Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Tesis para obtener el grado de

Licenciado en Ingeniería en Ciencias de la Computación

Una aproximación a un algoritmo clasificador para la identificación de interacciones medicamentosas

Autor: Luis Daniel Oidor Juárez

Asesores: PhD Luis Enrique Colmenares Guillén

Dr. José Gustavo López y López

Puebla, Puebla, México

Verano de 2015

Agradecimientos

El presente trabajo es el resultado de años de esfuerzo. Y no me refiero a la elaboración de esta tesis como tal, sino a todo el proceso desde que llegué a la Facultad de Ciencias de la Computación hasta el día de hoy. Cierto es que el proceso no se llevó a cabo conforme a lo planeado. En el camino surgieron situaciones súbitas que me hicieron cambiar el rumbo una y otra vez.

Debo a mis padres más que a nadie en este mundo. Gracias a ellos soy lo que soy. Ellos han sido mi inspiración y mi respaldo en todos los momentos de mi vida. Hoy, parte de su esfuerzo por darme un futuro se ve recompensado.

Tenía poco menos de 4 años cuando nació una bebé llorona que mis padres tendrían a bien llamar Alma Raquel. Ella me hace bromas todo el tiempo y parece demasiado gruñona, pero en realidad es la ternura en dos pies. Siempre ha pensado que su hermano mayor – es decir, yo – es una especie de superhéroe invencible. Esa idea me hace esforzarme por parecerlo aunque sea un poco.

Y como si no tuviera suficiente en esta vida, hace tiempo que apareció en mi vida la mujer con la que he decidido compartir los días que me quedan en esta breve estancia que suele llamarse vida. Su nombre es Ani. Ella ha sabido hacerme mejor persona y me ha dado más razones para nunca detenerme, además de ayudarme con los devaneos del idioma.

Recuerdo que una mañana fría de agosto conocí al Dr. Colmenares. Al principio fue duro conmigo, pero me dio la oportunidad de mostrarle lo buen estudiante que podía ser. Luego me brindó la confianza para realizar el presente trabajo bajo su asesoría. Él fue quien me presentó al Dr. Gustavo y quien le solicitó fuera co-asesor de la tesis que ahora presento.

En definitiva muchas son las personas que se han cruzado en mi camino. Muchas también han sido las que me ayudaron, de una o de otra forma, a alcanzar este objetivo planteado desde hace tiempo. No los menciono por temor a omitir a alguien, pero sé que sabrán perdonar mi desliz.

A todos gracias.

Divulgación científica

El presente trabajo de tesis fue tratado en los siguientes artículos en los que colaboré:

Colmenares Guillén, L. E., López y López, J. G., & Oidor Juárez, L. D. (2014). Identificación probabilística de interacciones medicamentosas. (D. Pinto, & D. Vilariño, Edits.) *Research in Computing Science*(88), 61-74., ISSN: 1870-4069.

Colmenares Guillén, L. E., López y López, J. G., Pérez de Celis, M. d., & Oidor Juárez, L. D. (2013). Una aproximación a un algoritmo clasificador para la identificación de interacciones medicamentosas. En B. U. Puebla, M. J. Somodevilla García, & M. d. Pérez de Celis Herrero (Edits.), *Tratamiento del lenguaje y del conocimiento* (Primera ed., págs. 41-55). España: Bubok Publishing S.L., ISBN: 978-84-686-4292-5.

Contenido

Lista de tablas

Lista de figuras

Introducción

Desde tiempos ancestrales, el ser humano se encuentra en una búsqueda permanente de la preservación de su salud y la curación de sus enfermedades, a partir del inicio de la civilización, en que sus males eran atribuidos a seres malignos y hechos mágicos, hasta nuestros días, en que utilizamos las herramientas que la ciencia y la tecnología nos brindan para la cura de enfermedades. De esta forma, se ha desarrollado un conjunto de técnicas y conocimientos, que conocemos, en su sentido más amplio, como medicina.

Hasta la primera mitad del siglo pasado, el hombre utilizó remedios que, en su gran mayoría, no alteraban de forma importante sus mecanismos fisiológicos. Posteriormente la medicina cambió, introduciendo una inmensa gama de medicamentos, capaces de modificar de manera favorable el curso de las enfermedades y la aparición de síntomas y signos. El papel que juegan los medicamentos en las sociedades actuales es tan relevante que hoy en día, la industria farmacéutica es una de las más dinámicas e importantes para la economía mundial. Tan solo en México, la industria farmacéutica tiene una participación de 1.30% del Producto Interno Bruto (PIB), según datos del Instituto Nacional de Estadística, Geografía e Informática (Caso Prado, 2011).

Ante este panorama, cada medicamento desarrollado debe seguir un minucioso proceso de pruebas para el aseguramiento de la eficacia del mismo, no sólo en términos de su calidad farmacéutica, sino también en función de la gravedad de los efectos secundarios y de las reacciones que estos puedan provocar en el ser humano.

Llegados a este punto, es de gran importancia definir claramente qué es un medicamento y sus diferencias con otros términos con los que comúnmente se confunde: fármaco y droga.

El concepto de droga designa a cualquier sustancia química que cause un efecto al ingresar al organismo humano. Las drogas se clasifican de acuerdo a la fuente de obtención en naturales, sintéticas y semi-sintéticas.

- **Drogas naturales.** Son sustancias de origen mineral, vegetal, animal e incluso del propio cuerpo humano y que son aplicadas a los pacientes sin un proceso previo de modificación química.
- **Drogas sintéticas.** Son aquellas que se fabrican en un laboratorio mediante síntesis química.
- **Drogas semi-sintéticas.** Se refiere a las drogas que se obtienen de una fuente natural y que posteriormente son sometidas a modificaciones químicas.

De acuerdo con los efectos que las drogas tienen sobre el organismo del ser humano al momento de su ingreso, éstas se dividen en fármacos, si estos efectos son mayoritariamente positivos, y venenos en caso contrario.

Una sustancia química cualquiera no es propiamente un fármaco o un veneno, ya que el factor diferenciador reside en la dosis aplicada. Por ejemplo, es posible la intoxicación mediante una sobredosis de ácido acetilsalicílico, que es el principio activo de la aspirina. Por el contrario, una sustancia que usualmente relacionamos con el concepto de veneno, como el arsénico, puede ser utilizada en cantidad pequeñas para procesos curativos.

La gran mayoría de las sustancias que utilizamos como fármacos tienen caracteres organolépticos desagradables para el ser humano. Un carácter organoléptico es cualquier propiedad que pueda ser percibida por los sentidos, como el color, la textura o el sabor. Algunas veces, el organismo humano rechaza de manera inmediata las sustancias que se le suministran, lo que representa un mecanismo biológico de autodefensa.

Con el fin de propiciar el suministro de fármacos al paciente, a éstos se deben añadir otras sustancias. Como resultado de este proceso se obtiene lo que conocemos como medicamento.

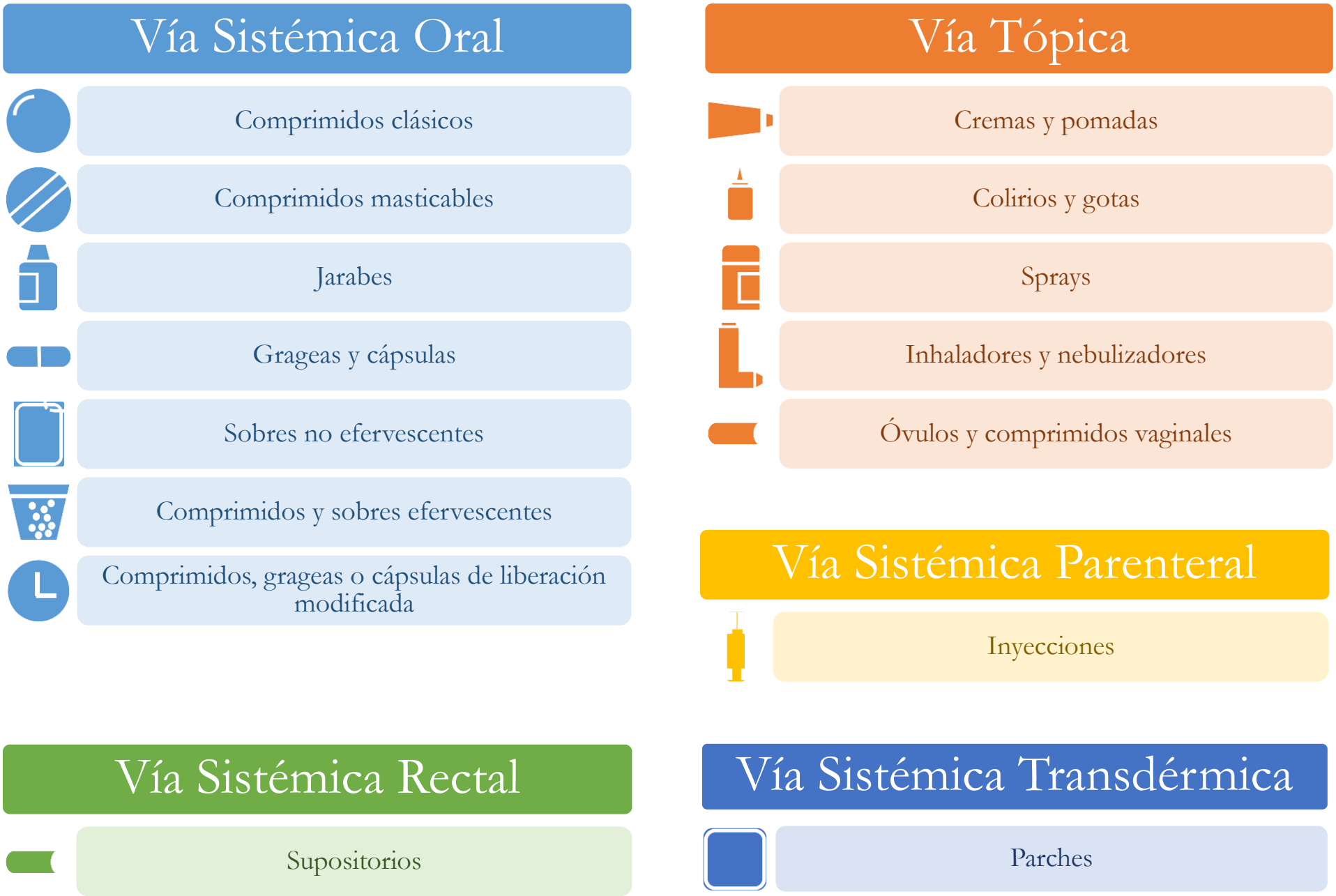


Fig. 1. Formas farmacéuticas más comunes.

Un medicamento es, entonces, una mezcla de un fármaco, que es el principio activo que produce el efecto terapéutico, en conjunto con excipientes y materiales de relleno. Un excipiente es una sustancia que facilita el ingreso y absorción por parte del organismo humano, en tanto que los materiales de relleno cumplen la función de dar una forma farmacéutica adecuada al medicamento.

En la **Fig. 1** se muestran las formas farmacéuticas más comunes (Organización de Consumidores y Usuarios, 2006).

En la práctica de la prescripción médica, la indicación de uso de un solo medicamento es inusual, por lo general un paciente debe administrarse dos o más. Es por ello que los desarrolladores de medicamentos deben realizar también pruebas de combinaciones de estos, de modo que sea posible evitar una interacción que afecte de manera negativa al paciente, ya sea por inhibición de los efectos de uno de ellos, la generación de efectos adversos o el aumento de toxicidad de alguna de las sustancias activas. A esta acción recíproca entre medicamentos se le conoce como **interacción medicamentosa**.

El estudio denominado *Detección de interacciones medicamentosas en el servicio de medicina interna del Hospital General Regional de Orizaba, Veracruz* (Campos Garza, y otros, 2006) obtuvo como resultado una incidencia del 31.87% de interacciones medicamentosas en un total de 342 farmacoterapias. Otro estudio, denominado *Uso de medicamentos, reacciones adversas e interacciones farmacológicas en un hospital obstétrico de Puebla, México* (Brito Barrera & Serrano Martínez, 2011) obtuvo un resultado muy similar, 37.5% de pacientes presentaron una interacción medicamentosa, de un total de 18.

Con la finalidad de evitar que estas interacciones se presenten de forma inesperada en los seres humanos, se han realizado múltiples estudios por parte de laboratorios, centros de investigación y empresas privadas en las áreas de química y medicina.

Pero, como se ha dicho antes, esta industria es una de las más grandes de la economía mundial. ¿Se podría aseverar que el médico conoce (y recuerda) todas y cada una de las posibles interacciones entre medicamentos que pudieran afectar la salud de un paciente? Definitivamente no. A pesar de los grandes avances tecnológicos, la prescripción médica es una actividad que sigue siendo aplicada por el profesional de la salud de forma "manual", a través del uso de los conocimientos generados por la investigación científica en materia farmacéutica, sin asistencia de dispositivos o mecanismos automáticos, como ya sucede en otras áreas de la medicina como los análisis clínicos y el diagnóstico médico.

La identificación de las interacciones medicamentosas es uno de los problemas más complejos dentro de la medicación. Actualmente se han desarrollado procesos que implican la evaluación del perfil farmacoterapéutico con el fin de identificar problemas

relacionados con los medicamentos (idoneidad de la prescripción y conciliación de la medicación), que mediante un proceso de intervención antes de la aplicación se evitan errores de medicación en el paciente hospitalizado.

Sin embargo, aún son numerosas las unidades de salud en las que esta identificación se lleva a cabo de forma posterior a la prescripción, es decir, cuando el paciente manifiesta síntomas relacionados con la modificación de los efectos de un medicamento a causa de otro.

Es necesario, entonces, diseñar métodos basados en el cómputo automático de grandes cantidades de datos, que sirvan como base para el desarrollo de herramientas que provean a los profesionales de la salud y a los desarrolladores de fármacos, apoyo en la identificación de las interacciones medicamentosas. Estos métodos deberán ser diseñados para explotar la información generada por los estudios e investigaciones realizados por los especialistas en la materia, siendo su principal aporte la automatización del procesamiento de información y la aplicación de la ciencia computacional en el desarrollo de soluciones que potencien la actividad humana.

Objetivos

Con el fin de identificar correctamente las posibles interacciones entre los medicamentos prescritos a un paciente determinado, es necesario que el profesional de la salud consulte grandes cantidades de texto, incluidas en publicaciones impresas o digitales, en busca de información relacionada con interacciones reportadas, así como las condiciones en que su ocurrencia es más o menos probable. De esta forma, para cada uno de los medicamentos incluidos en la prescripción, el profesional de la salud debe realizar una consulta de información.

La cantidad de tiempo que es necesaria para llevar a cabo esta actividad, en conjunto con la posibilidad de cometer errores humanos durante la búsqueda e identificación de interacciones hacen que sea inviable de realizar durante una consulta médica.

Objetivo general

El objetivo principal del presente trabajo de tesis es el siguiente:

> *Proponer una aproximación a un método clasificador que permita la identificación de párrafos de texto que enuncien alguna interacción medicamentosa a través de un diseño orientado a que la identificación se lleve a cabo de forma previa a la prescripción médica, con la finalidad de evitar la manifestación de efectos adversos en los pacientes.*

Objetivos específicos

Para alcanzar este objetivo será necesario llevar a cabo algunas tareas que conforman los objetivos específicos del presente trabajo de tesis, las cuales se enumeran a continuación.

1. Realizar una investigación de los métodos actuales para la identificación de interacciones medicamentosas.

2. Identificar las fuentes confiables de información sobre las características, efectos y contraindicaciones de medicamentos.

3. Analizar las ventajas y características generales de los diferentes tipos de clasificadores.

4. Determinar el tipo de clasificador en que se basará el diseño del algoritmo que será producto del proyecto.

Estructura del documento

Para conseguir una aproximación a un algoritmo con las características mencionadas en las secciones previas, se seguirá un proceso de investigación y análisis que será la base de los capítulos que conforman esta tesis.

En el **capítulo 1** se identificarán las fuentes de información sobre medicamentos, así como la forma en que estas exponen las posibles interacciones con otros medicamentos.

En el **capítulo 2** se revisarán las diferentes técnicas de clasificación y los tipos de algoritmos clasificadores, reconociendo sus ventajas y desventajas con la finalidad de elegir, más adelante, el tipo de clasificador idóneo para nuestro objetivo.

El **capítulo 3** discutirá el estado del arte con relación a la identificación de interacciones medicamentosas. En este capítulo se describirán las diferentes estrategias utilizadas en la actualidad.

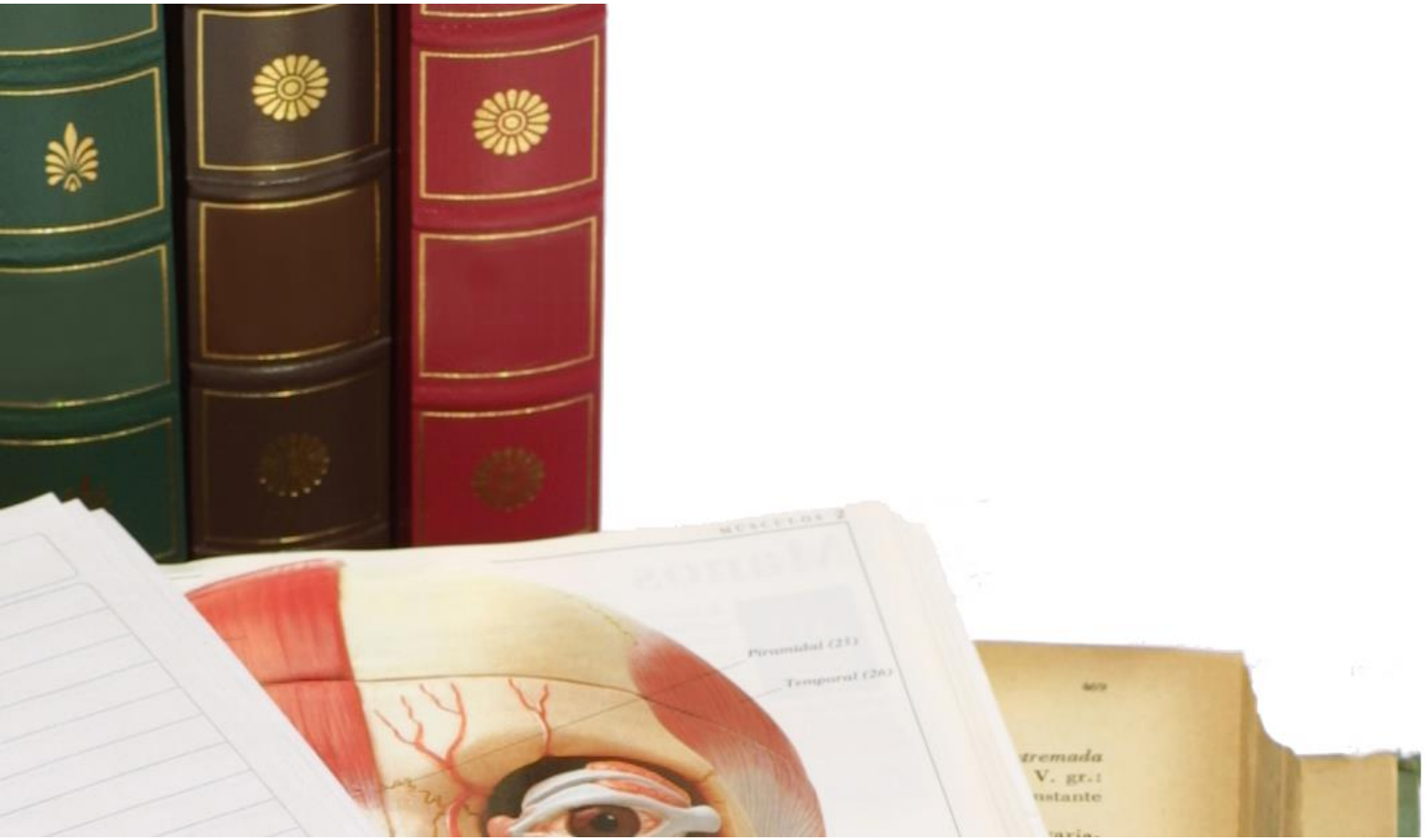
En el **capítulo 4** se expondrán algunas consideraciones relacionadas con el uso del lenguaje natural, mismas que deben tomarse en cuenta durante el desarrollo del presente trabajo.

El **capítulo 5** propondrá la aproximación a un método clasificador que es objetivo del presente trabajo de tesis. Se describirá el proceso de investigación seguido y los resultados esperados.

En el **capítulo 6** se describirán las pruebas realizadas para comprobar la eficiencia del método propuesto.

Por último, en el **capítulo final** se expondrán las conclusiones de la tesis y se propondrá el trabajo futuro que podrá desarrollarse en la misma línea de este trabajo.

Capítulo 1
Fuentes de información sobre medicamentos



En la actualidad, la información es un elemento poderoso y fundamental en el desarrollo de cualquier actividad humana de cualquier índole. Contar con la información correcta y suficiente es un requisito indispensable para el progreso, además de jugar un papel como factor diferencial de los procesos productivos y como componente esencial del bienestar humano.

Por lo regular, los conceptos información y conocimiento se encuentran directamente relacionados.

Cuando hablamos de información, nos referimos a un conjunto de datos relacionados entre sí sobre una materia determinada. La información que poseemos sobre un hecho en particular puede ser calificada, de tal manera que ésta puede ser correcta o incorrecta, según el grado en que describa tal hecho.

Una vez que la información es adquirida y asimilada, decimos que se ha transformado en conocimiento. Con base en los conocimientos es como los seres humanos tenemos la capacidad de tomar decisiones.

De modo que, para generar conocimientos útiles que permitan tomar decisiones adecuadas cuyo resultado sea el esperado, es necesario disponer de información correcta y completa.

En el caso de la medicina, y en específico de la prescripción médica, contar con la información de manera oportuna es determinante para alcanzar sus fines. Toda la

información relacionada con la medicina está incluida en un campo conocido como **información biomédica**.

De acuerdo con (Fundació Víctor Grífols i Lucas, 2007), los cuatro fines de la medicina son:

1. La prevención de enfermedades y lesiones, y la promoción y la conservación de la salud.

2. El alivio del dolor y el sufrimiento causado por males.

3. La atención y el cuidado a los enfermos y los cuidados a los incurables.

4. La evitación de la muerte prematura y la busca de una muerte tranquila.

Con la finalidad de apoyar a la medicina, durante los últimos años, las tecnologías de la información han aportado mecanismos para proveer de información a los profesionales en ciencias médicas, a través de los cuales es posible contar con una mayor cantidad de información en un tiempo menor al requerido por los métodos tradicionales no informatizados.

De esta forma, la ciencia computacional ha colaborado con la ciencia médica en lo relacionado con los procesos clínicos, educacionales, de gestión y de investigación, facilitando el manejo de grandes cantidades de datos y la ejecución de análisis computacional.

Sin embargo, para que la ciencia computacional esté en posibilidad de ofrecer métodos y herramientas que apoyen a alcanzar los fines de la medicina es necesario que ésta última le proporcione fuentes de información confiables que sirvan como base de dichas herramientas.

Gracias a los actuales avances tecnológicos, encontrar información de manera sencilla y fácil no representa problema alguno. Basta realizar una consulta a través de un motor de búsqueda en la web para encontrar miles, tal vez millones, de resultados.

No obstante, la publicación de información en Internet es una actividad altamente asequible, lo que permite que existan múltiples sitios web que no necesariamente cuentan con información correcta. Es por ello que, en la actualidad, el principal problema es seleccionar la información más relevante y de mayor calidad (Rancaño García, y otros, 2003).

1.1. Pirámide de Hynes

Tradicionalmente, las fuentes de información biomédica eran clasificadas en primarias, secundarias y terciarias. Dentro de las fuentes primarias estaban consideradas las revistas, en tanto que las bases de datos conformaban las fuentes secundarias. Las fuentes de información terciarias la constituían los libros.

Ante la creciente complejidad en la clasificación y sistematización de los recursos biomédicos, el esquema tradicional presenta ciertas debilidades.

Como respuesta a este problema, en el año 2001 el profesor *Brian Haynes* propuso un modelo piramidal de clasificación, mejor conocido como **Pirámide de las 4s**, debido a las iniciales, en inglés, de sus cuatro tipos de recursos: s*ystems* (sistemas), *synopses* (sinopsis), *syntheses* (síntesis) y *studies* (estudios).

En 2006, el mismo Haynes incluyó en su modelo un nivel más: *summaries* (sumarios), con lo cual pasó a denominarlo **Pirámide de las 5s**.

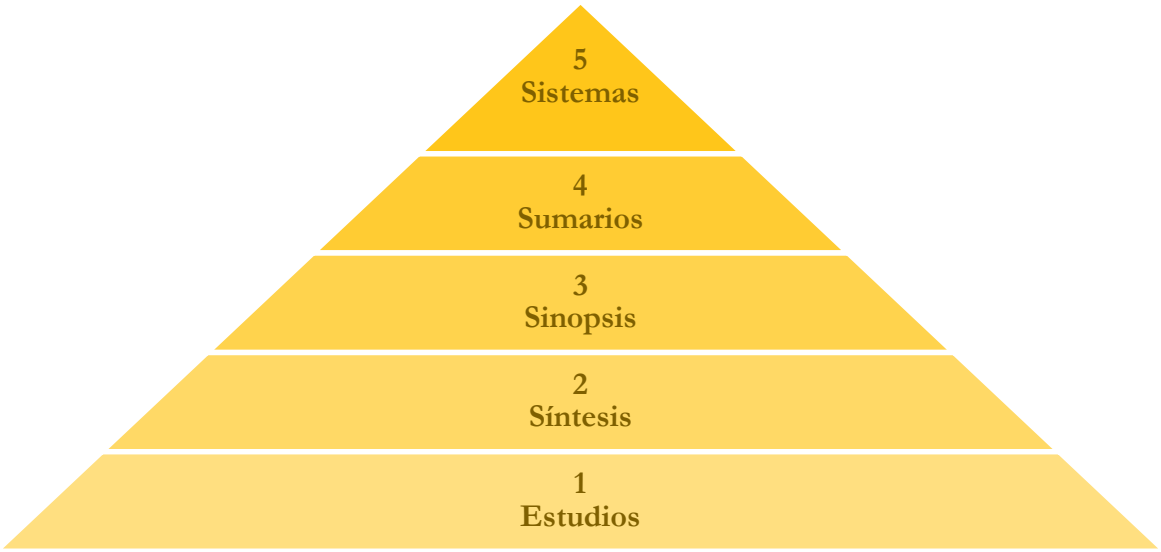


Fig. 2. Modelo piramidal de las 5s de Haynes.

Este modelo considera, en su nivel más bajo, a los **estudios** originales. Estos estudios pasan por revisiones sistemáticas, como la que se realiza con ayuda del *Manual Cochrane* (The Cochrane Collaboration, 2011), lo que da como resultado el segundo nivel de la pirámide, las **síntesis**.

También es posible, a partir de los estudios originales o de las síntesis, realizar descripciones breves, tales como las incluidas en las revistas secundarias, mismos que conforman las **sinopsis**, el tercer nivel del modelo.

En el cuarto nivel se encuentran los **sumarios**, que desarrollan guías de práctica clínica y documentos basados en la evidencia a partir de la integración de la información de mayor calidad de las capas inferiores.

Por último, en la cima de la pirámide se encuentran los **sistemas**, que son soluciones computacionales que implementan herramientas de apoyo a la toma de decisiones y basados en la información de los niveles que se encuentran bajo éste.

Llegados a este punto, algunos profesionales en ciencias médicas cuestionaron la equivalencia de una sinopsis de un estudio original y una sinopsis de una síntesis, dado que aparecen en un solo nivel del modelo de las 5s.

La respuesta a este cuestionamiento llegó en 2011, cuando el estrato de sinopsis se dividió en dos: **sinopsis de estudios** y **sinopsis de síntesis**. De esta forma el modelo pasó a conformarse de seis niveles, motivo por el que es conocido como **Pirámide de las 6s,** que es el que puede observarse en la **Fig. 3**. Este modelo también es conocido como **Pirámide de la Evidencia**.

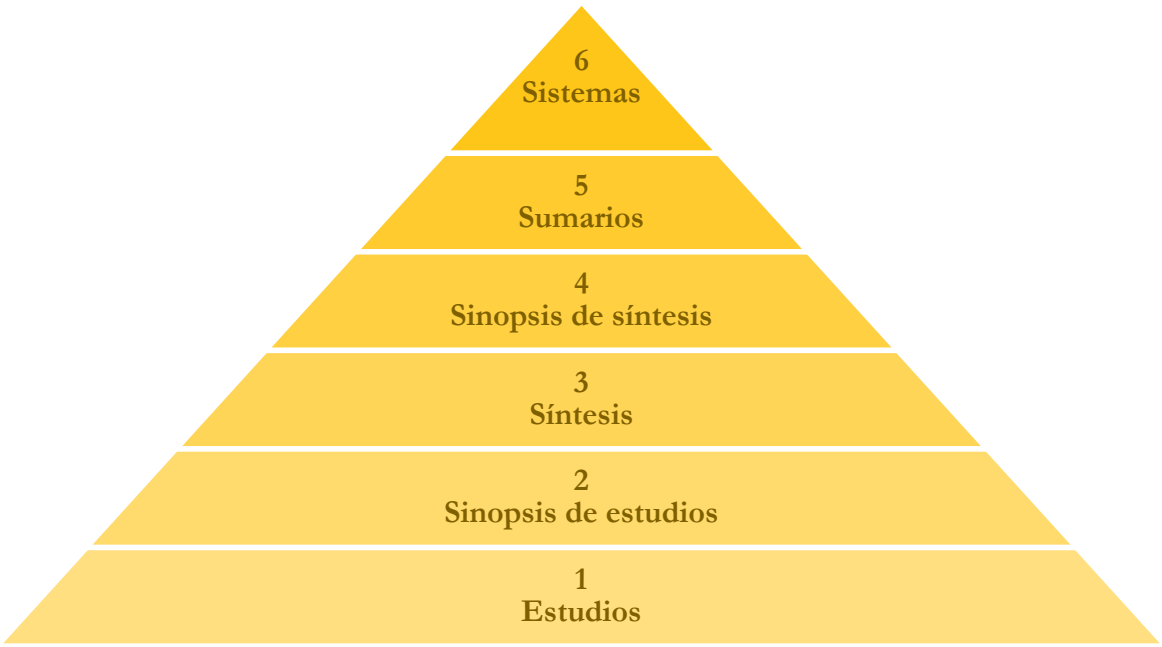


Fig. 3. Modelo piramidal de las 6s de Haynes.

El principal objetivo del modelo piramidal de las 6s es facilitar el uso de los recursos de información denominados pre-evaluados.

A los recursos pre-evaluados se les ha aplicado un filtro previo, de modo que éstos sean seleccionados de entre otros que no cuentan con la calidad requerida para ser considerados como fuentes de información confiables. Adicionalmente, debe considerarse la periodicidad de actualización de dichos recursos, a fin de garantizar que la información a la que se accede se encuentra *al día*.

La forma de utilizar el modelo en la toma de decisiones clínicas, inicia en el nivel más alto de la pirámide, conformado por los **sistemas**. Idealmente, en esta capa se encontrarán sistemas de información que integren y resuman las evidencias obtenidas como resultado del proceso de investigación, con procesos de actualización continua que permita contar con información reciente. Su eficiencia y eficacia como herramientas de soporte a la toma de decisiones han sido demostradas en la práctica clínica, sin embargo, generalmente se trata de herramientas poco asequibles.

Si, para un problema clínico específico, un sistema incorpora de manera adecuada la información basada en la evidencia y representa un apoyo confiable en la práctica, no será necesario ir más abajo en el modelo piramidal de Haynes.

Si no existen sistemas de información o éstos son inadecuados, el siguiente paso es consultar los **sumarios**. En los sumarios se incluye información basada en la evidencia sobre problemas clínicos específicos, que se actualiza de manera constante.

Dentro del nivel de sumarios, se incluyen las Guías de Práctica Clínica (GPC) basadas en la evidencia, que son "recomendaciones desarrolladas sistemáticamente para ayudar a los médicos y pacientes a tomar decisiones sobre la atención sanitaria adecuada en circunstancias clínicas específicas" (Registered Nurses Association of Ontario, 1990).

En caso de que no exista un sumario para el problema clínico, las consecuentes opciones son las **sinopsis** o las **síntesis**.

Una síntesis es una recopilación de una investigación completa sobre un problema clínico específico. El proceso para obtener una síntesis inicia con la formulación de una pregunta, luego se identifican los estudios que contienen información relevante, se valoran de acuerdo a su calidad y se extraen sus resultados. Finalmente se enuncian conclusiones al respecto.

Muchas veces, los médicos no cuentan con el tiempo suficiente para revisar de manera detallada una síntesis. En esos casos puede ser útil consultar una **sinopsis de síntesis** que contenga información resumida y relevante.

Si es necesario contar con información más detallada o no existe una sinopsis relacionada con el problema específico, pueden consultarse las **síntesis**.

En algunas ocasiones, no existen ni sistemas, ni sumarios, ni síntesis, ni sinopsis. Entonces, el siguiente nivel a consultar lo conforman las **sinopsis de estudios** individuales. Las sinopsis con respecto a los estudios individuales son documentos breves pero relevantes, de manera análoga a las sinopsis con respecto a las síntesis.

Las ventajas que tienen las sinopsis, es la seguridad de que el estudio o síntesis base tiene la calidad y relevancia suficiente como para ser resumido, así como la riqueza de los comentarios añadidos.

Finalmente, si no existen documentos de los niveles superiores del modelo de Haynes, la última opción son los **estudios** originales.

El modelo de Haynes es, actualmente, el más aceptado para la clasificación de fuentes de información relacionadas con biomedicina.

1.2. Principales fuentes de información sobre medicamentos

Como parte del proceso de investigación comprendido en el presente trabajo, se identificaron diversas fuentes de información referentes a medicamentos. Estas fuentes están incluidas dentro de las fuentes de información biomédica, por lo que es aplicable el modelo de las 6s de Haynes.

La revisión y evaluación de las fuentes de información es un proceso complejo en el que se deben tomar en cuenta aspectos tales como periodicidad de actualización, idiomas disponibles o institución patrocinadora, entre otras.

Dado que este proceso excede el alcance del presente trabajo, se utilizó como apoyo la información publicada por la **Secretaría de Salud** de México, cuya misión es "contribuir a un desarrollo humano justo incluyente y sustentable, mediante la promoción de la salud como objetivo social compartido y el acceso universal a servicios integrales y de alta calidad que satisfagan las necesidades y respondan a las expectativas de la población" (Secretaría de Salud, 2013).

En el portal de la Secretaría de Salud, en la sección titulada *Obras de consulta* (COFEPRIS, 2014), es posible encontrar diversas fuentes de información biomédica. De acuerdo con este sitio web, la información fue proporcionada por la Comisión Federal para la Protección contra Riesgos Sanitarios, COFEPRIS.

Entre las obras de consulta publicadas por la COFEPRIS en el sitio web de la Secretaría de Salud, se encuentran publicaciones relacionadas con enfermedades, anatomía y química, así como vocabularios y glosarios de uso general en medicina.

En cuanto a lo relacionado con medicamentos, se encuentran algunas publicaciones, entre las que destacan tres productos principales: el **Diccionario de Especialidades Farmacéuticas de Thomson**, la base de datos **Micromedex** y el catálogo **Vademécum**. En las siguientes líneas se describen brevemente estos recursos de información.

1.2.1. Diccionario de Especialidades Farmacéuticas de Thomson

Es mejor conocido como PLM, de las siglas de **Productos de Laboratorios Médicos**. Es una de las fuentes más utilizadas por los médicos y farmacias en México.

Incluye una lista completa de todos los fármacos aprobados por la **Secretaría de Salud** en México, lo que conforma un total de más de 3,500 productos farmacéuticos. Se encuentra disponible en su versión impresa y, gracias a los avances tecnológicos actuales, en versiones digitales en Internet y en forma de aplicaciones para dispositivos móviles.

PLM México tiene a disposición de cualquier persona con acceso a Internet una versión web disponible en la dirección electrónica http://www.medicamentosplm.com/ (PLM México, 2013). Una captura de pantalla se muestra en la **Fig. 4**.

Fig. 4. Página principal del Diccionario de Especialidades Médicas 2013 (PLM).

1.2.2. Micromedex

Micromedex es una base de datos norteamericana que contiene amplia información de medicamentos y sustancias relacionadas, así como de pruebas de laboratorio e interacciones medicamentosas.

Su contenido es actualizado constantemente mediante la revisión sistemática (sinopsis) de estudios médicos originales, siendo así una fuente confiable de información.

Consta de diversos productos, entre los cuales se encuentra DRUGDEX, un sistema de información específico sobre medicamentos, su administración, efectos adversos e interacciones (Truven Health Analytics, 2014). En la **Fig. 5** se muestra una captura de pantalla del sitio web internacional de Micromedex.

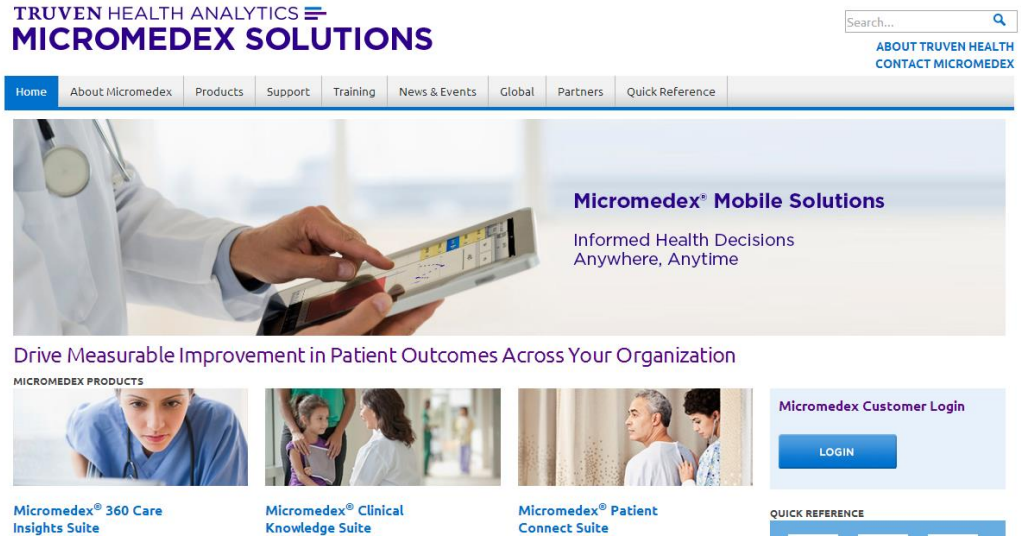


Fig. 5. Página principal del sitio web de Micromedex.

La **Benemérita Universidad Autónoma de Puebla** (BUAP), a través de la **Dirección General de Bibliotecas** (Benemérita Universidad Autónoma de Puebla, 2011), ha puesto a disposición de sus estudiantes, personal docente y colaboradores, una implementación de **Micromedex 2.0** de consulta restringida mediante el número de matrícula de estudiante o número de trabajador.

1.2.3. Vademécum

Vademécum es un catálogo de especialidades, entre las que se encuentra la información relacionada a los medicamentos. Puede encontrarse en versión impresa o en versión digital, que es distribuida por medio de un CD-ROM que contiene todos los archivos necesarios para ser instalado en un equipo de cómputo.

Contiene información relevante relacionada con productos farmacéuticos, y está destinado a profesionales en ciencias médicas, tales como médicos y farmacéuticos (Informed, S.A. de C.V., 2014).

En la **Fig. 6** puede observarse una captura de pantalla del sitio web de Vademécum en México.

Fig. 6. Sitio web de PR Vademécum en México.

Existen otras muchas fuentes de información, sin embargo las ya mencionadas son relevantes para el presente trabajo dada su alta aceptación y amplio uso por parte de los profesionales en ciencias médicas.

La identificación de estas fuentes como fiables y generalmente aceptadas es un paso fundamental en la elaboración de un método para la identificación de interacciones medicamentosas. Un método basado en información de calidad estará en condiciones de proporcionar resultados de calidad, como los que son requeridos en la práctica de la prescripción médica.

Para el desarrollo del presente trabajo, se utilizó el **Diccionario de Especialidades Médicas** como fuente de información, debido a su disponibilidad de manera libre y gratuita, así como por ser aceptado por la COFEPRIS como recurso de consulta de información.

1.3. Sistema de Clasificación ATC

Como ya se ha mencionado, la industria farmacéutica es una de las más grandes de la economía mundial, gracias a lo que, en la actualidad, el ser humano dispone de una amplia diversidad de medicamentos para su uso. Esta diversidad en los productos farmacéuticos existentes implica una creciente complejidad para obtener información sobre el uso de los medicamentos.

Ante ello, la **Organización Mundial de la Salud** instituyó el **Sistema de Clasificación Anatómica, Terapéutica y Química**, también conocido como **Sistema de Clasificación ATC** (Consejo Nacional de Salud, 2010). En este sistema, los principios activos se dividen en grupos con cinco diferentes niveles.

El primer nivel de clasificación está determinado por el **Grupo Anatómico**, es decir, el órgano, sistema o aparato del cuerpo humano sobre el que actúa, y es representando por una letra mayúscula. En total, existen catorce Grupos Anatómicos, que se muestran en la **Tabla 1**.

Grupos Anatómicos	
A	Tracto gastrointestinal y metabolismo
B	Sangre y órganos hematopoyéticos
C	Sistema cardiovascular
D	Dermatológicos
G	Aparato genito-urinario y hormonas sexuales
H	Hormonas sistémicas
J	Antiinfecciosos sistémicos en general
K	Soluciones de uso hospitalario
L	Antineoplásicos e inmunomoduladores
M	Sistema musculoesquelético
N	Sistema nervioso
P	Antiparasitarios
R	Aparato respiratorio
S	Órganos de los sentidos
V	Varios

Tabla 1. Grupos Anatómicos de la Clasificación ATC.

El segundo nivel lo conforman los **Grupos Terapéuticos Principales**, representados por dos caracteres numéricos. El tercer nivel está conformado por los **Grupos**

Farmacológicos, en tanto que el cuarto nivel está integrado por los **Grupos Químicos**. Ambos niveles son representados por una letra mayúscula, cada uno.

Finalmente, en el quinto nivel se encuentran las **Sustancias Químicas**, entre las que se encuentran los medicamentos. Estas sustancias son identificadas mediante dos dígitos numéricos.

Cada medicamento tiene asociado un código que lo identifica y que proporciona información sobre el grupo de sustancias al que pertenece. En la **Fig. 7** se muestra la estructura de un **Código ATC**.

Fig. 7. Código ATC del *ketorolaco*.

De esta manera, el **Código ATC** del *ketorolaco* es **M01AB15**.

1.3.1. Principios generales de la Clasificación ATC

La clasificación de los medicamentos mediante el sistema ATC se realiza de acuerdo con el uso terapéutico principal. De esta forma, el principio básico de la clasificación es asignar un solo **Código ATC** para cada vía de administración de cada medicamento. Esto es, las formas farmacéuticas con ingredientes y concentraciones similares tendrán un mismo **Código ATC**.

Por otro lado, un mismo medicamento puede tener más de un **Código ATC** si se encuentra disponible en más de una vía de administración con usos terapéuticos diferentes. En la **Tabla 2** se muestran algunos ejemplos de medicamentos con diferentes **Códigos ATC**.

Medicamento	Código 1		Código 2	
Metronidazol	D06BX01	Antiinfeccioso y antiséptico ginecológico	P01AB01	Agente contra la amebiasis y otras enfermedades por protozoarios
Acetilcisteína	R05CB01	Mucolítico	V03AB23	Antídoto
Aciclovir	J05AB01	Sistémico	S01AD03	Oftálmico

Tabla 2. Ejemplos de medicamentos con más de un Código ATC.

Un caso especial es la *prednisolona*, que en los productos farmacéuticos de un solo ingrediente tiene asignados varios **Códigos ATC**, debido a sus diferentes usos terapéuticos y fórmulas de aplicación. Este ejemplo se muestra en la **Tabla 3**.

Códigos ATC de la *prednisolona*	
A07EA01	Agente antiinflamatorio intestinal
C05AA04	Antihemorroidal de uso tópico (supositorio)
D07AA03	Preparado dermatológico
H02AB06	Corticosteroide para uso sistémico
R01AD02	Descongestivo nasal
S01BA04	Oftalmológico (gotas para los ojos)
S02BA03	Otológico (gotas para los oídos)

Tabla 3. Códigos ATC de la *prednisolona*.

Existen otras consideraciones sobre la **Clasificación ATC**. Por ejemplo, un mismo medicamento puede estar indicado para más de un uso, en tanto que el uso terapéutico de un medicamento puede variar entre países (Instituto Hondureño de Seguridad Social, 2009). Por otro lado, las sustancias que pertenecen a un mismo **Grupo Químico** no deben considerarse farmacoterapéuticamente equivalentes, ya que su modo de acción, efecto terapéutico, interacciones medicamentosas y efectos adversos pueden diferir.

Debido a estas variaciones, el **Sistema ATC** no es un sistema estricto de clasificación terapéutica.

Capítulo 2
Técnicas de clasificación

Toda la información biomédica, incluyendo la relacionada con los medicamentos, se encuentra disponible para ser consultada por los profesionales de ciencias médicas. No obstante, el proceso de consulta, lectura y análisis de esa información puede ser muy complejo.

Es, entonces, necesario que la ciencia computacional proporcione mecanismos de apoyo para la actividad de consulta de información relevante.

El proceso de búsqueda de información relevante puede expresarse, en su forma más general, como un problema de clasificación. En este tipo de problemas se accede a grandes cantidades de datos de entre los cuales se elige, bajo ciertos criterios establecidos de manera previa, a aquellos que cumplan con las características o atributos distintivos que se buscan. Los procesos de comprensión de esos atributos son denominados **procesos perceptuales**, en tanto que la selección de estos objetos es llamada **reconocimiento de patrones**.

Un patrón es una colección de descriptores que cumplen la función de representar a un objeto determinado.

En un problema típico de clasificación, se extraen las características del objeto, también denominado **instancia del problema**, con el fin de reconocer su patrón. Luego, este objeto es asignado a una categoría específica de acuerdo con el patrón que fue identificado. A esta categoría se le denomina **clase**.

2.1. Enfoques del reconocimiento de patrones

En la actualidad, no existe una teoría unificada de reconocimiento de patrones (Duda, Hart, & Stork, 1997). A lo largo de su corta historia, esta disciplina ha visto surgir diversos puntos de vista en lo referente a las herramientas, teorías y métodos que deben ser utilizados.

Los puntos de vista de las diferentes corrientes de pensamiento establecen formas diferentes de abordar los problemas de clasificación, y pueden agruparse en cuatro grandes enfoques:

1. El enfoque estadístico.
2. El enfoque neuro-reticular.
3. El enfoque sintáctico-estructural.
4. El enfoque lógico-combinatorio.

A continuación se explican brevemente estos cuatro enfoques.

2.1.1 El enfoque estadístico

La estadística ha sido, tradicionalmente, una de las principales herramientas utilizadas en el reconocimiento de patrones. La *Teoría Bayesiana de la Decisión*, el *Análisis Discriminante* y el *Análisis de Agrupamientos* son algunos de los fundamentos teóricos que han motivado esta relación.

A través de un **enfoque estadístico**, cada patrón es representando como un vector conformado por los valores resultantes del muestreo y cuantificación de las instancias del problema. Por su parte, una clase se puede conformar por uno o varios patrones base o prototipos.

De esta manera, un patrón es solamente un punto en un espacio cuya cantidad de dimensiones está determinado por la cantidad de variables consideradas (Bisquerra Alzina, 1989). A este espacio se le denomina **espacio de representación de los patrones**.

Con base en el enfoque estadístico se ha constituido el **reconocimiento estadístico de patrones**, el cual se basa en descripciones de objetos en términos de variables numéricas obtenidas a partir de la caracterización de sus mediciones. Se presupone que estas variables están definidas sobre un espacio vectorial normado (Iribarren, 1973).

La aplicación de este enfoque se ha extendido ampliamente, incluso hacia problemas en los que su uso no es recomendado.

Este enfoque puede ser implementado a través de diversas técnicas, entre las cuales destaca la de **n-gramas**, que será abordada más adelante en este mismo capítulo.

2.1.2. El enfoque neuro-reticular

El **enfoque neuro-reticular** es un enfoque más moderno, que tiene fundamento en la **teoría conexionista** que, a su vez, se fundamenta en el ámbito de la neuropsicología (Garson, 1997).

En este enfoque se utiliza una estructura de neuronas artificiales interconectadas denominada **red neuronal**. Estas neuronas se estimulan unas a otras y pueden ser *entrenadas* para obtener de ellas respuestas específicas ante estímulos predeterminados, como imitación del funcionamiento de las redes neuronales biológicas.

La principal ventaja de utilizar el enfoque neuro-reticular radica en la posibilidad de separar regiones no lineales de decisión complejas, de acuerdo con la cantidad de neuronas y capas.

Es por ello que las redes neuronales artificiales usualmente se aplican para resolver problemas de reconocimiento de patrones de alta complejidad.

2.1.3. El enfoque sintáctico-estructural

Si bien el reconocimiento estadístico de patrones es el más simple de los enfoques, esto es en gran parte debido a que no considera el contexto del patrón a reconocer, es decir, la relación existente entre ese patrón y otros.

Existen patrones que pueden descomponerse gradualmente en patrones cada vez más simples hasta llegar a componentes básicos. Cuando se toma en cuenta esta información de contexto estamos aplicando un enfoque sintáctico-estructural.

En el **reconocimiento sintáctico de patrones**, cada patrón es descrito a partir de sus elementos básicos y un conjunto de reglas sintácticas.

2.1.4. El enfoque lógico-combinatorio

Básicamente, el **enfoque lógico-combinatorio** parte del supuesto de que las instancias del problema que se desean identificar son descritas a través de una combinación de rasgos numéricos y no numéricos, cuyos valores pueden ser procesados por funciones numéricas. Este enfoque tiene fundamento teórico en la *Lógica Matemática*, la *Teoría Clásica de Conjuntos*, la *Teoría de los Subconjuntos Difusos*, la *Teoría Combinatoria* y, en general, la *Matemática Discreta*.

Una característica primordial del enfoque es que, de acuerdo a sus principios, el modelo de la instancia debe ser lo más cercano posible a la realidad. Esto evita la posibilidad de hacer suposiciones que no estén debidamente fundamentadas.

El **reconocimiento lógico de patrones** es una filosofía para afrontar los problemas, más que un conjunto de técnicas.

2.2. Modelo de n-gramas

Un **n-grama** es una secuencia de elementos dentro de un conjunto más grande, en este caso, de texto. La formación de este tipo de secuencias es la base del denominado modelo de n-gramas, una de las técnicas que implementan el enfoque estadístico del reconocimiento de patrones. Este modelo asume que, en un texto, la aparición de unas pocas secuencias condiciona la probabilidad de la siguiente secuencia.

La cantidad de elementos que serán considerados está determinada por el valor de n. Así, cuando n es igual a 1 se habla de **unigramas**, cuando n es igual a 2, se denominan **bigramas** y **trigramas** en aquellos casos en que n es igual a 3.

Los elementos que conforman los **n-gramas** pueden ser caracteres, palabras o cualquier otro elemento perfectamente identificable. Supóngase, por ejemplo, la siguiente frase

> *Coloca el dinero sobre la mesa.*

Si hacemos a n igual a 3, algunos posibles **trigramas** de la frase anterior serán:

1. Coloca/el/dinero
2. el/dinero/sobre
3. dinero/sobre/la
4. sobre/la/mesa

Los valores de n utilizados en la mayoría de sistemas varían entre 2 y 7, siendo 3 el más común (Charniak, 1993).

Los n-gramas de palabras suelen ser aplicados en los casos en que el dominio está limitado al léxico del texto base. No obstante, ésta técnica puede encontrarse con problemas, como el tratamiento de secuencias no encontradas en el texto base, ya que siempre existirán secuencias para las que no existe una probabilidad calculada o cuya probabilidad es cero. Para suavizar estas probabilidades, suele realizarse un cálculo de n-gramas con valores menores a n y mayores a cero.

El modelo de n-gramas está siendo, cada vez, más empleado en el reconocimiento de patrones (McNamee & Mayfield, 2004), debido a su simplicidad, eficiencia y robustez, además de mostrar independencia del idioma y del dominio.

2.3. Definición formal de clasificación

Las técnicas de reconocimiento de patrones sirven para identificar las instancias de un problema, que una vez identificadas se les asigna una categoría o clase determinada por sus características, es decir, *se clasifican*.

Se puede formalizar la **clasificación** como la aproximación de una función objetivo no conocida que describe la forma en la que instancias del problema deben ser clasificadas, mediante otra función, que es denominada **clasificador**. La función objetivo se puede representar como en **(1)**.

$$\Phi: I \; x \; C \; \rightarrow \; \{T, F\} \qquad \textbf{(1)}$$

Por otro lado, la función denominada **clasificador** se encuentra representada en la forma mostrada en **(2)**.

$$\Theta: I \; x \; C \; \rightarrow \{T, F\} \qquad \textbf{(2)}$$

En ambas formas, C es un conjunto predefinido de clases, en tanto que I es un conjunto de instancias del problema. Es común representar cada instancia $i_j \in I$ como una lista $A = \{a_1, a_2, ..., a_{|A|}\}$ de valores característicos, denominados **atributos**. Si Φ: i_j x $c_i \rightarrow T$, entonces i_j es un ejemplo positivo de la clase c_i. Si, por el contrario, Φ: i_j x $c_i \rightarrow F$, entonces i_j es un ejemplo negativo de c_i.

2.4. Clasificación supervisada y no supervisada

Es posible abordar el problema de la clasificación desde dos perspectivas completamente diferentes: la **clasificación supervisada** y la **clasificación no supervisada.** La principal diferencia entre ambos enfoques es la presencia de la variable dependiente, es decir, el conocimiento previo de la clase real de cada instancia.

A continuación se explican cada una de estas perspectivas de clasificación.

2.4.1. Clasificación supervisada

De manera provisional, podemos denominar como supervisado a un problema de clasificación en el que existe un conocimiento previo de las clases o categorías en que es posible clasificar instancias y que, además, cada clase contenga, al menos, un patrón ya clasificado.

En la **Fig. 8** se puede observar un diagrama conceptual de una clasificación supervisada típica.

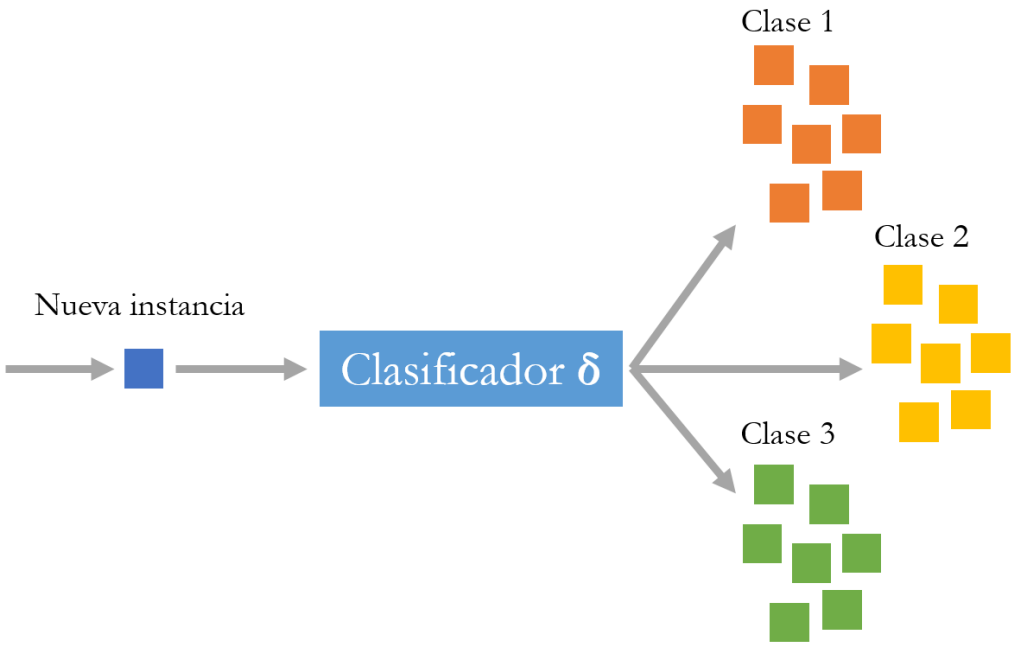


Fig. 8. Diagrama conceptual de la clasificación supervisada.

El proceso de clasificación supervisada implica la observación de los atributos de un conjunto de instancias clasificadas previamente, de modo que sea posible asignar una instancia no clasificada en una determinada categoría. El principal requisito para la construcción del clasificador es contar con una colección Ω, que es denominada *conjunto de entrenamiento*, de ejemplos tales que el valor de la función $\Phi(i_j, c_i)$ sea conocido para cada $(i_j, c_i) \in \Omega \times C$.

2.4.2. Clasificación no supervisada

A diferencia de la clasificación supervisada, los métodos de clasificación no supervisada no disponen de un conjunto de entrenamiento que caracterice las posibles clases de las instancias a clasificar. En su lugar, la clasificación no supervisada utiliza **algoritmos de agrupamiento** con el fin de construir este conjunto de clases. Al no existir un conjunto de entrenamiento, no se cuenta con conocimiento sobre descriptores de instancias clasificadas y, por tanto, no existen patrones predefinidos.

Los algoritmos de agrupamiento llevan a cabo un proceso de análisis previo a la clasificación cuyo objetivo es agrupar las instancias de forma tal que las instancias que pertenezcan a un mismo grupo posean un alto grado de semejanza, en tanto que, las instancias que pertenezcan a grupos diferentes se asemejen en un grado menor (Michie, Spiegelhalter, & Taylor, 2009).

La clasificación no supervisada es de gran importancia en aquellos problemas en los que no se cuenta con instancias previamente clasificadas, en los que el costo de clasificación por parte de un experto es alto y cuando los patrones existentes son susceptibles al cambio en el transcurso del tiempo.

Ambas perspectivas de clasificación, tanto la supervisada como la no supervisada, pueden ser aplicadas en conjunto. Por ejemplo, es posible aplicar la clasificación no supervisada con la finalidad de obtener un conjunto de entrenamiento, que luego puede ser utilizado para realizar una clasificación supervisada.

Existe, también, un tercer enfoque que en realidad es un híbrido entre el supervisado y el no supervisado, que fue nombrado como **semi supervisado**. Esto es, un problema en que el conocimiento sobre las posibles clases es parcial, es decir, no se conoce a todas las clases en que podemos dividir las instancias del problema.

2.5. Representación de un documento

Cuando se desea clasificar texto de manera automática a partir de un conjunto de entrenamiento, el resultado obtenido será una representación tipificada, susceptible de ser clasificada a través de diversos métodos, tales como un modelo vectorial o el modelado de espacio de palabras.

Sea d_j un documento dado, este puede ser representado por un vector $\overrightarrow{d_j} = (p_{1j} \dots p_{rj})$, donde t representa los términos y r el número total de términos presentes en el documento. Usualmente r es el número resultante después de haber excluido a las **palabras funcionales**, también denominadas **palabras vacías** (Aas & Eikvil, 1999).

Otra forma de realizar esta representación es mediante la contabilización de palabras de manera que se pueda asignarle un peso a cada término específico (Cornejo Aparicio, 2014). Para ello existen varias formas, entre las que se encuentran:

1. Ponderado booleano
2. Ponderado por frecuencia
3. Ponderado tf-idf

2.5.1. Ponderado booleano

En el tipo de ponderación booleana sólo son aceptados dos valores, uno (1) o cero (0), para representar la presencia de cada término en el documento. Esto es representado de la siguiente manera:

$$t_{ij} = \begin{cases} 1 \; si \; aparece \; en \; d_j \\ 0 \; en \; caso \; contrario \end{cases}$$

2.5.2. Ponderado por frecuencia

Cuando se pondera por frecuencia se contabiliza el número de veces que un término i aparece en el documento d_j, esto es, la frecuencia de aparición, representada por f_{ij}.

$$t_{ij} = f_{ij}$$

2.5.3. Ponderado tf-idf

Para realizar el ponderado tf-idf, se asigna el peso del término i en el documento j, en proporción al número de ocurrencias del término en el documento y en proporción inversa al número de documentos de la colección para los que existe al menos una ocurrencia para dicho término.

$$t_{ij} = f_{ij} * \log\frac{N}{n_i}$$

Donde N representa el número de documentos y n_i el número de documentos donde aparece el término.

Capítulo 3
Estado del arte

Las **interacciones medicamentosas** representan uno de los problemas más complejos relacionados con la administración de medicamentos a los pacientes (Iniesta Navalón, Urbieta Sanz, & Gascón Cánovas, 2011). Existen diversos motivos por los que esta actividad puede entregar resultados no deseados, entre los que se encuentran la alta demanda de los servicios médicos, la automedicación y la subestimación de los efectos adversos que pueden tener los medicamentos en el organismo del ser humano.

No obstante, las interacciones medicamentosas son sólo uno de los problemas relacionados con medicamentos de los que se encarga de atender la **farmacovigilancia**.

En primera instancia, es importante entender qué es y qué pretende la farmacovigilancia. De acuerdo con (Calderón Ospina & Urbina Bonilla, 2010) es "la disciplina encargada de la detección, evaluación, entendimiento y prevención de los efectos adversos y de cualquier otro problema relacionado con medicamentos".

Ante la relevancia del tema, diversos estudios se han realizado al respecto, obteniéndose así algunos métodos que buscan identificar las **reacciones adversas a medicamentos**, RAM por sus siglas, una vez que éstas se han manifestado en el paciente. Su principal objetivo es determinar si un medicamento específico está provocando la sintomatología al paciente y si éste debe suspender su administración.

Aunque la mayoría de las reacciones adversas no ponen en riesgo la vida del paciente, el profesional de ciencias médicas se enfrenta a una amplia variedad de efectos que pueden desviar su atención de reacciones más relevantes.

A continuación, se mencionan algunos de los algoritmos más importantes de identificación de reacciones adversas, incluidas las provocadas por interacción de medicamentos.

3.1. Algoritmo de Karch y Lasagna

El **Algoritmo de Karch y Lasagna**, a pesar de haber sido publicado en 1977, sigue siendo un estándar para la identificación de los efectos adversos de medicamentos, entre los que existen combinaciones de dos o más fármacos.

Contempla la secuencia temporal entre el cuadro clínico que presenta un paciente y los medicamentos que se presume son responsables de dicha sintomatología, mediante la evaluación de la relación causa-efecto. Esta relación puede clasificarse como *Definida*, *Probable*, *Posible* o *Condicional* (Armijo & González, 2001).

En la **Tabla 4** se pueden observar las cuestiones básicas que conforman el **Algoritmo de Karch y Lasagna**.

	Variación de la relación causal			
Criterio	Definida	Probable	Posible	Condicional
Secuencia temporal	Sí	Sí	Sí	Sí
Respuesta al fármaco conocida	Sí	Sí	Sí	No
Presencia de una explicación alternativa para la reacción	No	No	Sí	No
Mejora al retirar el medicamento	Sí	Sí	Sí o no	Sí o no
Reaparece al reintroducirlo	Sí	¿?	¿?	¿?

Tabla 4. Algoritmo de Karch y Lasagna.

En uno de los casos de éxito más conocidos, en España el **Ministerio de Salubridad**, a través del **Sistema Español de Farmacovigilancia**, lo ha implementado, aunque con modificaciones realizadas para asignar valores numéricos con la finalidad de cuantificar el resultado. El modelo ha sido seguido por órganos de otros países, tal es el caso de Perú, en donde el **Ministerio de Salud** recomienda la aplicación del **Algoritmo de Karch y Lasagna** modificado (Dirección General de Medicamentos, Insumos y Drogas, 2000).

3.2. Algoritmo de Kramer

El **Algoritmo de Kramer** fue publicado 1979 por *Michael S. Kramer*. Básicamente, consiste en una secuencia de preguntas y una escala de calificación que permite, al final de la aplicación del cuestionario, establecer la causalidad por categorías.

Este método consta de cincuenta y seis preguntas dicotómicas, es decir, que tienen dos posibles respuestas, en este caso *sí* o *no*. Es también un algoritmo diseñado para determinar si una reacción fue generada por un medicamento o un conjunto de medicamentos en específico (Kramer, Leventhal, Hutchinson, & Feinstein, 1979).

3.3. Algoritmo de Naranjo y colaboradores

El **Algoritmo de Naranjo y coaboradores** tiene fundamento en el de **Karch y Lasagna** y consta de un cuestionario similar al **algoritmo de Kramer**, pero de menor cantidad de cuestiones, solamente diez preguntas dicotómicas.

	Sí	No	No se sabe
1. ¿Hay informes previos concluyentes sobre ésta reacción?	+1	0	0
2. ¿El evento adverso apareció cuando se administró el medicamento sospechoso?	+2	-1	0
3. ¿La reacción mejoró cuando se suspendió el medicamento o se administró un antagonista?	+1	0	0
4. ¿Reapareció la reacción adversa cuando se volvió a administrar el medicamento?	+2	-1	0
5. ¿Hay causas alternativas que pudieron, por si solas, haber causado la reacción?	-1	+2	0
6. ¿Reapareció la reacción cuando se administró un placebo?	-1	+1	0
7. ¿El medicamento se detectó en la sangre (u otro fluido) en concentraciones tóxicas?	+1	0	0
8. ¿La reacción fue más severa cuando se aumentó la dosis o menos severa cuando se disminuyó?	+1	0	0
9. ¿El paciente ha tenido una reacción similar con el mismo medicamento u otros similares?	+1	0	0
10. ¿El evento adverso fue confirmado por medio de una evidencia objetiva?	+1	0	0
Total			

Tabla 5. Algoritmo de Naranjo y colaboradores.

Publicado en 1981, el **Algoritmo de Naranjo y colaboradores**, al igual que los algoritmos anteriores, no fue diseñado específicamente para determinar la interacción entre fármacos, sino de relacionar un efecto adverso con su causal. Resulta eficaz dada su simplicidad y su corta extensión (Naranjo, Busto, & Sellers, 1981).

En la **Tabla 5** se muestran las diez preguntas que conforman este algoritmo junto con las ponderaciones asignadas a cada respuesta posible. Con base en el puntaje total obtenido es posible evaluar la causalidad del efecto adverso, tal como se muestra en la **Tabla 6**.

Categoría	Puntaje total
Probada	>9
Posible	5 – 8
Dudosa	1 – 4
Incondicional	≤ 0

Tabla 6. Causalidad según el Algoritmo de Naranjo y colaboradores.

La particularidad que comparten los tres métodos antes mencionados es que su objetivo es identificar de manera general la causa de un efecto adverso en un paciente, lo que hace pensar en ellos como métodos *a posteriori* con relación a la prescripción médica. Esto implica que una interacción medicamentosa deberá presentarse al menos una vez para poder ser evaluada e identificada como potencialmente negativa.

3.4. Sistemas de consulta

Durante las últimas décadas, la ciencia computacional ha experimentado un vertiginoso desarrollo, con rápidos y continuos cambios que han producido, entre otras herramientas, computadoras de propósito general con mayor procesamiento de cómputo y movilidad.

No hay duda de que la ciencia computacional se ha convertido en una herramienta fundamental en la resolución de problemas de toda índole.

Como parte de esta incorporación de las nuevas tecnologías a ámbitos de la actividad humana, la medicina ha encontrado en la computación una herramienta poderosa para la agilización y mejoramiento de los procesos de apoyo biomédico. La lista de usos de la computación en el campo de la medicina es amplia.

Una de las principales aportaciones de la computación a la medicina lo representan la detección e identificación de alteraciones, por ejemplo, la tomografía axial

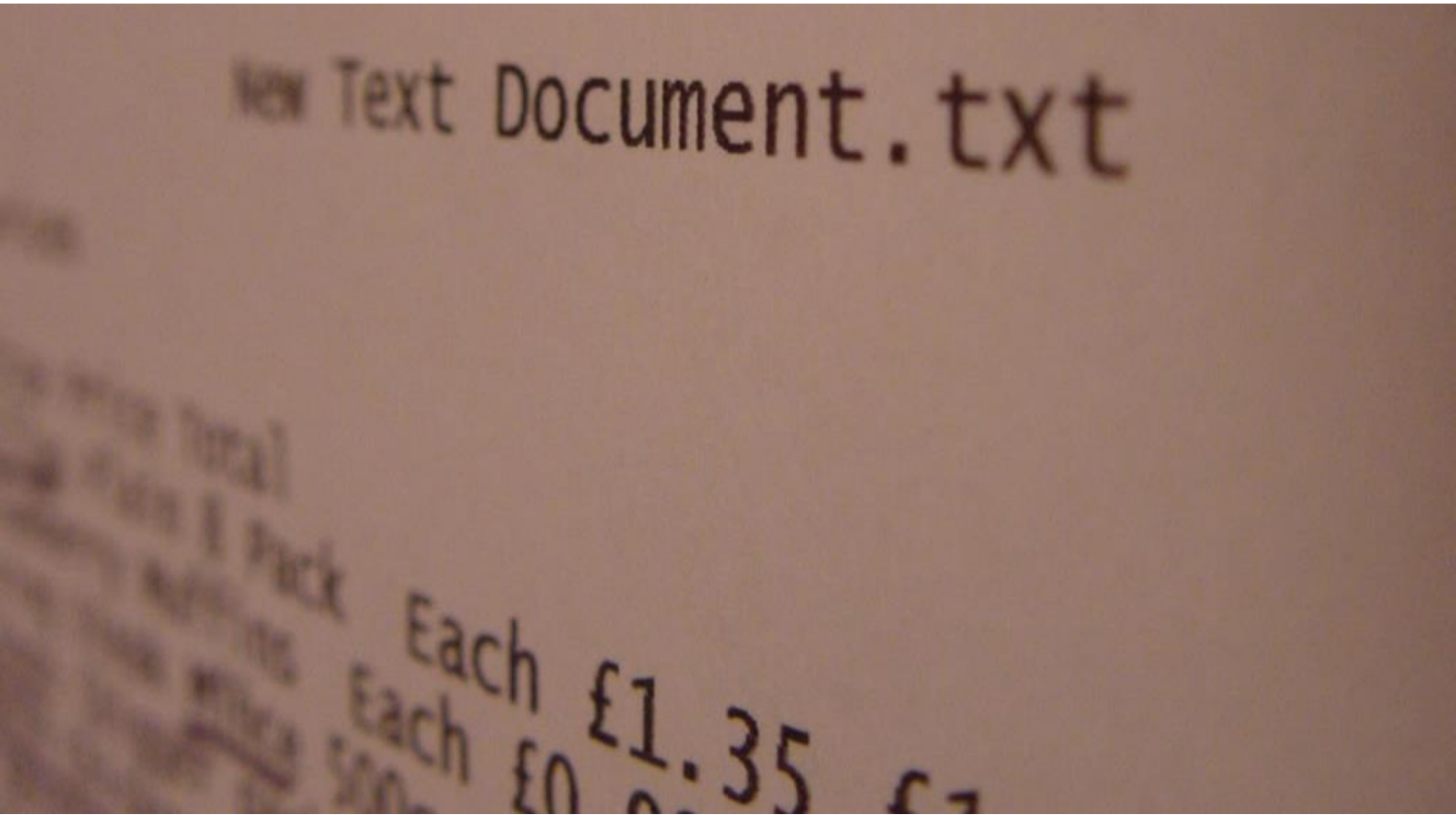
computarizada o el monitoreo de procesos fisiológicos (Freer Bustamante & Chavarría Cerdas, 1992).

Las líneas de proceso en los laboratorios también han sido automatizadas permitiendo así la transferencia y procesamiento de información a una mayor velocidad.

Por otro lado, existen sistemas de consulta que contienen información relacionada con medicamentos, a través de los cuales es posible realizar búsquedas a partir de ciertos términos. De esta manera un profesional de ciencias médicas puede encontrar información sobre interacciones entre un medicamento específico y otras sustancias. La desventaja que presenta éste método es que, una vez realizada la búsqueda mediante el sistema, el resultado presentado contendrá una gran cantidad de texto entre la que se debe buscar la porción que es de interés.

También existen sistemas de consulta en los que es posible indicar un conjunto de medicamentos y obtener como resultado las posibles interacciones entre ellos, sin embargo, como se tratará más adelante en este trabajo de tesis, estos sistemas requieren de un gran trabajo de intervención para mantener actualizada la información.

Capítulo 4
Consideraciones sobre el lenguaje natural



El origen de la medicina se remonta a épocas lejanas en el tiempo, incluso previas a la invención de la escritura. Cuando el ser humano aprendió a comunicarse mediante lenguaje escrito, la ciencia médica encontró un excelente medio para registrar y comunicar sus descubrimientos.

Desde ese momento y hasta la actualidad, se han generado grandes cantidades de información biomédica que hoy conforman el gran volumen de texto disponible en todos los tipos de fuentes de información, como los incluidos en la **Pirámide de las 6s de Haynes**, estudiada en el **Capítulo 1**.

El ser humano cuenta con una sorprendente capacidad para descifrar e interpretar los símbolos que conforman un texto, pero con un costo importante de tiempo y atención.

En el caso de la prescripción médica, se han desarrollado procesos que implican la evaluación del perfil farmacoterapéutico con el fin de identificar problemas relacionados con los medicamentos. Uno de estos procesos es el conocido como **evaluación de idoneidad de la prescripción**.

Un procedimiento típico para evaluar la idoneidad de la prescripción verifica la frecuencia de administración del medicamento, la duplicidad farmacéutica, posibles alergias y sensibilidades e interacciones medicamentosas, entre otras (Morales Bustamante, 2012).

La forma más común de realizar estas y otras verificaciones es mediante la consulta directa en textos médicos, aun cuando se realice mediante el uso de sistemas

informáticos. Sin embargo, como ya se ha mencionado, la lectura e interpretación de textos es una tarea costosa en términos de tiempo.

Una alternativa de solución a este problema lo representan las técnicas de clasificación automática de texto, cuyo objetivo es asignar documentos de texto, con base en su contenido, a una o varias clases predefinidas.

Sin embargo, esta tarea no es sencilla. Los textos que contienen información biomédica están codificados en lo que se denomina **lenguaje natural**, que es el que utilizan los seres humanos para propósitos generales de comunicación. Algunas de las propiedades del lenguaje natural dificultan la clasificación de textos, por ejemplo, cuando existen diferentes formas de expresar la misma idea y palabras que pueden interpretarse de diferentes maneras de acuerdo al contexto.

4.1. Variación y ambigüedad lingüísticas

Se habla de **variación lingüística** cuando es posible utilizar diferentes palabras o expresiones para comunicar una misma idea (Mayoral Asensio, 1997). Por otro lado, la **ambigüedad lingüística** se refiere a la posibilidad de que una misma palabra o frase tenga más de una interpretación.

Ambos fenómenos tienen un efecto no deseado en la clasificación de texto. En el caso de la **variación lingüística**, se produce lo que se denomina el **silencio documental**, esto es, la no obtención de resultados relevantes por no haber utilizado los términos exactos en que el documento se encuentra escrito. Por su parte, la **ambigüedad lingüística** produce el denominado **ruido documental**, es decir, la obtención de resultados no relevantes debido a que los textos contienen el término utilizado para clasificarlo pero con un uso diferente al deseado.

En las siguientes secciones se muestran algunos ejemplos de estos fenómenos lingüísticos (Lim, 1998).

4.1.1. Roles diferentes en función del contexto

Es posible que una misma palabra adopte diferentes roles en función del contexto, provocando ambigüedad lingüística. Por ejemplo:

Coloca el dinero que te sobre en el sobre que se encuentra sobre la mesa.

Como se puede observar, en el enunciado anterior, la palabra *sobre* puede adoptar tres diferentes roles:

a) Como forma conjugada del verbo sobrar: el dinero que te *sobre*.
b) Como sustantivo masculino singular: en el *sobre*.
c) Como preposición: *sobre* la mesa.

4.1.2. Diferentes asociaciones de frases

Con el fin de transmitir una idea, se suelen relacionar diferentes frases. Cada vez que se agrega una frase más surge ambigüedad ante la posibilidad de establecer las relaciones en un orden diferente. Por ejemplo:

Ayer vi a un niño con unos lentes en la escuela.

El enunciado anterior puede interpretarse de diferentes maneras:

a) Ayer vi a un niño que estaba en la escuela y que tenía unos lentes.
b) Ayer estaba en la escuela donde vi a un niño que tenía unos lentes.
c) Ayer estaba en la escuela donde, con ayuda de unos lentes, vi a un niño.

4.1.3. Polisemia

La polisemia es un fenómeno del lenguaje en el que una misma palabra puede tener más de un significado. Por ejemplo:

A María no le gustan las llamas.

En el ejemplo anterior, la palabra *llamas* puede tener dos significados distintos:

a) *Llama:* masa de gas que expiden los cuerpos en combustión.
b) *Llama:* Mamífero rumiante originario de Perú.

4.1.4. Sinonimia

Otro fenómeno común del lenguaje es la posibilidad de encontrar palabras diferentes que tienen un mismo significado. Tal es el caso de:

Principio/Inicio/Comienzo

Cuando determinadas palabras tienen el mismo significado de manera independiente a su contexto se denomina **sinonimia estricta**.

4.1.5. Lenguaje figurado

Existen casos en los que no es posible realizar una interpretación literal de las expresiones debido al uso del denominado **lenguaje o sentido figurado**. Un ejemplo de esto es la siguiente frase del famoso poema de *Jaime Sabines*, *Los amorosos* (Sabines, 2011):

Los amorosos son la hidra del cuento.

Tienen serpientes en lugar de brazos.

4.1.6. Anáforas

Otro fenómeno que produce ambigüedad es la presencia de **anáforas**, es decir, el uso de pronombres y adverbios que hacen referencia a algo mencionado con anterioridad. Tal es el caso de:

¿Cuándo dije eso?

¿Quién habló?

¿Debajo de dónde?

En estos casos la frase por sí misma no tiene un significado relevante, es necesario recurrir al contexto para hacer una interpretación correcta.

A través de los ejemplos anteriores ha quedado de manifiesto lo complejo que puede ser el lenguaje natural, lo que debe ser considerado durante el desarrollo de un método de clasificación automática de texto.

4.2. Corpus lingüísticos

Uno de los elementos más utilizados como referencia en el estudio de una lengua, lo constituyen los denominados **corpus lingüísticos**.

La definición más simple de **corpus** nos refiere a éste como una colección, generalmente amplia, de textos. Sin embargo, cuando el término es usado en el ámbito de la lingüística computacional, éste tiene más implicaciones.

La primera implicación se refiere al lugar donde este conjunto de textos está almacenado. (Leech, 1992) introduce el concepto de corpus *como un emocionante fenómeno, una magnífica gran cantidad de texto, almacenada en una computadora*.

Por su parte, (Francis, 1982) agrega a su definición de corpus el que esta colección se asume como representativa de un determinado idioma, dialecto o subconjunto de un idioma para ser usado en análisis lingüístico.

Pero, tal vez, la mejor definición la proporciona el grupo de trabajo que está dedicado a los corpus de texto. Denominado como EAGLES (Expert Advisory Group on Language Engineering, 1996), este grupo define un **corpus** como una *colección de piezas de un idioma seleccionadas y ordenadas de acuerdo a criterios lingüísticos explícitos con el fin de ser usados como ejemplo de un idioma*.

Actualmente, existe una gran variedad de corpus de diferentes dominios y disponibles en una gran cantidad de idiomas. Su importancia es tal, que existe una disciplina al interior de la lingüística, denominada **lingüística de corpus**, dedicada a estudiar la lengua a través de estas muestras de texto.

Capítulo 5
Una aproximación a un método clasificador para la identificación de interacciones medicamentosas

El método de identificación de párrafos que enuncien interacciones mediante un clasificador automático que es objeto del presente trabajo, fue desarrollado con base en la metodología de categorización de texto propuesta por (Aas & Eikvil, 1999) bajo un **enfoque estadístico de reconocimiento de patrones** mediante la aplicación de un modelo de **n-gramas**. De esta forma, se obtuvo un **clasificador de tipo supervisado**, a través de un **ponderado de términos por frecuencia** mediante el uso de **n-gramas de palabras**. El algoritmo desarrollado fue denominado **CIM**, por sus siglas, **Clasificador de Interacciones Medicamentosas**.

Éste método está conformado, en su sentido más general, por las siguientes fases y etapas de procesamiento textual.

1. Fase de entrenamiento
    a. Preprocesado
    b. Indexado
    c. Ponderación por frecuencia
    d. Generación del modelo
2. Fase de prueba
    a. Preprocesado
    b. Indexado
    c. Selección de resultado

Para la implementación del algoritmo clasificador en mención, se desarrolló un prototipo usando el lenguaje de programación **AWK**, herramienta básica para el

procesamiento de información basada en texto. Todos los códigos de ejemplo incluidos en el presente capítulo se encuentran codificados en AWK.

5.1. Fase de entrenamiento

En primera instancia, se requiere un conjunto de entrenamiento. Para ello, se generó un corpus conformado por párrafos extraídos del **Diccionario de Especialidades Farmacéuticas de Thomson** o PLM, donde se expresara alguna interacción entre medicamentos. De esta forma, se seleccionó un medicamento por cada letra del alfabeto, con el objetivo de generar una muestra aleatoria.

En la **Tabla 7** se muestran los 3**6 medicamentos** que conformaron el corpus base.

A Grin	Ebixa	Livial	QG5
Acavexal	Esclerovitan A O	M E Medic	Radiance
Antivipmyn	Evista	Mazda	Sabima
Argentafil	Fabitec	Nabian-K	Tabcin Active
Bactocin	Gablacotec	Nodescrón	Ulcevit
Bricanyl ex	Ganglioside		Valcyte
Brurem	Glypressin	Navildez	Wadil
C Cobistal	Haitrax	Oacerein	Xarelto
Collifrin	Ibacnol	Octalbin	Xuzal
Crinone	Jalra	Paclisan	Yadegal Compuesto
Dabex	K 50	Protamina 1000	Zaat
Dismedox	Lacdol-S		

Tabla 7. Medicamentos utilizados para la generación del corpus base.

A partir de los párrafos que enuncian alguna interacción medicamentosa, incluidos en el corpus, se realiza la fase de entrenamiento y se genera el modelo que servirá para clasificar nuevos textos. El algoritmo de entrenamiento se encuentra en el archivo `cim_entrena.awk`.

5.1.1. Preprocesado

Ocasionalmente, la información de entrada debe ser tratada de manera previa, con el fin de corregir posibles deficiencias, o bien de preparar los datos para su posterior procesamiento.

Por ello, el primer paso del proceso de categorización textual es el **preprocesado de documentos**, que consiste en transformar el texto desde su forma original hasta

llevarlo a un formato adecuado para el procesamiento mediante técnicas de clasificación automática.

Para el presente trabajo, la fase de preprocesado se conformó de las siguientes etapas:

a) Eliminación de etiquetas
b) Eliminación de palabras vacías o stopwords

Una vez ejecutadas las etapas mencionadas se obtiene un texto del que es posible extraer características que permitan su clasificación, lo que constituye la siguiente fase del procesamiento textual.

5.1.1.1. Eliminación de etiquetas

Uno de los principales objetivos del presente trabajo es automatizar tareas repetitivas cuya complejidad es mínima. Para tal efecto, es deseable que la información de entrada tenga el tratamiento manual mínimo posible, esto es, minimizar las tareas que tenga que llevar a cabo un ser humano para que el texto de entrada sea tratado por el clasificador.

Uno de los métodos más comunes para codificar texto es el denominado **lenguaje de marcado**, que se auxilia de etiquetas o marcas para proveer información adicional sobre la estructura o la presentación del texto. La codificación más extendida es la conocida como **Lenguaje de Marcado de Hipertexto** o **HyperText Markup Language** (HTML por sus siglas). En éste y otros tipos de codificación, la forma más común de indicar la presencia de una etiqueta es mediante los símbolos conocidos como *menor que* (<) y *mayor que* (>), correspondientes a los códigos ASCII 3C y 3E hexadecimal, respectivamente. En la **Fig. 12** se muestra un ejemplo de texto con etiquetas HTML.

```
<p class="normal">
  <span class="rubros-azules">FARMACOCINÉTICA Y
  FARMACODINAMIA:</span>
</p>
<p class="normal">
  "El ketorolaco es un agente antiinflamatorio
  no esteroide, que muestra actividad
  analgésica, antiinflamatoria y débil actividad
  antipirética."
</p>
<p class="normal">
  "El ketorolaco inhibe la síntesis de
  prostaglandinas y no tiene ningún efecto sobre
  los receptores de los opiáceos."
</p>
<p class="normal">…</p>
```

Fig. 9. Texto con etiquetas HTML.

Con la finalidad de eliminar todas las etiquetas del texto de entrada que cumplan con la condición de aparecer entre los símbolos *menor que* y *mayor que*, antes mencionados, se utiliza la función **gsub** de AWK, que sirve para sustituir una cadena por otra. En la **Tabla 8** se muestra el código que realiza dicha sustitución.

```
1: gsub(/<[^>]+>/,"");
```

Tabla 8. Código para eliminar etiquetas delimitadas por < y >.

Para eliminar etiquetas delimitadas por caracteres distintos a *menor que* y *mayor que*, basta con reproducir el comando mostrado en la **Tabla 8** y sustituir los símbolos mencionados por un carácter nulo.

5.1.1.2. Eliminación de palabras vacías o stopwords

Se denomina palabras vacías o stopwords, a aquellas palabras que no proporcionan información relevante para la clasificación del texto, tales como preposiciones, conjunciones, disyunciones y verbos copulativos, principalmente.

Si no se realiza un filtrado previo de estas palabras vacías, es muy probable que durante el proceso de búsqueda se genere el ya mencionado **ruido documental**.

Para realizar el filtrado de palabras vacías, se utilizó la lista proporcionada por el proyecto **Snowball** (Porter, 2014), que ha desarrollado listas de palabras vacías en diferentes idiomas como parte de un proyecto de software de **stemming**.

Para poder procesar el texto, se genera el arreglo `stopwords[]`, que contiene las palabras vacías que serán filtradas. Posteriormente se genera el arreglo `palabras[]`, que contendrá solamente las palabras cuyo significado es relevante para la clasificación, es decir, ya filtradas. Para ello, el prototipo lee cada palabra del **conjunto de entrenamiento**, comparándola con la lista de palabras vacías, a fin de discriminar aquellas que coincidan. En la **Tabla 9** se muestra el segmento de código que realiza esta tarea.

```
 1: for (i=1; i<NF; i++)
 2: {
 3:    j=0;
 4:    while(j<num_stopwords)
 5:    {
 6:       j++;
 7:       es_stopword=0;
 8:       if (tolower($i)==tolower(stopwords[j]))
 9:       {
10:          es_stopword=1;
11:          break;
12:       }
13:    }
14:    if (es_stopword==0)
15:       palabras[num_palabras++]=tolower($i);
16: }
```

Tabla 9. Código para filtrar palabras vacías.

5.1.2. Indexado

En la etapa de indexado son obtenidos los **n-gramas** a partir del arreglo `palabras[]`. El valor de *n* será definido por la variable `NGRAM_LENGTH`. De esta forma, el prototipo calcula los n-gramas con *n* términos y con (*n-1, n-2, ..., 1*). Las frecuencias son multiplicadas por *n*, a fin de asignar un peso mayor a los bigramas y trigramas, y así privilegiar el orden de aparición conjunta de los términos. AWK permite utilizar índices no numéricos. Haciendo uso de esta característica, se genera el arreglo `ngrams[]`, utilizando como índice el mismo n-grama. Esto permitirá calcular en el mismo segmento de código la frecuencia de cada n-grama, como se muestra en la **Tabla 10**.

```
1:  for (i=1; i<=num_palabras; i++)
2:  {
3:     for (j=0; j<NGRAM_LENGTH; j++)
4:     {
5:        if (j==0) text_ngram = palabras[i];
6:        if (j>0) text_ngram = text_ngram "/" palabras[i+j];
7:        if (i <= (num_palabras-j))
8:        {
9:           #Se almacena la frecuencia global del n-grama
10:          ngrams[text_ngram] = text_ngram;
11:         frecuencias[text_ngram]=(frecuencias[text_ngram]+1)*j;
12:       }
13:    }
14: }
```

Tabla 10. Código para generar los n-gramas y su frecuencia.

5.1.3. Ponderación por frecuencia

Una vez que han sido obtenidos la frecuencia de los n-gramas, se calcula la probabilidad de cada n-grama dado el conjunto de entrenamiento. Para ello, se divide la frecuencia obtenida para cada n-grama entre la cantidad total de palabras del documento. Este dato es obtenido de la variable `num_palabras`, que fue actualizada durante la generación del arreglo `palabras[]`.

En la **Tabla 11** se muestra el código que realiza la **ponderación por frecuencia**.

```
1:  for (x in frecuencias)
2:  {
3:     prob[x] = frecuencias[x]/num_palabras;
```

```
4:     model[x] = model[x] + prob[x];
5: }
```

Tabla 11. Código que calcula la ponderación por frecuencia de cada n-grama.

5.1.4. Generación del modelo

Finalmente, se genera el modelo de n-gramas ponderado por frecuencia, para lo que se crea el archivo `modelo.txt`. Esto se realiza con ayuda del código mostrado en la **Tabla 12**.

```
1: for (ngrama in prob)
2:    if ((ngrama != "/" A) && (ngrama != "//"))
3:       print ngrama " " prob[ngrama] > "modelo.txt";
```

Tabla 12. Código que genera el modelo de n-gramas ponderados por frecuencia.

De esta forma, el archivo `modelo.txt`, tendrá una estructura de dos columnas, en donde la primera columna contendrá los n-gramas obtenidos a partir del conjunto de entrenamiento, y en la segunda columna se encontrará la probabilidad calculada dado el total de palabras contenidas en el documento. En la **Fig. 10** se muestra un ejemplo de la estructura de `modelo.txt`.

	A	B
1	inhibidores/monoaminooxidasa/(imao)	0.00035051
2	alterna,/decir,	0.00035051
3	tubular,/incremento/porcentaje	0.00035051
4	utilice/topiramato/evitarse	0.00035051
5	modelo/in	0.00035051
6	4.5	0.00035051
7	respectivamente	0.00035051
8	medicamentos/aumentar	0.00035051
9	status/epilepticus.	0.00035051
10	incrementa/concentraciones	0.00035051
11	utilizarse/valganciclovir	0.00035051

Fig. 10. Estructura del archivo modelo.txt.

Este modelo de n-gramas será utilizado para clasificar nuevas instancias con base en la probabilidad de aparición de los mismos en textos no clasificados.

5.2. Fase de prueba

Una vez que se ha realizado el entrenamiento, el clasificador está en condiciones de realizar su función principal: clasificar párrafos de texto con el fin de identificar aquellos en donde se enuncien interacciones medicamentosas.

El algoritmo de entrenamiento se encuentra en el archivo `cim_prueba.awk`.

5.2.1. Preprocesado

Con el fin de clasificar las nuevas instancias bajo condiciones similares a las utilizadas para la generación del modelo, es necesario realizar el preprocesado del texto que será clasificado de manera análoga a la descrita en la sección **5.1.1 (Preprocesado)**.

Como se ha visto antes, el preprocesado consta de dos actividades: la eliminación de etiquetas y la eliminación de palabras vacías o stopwords.

La eliminación de etiquetas se realiza de la misma manera que en la **fase de entrenamiento**, por lo que no se explicará nuevamente en esta fase.

Por su parte, la eliminación de palabras vacías es abordada con un enfoque distinto, debido a que, a diferencia de la fase de entrenamiento, en la fase de prueba es necesario ponderar las probabilidades por párrafo. Para ello, se utiliza el código que se muestra en la **Tabla 13**.

```
 1: while (getline < ARGV[1])
 2: {
 3:    num_lineas++;
 4:    frases[num_lineas]=$0;
 5:    num_palabras=0;
 6:
 7:    for (i=1; i<NF; i++)
 8:    {
 9:       j=0;
10:       while(j<num_stopwords)
11:       {
12:          j++;
13:          es_stopword=0;
14:          if (tolower($i)==tolower(stopwords[j]))
15:          {
16:             es_stopword=1;
17:             break;
18:          }
19:       }
20:       if (es_stopword==0)
```

```
21:         {
22:             palabras[num_lineas, num_palabras++]=tolower($i);
23:             parrafos[num_lineas] = num_palabras;
24:         }
25:     }
26: }
```

Tabla 13. Código de eliminación de palabras vacías de la fase de prueba.

En primera instancia, se agregó la instrucción en el **renglón 3**, que contabilizará el número de líneas. En el ámbito de AWK, cada línea está separada por un salto de línea, por lo tanto, se puede entender cada línea como un párrafo.

El segundo cambio se encuentra en el **renglón 4**, lo constituye la generación del arreglo `frases[]`, que servirá para almacenar cada párrafo original, antes de la eliminación de palabras vacías. Este arreglo será utilizado para mostrar el párrafo con mayor probabilidad de expresar interacciones medicamentosas.

En el **renglón 22** se genera el arreglo `palabras[]`, cuyo índice se compone del número de línea y el número de palabra en la línea. De esta forma se podrá recorrer el arreglo e identificar su posición exacta en el texto.

Por último, en el **renglón 23** se genera el arreglo `párrafos[]`, que contendrá el número de palabras por párrafo, cuyo valor será necesario para recorrer el arreglo `palabras[]`.

También es necesario generar el arreglo `modelo[]`, a partir del archivo `modelo.txt`, generado durante la fase de entrenamiento. Este arreglo contendrá la probabilidad de cada n-grama.

5.2.2. Indexado

De manera similar a lo realizado durante la fase de entrenamiento, en el indexado se calculan los n-gramas a partir del arreglo `palabras[]`, con la diferencia de que en ésta fase se calcula la probabilidad de cada párrafo. Este cálculo de probabilidades se realiza buscando cada n-grama en el arreglo `modelo[]` y acumulando su probabilidad por párrafo. En la **Fig. 11** se muestra una representación de esta actividad.

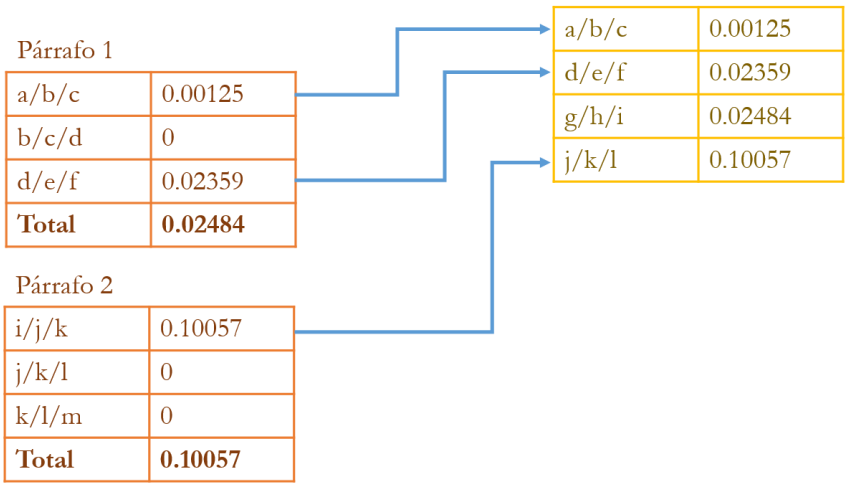


Fig. 11. Suma de probabilidades de cada n-grama en modelo.txt.

En la **Tabla 14** se muestra el código mediante el que se realiza esta acción.

```
1:  for (i=1; i<num_lineas; i++)
2:     for (j=1; j<parrafos[i]; j++)
3:     {
4:        num_ngramas=0;
5:        for (k=0; k<NGRAM_LENGTH; k++)
6:        {
7:           if (k==0) text_ngram = palabras[i,j];
8:           if (k>0)
9:              text_ngram = text_ngram "/" palabras[i,j+k];
10:          if (j <= (parrafos[i]-k))
11:          {
12:             #Se calcula la probabilidad del párrafo
13:             prob_parrafo[i] = prob_parrafo[i] +
                    modelo[text_ngram];
14:          }
15:       }
16:    }
```

Tabla 14. Código para la suma de probabilidades por párrafo.

El principal cambio del indexado en la fase de prueba, con relación a la fase de entrenamiento, lo constituye la instrucción del **renglón 13**, donde se acumula la probabilidad de n-grama en el modelo, por cada párrafo (arreglo `prob_parrafo[]`).

5.2.3. Selección de resultado

Una vez que el arreglo `prob_parrafo[]` contiene la probabilidad de cada n-grama acumulada por párrafo, sólo resta seleccionar el párrafo con la mayor probabilidad de expresar alguna interacción medicamentosa.

```
1:  max=0;
2:  mayor=0;
3:
4:  for (i=1; i<num_lineas; i++)
5:  {
6:     if (prob_parrafo[i] > max)
7:     {
8:        max = prob_parrafo[i];
9:        mayor = i;
10:    }
11: }
```

Tabla 15. Código para seleccionar el párrafo de mayor probabilidad.

Esto se realiza con un recorrido sencillo por el arreglo `prob_parrafo[]`, actualizando en la variable `max` el valor máximo de probabilidad y en la variable `mayor` el número del párrafo correspondiente. El código para realizar esta acción se muestra en la **Tabla 15**.

Finalmente, se muestra el párrafo original que fue calificado con la mayor probabilidad de enunciar alguna interacción medicamentosa, con base en la suma de las probabilidades de los n-gramas que lo componen.

Capítulo 6
Pruebas

Una vez desarrollado el prototipo objeto del presente trabajo, es de suma importancia verificar su funcionamiento, a través de la ejecución de algunas pruebas funcionales con información real.

El desarrollo de estas pruebas es de suma importancia, ya que permitirá conocer el grado de eficiencia de la aproximación propuesta, así como identificar áreas de oportunidad para futuras mejoras.

De acuerdo con la definición proporcionada por (International Organization for Standardization, 2010), las pruebas tienen como objetivo verificar la funcionalidad del sistema a través de sus interfaces externas comprobando que dicha funcionalidad sea la esperada en función de los requisitos.

Para el caso del presente trabajo, el requisito principal es la identificación probabilística de párrafos que contengan información relacionada con interacciones medicamentosas, por lo cual, las pruebas realizadas fueron orientadas a determinar el éxito en la consecución de este objetivo.

6.1. Validación de resultados

Como se mencionó en el **Capítulo 1 (Fuentes de información sobre medicamentos)**, para el desarrollo del presente trabajo, se utilizó el **Diccionario de Especialidades Médicas** (PLM México, 2013) como fuente de información.

La estructura en que es mostrada la información contenida en dicha fuente facilita la identificación de párrafos que contienen información relacionada con interacciones entre medicamentos, ya que éstas se encuentran señaladas de manera expresa, como se muestra en la Fig.

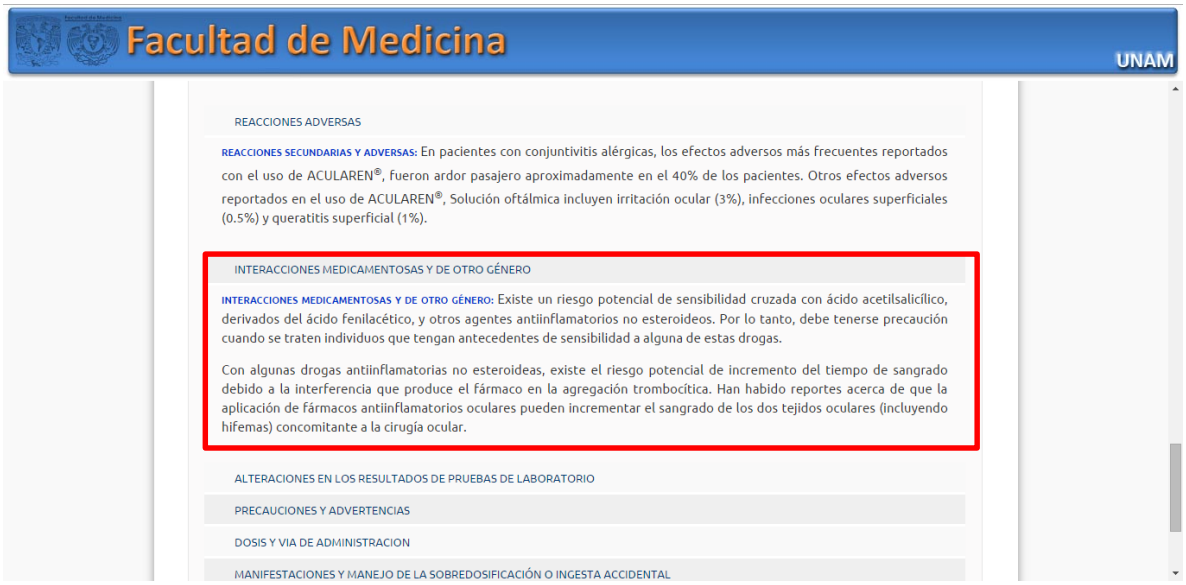


Fig. 12. Sección de interacciones medicamentosas en el Diccionario de Especialidades Médicas.

De esta forma, la verificación de resultados se realizó consultando la ficha de cada medicamento en el **Diccionario de Especialidades Médicas**.

Para facilitar la validación del resultado, el prototipo desarrollado mostrará, además del párrafo seleccionado, el orden que guarda con relación al resto de párrafos y la probabilidad calculada.

6.2. Sintaxis de ejecución

El prototipo desarrollado consta de dos archivos ejecutables, codificados en AWK:

a) Entrenamiento: `cim_entrena.awk`
b) Prueba: `cim_prueba.awk`

En las siguientes secciones se explica la sintaxis de ejecución para cada uno de los archivos ejecutables.

6.2.1. cim_entrena.awk

El archivo `cim_entrena.awk` requiere los siguientes datos de entrada:

a) Corpus base (archivo `enunciados.txt`)
b) Lista de palabras vacías (archivo `stopwords_es.txt`)

En la **Tabla 16** se muestra la sintaxis requerida para el archivo `cim_entrena.awk`. El parámetro `-f` le indica al intérprete de AWK que el código a ejecutar se encuentra en un archivo.

```
> awk -f cim_entrena.awk enunciados.txt stopwords_es.txt
```

Tabla 16. Sintaxis de ejecución para cim_entrena.awk.

En este caso, se ha omitido la salida por consola, ya que ésta es dirigida al archivo `modelo.txt`.

6.2.2. cim_prueba.awk

El archivo `cim_prueba.awk` requiere los siguientes datos de entrada:

a) Texto que será clasificado
b) Lista de palabras vacías (archivo `stopwords_es.txt`)
c) Modelo generado por `cim_entrena.awk` (archivo `modelo.txt`)

En la **Tabla 17** se muestra la sintaxis requerida para el archivo `cim_prueba.awk`.

```
> awk -f cim_prueba.awk [texto.txt] stopwords_es.txt modelo.txt
```

Tabla 17. Sintaxis de ejecución para cim_prueba.awk.

La información generada por `cim_prueba.awk` consta del archivo de texto `probabilidades.txt`, conteniendo la probabilidad de todos los párrafos del texto de entrada, así como una salida en pantalla, con la estructura que se indica en la **Tabla 18**.

```
> [Número de párrafo] [Párrafo original] [Probabilidad]
```

Tabla 18. Estructura de salida de cim_prueba.awk.

6.3. Pruebas

Para todos los casos de prueba que se enuncian a continuación, dado que el prototipo desarrollado selecciona únicamente el párrafo con mayor probabilidad de enunciar alguna interacción medicamentosa, se comprobará que dicho párrafo se encuentre en la sección denominada **Interacciones medicamentosas y de otro género**, dentro de la ficha del medicamento, tal como se incluye en el **Diccionario de Especialidades Médicas**.

Las pruebas realizadas se ejecutaron a partir de un conjunto de 10 medicamentos, cuyas marcas fueron seleccionadas de manera aleatoria, que contuvieran sustancias activas indicadas en el **Cuadro Básico de Medicamentos** del **Instituto Mexicano del Seguro Social** (Instituto Mexicano del Seguro Social, 2015). Las sustancias activas también fueron seleccionadas aleatoriamente.

La lista con los medicamentos utilizados se muestra en la **Tabla 19**.

No.	Sustancia activa	Medicamento seleccionado	Grupo
1	Ácido acetilsalicílico	Dalabul	Analgesia (1)
2	Fentanilo	Fenodid	Anestesia (2)
3	Amiodarona	Braxan	Cardiología (3)
4	Hidrocortisona	Nositrol	Dermatología (4)
5	Atorvastatina	Blodivit	Endocrinología y metabolismo (5)
6	Amikacina	AMK	Enfermedades infecciosas y parasitarias (6)
7	Loratadina	Lorimox	Enfermedades inmunoalérgicas (7)
8	Ranitidina	Ranulin	Gastroenterología (8)
9	Metronidazol	Metricom	Gineco-obstetricia (9)
10	Dexametazona	Decorex	Hematología (10)

Tabla 19. Medicamentos utilizados en la realización de pruebas.

Para verificar el resultado de la clasificación, se consideraron las siguientes comprobaciones:

a) Identificación visual del párrafo seleccionado en pantalla (salida por consola)
b) Verificación mediante la ficha del medicamento (PLM México, 2013)

Los resultados de las pruebas realizadas se enuncian de forma dicotómica, esto es, asignando un valor positivo si el párrafo seleccionado efectivamente enuncia una interacción medicamentosa y un valor negativo en el caso contrario.

A continuación se muestran dos casos de prueba, con la finalidad de ejemplificar las evidencias obtenidas para resultados satisfactorios y no satisfactorios. Posteriormente, se presenta un resumen de los resultados finales.

6.3.1. Caso satisfactorio: Fenodid

Fenodid es un medicamento que contiene **fentanilo**, indicado como agente analgésico opioide complementario en anestesia general, regional o local.

Se realizó la clasificación mediante `cim_prueba.awk`, sobre el archivo `fenodid.txt`, obteniendo como resultado la selección del **párrafo 105** con una ponderación sumarizada de probabilidades de **0.00898876**. Este resultado es mostrado en la **Fig. 13**.

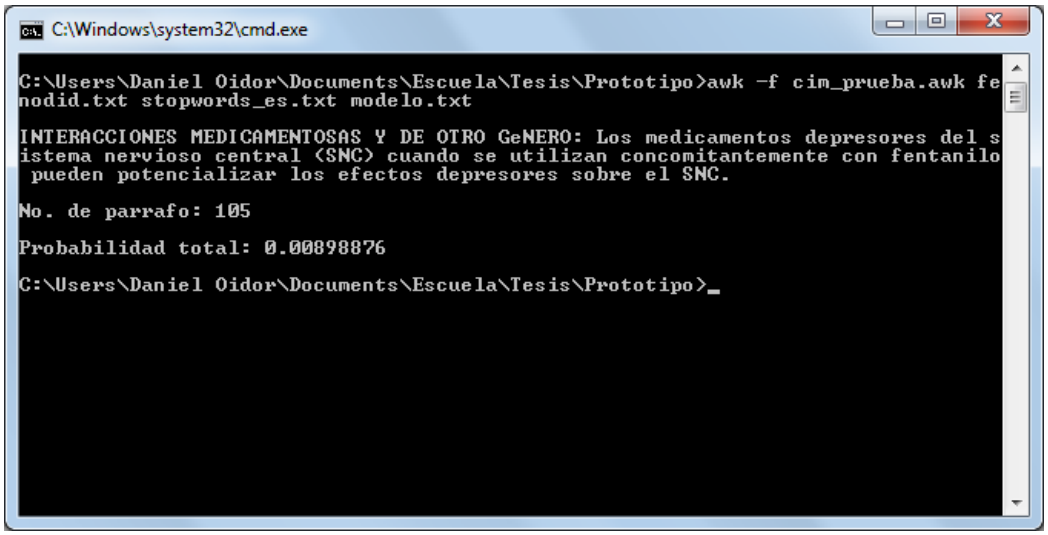
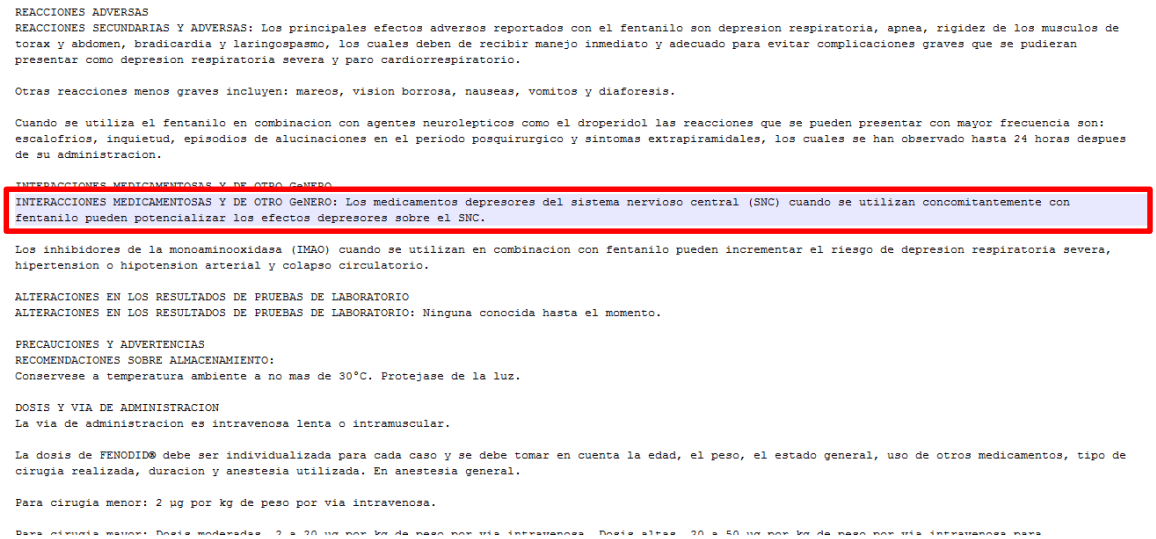


Fig. 13. Resultado de la clasificación del archivo fenodid.txt.

El párrafo seleccionado, de acuerdo con la ficha original, **sí expresa una interacción medicamentosa**, como puede observarse en la **Fig. 14**.



REACCIONES ADVERSAS
REACCIONES SECUNDARIAS Y ADVERSAS: Los principales efectos adversos reportados con el fentanilo son depresion respiratoria, apnea, rigidez de los musculos de torax y abdomen, bradicardia y laringospasmo, los cuales deben de recibir manejo inmediato y adecuado para evitar complicaciones graves que se pudieran presentar como depresion respiratoria severa y paro cardiorrespiratorio.

Otras reacciones menos graves incluyen: mareos, vision borrosa, nauseas, vomitos y diaforesis.

Cuando se utiliza el fentanilo en combinacion con agentes neurolepticos como el droperidol las reacciones que se pueden presentar con mayor frecuencia son: escalofrios, inquietud, episodios de alucinaciones en el periodo posquirurgico y sintomas extrapiramidales, los cuales se han observado hasta 24 horas despues de su administracion.

INTERACCIONES MEDICAMENTOSAS Y DE OTRO GeNERO
INTERACCIONES MEDICAMENTOSAS Y DE OTRO GeNERO: Los medicamentos depresores del sistema nervioso central (SNC) cuando se utilizan concomitantemente con fentanilo pueden potencializar los efectos depresores sobre el SNC.

Los inhibidores de la monoaminooxidasa (IMAO) cuando se utilizan en combinacion con fentanilo pueden incrementar el riesgo de depresion respiratoria severa, hipertension o hipotension arterial y colapso circulatorio.

ALTERACIONES EN LOS RESULTADOS DE PRUEBAS DE LABORATORIO
ALTERACIONES EN LOS RESULTADOS DE PRUEBAS DE LABORATORIO: Ninguna conocida hasta el momento.

PRECAUCIONES Y ADVERTENCIAS
RECOMENDACIONES SOBRE ALMACENAMIENTO:
Conservese a temperatura ambiente a no mas de 30°C. Protejase de la luz.

DOSIS Y VIA DE ADMINISTRACION
La via de administracion es intravenosa lenta o intramuscular.

La dosis de FENODID® debe ser individualizada para cada caso y se debe tomar en cuenta la edad, el peso, el estado general, uso de otros medicamentos, tipo de cirugia realizada, duracion y anestesia utilizada. En anestesia general.

Para cirugia menor: 2 µg por kg de peso por via intravenosa.

Para cirugia mayor: Dosis moderadas, 2 a 20 µg por kg de peso por via intravenosa. Dosis altas, 20 a 50 µg por kg de peso por via intravenosa para

Fig. 14. Captura de pantalla de la ficha de Fenodid.

6.3.2. Caso no satisfactorio: Dalabul

Dalabul es un medicamento que contiene **ácido acetilsalicílico** y es utilizado como analgésico, antipirético y antiinflamatorio.

Se realizó la clasificación mediante `cim_prueba.awk`, sobre el archivo `dalabul.txt`, obteniendo como resultado la selección del **párrafo 97** con una ponderación sumarizada de probabilidades de **0.161421**. Este resultado es mostrado en la **Fig. 15**.

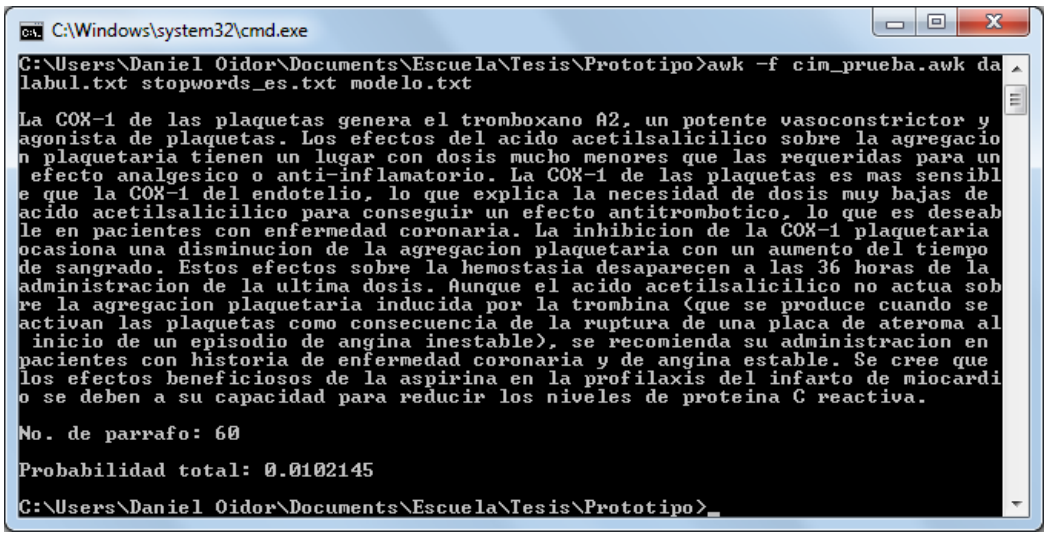
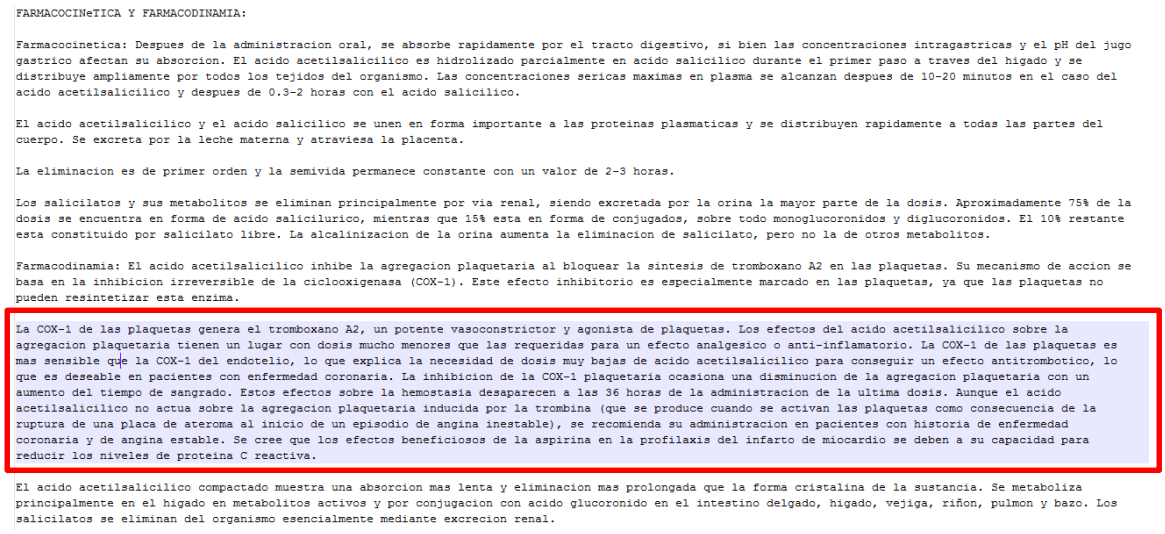


Fig. 15. Resultado de la clasificación del archivo dalabul.txt.

El párrafo seleccionado, de acuerdo con la ficha original, **no expresa una interacción medicamentosa**, como puede observarse en la **Fig. 16**.



FARMACOCINeTICA Y FARMACODINAMIA:

Farmacocinetica: Despues de la administracion oral, se absorbe rapidamente por el tracto digestivo, si bien las concentraciones intragastricas y el pH del jugo gastrico afectan su absorcion. El acido acetilsalicilico es hidrolizado parcialmente en acido salicilico durante el primer paso a traves del higado y se distribuye ampliamente por todos los tejidos del organismo. Las concentraciones sericas maximas en plasma se alcanzan despues de 10-20 minutos en el caso del acido acetilsalicilico y despues de 0.3-2 horas con el acido salicilico.

El acido acetilsalicilico y el acido salicilico se unen en forma importante a las proteinas plasmaticas y se distribuyen rapidamente a todas las partes del cuerpo. Se excreta por la leche materna y atraviesa la placenta.

La eliminacion es de primer orden y la semivida permanece constante con un valor de 2-3 horas.

Los salicilatos y sus metabolitos se eliminan principalmente por via renal, siendo excretada por la orina la mayor parte de la dosis. Aproximadamente 75% de la dosis se encuentra en forma de acido salicilurico, mientras que 15% esta en forma de conjugados, sobre todo monoglucoronidos y diglucoronidos. El 10% restante esta constituido por salicilato libre. La alcalinizacion de la orina aumenta la eliminacion de salicilato, pero no la de otros metabolitos.

Farmacodinamia: El acido acetilsalicilico inhibe la agregacion plaquetaria al bloquear la sintesis de tromboxano A2 en las plaquetas. Su mecanismo de accion se basa en la inhibicion irreversible de la ciclooxigenasa (COX-1). Este efecto inhibitorio es especialmente marcado en las plaquetas, ya que las plaquetas no pueden resintetizar esta enzima.

La COX-1 de las plaquetas genera el tromboxano A2, un potente vasoconstrictor y agonista de plaquetas. Los efectos del acido acetilsalicilico sobre la agregacion plaquetaria tienen un lugar con dosis mucho menores que las requeridas para un efecto analgesico o anti-inflamatorio. La COX-1 de las plaquetas es mas sensible que la COX-1 del endotelio, lo que explica la necesidad de dosis muy bajas de acido acetilsalicilico para conseguir un efecto antitrombotico, lo que es deseable en pacientes con enfermedad coronaria. La inhibicion de la COX-1 plaquetaria ocasiona una disminucion de la agregacion plaquetaria con un aumento del tiempo de sangrado. Estos efectos sobre la hemostasia desaparecen a las 36 horas de la administracion de la ultima dosis. Aunque el acido acetilsalicilico no actua sobre la agregacion plaquetaria inducida por la trombina (que se produce cuando se activan las plaquetas como consecuencia de la ruptura de una placa de ateroma al inicio de un episodio de angina inestable), se recomienda su administracion en pacientes con historia de enfermedad coronaria y de angina estable. Se cree que los efectos beneficiosos de la aspirina en la profilaxis del infarto de miocardio se deben a su capacidad para reducir los niveles de proteina C reactiva.

El acido acetilsalicilico compactado muestra una absorcion mas lenta y eliminacion mas prolongada que la forma cristalina de la sustancia. Se metaboliza principalmente en el higado en metabolitos activos y por conjugacion con acido glucoronido en el intestino delgado, higado, vejiga, riñon, pulmon y bazo. Los salicilatos se eliminan del organismo esencialmente mediante excrecion renal.

Fig. 16. Captura de pantalla de la ficha de Dalabul.

Una vez mostrada la forma en que fueron evaluados los diez casos que conformaron la prueba, se muestran a continuación los resultados obtenidos mediante el método clasificador diseñado.

6.4. Resultados

Después de ejecutadas las pruebas, se procedió a contabilizar los resultados obtenidos, mismos que se muestran en la **Tabla 20**.

No.	Medicamento	Sustancia activa	Resultado
1	Dalabul	Ácido acetilsalicílico	No satisfactorio
2	Fenodid	Fentanilo	Satisfactorio
3	Braxan	Amiodarona	No satisfactorio
4	Nositrol	Hidrocortisona	No satisfactorio
5	Blodivit	Atorvastatina	Satisfactorio
6	AMK	Amikacina	Satisfactorio
7	Lorimox	Loratadina	Satisfactorio
8	Ranulin	Ranitidina	Satisfactorio
9	Metricom	Metronidazol	Satisfactorio
10	Decorex	Dexametazona	No satisfactorio
		Efectividad	**60%**

Tabla 20. Resultados de las pruebas.

Con base en los resultados mostrados, la aproximación a un algoritmo clasificador de interacciones medicamentosas objeto del presente trabajo, tiene una **efectividad del 60%**.

De los cuatro casos cuyo resultado fue no satisfactorio, en tres de ellos el párrafo seleccionado se encontró en la sección de **Farmacocinética y farmacodinamia** y uno de ellos en **Precauciones generales**.

Conclusiones

El principal objetivo del presente trabajo fue proponer una aproximación a un método clasificador que permita la identificación de interacciones medicamentosas a través de un diseño orientado a que ésta se lleve a cabo de forma previa a la prescripción médica.

Para alcanzar el objetivo citado, se realizó una investigación de los métodos actuales utilizados para reconocer interacciones medicamentosas, mediante la que fue posible determinar algunos algoritmos de evaluación de las reacciones adversas que un paciente presenta ante el consumo de determinados medicamentos. Por otro lado, se encontraron fuentes de información disponible en medios electrónicos, cuya actualización requiere de una constante revisión del contenido.

De igual manera, se identificaron aquellas fuentes de información sobre medicamentos, que fueran confiables. Fue así como se seleccionó al **Diccionario de Especialidades Farmacéuticas de Thomson** como la principal fuente del presente trabajo.

Se analizaron las ventajas y características generales de los diferentes tipos de clasificadores, lo que permitió elegir un **enfoque estadístico de reconocimiento de patrones**, mediante **n-gramas de palabras**, como base para la aproximación del algoritmo en mención.

Una vez concluido el desarrollo y realizadas las pruebas, los resultados mostraron una **efectividad del 60%**, lo que representa una densidad de error de 40%.

Durante el presente trabajo, se obtuvieron algunas conclusiones importantes, que se enuncian a continuación:

- **La organización del texto es relevante para la ponderación de probabilidades.** Como pudo observarse, los resultados obtenidos fueron satisfactorios, dado que la organización del texto en la fuente original contaba con una separación en párrafos de tamaños (cantidad de palabras) homogéneos.
- **La frecuencia de aparición de las palabras es de mayor importancia que el orden de las mismas.** Para el caso de identificar el objetivo de un texto, los **n-gramas** con mayor repetición tuvieron, generalmente, un valor de n menor a 3, lo que muestra que la ponderación de probabilidad se dio, principalmente, por la aparición de palabras.
- **El orden de aparición de las palabras no es relevante cuando los documentos por clasificar no tienen un estilo de redacción identificable.** La fuente de información seleccionada contiene textos escritos por diferentes personas en un lenguaje neutral con un objetivo informativo. En estos casos, el orden de aparición de las palabras puede variar, perdiendo así relevancia en el proceso de identificación.

Trabajo futuro

Finalmente, el presente trabajo puede ser complementado con el fin de obtener mejoras en la efectividad del mismo, mediante acciones como las que se enuncia a continuación:

- **Evaluar el rendimiento del clasificador con valores de n distintos a 3.** Con el fin de seleccionar un valor de n cuyos resultados sean más certeros, se requiere evaluar el grado de efectividad con valores menores y mayores a 3, que fue el valor seleccionado en el desarrollo del presente trabajo.
- **Incorporar al corpus base información dicotómica.** En el presente trabajo se construyó un corpus base que contuviera párrafos precalificados con información de interacciones medicamentosas. De esta forma, el resultado de la clasificación es la selección del párrafo con mayor probabilidad de enunciar interacciones medicamentosas. Integrar al corpus base, párrafos precalificados que no enuncien interacciones medicamentosas, permitirá obtener, adicionalmente, la probabilidad de que cada párrafo clasificado no enuncie interacciones entre medicamentos. Con ambas probabilidades, es posible calcular una brecha y mejorar el rendimiento de la clasificación.

- **Establecer un valor mínimo de probabilidad.** La aproximación desarrollada permite seleccionar un párrafo que enuncie interacciones entre medicamentos. Es necesario establecer un valor mínimo de probabilidad para considerar que un párrafo enuncia alguna interacción. Esto es necesario dado que, en textos donde no sean expresadas interacciones medicamentosas el valor de probabilidad más alto sea discriminado si no supera el mínimo establecido.

Referencias

Aas, K., & Eikvil, L. (1999). *Text Categorisation: A survey.* Oslo, Noruega: Norwegian Computing Center.

Armijo, J., & González, M. (2001). Estudios de seguridad de medicamentos: Métodos para detectar las reacciones adversas y la valoración de la relación causa-efecto. (Farmaindustria, Ed.) *El ensayo clínico en España*, 161-190.

Awad, E. M., & Ghaziri, H. M. (2004). *Knowledge Management.* Prentice Hall.

Benemérita Universidad Autónoma de Puebla. (2011). *Dirección General de Bibliotecas BUAP.* Recuperado el 11 de Enero de 2014, de http://www.bibliotecas.buap.mx/portal/index.php

Bisquerra Alzina, R. (1989). *Introducción conceptual al análisis multivariable : un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD* (Vol. III). España: McGraw Hill.

Brito Barrera, Y., & Serrano Martínez, P. (Agosto de 2011). Uso de medicamentos, reacciones adversas e interacciones farmacológicas en un hospital obstétrico de Puebla, México. *InFÁRMAte*(27), 58-98.

Calderón Ospina, C. A., & Urbina Bonilla, A. (Enero-Abril de 2010). La Farmacovigilancia en los últimos 10 años: actualización de conceptos y clasificaciones. Logros y retos para el futuro en Colombia. *Médicas UIS*(24), 57-73.

Campos Garza, J. F., Aquino Arteaga, A., Uc Morales, D. N., Herrera Huerta, E. V., Velázquez Hernández, F., & Hernández Cruz, R. (2006). Detección de interacciones medicamentosas en el servicio de medicina interna del Hospital General Regional de Orizaba, Veracruz. *Revista de Salud Pública y Nutrición*(11).

Caso Prado, P. (2011). *Esquema regulatorio de medicamentos en México: Oportunidades y retos.* Ciudad de México.

Charniak, E. (1993). *Statistical language learning.* Cambridge: The MIT Press.

COFEPRIS. (Abril de 2014). *Obras de consulta*. Recuperado el 15 de Marzo de 2014, de Secretaría de Salud de México: http://www.salud.gob.mx/unidades/cofepris/bv/mconsulta.htm

Consejo Nacional de Salud. (2010). Cuadro Nacional de Medicamentos Básicos y Registro Terapéutico. Quito, Ecuador.

Cornejo Aparicio, V. M. (2014). Modelamiento de espacio de palabras en la clasificación de documentos. *Ingetecno, 3*(1), 42-55.

Dirección General de Medicamentos, Insumos y Drogas. (2000). *Resolución Directoral Nº 813 - 2000-DG-DIGEMID*. Resolución, Estado Peruano, Ministerio de Salud, Lima.

Duda, R. O., Hart, P. E., & Stork, D. G. (1997). *Pattern classification* (Segunda ed.). Menlo Park, California, U.S.A.: Ricoh California Research Center.

Expert Advisory Group on Language Engineering. (1996). Text Corpora Working Group Reading Guide.

Francis, W. (1982). Problems Assembling and Computerizing Large Corpora. (S. Johansson, Ed.) 124-136.

Freer Bustamante, E., & Chavarría Cerdas, J. (Marzo-Junio de 1992). El desarrollo de la computación y su influencia en la medicina. *Revista costarricense de ciencias médicas, 13*, 59-70.

Fundació Víctor Grífols i Lucas. (2007). *Los fines de la medicina* (Segunda ed.). (A. Translations, Trad.) Barcelona, España: Fundació Víctor Grífols i Lucas.

Garson, J. (18 de Mayo de 1997). Connectionism. *Stanford Encyclopedia of Philosophy*.

Gómez Villegas, M. A. (1996). Origen de la teoría de la probabilidad. El teorema de Bayes. *Seminario "Orotava" de Historia de la Ciencia.* La Orotava, Canarias.

Haynes, R. B. (2001). Of studies, syntheses, synopses, and systems: the "4S" evolution of services for finding current best evidence. *Artículo(3), A11*, 134. ACP J Club.

Informed, S.A. de C.V. (2014). *PR Vademécum México*. Recuperado el 13 de Abril de 2014, de http://mx.prvademecum.com/

Iniesta Navalón, C., Urbieta Sanz, E., & Gascón Cánovas, J. J. (2011). Análisis de las interacciones medicamentosas asociadas a la farmacoterapia domiciliaria en pacientes ancianos hospitalizados. *Revista Clínica Española*. doi:10.1016/j.rce.2011.04.005

Instituto Hondureño de Seguridad Social. (Marzo de 2009). Cuadro Básico de Medicamentos. Tegucigalpa, Honduras.

Instituto Mexicano del Seguro Social. (Marzo de 2015). Cuadro Básico de Medicamentos.

International Organization for Standardization. (Diciembre de 2010). Systems and software engineering -- Vocabulary. *SO/IEC/IEEE 24765:2010(E)*. doi:10.1109

Iribarren, I. L. (1973). *Topología de espacios métricos* (Primera ed.). México: Limusa Wiley S.A.

Kramer, M. S., Leventhal, J. M., Hutchinson, T. A., & Feinstein, A. R. (1979). An Algorithm for the operational assessment of adverse drug reactions. I. Background, description, and instructions for use. *The Journal of the American Medical Association*(7), 623-632. doi:10.1001/jama.1979.03300070019017

Laplace, P. S. (1774). *Mémoire sur la probabilité des causes par le évènemens* (Vol. 8). París, Francia.

Laplace, P. S. (1812). *Théorie analytique des probabilités* (Primera ed.). (Courcier, Ed.) París, Francia.

Leech, G. (1992). Corpora theories of linguistic performance. (J. Svartvik, Ed.) *Directions in Corpus Linguistics*, 105-122.

Lim, H. S. (1998). Estudio sintáctico-semántico de la ambigüedad del sintagma nominal del español. *ASELE. IX*. Centro Virtual Cervantes.

Mayoral Asensio, R. (Junio de 1997). La traducción de la variación lingüística. *Tesis Doctoral, I*. Granada, España.

McNamee, P., & Mayfield, J. (2004). Character n-gram tokenization for european language text retrieval. *Information retrieval*(7), 73-97.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (Edits.). (2009). *Machine Learning, Neural and Statistical Classification*. Overseas Press.

Morales Bustamante, Ó. Á. (Enero de 2012). Procedimiento para verificar la idoneidad de la prescripción dentro de los servicios médicos. *Manual de procedimientos*. México, D.F., México.

Moreno Seco, F. (24 de Febrero de 2004). Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano. *Tesis*. Alicante, España.

Naranjo, C. A., Busto, U., & Sellers, E. M. (1981). A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*(30), 239-245.

Ongallo Chanclón, C., & Gallego Gil, D. J. (2003). *Conocimiento y gestión: La gestión del conocimiento para la mejora de las personas y las organizaciones* (Primera ed.). Madrid, España: Pearson Alambra.

Organización de Consumidores y Usuarios. (Octubre-Noviembre de 2006). Las formas farmacéuticas. *OCU-Salud*(68), 38.

PLM México. (2013). *Diccionario de Especialidades Farmacéuticas 2013*. Recuperado el 16 de Diciembre de 2013, de http://www.medicamentosplm.com/

Porter, M. (2014). *Snowball.* Recuperado el 8 de Enero de 2015, de http://snowball.tartarus.org

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.

Rancaño García, I., Rodrigo Pendás, J. A., Villa Estébanez, R., Abdelsater Fayad, M., Díaz Pérez, R., & Álvarez García, D. (2003). Evaluación de las páginas web en lengua española útiles para el médico de atención primaria. *Aten Primaria, 31*(6), 575-584.

Registered Nurses Association of Ontario. (1990). Promoting asthma control in children. *Clinical practice guidelines: directions for a new program, 2004*. (M. J. Field, & K. N. Lohr, Edits.) Toronto, Canada: National Academy Press.

Rotaetxe, R., Vicente, D., Etxeberria, A., Mozo, C., Larrañaga, M., Valverde, E., . . . Iturrioz, P. (2000). *Evaluación de la variabilidad e idoneidad en la prescripción de antimicrobianos en atención primaria en la Comunidad Autónoma del País Vasco. Recomendaciones de uso apropiado.* Investigación Comisionada, Gobierno Vasco, Departamento de Sanidad.

Ruiz-Shulcloper, J. (Septiembre de 2000). Logical Combinatorial Pattern Recognition. *Foro Iberoamericano de Reconocimiento de Patrones*, 123-128.

Sabines, J. (2011). *Antología poética* (Cuarta ed.). (G. Flores Liera, Ed.) México: Fondo de Cultura Económica.

Secretaría de Salud. (4 de Abril de 2013). *Misión y visión de la Secretaría de Salud*. Recuperado el 13 de Marzo de 2014, de Secretaría de Salud de México: http://portal.salud.gob.mx/contenidos/conoce_salud/mision_y_vision/misionvision.html

The Cochrane Collaboration. (2011). Manual Cochrane de revisiones sistemáticas de intervenciones. (J. P. Higgins, S. Green, Edits., & C. C. Iberoamericano, Trad.) The Cochrane Collaboration.

Truven Health Analytics. (2014). *Evidence-based clinical decision support*. Recuperado el 26 de Febrero de 2014, de Micromedex solutions: http://micromedex.com/