

Normalización de datos en registros de historia clínica electrónica y aplicación práctica de procesamiento del lenguaje natural para las instituciones de salud colombianas.

(agosto de 2022)

Mojica Mugno, Hernando Miguel.

hmojicam@unbosque.edu.co - mojicamugno.miguel@gmail.com

Universidad El Bosque
Bogotá Colombia.

RESUMEN – Las instituciones de salud en Colombia han implementado sistemas informáticos de historia clínica, en principio se fundamentaban en los formatos preimpresos usados en servicios de atención, allí se registraba la información de atenciones por parte del profesional tratante narrando de manera cualitativa los datos referentes a la atención, surge el interrogante ¿Puede llegar a ser posible utilizar esta información “no estructurada” con fines estadísticos, de interoperabilidad o investigación?

El presente artículo aborda el tema con dos ejercicios, el primero presenta un análisis sobre campos textuales en registros de diferentes etapas de la atención clínica y la posibilidad de normalizarlos para facilitar la minería de datos, el segundo es la elaboración de un prototipo de software que realiza una consulta sobre el contenido de una descripción quirúrgica y con técnicas de procesamiento de lenguaje natural obtiene información para ser comparada con una terminología válida a nivel mundial como es SNOMED CT.

PALABRAS CLAVE - Historia Clínica electrónica, interoperabilidad, Procesamiento del Lenguaje natural, Normalización, SNOMED CT.

ABSTRACT – Health institutions in Colombia have implemented computerized clinical history systems, in principle they are based on the preprinted formats used in care services, there the care information was recorded by the treating professional, qualitatively narrating the data From reference to attention, the question arises: Can it be possible to use this "unstructured" information with fine statistics, interoperability or research?

Documento recibido el XX/XX/XXX. Aprobado: XX/XX/XXXX

Tipo de Artículo: Investigación científica y tecnológica.

Afiliación institucional del autor: Universidad El Bosque, estudiante Maestría en informática biomédica.

Autor para Comunicaciones: hmojicam@unbosque.edu.co – mojicamugno.miguel@gmail.com

El autor declara que no tiene conflicto de interés.

This article deals with the subject with two exercises, the first presents an analysis of textual fields in records from different stages of clinical care and the possibility of normalizing them to facilitate data mining, the second is the development of a software prototype that makes a query about the content of a surgical description and with natural language processing techniques obtains information to be compared with a worldwide valid terminology such as SNOMED CT.

KEY WORDS - Electronic Medical Record, interoperability, Natural language processing, Normalization, SNOMED CT.

I. INTRODUCCIÓN

La historia clínica es el registro de la atención en salud en cualquiera de sus ámbitos o situaciones, la historia de este registro se remonta a las primeras descripciones de las actividades realizadas por los primeros médicos (curanderos) de la era medieval. Aunque su composición y medios de registro o almacenamiento han ido cambiando a lo largo del tiempo, sus propósitos se han mantenido desde un principio, en este sentido podemos enumerar el historial de atenciones y patologías de un paciente, la necesidad de crear un repositorio de conocimiento para la investigación en aras de mejorar las condiciones de salud de la población y brindar una atención efectiva a futuros pacientes.[1]

En la actualidad la tecnología informática ha potenciado a la Historia clínica convirtiéndose en una nueva rama del conocimiento y abriendo un abanico de posibilidades que permiten la interacción de profesionales en diferentes ubicaciones, la facilidad de generación de datos para análisis estadísticos y para proyectos de investigación, el computador se ha erigido como una herramienta para el profesional médico desde cualquier punto de vista.

II. OBJETIVOS

- A) Objetivo general: Realizar una propuesta metodológica orientada a estructurar los datos mediante criterios unificados para lograr una adecuada normalización en el registro de atenciones en las historias clínicas electrónicas para Colombia.
- B) Objetivos específicos:
- Realizar un análisis y diagnóstico de los distintos plantillas, formatos y entradas de datos e identificar las brechas para la calidad del registro en la historia clínica electrónica.
 - Proponer una metodología para estructurar los datos en modelos de registro clínico.
 - Presentar una prueba de concepto orientada al diseño de un algoritmo capaz de interpretar el contenido de un documento clínico, para ejemplo una descripción quirúrgica, que esté en capacidad de comparar los datos consignados por el médico responsable del procedimiento contra un listado terminológico válido en salud y extraer en forma de datos estructurados la información registrada.

III. PLANTEAMIENTO DEL PROBLEMA

CONTEXTO COLOMBIANO

A nivel local, desde inicios de la década 2000, ha emergido una industria del software para el registro de la historia clínica digital, incorporando nuevas funcionalidades pero con una marcada prioridad para agilizar los procesos de facturación, dejando de lado de manera desafortunada la visión frente a diferentes aspectos que se pueden optimizar frente a lo clínico, las plantillas de historia clínica electrónica guardan aun marcada similitud con los formatos preimpresos que le anteceden dejando extensos campos de texto libre para que el profesional en salud pueda incluso copiar y pegar sus hallazgos escritos desde procesadores de palabra como Word o el mismo block de notas. Esta información no estructurada difícilmente puede ser utilizada en proyectos de minería de datos o interoperabilidad, por eso se evidencia la necesidad de promover un estándar en el diseño de formatos de registro clínico que permita normalizar los datos al máximo nivel posible.[9]

Sin embargo, debe reconocerse que hay una parte de la información registrada por el médico que claramente es descriptiva, aquella parte donde el médico debe relatar en sus propias palabras un hallazgo, narrar un evento o argumentar una decisión, en esa instancia la tecnología informática debe dar un paso adelante y ofrecer alternativas que permitan utilizar estas descripciones en forma de datos estructurados y del mismo modo obtener resultados de información útiles para proyectos de investigación, comunicación de servicios de

atención e incluso la gestión de cuentas médicas.

IV. METODOLOGÍA

En este sentido el presente documento aborda dos puntos de vista en la gestión del dato registrado en Historia clínica electrónica, a saber:

- 1) *Propuesta para normalización de datos en plantillas de historia clínica electrónica. (Objetivos específicos 1 y 2)*
- 2) *Prueba de concepto sobre el desarrollo de un algoritmo que realice la lectura de una descripción quirúrgica y presente con base en un método de interpretación del lenguaje natural una lista de términos que pueda comparar con la base de datos de la terminología SNOMED CT y de este modo identificar procedimientos realizados en la atención.*

A continuación, se aborda el desarrollo de estas temáticas:

PARTE I

NORMALIZACIÓN DE DATOS EN PLANTILLAS DE HISTORIA CLÍNICA ELECTRÓNICA:

Esta parte de la experiencia busca definir las mejores alternativas para establecer los datos de captura en las diferentes plantillas de historia clínica electrónica llevando a una captura de datos eficiente, manteniendo un control sobre el tiempo invertido por el profesional de salud frente al computador mediante datos de selección en los que su volumen de datos escritos disminuya en consonancia al uso de datos tipo selección, al final el conjuntos de datos obtenido brindará una mayor eficiencia en el uso de herramientas de minería de datos, para lograr estos propósitos y cumplir los dos primeros objetivos específicos se realizaron los siguientes pasos:

- a. Paso 1: Definición de plantillas a utilizar en el marco de un ciclo de atención hospitalarios, con el fin de abarcar por lo menos 40 tipos de atención en todos los ámbitos de un hospital cuya atención vaya desde el nivel básico hasta la alta complejidad, se estableció realizar el análisis con un 0.25% del total de plantillas, es decir 10 formatos de registro clínico distribuidos en los 4 ámbitos de atención:
 - i. Atención ambulatoria: 3 plantillas.
 - ii. Atención en urgencias: 3 plantillas.
 - iii. Atención hospitalaria: 3 plantillas.
 - iv. Atención en salas de cirugía: 1 plantilla.

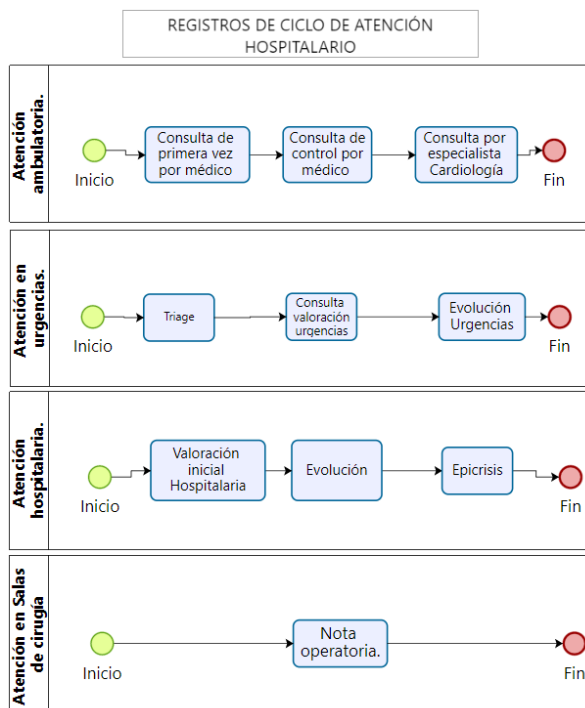


Fig. 1. Plantillas seleccionadas dentro del ciclo de atención hospitalaria, abarcan los 4 tipos de servicio tradicionales.

- b. Paso 2. Búsqueda de plantillas disponibles para consulta pública, en este sentido se presentó uno de los grandes retos del proyecto no obstante que no existe como tal un repositorio oficial estatal de plantillas de historia clínica, tampoco se ha promulgado una norma o documento público en Colombia que indique el contenido de las plantillas de registro clínico para el sector hospitalario, sin embargo algunas instituciones académicas han publicado plantillas reales de uso en el sector salud nacional, a continuación las escogidas para el desarrollo del proyecto:

- i. Corporación Universitaria Rafael Núñez (Cartagena).
- ii. Universidad del Sinú.
- iii. Universidad Simón Bolívar.

Estas plantillas se encuentran principalmente en portales de apoyo académico como: Comunidad de estudio latinoamericana UDOCZ y el Portal StuDocu.

Incluso en estos portales se encuentran datos de ejemplo de registros clínicos, como es el caso de descripciones quirúrgicas, utilizadas en la segunda parte del proyecto.

- c. Paso 3. Organización y clasificación de los

datos de historia clínica; en este sentido se aplicó uno de los conceptos base en la elaboración de plantillas de registro clínico propuesto por el Dr. Lawrence Weed (E.E. U.U) en los años sesenta que ofrece una línea de razonamiento al orden lógico de los datos de un proceso de atención en salud, ese orden es conocido como el modelo “P.S.O.A.P.”, a continuación, su significado: (P) Paciente: Hace referencia a los datos de información demográfica del paciente, incluye su ubicación si se trata de un paciente hospitalizado, la información de afiliación, género y edad son de gran importancia en el proceso de atención en cualquier tipo de servicio.

(S) Subjetivo: Los datos de esta sección son todas aquellas apreciaciones de primera mano que el médico obtiene del paciente, lo que refiere y lo que siente, por parte del médico hace referencia a sus primeras observaciones.

(O) Objetivo: Datos detallados de mayor exactitud que se obtiene de la exploración física, la toma de signos vitales, esta parte incluye la interpretación de resultados de apoyo diagnóstico.

(A) Análisis: Los diagnósticos, principal y secundarios, del mismo modo se especifican las conclusiones del profesional de salud.

(P) Plan: En esta sección se describen el tratamiento y las recomendaciones, si es paciente hospitalizado aquí se incluye el destino, el procedimiento o el tratamiento con base en medicamentos a realizar. [1]

Al realizar la selección de formatos y el mapeo de los datos, cada uno fue clasificado en las dimensiones de esta metodología, como resultado preliminar de este diagnóstico se encontró:

Plantillas analizadas: 10
Sumatoria de datos: 371
Consolidación de datos: 87
Tipos de datos encontrados
Tipo texto no estructurado: 70 (80,45%)
Númerico: 7 (8,05%)
Booleano: 5 (5,75%)
Lista de selección: 5 (5,75%)

Tabla 1: Resultados de la revisión de las plantillas utilizadas en el ejercicio.

De esta revisión inicial de plantillas se puede inferir que al tener un 80,45% de datos tipo texto no estructurado, la mayor parte de la información recogida por estas plantillas no estará atomizada y por ende su uso directo en consultas estadísticas

no será amigable, del mismo modo que en entornos de interoperabilidad viajará en alto volumen generando bajos rendimientos de operación al sistema informático.

Claramente se hace necesario un rediseño en la captura y en los tipos de datos que permita mejorar los aspectos mencionados y a la vez facilitar la labor de registro médica sin disminuir la posibilidad de reflejar toda la información necesaria desde los puntos de vista cualitativo y descriptivo.

- d. Paso 4. Codificación y tipificación de datos; Una vez se organizaron los datos de las plantillas en listados individuales, se procedió a extraer cada uno de ellos en un listado adjunto donde se tomó cada dato de manera individual, buscando crear un modelo de referencia donde cada dato a usar en las plantillas de historia clínica tuviera la posibilidad de ser usado en las plantillas que se requiera con sus mismas características, acto seguido se buscó la equivalencia de cada uno en la lista del buscador terminológico LOINC, posterior a esto y con las definiciones dadas por la terminología se indicó a cada campo cuál sería su tipo de dato apropiado sugerido, obteniendo el siguiente resultado:

Nuevos tipos de datos propuestos.
Booleano: 14 (16,13%)
Texto: 16 (18,39%)
Casillas de selección: 28 (32,18%)
Arquetipos: 29 (33,3%)

Tabla 2: Tipos de datos obtenidos tras la validación contra LOINC y el análisis de contexto.

Tras la realización de este ejercicio y el resultado obtenido se puede establecer que el uso de los datos en esta forma proveerá mayores beneficios en el posterior uso de la información obtenida, incluso mejorando el tiempo disponible para que el médico pueda realizar la consulta e interactuar con el paciente.[4]

La comparación y el mapeo propuesto optimizaría también los procesos de búsqueda y extracción de datos, la interoperabilidad e incluso la facilidad para obtener datos hacia proyectos de investigación.

NUEVOS CONCEPTOS APLICADOS DURANTE EL EJERCICIO DE CONSOLIDACIÓN DE DATOS.

En esta verificación se introducen a este proyecto dos conceptos que deben ser ampliados en este documento, el primero es la terminología LOINC (Logical Observation Identifiers Names and Codes: Vocabulario universal para identificar observaciones clínicas y de laboratorio, contiene códigos para cada prueba, medición u observación clínica

existente, es uno de los listados terminológicos más completos que existen de uso a nivel mundial, no solo permite reconocer el significado de los valores, también ofrece información aplicable a su composición, a sus mecanismos de interoperabilidad y a los datos que podría utilizar en el caso de campos cuyo valor sea recogido desde una lista de selección.

ARQUETIPOS CLÍNICOS[6]

Según el listado obtenido del ejercicio, el 33,3% de los datos reestructurados responden a la característica denominada Arquetipo, este vocablo muestra al concepto clínico como el producto de una composición de características y valores, de manera mas explícita indica cuales son los datos que deben estar incluidos en un concepto clínico dentro de una historia clínica, a manera de ejemplo; El campo descripción quirúrgica dentro de una nota operatoria debe tener unos valores obligatorios que el médico debe incluir dentro del campo de texto, el sistema idealmente debiera estar en capacidad de de identificar si uno de estos valores hiciera falta y luego de estructurarlos para un posterior uso de la información.

Con este concepto se acepta que el profesional de salud debe tener la posibilidad de ingresar datos descriptivos con el fin que pueda, en lenguaje humano, informar sobre un procedimiento realizado o un hallazgo, sin embargo se mantiene latente el interrogante sobre la manera adecuada en que el sistema informático facilite trabajar con estos datos en ambientes de exploración, en este sentido el presente proyecto aborda un ejercicio que busca llegar al tratamiento normalizado y estructurado de los datos clínicos aun si provienen de una larga descripción.

METODOLOGÍA PROPUESTA PARA ESTRUCTURAR LOS DATOS EN MODELOS DE REGISTRO CLÍNICO

En este sentido es claro que la búsqueda de un manejo ideal para dar solución a todos los aspectos identificados tiene que propender al uso de estándares de interoperabilidad semántica válidos a nivel internacional para buscar que el sistema de registro clínico (EHR – Electronic Health Record) pueda hablar y reconocer el mismo idioma de cualquier interlocutor a nivel organizacional, nacional e internacional, en este sentido son múltiples las fuentes conceptuales que pueden apoyar esta elaboración de plantillas, a continuación se presentan aquellas que fueron objeto del componente de investigación de este proyecto:

- a) OpenEHR: Nacida en 2007, es un estándar de uso libre que especifica la manera de organizar los datos de una historia clínica, tiene una clara representación de arquetipos que facilita su comprensión, de hecho propone su propia sintaxis bajo la denominación ADL (Archetype Definition Language). Cuenta con una colección de cursos en su pagina web, no son gratuitos estos cursos. A nivel organizacional tiene su sede en el Reino Unido.

OpenEHR cuenta con implementaciones en Brasil, Holanda, Noruega, Australia, Filipinas, Portugal, Rusia, Eslovenia, Suecia y el Reino Unido. Un ejemplo de utilización en Latinoamérica es el caso de Brasil que desde 2010 reglamentó a nivel estatal su uso en la implantación de los sistemas de información con miras a su interoperabilidad a nivel nacional.[1]

El estándar OpenEHR no ha sido adoptado en Colombia a la fecha de elaboración del presente artículo.[12]

b) ISO 13606: Norma técnica creada por el comité europeo de normalización, aprobada por la ISO (International Organization for Standardization), también concebida para facilitar interoperabilidad de sistemas de información clínicos, propone un modelo de arquetipos y una lista de términos, todos enmarcados dentro de su propio modelo de referencia. Permite el uso de LOINC y CIE10, entre otros listados terminológicos. Un ejemplo de su utilización en Latinoamérica es su adopción como NTE (Norma técnica ecuatoriana) codificada INEN-ISO 13606-1, es interesante realizar su revisión y encontrar el valioso aporte que puede significar su revisión dado que está expresada de una manera sencilla con el fin de facilitar su implementación.

No ha sido adoptada en el sistema de salud colombiano.[12]

c) HL7 V3 CDA RIM (Modelo de referencia de información), provee una estructura de datos que permite la identificación de todos los aspectos clínicos, incluso administrativos, de amplia difusión a nivel mundial, presenta la composición de registros clínicos en los denominados CDA (Clinical document arquitectura), permite la integración con diferentes listados terminológicos como SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms) y LOINC, este estándar ha sido utilizado en Colombia en proyectos como el sistema Bogotá Salud Digital implementado por la Secretaría distrital de Salud de Bogotá en 2019. Otro ejemplo de uso es el sistema Uruguayo Salud.uy donde hace parte de su CMD (Conjunto mínimo de datos).

d) FHIR HL7 – C CDA: (Fast Healthcare Interoperability Resources), es la versión más reciente y evolucionada de HL7 como estándar de interoperabilidad semántica, presenta un detalle minucioso de la estructura de campos de la Historia clínica y registros complementarios, a lo que se suma su propuesta de estructuras de datos compuestas, es decir, muestra el dato y propone en qué parte del registro debe estar ubicado.[2]

En Colombia se ha comenzado su implementación en proyectos como el sistema de Interoperabilidad clínica nacional resultante de la aplicación de la Ley 2015 de 2020 y la resolución 866 de 2021. Otro ejemplo de su uso a nivel mundial es su adopción por parte de los centros de Servicios de Medicare y Medicaid en EE UU. [9][10]

En conclusión, de lo anterior, la entidad de salud que implementa un sistema informático clínico debe priorizar el adherir al estándar que más le convenga según su localía y los entornos en que vaya a interactuar, no siendo prioridad la herramienta tecnológica o la tendencia del apoyo técnico adscrito.

Antes de llevar a cabo cualquier proceso de implementación este aspecto debe ser objeto de análisis y planeación por la entidad que adquiere o desarrolla sistema de información EHR.

PARTE II

PRUEBA DE CONCEPTO SOBRE EL DESARROLLO DE UN ALGORITMO QUE REALICE LA LECTURA DE UNA DESCRIPCIÓN QUIRÚRGICA Y PRESENTE CON BASE EN UN MÉTODO DE INTERPRETACIÓN DEL LENGUAJE NATURAL;

En esta sección se aborda un ejercicio que busca tomar la información registrada en una nota operatoria y compararla con una lista de términos clínicos válidos para inferir que procedimientos y actividades realizó el médico según su información registrada.

Esta etapa del proyecto se presenta en los siguientes pasos:

- a. Paso 1: Inscripción a Snomed-CT; dicha inscripción se puede realizar en las páginas Snomed de cualquiera de los países latinoamericanos inscritos, en nuestro caso se realizó como estudiante, sin obstáculo ni costo para ser admitido.

Estando allí se pudo acceder a las últimas versiones de Snomed Ct. El propósito de permitir la descarga de estos archivos es poder ser incluidos en sistemas E.H.R. (Electronic Health Record) para facilitar la codificación de términos desde los mismos aplicativos, dentro de la página se encuentran las más recientes.

Se ofrecen los paquetes en dos modalidades:

- Paquete de extensión: Contiene los datos de las últimas actualizaciones, su descarga es de menor peso, resulta ideal para actualizaciones.
- Paquete de edición: Contiene los archivos completos incluyendo la versión extensa en Español de mayor tamaño y completitud, ideal para carga inicial y utilizada en el presente proyecto. [8]

El paquete Snomed CT tiene diversos archivos en diferentes carpetas contenedoras, para lograr su comprensión, dentro de las actividades del proyecto, se accedió a curso publicado en youtube por parte del canal N.H.S. Digital. N.H.S. es la sigla que identifica al National Health Service de Inglaterra, es decir al servicio nacional de salud del

Reino Unido, creadores de SNOMED CT.[5]

SNOMED CT

Es un acrónimo que significa Systematized Nomenclature of Medicine – Clinical Terms, es un compendio de términos que se usa con el fin de codificar y en general representar información de una manera organizada para facilitar los procesos de análisis y comunicación de los registros médicos, su uso y estandarización es de nivel mundial.[2]

Los archivos de los paquetes Snomed Ct vienen separados en tres grupos:

FULL: Dentro de este contenedor están los archivos de la versión completa, allí vienen todos los términos acuñados desde 2002 (Nacimiento de Snomed Ct), se utiliza este paquete cuando se carga por primera vez en un sistema de información.

DELTA: En este contenedor se presentan únicamente los términos registrados y o modificados entre la penúltima y la última versión.

SNAPSHOT: Aquí se encuentran los archivos que contienen únicamente la versión más reciente.

Los archivos Delta y Snapshot se utilizan únicamente para actualizar parametrizaciones existentes según necesidad.

Dentro de cada uno de estos paquetes se encuentran diferentes archivos cuya nomenclatura también responde a una clasificación, a saber:

CONCEPT: Contiene los conceptos clínicos que fundamentan la terminología.

DESCRIPTION: Contiene las descripciones que se asocian a cada concepto.

RELATIONSHIPS: Contiene las relaciones entre los conceptos y las descripciones. [5]

b. Paso 2: Creación de la base de datos.

En esta parte del ejercicio práctico se utilizó el motor de base de datos MYSQL en su versión para instalación local, aunque este motor ahora es un producto adscrito a Oracle, continúa siendo de distribución libre, ahora con una nueva interfaz más parecida a las versiones de Oracle.

En dicha base de datos denominada “lector_descripción” se agregaron las tablas base de la estructura Snomed, cargada inicialmente con la versión Full, y una tabla adicional para contener datos de descripciones quirúrgicas objeto del ejercicio.

Cabe mencionar que este proceso de carga solo puso realizarse en la línea de comandos mediante la ejecución de la instrucción LOAD INTO. Los archivos planos previamente debieron ser ajustados a separación por comas y extensión. Csv.

A continuación, los tamaños en registros de cada uno de los repositorios subidos:

- Conceptos: 699.250 registros.
- Descripciones: 5.164.592 registros.
- Relaciones: 6.425.049 registros.

Los elevados tamaños de estas tablas incidieron en el rendimiento del motor haciendo que el desarrollo del ejercicio se diera en tiempos más extensos, sin embargo, se agregaron algunos índices a las tablas para mejorar en cierta medida el tiempo de ejecución.

A la construcción de la base de datos se añadieron las tablas: Descripción quirúrgica, que contiene datos de las notas operatorias utilizadas para el desarrollo del ejercicio, la tabla Códigos_Cups, en la cual se cargó la información del anexo de la resolución 2077 de 2021, con el fin de obtener la identificación de procedimientos quirúrgicos acorde a la terminología válida para Colombia.[11]

De igual manera se crearon tres vistas de apoyo, estas vistas son consultas complejas que se elaboran dentro del motor de base de datos y su ejecución se da en dicha instancia, facilitan su utilización en la aplicación pues reducen la necesidad de detallar su estructura y simplemente pueden ser ejecutadas llamando a su denominación dentro de una consulta SQL.

- New_view: Presenta de manera depurada los datos de la resolución RIPS.
- Procsnomed: Presenta la combinación de datos de las tablas Conceptos de SNOMED y Descripciones de SNOMED.
- Vista_expresiones_snomed: Presenta el diagrama de expresiones precoordinaadas y poscoordinaadas de un procedimiento seleccionado.

Una tendencia aplicada en el presente proyecto fue buscar que el desarrollo referido a los datos almacenados en las tablas fuera llevado a cabo mediante instrucciones SQL almacenadas en estas vistas, lo cual es de gran ayuda para simplificar las estructuras de código de la aplicación de usuario final.

```

1 CREATE
2 ALGORITHM = UNDEFINED
3 DEFINER = 'root'@'localhost'
4 SQL SECURITY DEFINER
5 VIEW `vista_expresiones_snomed` AS
6 SELECT DISTINCT
7   `relaciones_snomed`.`sourceId` AS `PROCEDIMIENTO`,
8   IF((`relaciones_snomed`.`relationshipGroup` < 1),
9     'Expresión precoordinada') AS TIPO,
10  `relaciones_snomed`.`destinationId`,
11  IF((`relaciones_snomed`.`relationshipGroup` < 1),
12    `relaciones_snomed`.`destinationId`,
13    ' ') AS PRECOORDINADA,
14  IF((`relaciones_snomed`.`relationshipGroup` = 1),
15    `relaciones_snomed`.`typeId`,
16    ' ') AS POSCOORDINADA,
17  IF((`relaciones_snomed`.`relationshipGroup` = 1),
18    `relaciones_snomed`.`destinationId`,
19    ' ') AS VALOR
20 FROM
21   (`relaciones_snomed`
22  JOIN `descripciones_snomed` ON ((`relaciones_snomed`.`destinationId` = d
23 WHERE
24   ((`relaciones_snomed`.`active` = 1)
25    AND (`descripciones_snomed`.`languageCode` = 'es'))

```

Fig. 2. Código SQL que presenta diagrama de expresiones precoordinadas y poscoordinadas SNOMED en la aplicación.

- c. Paso 3: Obtención de notas operatorias para prueba. Para esta actividad se recurrió al portal StuDocu donde se encuentran registros clínicos publicados con fines académicos, los contenidos utilizados fueron precargados en la base de datos elaborada denominada Descripción_quirúrgica, uno de los campos establece el identificador del procedimiento, allí se incluyó el respectivo código cups para la cirugía realizada.
- d. Paso 4: Elaboración del algoritmo para interpretación y comparación.

Este es el componente de cara al manejo de los datos y obtención de resultados del proyecto, para esta parte se utilizó el entorno de programación Python, muy utilizado en la actualidad por ser de distribución libre, facilidad de uso y adaptabilidad al uso de librerías para realizar ejercicios de Machine learning y manejo de grandes volúmenes de datos, en el ejercicio se utilizó dentro del marco Visual Studio Core de Microsoft cuya distribución es libre. [3]

En el inicio de esta parte del ejercicio se realizó la correspondiente carga de base de datos en el código mediante el uso de la librería MySQLdb, diseñada para facilitar la conexión a bases de datos sobre el motor MySQL.[3]

```

##Conexión a base de datos
import MySQLdb

miConexion = MySQLdb.connect (
host='localhost', user = 'root', passwd =
'W3syntac', db='lector_descripcion')
##cur = miConexion.cursor()

```

```

cur = miConexion.cursor()
cur.execute ("Select texto_desci from
descripcion_quirurgica")

```

Fig. 3. Código Python para carga de base de datos MySQL en la aplicación. [3]

De igual manera se utilizó la librería NLTK, muy utilizada en el ámbito del procesamiento de texto en el procesamiento de lenguaje natural.

```

##Procesamiento descripción quirúrgica
##Conexión a base de datos
import MySQLdb
import nltk ##Librería para procesamiento de lenguaje natural
import os #
os.system('cls')##Instrucción para borrar pantalla de ejecuciones anteriores
admisión= int (input("Inserte número de admisión: "))
print("El número de admisión es :", admisión)
miConexion = MySQLdb.connect ( host='localhost', user = 'root', passwd = 'Viv12824562', db='lector_descripcion')
cur = miConexion.cursor()
cur2 = miConexion.cursor()
cur2.execute ("Select CODIGOPRO, NOMBREPRO from new_view where admisión = " + str(admisión))

```

Fig. 4. Inserción de la librería NLTK en la aplicación Python.

- e. Paso 5: Establecimiento del método para realizar minería de texto mediante un algoritmo para procesamiento del lenguaje natural.

Aquí se establece la manera en la cual el algoritmo realizará el recorrido por el texto de la descripción quirúrgica, para esto se utiliza una serie de técnicas de uso común ordenadas en pasos que facilita este proceso, a saber:

- I. CASE NORMALIZATION: Paso de letras mayúsculas a minúsculas.
- II. PUNCTUATION ERASURE: Eliminar signos de puntuación.
- III. STOP WORDS: Eliminar palabras vacías.
- IV. TOKENIZACIÓN: Separar palabras para encontrar significados diferentes.
- V. FILTRADO Y CONTEO: Con las palabras no vacías separadas se realiza un conteo y se representan en orden de mayor a menor repetición.

En este ejercicio se logra obtener un listado de las palabras que más se repiten por parte del médico en el registro, de gran utilidad para procesos de auditoría clínica y auditoría de cuentas médicas, siendo este el primer producto de este ejercicio.

```

#PIL PASO 1 CASE NORMALIZATION: Convertir todo el texto de la descripción en minúscula.
print ("\nCONVERSION DE LETRAS A MINUSCULA\n")
desq = str(desq)
desq = desq.lower()
print (desq)

```

Fig. 5. Conversión a letras minúsculas de todo el texto de la descripción.

```

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import string
stop_words= set (stopwords.words('spanish'))
word_tokens = word_tokenize (desq)
word_tokens = list(filter(lambda token: token not in string.punctuation, word_tokens))

```

Fig. 6. En esta sección se dan las especificaciones a la librería nltk para eliminar signos de puntuación, palabras vacías y separar las palabras no vacías (Tokenización).


```
for palabra in word_tokens:
    if palabra not in stop_words:
        filtro.append(palabra)
print(word_tokens)
print('\nFILTRADO Y CONTEO DE LAS PALABRAS QUE MÁS SE REPITEN EN LA DESCRIPCIÓN QUIRÚRGICA\n')
from collections import Counter
c=Counter(filtro)
print(c.most_common(20))
```

Fig. 7. En este segmento el programa separa las palabras no vacías y acto seguido realiza el conteo por repetición, la función mostrará las 20 primeras.

Tras la ejecución del código Python establecido para la primera parte del ejercicio se obtienen los siguientes resultados:

```
CODIGO CUPS Y NOMBRE DEL PROCEDIMIENTO ('471102', 'APENDICECTOMÍA VÍA ABIERTA')

DESCRIPCIÓN QUIRÚRGICA REGISTRADA ORIGINALMENTE:
('PACIENTE EN DECÚBITO SUPINO PROTOCOLO DE CIRUGIA SEGURA ASEPSIA Y ANTISEPSIA
COLOCACIÓN DE CAMPOS QUIRÚRGICOS ESTÉRILES. DIÉRESIS 1 DESCRITA PROFUNDIZACIÓN
POR PLANOS HASTA VIDA ABDOMINAL HALLAZGO
S DESCRITOS IDENTIFICACION MESOY ARTERIA APÉNDICE, LIGADURA DE LOS MISMOS POLIGL
ACTINA 2/0 Y SECCION DIST IDENTIFICACION BASE APENDICULAR, COLOCACION PUNTO TRAN
SECTIVO ESTA POLIGLACTINA 2/0 Y DIST SE
CCIÓN SECADO DE CAVIDA CON MONTEADAS REVISIÓN DE HEMOSTASIA SE CONFIRMA MATERIAL
BLANCO Y QUIRÚRGICO COMPLETO CIERRE DE PUN ROJO ABDOMINAL CON POLIGLACTINA1/0 C
IERRE POR PLANOS COLOCACIÓN DE APÓSITO S
SOBRE HERIDA QUIRÚRGICA',)
```

Fig. 8. Texto original escrito por el médico cirujano.

```
CONVERSION DE LETRAS A MINUSCULA
('paciente en decubito supino protocolo de cirugía segura asepsia y antisepsia
colocación de campos quirúrgicos estériles. diéresis 1 descrita profundización p
or planos hasta vida abdominal hallazgos
desritos identificación mesoy arteria apéndice, ligadura de los mismos poligla
ctina 2/0 y seccion dist identificación base apendicular, colocacion punto trans
sectivo esta poliglactina 2/0 y dist seccion secado de cavida con monteadas revisión de hemostasia se confirma material
blanco y quirúrgico completo cierre de pun rojo abdominal con poliglactina1/0 ci
erre por planos colocación de apósito s
obre herida quirúrgica',)
```

Fig. 9. Conversión de texto a letras minúsculas.

```
EXTRACCIÓN DE PALABRAS Y TOKENIZACIÓN
['paciente', 'en', 'decubito', 'supino', 'protocolo', 'de', 'cirugia', 'segura',
'asepsia', 'y', 'antisepsia', 'colocación', 'de', 'campos', 'quirúrgicos', 'e
stériles', 'diéresis', '1', 'descrita',
'profundización', 'por', 'planos', 'hasta', 'vida', 'abdominal', 'hallazgos', 'd
escritos', 'identificación', 'mesoy', 'arteria', 'apéndice', 'ligadura', 'de', '
los', 'mismos', 'poliglactina', '2/0',
'y', 'seccion', 'dist', 'identificación', 'base', 'apendicular', 'colocacion', 'p
unto', 'transsectivo', 'esta', 'poliglactina', '2/0', 'y', 'dist', 'sección', 's
ecado', 'de', 'cavida', 'con', 'monteada
```

Fig. 10. Extracción de palabras vacías y separación de las no vacías.

```
FILTRADO Y CONTEO DE LAS PALABRAS QUE MÁS SE REPITEN EN LA DESCRIPCIÓN QUIRÚRGIC
A
[( 'colocación', 2), ('planos', 2), ('abdominal', 2), ('identificación', 2), ('po
liglactina', 2), ('2/0', 2), ('dist', 2), ('cierre', 2), ('paciente', 1), ('dec
ubito', 1), ('supino', 1), ('protocolo',
1), ('cirugía', 1), ('segura', 1), ('asepsia', 1), ('antisepsia', 1), ('campos
', 1), ('quirúrgicos', 1), ('estériles', 1), ('diéresis', 1)]
```

Fig. 11. Filtrado y conteo de palabras por el número de veces que se repiten.

f. Paso 6: Utilización del modelo SNOMED para apoyar el conocimiento clínico en el ejercicio.

En esta parte el análisis realizado permitió establecer que la mayor utilidad estaría en la búsqueda del nombre del procedimiento realizado dentro del listado de conceptos SNOMED-CT, para el ejemplo registrado en el presente artículo, la actividad seleccionada con el código CUPS 471102 denominada Apendicectomía Vía Abierta, arrojó los siguientes resultados:

```
('174048001', 'apendicectomía endoscópica no clasificada en otra parte (procedi
miento)')
('235313004', 'apendicectomía no urgente (procedimiento)')
('235313004', 'apendicectomía no de emergencia (procedimiento)')
('235314005', 'apendicectomía por inversión (procedimiento)')
('303400009', 'apendicectomía no clasificada en otra parte (procedimiento)')
('307581005', 'apendicectomía de intervalo laparoscópica (procedimiento)')
('708876004', 'apendicectomía laparoscópica con asistencia robótica (procedimien
to)')
```

Fig. 12. Listado de alternativas resultantes de la comparación de la denominación CUPS del procedimiento analizado con la combinación de las tablas de Conceptos y descripciones de SNOMED CT.

En la nomenclatura de descripciones de SNOMED se encuentra el campo TERM que contiene los textos de la terminología, una de las formas de su desarrollo establece que para identificar algunos tipos de término se utiliza su identificador al final entre paréntesis, de esta manera fue sencilla la elaboración de la consulta que obtiene los procedimientos agregando de manera simple el filtro que al final presentara únicamente aquellos que concluyeran con “%(procedimiento).”

Dado que el ejercicio realizado demostró que en la mayoría de los casos los listados resultantes de esta combinación generaron amplios abanicos de opciones, se agregó a la aplicación la posibilidad que el usuario final registrara el código SNOMED de la actividad con mayor aplicabilidad al procedimiento realizado, de esta manera un componente adicional de la aplicación realizada presenta la consulta que muestra el diagrama de conceptos y dependencias según la terminología SNOMED CT para el procedimiento seleccionado, así se presenta dicho diagrama en modo texto, las siguientes gráficas (13 y 14) muestran la comparación de la diagramación del aplicativo realizado frente a la presentada por el visor de SNOMED CT Internacional.

```
('80146002', 'Parents - Jerarquías superiores...', '27010001', ' ', ' ')
('80146002', 'Parents - Jerarquías superiores...', '8613002', ' ', ' ')
('80146002', 'Atributo.....', '405813007', '66754008')
('80146002', 'Atributo.....', '260686004', '129304002')
```

Fig. 13. Visualización del diagrama de Jerarquías superiores y atributos de SNOMED CT, generado por el aplicativo realizado para el presente proyecto.

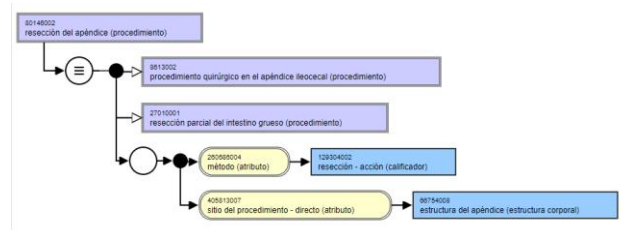


Fig. 14. Visualización del diagrama de Jerarquías superiores y atributos de SNOMED CT, generada dentro del visor SNOMED Internacional.

Como se puede apreciar el resultado obtenido presenta los mismos códigos resultantes de la búsqueda de la

Apendicectomía cuyo código en la terminología es el 80146002.

La realización del código que permite visualizar el diagrama se realizó del lado de la base de datos en la codificación de vistas que son invocadas desde el código del programa python.

En este sentido es de importancia agregar en el análisis conceptual del proyecto realizado los siguientes términos:

Expresiones precoordinadas: Es conocida como una técnica inmersa en la terminología que presenta un nivel de detalle avanzado frente al significado y composición de un término seleccionado presentando ideas que deben ser recordadas por el médico que registra el dato en el sistema E.H.R. [7]

Expresiones poscoordinadas: En complemento al concepto anterior, las expresiones poscoordinadas ofrecen dos o más alternativas de selección al profesional que realiza un registro, por ejemplo para indicar el posible lugar del cuerpo en que se realiza un procedimiento o la manera que se realiza la actividad, entre muchas alternativas posibles.[7]

CONCLUSIONES

A nivel general el ejercicio realizado compuesto de dos secciones diferentes pero complementarias deja las siguientes conclusiones que permiten establecer propuestas prácticas en procesos de sistematización clínica a implementar en Colombia, estas son:

Para la primera parte del ejercicio consistente en proponer la adecuada definición de los tipos de datos en plantillas de historia clínica electrónica:

1. Es importante, en aras de facilitar el proceso de consulta que los campos dentro de la plantilla estén dispuestos en el orden propuesto por el método PSOAP
2. Aquellos sistemas de información EHR (Electronic Health Record) que permitan personalizar las plantillas de registro de forma específica para cada tipo de atención, darán una mayor facilidad al profesional que realiza el registro pues mantendrán el orden lógico de la consulta y además solo pedirán el registro de datos específicos en el momento que sea necesario.
3. Existen datos tipo descripción pero que por conveniencia debieran consultar primero dentro de una lista de opciones, ejemplo: Si/No y luego la descripción de lo visualizado, esto facilitará en el proceso de minería de datos a filtrar condiciones y poblaciones específicas.
4. Datos de múltiple utilización en diferentes formatos, siendo el mismo dato, deben tener la misma estructura en las diferentes plantillas, por ejemplo; el diagnóstico principal si es una lista de selección, deberá ser del mismo modo en todas las plantillas donde sea utilizado.
5. La creación de plantillas de historia clínica digital debe ser estratégica, se debe considerar la cantidad de datos y el tiempo requerido para diligenciarlos versus el tiempo definido

para la atención, en este sentido es importante tener en cuenta que el profesional de la salud requiere tiempo para interactuar con el paciente, si un dato es de mayor importancia en una fase posterior de atención debe ser registrado en dicha fase.

Para la segunda parte, consistente en cruzar la información referida por el médico en una descripción quirúrgica con la lista de términos Snomed Ct.

1. La importancia que Colombia considere desde el ámbito gubernamental incursionar en una afiliación nacional a Snomed y su correspondiente adopción, seguida de procesos formativos, en aras de aprovechar estas tecnologías para hacer más útil el registro clínico, no solo con fines de favorecer procesos de cobro, lo que por defecto debe evidenciarse, sino favorecer el desempeño clínico apoyado tecnologías informáticas de alto nivel.
2. La utilización de herramientas de procesamiento del lenguaje natural puede proveer al proceso de atención guías que permitan crear alertas de mayor sofisticación en el proceso de atención, en el sentido más avanzado incluso pueden proveer herramientas de apoyo en la toma de decisiones clínicas (HATD), naturalmente siendo objeto de espacios de prueba que garanticen su adecuada inclusión en el quehacer médico y entendiendo que siempre las decisiones finales deben ser tomadas por el médico tratante, especialista o general y que los resultados de un algoritmo son básicamente un apoyo.
3. El ejercicio demuestra en una de las muchas formas posibles que puede obtenerse información estructurada de textos no estructurados, lo cual es apoyo para la analítica de datos con todas las ventajas y beneficios que se pueden derivar de su uso.

RECONOCIMIENTO

Los agradecimientos del autor para las siguientes personas por sus aportes conceptuales y de apoyo a la investigación presentada; Jefe Viviana Andrea Gómez Torres – Jefe de la Unidad Funcional de Calidad Fundación Instituto Roosevelt (Bogotá), por sus aportes en la comprensión de los datos clínicos y la estructuración del ciclo de atención presentado, Ingeniero Fernando A. Portilla V. Digital Health Senior Advisor and Health Interoperability standards specialist y Docente Universidad El Bosque, quien ha sido el tutor del proyecto y ha orientado la metodología y conclusiones con base en su amplia experiencia en el tema, Del mismo modo a la Doctora Lucía Mora Quiñonez, médica Jefe de la oficina de Gestión del conocimiento en la Subred Integrada de servicios de Salud Sur Occidente por sus aportes en métodos de investigación y la estructura para elaborar el presente artículo.

REFERENCIAS BIBLIOGRÁFICAS

- [1] D. Luna, C. Otero, F. Piazzotta y F. Campos, *Sistemas De Información Para La Salud*. Departamento de Informática en Salud – Hospital Italiano de Buenos Aires Argentina, Fernán González Bernaldo de Quirós, 2018.
- [2] F. Campos, D. Kaminker y C. Otero, *Principios de Interoperabilidad en Salud y Estándares*. Ciudad Autónoma de Buenos Aires: Sociedad Italiana de Beneficencia en Buenos Aires, Daniel Luna, Fernán González Bernaldo de Quirós, 2018.
- [3] R. Rodríguez Guerrero, C.A. Vanegas y G.A. Castang Montiel, *Phyton a su alcance*, Bogotá – Colombia, Editorial UD Universidad Distrital Francisco José de Caldas, 2020.
- [4] A. Lopez Osornio, S.D. Montenegro, S. García Martí y L. Toselli, “Codificación múltiple de una lista de problemas utilizando la CIAP-2, CIE-10 y SNOMED CT”, [online] Researchgate.net, 2004, 266040780.
- [5] Canal NHS Digital, NHS Digital, Whats SNOMED CT?, 14 jun 2021, youtube, <https://www.youtube.com/playlist?list=PLNZyoy6rRz4jirV6miIqwlRb42h0nbQKt>, 17 abr 2022.
- [6] J.C. Castrillón Betancur, J.F. Flórez Arango, “Arquetipos, terminologías e interoperabilidad”, [online], Revista Cubana de Investigaciones Biomédica, 2015, <http://scielo.sld.cu>.
- [7] J.A. Miñaro Giménez, C. Martínez Costa, P. López García y S. Schulz, “Building SNOMED CT Post_Coordinated Expressions from Annotation Groups”, [online], National Library of Medicine, 2017, <https://pubmed.ncbi.nlm.nih.gov/28423832>.
- [8] SNOMED International, “Guía de Introducción a SNOMED CT”, [online], International Health Terminology Standards Development Organisation, 2018, <http://snomed.org/doc>.
- [9] Congreso de la República de Colombia, “Ley 2015 2020”, [online], Ministerio de Salud y Protección Social de la República de Colombia, 2020, https://www.minsalud.gov.co/Normatividad_Nuevo/Forms/DispForm.aspx?ID=5897
- [10] Ministerio de salud y Ministerio de tecnologías de la Información y las Comunicaciones de la República de Colombia, “Resolución 866 de 2021”, [online], Ministerio de Salud y Protección Social de la República de Colombia, 2021, chrome-extension://efaidnbmnnnibpajpcglclefindmkaj/https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DII/resolucion-866-de-2021.pdf
- [11] Ministerio de salud y Protección Social de la República de Colombia, “Resolución 2077 de 2021”, [online], Ministerio de Salud y Protección Social de la República de Colombia, 2021, chrome-extension://efaidnbmnnnibpajpcglclefindmkaj/https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DII/resolucion-2077-de-2021.pdf.
- [12] F.A. Zambrano Villar, J.C. García Ojeda, “Implementación del estándar CDA HL7 para la consulta de la historia clínica universitaria. Caso de Estudio: Medicina general de la Universidad Industrial de Santander”, Universidad Autónoma de Bucaramanga, 2014, chrome-extension://efaidnbmnnnibpajpcglclefindmkaj/https://repository.unab.edu.co/bitstream/handle/20.500.12749/3352/2014_Articulo_Fernando_Andres_Zambrano_Villar_1.pdf?sequence=2&isAllowed=y.

Biografía Autor: Hernando Miguel Mojica Mugno, Ingeniero en redes de computadores egresado Universidad Distrital Francisco José de Caldas, especialista en Gerencia de tecnología de la Universidad Autónoma de Colombia y Candidato a Magister en Informática Biomédica de la Universidad El Bosque. Correo electrónico hmojicam@unbosque.edu.co y mojicamugno.miguel@gmail.com.