



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS

FACULTAD 2

TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO EN
CIENCIAS INFORMÁTICAS

***Técnica de Reconocimiento de Patrones para identificar la semejanza entre
medicamentos teniendo en cuenta las reacciones adversas que causan***

AUTORES:

Tania Garcia González

Ernesto Alejandro Yero Oses

TUTORES:

MSc. Annia Arencibia Morales

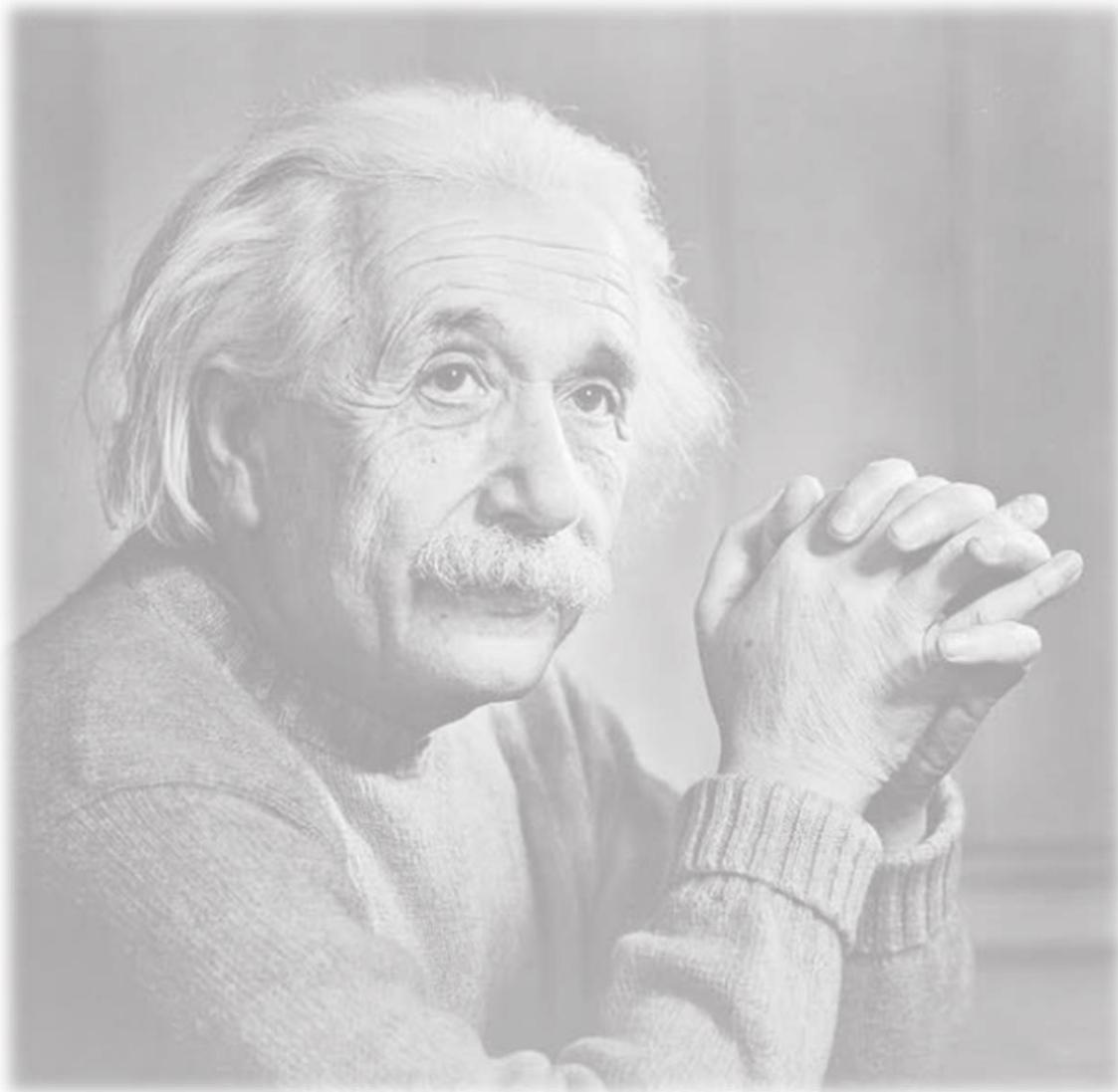
Ing. Yoiler Joaquín Frometa Moreno

La Habana, mayo de 2015

“Año 57 de la Revolución”



"... conectados al futuro, conectados a la Revolución."



*"Toda nuestra ciencia, comparada con la realidad, es primitiva e infantil
y sin embargo es lo máspreciado que tenemos"*

Albert Einstein.

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo. Para que así conste firmamos la presente a los ____ días del mes de junio del año 2015.

Tania Garcia González
Firma del Autor

Ernesto Alejandro Yero Oses
Firma del Autor

MSc. Annia Arencibia Morales
Firma del Tutor

Ing. Yoiler Joaquín Frometa Moreno
Firma del Tutor

DATOS DE CONTACTO

MSc. Annia Arencibia Morales (aarencibia@uci.cu): graduada de Ingeniero en Ciencias Informáticas en la Universidad de las Ciencias Informáticas (UCI) en el año 2007. Pertenece al Centro de Informática Médica (CESIM), actualmente posee la categoría docente de profesor Asistente. Máster en Informática Aplicada. Imparte Postgrado de Ingeniería de Software. Pertenece al Departamento de Desarrollo de Aplicaciones.

Ing. Yoiler Joaquín Frometa Moreno (yfrometa@uci.cu): graduado de Ingeniería en Ciencias Informáticas en la Universidad de las Ciencias Informáticas en el año 2007. Profesor de la Universidad de las Ciencias Informáticas, posee categoría docente de Profesor Instructor. Se desempeña como profesor de Almacenes de datos en la Facultad 2.

AGRADECIMIENTOS

Agradecimientos generales

A los tutores, por su constante orientación, valiosos comentarios y sugerencias en cada una de las etapas, así como por su disposición en todo momento para resolver cualquier duda surgida.

A las doctoras Tania Yero y Amarilis Valdés, que a pesar de todas sus responsabilidades sacaron un tiempito para la elaboración y aplicación de la encuesta.

A las profes Natalia y Yunia, por enseñarnos los conocimientos básicos necesarios para poder enfrentar el desarrollo de la investigación.

Agradecimientos de Tania García González

A mis hermanos que, por tratar siempre de ser un buen ejemplo a seguir, son los que han hecho que no haya pensado en mirar para atrás, sino siempre al frente.

A mis padres por ser mi guía todos estos años y por enseñarme que con esfuerzo se logra todo lo que uno quiere.

A mis abuelos, que me criaron y malcriaron pero siempre enseñándome desde sus experiencias con todo el amor del mundo.

A mis suegros, Cecilia y Rolando, que han sido como mis padres los últimos cuatro años, me han aconsejado, apoyado y acogido como una hija.

A mi madrina, Lioska, que es un motor impulsor grandísimo en mi vida, siempre alentándome a alcanzar un futuro mejor.

A mis tías, Elinor, Margot, Esperanza, Nohalys, Martha que me han apoyado con todo su cariño estos 5 años.

Agradecimientos de Ernesto A. Yero Osés

A mi familia por todo el esfuerzo que han hecho para lograr mis sueños y metas.

A la Revolución por brindar la posibilidad de estudiar en esta Universidad de Excelencia.

A mis amigos y compañeros que siempre estuvieron ahí para apoyar.

A todos aquellos que de una forma u otra ayudaron en el desarrollo de este trabajo.

DEDICATORIA

Dedicatoria de Tania Garcia González

Dedico el trabajo a mis padres y mis hermanos. Mamá y papi espero hacer recíproco todo lo que han hecho por mí. Liset y Marquito espero ser un ejemplo digno a seguir y que vean que con esfuerzo todo se logra. ¡Si yo pude ustedes también!

Dedicatoria de Ernesto A. Yero Oses

A mi familia, por el cariño y la paciencia con que me han educado, por guiarme siempre por el camino correcto al cumplimiento de mis metas.

A todas aquellas personas que han contribuido de una forma u otra con mi formación profesional, a mis tutores, mis compañeros, mis profesores y mis amigos.

A mi hermano, por obligarme a ser cada vez mejor, por tenerme como meta seguir aunque no lo diga, y porque no tengo otro.

RESUMEN

Los medicamentos son muy utilizados en la actualidad, estos curan enfermedades y al mismo tiempo pueden causar alguna reacción adversa a medicamentos (RAM) a quienes los consumen. Al manifestarse en un paciente una RAM, el personal facultativo debe ser capaz de cambiar el medicamento por un sustituto con igual o mejor efectividad sobre la enfermedad. Esta tarea se realiza de forma manual, consultando el Formulario Nacional de Medicamentos (FNM), libro extenso que no cuenta con la semejanza entre los medicamentos.

Con el objetivo de identificar la semejanza de los medicamentos según las RAM que estos causan se le adiciona al producto Sistema para el Control Farmacológico: Synta, del Centro de Informática Médica de la Universidad de las Ciencias Informáticas, una funcionalidad para determinar la semejanza existente entre los medicamentos atendiendo a las reacciones adversas que ocasionan. El desarrollo de esta funcionalidad se realizó aplicando, dentro de las técnicas de Reconocimiento de Patrones, la selección de rasgos y la clasificación no supervisada. Para ello se agruparon los medicamentos, según la RAM que causan, mediante el algoritmo de LC-Conceptual. Se calculó el peso informacional asociado a cada rasgo (RAM) calculando los testores típicos, mediante el algoritmo Fast-BR. Se determinó la semejanza entre los medicamentos atendiendo a los efectos adversos que causan.

El sistema permite conocer la semejanza que posee el medicamento causante de la reacción adversa con el conjunto de medicamentos suplentes, atendiendo a la RAM que causan facilitando la prescripción de los mismos.

Palabras clave: agrupamiento, Formulario Nacional de Medicamentos, función de semejanza, rasgo, Reacción adversa a medicamentos, Reconocimiento de Patrones.

ÍNDICE

Introducción.....	9
Capítulo 1 Conceptos del Reconocimiento de Patrones, sistemas y herramientas asociados a la determinación de semejanza entre medicamentos según RAM	16
1.1 Reconocimiento de patrones.....	16
1.2 Peso Informacional de los Rasgos.....	19
1.3 Agrupamiento de los datos (<i>Clustering</i>)	23
1.4 Soluciones existentes a nivel internacional y nacional	29
1.5 Herramientas y tecnologías a utilizar en la propuesta de solución	32
Capítulo 2 Propuesta de solución para el cálculo de la semejanza entre medicamentos según RAM	38
2.1 Propuesta de solución.....	38
2.2 Definición de objetos, rasgos y dominios del problema	39
2.3 Agrupar medicamentos según RAM (Paso 1)	40
2.4 Cálculo del peso informacional de los rasgos (Paso 2)	42
2.5 Propuesta de función de semejanza a aplicar a los medicamentos (Paso 3)	47
2.6 Requisitos no funcionales (RNF).....	48
Capítulo 3 Validación de la solución propuesta	50
3.1 Tamaño de la muestra de medicamentos seleccionados	50
3.2 Validación de los resultados de la función de semejanza.....	52
Conclusiones.....	62
Recomendaciones.....	63
Referencias	64
Bibliografía	69
Anexos	73
Anexo 1 Fases del desarrollo de un nuevo medicamento y relación con su seguridad	73
Anexo 2 Encuesta realizada a los especialistas en medicina.....	73

Anexo 3 Descripción del algoritmo LC-Conceptual	75
Anexo 4 Descripción del algoritmo Fast-BR	75
Anexo 5 Medicamentos incluidos para las pruebas del sistema.....	77
Anexo 6 Reportes de reacciones adversas en Cuba 2000-2014.....	78

INTRODUCCIÓN

Todos los seres humanos enferman en algún momento de su vida y los medicamentos constituyen la herramienta terapéutica más utilizada para su curación. La Organización Mundial de Salud (OMS), establece que las palabras fármacos o medicamentos son sinónimos y lo define como cualquier sustancia, presente en un producto farmacéutico que se utiliza para explorar o modificar aparatos fisiológicos o enfermedades en beneficio del receptor. [1]

Desde los primeros estudios de utilización de medicamentos quedó patente la necesidad de disponer de un sistema internacional único de clasificación de los mismos. El establecimiento de un sistema de clasificación aceptado internacionalmente fue esencial para: [2]

- comparar el consumo de un país a otro,
- el seguimiento del patrón nacional de consumo,
- el análisis de los cambios en el consumo a lo largo del tiempo en un país determinado,
- la farmacovigilancia y cuando se necesite información sistemática sobre los medicamentos y sobre su uso.

Existen varias clasificaciones para los medicamentos, de las cuales la más aceptada es aquella que se basa en las principales acciones farmacológicas¹ o usos terapéuticos². La clasificación Anatómica Terapéutica Química (ATC, por sus siglas en inglés) es un sistema europeo de codificación de sustancias farmacéuticas y medicamentos en niveles con arreglo al sistema u órgano efector, efecto farmacológico, indicaciones terapéuticas y la estructura química de un fármaco [3]. Además, los medicamentos, se pueden clasificar también por su constitución química, sus formas de aplicación, sus efectos o su grado de efectividad. [4]

La OMS, elaboró y publicó en 1977 la primera Lista Modelo de Medicamentos Esenciales³ permitiendo identificar las necesidades específicas de medicamentos en los diversos niveles del sistema de salud; también en el terreno económico, reduciendo el impacto del gasto en medicamentos [5]. Los Medicamentos esenciales son los que satisfacen las necesidades prioritarias de salud de la población. Se seleccionan teniendo debidamente en cuenta su pertinencia para la

¹ Mecanismo por el que un fármaco ejerce la acción deseada.

² Relativo a la terapéutica. Medicina: relativo al tratamiento médico de las enfermedades Algunas bacterias son usadas con fines terapéuticos.

³ Lista Modelo de Medicamentos Esenciales: sustancias terapéuticas que sirven para satisfacer las necesidades de atención de salud de la mayoría de la población. Se trata de medicamentos genéricos con una comprobada eficacia, un aceptable rango de seguridad y un coste económico asequible.

salud pública, pruebas de su validez y seguridad, y su eficacia comparativa en relación con el costo. Estos deben estar disponibles en los sistemas de salud en todo momento, en cantidades suficientes, en las formas farmacéuticas apropiadas, con garantía de la calidad e información adecuada, a un precio que los pacientes y la comunidad puedan pagar. La aplicación del concepto de medicamentos esenciales ha de ser flexible y adaptable a muchas situaciones diferentes. La determinación de los medicamentos que se consideran esenciales es responsabilidad de cada país. [6]

Países como España, Chile, Perú, Cuba, entre otros, además de contar con los medicamentos esenciales tienen un Formulario Nacional de Medicamentos (FNM); el cual es un listado de medicamentos elaborados por la autoridad de salud correspondiente a cada país, o sociedades médicas y farmacológicas. Los medicamentos se identifican por la Denominación Común Internacional. [7] [8]

El FNM cubano es un documento que reúne la lista de los medicamentos y productos medicamentosos que han sido seleccionados para ser utilizados en el país, con la información farmacológica para promover su uso seguro y efectivo [9]. Su objetivo principal es satisfacer las necesidades de médicos, farmacéuticos, enfermeros y todos los profesionales relacionados con la información sobre medicamentos y su uso [8]. Hasta la actualidad este cuenta con tres ediciones, la primera en el 2003, la segunda en el 2006 y por último la del 2011 [8], en formato duro. La versión más reciente y actualizada es la del 2014 que se encuentra digital en los portales de la salud cubana [10].

En la práctica, el uso de la mayor parte de los medicamentos implica un cierto riesgo que debe intentar reducirse al mínimo por lo que al desarrollarse un medicamento se tienen en cuenta cuatro fases (**Anexo 1**) [11]:

- En la fase preclínica se estudian los efectos farmacológicos del nuevo medicamento y su toxicidad y se establece un índice terapéutico que es esencial para que se inicie el estudio clínico.
- En la fase I se estudia la seguridad del nuevo medicamento en voluntarios sanos y los resultados de estos ensayos clínicos pueden ser críticos para la continuidad del estudio.
- En los ensayos clínicos de las fases II y III se valoran conjuntamente la eficacia y seguridad del medicamento y se establece una relación eficacia-toxicidad que será decisiva para el registro del nuevo medicamento.
- En la fase IV o fase de pos-comercialización se perfilan aspectos específicos de la toxicidad del medicamento.

Como resultado del proceso son detectados los efectos indeseados del medicamento o Reacciones Adversas al Medicamento (RAM) las cuales son archivadas por los diferentes países para su conocimiento y prevención.

Una RAM es una respuesta nociva y no intencionada, que se produce a dosis aplicadas normalmente en el ser humano para la profilaxis⁴, el diagnóstico o terapia de una enfermedad, o para la modificación de la función fisiológica. [12]

Las RAM afectan a personas de todos los países; por lo que el conocimiento incompleto sobre la frecuencia y gravedad de los efectos adversos de los medicamentos es una de las mayores debilidades de la terapéutica moderna, no solo en Cuba sino internacionalmente. [13] Por este motivo en 1968 se puso en marcha el Programa OMS de Vigilancia Farmacéutica Internacional, coordinado por la OMS y su Centro Colaborador de Uppsala, con la idea de aglutinar los datos existentes sobre las reacciones adversas a los medicamentos. Inicialmente, el programa, comenzó con un estudio piloto de 10 países pero en la actualidad forman parte de él más de 180 países [14]:



Figura 1 Composición de la red del Programa OMS de Vigilancia Farmacéutica Internacional

En Cuba se ha implementado un programa de farmacovigilancia⁵ para todos los profesionales y técnicos de la salud, con un único método de validación e identificación de los riesgos asociados con los medicamentos comercializados. El Departamento de Farmacoepidemiología (DF) del Ministerio de Salud Pública de Cuba (MINSAP) es el encargado de realizar acciones que permitan

⁴ Conjunto de medidas que se aplican para prevenir las enfermedades

⁵ La ciencia y las actividades relativas a la detección, evaluación y prevención de los riesgos asociados a los medicamentos o cualquier otro problema relacionado con los fármacos.

alcanzar un uso racional de los medicamentos por medio de la coordinación y dirección metodológica de la red cubana de Farmacoepidemiología. [15]

En el DF del MINSAP en el periodo de enero a diciembre del 2014 se notificaron 20 381 sospechas de reacciones adversas medicamentosas superior a la media anual de 15 534 reportes (**Anexo 6**), para una tasa de reportes de 1 812 por millón de habitantes (Figura 2). [16]

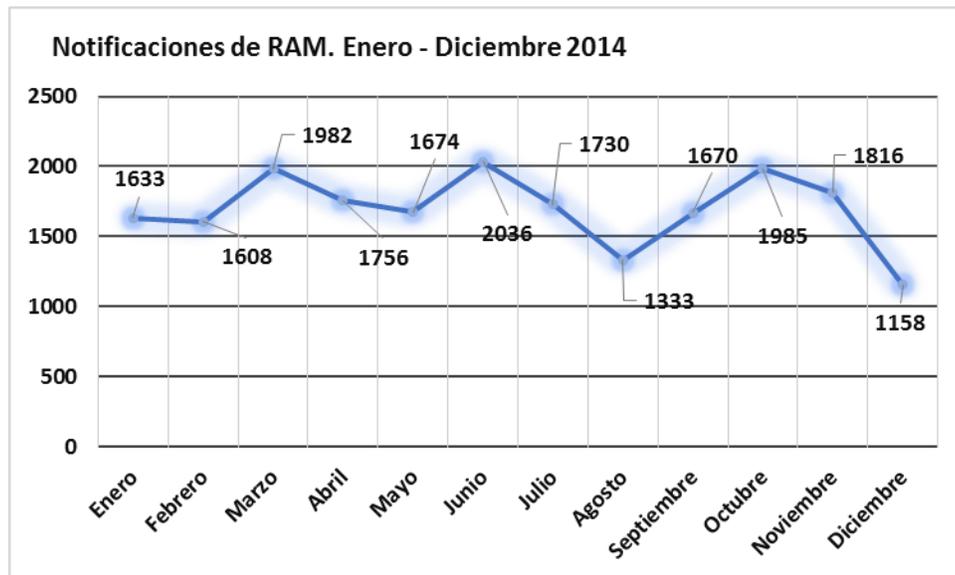


Figura 2 Notificaciones de RAM. Enero - Diciembre 2014

El continuo avance de la tecnología, la informática, y su notoria gravitación en los diversos campos de los Sistemas de Salud, ofrece nuevas perspectivas en relación con la distribución y control de los medicamentos; razón por la cual diversas empresas de desarrollo y muchos países han dedicado tiempo a la creación y estudio de sistemas que traten temas de medicamentos o gestión farmacéutica.

En la Universidad de las Ciencias Informáticas (UCI) se encuentra el Centro de Informática Médica (CESIM). Uno de los proyectos desarrollados es un producto llamado Sistema para el Control Farmacológico: Synta. Dicho producto cuenta con los módulos: Control de Recetas Médicas, Reacciones Adversas a Medicamentos, Consumo de Medicamentos y Nomenclador, encargados de informatizar los procesos principales del DF del MINSAP. [17]

El módulo Reacciones Adversas a Medicamentos (MRAM) es el encargado de recopilar todas las RAM ocurridas en Cuba, recolectándolas desde la misma Unidad de Salud (US), responsabilidad que debe tener el médico que identifica la RAM en los pacientes. [17]

Cuando un paciente llega a una consulta sufriendo una RAM, ocasionada por el consumo de un medicamento prescrito, el médico debe ser capaz de cambiar el medicamento que le ocasiona la RAM por un sustituto con igual o mejor efectividad sobre la enfermedad. Para que este proceso sea desarrollado con éxito el personal facultativo debe conocer la semejanza que posee el medicamento causante de la reacción adversa con el conjunto de medicamentos suplentes atendiendo a las RAM que causan.

No existe en estos momentos a disposición del personal facultativo ninguna fuente en la cual puedan apoyarse para obtener los resultados de semejanza entre medicamentos según RAM, para ello se apoya en el FNM desarrollado por el MINSAP, que debe ser actualizado cada dos años, donde se almacenan los medicamentos y sus reacciones adversas teniendo en cuenta su probabilidad de aparición. A partir de estos datos, y del conocimiento del médico para determinar para las diferentes situaciones de cuál debe ser el nuevo medicamento a suministrar en caso de aparición de una RAM, tarea que se dificulta debido a que:

- Los médicos, no siempre cuentan con la última versión del FNM; además de no poderlo consultar en todo momento.
- Muchas veces los médicos que están en las consultas no tienen la suficiente experiencia práctica para realizar este tipo de operación, con los mejores resultados.
- El FNM es un libro extenso escrito en lenguaje natural que el médico debe interpretar para poder realizar un cambio de medicamento.

El FNM no contiene la relación de semejanzas existentes entre los medicamentos teniendo en cuenta sus reacciones adversas. La búsqueda de la información en el FNM resulta compleja, pues se encuentra dispersa y no del todo concisa, haciendo el trabajo del médico engorroso, lento y no del todo confiable, teniendo en cuenta la gran cantidad de medicamentos existentes (alrededor de 700 medicamentos). Por ello en ocasiones la decisión tomada no es la más acertada ocasionando que el nuevo medicamento suministrado:

- mantenga la RAM.
- provoque una nueva RAM (ocasionando complicaciones más graves y en algunos casos hasta la muerte).

Al detectar todos los problemas planteados anteriormente se formula el siguiente **problema a resolver**: en el Formulario Nacional de Medicamentos se hace engorrosa la identificación de la semejanza de los medicamentos según las reacciones adversas que estos causan, dificultando la prescripción de un nuevo fármaco a los pacientes que tienen este padecimiento.

Este problema está enmarcado en el **objeto de estudio**: técnicas de Reconocimiento de Patrones en las reacciones adversas a medicamentos, centrándose en el **campo de acción**: la semejanza entre los medicamentos del Formulario Nacional de Medicamentos atendiendo a sus reacciones adversas.

Teniendo como **objetivo general**: identificar la semejanza de los medicamentos según las RAM que estos causan en el Formulario Nacional de Medicamentos, mediante el uso de Reconocimiento de Patrones, facilitando la prescripción de los mismos.

Como tareas de la investigación se proponen las siguientes:

1. Realizar un análisis de los datos existentes en el FNM que permitan ser usados para establecer semejanzas entre medicamentos según RAM.
2. Realizar un análisis de los sistemas informáticos que gestionan información relacionada con la farmacovigilancia, a nivel nacional e internacional, identificando aquellos que determinan la semejanza entre medicamentos atendiendo a sus RAM.
3. Identificar las técnicas utilizadas en la determinación de semejanzas entre objetos, definiendo los conceptos necesarios.
4. Crear una Base de Conocimientos, donde se almacene la información referente a los medicamentos, sus grupos farmacológicos y sus Reacciones Adversas a Medicamentos, para ser utilizada en los algoritmos seleccionados.
5. Implementar los algoritmos seleccionados que permitan conocer la semejanza entre los medicamentos de un mismo grupo farmacológico y crear grupos de medicamentos teniendo en cuenta su semejanza en cuanto a reacciones adversas.
6. Integrar la solución al MRAM del producto Synta, siguiendo las pautas definidas por el Departamento de Desarrollo de Aplicaciones del CESIM para la realización de este proyecto.

Entre los beneficios esperados con el desarrollo de la presente investigación, se encuentran:

- Ofrecer información a los médicos que les permita conocer las semejanzas entre los medicamentos del FNM, teniendo en cuenta las reacciones adversas que estos causan.
- Facilitar al médico una herramienta, que le permita mejorar la prescripción de pacientes que acudan a las consultas con una RAM.

Para el desarrollo del trabajo se utilizarán métodos de investigación; a continuación se listan los mismos:

Métodos Teóricos

- Analítico sintético: para realizar la valoración de los algoritmos de reconocimiento de patrones existentes, así como las investigaciones similares realizadas con anterioridad; además para analizar la estructura de las RAM en el FNM.
- Inductivo-Deductivo: en la aplicación de casos de pruebas al sistema, llegando a conclusiones a partir de las respuestas proporcionadas por este.

Métodos Empíricos

- Encuesta: se realizaron entrevistas a diferentes médicos para determinar las tendencias de prescripción y compararlas con las respuestas que brinda la funcionalidad realizada.

El documento está estructurado en tres capítulos, siendo estos:

Capítulo 1: Conceptos del Reconocimiento de Patrones, sistemas y herramientas asociados a la determinación de semejanza entre medicamentos según RAM, incluye un análisis y revisión del estado del arte, a nivel internacional y nacional de Reconocimiento de Patrones en el campo de la salud, específicamente determinando la semejanza entre los medicamentos del FNM según sus RAM. Se realiza una descripción de los principales conceptos necesarios para la mejor comprensión de la investigación; además se describen las herramientas a utilizar en el desarrollo de la propuesta de solución.

Capítulo 2: Propuesta de solución para el cálculo de la semejanza entre medicamentos según RAM, se describen los principales conceptos asociados a las relaciones de semejanza entre los medicamentos y las técnicas seleccionadas para la creación de los algoritmos utilizados. Se desarrolla la propuesta de solución centrándose en las características de la función de semejanza empleada. Se hace alusión a los principales aspectos a tener en cuenta para la correcta implementación del sistema

Capítulo 3: Validación de la solución propuesta, tiene como objetivo, mostrar los resultados de la validación la propuesta de función para la determinación de semejanza entre medicamentos según sus RAM. Para lograrlo se demuestra apoyados en las técnicas de metodología de investigación científica la validez de la muestra de datos seleccionada para realizar el agrupamiento. Posteriormente, se muestran los resultados de una encuesta aplicada a los profesionales de la salud, sobre situaciones casos reales para comparar los resultados de los médicos por los brindados por el sistema.

CAPÍTULO 1 CONCEPTOS DEL RECONOCIMIENTO DE PATRONES, SISTEMAS Y HERRAMIENTAS ASOCIADOS A LA DETERMINACIÓN DE SEMEJANZA ENTRE MEDICAMENTOS SEGÚN RAM

En el presente capítulo se abordan temas relacionados al Reconocimiento de Patrones (RP) y a la Teoría de Testores, enfocándose en su aplicación para la determinación de semejanzas entre objetos. Se hace alusión a las características generales de los algoritmos de agrupamientos, necesarios para la identificación de los pesos informacionales asociados a los medicamentos. Del mismo modo se incluye una descripción de las tecnologías, herramientas y lenguajes utilizados. Se realiza un análisis sobre los sistemas similares existentes vinculados a la farmacovigilancia.

1.1 Reconocimiento de patrones.

Reconocer patrones⁶ es una actividad casi involuntaria en muchos seres vivos, pero el RP, no se comienza a conformar como ciencia hasta finales de los años 50 con la publicación del libro *Principios de la Neurodinámica de Rosenblatt* [18].

Desde sus primeros momentos el RP estuvo fuertemente ligado a la identificación, al reconocimiento de imágenes. Por ello no es de extrañar que esta disciplina se haya asociado por muchos especialistas exclusivamente con el procesamiento y análisis de imágenes. El mismo ha sido definido por diferentes autores como:

Según Diego Sebastián Comas y Gustavo Javier Meshino: es el proceso en el cual se agrupa un conjunto de datos en una determinada cantidad de grupos (*clusters*) o etiquetas mediante un mecanismo de decisión. Este mecanismo depende de la técnica particular que se utilice y es el que más define al algoritmo de reconocimiento de patrones. [19]

Según Shulcloper: es la ciencia de carácter multidisciplinario que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos y/o abstractos, con el propósito de extraer –mediante dispositivos computacionales y/o el hombre– la información que permita establecer propiedades y/o vínculos de o entre conjuntos de dichos objetos. [20]

⁶ Punto de referencia. Patrón es sinónimo de objeto, en algunas ocasiones es conveniente establecer diferencias entre los objetos acerca de los cuales se conocen ciertas propiedades (por ejemplo, la pertenencia a una clase, tipo, etc.). Al identificar entre muchos objetos a uno conocido se está identificando, o reconociendo, un patrón.

Según Anil K. Jain: es el estudio de cómo las máquinas pueden observando el ambiente aprender a distinguir patrones de interés de un fondo y realizar decisiones razonables sobre las categorías de los mismos. [21]

En esta investigación, se adopta el concepto de José Ruíz Shulcloper.

1.1.1 Conceptos básicos asociados

Se hace necesario para un mejor entendimiento de las principales características y definiciones que se abordarán del RP la definición de un grupo de conceptos básicos:

Se denomina **objeto** al término empleado en diferentes disciplinas para referirse a los entes sujetos a estudio. [18]

Se entiende el conjunto de todos los posibles objetos admisibles para los propósitos del estudio en cada caso particular como **universo de objetos**. [18]

Se llama **clase** a un conjunto de objetos que constituyen un tipo, una categoría, en el contexto de los objetos del universo. Siempre será un subconjunto propio del universo de objetos. [18]

Rasgo (atributo, variable, característica, propiedad, primitivas de un alfabeto) es un factor a tener en cuenta en el estudio de los objetos dados. Constituyen la vía real para poder estudiar, procesar, analizar a los objetos. Es decir, se tienen que estudiar descripciones de los objetos en términos de un conjunto de rasgos. [18]

Dominio de definición es el conjunto de valores admisibles que tiene asociado un rasgo. [18]

1.1.2 Problemas fundamentales de Reconocimiento de Patrones

Por problemas de RP según la bibliografía consultada [20] [18] se entienden todos aquellos relacionados con la clasificación de objetos y con la determinación de los factores que inciden en los mismos. Así, se consideran cuatro familias de problemas que se denominan respectivamente:

~ Selección de rasgos

Los problemas de selección de rasgos están basados en la descripción de los objetos en términos de un conjunto de rasgos teniendo dos usos principales:

- para reducir el número de rasgos, en términos de describir los objetos en modo eficiente.
- para encontrar los rasgos que inciden en el problema de manera determinante.

Son situaciones prácticas en las cuales el interés se centra en las descripciones de los objetos, en la importancia de las variables para describir el problema, en su poder descriptivo donde nos interesa conocer la influencia de los rasgos en el fenómeno o conjunto de objetos que se están estudiando.

~ **Clasificación supervisada**

En este tipo de problema dado un universo de objetos y el conocimiento acerca de la existencia de ciertas clases con características de especial interés y una muestra de objetos que pertenecen a cada una de ellas, el problema consiste en determinar para cada uno de los objetos no clasificados las relaciones de pertenencia de los mismos con cada una de las clases.

~ **Clasificación no supervisada**

En este tipo de problemas lo principal a resolver es encontrar las relaciones entre los objetos de un universo en términos de sus rasgos. Estas relaciones se establecen sobre la base del concepto de similaridad teniendo como propósito agrupar los objetos según su analogía pudiéndose conocer el número de grupos de objetos por los que se agruparía o no.

~ **Clasificación parcialmente supervisada**

Este tipo de problema consiste en una combinación de los problemas de clasificación anteriormente descritos. En el universo de objetos dado, se conoce la existencia de ciertas clases e incluso se tienen muestras de algunas de ellas pero no de todas. El problema central es clasificar nuevos objetos en estas circunstancias, en la que no se tienen muestras de todas las clases y en las que incluso pudieran existir clases que se desconocen.

1.1.3 Distancias o Métricas.

En el RP, la distancia entre dos objetos cualesquiera del espacio es una medida de cuán similares son de acuerdo a sus características, ya que éstas se escogen de forma tal que mientras más parecidos sean los objetos menor debe ser la distancia entre ellos y, por el contrario, los objetos muy lejanos deben ser poco similares. [18]

Definición 1: Un espacio métrico es un par (X, β) donde X es un conjunto ($X \neq \emptyset$) y β una distancia o métrica definida sobre X . Una función $d: X \times X \rightarrow \mathfrak{R}^+$, se dice que es una distancia o métrica si satisface los siguientes axiomas:

- 1 $\beta(x, y) \geq 0 \forall x, y \in X$, y $\beta(x, y) = 0$ si y solo si $x = y$
- 2 $\beta(x, y) = \beta(y, x) \forall x, y \in X$ (simetría)
- 3 $\beta(x, z) \leq \beta(x, y) + \beta(y, z) \leq \forall x, y, z \in X$ (desigualdad triangular)

La métrica más frecuentemente utilizada en toda la literatura es la métrica Euclídea [22]:

$$\beta(O_i, O_j) = \sqrt{\sum_{k=1}^n (x_k(O_i) - x_k(O_j))^2}$$

Ecuación 1 Distancia Euclidiana

Donde O_i y O_j son los objetos para los cuales se desea calcular la distancia, n es el número de características de los objetos del espacio y $x_k(O_i)$, $x_k(O_j)$ es el valor del atributo k -ésimo en los objetos O_i y O_j , respectivamente.

Ejemplos de Medidas de Semejanza

Otras de las funciones de semejanzas más utilizadas y descritas en la bibliografía consultada [20] [22] [23] son:

$$\beta(O_i, O_j) = \frac{\sum_{k=1}^n w_k * \delta_k(O_i, O_j)}{\sum_{k=1}^n w_k}$$

Ecuación 2 Distancia basada en la suma pesada de la comparación por rasgos

Donde $\delta(O_i, O_j)$ representa la función de similitud por rasgos, n es el número de rasgos a tener en cuenta y w_k la importancia asociada al rasgo k .

También se observa una más sencilla empleada fundamentalmente en algoritmos de pre-procesamiento de información, o en casos en los que no se puede contar con el peso informacional de cada uno de los rasgos:

$$\beta(O_i, O_j) = \sum_{k=1}^n \delta_k(O_i, O_j)$$

Ecuación 3 Distancia basada en la suma de la comparación por rasgos

1.2 Peso Informacional de los Rasgos

1.2.1 Teoría de Testores

El concepto de testor aparece por primera vez en un trabajo de los científicos rusos Cheguis y Yablonskii [24] en un intento por tratar de simplificar la tarea de encontrar desperfectos en circuitos eléctricos que podían ser modelados a través de funciones del álgebra de la lógica.

La determinación de todos los *testores típicos* (TT) de una matriz de entrenamiento⁷ (ME) encuentra gran aplicación en el RP, porque ellos permiten determinar el peso informacional de los rasgos.

Medida de la importancia informacional de los rasgos en función del concepto de testor

El enfoque basado en la *Teoría de Testores*, para definir una medida de la importancia de los rasgos, se basa en la idea de que un TT, según Zhuravliov, es un conjunto de columnas (rasgos) de la ME para las cuales se cumple que no existen filas iguales en clases diferentes, pero además si eliminamos una cualquiera de estas columnas las restantes perderán esta propiedad. [18]

La idea básica consiste en lo siguiente: dada la definición de testor y TT, se tiene que cada objeto se pone en correspondencia con un s-uplo⁸ que tiene la propiedad de ser diferente en al menos alguna coordenada con respecto a cualquier otro objeto que pertenezca a una clase diferente. Al pasar de un testor a un TT (eliminando rasgos) se llega a un conjunto irreducible de rasgos, donde cada rasgo resulta imprescindible para mantener las diferencias entre las clases. Esta es la propiedad que distingue a los TT. Es natural suponer, que si un rasgo aparece en muchos conjuntos irreducibles, resulta más difícil prescindir de él para describir de manera diferenciante las clases, o sea que él es más diferenciante. Sobre la base de esta idea, Zhuravliov formula su definición de peso informacional de un rasgo como la frecuencia relativa de aparición de ese rasgo en la familia de los TT. [18]

Sea τ el número de TT que tiene una cierta matriz de entrenamiento de un problema de reconocimiento de patrones y sea $\tau(i)$ el número de aquellos TT en los que aparece la columna correspondiente al rasgo x_i . Se dice que el peso informacional de este rasgo viene dado por la magnitud [18]:

$$P(x_i) = \frac{\tau(i)}{\tau}, \text{ para } i = 1, \dots, n, x_i \in \mathbb{R}$$

Ecuación 4 Peso informacional del rasgo

Haciendo uso de esta definición se han resuelto una serie de problemas de gran interés, sin embargo, la expresión anterior propuesta por Zhuravliov para la evaluación del peso informacional de los rasgos, si bien resulta intuitiva, al considerar más importante un rasgo mientras más veces

⁷ Se define como la matriz que almacena a la relación de todos los objetos para el entrenamiento y sus rasgos, donde cada fila describe a un objeto a través de sus rasgos (columnas).

⁸ En matemáticas, una tupla, es una lista ordenada de elementos. Una s-tupla es una secuencia (o lista ordenada) de s elementos, siendo s un número natural (entero no-negativo).

aparece en la familia de los TT, no considera las longitudes de los testores que contienen al rasgo en cuestión, de modo que por ejemplo, si la familia de TT de un problema fuera:

$$Q = \{\{x_1\}, \{x_2, x_5\}, \{x_2, x_3, x_4\}\}$$

Se tiene que:

$$P(x_1)=P(x_3)=P(x_4)=P(x_5)=0.333 \text{ y } P(x_2)=0.666.$$

Resulta natural pensar que el rasgo x_1 es más relevante que x_3 , x_4 y x_5 pues aunque sólo aparece en un TT no requiere de ningún otro rasgo para lograr diferenciar los objetos que están en clases diferentes, x_5 requiere de otro rasgo y por su parte x_3 y x_4 requieren de otros dos; luego si se denotara por $D(x)$ la relevancia del rasgo x , podría esperarse que:

$$D(x_1) > D(x_5) > D(x_3)=D(x_4).$$

Este análisis conduce a concluir que considerar la frecuencia de aparición de un rasgo en la familia de los TT como medida de su importancia, aunque resulta bastante natural a partir de la interpretación de un TT como un conjunto irreducible de rasgos diferenciadores, en algunos casos puede llevar a interpretaciones erróneas.[18]

También resulta natural considerar un rasgo más importante en la medida en que es menor la longitud de los TT en que aparece.

Se asociará a cada rasgo una magnitud, que depende de estas longitudes y que se denotará por $L(x)$ [18]:

$$L(x) = \frac{\sum_{t \in \psi^*(x)} \frac{1}{|t|}}{|\psi^*(x)|}$$

Ecuación 5 Peso del rasgo dependiente de la longitud del testor

Para el ejemplo anterior, se tiene que

$$L(x_1)=1, L(x_2)=0.418, L(x_3)=L(x_4)=0.333 \text{ y } L(x_5)=0.5.$$

Estos valores no se corresponden "totalmente" con lo que se señalaba como deseado con respecto a la relevancia, pues aunque x_1 aparece con máximo valor, x_5 tiene asociado un valor mayor que el de x_2 y debía esperarse lo contrario.

Como se ha dicho, $P(x)$ toma en cuenta sólo la frecuencia de aparición del rasgo en la familia de TT; $L(x)$ por su parte considera únicamente la longitud de los TT en los que aparece el rasgo.

Se definirá una nueva medida de la relevancia del rasgo en función de estas dos magnitudes; es decir, en general se considerará que $D(x) = \rho(P(x), L(x))$

Considérese la siguiente expresión [18]:

$$D(x) = \alpha P(x) + \beta L(x) \text{ con } \alpha, \beta > 0 \text{ y } \alpha + \beta = 1$$

α y β son dos parámetros que ponderan la participación o influencia de $P(x)$ y de $L(x)$ respectivamente en D . Esto es, la importancia que se le concede para la relevancia a la frecuencia de aparición y a la longitud de los testores.

Si se le concede a ambos la misma importancia se puede considerar $\alpha = \beta = 0.5$. Para estos valores, en el ejemplo anterior se tendrá.

$$D(x_1) = 0.666, D(x_2) = 0.542, D(x_3) = D(x_4) = 0.333 \text{ y } D(x_5) = 0.416$$

Resulta necesario señalar que el cálculo del peso informacional de los rasgos de los objetos haciendo uso de la *Teoría de Testores*, implica un conocimiento previo de las clases (grupos) a los que pertenecen los mismos, por tal tratar de determinar los TT en un conjunto de objetos no clasificados es una actividad que va en contra del propio concepto de TT.

1.2.2 Teoría de la información

La teoría de la información es una rama de la teoría matemática de la probabilidad y la estadística, que estudia la información y todo lo relacionado con ella: canales, compresión de datos, criptografía, aprendizaje, etc. Forma la piedra angular sobre la que se ha desarrollado toda la teoría actual de la comunicación y la codificación. La teoría de la información abarca todas las formas de transmisión y almacenamiento de información. Un concepto fundamental en la teoría de la información es la cantidad de información contenida en un mensaje, ésta es un valor matemático bien definido y medible. El término Entropía ha sido definido para designar la cantidad de información media de esos mensajes, puede ser entendida intuitivamente como el grado de "desorden" en un sistema. [25]

Para relacionar la cantidad de información con la probabilidad, Shannon presentó la siguiente fórmula:

$$l = \log_2 \frac{1}{p}$$

Ecuación 6 Cantidad de Información según Shannon

Donde p es la probabilidad del mensaje que se transmite.

Si en un conjunto de mensajes, sus probabilidades son iguales, la fórmula para calcular la entropía sería: $H = \log_2 [N]$ donde N es el número de mensajes posibles en el conjunto.

Para definir formalmente la entropía de una variable, se considera: X una variable aleatoria discreta con posibles valores x_1, \dots, x_N , y con probabilidades asociadas $p(x_1), \dots, p(x_N)$, $I(X)$ la variable aleatoria cantidad de información asociada a X . Se define la entropía de Shannon de la variable aleatoria X , como la esperanza matemática de la variable aleatoria $I(X)$ asociada [26]:

$$H(X) = E[I(X)] = - \sum_{i=1}^x p(x_i) \log_2 p(x_i)$$

Ecuación 7 Entropía de Shannon

Si $p(x_i) = 0$ para algún valor de i , la indeterminación se resuelve asignándole el valor 0.

La entropía de X se interpreta como el valor esperado de $\log_2 \left[\frac{1}{p(x)} \right]$, donde X es tomada de acuerdo a la densidad de probabilidad $p(x)$.

Propiedades de la entropía:

- $0 \leq H(x) \leq \log_2 N$
- $H(X) = 0 \Leftrightarrow \exists x_i \text{ con } p(x_i) = 1$

1.3 Agrupamiento de los datos (*Clustering*)

El agrupamiento como técnicas dentro del Reconocimiento de Patrones se encarga de organizar una colección de objetos en clases o grupos, de forma tal que los objetos pertenecientes a un mismo grupo sean lo suficientemente similares como para poder inferir que son del mismo tipo y los objetos pertenecientes a grupos distintos sean lo suficientemente diferentes como para poder afirmar que son de tipos diferentes. [27]

Según Pons Porrata: dado un conjunto de objetos definidos en términos de un conjunto de rasgos, los algoritmos de agrupamiento intentan construir particiones o cubrimientos de este conjunto, donde la semejanza intra-grupo sea máxima y la semejanza inter-grupos sea mínima. [28]

Según la bibliografía consultada los algoritmos de agrupamiento pueden ser clasificados atendiendo a diversos criterios [29]. Entre los más utilizados se encuentran:

- Según el tipo de grupos que forma el algoritmo.
- Según la capacidad que tiene el algoritmo para procesar los cambios.
- Según el tipo de estructuración que forma con los objetos.
- Según el enfoque utilizado para formar los grupos.

1.3.1 Conceptos básicos asociados

Sean ϑ una colección de objetos, β una función de semejanza simétrica y β_0 un umbral de semejanza con un valor definido por el usuario.

Definición 2: El umbral de semejanza β_0 es la cota inferior de los posibles valores de semejanza existentes para cada objeto con respecto a otro, pudiendo ser el mismo especificado por un especialista o calculado [18].

Definición 3: Se dice que dos objetos $O_i, O_j \in \vartheta$ son β_0 -semejantes si cumplen que $\beta(O_i, O_j) \geq \beta_0$, en otro caso si $\forall O_j \in \vartheta$ se cumple que $\beta(O_i, O_j) < \beta_0$ entonces llamaremos al objeto O_i como β_0 -aislado [30].

Definición 4: Diremos que $C \subseteq \vartheta$, $C \neq \emptyset$ es una componente β_0 -conexa si se cumple que [28]:

$$a) \quad \forall O_i, O_j \in C, O_i \neq O_j, \exists O_{i_1}, \dots, O_{i_q} \in C \left[O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p \in \{1, \dots, q-1\} \beta(O_p, O_{p+1}) \geq \beta_0 \right]$$

$$b) \quad \forall O_i \in \vartheta \left[O_j \in C \wedge \beta(O_i, O_j) \geq \beta_0 \right] \rightarrow O_i \in C$$

c) Todo elemento β_0 -aislado es una componente conexa (degenerada).

Definición 5: Diremos que $C \subseteq \vartheta$, $C \neq \emptyset$, es un conjunto β_0 -compacto si se cumple que [28]:

$$a) \quad \forall O_j \in \vartheta \left[O_i \in C \wedge \max_{\substack{O_t \in \vartheta \\ O_t \neq O_p}} \{\beta(O_i, O_t)\} = \beta(O_i, O_j) \geq \beta_0 \right] \rightarrow O_j \in C$$

$$b) \quad \left[\max_{\substack{O_t \in \vartheta \\ O_t \neq O_p}} \{\beta(O_p, O_t)\} = \beta(O_p, O_t) \geq \beta_0 \wedge O_t \in C \right] \rightarrow O_p \in C$$

c) $|C|$ es la mínima.

d) Todo elemento β_0 - aislado constituye un conjunto compacto (degenerado).

A partir de las definiciones anteriores se pueden extraer dos criterios agrupacionales:

- **β_0 -conexo**, realiza una búsqueda de aquellos objetos que son semejantes unos con otros y los coloca en un mismo grupo. Todos los objetos cuyas semejanzas sean igual o estén por encima del umbral serán considerados semejantes.
- **β_0 - compacto**, realiza una búsqueda de aquellos objetos que son más semejantes entre ellos y los coloca en un mismo grupo. Para ello escoge de los objetos semejantes que cumplen con la condición de que su valor de semejanza es mayor o igual al umbral, aquellos que tengan mayor valor.

1.3.2 Clasificación general de los algoritmos de agrupamiento

En la literatura consultada [20] [31] [32] existe una gran cantidad de técnicas de agrupamiento que varían de acuerdo a la arquitectura y los datos que pueden manejar. Una clasificación general divide los algoritmos en: agrupamiento particional, agrupamiento jerárquico, agrupamiento basado en densidad y agrupamiento basado en grid.

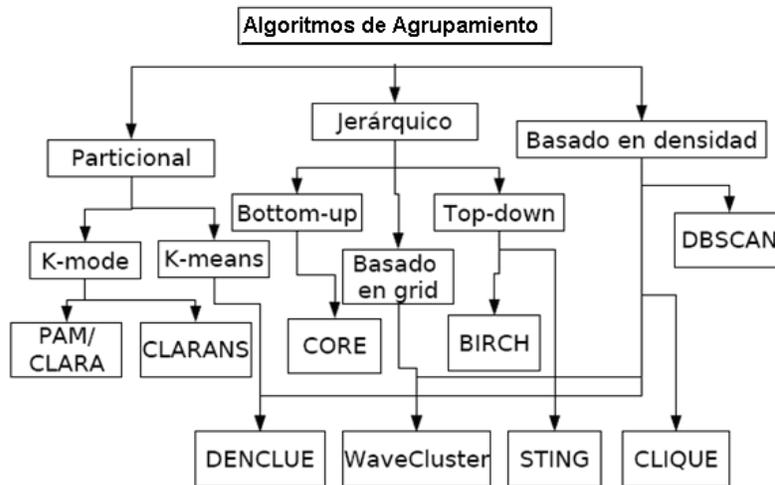


Figura 3 Algoritmos básicos de agrupamiento [32]

Algoritmos jerárquicos

Los algoritmos jerárquicos, como su nombre indica, construyen una jerarquía de agrupamientos, uniendo o dividiendo los grupos de acuerdo a una cierta función de similitud/disimilitud entre los grupos. En otras palabras, construyen un árbol de *clusters* llamado dendograma (Figura 4), en este cada hoja representa un agrupamiento, y el resto representan una distinta agrupación. Tal enfoque permite estudiar los datos con diferentes niveles de granularidad.

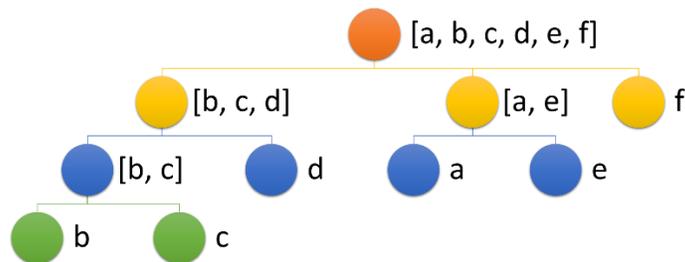


Figura 4 Ejemplo de dendograma (Elaboración propia)

Los métodos de agrupamiento jerárquicos se categorizan en aglomerativos (*bottom-up*) y divisivos (*top-down*). Un agrupamiento aglomerativo, generalmente, comienza con grupos unitarios (*singleton clusters*) y, recursivamente, une dos o más *clusters* apropiados. Un agrupamiento divisivo, generalmente, comienza con un sólo *cluster* en el que están todos los puntos o datos y, recursivamente, divide el *cluster* más apropiado. El proceso continúa hasta que se alcanza algún criterio de parada (frecuentemente el número *k* de *clusters*).

Algunos algoritmos de agrupamiento que pertenecen a esta clasificación son: CURE (*Clustering Using Representatives*), CHAMALEON, BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchical*) y ROCK (*RObust Clustering algorithm using linKs*).

Métodos por partición

Los métodos por partición emplean diferentes técnicas de reubicación para asignar los puntos o datos a uno de los *k clusters*; algunos buscan puntos específicos para luego reubicar los restantes, otros tienen un enfoque probabilístico. A diferencia de los métodos jerárquicos tradicionales, en algunos casos, los *clusters* son revisitados después de construidos y los puntos reubicados para mejorar el agrupamiento.

Estos algoritmos muchas veces asumen un conocimiento a priori del número de *clusters* en que debe ser dividido el conjunto de datos. La idea más usada es hallar los centroides, uno para cada *cluster* y, luego, ubicar los restantes puntos en el grupo del centroide más cercano. La figura muestra un agrupamiento realizado con un algoritmo particional:

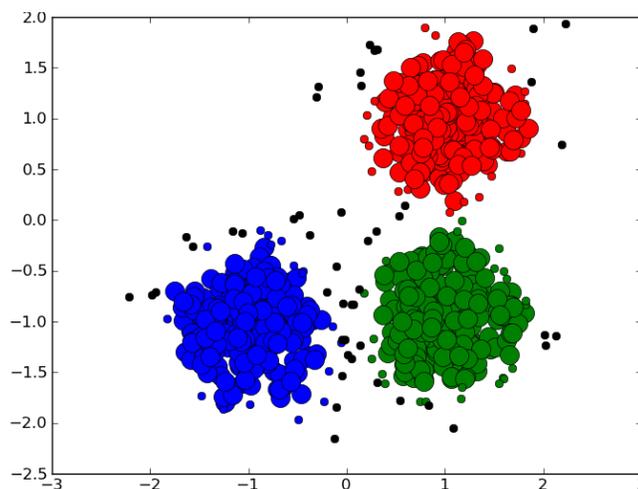


Figura 5 *Clusters* creados con un algoritmo de particionado [33]

Algunos algoritmos de agrupamiento que pertenecen a esta clasificación son: CLARA (*Clustering Large Applications*), CLARANS (*Clustering Large Applications based on Randomized Search*), K-prototypes, K-mode, K-Means.

Agrupamiento basado en densidad

Los algoritmos basados en densidad obtienen grupos basados en regiones densas de objetos en el espacio de datos que están separados por regiones de baja densidad (estos elementos aislados representan ruido).

Este tipo de métodos es muy útil para filtrar ruido y encontrar grupos de diversas formas. La mayoría de los métodos de particionamiento, realizan el proceso de agrupamiento con base en la distancia entre dos objetos.

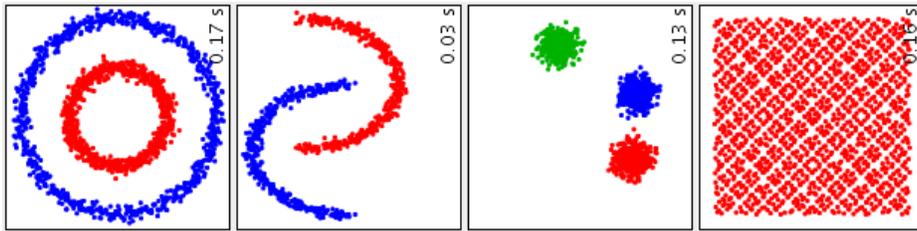


Figura 6 Clusters creados usando un algoritmo de densidad [34]

Algunos algoritmos de agrupamiento que pertenecen a esta clasificación son: DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), OPTICS (*Ordering Points To Identify the Clustering Structure*) y DENCLUE (*DENSity-based CLUstEring*).

Agrupamiento basado en grid

Recientemente un número de algoritmos de agrupamiento han sido presentados para datos espaciales, estos son conocidos como algoritmos basados en grid. Los mismos cuantifican el espacio en un número finito de celdas y aplican operaciones sobre dicho espacio. La mayor ventaja de este método es su veloz procesamiento del tiempo, el cual generalmente es independiente de la cantidad de objetos a procesar.

Algunos algoritmos de agrupamiento que pertenecen a esta clasificación son: STING (*Statistical Information Grid-based method*), CLIQUE y Wavecluster.

La figura muestra el resultado de aplicar sobre una misma base de datos distintos tipos de *cluster*:

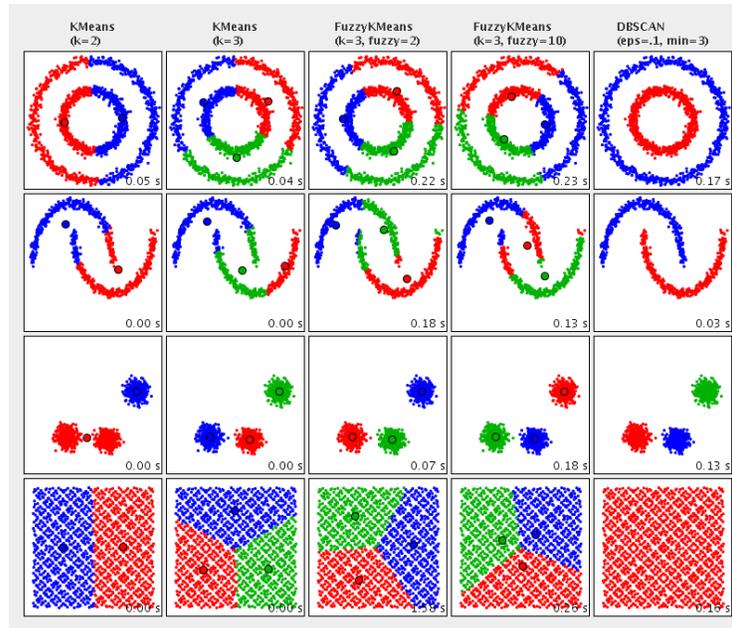


Figura 7 Resultado de cinco algoritmos sobre cuatro bases de datos [34]

1.3.3 Algoritmos de Agrupamiento Conceptual

Los algoritmos de agrupamiento, según el enfoque utilizado para formar los grupos, se clasifican en clasificatorios o conceptuales. Los clasificatorios generalmente construyen los grupos a partir de las relaciones de semejanzas de los objetos y como resultado obtienen solo la estructuración de los objetos. Por otra parte, los algoritmos conceptuales son aquellos que, además de organizar la colección de objetos en grupos, se encargan de encontrar una descripción intencional de los grupos. De forma general, el estudio de los métodos de agrupamiento se ha centrado más en los métodos clasificatorios. Estos métodos tienen como inconveniente el hecho de que dejan al especialista la tarea de interpretar los agrupamientos. Esto es una limitación para muchas aplicaciones en las que, además de conocer los grupos, los usuarios requieren conocer el significado subyacente de los mismos. [27]

Los algoritmos de agrupamiento conceptual se pueden dividir en dos grandes grupos: los algoritmos incrementales y los no incrementales [30]:

- Los algoritmos incrementales basan su funcionamiento en la adaptación de los grupos (o conceptos) con los nuevos objetos que se le van presentando, es decir, cada vez que se analiza un nuevo objeto éste se clasifica, mediante cierta estrategia, en uno de los grupos ya existentes o se crean nuevos grupos. Por lo general, estos algoritmos se han desarrollado para los casos en que el conjunto de objetos a estructurar no está completamente dado, es decir, es dinámico.

Ejemplos de algoritmos de este tipo son: UNIMEM, COBWEB, Galois, EPAM, CLASSIT y LINNEO+.

- Por otro lado, los algoritmos no incrementales estructuran una muestra de objetos utilizando el conjunto de datos completo. Ejemplos de estos algoritmos son: CLUSTER/PAF, CLUSTER/2, y CLUSTER/S, WITT, K-Means conceptual, LC conceptual y RGC.

Disímiles son las aplicaciones que en el mundo hacen uso de las técnicas de RP, aplicando algoritmos para la determinación de semejanzas, en dependencia de las necesidades específicas de cada una. Para el campo de la salud se han desarrollado muchos sistemas informáticos con diferentes objetivos, entre ellos: el diagnóstico de enfermedades, recomendaciones de tratamientos, detección de reacciones adversas entre otros. Luego de una revisión bibliográfica no se encontró ningún sistema que determine la semejanza entre medicamentos; a continuación se describen algunos que gestionan información relacionada con la farmacovigilancia.

1.4 Soluciones existentes a nivel internacional y nacional

1.4.1 Synta

El Sistema para el control Farmacológico (Synta) [17] surge con la necesidad de gestionar un conjunto de información que se maneja en el DF del MINSAP, para ello se apoya el mismo en la Universidad de las Ciencias Informáticas. Este cuenta con cinco componentes relacionados con la información que se genera en los procesos fundamentales de este departamento:

- Reacciones Adversas a Medicamentos
 - Sistema inteligente para la predicción de RAM.
 - *Data Marts* para el análisis de la información.
- Tarjeta Control
- Recetas Médicas
- Consumo de Medicamentos
- Configuración

El sistema cuenta también con un subsistema independiente que ayuda a manejar la información común entre los componentes: el sistema Nomencladores. Permitiendo este subsistema que la información no se duplique entre todos los componentes del sistema.

Reacciones Adversas a Medicamentos (MRAM)

Este componente es el encargado de recopilar todas las RAM ocurridas en Cuba, recolectando las mismas desde la misma Unidad de Salud (US) [35]. Una de las funcionalidades esenciales con que

cuenta el componente es la posibilidad de gestionar gráficas de comportamiento, teniendo en cuenta los datos seleccionados, pudiendo ser estos:

- RAM por años
- RAM por medicamentos
- Comparación de RAM ente distintos años, entre otros.

Este sistema también cuenta con dos subsistemas internos, un Sistema inteligente para la predicción de RAM, que le posibilita al médico, predecir que RAM podría ocasionar si a un tratamiento se le incorpora otro medicamento, teniendo en cuenta un conjunto de datos de un paciente determinado. De la misma forma cuenta con un *Data Marts*, que agiliza la obtención de reportes e información especializada respecto al componente RAM.

1.4.2 Sistema basado en casos para predecir la ocurrencia de reacciones adversas a medicamentos en la consulta médica.

En el año 2013 se crea un Sistema Basado en Casos (SBC) para predecir una posible RAM en la consulta médica, que le permite al médico consultar qué RAM puede ocurrir en el paciente ante un nuevo tratamiento. El SBC se integró al Módulo de Reacciones Adversas a Medicamentos, del producto Synta, desarrollado por el Departamento de Desarrollo de Aplicaciones, del cual obtiene la información necesaria referente a las RAM; utilizando esta para la predicción. [36]

El mismo posee un motor de inferencia que utiliza la función de semejanza descrita anteriormente (Ecuación 2), permitiendo al sistema de conjunto con la implementación de un algoritmo de recuperación la determinación de la semejanza entre el nuevo objeto (medicamento a recetar) y los existentes en la base de casos. Los rasgos que se tiene en cuenta no toman en cuenta las RAM que produce el medicamento.

1.4.3 Sistema basado en reglas para detectar interacción medicamentosa desde la consulta médica

En el año 2014, en el mismo departamento se crea un Sistema Basado en Reglas (SBR) para identificar la interacción entre medicamentos, que le permite al médico conocer qué RAM puede sufrir un paciente por consumir más de un medicamento al mismo tiempo. Este sistema se integró al Módulo de Reacciones Adversas a Medicamentos, del producto Synta, del cual se obtiene la información necesaria referente a los medicamentos y las RAM. [37]

1.4.4 Formulario Nacional de Medicamentos (Android)

Esta aplicación surge como complemento del trabajo realizado por los especialistas del Comité Editorial del Formulario Nacional de Medicamentos de Cuba, cuyo contenido está disponible en formato impreso y digital online, como fuente de información de la Biblioteca Virtual en Salud de Cuba, perteneciente al Centro Nacional de Información de Ciencias Médicas-Infomed. [38]

1.4.5 Formulario Nacional de Medicamentos (Web)

El Formulario Nacional de Medicamentos, versión web, es una obra de referencia que registra los medicamentos disponibles en Cuba, creado por un colectivo de profesionales del Ministerio de Salud Pública, con el apoyo de especialistas de Infomed, editores de la ECIMED y profesionales del Centro de Información Farmacéutica (CINFA). [10]

Entre los objetivos que persigue este formulario se encuentran:

- Facilitar la elección crítica de los medicamentos, al proporcionar su información objetiva, contrastada y con acceso inmediato.
- Mejorar el perfil de prescripción farmacológica a través de una selección racional de los medicamentos.
- Impulsar la formación continuada.
- Contribuir a un uso racional de los medicamentos y a la calidad de la atención médica.

1.4.6 Algoritmo para detectar efectos adversos de los medicamentos en las redes sociales.

Investigadores de la Universidad Carlos III de Madrid bajo el proyecto europeo TrendMiner en colaboración con la empresa Daedalus han desarrollado un nuevo sistema de algoritmos para detectar los efectos adversos de los medicamentos siguiendo en tiempo real la información sobre salud que generan los pacientes en redes sociales como Twitter o blogs especializados.

Este prototipo, según explica uno de estos investigadores del departamento de Informática de la Universidad Carlos III, permite analizar los comentarios en redes sociales por medio de técnicas de procesamiento del lenguaje natural además incorpora un procesador lingüístico basado en MeaningCloud, una tecnología comercial para análisis de grandes volúmenes de datos, que reconoce las menciones de fármacos, efectos adversos y enfermedades.

La mayor desventaja que atenta contra el algoritmo es que gran número de personas en las redes escribe con faltas de ortografía o en abreviaturas para referirse a los medicamentos, además de no conocer el nombre comercial de los mismos, lo que imposibilita que cuando se mencionen sean

detectados por el algoritmo. Especialistas en Farmacología expresan: “muchos de los posibles pacientes confunden síntomas con patología, la terminología que emplean no es clara, hay comentarios duplicados porque se retuitean y los estándares de la industria para evaluar las reacciones adversas son más estrictos, han de estar documentados y poder confirmarlos con un profesional sanitario”. [39]

1.4.7 Solución informática para registrar alergias y reacciones adversas a medicamentos.

El Instituto Catalán de la Salud (ICS) desarrolló una solución informática para registrar alergias y reacciones adversas a medicamentos. El nuevo sistema posibilita que los profesionales incluyan en la historia clínica electrónica cualquier antecedente indeseado en el consumo de fármacos.

La información recopilada se registra en un formulario predeterminado que, si llega a detectar alguna posible reacción adversa, envía una notificación al Centro de Farmacovigilancia de Cataluña de manera automática. La solución permite que los médicos accedan a bases de datos reconocidas y a bibliografía actualizada sobre medicamentos. Según informaron los desarrolladores, la herramienta de Prescripción Farmacológica Segura se activa cuando se dé de alta un medicamento en la Estación Clínica. Además, comprueba automáticamente la compatibilidad del fármaco con los antecedentes de los pacientes, datos disponibles en la historia clínica electrónica. [40]

1.5 Herramientas y tecnologías a utilizar en la propuesta de solución

Dado que la solución a la problemática planteada está concebida para ser integrada al Módulo de RAM perteneciente a Synta, las herramientas y tecnologías a utilizar en el transcurso del proceso de desarrollo son las mismas que se utilizaron en desarrollo de dicho módulo. Además se describe el lenguaje de programación Java ya que fue utilizado para la implementación de algunas características necesarias para el desarrollo de la solución.

1.5.1 Sistemas Gestores de Bases de Datos (SGBD)

El Sistema de Gestión de Bases de Datos (SGBD) [41], es el responsable de tratar todas las peticiones de información de dichos usuarios. Este sistema es un conjunto de programas de propósito general, que permite a los usuarios controlar el acceso y la utilización de la base de datos, para incluir, modificar o recuperar información, incluyendo prestaciones con el fin de conseguir la independencia, integridad y seguridad de los datos, y la concurrencia de los usuarios.

MySql 5.5

El software MySql proporciona un servidor de base de datos SQL (*Structured Query Language*) muy rápido, multi-hilos, multiusuario y robusto. El servidor MySql está diseñado para entornos de producción críticos, con alta carga de trabajo. Entre sus principales características se encuentran que es un servidor de base de datos muy rápido, fiable y fácil de usar. [42] La selección del mismo se fundamenta en que éste es el utilizado por el proyecto Synta.

EMS MySql Manager 3.7

EMS SQL Manager para MySql es una herramienta de alto rendimiento para la administración y manejo de *MySql Database Server*. Funciona con cualquier versión de MySql desde 4.1 a la más reciente y soporta todas las últimas características de MySql; incluyendo disparadores, vistas, procedimientos almacenados y funciones, claves foráneas. Permite crear y editar todos los objetos de base de datos MySql, diseñar bases de datos MySql de manera que sean más sencillas y fácil de tratar, y ejecutar scripts SQL. [43]

1.5.2 Servidores Web

Servidores Web (Web Servers): Básicamente, un servidor web sirve contenido estático a un navegador, carga un archivo y lo sirve a través de la red. Los servidores web son aquellos cuya tarea es alojar sitios y aplicaciones, las cuales son accedidas por los clientes utilizando un navegador que se comunica con el servidor utilizando el protocolo HTTP (*Hypertext Transfer Protocol*). [44]

Apache 2.2

Representa el complemento perfecto para el desarrollo de páginas dinámicas con PHP. Caracterizado principalmente por ser gratuito, de alta popularidad, sencillez de manejo y versatilidad. Puede ser instalado en plataformas Linux y Windows. [45]

Algunas de las características que presenta son:

- **Fiabilidad:** un gran número de los servidores con más alta disponibilidad funcionan bajo un servidor Apache.
- **Software libre:** el servidor Apache es totalmente gratuito y es distribuido bajo la licencia de Apache, en la cual se permite realizar cambios al código fuente.
- **Extensibilidad:** se pueden añadir módulos para ampliar las capacidades del servidor.

1.5.3 Lenguaje de Programación

PHP

HyperText Pre-Processor (PHP, por sus siglas en inglés), es un lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas. [46]

Características:

- Es un lenguaje multiplataforma
- Completamente orientado al desarrollo de aplicaciones web dinámicas con acceso a información almacenada en una base de datos
- El código fuente escrito en PHP es invisible al navegador y al cliente. El servidor es el que se encarga de ejecutar el código y enviar su resultado HTML al navegador. Esto hace que la programación en PHP sea segura y confiable.
- Capacidad de conexión con la mayoría de los motores de base de datos que se utilizan en la actualidad. Destaca su conectividad con MySQL y PostgreSQL.
- Es libre, por lo que se presenta como una alternativa de fácil acceso para todos
- Permite aplicar técnicas de programación orientada a objetos

Java V8

Lenguaje de programación orientado a objetos (POO), de alto nivel y semicompilado que funciona con una máquina virtual. Puede ser ejecutado en la mayoría de los sistemas operativos, software de distribución libre, rápido, multitarea, seguro y fiable. [47]

Características:

- Independencia de la plataforma: Java funciona con las principales plataformas de hardware y sistemas operativos, o bien con el software JVM directamente desde Oracle, o como parte de la comunidad OpenJDK.
- Alto rendimiento: Las optimizaciones integradas para entornos multiproceso lo hacen aún más rápido
- Fácil de aprender: El modelo de Java para la gestión de la memoria, los procesos múltiples y la gestión de excepciones lo convierte en un lenguaje eficaz para los desarrolladores nuevos y para los más experimentados.
- Basado en estándares: El lenguaje Java y la tecnología relacionada evolucionan a través de *Java Community Process*, un mecanismo que permite desarrollar especificaciones técnicas para la tecnología Java.

- Entornos de ejecución coherentes: Permite realizar despliegues con confianza con entornos de tiempo de ejecución que van de Java SE en equipos de sobremesa a *Java SE for Embedded Devices* y *Oracle Java Micro Edition Embedded Client*.
- Modelo con seguridad probada: ofrece un entorno de aplicaciones avanzado con un alto nivel de seguridad que es idóneo para las aplicaciones de red.

JavaScript 1.1

JavaScript es un lenguaje script u orientado a documento, como pueden ser los lenguajes de macros, que tienen muchos procesadores de texto y planillas de cálculo. JavaScript es una de las múltiples maneras que han surgido para extender las capacidades del lenguaje HTML, al ser la más sencilla, es por el momento la más extendida. No se puede desarrollar un programa con JavaScript que se ejecute fuera de un navegador, es un lenguaje interpretado que se embebe en una página web HTML. [48]

HTML

HyperText Mark-Up Language (HTML), se conoce como "lenguaje de marcado", cuya función es preparar documentos escritos aplicando etiquetas de formato. HTML se usa también para la lectura de documentos en Internet desde diferentes equipos, gracias al protocolo HTTP, que permite a los usuarios acceder de forma remota a documentos almacenados en una dirección específica de la red. [49]

CSS

Cascade Style Sheet (CSS), traducido significa: Hojas de Estilo en Cascada. Permiten controlar la apariencia de una página web; describiendo cómo los elementos dispuestos en la página son presentados al usuario. CSS es un gran avance que complementa el HTML y es un estándar aceptado por toda la industria relacionada con la Web, o por lo menos, gran parte de los navegadores. [50]

1.5.4 Marco de desarrollo

Los marcos de desarrollo (*frameworks*) se pueden considerar como soluciones completas que contemplan herramientas de apoyo a la construcción (ambiente de trabajo o desarrollo) y motores de ejecución (ambiente de ejecución). Estos aceleran el proceso de desarrollo, permiten reutilizar código ya existente y promover buenas prácticas de desarrollo, como el uso de patrones. [51]

Symfony 1.4

Es un *frameworks* de desarrollo web, que simplifica la creación de una aplicación, mediante la automatización de algunos de los patrones utilizados para resolver tareas comunes. Proporciona estructura al código fuente, forzando al desarrollador a crear códigos más legibles y fáciles de mantener. Posibilita la programación de sistemas y encapsula operaciones complejas en instrucciones sencillas. Symfony separa la lógica de negocio, la lógica de servidor y la presentación del sistema. Proporciona varias herramientas y clases, encaminadas a reducir el tiempo de desarrollo de una aplicación web compleja. Además, automatiza las tareas más comunes, permitiendo al desarrollador dedicarse por completo a los aspectos específicos de cada aplicación. [51]

Características de Symfony:

- Fácil de instalar y configurar en la mayoría de las plataformas
- Independiente del sistema gestor de base de datos
- Código fácil de leer, permitiendo un mantenimiento sencillo
- Permite la integración con librerías desarrolladas por terceros

ExtJS 3.1

ExtJS es una librería JavaScript que mantiene el código bien estructurado. Incluso las aplicaciones más grandes pueden ser de fácil mantenimiento. Ofrece una recopilación de elementos de interfaz de usuario con temas elegantes de partida. Una potencia de la base de ExtJS, es su diseño de componentes, que permite ampliar fácilmente los componentes predeterminados para satisfacer las necesidades del cliente. Utiliza tecnologías de navegación específica, para proporcionar una experiencia consistente y de alto rendimiento al usuario. Hace un amplio uso de imágenes para la tematización, y de tecnologías específicas para el manejo de datos. [52]

1.5.5 Entorno de Desarrollo Integrado

NetBeans V 8.0.1

Entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. El NetBeans IDE es un IDE de código abierto escrito completamente en Java usando la plataforma NetBeans. El NetBeans IDE soporta el desarrollo de todos los tipos de aplicación Java. [53]

PHPStorm

PHPStorm cuenta con un editor de código inteligente para PHP que realmente "entiende" su código y entiende profundamente su estructura, soporte PHP 5.3, 5.4, 5.5 y 5.6 para los proyectos modernos y antiguos. El IDE proporciona finalización de código inteligente, resaltado de sintaxis, la configuración de formato de código extendido, sobre la marcha de la comprobación de errores, plegado de código, apoya las mezclas de idiomas etc. [54]

CAPÍTULO 2 PROPUESTA DE SOLUCIÓN PARA EL CÁLCULO DE LA SEMEJANZA ENTRE MEDICAMENTOS SEGÚN RAM

En este capítulo se abordan los aspectos relacionados con el dominio de la investigación y el sistema. Se desarrolla la propuesta de solución centrándose en las características de la función de semejanza empleada, así como los elementos y conceptos asociados al cálculo del peso informacional de los rasgos apoyados de la *Teoría de Testores*. Se hace alusión sobre los principales aspectos a tener en cuenta para la correcta implementación del sistema.

2.1 Propuesta de solución.

A pesar de que en el Synta se mantienen nombrados los datos de los medicamentos, las reacciones adversas de los medicamentos y las frecuencias de aparición de las mismas; este no posee la relación existente entre ellos. Por tanto es necesario adicionar al MRAM esta relación para realizar comparaciones atendiendo a sus semejanzas según las reacciones adversas. Por lo que se propone implementar funcionalidades para gestionar la relación *Medicamento-RAM-Frecuencia* y para el cálculo de la semejanza entre medicamentos atendiendo a sus reacciones adversas. (Figura 8)

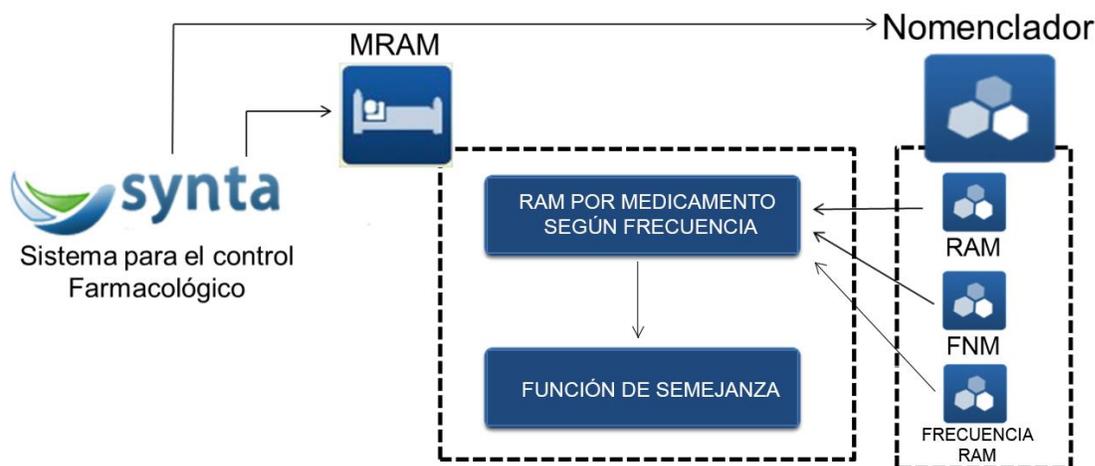


Figura 8 Integración de la Funcionalidad para la determinación de Semejanzas en el MRAM del producto Synta (Elaboración propia)

El flujo de trabajo en el sistema después de desarrollada la funcionalidad será el siguiente:

- El paciente acude a la consulta médica al notar en él la aparición de una reacción adversa al medicamento, o a los medicamentos prescritos.

- Para que el médico pueda identificar cual medicamento debe ser el idóneo para sustituir al antes prescrito se debe introducir en el sistema el medicamento actual y las propuestas del médico.
- El sistema procesa los datos introducidos y le comunica al médico la relación de semejanza entre el medicamento actual y todas las propuestas. El médico debe determinar atendiendo a las características del paciente la factibilidad de las propuestas realizadas, para realizar la prescripción.

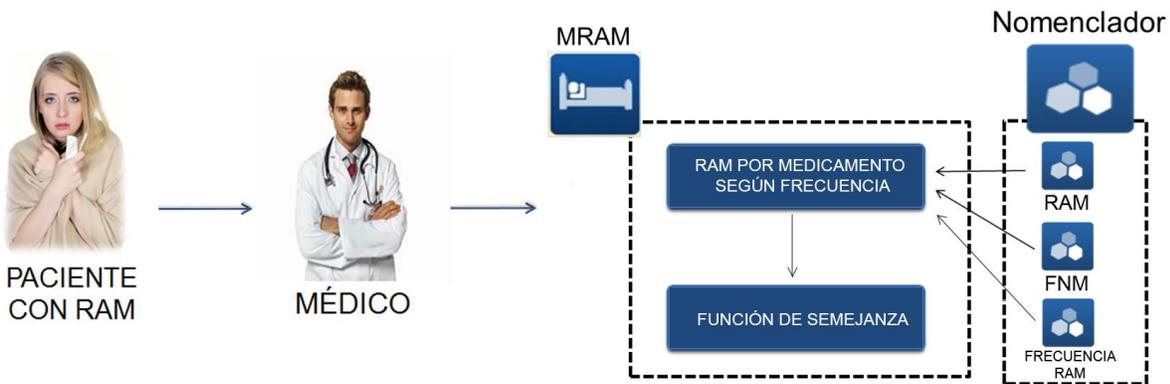


Figura 9 Propuesta de flujo a seguir para identificar la semejanza entre medicamentos, desde el MRAM (Elaboración propia)

La implementación de funcionalidades para la determinación de semejanzas entre medicamentos tiene implícito la aplicación de las técnicas para el RP. Es el proceso para la determinación de semejanzas entre medicamentos haciendo uso de del enfoque lógico-combinatorio del RP precisamente donde se centran las ideas fundamentales de este trabajo. Para una mejor comprensión se ha decidido dividir la propuesta de solución en pasos:

- **Paso 1:** Agrupar medicamentos según la RAM que causan, para obtener los medicamentos del FNM particionados atendiendo a sus RAM.
- **Paso 2:** Calcular el peso informacional asociado a cada rasgo (RAM).
- **Paso 3:** Propuesta de función de semejanza a aplicar a los medicamentos.

2.2 Definición de objetos, rasgos y dominios del problema

Se definen para el problema de determinación de semejanzas entre medicamentos atendiendo a sus reacciones adversas como objeto de trabajo: **los medicamentos**, teniendo como rasgos descriptivos: **las reacciones adversas**, enmarcadas en el dominio: **frecuencias con que pueden aparecer las mismas**.

2.3 Agrupar medicamentos según RAM (Paso 1)

Para lograr un mejor resultado en la relación de semejanza entre medicamentos según sus RAM se hace necesario tener agrupados a los mismos, atendiendo a esta característica. Sin un conocimiento a priori de cuantos grupos pudiesen formarse al aplicar cualquier técnica de agrupamiento, y sin conocer un valor previo del peso informacional de cada una de las reacciones adversas se recurre a utilizar un algoritmo de clustering, o agrupamiento de datos que sirva para pre-procesar la información referente a los medicamentos.

Para ello se propone el uso de un algoritmo de agrupamiento no supervisado:

Algoritmo 1. Agrupamiento no supervisado [18]

Entrada: Matriz inicial⁹.

Salida: Matriz de Entrenamiento¹⁰.

P1: Construir la Matriz de Semejanza¹¹ utilizando una función de semejanza.

$$\beta(O_i, O_j) = \sum_{k=1}^n \delta_k(O_i, O_j) \quad [22]$$

P2: Calculo del umbral de semejanza

$$\beta_0 = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \beta(O_i, O_j) \quad [22]$$

donde m: número de objetos, i y j recorren las filas y columnas respectivamente.

P3: Agrupar siguiendo un criterio agrupacional.

2.3.1 Algoritmo de *clustering*

Se decide utilizar para agrupar a los medicamentos como algoritmo de agrupamiento a los que atendiendo al enfoque utilizado para formar los grupos se clasifican en conceptuales, pues no solo son capaces de crear la estructuración de los objetos, sino que brindan una caracterización comprensible para los expertos sobre la estructuración antes creada, información que resultará necesaria para una mejor comprensión de los grupos formados. Los algoritmos antes seleccionados,

⁹ Matriz con los datos de los medicamentos sin formar agrupaciones

¹⁰ Matriz con los medicamentos y su grupo asociado

¹¹ Matriz con la relación de semejanza existente entre todos los medicamentos

como con anterioridad se había expuesto, se subdividen en dos categorías incrementales y no incrementales.

Algoritmos incrementales según la bibliografía consultada [18] [55] [27] tienen como principal desventaja la dependencia del resultado en función del orden de procesamiento de los objetos. Además de las limitaciones propias que presenta cada algoritmo, entre los que se agrupan en este criterio se encuentran el EPAM y COBWEB que tienen como desventaja la que solo manejan variables cualitativas, en otro caso está el CLASSIT que únicamente maneja atributos cuantitativos. Otras variantes como COBWEB/3 maneja información mezclada pero está basado en probabilidades, medias y varianzas, lo que resulta complicado de interpretar sobre todo para usuarios que no estén familiarizados con conceptos de probabilidad y estadísticas. UNIMEM no puede procesar objetos descritos por rasgos incompletos y comparte la limitante de LINNEO+ de que la estructura obtenida depende del orden de análisis de los objetos y por lo tanto, pueden obtenerse diferentes jerarquías a partir de una misma colección, dependiendo del orden en que se analicen sus objetos.

Algoritmos no incrementales según la bibliografía consultada [18] [55] [27] aun cuando no presentan la problemática anterior al realizar un análisis de cada uno de ellos por separado muestra algunas características que se deben considerar:

- CLUSTER/PAF, CLUSTER/2, CLUSTER/S y k-MEANS conceptual tienen como característica principal el hecho de requerir semillas para su funcionamiento, el número de semillas iniciales determina el número de agrupamientos. Este dato en la mayoría de los problemas prácticos es una incógnita, por lo que estos algoritmos en problemas donde no se conozca el número de agrupamientos a formar no resuelven el problema de manera satisfactoria. Este tipo de algoritmo presentan como dificultad adicional que pueden caer en máximos locales.
- WITT no es capaz de procesar objetos descritos por atributos numéricos, mezclados, ni incompletos.
- LC-Conceptual está basado en el Enfoque Lógico Combinatorio del Reconocimiento de Patrones es capaz de procesar objetos descritos por atributos mezclados e incompletos. Usar el criterio de agrupamiento basado en componentes β -conexas puede formar grupos con baja semejanza intra-grupo.
- RGC está basado en el Enfoque Lógico Combinatorio del Reconocimiento de Patrones es capaz de procesar objetos descritos por atributos mezclados e incompletos. La principal limitación del algoritmo es la complejidad computacional, la cual puede ser exponencial. Además, los

conceptos que se generan pueden ser muy grandes y por lo tanto difíciles de interpretar en problemas reales.

Por lo anteriormente expuesto se decide emplear como algoritmo para la determinación de los cluster al algoritmo LC-Conceptual (**Anexo 3**) usando como criterio de agrupamiento el β_0 -compacto.

Algoritmo LC Conceptual

El algoritmo LC-Conceptual es capaz de trabajar tanto con atributos cualitativos como cuantitativos, además de admitir la ausencia de información. No requiere especificar el número de agrupamientos a priori ni semillas, por lo que representa una ventaja en la investigación en curso pues es deseable organizar a los medicamentos en agrupaciones considerando las semejanzas entre los mismos según sus RAM sin presuponer una determinada cantidad de grupos.

Debido a sus características, este algoritmo permite resolver una gama amplia de problemas de estructuración conceptual que otros algoritmos no pueden resolver satisfactoriamente y aunque la determinación de la descripción extensional de los grupos es computacionalmente poco compleja, la etapa de caracterización que ejecuta para formar los conceptos es computacionalmente costosa, principalmente debido al cálculo de los testores, a pesar de tener esta limitante es muy eficiente y aunque el criterio de agrupamiento basado en componentes β_0 -conexas que propone LC puede formar grupos con baja semejanza intra-grupo, esto puede mitigarse proponiendo una variante mucho mejor como es el β_0 -compacto.

2.4 Cálculo del peso informacional de los rasgos (Paso 2)

El concepto de testor aparece por primera vez en un trabajo de los científicos rusos Cheguis y Yablonskii [24] en un intento por tratar de simplificar la tarea de encontrar desperfectos en circuitos eléctricos que podían ser modelados a través de funciones del álgebra de la lógica. La determinación de todos los testores típicos de una matriz de entrenamiento encuentra gran aplicación en el RP, porque ellos permiten determinar el peso informacional de los rasgos.

El enfoque basado en la *Teoría de Testores*, para definir una medida de la importancia de los rasgos, se basa en la idea de que un testor típico, según Zhuravliov, es un conjunto de columnas (rasgos) de la matriz de entrenamiento para las cuales se cumple que no existen filas iguales en clases diferentes, pero además si se elimina aleatoriamente una de las columnas restantes perderán esta propiedad.

Varios algoritmos han sido desarrollados para calcular el conjunto de todos los testores típicos. Ellos pueden clasificarse de acuerdo con su estrategia computacional en dos categorías: algoritmos de

escala exterior y de escala interior. Los primeros calculan los testores típicos generando los elementos del conjunto potencia en un orden predefinido, pero tratando de evitar el análisis de subconjuntos irrelevantes. Los segundos exploran la estructura interna de la Matriz de Entrenamiento (ME) encontrando condiciones que garantizan la propiedad de testor. [56]

Algoritmo 2. Calculo del peso informacional de los rasgos [18]

Entrada: Matriz de Entrenamiento.

Salida: Peso Informacional de los Rasgos.

P1: Formar la Matriz de Diferencia.

P2: Formar la Matriz de Básica.

P3: Seleccionar el conjunto de testores.

P4: Calcular el peso de los rasgos según la teoría de testores

$$P(x_i) = \frac{\tau(i)}{\tau}, L(x) = \frac{\sum_{t \in \psi^*(x)} \frac{1}{|t|}}{|\psi^*(x)|}, D(x) = \alpha P(x) + \beta L(x) \text{ con } \alpha, \beta = 0.5$$

Para la selección del conjunto de todos los TT se analizaron los algoritmos que se mencionan a continuación, concluyendo como algoritmo a utilizar el Fast-BR (**Anexo 4**).

Algoritmo BT

Algoritmo de escala exterior donde el orden que lo caracteriza es el generado por los números naturales de forma ascendente en su notación binaria mediante un n-uplo booleano. Éste constituye un orden total sobre los elementos no nulos del conjunto potencia del conjunto de rasgos de la ME [57]. Este algoritmo presenta dos inconvenientes:

- genera muchos n-uplos que son supra conjuntos de TT. Los saltos que tiene definido sólo pueden evitar los supra conjuntos consecutivos al que acaba de detectar como testor y muchos de los supra conjuntos restantes serán examinados innecesariamente.
- para comprobar si un nuevo n-uplo es o no testor siempre es preciso analizar las filas de la Matriz Básica (MB) desde el principio, como si nunca antes se hubiese hecho.

Algoritmo CT

Algoritmo de escala interior y trabaja basado en los conceptos de conjunto de filas independientes¹² y conjunto completo¹³. El método del CT consiste en explorar las posiciones unitarias de la MB encontrando los conjuntos completos y comprobando, entonces, si son TT.

Éste como los demás algoritmos de escala interior, presenta la deficiencia de que repiten los TT encontrados. Este problema es inherente a su estrategia. Para que un conjunto τ sea un TT debe existir para cada rasgo de τ al menos una fila que tenga un 1 en esa columna y cero en todas las restantes de τ , pero esta fila no tiene porqué ser única; de hecho, es más probable que no lo sea. Esto se traduce para el CT en que un mismo TT se origina a partir de más de un par conjunto completo-conjunto de filas independientes. [58]

Algoritmo CC

Algoritmo de escala interior y está basado en los conceptos de Conjunto Compatible¹⁴ y Conjunto Regular¹⁵. El CC tiene como peculiaridad, a diferencia del CT, que su objetivo es garantizar la tipicidad. Esto está implícito en la definición de conjunto compatible, luego tiene que comprobar si se cumple la propiedad de testor.

El CC presenta la misma limitación señalada anteriormente al algoritmo CT, pues en este caso, un mismo TT se origina a partir de más de un conjunto compatible. Además el algoritmo tiene otra deficiencia: para obtener un TT es necesario explorar un gran número de combinaciones de las posiciones unitarias de la MB, es decir, de conjuntos compatibles. [59]

Algoritmo REC

Algoritmo de escala exterior que tiene la particularidad de que trabaja directamente con la ME, lo cual lo sitúa en desventaja con relación a los restantes algoritmos por el enorme manejo de información que tiene que realizar. [60]

¹² Cuando ninguno de los vectores se puede escribir como combinación lineal de los restantes.

¹³ Todo conjunto completo es un testor.

¹⁴ Dos elementos unitarios de MB a_{i_1, j_1} , a_{i_2, j_2} son compatibles si $a_{i_1, j_1} = a_{i_2, j_2} = 0$. El conjunto $D = \{ a_{i_1, j_1}, a_{i_2, j_2}, \dots, a_{i_s, j_s} \}$ de valores unitarios de MB es compatible si $s = 1$ ó sus elementos son compatibles dos a dos.

¹⁵ Un conjunto regular es un conjunto compatible, en el cual las columnas de MB asociadas a él constituyen un testor.

Algoritmo CER

Algoritmo de escala exterior que fue encaminado a resolver el problema presentado por el REC y lo hizo introduciendo un nuevo orden en el conjunto potencia de los rasgos. [61]

Algoritmo LEX

Clasificado como de escala exterior la idea del algoritmo es ir construyendo listas de rasgos que posean la propiedad de tipicidad y luego, comprobar si este conjunto de rasgos constituye un testor. Cuando se obtiene un testor típico se producen saltos.

Con el objetivo de lograr un mejor desempeño del algoritmo se realiza previamente un ordenamiento de MB que consiste en colocar como primera fila aquella que tenga la menor cantidad de unos y, en esta fila encontrada, situar todos los unos a la izquierda. Si existiera más de una fila con mínima cantidad de unos se toma cualquiera de ellas. De esta forma, el algoritmo puede terminar cuando, durante el proceso de retroceso, la lista actual esté vacía y el rasgo que le corresponda analizarse posea un cero en la primera fila, ya que ninguno de los posibles subconjuntos de rasgos que se pueden formar con los rasgos que restan pueden formar un testor típico por no cumplir la propiedad de ser testor. [24]

Algoritmo CT-EXT

Se obtiene el (CT-EXT) nombre del algoritmo combinando CT (debido a las similitudes en el comportamiento que presenta con respecto al algoritmo CT) y EXT (porque lleva a cabo su búsqueda sobre el conjunto potencia de características).

CT-EXT es un algoritmo de tipo externo que utiliza el mismo orden lexicográfico que el algoritmo LEX. El algoritmo CT-EXT genera combinaciones de características incrementales reduciendo, paso a paso, el número de objetos que pertenecen a diferentes clases que son confusas, hasta que una combinación sea un testor. Posteriormente, CT-EXT verifica si la combinación generada es un testor típico. [62]

Algoritmo BR

Algoritmo de escala exterior que evita el análisis de un mayor número de subconjuntos irrelevantes y que eficientemente verifica la propiedad de testor tomando como ventaja las operaciones a nivel de bits de la computadora. La idea general del algoritmo BR es primeramente generar listas de rasgos que satisfagan la propiedad de tipicidad y luego, verificar la condición de testor. [61]

El algoritmo BR para calcular los TT en los resultados experimentales demuestra que este supera significativamente al resto de los mencionados hasta este momento en la investigación.

Las principales contribuciones que aseguran la aceleración en el cálculo de todos los testores típicos son:

- Un nuevo método para verificar las propiedades de tipicidad y testor, las cuales se basan en la lógica binaria y aprovecha las operaciones a nivel de bits de la computadora.
- La introducción de un mecanismo para generar subconjuntos de rasgos candidatos que evita el análisis de un mayor número de subconjuntos irrelevantes.
- El orden previo a la matriz básica que garantiza que el método finalice lo antes posible.

Algoritmo Fast-BR

El algoritmo Fast-BR según la bibliografía consultada [63] resulta de una implementación del BR con reducción de columnas redundantes en la MB, y el empleo de t-uplas binarias para la representación de las filas y testores. La reducción de columnas redundantes consiste en agrupar en la misma MB inicial a todas las columnas iguales, creando una nueva MB reducida en la cual todos los TT obtenidos serán denominados pseudo-TT, al ser los mismos obtenidos de una MB reducida.

Según resultados experimentales [63] de la comparación de los tiempos de ejecución de los algoritmos con mejores resultados (Tabla 1) se muestran las dos variantes BR y Fast-BR.

Tabla 1 Tiempo de ejecución del Fast-BR y otros algoritmos en bases de datos reales

Dimensiones MB	CT_EXT	LEX	BR	Fast-BR	Columnas Redundantes
14x17	26ms	13ms	11ms	9ms	12%
120x35	58s 927ms	50s 222ms	2s 338ms	2s 341ms	0%
30x22	48ms	40ms	38ms	18ms	5%
100x57	5s 107ms	780ms	167ms	189ms	0%
250x57	59s 284ms	6s 564ms	335ms	352ms	0%
500x57	5m 13s	29s 129ms	853ms	888ms	0%
1000x57	26m 26s	2m 15s	2s 887ms	2s 931ms	0%

Se hace necesario señalar porqué aun cuando en la mayor parte de las pruebas se muestra como más veloz al BR, en la presente investigación se decide optar por el Fast-BR. En los casos donde no existen columnas redundantes el tiempo de demora del BR respecto al Fast-BR media sobre 95,59%, sin embargo cuando existe redundancia el tiempo del Fast-BR con respecto al BR media el 64,58%. Como la composición de la MB generada por el análisis de la ME generada **Paso 1**, es impredecible se decide optar por el Fast-BR; a modo de que en una posible ocurrencia de

redundancia la ganancia en tiempo sea superior al caso contrario con redundancia 0 en el cual la pérdida sería mínima, dado que la diferencia mostrada de un 5% representa la revisión de las posibles columnas redundantes.

Por el gran consumo de recursos que representa la realización computacional de los **Pasos 1 y 2**, se decide la utilización del lenguaje de programación Java, para el trabajo con matrices de alta dimensionalidad.

2.5 Propuesta de función de semejanza a aplicar a los medicamentos (Paso 3)

Determinados los grupos de los medicamentos según sus RAM, y calculados los pesos informacionales de cada uno de los rasgos presentes en los medicamentos, se propone como función de semejanza a ser utilizada en la propuesta de solución: [20]

$$\beta(O_i, O_j) = \frac{\sum_{k=1}^n P_k * \delta(O_i, O_j)}{\sum_{k=1}^n P_k}$$

Ecuación 8 Distancia basada en la suma pesada de la comparación por rasgos

Donde $\delta(O_i, O_j)$ representa la función de similitud por rasgos y P_k representa el peso del rasgo k. Entre las principales ventajas que pudiera acarrear el uso de esta función es que: bajo el criterio de un experto para agrupar los medicamentos según las reacciones adversas que estos ocasionan el peso informacional de los rasgos le darían más importancia a las RAM más significativas para los médicos. Como en la propuesta actual la clasificación es realizada por un algoritmo no supervisado, esta función contribuye a la mejor diferenciación de los medicamentos según sus RAM.

2.5.1 Función de similitud por rasgos

La función de similitud por rasgo no puede ser realizada de forma estática, pues los valores correspondientes al dominio de los rasgos es variable por ello se propone.

$$\delta(O_i, O_j) = \begin{cases} 1 & \text{ssi } O_i = O_j \\ [0,1] & \text{definido por el usuario} \\ 0 & \text{eoc} \end{cases}$$

Ecuación 9 Función de similitud por rasgos

Donde el sistema colocará 1 en el caso de que ambos rasgos tengan igual valor, asignará valores entre 0 y 1 según los definidos por el usuario y colocará 0 en otro caso que no se encuentre definido en las reglas anteriores.

2.6 Requisitos no funcionales (RNF)

Los requisitos no funcionales que se describen a continuación, son las condiciones o capacidades recomendadas para que el producto Synta funcione correctamente. Al proponerse la integración de la propuesta de solución a un módulo del producto Synta, este sigue los RNF definidos por los desarrolladores de dicho producto.

Usabilidad

RNF1 El sistema solo podrá ser utilizado por los usuarios definidos en el Módulo de Administración, que tengan acceso al Módulo de Reacciones Adversas a Medicamentos.

RNF2 El usuario definido como administrador, debe tener conocimientos básicos de informática.

RNF3 El sistema debe presentar un acceso fácil y rápido, para facilitar el uso del mismo por usuarios con pocos conocimientos en el campo de la informática.

Seguridad

RNF4 Se harán validaciones de la información, tanto en el cliente como en el servidor, contra ataques de inyección HTML o SQL.

Confidencialidad

RNF5 La información solo podrá ser usada por personal autorizado (usuarios definidos en el Módulo de Administración de Synta).

Integridad

RNF6 La información manejada por el sistema será objeto de protección contra estados inconsistentes.

Eficiencia

RNF7 El sistema deberá ser rápido ante las solicitudes de los usuarios, el tiempo de respuesta deberá ser menor que 5 segundos.

Soporte

RNF8 El personal encargado de dar soporte al sistema debe contar con el nivel técnico requerido.

Software del servidor

RNF9 Se utilizará MySql versión 5.5.

RNF10 Se utilizará tecnología Apache versión 2.2 o superior, para el servidor web.

RNF11 Se utilizará para el desarrollo PHP 5.3, el *framework* Symfony 1.4 y el lenguaje de programación Java 8.

RNF12 Se utilizará un servidor con el sistema operativo Windows 7 o superior, o con un sistema operativo GNU/Linux. Ubuntu v12.04 preferentemente.

Software del cliente

RNF13 En las computadoras de los clientes se requiere de un navegador web (Internet Explorer versión 8.0 o superior, Mozilla Firefox versión 3.0 o superior).

Hardware para el cliente

RNF14 Se recomienda un ordenador Pentium IV, o superior, con al menos 1GB de memoria RAM, 1GB de espacio libre en disco duro.

RNF15 Se requiere tarjeta de red.

Hardware para el servidor

RNF16 Se recomienda un ordenador Core i3 que tenga al menos 2GB de memoria RAM y 20GB de disco duro libre.

Interfaz

RNF17 El sistema debe tener una interfaz fácil de usar y amigable para que pueda ser utilizada sin mucho entrenamiento por el usuario.

RNF18 Las interfaces de usuario deben estar definidas y creadas siguiendo las pautas establecidas por los desarrolladores del producto Synta. (Ejemplo: íconos, mensajes del sistema, etiquetado).

Estándares aplicables

RNF19 Para la implementación del sistema se deben seguir los estándares de codificación y diseño, definidos por los desarrolladores del producto Synta.

CAPÍTULO 3 VALIDACIÓN DE LA SOLUCIÓN PROPUESTA

El presente capítulo tiene como objetivo, mostrar los resultados de la validación de la propuesta de función para la determinación de semejanza entre medicamentos según sus RAM. Para lograrlo se demuestra apoyados en las técnicas de metodología de investigación científica la validez de la muestra de datos seleccionada para realizar el agrupamiento. Posteriormente, se muestran los resultados de una encuesta aplicada a los profesionales de la salud, sobre casos reales para comparar los resultados de los médicos por los brindados por el sistema.

3.1 Tamaño de la muestra de medicamentos seleccionados

Según Sampieri, en su libro Metodología de la Investigación [64], una muestra representa un subgrupo de la población que se investiga. Teniendo en cuenta que la población escogida son los medicamentos presentes en el FNM, y utilizando las fórmulas propuestas por el autor del libro para el cálculo de una muestra no probabilística:

$$n = \frac{k^2 * p * q * N}{(e^2 * (N - 1)) + k^2 * p * q}$$

Ecuación 10 Tamaño de la muestra

Donde:

N: es el tamaño de la población.

k: es una constante que depende del nivel de confianza.

e: es el error deseado.

p: es la proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que $p=q=0.5$ que es la opción más segura.

q: es la proporción de individuos que no poseen esa característica, es decir, es $1-p$.

n: es el tamaño de la muestra.

Teniendo como población los cerca de 700 medicamentos del FNM, con un margen de error del 4.5% y un nivel de confianza del 99% se obtuvo que para el agrupamiento se deben utilizar 103 medicamentos (Tabla 2).

Tabla 2 Valores para el cálculo del tamaño de la muestra

Datos	Valor
N	700
K	0,99
E	0,045
P	0,5
Q	0,5
N	103,2926829

Los medicamentos escogidos para el agrupamiento son resultado de un análisis realizado a los datos recogidos por el DF donde se seleccionaron a los 30 medicamentos con mayor número de reporte de apariciones de RAM (Tabla 3), posteriormente consultando a un especialista se realizó en conjunto de problemas tipos en los cuales aparecen relacionados 44 medicamentos, entre ellos algunos de los antes mencionados. Para completar el tamaño de la muestra se siguió añadiendo medicamentos según el orden de aparición en los reportes (**Anexo 5**).

Tabla 3 Medicamentos según cantidad de reportes

MEDICAMENTO	CANTIDAD DE APARICIONES	MEDICAMENTO	CANTIDAD DE APARICIONES
CAPTOPRIL	3795	ERITROMICINA	622
AMOXICILINA	2419	AMINOFILINA	576
CIPROFLOXACINA	2256	NAPROXEN	551
PENICILINA RAPILENTA	2123	NIFEDIPINA	550
DIPIRONA	2061	CEFAZOLINA	525
IBUPROFEN	1968	ACIDO ACETILSALICILICO	508
ESPASMOFORTE	1273	HALOPERIDOL	508
METRONIDAZOL	1273	ATENOLOL	495
CEFALEXINA	1273	PIROXICAM	445
ENALAPRIL MALEATO	1221	DIFENHIDRAMINA	437
CO-TRIMOXAZOL	1207	CEFTRIAXONA SODICA	433
AZITROMICINA	1152	AMPICILLIN	418
PENICILINA G SODICA	969	TETRACICLINA	412
METOCLOPRAMIDA	962	PARACETAMOL	785
SALBUTAMOL	687		

El algoritmo propuesto para la cantidad de medicamentos introducidos (ciento tres), fue capaz de generar 19 grupos, se analizaron en total de 502 RAM asociadas a los medicamentos. Generándose los testores típicos permitiendo calcular los diferentes pesos asociados a las RAM.

3.2 Validación de los resultados de la función de semejanza

Con ayuda de un profesional de la medicina, y haciendo uso de 44 de los 103 medicamentos introducidos en el sistema, lo que representa el 42% del total de medicamentos, se realizó una encuesta (**Anexo 2**), orientada a los médicos, con cinco casos reales de ocurrencia de reacciones adversas.

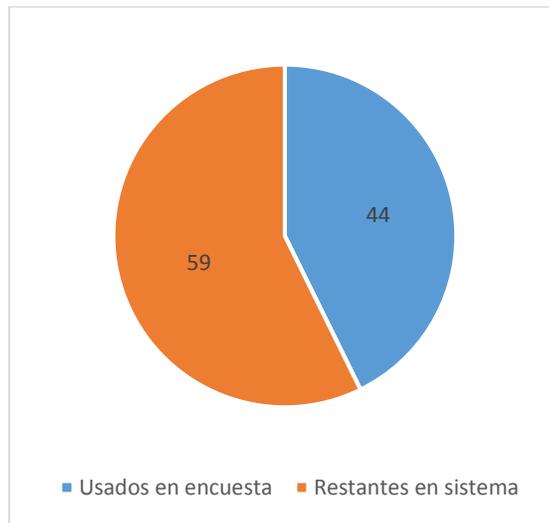


Figura 10 Distribución de medicamentos en el sistema y la encuesta

Cada pregunta de la encuesta aplicada (**Anexo 2**) consistente en un problema real que los médicos pueden enfrentar en consulta, se describe a un paciente con una RAM causada por un medicamento y dada una lista de medicamentos deben seleccionar cuáles considera como posibles sustitutos. El objetivo de la misma es tener un punto de comparación real que sirva para evaluar la veracidad de la información brindada por el sistema.

Dicha encuesta se aplicó a una muestra de ochenta profesionales vinculados con el Sistema Nacional de Salud (SNS), entre estos resaltan: treinta y siete Médicos General Integral (MGI), veinte Pediatras y veintitrés médicos de otras especialidades (Medicina interna, Anatomía Patológica, Cirugía, Gastroenterología, Ginecología, Máxilo facial, Toxicología, Reumatología, Psicología, Imaginología, Urología y Otorrino-laringología). La cantidad de encuestados y su diversidad atendiendo a su especialidad y provincia se muestran en (Figura 12) (Figura 11).

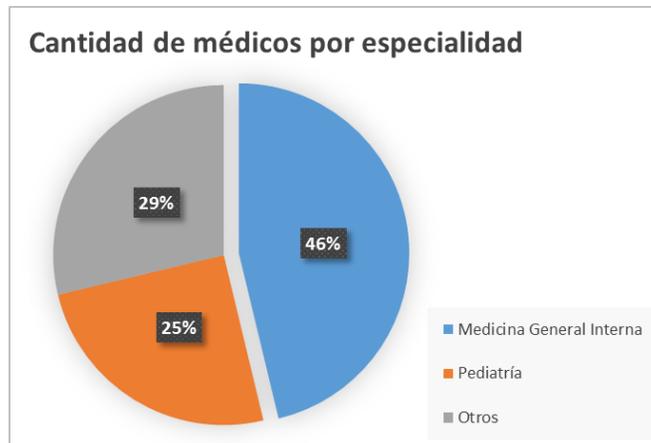


Figura 11 Médicos por especialidad encuestados

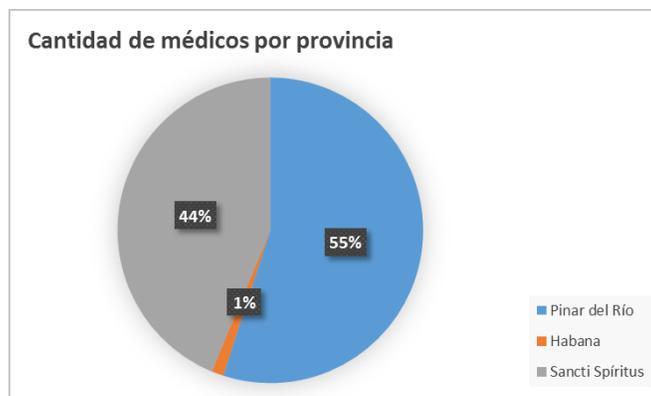


Figura 12 Médicos por provincia encuestados

Al aplicar la encuesta se obtuvieron un total de 400 respuestas que respondieron los médicos ante situaciones reales, a partir de estos se puede determinar el nivel de veracidad del sistema.

En todos los casos es necesario apreciar que la propuesta de solución implementada constituye una herramienta de apoyo a los médicos, por ello los resultados de la encuesta deben analizar las características de los problemas propuestos. Para un correcto análisis de la encuesta para la validación de los resultados se deben tener en cuenta algunos criterios médicos importantes:

- Factibilidad del medicamento: aun cuando dos medicamentos son totalmente diferentes en cuanto sus RAM, es posible que los médicos no decidan darlo, pues no solamente el médico analiza a la semejanza según sus RAM. Existen medicamentos que tienen una *velocidad* de acción en el sistema inferior a otros, y aun cuando se sabe que es preciso evitar las RAM en circunstancias es necesario “curar” por sobre todas las decisiones secundarias. Relación riesgo beneficios.
- Pacientes, no enfermedades: es común oír hablar a los médicos de los pacientes no de las enfermedades, de allí que los médicos prescriben según el paciente con el objetivo de impedir,

o curar, una enfermedad. Es lógico pensar que los médicos prescriban según a lo que están acostumbrados a tratar, si es un médico que solo atiende adultos resulta que es probable que resulte un enfoque distinto al de un pediatra.

Tomando en cuenta estas y otras consideraciones se procede a realizar un análisis de cada uno de los problemas tipos propuestos.

3.2.1 Análisis de los resultados de la pregunta 1

El primero de los problemas tipos propuestos pretende ver cual resulta el mejor remplazo a la Penicilina Rapilenta, para tratar a un paciente de 5 años de edad (ver **Anexo 2**). Como resultado se obtuvo (Figura 13) que en ninguna circunstancia los médicos recetan ni Ampicilina, ni Oxacilina, además consideramos como muy baja la selección de la Amoxicilina (2.05%), Ciprofloxacina (3.75%), Tetraciclina (7.50%) y Sulfaprín (11.25%).

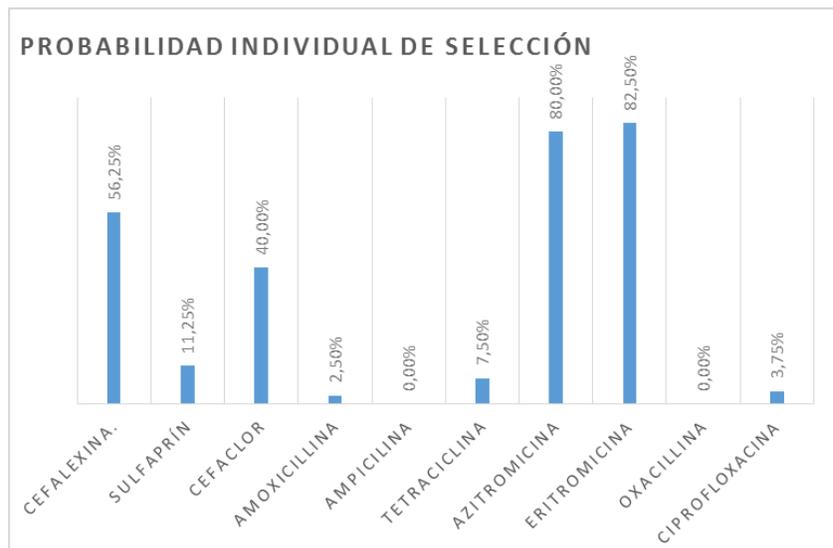


Figura 13 Probabilidad individual de selección de un medicamento en pregunta 1

Teniendo en cuenta el análisis anterior se propone desechar del análisis en el sistema a los medicamentos antes mencionados, pues un estudio posterior del Formulario Nacional del Medicamento demostró que los mismos tienen entre sus contraindicaciones el uso en niños, por tal es lógica la no selección de los mismos por los médicos. En un análisis de los resultados de la selección por especialidad (Figura 14) se puede ver como en el caso de los pediatras especialistas en la atención de niños (paciente de 5 años propuesto), no marcan resultados iguales a los MG, otros y la media, los cuales concuerdan entre sí.

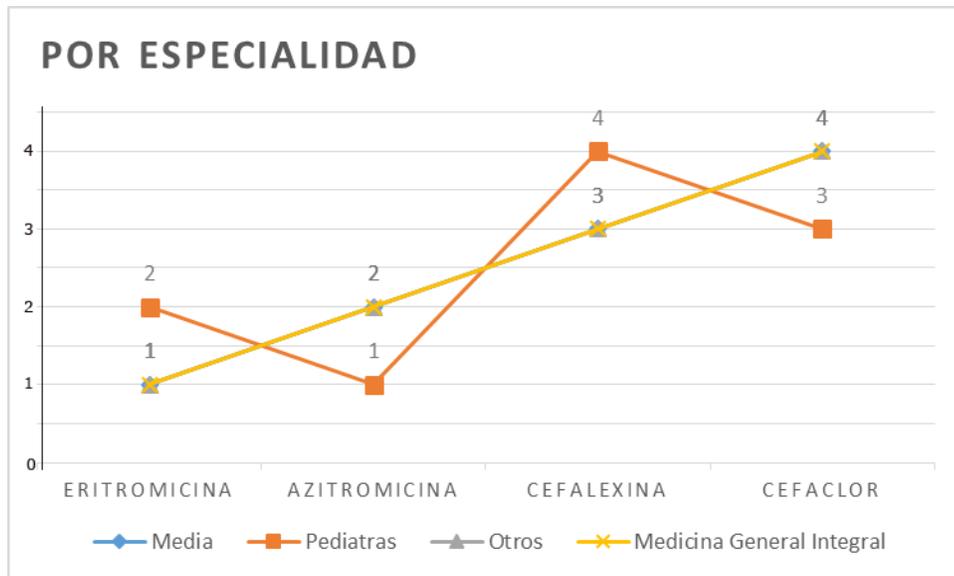


Figura 14 Selección de los medicamentos en la pregunta 1 según especialidad

Otro análisis de esta misma información enfocado en el tiempo de graduados de los especialistas da muestra que los especialistas de más de 20 años de graduados seleccionan de forma distinta a la media (Figura 15)

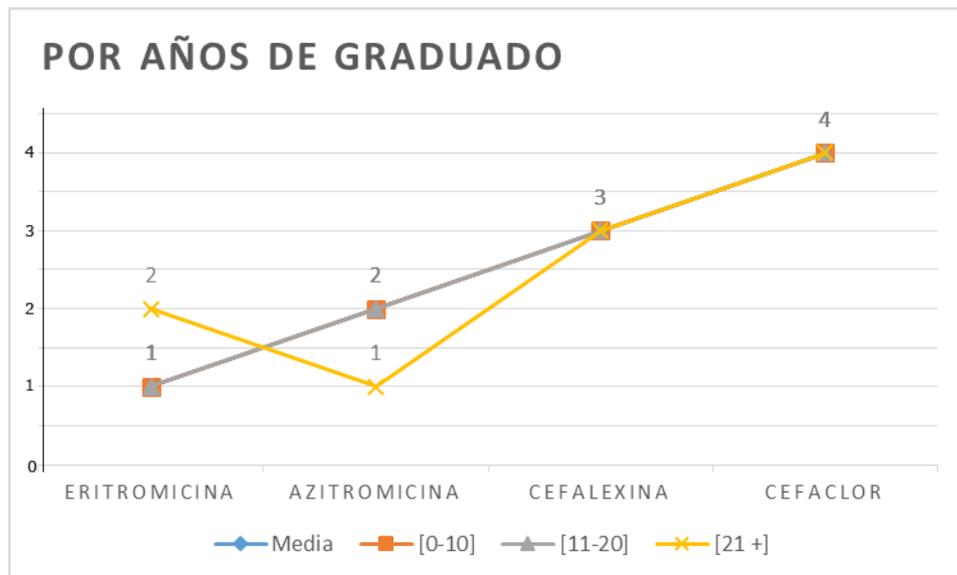


Figura 15 Selección de los medicamentos en la pregunta 1 según años de graduado

Un posterior análisis incluyendo los resultados de semejanza brindados por el sistema (Figura 16) da muestra de que a pesar de los resultados no estar totalmente en semejanza con el resultado de la media presenta iguales resultados que los especialistas con más de 20 años de graduado y de igual forma coincide en la primera y segunda opción con los pediatras.

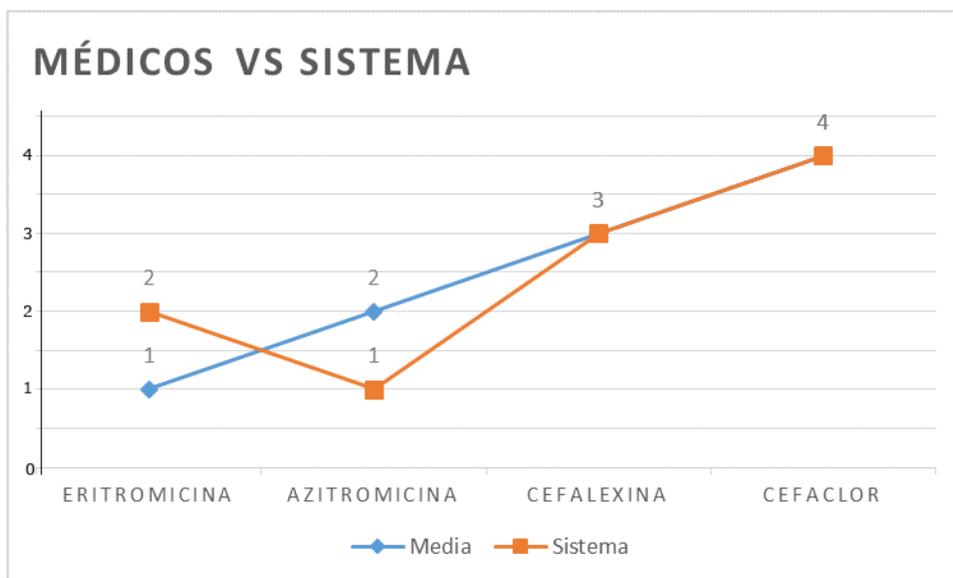


Figura 16 Resultados de la selección de medicamentos contra la semejanza en el sistema, pregunta 1

3.2.2 Análisis de los resultados de la pregunta 2

El segundo de los problemas tipos propuestos pretende ver cual resulta el mejor remplazo a al Ibuprofeno, para tratar a un paciente de 32 años de edad (**Anexo 2**). Como resultado se obtuvo (Figura 17Figura 13), según el análisis realizado que los médicos optan mayormente por el Metocarbamol y la Dipirona para realizar el cambio.

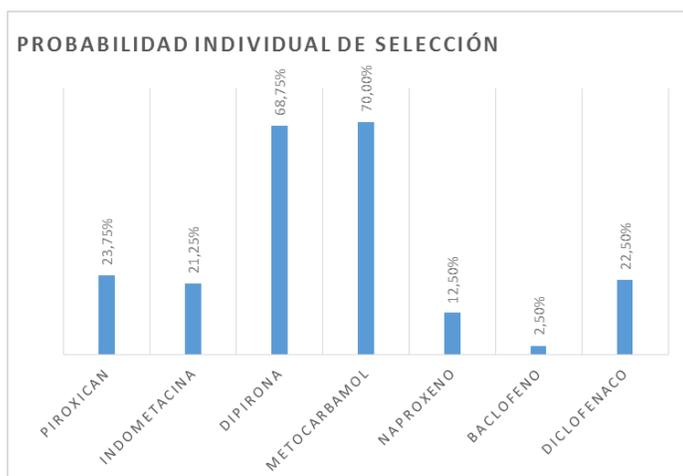


Figura 17 Probabilidad individual de selección de un medicamento en pregunta 2

Se puede observar que los médicos escogen al Baclofeno y al Naproxeno muy poco; al realizar un análisis del FNM se observa que pertenecen a la familia del Ibuprofeno, estando integrados por similares compuestos químicos. Los médicos evitan cambiar el medicamento por uno de similar composición, por este motivo en las observaciones realizadas no se tienen en cuenta al Baclofeno y al Naproxeno. Un análisis, por especialidades, de los resultados (Figura 18) muestra que los

resultados del sistema concuerdan con los pediatras y los que clasifican dentro de otras especialidades, en la selección de la Dipirona antes que el Metocarbamol. Un segundo resultado agrupado por los años de graduado, da muestra de una concordancia en todos los casos con los médicos de más de 21 años de graduado.

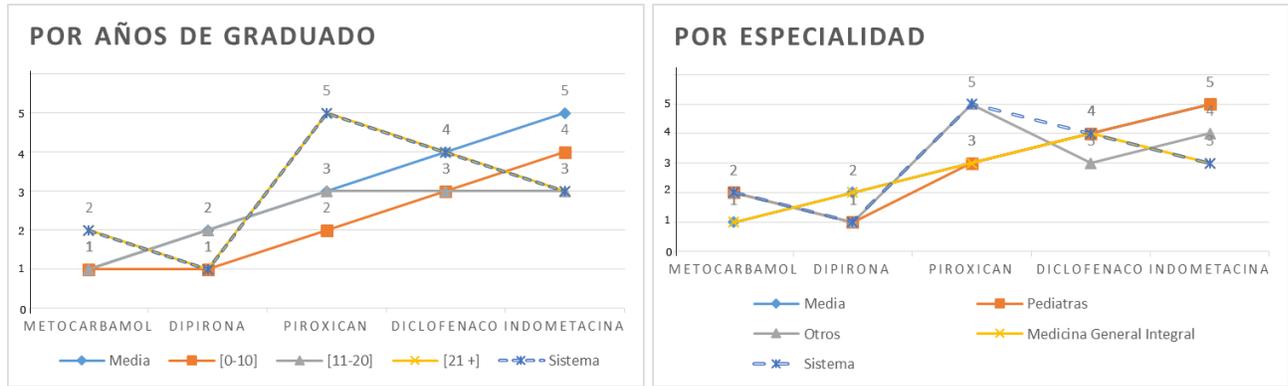


Figura 18 Selección de los medicamentos en la pregunta 2 según años de graduado y especialidad

3.2.3 Análisis de los resultados de la pregunta 3

El tercero de los problemas tipos propuestos pretende ver cual resulta el mejor remplazo a la Difenhidramina, para tratar a un paciente de 3 años de edad (**Anexo 2**). Como resultado se obtuvo (Figura 19) valores superiores todos al 20% de probabilidad de aparición por tal se consideraran a todos los medicamentos para el análisis.

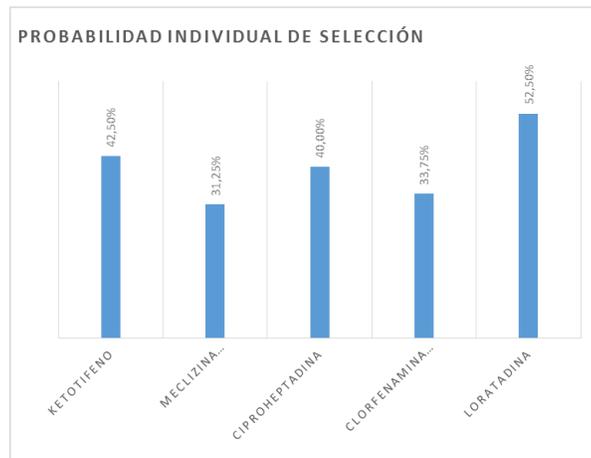


Figura 19 Probabilidad individual de selección de un medicamento en pregunta 3

Analizando los resultados agrupados atendiendo a los criterios de las especialidades y los años de graduado se muestra una discordancia en la gran mayoría de los puntos, lo cual dificulta la selección de un criterio más que el de la media.

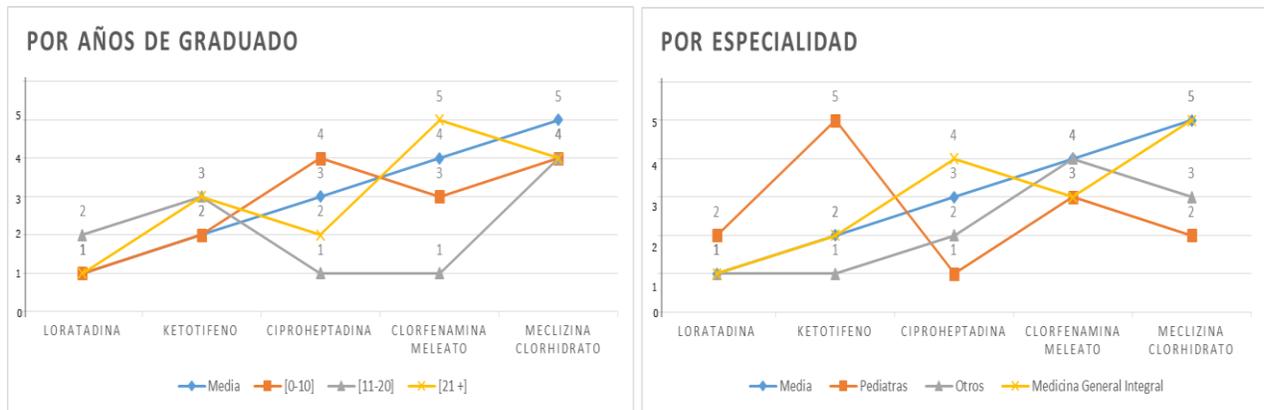


Figura 20 Selección de los medicamentos en la pregunta 3 según años de graduado y especialidad

Un análisis de los datos antes expuestos de conjunto con los del sistema (Figura 21), da muestra de que en relación a la media la semejanza en todos los casos dista en más menos uno en relación a la propuesta media de los médicos.

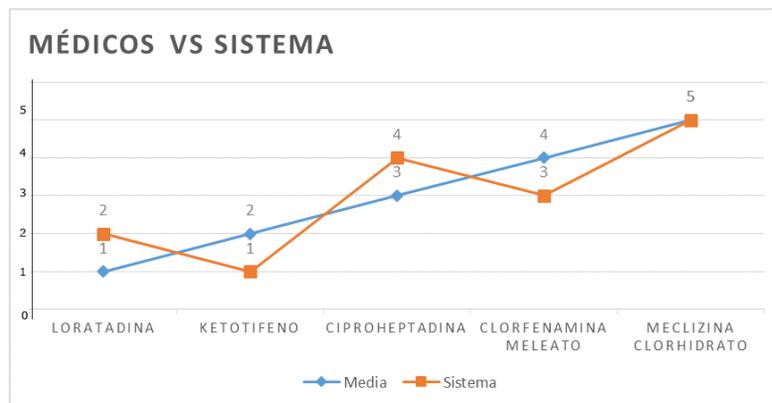


Figura 21 Resultados de la selección de medicamentos contra la semejanza en el sistema, pregunta 3

En este caso aun cuando los resultados aparentan estar alejados de la media es necesario tener en cuenta otros criterios, como el tiempo de acción del medicamento y la edad del paciente. Con relación a los MGI el sistema tiene un 60% de semejanza, por ser ellos la mayoría porcentual, y al existir tanta variedad de criterio y todos por debajo del 50% se puede considerar aceptables los resultados del sistema.

3.2.4 Análisis de los resultados de la pregunta 4

El cuarto de los problemas tipos propuestos pretende ver cual resulta el mejor remplazo a la Captopril, para tratar a un paciente de 54 años de edad (**Anexo 2**). En los resultados (Figura 22) se muestra a la Reserpina y la Dihidralizina, como medicamentos pocos seleccionados dado que en el Formulario Nacional de Medicamentos recomiendan no ser empleado en personas mayores.

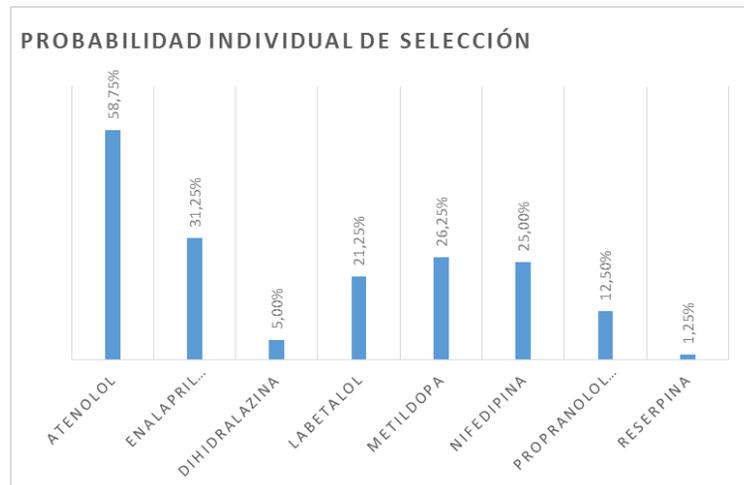


Figura 22 Probabilidad individual de selección de un medicamento en pregunta 4

En una comparación de la selección de los médicos agrupados según especialidad y años de graduado (Figura 23) se puede notar la discrepancia de criterios existentes respecto a la prescripción del caso propuesto.

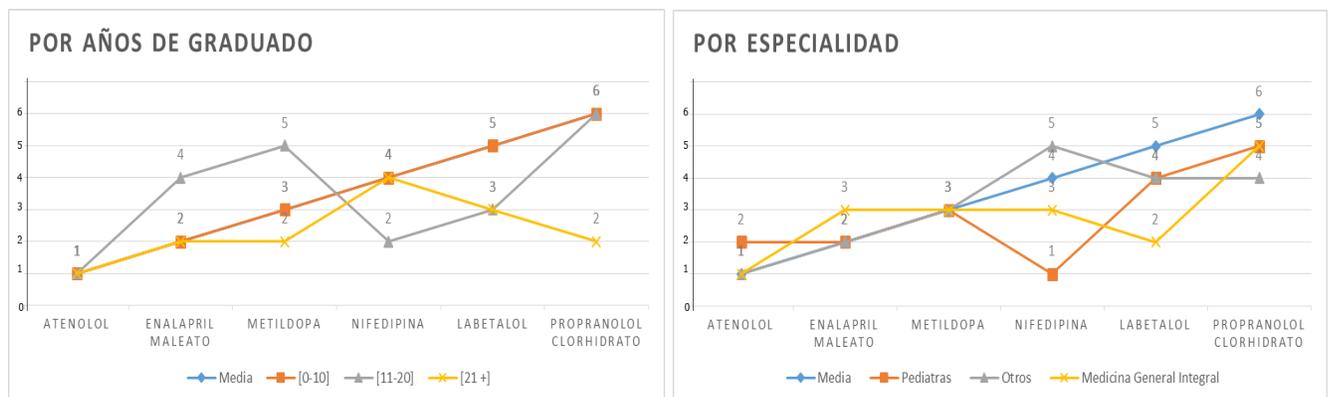


Figura 23 Selección de los medicamentos en la pregunta 4 según años de graduado y especialidad

Comparando los datos de semejanza ofrecidos por el sistema (Figura 24) se nota una cierta separación de la media propuesta por los médicos, pero en una revisión a priori del Formulario Nacional de Medicamentos el Labetalol tiene entre sus precauciones el uso en personas mayores, que tienden a estar propensos a la ocurrencia de RAM. Este dato no es tenido en cuenta por la función de semejanza, mas por los médicos sí.

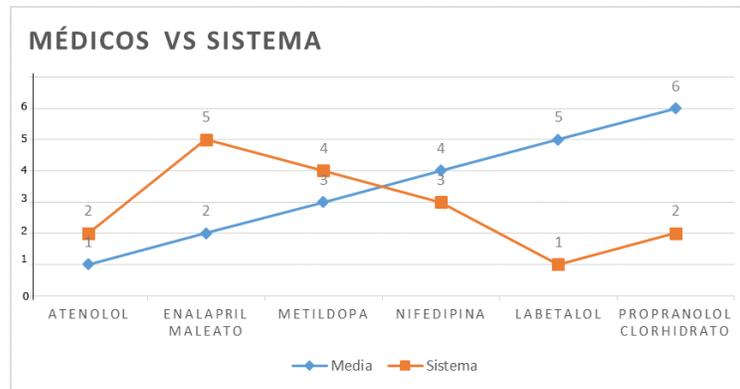


Figura 24 Resultados de la selección de medicamentos contra la semejanza en el sistema, pregunta 4

Teniendo esto en cuenta puede decirse que la aplicación tiene puntos de coincidencia con el 25% de la muestra (Pediatras) al seleccionar al Atenolol, como el segundo más propenso a ser dado al paciente, de igual forma comparte relación con los MGI los cuales proponen como segunda posible opción al Labetalol mientras este es el primero a dar según el sistema.

3.2.5 Análisis de los resultados de la pregunta 5

El quinto de los problemas tipos propuestos pretende ver cual resulta el mejor remplazo al tratamiento con Penicilina Cristalina, para tratar a un paciente de 7 años de edad (**Anexo 2**).

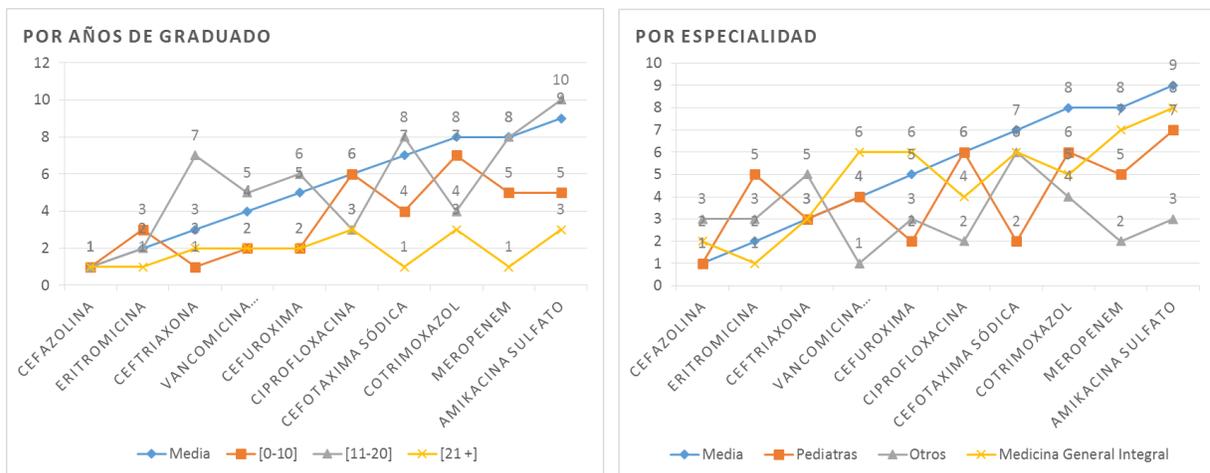


Figura 25 Selección de los medicamentos en la pregunta 5 según años de graduado y especialidad

En los datos resultados del análisis de lo propuesto por los médicos no se evidencia un criterio unificado en relación a cuál debe ser la prescripción, pues a pesar de parecer que existe una cohesión en los datos respecto a la Cefazolina este no es un criterio fuerte, pues un análisis sobre las gráficas anteriores muestra que no solo recae en ella el criterio de ser elegida como primera opción. La inclusión de los resultados del sistema en la gráfica (Figura 26) muestra como los resultados del sistema oscila acertadamente en los distintos criterios de selección.

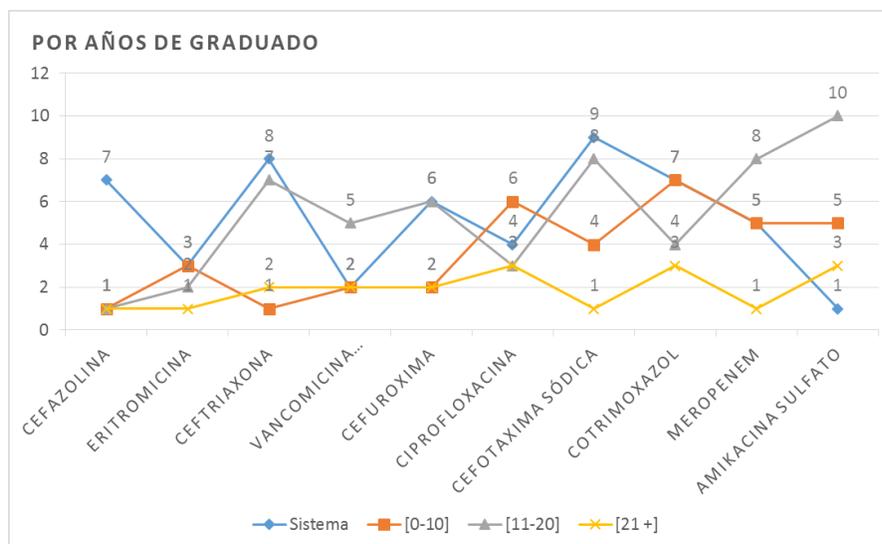


Figura 26 Criterios de los médicos según años de graduados y el sistema

3.2.6 Resultados del análisis

- Los médicos utilizan diferentes métodos para la prescripción; ya sea teniendo en cuenta la especialidad a la que se dedican, las características del paciente con que se encuentren en ese momento y los años de experiencia que tengan.
- La funcionalidad tiene como propósito brindar una herramienta de ayuda a la prescripción, y de acuerdo con el FNM, el médico debe tener en cuenta que el sistema muestra los resultados ordenados siempre empezando de los menos semejantes.

CONCLUSIONES

Se describe el resultado de una investigación que comenzó con el objetivo de obtener una herramienta que facilitará al personal facultativo la prescripción de medicamentos a pacientes que padecen de una RAM. Con dicho propósito se abordaron temas relacionados con los algoritmos de agrupamientos, y las técnicas de Reconocimiento de Patrones en la determinación de semejanzas entre objetos. Además se concluye que:

- Debido a que la información en el FNM se encuentra dispersa y no del todo concisa la búsqueda en él resulta compleja, haciendo el trabajo del médico engorroso; con el uso de la aplicación se le facilita al médico la determinación de la semejanza entre medicamentos para apoyar así la prescripción.
- Se realizó un análisis de sistemas informáticos y algoritmos que determinan la semejanza entre los medicamentos, atendiendo a las reacciones adversas que causan, y no se encontró ninguno que realice esta función.
- El Synta del CESIM posee nombrados los medicamentos del FNM, las reacciones adversas de los medicamentos y la frecuencia de aparición de las mismas, por lo que se decide agregarle la funcionalidad para determinar la semejanza entre medicamentos atendiendo a sus reacciones adversas.
- Se decide utilizar, dentro de las técnicas del Reconocimiento de Patrones, la selección de rasgos y la clasificación no supervisada permitiendo las mismas identificar la semejanza entre medicamentos atendiendo a sus reacciones adversas en el Formulario Nacional de medicamentos.
- Se identificó la información referente al objeto, rasgo y dominio del problema planteado, y se creó una Base de Conocimientos que se incorporó en el MRAM del Synta con dicha información necesaria para el cálculo de la semejanza entre medicamentos atendiendo a las reacciones adversas que causan.

RECOMENDACIONES

- Utilizar como algoritmo para el cálculo de los testores típicos el algoritmo FAST-BR con tuplas binarias acumulativas.
- Determinar la semejanza entre los medicamentos no solo teniendo en cuenta las reacciones adversas que ocasionan sino también la composición química del medicamento y las contraindicaciones que presentan.
- Mostrar los conceptos resultados de aplicar el algoritmo LC-Conceptual para agrupar medicamentos atendiendo a las reacciones adversas que causan a los profesionales de la medicina.
- Extraer los conocimientos de la semejanza entre medicamentos en un formato que puedan ser utilizados por el FNM web y la aplicación para Android.

REFERENCIAS

- [1]. Holloway, K. y Green, T. Comités de farmacoterapia Guía práctica. . 2004,
- [2]. González Podio, A.M. Características del consumo de antibióticos y de la resistencia bacteriana en la Ciudad de Santa Fé: estimación del gasto en antibióticos en un servicio de salud (1999-2001). *Características del consumo de antibióticos y de la resistencia bacteriana en la Ciudad de Santa Fé: estimación del gasto en antibióticos en un servicio de salud (1999-2001)*. Ediciones ISALUD, 2005,
- [3]. Saladrigas, M.V. El sistema de clasificación ATC de sustancias farmacéuticas para uso humano. *Panace*. 2004, Vol. 5, núm. 15, p. 59.
- [4]. Pedrás, J.R.L. y Tognoni, G. *Principios de epidemiología del medicamento*. Ediciones Científicas y Técnicas, 1993.
- [5]. Hegoa. Diccionario de Acción Humanitaria y Cooperación al Desarrollo. [en línea]. 2006. [Accedido: 10 abril 2015]. Recuperado: <http://www.dicc.hegoa.ehu.es/>.
- [6]. Shankar, P.R. Essential medicines and health products information portal. *Journal of pharmacology & pharmacotherapeutics*. Medknow Publications, 2014, Vol. 5, núm. 1, p. 74.
- [7]. Ministerio de Salud Nicaragua. *Formulario Nacional de Medicamentos*. 2014.
- [8]. Ministerio de salud Pública. Centro para el Desarrollo de la Farmacoepidemiología. *Formulario Nacional de Medicamentos*. Ed. Ciencias Médicas, Cuba, 2011. ISBN 978-959-212-706-7.
- [9]. Benítez Gener, A. et al. Glosario de términos farmacológicos. *Revista de Ciencias Médicas de Pinar del Río*. 1999, Editorial Ciencias Médicas, 2007, Vol. 11, núm. 5, p. 93–98.
- [10]. Formulario Nacional de Medicamentos. *Centro Nacional de Información de Ciencias Médicas* [en línea]. 2015. [Accedido: 19 enero 2015]. Recuperado: <http://fnmedicamentos.sld.cu/>.
- [11]. Armijo, J.A. et al. Estudios de seguridad de medicamentos: métodos para detectar las reacciones adversas y valoración de la relación causa-efecto. *El ensayo clínico en España*. Farmaindustria Madrid, 2001, p. 161–190.
- [12]. Upsala Monitor Centre UMC. *Upsala Monitor Centre* [en línea]. 2015. [Accedido: 2 febrero 2015]. Recuperado: <http://www.who-umc.org/>.
- [13]. García, F.D. y Alcalde, S.L. La farmacovigilancia, una reflexión 23 años después. *Resumed*. 2000, Vol. 13, núm. 5, p. 216–224.
- [14]. OMS. Portal de Información - Medicamentos Esenciales y Productos de Salud. [en línea]. 2015. [Accedido: 1 mayo 2015]. Recuperado: <http://apps.who.int/medicinedocs/es/d/Js6166s/4.html>.

- [15]. Ministerio de salud Pública. Centro Nacional de Información de Ciencias Médicas. Farmacoepidemiología. [en línea]. 2015. [Accedido: 1 mayo 2015]. Recuperado: <http://www.cdfc.sld.cu/node/2>.
- [16]. Unidad Coordinadora Nacional de Farmacovigilancia. INFORME DE FARMACOVIGILANCIA. 2014. . 2014.
- [17]. Arencibia Morales, A. et al. SLD137 SISTEMA INTEGRAL PARA EL CONTROL FARMACOLÓGICO. *Informática Salud 2013*. 2012,
- [18]. Ruiz Shulcloper, J. *Reconocimiento Lógico Combinatorio de Patrones: Teoría y Aplicaciones*. Tesis doctoral. Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV), 2009.
- [19]. Comas, D.S. y Meschino, G.J. Segmentación de Imágenes mediante Reconocimiento de Patrones. .
- [20]. Ruiz Shulcloper, J., Arenas, A.G. y Trinidad, J.F.M. *Enfoque Lógico Combinatorio al Reconocimiento de Patrones, I*. 1999. ISBN 970-18-2384-1.
- [21]. Jain, A.K., Duin, R.P.W. y Mao, J. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. IEEE, 2000, Vol. 22, núm. 1, p. 4–37.
- [22]. Wilson, D.R. y Martinez, T.R. Reduction techniques for instance-based learning algorithms. *Machine learning*. Springer, 2000, Vol. 38, núm. 3, p. 257–286.
- [23]. Sánchez, N.M., Lorenzo, M.M.G. y Valdivia, Z.Z.G. Modelo para diseñar sistemas de enseñanza-aprendizaje inteligentes utilizando el razonamiento basado en casos. *Revista Avances en Sistemas e Informática*. Universidad Nacional de Colombia, 2009, Vol. 6, núm. 3, p. 67–78.
- [24]. Alganza, Y.S. y Porrata, A.P. LEX: A NEW ALGORITHM FOR THE CALCULUS OF ALL TYPICAL TESTORS. . 2000,
- [25]. Pascual González, D. *Algoritmos de agrupamiento basados en densidad y validación de clusters*. Universitat Jaume I, Departament de Llenguatges i Sistemes Informàtics, 2010.
- [26]. Boyajian, W.L. Una propuesta de distancia entre procesos cuánticos. . 2011,
- [27]. Pérez Suárez, A. y Medina Pagola, J.E. Algoritmos para el agrupamiento conceptual de objetos. . La Habana, Cuba: 2014.
- [28]. Pons Porrata, A. *Desarrollo de algoritmos para la estructuración dinámica de información y su aplicación a la detección de sucesos*. Doctoral thesis, University Jaume I, Spain, 2004.
- [29]. Jain, A.K., Murty, M.N. y Flynn, P.J. *Data clustering: a review*. 1999. 1999. ISBN 0360-0300. DOI 10.1145/331499.331504.
- [30]. Gandón, A.G., Pons, S.V. y Ruiz Shulcloper, J. Algoritmos de agrupamiento conceptuales: un estado del arte. . La Habana, Cuba: 2012. Serie Azul.

- [31]. Reyes, A.J.O. y García, A.O. *Vista de Análisis usando la Técnica de Agrupamiento para el Sistema Integral para la Atención Primaria de Salud*. Universidad de las Ciencias Informáticas, 2012.
- [32]. Silveira Martineaux, K. y Fernández Pérez, R. Comparación de algoritmos de clasificación y agrupamiento aplicando técnicas de Minería de datos. . La Habana, Cuba: 2008.
- [33]. Fast Threshold Clustering Algorithm (FTCA). [en línea]. 2013. [Accedido: 26 mayo 2015]. Recuperado: <https://cssanalytics.wordpress.com/2013/11/26/fast-threshold-clustering-algorithm-ftca/>.
- [34]. Apache Commons. [en línea]. [Accedido: 10 abril 2015]. Recuperado: <http://commons.apache.org/proper/commons-math/userguide/ml.html>.
- [35]. Suarez Lorenzo, G. y Hernández Cárdenas, Y. Módulos Consumo de Medicamentos y Reacciones Adversas a Medicamentos (RAM) para el producto Synta. . La Habana, Cuba: 2011.
- [36]. Trujillo Irarragorri, L.E. y Corona Prendes, F.D. *SISTEMA BASADO EN CASOS PARA PREDECIR LA OCURRENCIA DE REACCIONES ADVERSAS A MEDICAMENTOS EN LA CONSULTA MÉDICA*. Universidad de las Ciencias Informáticas, 2013.
- [37]. Reyes García, J. y Kindelán Colmenero, J.C. *SISTEMA BASADO EN REGLAS PARA DETECTAR INTERACCIÓN MEDICAMENTOSA DESDE LA CONSULTA MÉDICA*. Universidad de las Ciencias Informáticas, 2014.
- [38]. Villa Jiménez, O.M. Formulario Nacional de Medicamentos. *Infomed* [en línea]. 2015. [Accedido: 16 febrero 2015]. Recuperado: <http://fnmedicamentos.sld.cu/index.php?P=aplicacionAndroid>.
- [39]. Asociación de Medicina de la Industria Farmacéutica de España. AMIFE. [en línea]. 2015. [Accedido: 14 abril 2015]. Recuperado: www.amife.com.
- [40]. EHealth Reporter Latin America. EHealth Reporter Latin America. [en línea]. 2014. [Accedido: 15 abril 2015]. Recuperado: www.ehealthreporter.com.
- [41]. Alegsa, L. Alegsa.com.ar. [en línea]. 2015. [Accedido: 21 mayo 2015]. Recuperado: <http://www.alegsa.com.ar/Dic/sgbd.php>.
- [42]. MySQL, A.B. The world's most popular open source database. *MySQL AB*. 2007,
- [43]. EMS Database Management Solutions. EMS Database Management Solutions. *EMS Database Management Solutions* [en línea]. 2015. [Accedido: 4 mayo 2015]. Recuperado: <http://www.sqlmanager.net>.
- [44]. Sierra García, M. Comunidad de diseñadores de España. [en línea]. 2015. [Accedido: 21 mayo 2015]. Recuperado: <http://www.aprenderaprogramar.com/>.
- [45]. Gerner, J. et al. *Professional LAMP: Linux, Apache, MySQL, and PHP Web Development*. 2006. ISBN 9780764597237.

- [46]. ACHOUR, M. y Others. *Manual de PHP*. [en línea] *Manual de PHP, 2012*. [Consultado el: 29 de septiembre de 2012].
- [47]. Jendrock, E. et al. *The Java EE 7 Tutorial*. Addison-Wesley Professional, 2014.
- [48]. MacCaw, A. *JavaScript Web Applications*. 2011. ISBN 9781449303518.
- [49]. Kleinfeld, S. *HTML5 for Publishers* [en línea]. 2011. ISBN 9781449314606. Recuperado: <http://shop.oreilly.com/product/0636920022473.do>.
- [50]. Cederholm, D. *CSS3 FOR WEB DESIGNERS*. 2010. ISBN 9780984442522. DOI 10.1016/j.ijhcs.2006.06.002.
- [51]. Potencier, F. y Zaninotto, F. *Symfony 1.2, la guía definitiva*. 2008. Apress, 2008.
- [52]. Chandan, R. *Sencha/ExtJS: Object Oriented JavaScript*. 2011. 2011.
- [53]. Oracle Corporation. NetBeans Release Roadmap. [en línea]. 2015. [Accedido: 25 mayo 2015]. Recuperado: <https://netbeans.org/community/releases/roadmap.html>.
- [54]. JetBrains s.r.o. PHP Storm. [en línea]. 2015. [Accedido: 5 febrero 2014]. Recuperado: <https://www.jetbrains.com/phpstorm/features/>.
- [55]. Pascual, D., Pla, F. y Sánchez, S. Algoritmos de agrupamiento. *Unpublished*. p. 13.
- [56]. Ortiz-Posadas, M.R., Martínez Trinidad, J.F. y Ruiz Shulcloper, J. A new approach to differential diagnosis of diseases. *International journal of bio-medical computing*. Elsevier, 1996, Vol. 40, núm. 3, p. 179–185.
- [57]. Ruiz Shulcloper, J., Bravo, M. y Aguila, F. Algoritmos BT y TB para el cálculo de todos los tests típicos. *Revista Ciencias Matemáticas*. 1982, Vol. 6, núm. 2.
- [58]. Bravo, A. Algoritmo CT para el cálculo de los testores típicos de una matriz k-valente. *Revista Ciencias Matemáticas*. 1983, Vol. 4, núm. 2, p. 123–144.
- [59]. Águila, L. y Ruiz Shulcloper, J. Algoritmo CC para la elaboración de la información k-valente en problemas de Reconocimiento de Patrones. *Revista Ciencias Matemáticas*. 1984, Vol. 5, núm. 3.
- [60]. Ruiz Shulcloper, J., Alba, C. y Lazo Cortés, M. Introducción a la teoría de Testores Típicos. *Serie verde*. 1995, núm. 50.
- [61]. Lias Rodríguez, A. y Pons Porrata, A. Un Nuevo Algoritmo de Escala Exterior para el Cálculo de los Testores Típicos. .
- [62]. Sanchez Díaz, G. y Lazo Cortés, M. CT-EXT: an algorithm for computing typical testor set. *Progress in Pattern Recognition, Image Analysis and Applications*. Springer, 2007, p. 506–514.

- [63]. Lias Rodríguez, A. y Sanchez Díaz, G. AN ALGORITHM FOR COMPUTING TYPICAL TESTORS BASED ON ELIMINATION OF GAPS AND REDUCTION OF COLUMNS. . 2013, Vol. 27, núm. 8. DOI 10.1142/S0218001413500225.
- [64]. Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. *Metodología de la investigación* [en línea]. 2006. ISBN 9701057538. Recuperado: http://www.univo.edu.sv:8081/tesis/020090/020090_Cap1.pdf.

BIBLIOGRAFÍA

1. ACHOUR, M. y Others. *Manual de PHP*. [en línea] *Manual de PHP, 2012*. [Consultado el: 29 de septiembre de 2012].
2. Águila, L. y Ruiz Shulcloper, J. Algoritmo CC para la elaboración de la información k-valente en problemas de Reconocimiento de Patrones. *Revista Ciencias Matemáticas*. 1984, Vol. 5, núm. 3.
3. Alegsa, L. Alegsa.com.ar. [en línea]. 2015. [Accedido: 21 mayo 2015]. Recuperado: <http://www.alegsa.com.ar/Dic/sghbd.php>.
4. Alganza, Y.S. y Porrata, A.P. LEX: A NEW ALGORITHM FOR THE CALCULUS OF ALL TYPICAL TESTORS. . 2000,
5. Anaya Sánchez, H., Pons Porrata, A. y Berlanga Llavori, R. A new document clustering algorithm for topic discovering and labeling. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2008,p. 161–168. ISBN 3540859195. DOI 10.1007/978-3-540-85920-8_20.
6. Anaya Sánchez, H., Pons Porrata, A. y Berlanga Llavori, R. A document clustering algorithm for discovering and describing topics. *Pattern Recognition Letters*. 2010, Vol. 31, núm. 6, p. 502–510. ISSN 01678655. DOI 10.1016/j.patrec.2009.11.013.
7. Arencibia Morales, A. et al. SLD137 SISTEMA INTEGRAL PARA EL CONTROL FARMACOLÓGICO. *Informática Salud 2013*. 2012,
8. Armijo, J.A. et al. Estudios de seguridad de medicamentos: métodos para detectar las reacciones adversas y valoración de la relación causa-efecto. *El ensayo clínico en España*. Farmaindustria Madrid, 2001, p. 161–190.
9. Benítez Gener, A. et al. Glosario de términos farmacológicos. *Revista de Ciencias Médicas de Pinar del Río*. 1999, Editorial Ciencias Médicas, 2007, Vol. 11, núm. 5, p. 93–98.
10. Bishop, C.M.C.C.M. *Pattern Recognition and Machine Learning* [en línea]. Springer, 2006. Information science and statistics. ISBN 978-0387310732. DOI 10.1117/1.2819119. Recuperado: <http://www.library.wisc.edu/selectedtoocs/bg0137.pdf>.
11. Boyajian, W.L. Una propuesta de distancia entre procesos cuánticos. . 2011,
12. Bravo, A. Algoritmo CT para el cálculo de los testores típicos de una matriz k-valente. *Revista Ciencias Matemáticas*. 1983, Vol. 4, núm. 2, p. 123–144.
13. Cederholm, D. *CSS3 FOR WEB DESIGNERS*. 2010. ISBN 9780984442522. DOI 10.1016/j.ijhcs.2006.06.002.
14. Chandan, R. *Sencha/ExtJS: Object Oriented JavaScript*. 2011. 2011.
15. Comas, D.S. y Meschino, G.J. Segmentación de Imágenes mediante Reconocimiento de Patrones. .
16. EMS Database Management Solutions. EMS Database Management Solutions. *EMS Database Management Solutions* [en línea]. 2015. [Accedido: 4 mayo 2015]. Recuperado: <http://www.sqlmanager.net>.
17. Gandón, A.G., Pons, S.V. y Ruiz Shulcloper, J. Algoritmos de agrupamientos conceptuales: un estado del arte. . La Habana, Cuba: 2012. Serie Azul.

18. García, F.D. y Alcalde, S.L. La farmacovigilancia, una reflexión 23 años después. *Resumed*. 2000, Vol. 13, núm. 5, p. 216–224.
19. Gerner, J. et al. *Professional LAMP: Linux, Apache, MySQL, and PHP Web Development*. 2006. ISBN 9780764597237.
20. González Podio, A.M. Características del consumo de antibióticos y de la resistencia bacteriana en la Ciudad de Santa Fé: estimación del gasto en antibióticos en un servicio de salud (1999-2001). *Características del consumo de antibióticos y de la resistencia bacteriana en la Ciudad de Santa Fé: estimación del gasto en antibióticos en un servicio de salud (1999-2001)*. Ediciones ISALUD, 2005,
21. Hegoa. Diccionario de Acción Humanitaria y Cooperación al Desarrollo. [en línea]. 2006. [Accedido: 10 abril 2015]. Recuperado: <http://www.dicc.hegoa.ehu.es/>.
22. Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. *Metodología de la investigación* [en línea]. 2006. ISBN 9701057538. Recuperado: http://www.univo.edu.sv:8081/tesis/020090/020090_Cap1.pdf.
23. Holloway, K. y Green, T. *Comités de farmacoterapia Guía práctica*. . 2004,
24. Jain, A.K., Duin, R.P.W. y Mao, J. Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. IEEE, 2000, Vol. 22, núm. 1, p. 4–37.
25. Jain, A.K., Murty, M.N. y Flynn, P.J. *Data clustering: a review*. 1999. 1999. ISBN 0360-0300. DOI 10.1145/331499.331504.
26. Jendrock, E. et al. *The Java EE 7 Tutorial*. Addison-Wesley Professional, 2014.
27. JetBrains s.r.o. PHP Storm. [en línea]. 2015. [Accedido: 5 febrero 2014]. Recuperado: <https://www.jetbrains.com/phpstorm/features/>.
28. Kleinfeld, S. *HTML5 for Publishers* [en línea]. 2011. ISBN 9781449314606. Recuperado: <http://shop.oreilly.com/product/0636920022473.do>.
29. Lias Rodríguez, A. y Pons Porrata, A. Un Nuevo Algoritmo de Escala Exterior para el Cálculo de los Testores Típicos. .
30. Lias Rodríguez, A. y Pons Porrata, A. BR: A New Method for Computing All Typical Testors. . 2009, p. 433–440.
31. Lias Rodríguez, A. y Pons Porrata, A. Un Nuevo Algoritmo de Escala Exterior para el Cálculo de los Testores Típicos.
32. Lias Rodríguez, A. y Sanchez Díaz, G. AN ALGORITHM FOR COMPUTING TYPICAL TESTORS BASED ON ELIMINATION OF GAPS AND REDUCTION OF COLUMNS. . 2013, Vol. 27, núm. 8. DOI 10.1142/S0218001413500225.
33. MacCaw, A. *JavaScript Web Applications*. 2011. ISBN 9781449303518.
34. Ministerio de Salud Nicaragua. *Formulario Nacional de Medicamentos*. 2014.
35. Ministerio de Salud Nicaragua y Ministerio de salud Pública. Centro para el Desarrollo de la Farmacoepidemiología. *Formulario Nacional de Medicamentos* [en línea]. Ed. Ciencias Médicas, Cuba, 2014. ISBN 978-959-212-706-7. [Accedido: 19 enero 2015]. Recuperado: <http://fnmedicamentos.sld.cu/>.
36. Ministerio de salud Pública. Centro Nacional de Información de Ciencias Médicas. Farmacoepidemiología. [en línea]. 2015. [Accedido: 1 mayo 2015]. Recuperado: <http://www.cdfc.sld.cu/node/2>.

37. Ministerio de salud Pública. Centro para el Desarrollo de la Farmacoepidemiología. *Formulario Nacional de Medicamentos*. Ed. Ciencias Médicas, Cuba, 2011. ISBN 978-959-212-706-7.
38. MySQL, A.B. The world's most popular open source database. *MySQL AB*. 2007,
39. OMS. Portal de Información - Medicamentos Esenciales y Productos de Salud. [en línea]. 2015. [Accedido: 1 mayo 2015]. Recuperado: <http://apps.who.int/medicinedocs/es/d/Js6166s/4.html>.
40. Oracle Corporation. NetBeans Release Roadmap. [en línea]. 2015. [Accedido: 25 mayo 2015]. Recuperado: <https://netbeans.org/community/releases/roadmap.html>.
41. Ortiz-Posadas, M.R., Martínez Trinidad, J.F. y Ruiz Shulcloper, J. A new approach to differential diagnosis of diseases. *International journal of bio-medical computing*. Elsevier, 1996, Vol. 40, núm. 3, p. 179–185.
42. Pascual González, D. *Algoritmos de agrupamiento basados en densidad y validación de clusters*. Universitat Jaume I, Departament de Llenguatges i Sistemes Informàtics, 2010.
43. Pascual, D., Pla, F. y Sánchez, S. Algoritmos de agrupamiento. *Unpublished*. p. 13.
44. Pedrás, J.R.L. y Tognoni, G. *Principios de epidemiología del medicamento*. Ediciones Científicas y Técnicas, 1993.
45. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. ... *of Machine Learning* ... [en línea]. 2012, Vol. 12, p. 2825–2830. ISSN 15324435. Recuperado: <http://dl.acm.org/citation.cfm?id=2078195> \n <http://arxiv.org/abs/1201.0490>.
46. Pérez Suárez, A. y Medina Pagola, J.E. Algoritmos para el agrupamiento conceptual de objetos. . La Habana, Cuba: 2014.
47. Pons Porrata, A. *Desarrollo de algoritmos para la estructuración dinámica de información y su aplicación a la detección de sucesos*. Doctoral thesis, University Jaume I, Spain, 2004.
48. Pons Porrata, A., Berlanga Llavori, R. y Ruiz Shulcloper, J. *Topic discovery based on text mining techniques*. 2007. 2007. DOI 10.1016/j.ipm.2006.06.001.
49. Potencier, F. y Zaninotto, F. *Symfony 1.2, la guía definitiva*. 2008. Apress, 2008.
50. Reyes García, J. y Kindelán Colmenero, J.C. *SISTEMA BASADO EN REGLAS PARA DETECTAR INTERACCIÓN MEDICAMENTOSA DESDE LA CONSULTA MÉDICA*. Universidad de las Ciencias Informáticas, 2014.
51. Reyes González, Y. *Modelo para la adaptación de las soluciones en un Sistema Basado en Casos utilizando el agrupamiento conceptual*. Universidad de las Ciencias Informáticas, 2014.
52. Reyes, A.J.O. y García, A.O. *Vista de Análisis usando la Técnica de Agrupamiento para el Sistema Integral para la Atención Primaria de Salud*. Universidad de las Ciencias Informáticas, 2012.
53. Ruiz Shulcloper, J. Logical Combinatorial Pattern Recognition. *ScientificConnors CiteSeerX-Scientific Literature Digital Library and Search Engine (United States)*. 2002,
54. Ruiz Shulcloper, J. *Reconocimiento Lógico Combinatorio de Patrones: Teoría y Aplicaciones*. Tesis doctoral. Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV), 2009.
55. Ruiz Shulcloper, J., Alba, C. y Lazo Cortés, M. Introducción a la teoría de Testores Típicos. *Serie verde*. 1995, núm. 50.
56. Ruiz Shulcloper, J., Arenas, A.G. y Trinidad, J.F.M. *Enfoque Lógico Combinatorio al Reconocimiento de Patrones, I*. 1999. ISBN 970-18-2384-1.

57. Ruiz Shulcloper, J., Bravo, M. y Aguila, F. Algoritmos BT y TB para el cálculo de todos los tests típicos. *Revista Ciencias Matemáticas*. 1982, Vol. 6, núm. 2.
58. Saladrigas, M.V. El sistema de clasificación ATC de sustancias farmacéuticas para uso humano. *Panace*. 2004, Vol. 5, núm. 15, p. 59.
59. Sanchez Díaz, G. y Lazo Cortés, M. CT-EXT: an algorithm for computing typical testor set. *Progress in Pattern Recognition, Image Analysis and Applications*. Springer, 2007, p. 506–514.
60. Sanchez Díaz, G., Lazo Cortés, M. y Piza Dávila, H.I. A Fast Implementation for the Typical Testor Property Identification Based on an Accumulative Binary Tuple. . 2012, Vol. 5.
61. Sánchez, N.M., Lorenzo, M.M.G. y Valdivia, Z.Z.G. Modelo para diseñar sistemas de enseñanza-aprendizaje inteligentes utilizando el razonamiento basado en casos. *Revista Avances en Sistemas e Informática*. Universidad Nacional de Colombia, 2009, Vol. 6, núm. 3, p. 67–78.
62. Shankar, P.R. Essential medicines and health products information portal. *Journal of pharmacology & pharmacotherapeutics*. Medknow Publications, 2014, Vol. 5, núm. 1, p. 74.
63. Sierra García, M. Comunidad de diseñadores de España. [en línea]. 2015. [Accedido: 21 mayo 2015]. Recuperado: <http://www.aprenderaprogramar.com/>.
64. Silveira Martineaux, K. y Fernández Pérez, R. Comparación de algoritmos de clasificación y agrupamiento aplicando técnicas de Minería de datos. . La Habana, Cuba: 2008.
65. Suarez Lorenzo, G. y Hernández Cárdenas, Y. Módulos Consumo de Medicamentos y Reacciones Adversas a Medicamentos (RAM) para el producto Synta. . La Habana, Cuba: 2011.
66. Trujillo Irarragorri, L.E. y Corona Prendes, F.D. *SISTEMA BASADO EN CASOS PARA PREDECIR LA OCURRENCIA DE REACCIONES ADVERSAS A MEDICAMENTOS EN LA CONSULTA MÉDICA*. Universidad de las Ciencias Informáticas, 2013.
67. Unidad Coordinadora Nacional de Farmacovigilancia. INFORME DE FARMACOVIGILANCIA. 2014. . 2014.
68. Villa Jiménez, O.M. Formulario Nacional de Medicamentos. *Infomed* [en línea]. 2015. [Accedido: 16 febrero 2015]. Recuperado: <http://fnmedicamentos.sld.cu/index.php?P=aplicacionAndroid>.
69. Wilson, D.R. y Martinez, T.R. Reduction techniques for instance-based learning algorithms. *Machine learning*. Springer, 2000, Vol. 38, núm. 3, p. 257–286.
70. Formulario Nacional de Medicamentos. *Centro Nacional de Información de Ciencias Médicas* [en línea]. 2015. [Accedido: 19 enero 2015]. Recuperado: <http://fnmedicamentos.sld.cu/>.
71. Upsala Monitor Centre UMC. *Upsala Monitor Centre* [en línea]. 2015. [Accedido: 2 febrero 2015]. Recuperado: <http://www.who-umc.org/>.
72. Apache Commons. [en línea]. [Accedido: 10 abril 2015]. Recuperado: <http://commons.apache.org/proper/commons-math/userguide/ml.html>.
73. Fast Threshold Clustering Algorithm (FTCA). [en línea]. 2013. [Accedido: 26 mayo 2015]. Recuperado: <https://cssanalytics.wordpress.com/2013/11/26/fast-threshold-clustering-algorithm-ftca/>.

ANEXOS

Anexo 1 Fases del desarrollo de un nuevo medicamento y relación con su seguridad

Fase	Objetivo
Fase 0 ó preclínica	Estudios de toxicidad y establecimiento del índice terapéutico del fármaco.
Fase I en voluntarios	Seguridad del fármaco.
Fases II y III	Eficacia y seguridad del medicamento y establecimiento de la relación eficacia-toxicidad.
Fase IV poscomercialización	Valoración de la seguridad del fármaco en las condiciones habituales de uso.

Anexo 2 Encuesta realizada a los especialistas en medicina

Una RAM es una respuesta nociva no intencionada provocada por un fármaco tras su administración. A continuación se presentan 5 casos de estudio de situaciones médicas reales donde se pretende conocer la respuesta del personal facultativo ante la situación:

Especialidad _____ Edad _____
Años de graduado _____ Sexo _____

1. Paciente de 5 años de edad que fue atendido por los servicios médicos de salud por presentar cuadro de Amigdalitis Aguda con Exudados y Membranas para lo cual le indican tratamiento con Penicilina rapilenta, al segundo día de habersele administrado la paciente comienza a presentar un rash morbiliforme en todo el cuerpo de color rosado intenso, prurito asociado, se reevalúa y se interpreta como reacción medicamentosa, ante esta situación clínica se debe cambiar el medicamento

a) De las siguientes opciones de antibiótico marque con una X el o los antibióticos que se puede emplear como segunda opción:

Cefalexina	Sulfaprín	Cefaclor	Amoxicilina	Ampicilina
Tetraciclina	Azitromicina	Eritromicina	Oxacilina	Ciprofloxacina

2. Paciente de 32 años de edad que acudió a consulta hace 3 días por presentar dolores en región sacroilíaca, el médico le diagnostica Sacrolumbalgia orientando tomar Paracetamol 1 tab. C/ 6 horas e Ibuprofeno 1 tab. C/ 8 horas, 48 horas después el paciente es vuelto a valorar por presentar aumento de volumen en el cuello sin otros signos flogísticos que sugieran infección, al analizar los

medicamentos empleados se demuestra que estos tienen sinergismo, el galeno decide cambiar el tratamiento

a) De los siguientes antiinflamatorios y/o analgésicos marque con una X aquel o aquellos que se puedan emplear ante esta situación en el paciente

Piroxicam	Dipirona	Naproxeno	Diclofenaco
Indometacina	Metocarbamol	Baclofeno	

3. Paciente de 3 años de edad alérgico desde edades tempranas, que presenta cuadro de varicela, valorado por el especialista quien decide indicar difenhidramina 1 tab c/ 12 horas por 10 días, al tercer día la madre viene asustada porque el niño se muestra agresivo e irritable con sus padres, acompañándose de insomnio, al revisar el medicamento impuesto se considera que el paciente está presentando síntomas de reacciones adversas al mismo, se decide realizar cambio al tratamiento

a) De los siguientes antihistamínicos marque con una X aquel o aquellos que puedan ser reemplazado, minimizando los síntomas para este paciente

Ketotifeno	Ciproheptadina	Clorfenamina maleato
Meclizina clorhidrato	Loratadina	

4. Paciente de 54 años de edad, masculino que debuta con hipertensión arterial, se comienza tratamiento con Captopril 1 tab. C/ 8 horas, a los 15 días de iniciado el mismo el paciente recurre nuevamente al consultorio médico por presentar cefalea mantenida que no se alivia con analgésicos habituales, el médico decide realizar cambios al tratamiento.

a) De los hipotensores presentes en nuestro mercado marque con una X cual o cuales pueden ser utilizados por el paciente minimizando los síntomas referidos

Atenolol	Dihidralazina	Metildopa	Propranolol clorhidrato
Enalapril maleato	Labetalol	Nifedipina	Reserpina

5. Paciente de 7 años de edad, con antecedentes de asma bronquial, que se le diagnostica neumonía de base pulmonar derecha virgen de tratamiento antibiótico, el doctor decide iniciar tratamiento con Penicilina Cristalina intramuscular a lo que la madre hace referencia que el niño es alérgico a este antibiótico desde los 3 años de edad pero que además no puede tomar medicamentos que contengan tartracina porque le exacerba la crisis de asma bronquial, ante esta situación el médico decide ingresar al paciente

a) De las siguientes opciones de medicamentos de uso intrahospitalarios en nuestro sistema de salud marque con una X cual o cuales pueden utilizarse en este tipo de paciente

Cefazolina	Ceftriaxona	Cotrimoxazol	Ciprofloxacina	Eritromicina
Cefotaxima sódica	Cefuroxima	Amikacina sulfato	Vancomicina clorhidrato	Meropenem

Anexo 3 Descripción del algoritmo LC-Conceptual

Sea $O = \{O_1, O_2, \dots, O_n\}$ una colección de objetos, descritos en términos de un conjunto de atributos que pueden ser cualitativos, numéricos o incluso puede desconocerse su valor. Sea además β una función tal que $\beta(O_i, O_j)$ establece el grado de semejanza que tiene un par de objetos $O_i, O_j \in O$. El algoritmo LC representa la colección de objetos como un grafo en el cual los objetos de la colección son los vértices del grafo y existe una arista entre dos objetos O_i y O_j si $\beta(O_i, O_j) \geq \beta_0$ o $\beta(O_j, O_i) \geq \beta_0$; donde β_0 es un número real en el intervalo $[0, 1]$ que determina el umbral mínimo de semejanza que deben tener dos objetos para ser considerados como semejantes. A partir del grafo anterior, los grupos son aquellos subconjuntos de O que satisfacen un conjunto de propiedades relativas a las semejanzas entre los objetos que constituyen cada subconjunto; estas propiedades se conocen como criterio de agrupamiento. El algoritmo LC propone dos criterios de agrupamiento que forman agrupamientos disjuntos. Uno de estos criterios forma grupos que son componentes β -conexas y el otro, grupos que son conjuntos β -compactos.

A partir de los grupos formados se construyen las propiedades (conceptos) que caracterizan a cada grupo de objetos. Se forma una ME a partir de MI y de los K_1, \dots, K_c , luego se calcula el conjunto de los testores más relevantes de ME $\tau = \{t_1, \dots, t_p\}$. Para cada clase $K_i, i = 1, \dots, c$ se calcula la estrella $G_\tau(K_i \setminus K_1, \dots, K_{i-1}, K_{i+1}, \dots, K_c)$ que estará formada por p l-complejos (tantos como testores relevantes). Para el cálculo de los $R \mid_{K_i}$ en cada $X \in t_q, t_q \in \tau$ se emplea el operador de refusión condicionada RUC. Los conceptos que caracterizan a cada agrupamiento K_i serán los l-complejos obtenidos en el paso anterior.

Anexo 4 Descripción del algoritmo Fast-BR

ALGORITMO FAST-BR

Entrada: Una matriz básica MB.

Salida: El conjunto de todos los testores típicos de MB.

1. Reducción de columnas

- a) Se seleccionan todas las columnas iguales.

b) Se forma una nueva MB, con una representación de cada columna

2. Ordenando filas y columnas en MB

a) Sea F el conjunto de filas que tienen el menor número de 1s.

b) Para cada fila $f \in F$ obtener el número de 1s en todas las columnas de MB que contengan un 1 en f . Poner la fila con mayor número como primera fila en MB. Si existe más de una fila con el mayor número, tomar cualquiera de ellas.

c) Sea $C1$ ($C0$) el conjunto de columnas con un 1 (0) en la primera fila en MB. Reordenar las columnas tal que las columnas de $C1$ estén a la izquierda y las de $C0$ estén a la derecha. Ordenar de forma descendente las columnas de $C1$ de acuerdo con el número de 1s. Las columnas en $C0$ se ordenan de la misma forma.

3. Inicialización:

a) $L = []$

b) Sea TTR la lista de testores típicos, $TTR = []$. Note que TTR es una lista de listas.

c) Sea R la lista de todos los rasgos en MB y $TL = [R]$. Note que TL es también una lista de listas.

4. Proceso:

a) Sea LR la última lista de rasgos en TL , esto es, $LR = \text{último}(TL)$.

b) Sea X el primer rasgo de LR .

c) Si $|TL| = 1$, entonces

Si la columna correspondiente a X (cX) tiene un cero en la primera fila de MB, entonces retornar ir al paso 5.

Si no, si $cX = (1, \dots, 1)$, entonces $TTR = TTR + [X]$, $LR = LR \setminus [X]$ e ir al Paso 4b.

d) $L = \text{sust}(L, X, |TL|)$

e) Eliminar el último elemento (lista) de TL , esto es, $TL = TL \setminus [\text{último}(TL)]$.

f) $LR = LR \setminus [X]$

g) $LP = \text{noExcl}(LR, L)$

h) $TTR = TTR + \text{TipL}(LR, L)$

i) Si $|LR| > 1$ entonces

$TL = TL + [LR]$

Si $|LP| > 1$ entonces

a. $TL = TL + [LP]$

Si no,

Sea X_i el único rasgo de LR.

Si $X_i \in \text{último}(TTR)$ ó $X_i \in LP$ entonces

Si existe hueco de TL entonces

$$TL = \text{ElimH}(TL)$$

j) Ir al paso 4.

5. Generación de TTS:

a) Los TTS se obtienen a partir de los pseudo-TT almacenados en TTR, basados en la reducción de columnas aplicadas en el primer paso.

Anexo 5 Medicamentos incluidos para las pruebas del sistema

Medicamentos en el Sistema		
ACIDO ACETILSALICILICO	DIPIRONA _TAB	METRONIDAZOL _TAB
ESPASMOFORTE	DIPIRONA _AMP	METRONIDAZOL _OVU
ESTREPTOQUINASA	NAPROXENO	AMINOFILINA
ALOPURINOL	DICLOFENACO DE SODIO	METOCLOPRAMIDA _TAB
CEFALEXINA _CAP	DICLOFENACO SODICO	METOCLOPRAMIDA _GOT
CEFALEXINA _SUS	INDOMETACINA	SALBUTAMOL _TAB
SULFAMETOXAZOL TRIMETOPRIM _SUS	Y METOCARBAMOL	SALBUTAMOL _JAR
CEFACLOR	BACLOFENO	SALBUTAMOL _FSC
AMOXICILINA _CAP	DIFENHIDRAMINA _TAB	SALBUTAMOL _AMP
AMPICILINA _CAP	DIFENHIDRAMINA _AMP	SALBUTAMOL _AER
AMPICILINA _POL	DIFENHIDRAMINA _JAR	TEOFILINA _SUP
AMPICILINA _BUL	KETOTIFENO _COL	HEBERPROT
TETRACICLINA _UNG	KETOTIFENO _JAR	SALAZOSULFAPIRIDINA HC
TETRACICLINA _CAP	MECLIZINA CLORHIDRATO	FUROSEMIDA
TETRACICLINA _TAB	LORATADINA	EUPARIN
AZITROMICINA _TAB	LORATADINA _JAR	CODEINA _GOT
AZITROMICINA _CAP	CAPTOPRIL	CODEINA _TAB
ERITROMICINA _TAB	ATENOLOL _AMP	DOBESILATO DE CALCIO
ERITROMICINA _BUL	ATENOLOL _TAB	KETOTIFENO _TAB
OXACILINA	ENALAPRIL	CIPROHEPTADINA
PENICILINA RAPILENTA	DIHIDRALAZINA _AMP	METILDOPA
CIPROFLOXACINA _TAB	DIHIDRALAZINA _TAB	DOMPERIDONA _TAB
CIPROFLOXACINA _BUL	LABETALOL	DOMPERIDONA _SUS
CIPROFLOXACINA _COL	NIFEDIPINA	TRIHEXIFENIDILO

PARACETAMOL _TAB	RESERPINA	METIMAZOL
PARACETAMOL _SUP	CEFAZOLINA	AZTREONAM
PARACETAMOL _GOT	CEFOTAXIMA SODICA	CLOFAZIMINA
IBUPROFENO _TAB	CEFTRIAXONA	CARVEDIOL
IBUPROFENO _AMP	CEFUROXIMA	ETIONAMIDA
IBUPROFENO _JAR	AMIKACINA	GEMCITABINA
PIROXICAM _TAB	VANCOMICINA CLORHIDRATO	MISOPROSTOL _TAB
PIROXICAM _SUP	MEROPENEM	IDARRUBICINA
DIPIRONA INFANTIL	PIPERACILINA	DEXCLORFENIRAMINA
METRONIDAZOL _FSC	PROPRANOLOL _TAB	GLIMEPIRIDA
METRONIDAZOL _JAL	PROPRANOLOL _AMP	TOLBUTAMIDA

Anexo 6 Reportes de reacciones adversas en Cuba 2000-2014

En el DF del MINSAP entre los años 2000 y 2014 se llevó un registro de todas las RAM reportadas en el país por provincia, municipio, unidad que realizó el reporte, entre otros aspectos de interés para ellos, donde se evidencia un aumento de los reportes realizados teniéndose un total de 206812 reacciones adversas en los siete años.[15]

