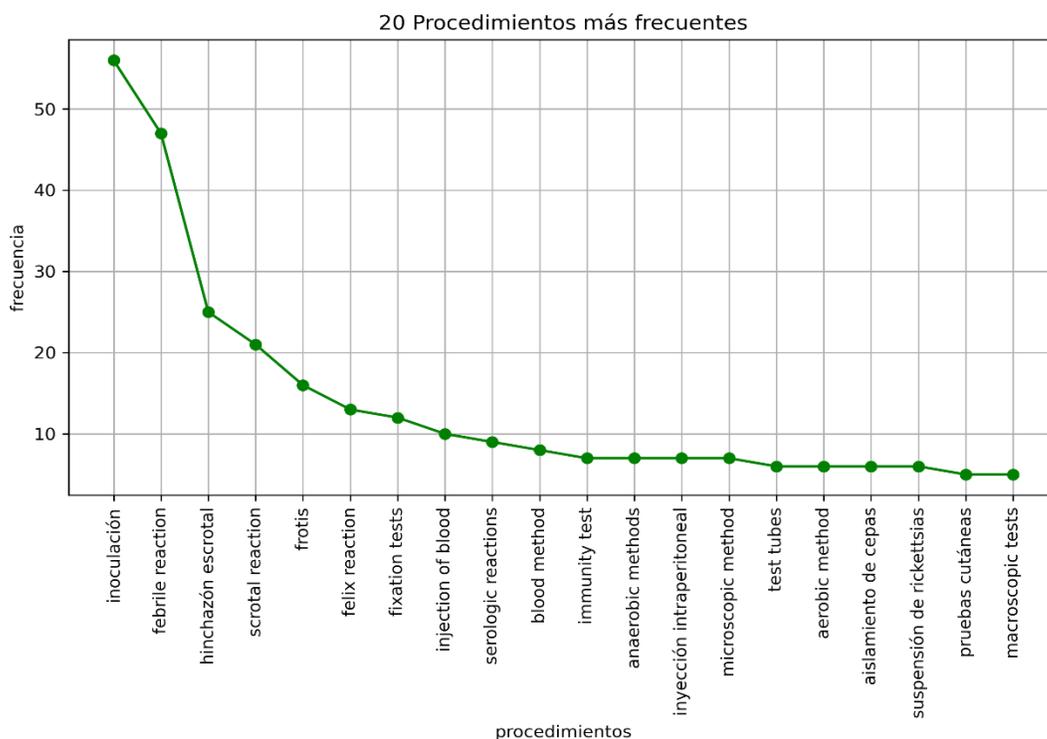


Figura 36. Frecuencia de palabras (procedimientos)

Fuente: Elaborada por los autores

4.2.3 Instrumentos

La categoría instrumentos se encuentra conformada por 35 palabras (Anexo 1), las cuales representan un 9.25 % del total de términos. En la Figura 37 se visualizan las palabras más frecuentes relacionadas con las herramientas y utensilios que son utilizados como recursos de apoyo en el análisis de los objetos de estudio y, por ende, que son implementados en los procedimientos.

Las palabras de esta categoría se pueden dividir en dispositivos de laboratorio, tales como tubes, flask, frasco, microscopic., y en sustancias como alcohol, serum, ácido, etc.

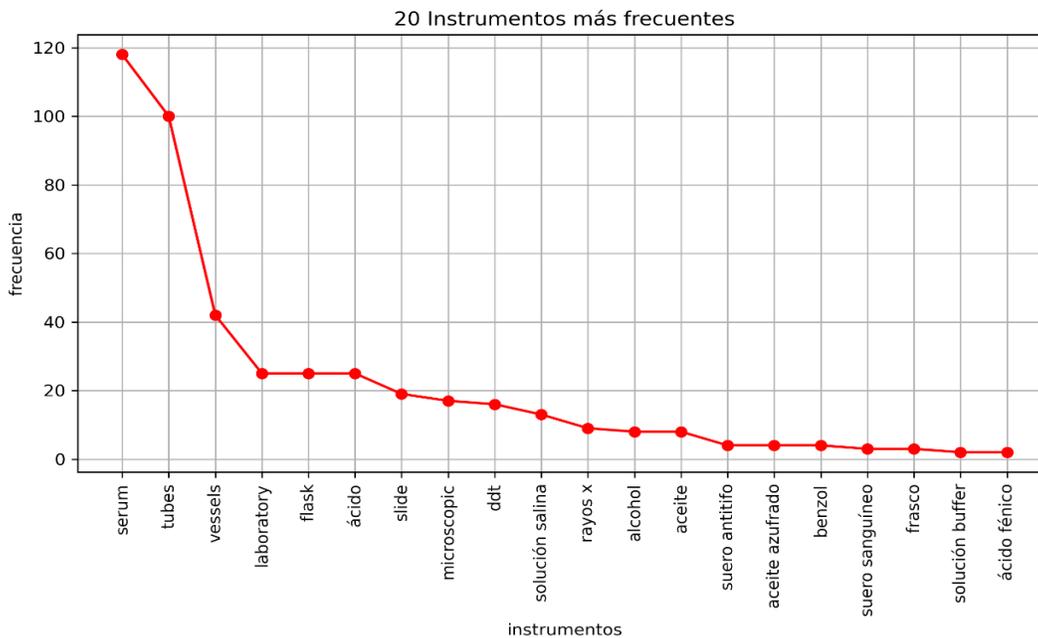


Figura 37. Frecuencia de palabras (instrumentos)

Fuente: Elaborada por los autores

De este modo, y como resultado de la interpretación de los primeros términos de cada categoría, podemos inferir que, dentro del dominio de conocimiento del tifo en México, el procedimiento que más se aplicó fue la inoculación de suero y bacterias del tifo a los diversos modelos de experimentación (animales) utilizados en el laboratorio. Estos componentes mencionados son parte de la materialidad que hizo posible una práctica de investigación científica que permitió la experimentación con la finalidad de encontrar resultados aplicados a desarrollar vacunas para erradicar la enfermedad. Las relaciones entre tipos distintos de entidades nombradas, tales como *objetos de estudio*, *procedimientos* e *instrumentos* se encuentran interrelacionadas en la red semántica que se muestra en la Figura 38. En esta se visualizan las relaciones y estructuras formadas entre las palabras. La red semántica se generó en Python con el algoritmo *kamada_kawai*, los nodos representan las entidades de las categorías antes mencionadas y los vértices establecen las relaciones que existen entre estas. Cabe mencionar que el tamaño de los nodos es proporcional a la coocurrencia de palabras y no depende de su frecuencia de aparición en el corpus textual.

La Figura 38 está orientada a mostrar las relaciones semánticas existentes entre tipos distintos de entidades nombradas, en este caso, se muestran los objetos que se estudiaron a través de diferentes procedimientos y estos a su vez aparecen relacionados con el tipo de instrumentos utilizados. De este modo, las palabras con nodos más visibles y con más relaciones, son los que más relevancia tuvieron en el dominio del conocimiento del tifo en México.

4.2.4 Lugares

Esta categoría está conformada por 152 palabras (Anexo 1) que corresponden a un 40.1% del total de términos de las categorías. La Figura 39 muestra las palabras más frecuentes referentes a las zonas geográficas donde se realizaron los procedimientos científicos. La información sobre el origen de estos lugares permitió el desarrollo de indicadores de espacialidad en el tema del tifo en México. Cabe mencionar, que las palabras de esta categoría están vinculadas con zonas geográficas de diferentes partes del mundo, tal y como se visualiza en el mapa 2.

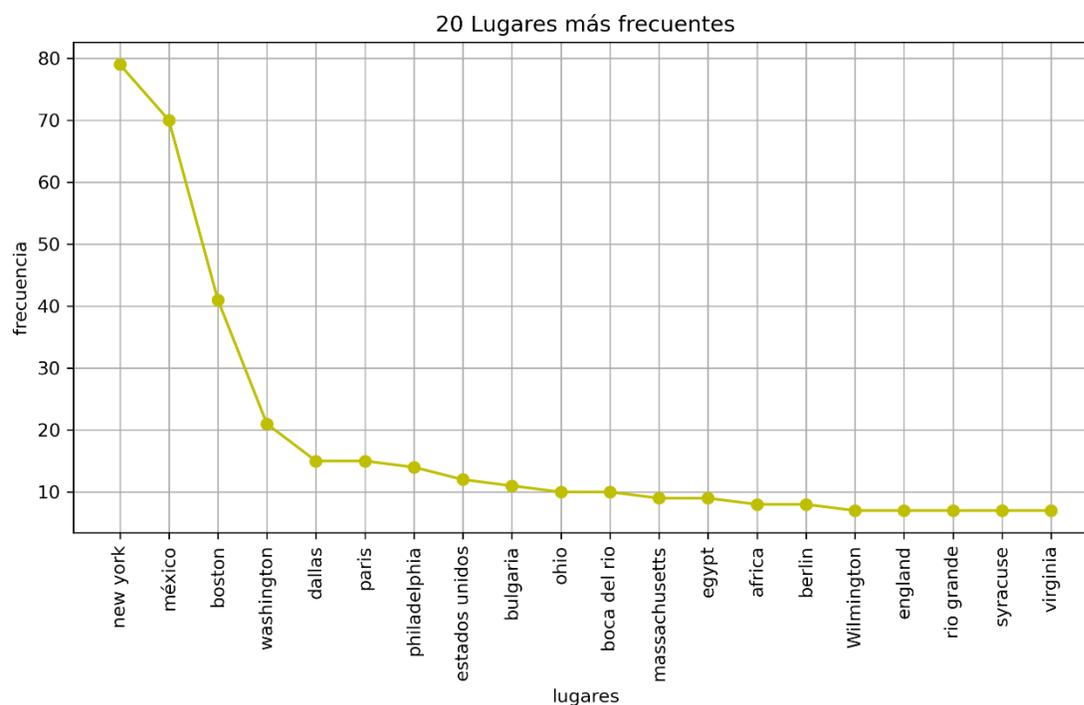
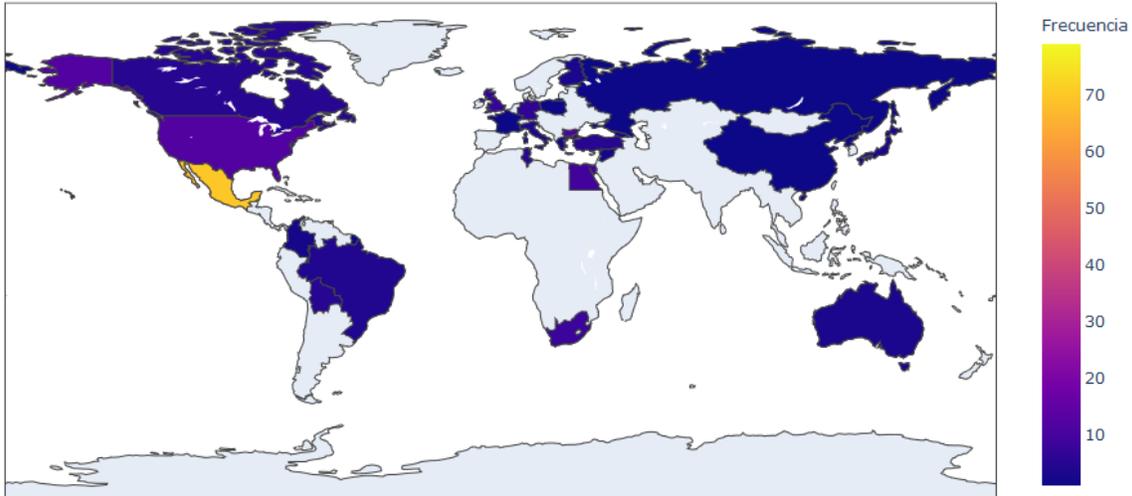


Figura 39. Geografía de las palabras referentes a lugares

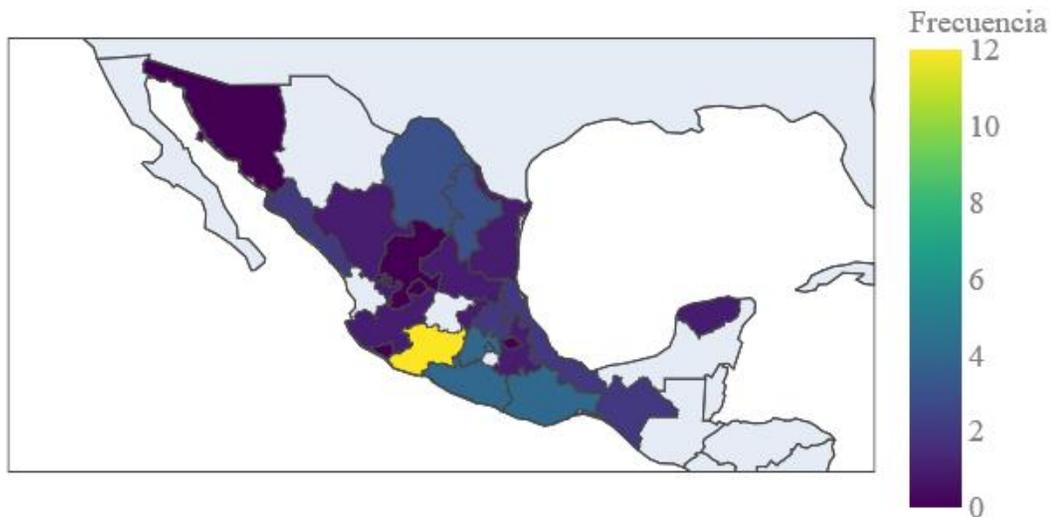
Fuente: Elaborada por los autores



Mapa 2. Geografía de las palabras referentes a lugares

Fuente: Elaborada por los autores

El mapa 2 muestra las palabras de los lugares asociados a los países más relevantes con base a su frecuencia de aparición en el corpus de información, lo cual infiere el papel importante que estos países desempeñaron en el desarrollo de la investigación científica sobre el tifo (Anexo 2). En el mapa 3, se visualiza un mapa correspondiente sólo al país de México, que muestra la distribución de los estados más relevantes en el tema del tifo, con base en la frecuencia de aparición.



Mapa 3. Frecuencia de nombres de Estados de México mencionados en los textos

Fuente: Elaborado por los autores

4.2.5 Instituciones

Esta categoría está conformada por 70 nombres de instituciones (Anexo 1) que representan un 18.4% del total de términos de las categorías. En la Figura 40 se muestran las instituciones que adquirieron roles como sedes de trabajos de investigación, de las cuales la institución más frecuente fue el Hospital General de México. Estas instituciones también jugaron un papel muy importante en la formación académica de recursos humanos.

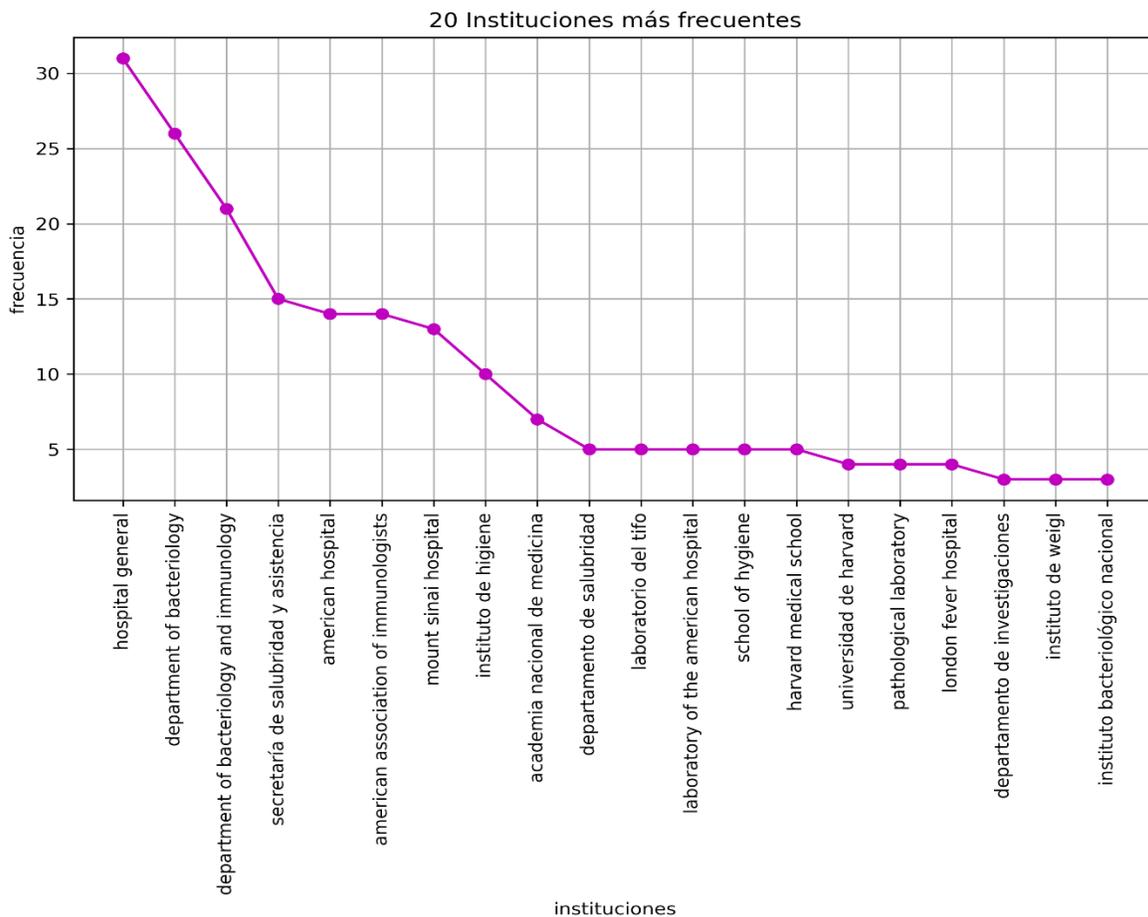


Figura 40. Frecuencia de palabras (instituciones)

Fuente: Elaborada por los autores

La Figura 41 muestra un grafo donde se observan las relaciones de las categorías *lugares e instituciones* más representativas en torno al tema del tifo en México.

De igual manera, el grafo se generó en Python utilizando el algoritmo de dispersión de nodos *kamada_kawai*, así mismo, el tamaño de los nodos representa sus relaciones y no su frecuencia en el corpus textual. Como se puede observar, se destaca ampliamente la relevancia de México como sede geográfica de la mayor parte de las instituciones, así como Estados Unidos. En el caso de México, las instituciones pertenecen a diferentes sectores, como centros de investigación, instituciones de salud y universidades, solo por mencionar algunas.

El caso de Estados Unidos no es diferente, también podemos identificar instituciones pertenecientes al sector salud, universidades y organizaciones sin fines de lucro. Otras instituciones importantes en torno al tema del tifo tienen su sede en los países europeos, así como del continente africano y del medio oriente.

4.3 Resultados en Protégé

Aspectos generales

La ontología como sistema de representación de información de la enfermedad del tifo en México, es un esquema conceptual que se encuentra desarrollado mediante el lenguaje OWL, el cual implementa a otros lenguajes como XML y RDF. La ontología está constituida principalmente como un sistema que permite la recuperación de información específica sobre este dominio de conocimiento, girando en torno a las categorías: *objetos de estudio, procedimientos, instrumentos, lugares e instituciones*.

Generalmente, se puede acceder a la información de una ontología mediante un sistema de información que contenga una ontología implícita que permita interactuar con datos, información y conocimiento de un dominio específico en forma semántica. Una ventaja de un sistema de información basado en ontologías es facilitar la realización de las consultas y la navegación (Barachini, Alvarez & Fortea, 2009). Para motivos de esta investigación y a falta de un sistema de información basado en ontologías, la información contenida en este esquema conceptual se mostrará por medio del software Protégé.

En primer lugar, la Figura 42 y 43 muestran las palabras asociadas a las categorías de análisis: *objetos de estudio, procedimientos, instrumentos, lugares e*

instituciones más relevantes mencionados en los textos. Como se puede observar, las palabras se visualizan de manera jerárquica, lo que representa las subdivisiones que una palabra tiene. Estas subdivisiones corresponden con la estructuración de las taxonomías, las cuales se definen como instancias.

En términos generales, las palabras se encuentran divididas en distintas tipologías, lo que representa la variedad de elementos considerados dentro del desarrollo científico de la enfermedad del tifo. Dentro de la categoría *objetos de estudio* se puede identificar vocabulario que está relacionado con distintos tipos de animales domésticos (*vectores de transmisión*), así como insectos (*reservorios*), bacterias (*agente infeccioso*) que sirvieron como base para la diversificación de la epidemia.



Figura 42. Jerarquía del vocabulario (objetos de estudio, procedimientos e instrumentos)

Fuente: Elaborada por los autores



Figura 43. Jerarquía del vocabulario (lugares e instituciones)

Fuente: Elaborada por los autores

En la Figura 43 se presenta el análisis de las categorías de *lugares* e *instituciones* que se encuentra jerarquizada dependiendo de si una institución desempeñó las funciones de investigación, formación o adscripción. Como se puede observar, esta categoría engloba instituciones de diversos sectores, como universidades, hospitales y organizaciones pertenecientes a un gremio en particular. Por otro lado, el vocabulario de la categoría *lugares* se encuentra estructurado de lo general a lo particular dependiendo de la zona geográfica. La Figura 44, muestra un grafo de relaciones semánticas entre palabras de la categoría *objetos de estudio* generado en el software Protégé. Los nodos se refieren a las palabras, que incluyen a los animales domésticos, roedores, mamíferos, ovíparos, insectos y bacterias. Los vértices establecen el tipo de relación que los nodos tienen. El grafo está orientado a identificar el tipo de relaciones que se establecen dentro de los conceptos de las diferentes categorías, en el caso de objetos de estudio, las relaciones están sujetas a los siguiente:

Agente infeccioso → Reservorios → Vectores de transmisión → Huésped

Grafo de relaciones semánticas entre las palabras de la categoría “Objetos de estudio” en Protégé

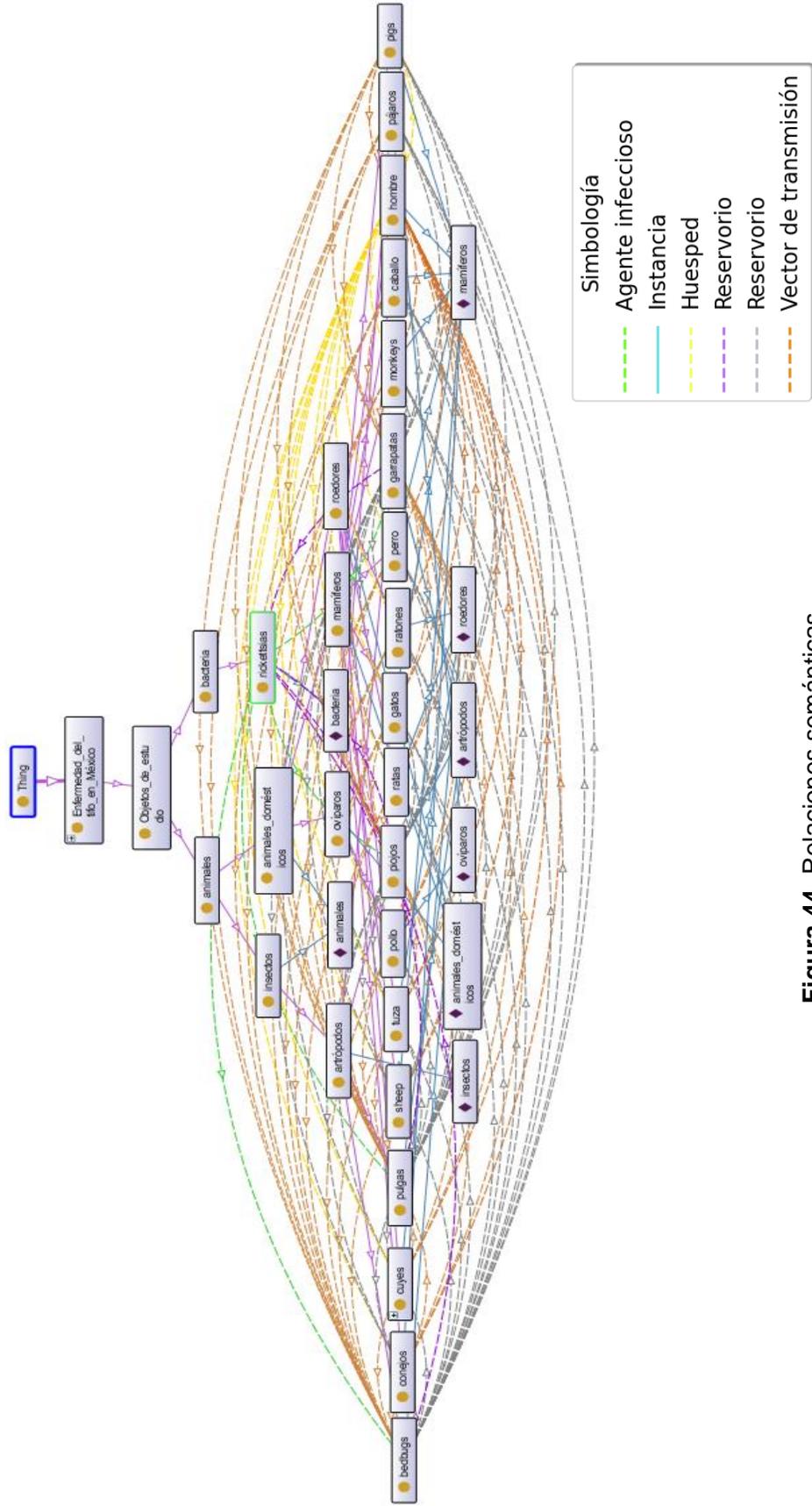


Figura 44. Relaciones semánticas

Fuente: Elaborada por los autores

Consultas SPARQL Query

La pestaña SPARQL Query permite redactar y ejecutar consultas. La pestaña se puede agregar desde “Window”, seguido de “Query views” y por último SPARQL query, tal como se muestra en la Figura 45.

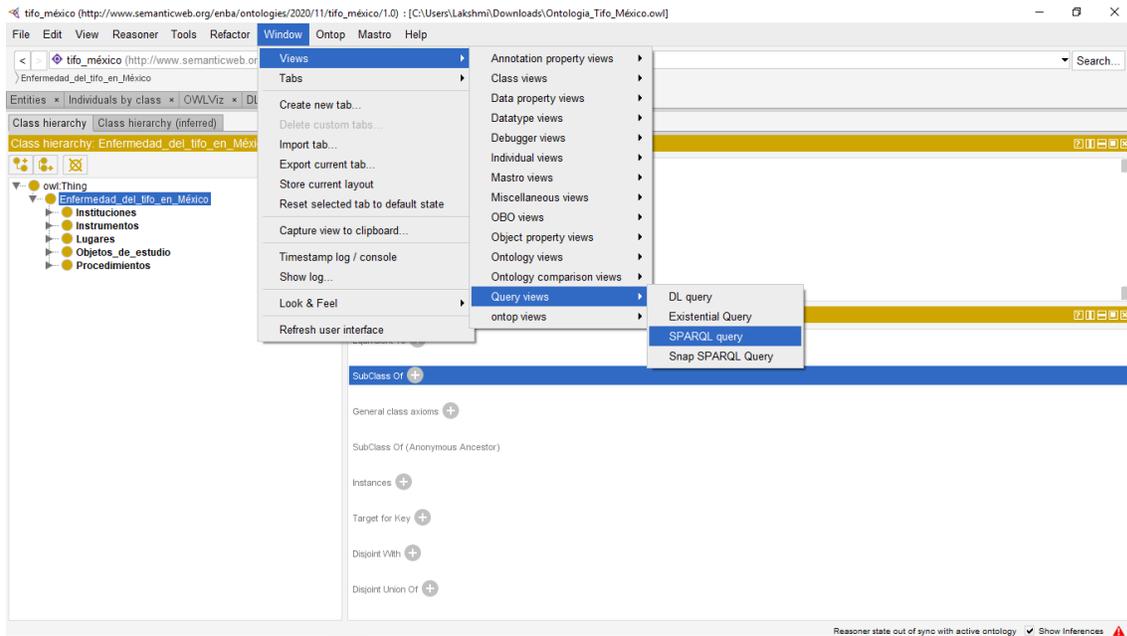


Figura 45. Consulta SPARQL Query (inicio)

Fuente: Elaborada por los autores

SPARQL es un lenguaje de consulta de datos estructurados mediante grafos en formato RDF, el cual tiene como base las tripletas *sujeto-predicado-objeto*. La estructura general de una consulta en SPARQL comienza con prefijos establecidos (PREFIX) que son aquellos que sirven para abreviar las URI (Identificador de Recursos Uniforme), así mismo “SELECT”, se utiliza para definir los ítems de datos sobre los que se realiza la consulta y “WHERE”, especifica el patrón de tripletas antes mencionadas contra el cual se contrastarán los resultados.

Para la recuperación de información específica dentro de este dominio de conocimiento es necesario escribir dentro del apartado “WHERE” una instrucción que refleje el sentido de la consulta. Estas instrucciones se encuentran en el vocabulario *RDF Schema 1.1*.

La primera consulta que se presenta se muestra en la Figura 46, está con la instrucción *rdfs:Sub Class* que muestra los conceptos (clases) en la parte izquierda y sus subclases correspondientes a la derecha. De este modo, se pueden conocer las categorías a las que pertenecen ciertos conceptos.

The screenshot shows a web interface for a SPARQL query. At the top, there are tabs for 'Entities', 'Classes', 'Individuals by class', 'OWL Viz', 'DL Query', and 'SPARQL Query'. The 'SPARQL Query' tab is active, displaying the following query:

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?subject ?object
WHERE { ?subject rdfs:subClassOf ?object }

```

Below the query, the results are displayed in a table with two columns: 'subject' and 'object'.

subject	object
animales_domésticos	animales
termómetro	dispositivos_médicos
hospital_general	Investigación_y_Adscripción
suspensión_fénica	aplicados_a_animales_domésticos
university_of_cincinnati	Formación_y_Adscripción
coloración_de_castañeda	aplicados_a_insectos
scrotal_reacion	aplicados_a_animales_domésticos
serologic_reactions	aplicados_a_animales_domésticos
Zacatecas	México
secretaría_de_salubridad_y_asistencia	Adscripción
Lugares	Enfermedad_del_tifo_en_México
rickettsias_intracelulares_en_frotis_de_túnica_vaginal_de_cuyes	aplicados_a_animales_domésticos
Tamaulipas	México

At the bottom right of the table, there is an 'Execute' button.

Figura 46. Consulta 1 SPARQL Query

Fuente: Elaborada por los autores

La segunda consulta se ejecuta para conocer la descripción o definición de los conceptos. Para esto, se utilizó la instrucción *rdfs:isDefinedBy* la cual se interpreta como “es definido por”. Esto dará como resultado, en la parte izquierda los conceptos y en la derecha su definición correspondiente, tal como se muestra en la Figura 47.

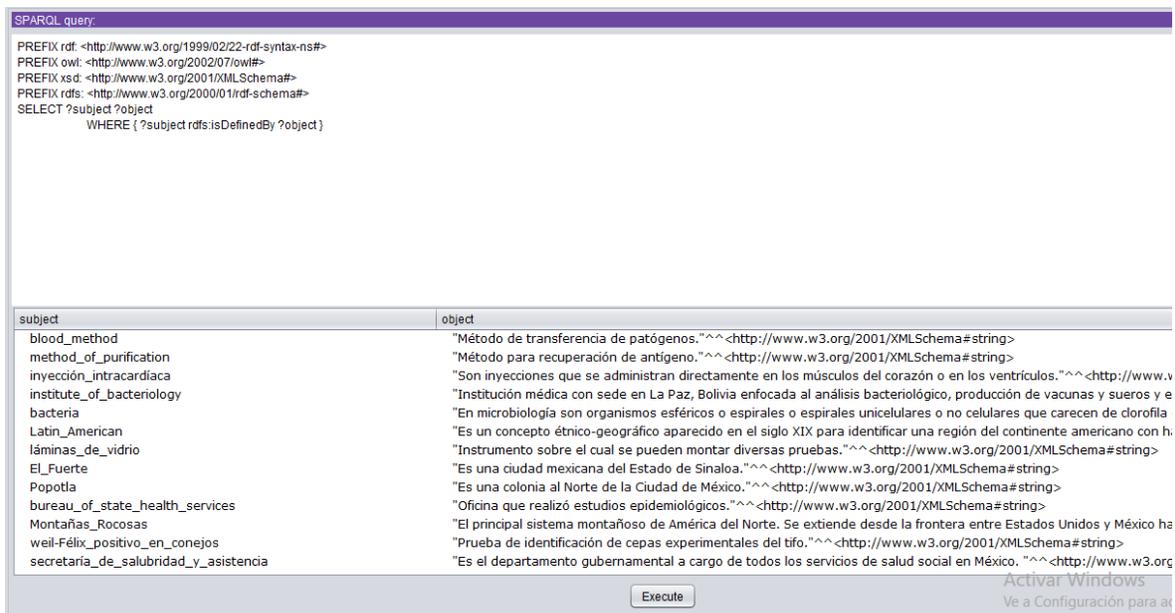


Figura 47. Consulta 2 SPARQL Query

Fuente: Elaborada por los autores

En la Figura 48, se muestra una consulta con la instrucción *rdfs:seeAlso* la cual muestra como resultado la variante de los términos implementados en la ontología. La palabra *seeAlso* significa “ver también”.

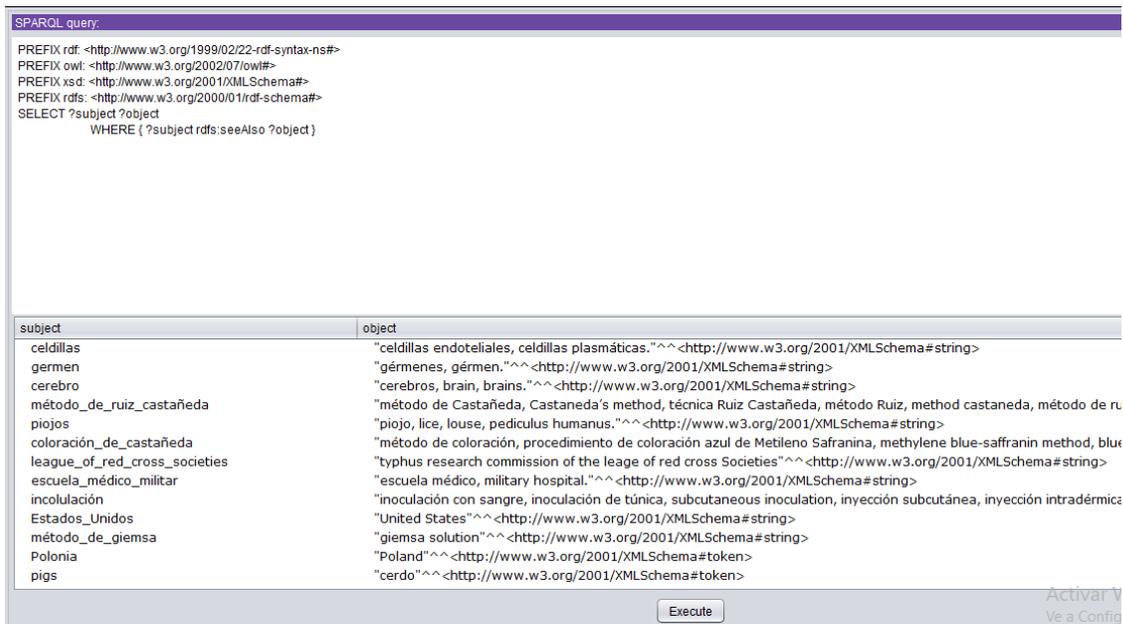


Figura 48. Consulta 3 SPARQL Query

Fuente: Elaborada por los autores

La Figura 49 presenta la consulta que se definió a partir de la instrucción *rdfs:subPropertyOf*, la cual muestra como resultado los términos de la ontología, así como a qué categoría pertenecen.

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?subject ?object
WHERE { ?subject rdfs:subPropertyOf ?object}
```

subject	object
El_Paso	Lugares
Kyoto	Lugares
Paracho	Lugares
ectoparásitos	Objetos_de_estudio
Etiá	Lugares
aceite_alcanforado	Instrumentos
Wilmington	Lugares
Buenos_Aires	Lugares
Fresnillo	Lugares
intestino	Objetos_de_estudio
párasitos	Objetos_de_estudio
cepas	Objetos_de_estudio
Popotla	Lugares

Figura 49. Consulta 4 SPARQL Query

Fuente: Elaborada por los autores

Las consultas anteriores se usaron para ejemplificar el uso de la ontología como sistema de representación y recuperación de la información referente al tema de la enfermedad del tifo en México.

Consideramos importante mencionar que, en lo particular, la ventaja principal del uso de ontologías como herramientas de extracción de información es que estas están conformadas por datos vinculados, mediante lenguajes que permiten la interoperabilidad del usuario y la recuperación explícita de datos específicos de una disciplina.

Conclusiones

El reconocimiento de entidades y la frecuencia de palabras puede considerarse como la construcción de indicadores alternativos en las métricas de la información. Estos indicadores están más orientados al contenido de los textos, permitiendo cuantificar una palabra o secuencia de palabras relacionadas al desarrollo de un producto de investigación o ser incluidas como tópicos relevantes.

El objetivo principal del desarrollo de los indicadores de contenido y temas relevantes es incluirlos en los sistemas de información actuales y que ejerzan como puntos de acceso con base a datos o información de un texto, como indicadores alternativos a los indicadores tradicionales de producción e impacto científico basados en la hegemonía del autor.

El vocabulario utilizado en la ontología comprende 379 palabras que se encuentran divididas entre las categorías: *objetos de estudio, procedimientos, instrumentos, lugares e instituciones*. Las palabras con mayor frecuencia de cada categoría corresponden a *typhus, inoculación, serum, new york, Hospital General*.

Las relaciones semánticas entre las palabras son de dos tipos: (i) relaciones de hechos entre objetos de estudio, procedimientos, e instrumentos, utilizados en la obtención de resultados de los trabajos de tipo experimental; (ii) relaciones de espacialización (geográficas) entre instituciones de formación, investigación, adscripción y las zonas geográficas nacionales y extranjeras como sedes de investigación en el tema del tifo.

De acuerdo con la coocurrencia de palabras entre las categorías mencionadas, los términos *cuyes, ratas, hombres, frotis, coloración de Castañeda, método de Giemsa, jeringas, microscopio, México, Estados Unidos, universidad de Harvard, UNAM*, pueden considerarse como palabras estructurantes de la conformación genealógica del texto científico en el tema del tifo en México.

El vocabulario de 379 palabras clasificadas bajo cada una de las categorías de las palabras estudiadas (*objetos de estudio, procedimientos e instrumentos, lugares e instituciones*), y las relaciones semánticas, de espacialización (geográficas) y de

jerarquía, establecidas entre ellas, configuran la ontología buscada en esta investigación a través de la aplicación del método de bibliominería. Esta ontología es un sistema de representación del conocimiento de la enfermedad del tifo en México en el periodo 1904-1977.

Con la obtención de la ontología mencionada, como lenguaje (vocabulario) representativo del conocimiento producido en la enfermedad estudiada, se cumple con el objetivo perseguido en este trabajo.

La metodología empleada facilitó el análisis de los textos permitiendo una mejor cuantificación de los resultados mediante la automatización de procesos, considerando que estos documentos históricos cuentan con una estructura diferente a la de una publicación actual. Siguiendo esta metodología, se puede aplicar al análisis de otros campos de conocimiento, así como consolidar sistemas de información históricos como el caso del Atlas Histórico de la Ciencia Mexicana.

Esta investigación también pretende dar otro enfoque respecto a formas alternativas de análisis de corpus de información en temas relacionados al desarrollo de la ciencia médica en México.

El desarrollo de ontologías permite el análisis e investigación de nuevos campos del conocimiento, implementando tecnologías actuales. Para poder lograr esos desarrollos tecnológicos, se requieren ciertas habilidades que pueden ser desarrolladas por los profesionales de la información (bibliotecarios) que permitan la integración de estos temas en la enseñanza de la bibliotecología y ciencias de la información en México.

Debilidades de esta investigación. Están relacionadas con problemas propios de la información histórica, específicamente con la digitalización de los documentos que no están en línea y con el preprocesamiento (limpieza del texto).

La investigación presentada, hizo posible el desarrollo metodológico para elaborar la ontología del dominio de conocimiento de la enfermedad del tifo en México en el periodo 1904-1977, así como el análisis bibliométrico de la literatura especializada en el tema.

Referencias

Barchini, G., Alvarez, M. M., & Fortea, G. (2009). *Evaluación de la calidad de los sistemas de información basados en ontologías*. Trabajo presentado en IX Congreso ISKO-España: Nuevas perspectivas para la difusión y organización del conocimiento, Valencia, España.

Saif, H., Fernández, M., He, Y., & Alani, H. (2014). *On stopwords, filtering and data sparsity for sentiment analysis of twitter*.

Anexos

1. Tabla de frecuencia de las palabras pertenecientes a las categorías objeto de estudio, procedimientos, instrumentos, lugares e instituciones.

➤ Objetos de estudio

	Entidad	Frecuencia	Porcentaje %
1	Typhus	2061	33.73
2	Blood	1176	19.24
3	Virus	444	7.27
4	Organism	317	5.19
5	Rickettsias	278	4.55
6	Cuyes	254	4.16
7	Animales	172	2.81
8	Monkeys	170	2.78
9	Ratas	124	2.03
10	Cepas	114	1.87
11	Piojos	102	1.67
12	Hombre	72	1.18
13	Cerebro	71	1.16
14	Túnica Vaginal	67	1.10
15	Bedbugs	64	1.05
16	Scrotum	54	0.88
17	Organs	52	0.85
18	Celdillas	51	0.83
19	Células	39	0.64
20	Peritoneo	36	0.59
21	Pulgas	31	0.51
22	Ratones	31	0.51
23	Microorganisms	30	0.49
24	Germen	28	0.46
25	Orina	27	0.44

26	Conejos	21	0.34
27	Testicle	20	0.33
28	Corpúsculos	19	0.31
29	Pulmones	18	0.29
30	Sheep	18	0.29
31	Bacterias	15	0.25
32	Alveolos	14	0.23
33	Parásitos	11	0.18
34	Médula Ósea	10	0.16
35	Roedores	9	0.15
36	Piel	7	0.11
37	Garrapatas	7	0.11
38	Intestinos	6	0.10
39	Mamíferos	6	0.10
40	Gatos	6	0.10
41	Estómago	5	0.08
42	Pollo	5	0.08
43	Reservorio	5	0.08
44	Leucocytes	4	0.07
45	Perros	4	0.07
46	Tuza	4	0.07
47	Embriones	4	0.07
48	Ectoparásitos	4	0.07
49	Insectos	3	0.05
50	Caballos	3	0.05
51	Artrópodos	3	0.05
52	Pigs	3	0.05
53	Genitals	3	0.05
54	Citoplasma Hidrópico	2	0.03
55	Tejido Pulmonar	2	0.03
56	Citología Medular	2	0.03
57	Gargantas	1	0.02
58	Hígado	1	0.02
59	Pájaros	1	0.02

➤ **Procedimientos**

	Entidad	Frecuencia	Porcentaje %
1	Inoculación	56	15.77
2	Febrile Reaction	47	13.24
3	Hinchazón Escrotal	25	7.04
4	Scrotal Reaction	21	5.92
5	Frotis	16	4.51
6	Felix Reaction	13	3.66
7	Fixation Tests	12	3.38
8	Injection Of Blood	10	2.82
9	Serologic Reactions	9	2.54
10	Blood Method	8	2.25
11	Immunity Test	7	1.97
12	Anaerobic Methods	7	1.97
13	Inyección Intraperitoneal	7	1.97
14	Microscopic Method	7	1.97
15	Test Tubes	6	1.69
16	Aerobic Method	6	1.69
17	Aislamiento De Cepas	6	1.69
18	Suspensión De Rickettsias	6	1.69
19	Pruebas Cutáneas	5	1.41
20	Macroscopic Tests	5	1.41
21	Mielograma	5	1.41
22	Agglutination Tests	4	1.13
23	Prueba Opsónica	4	1.13
24	Diet Method	4	1.13
25	Método De Ruiz Castañeda	3	0.85
26	Método De Giemsa	3	0.85
27	Coloración De Castañeda	3	0.85
28	Pruebas Intracutáneas	3	0.85
29	Prueba Clásica	2	0.56
30	Inyección Rectal	2	0.56
31	Inyección Intracardiaca	2	0.56
32	Observación Directa	2	0.56
33	Dale Method	2	0.56

34	Dilución De Sangre	2	0.56
35	Subcutaneous Method	2	0.56
36	Weil-Félix Positivo En Conejos	2	0.56
37	Lesiones Histológicas Típicas En Cerebros De Cuyes	2	0.56
38	Pruebas De Esterilidad	2	0.56
39	Lung Method	2	0.56
40	Slide-Agglutinative Method	2	0.56
41	Vacunación Profiláctica	1	0.28
42	Inyección De Suspensiones	1	0.28
43	Pruebas De Neutralización	1	0.28
44	Método De Rata A Cuy	1	0.28
45	Prueba De Susceptibilidad	1	0.28
46	Pruebas Alérgicas	1	0.28
47	Inmunización Pasiva	1	0.28
48	Suspensión Félica	1	0.28
49	Precipitin Reaction	1	0.28
50	Injection Of Blood From Man	1	0.28
51	Method Of The-Obald Smith	1	0.28
52	Gram'S Method	1	0.28
53	Sellards' Technic	1	0.28
54	X-Ray Method	1	0.28
55	Procedimiento De Infección Intra-Rectal	1	0.28
56	Procedimiento De Luoto	1	0.28
57	Procedimiento De Diagnóstico Del Tifo	1	0.28
58	Rickettsias Intracelulares En	1	0.28

	Frotis De Túnica Vaginal De Cuyes		
59	Prueba De Shick	1	0.28
60	Maitland Method	1	0.28
61	Method Of Purification	1	0.28
62	Method Of Tissue Culture	1	0.28
63	Fertile Egg Method	1	0.28

➤ **Instrumentos**

	Entidad	Frecuencia	Porcentaje %
1	Serum	118	25.27
2	Tubes	100	21.41
3	Vessels	42	8.99
4	Laboratory	25	5.35
5	Flask	25	5.35
6	Ácido	25	5.35
7	Slide	19	4.07
8	Microscopic	17	3.64
9	DDT	16	3.43
10	Solución Salina	13	2.78
11	Rayos X	9	1.93
12	Alcohol	8	1.71
13	Aceite	8	1.71
14	Suero Antitifo	4	0.86
15	Aceite Azufrado	4	0.86
16	Benzol	4	0.86
17	Suero Sanguíneo	3	0.64
18	Frasco	3	0.64
19	Solución Buffer	2	0.43
20	Ácido Fénico	2	0.43
21	Alcohol Metílico	2	0.43
22	Suero De Un Animal	2	0.43
23	Aceite De Olivo	2	0.43
24	Suero De Cuy	2	0.43
25	Termómetro	2	0.43
26	Metileno	1	0.21
27	De Aceite Alcanforado	1	0.21

28	Papel Lustre	1	0.21
29	Solución De Formalina	1	0.21
30	Ácido Acético	1	0.21
31	Suero Antiproteo	1	0.21
32	Mortero Estéril	1	0.21
33	Gasa	1	0.21
34	Láminas De Vidrio	1	0.21
35	Jeringas	1	0.21

➤ **Lugares**

	Entidad	Frecuencia	Porcentaje %
1	New York	79	12.58
2	México	70	11.15
3	Boston	41	6.53
4	Washington	21	3.34
5	Dallas	15	2.39
6	Paris	15	2.39
7	Philadelphia	14	2.23
8	Estados Unidos	12	1.91
9	Bulgaria	11	1.75
10	Ohio	10	1.59
11	Boca Del Rio	10	1.59
12	Massachusetts	9	1.43
13	Egypt	9	1.43
14	Africa	8	1.27
15	Berlin	8	1.27
16	Wilmington	7	1.11
17	England	7	1.11
18	Rio Grande	7	1.11
19	Syracuse	7	1.11
20	Virginia	7	1.11
21	San Francisco	6	0.96
22	San Antonio	6	0.96
23	Huixtla	6	0.96
24	Germany	6	0.96
25	Bolivia	5	0.80
26	Sinaloa	5	0.80
27	North American	5	0.80
28	Túnez	5	0.80

29	El Paso	5	0.80
30	Turkey	5	0.80
31	Canada	5	0.80
32	Volhynia	5	0.80
33	Oaxaca	4	0.64
34	León	4	0.64
35	Brasil	4	0.64
36	Querétaro	4	0.64
37	Tlalnepantla	4	0.64
38	Europa	4	0.64
39	Australia	3	0.48
40	La Laguna	3	0.48
41	Durango	3	0.48
42	Michoacán	3	0.48
43	Finland	3	0.48
44	Mcallen	3	0.48
45	Longview	3	0.48
46	Japan	3	0.48
47	Zinacatepec	3	0.48
48	Pennsylvania	3	0.48
49	Zanzibar	3	0.48
50	Cincinnati	3	0.48
51	North Carolina	3	0.48
52	Italia	3	0.48
53	Asia Oriental	2	0.32
54	Puebla	2	0.32
55	El Fuerte	2	0.32
56	Sonora	2	0.32
57	Coahuila	2	0.32
58	Colombia	2	0.32
59	Puruándiro	2	0.32
60	Paracho	2	0.32
61	San Paulo	2	0.32
62	Grecia	2	0.32
63	Toluca	2	0.32
64	Tehuantepec	2	0.32
65	Tlaxcala	2	0.32
66	Jalisco	2	0.32
67	San Luis Potosí	2	0.32
68	Gulf Of Mexico	2	0.32
69	Syria	2	0.32

70	Maryland	2	0.32
71	Georgia	2	0.32
72	Indiana	2	0.32
73	Illinois	2	0.32
74	Aleppo	2	0.32
75	Alabama	2	0.32
76	Arcelia	2	0.32
77	Iguala	2	0.32
78	Wisconsin	2	0.32
79	Minnesota	2	0.32
80	Servia	2	0.32
81	Jungapeo	2	0.32
82	Orizaba	2	0.32
83	Zacatecas	2	0.32
84	México, D.F.	1	0.16
85	Latin American	1	0.16
86	Apodaca	1	0.16
87	Ejido Zaragoza	1	0.16
88	Veracruz	1	0.16
89	Colima	1	0.16
90	Cherán	1	0.16
91	Maravatío	1	0.16
92	Pátzcuaro	1	0.16
93	Tlalpujahuá	1	0.16
94	Zitácuaro	1	0.16
95	Zamora	1	0.16
96	La Piedad	1	0.16
97	Morelia	1	0.16
98	Panamá	1	0.16
99	Uruapan	1	0.16
100	Tamaulipas	1	0.16
101	Nuevo Laredo	1	0.16
102	Francia	1	0.16
103	Cuernavaca	1	0.16
104	New Orleans	1	0.16
105	New Brunswick	1	0.16
106	Rio De Janeiro	1	0.16
107	Aguascalientes	1	0.16
108	Estado De México	1	0.16
109	Coatzacoalcos	1	0.16
110	Rusia	1	0.16

111	Korea	1	0.16
112	Chiapas	1	0.16
113	Guerrero	1	0.16
114	Xochimilco	1	0.16
115	Guadalajara	1	0.16
116	Texas	1	0.16
117	Polonia	1	0.16
118	Varsovia	1	0.16
119	China	1	0.16
120	Balkanes	1	0.16
121	Oklahoma City	1	0.16
122	New Braunfels	1	0.16
123	Galveston	1	0.16
124	Gainesville	1	0.16
125	Kyoto	1	0.16
126	Kyushu	1	0.16
127	Okayama	1	0.16
128	Montañas Rocosas	1	0.16
129	Popotla	1	0.16
130	Tapachula	1	0.16
131	Buenos Aires	1	0.16
132	Keijo	1	0.16
133	San Nicolas Totolapa	1	0.16
134	San Jose	1	0.16
135	Acapulco	1	0.16
136	Tlapa	1	0.16
137	Hidalgo	1	0.16
138	Pachuca	1	0.16
139	Tepeitic	1	0.16
140	Metepec	1	0.16
141	General Teran	1	0.16
142	Linares	1	0.16
143	Mitla	1	0.16
144	Tlalixtac De Cabrera	1	0.16
145	Etla	1	0.16
146	Huimilpan	1	0.16
147	Zacatón	1	0.16
148	Mazatlán	1	0.16
149	Yucatán	1	0.16
150	Mérida	1	0.16

151	Fresnillo	1	0.16
152	San Lorenzo Oyamel	1	0.16

➤ **Instituciones**

	Entidad	Frecuencia	Porcentaje %
1	Hospital General	31	11.57
2	Department Of Bacteriology	26	9.70
3	Department Of Bacteriology And Immunology	21	7.84
4	Secretaría De Salubridad Y Asistencia	15	5.60
5	American Hospital	14	5.22
6	American Association Of Immunologists	14	5.22
7	Mount Sinai Hospital	13	4.85
8	Instituto De Higiene	10	3.73
9	Academia Nacional De Medicina	7	2.61
10	Departamento De Salubridad	5	1.87
11	Laboratorio Del Tifo	5	1.87
12	Laboratory Of The American Hospital	5	1.87
13	School Of Hygiene	5	1.87
14	Harvard Medical School	5	1.87
15	Universidad De Harvard	4	1.49
16	Pathological Laboratory	4	1.49
17	London Fever Hospital	4	1.49
18	Departamento De Investigaciones	3	1.12
19	Instituto De Weigl	3	1.12
20	Instituto Bacteriológico Nacional	3	1.12
21	Manicomio General	3	1.12
22	Facultad De Medicina	3	1.12
23	Baylor University	3	1.12
24	Division Of Epidemiology	3	1.12
25	Liverpool City Hospital	3	1.12
26	Cárcel De Belem	2	0.75
27	Lebanon Hospital	2	0.75
28	Instituto Pasteur	2	0.75
29	División De Química Orgánica De E. R. Squibb & Sons	2	0.75

30	League Of Red Cross Societies	2	0.75
31	Rockefeller Institute For Medical Research	2	0.75
32	Cárcel De Xochimilco	2	0.75
33	American Society Of Tropical Medicine	2	0.75
34	State Hygienic Laboratory	2	0.75
35	Harvard Laboratories	2	0.75
36	International Health Division Of The Rockefeller Foundation	2	0.75
37	Instituto De Enfermedades Tropicales	1	0.37
38	Instituto De Butantan	1	0.37
39	Escuela De Medicina	1	0.37
40	Laboratorio Wollbach	1	0.37
41	Instituto De Lwow	1	0.37
42	Instituto Micro Biológico	1	0.37
43	U.N.A.M	1	0.37
44	Comisión Sanitaria De La Cruz Roja Americana	1	0.37
45	Laboratorios De Investigaciones Médicas	1	0.37
46	Universidad De Chicago	1	0.37
47	Instituto De Nicolle	1	0.37
48	University Of Cincinnati	1	0.37
49	Department Of Bacteriology And Hygiene	1	0.37
50	Rocky Mountain Laboratory	1	0.37
51	Mexican Typhus Commission	1	0.37
52	Escuela Superior De Medicina De Keijo	1	0.37
53	Institute Of Bacteriology	1	0.37
54	Bureau Of State Health Services	1	0.37
55	Bureau Of Entomology And Plant Quarantine Of The United States	1	0.37
56	Insecticide Laboratory In Orlando	1	0.37
57	Catholic University	1	0.37
58	American Medical Association	1	0.37
59	Fordham Hospital	1	0.37
60	Alexian Brothers Hospital	1	0.37
61	Museum Of Natural History	1	0.37
62	Marine-Hospital Service	1	0.37
63	Escuela Médico Militar	1	0.37
64	Hygiene And Preventive Medicine	1	0.37
65	Christ Hospital	1	0.37
66	Research Department	1	0.37

67	Salmonella And Virus Laboratory	1	0.37
68	Laboratory Of Dr. Ernest Goodpasture	1	0.37
69	Laboratory Of Dr. Gerardo Varela	1	0.37
70	Typhus Research Laboratory Of Dr. Castaneda	1	0.37

2. Estados y municipios más relevantes con base a su frecuencia en el desarrollo científico en torno al tema del tifo en México.

Estado/ Municipio	Información específica
Edo. México	Se realizó un estudio del tifo, mediante el método Ruiz Castañeda
Metepéc	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Toluca	Experimentación en el hombre mediante la vacunación
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
	Fueron atendidos varios casos de personas enfermas con tifus
Tlalnepantla	Se realizó un tratamiento comparativo con polvo y jabón
	Se realizó una desparasitación a individuos en masa aplicada al dominio de un brote epidémico de tifo
	Se realizó despiojamiento a un poblado donde existía una epidemia de la enfermedad del tifo
	Se realizaron inspecciones de casos y parásitos en personas
San Lorenzo Oyamel	Se realizó despiojamiento a individuos en masa por DDT, aplicado al dominio de un brote epidémico de tifo
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
	Se realizaron inspecciones de casos y parásitos en personas
	Se administro tratamiento contra el tifo a personas y se observó un aumento de infestación
<i>Tamaulipas</i>	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Nuevo Laredo	Se obtuvieron 7 sueros de ratas capturadas, los cuales fueron positivos al tifo clásico

<i>Oaxaca</i>	Se inmunizaron a personas satisfactoriamente, suficiente para controlar algunos brotes epidémicos.
	Se examinaron a individuos que presentaban grandes probabilidades de haber padecido tifo
	Experimentación en el hombre mediante la vacunación
	Entidad más afectada por el tifo (defunciones)
	Entidad más afectada por brotes del tifo
	Estudio de la epidemia del tifus
	Experimentación con conejillo de indias
Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica	
Tlalixtac de Cabrera	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Etla	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Mitla	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Tehuantepec	Se realizó inoculación en una especie del mono
<i>Puebla</i>	Se inmunizaron a personas satisfactoriamente, suficiente para controlar algunos brotes epidémicos.
	Entidad más afectada por el tifo (defunciones)
	Entidad más afectada por brotes del tifo
Zinacatepec	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
	Estudio sobre las cepas del virus tifo
	Recolección de piojos de un paciente
<i>Ciudad de México</i>	Estudio sobre el suero de personas
	Investigación experimental, sobre bases epidemiológicas
	Diagnosticó en habitantes sobre la fiebre de las trincheras
	Experimentación en el hombre mediante la vacunación
	Pruebas en algunos infectados del tifo del Hospital General de la ciudad de México
	Análisis de resultados de algunas pruebas llevadas a cabo en distintos grupos de personas
	Realización de pruebas del tifo exantemático en el Hospital General
	Pruebas en residentes donde no ha sido observada la enfermedad
Estudio sobre el tifo exantemático	

	Experimentación de las cepas del tifo exantemático
	Aislamiento de cepas de tifo de enfermos
	Observaciones del tifo exantemático en casos registrados
	Experimentación del tifo en ratas y cuy
	Parasitación encontrada en algunas colonias de la Ciudad de México
	Se estudiaron algunos brotes de tifo orquíutico y no-orquíutico
	Estudio del virus del cerebro de ratas
	Estudio sobre las relaciones entre el tifo murino y el tifo clásico
	Experimento con un cuy con orquitis tifosa típica, inoculado con tifo endémico
	Investigaciones sobre todas las especies de pulgas, encontradas en ratas
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Coatzacoalcos	Experimento en monos
Xochimilco	Se realizó parasitación a individuos de la cárcel de Xochimilco
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Popotla	Experimentos sobre la conversión de cepas de tifus
San Nicolas	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Totolapa	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Jalisco</i>	Estudios sobre una infección exantemática endémica y fiebre tifo-exantemática benigna
	Estudio sobre una enfermedad infecciosa aguda en Zacoalco, considerado por los médicos de la región como de fiebre tifoidea.
Guadalajara	Experimentación en el hombre mediante la vacunación
	Experimentación en ratas y cuyes
<i>Guerrero</i>	Entidad más afectada por brotes del tifo
Tlapa	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Acapulco	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica

Arcelia	Se realizaron brigadas de salud y se otorgaron tratamientos para investigaciones
	Se realizó una observación clínica detallada y continua de los casos del tifo
Iguala	Se realizaron brigadas de salud y se otorgaron tratamientos para investigaciones
<i>Nuevo León</i>	Estudio de cepas Proteus X-19 y XK en el Instituto Instituto de Higiene
Apodaca	Estudio de cepas Proteus X-19 y XK en el Instituto Instituto de Higiene
General Terán	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Linares	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Yucatán</i>	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Mérida	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Hidalgo</i>	Entidad más afectada por el tifo (defunciones)
	Entidad más afectada por brotes del tifo
Pachuca	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Teipeitic	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Sinaloa</i>	Estudio de cepas Proteus X-19 y XK en el Instituto Instituto de Higiene
	Aparición de un brote de la fiebre manchada
	Estudio de casos del tifo en rancherías
	Estudio del tifo en zonas rurales
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
El Fuerte	Aparición de un brote de la fiebre manchada
	Estudio de casos del tifo en rancherías
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Mazatlán	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>San Luis Potosí</i>	Se realizó experimento a dos individuos
	Aplicación de vacuna a personas
	La inoculación en cinco personas, convaleciente de tifo

	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Zacatón	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Chiapas</i>	Entidad más afectada por brotes del tifo
Huixtla	Se realizaron campañas de salubridad para enfermedades tropicales Se realizó una observación clínica detallada y continua de los casos del tifo
Tapachula	Se realizó una observación clínica detallada y continua de los casos del tifo
<i>Michoacán</i>	Investigación de una fiebre petequial Se realizó una clasificación serológica Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Cherán	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
La piedad	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Puruándiro	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Maravatío	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Morelia	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Paracho	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Tlalpujahua	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Zamora	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Zitácuaro	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Jungapeo	Se realizó un diagnóstico clínico y un estudio en ranchos sobre el brote del tifus Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica

Uruapan	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
Pátzcuaro	Aplicación de la reacción de Weil Félix practicada con sueros de enfermos
<i>Querétaro</i>	Entidad más afectada por brotes del tifo
	Entidad más afectada por el tifo (defunciones)
	Participaron enfermeras en la Campaña Nacional contra el Tifo en zonas rurales
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Huimilpan	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Coahuila</i>	Estudio del tifo en zonas rurales
	Estudio del suero sanguíneo de animales
	Entidad más afectada por el tifo (defunciones)
	Entidad más afectada por brotes del tifo
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Ejido Zaragoza	Estudio de fiebre manchada en garrapatas colectadas en un perro
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
San José	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
La Laguna	Estudio del tifo en zonas rurales
	Zona infectada intensamente por la fiebre manchada
	Estudio del suero sanguíneo de animales
<i>Durango</i>	Estudio del tifo en zonas rurales
	Estudio del suero sanguíneo de animales
	Experimentación en el hombre mediante la vacunación
	Entidad más afectada por brotes del tifo
	Entidad más afectada por el tifo (defunciones)
La Laguna	Estudio del tifo en zonas rurales
	Zona infectada intensamente por la fiebre manchada
	Estudio del suero sanguíneo de animales
<i>Veracruz</i>	Estudio de la infección natural del amblyomma cajennense con fiebre manchada
	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica

Orizaba	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Boca del Río	Se realizó una observación clínica detallada y continua de los casos del tifo
<i>Tlaxcala</i>	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
<i>Zacatecas</i>	Se extrajeron de varias comunidades muestras de suero positivas de tifus, fiebre manchada, clásica
Sonora	Estudio de casos del tifo en rancherías
	Estudio del tifo en zonas rurales