

Implementación de servicios terminológicos en una red de atención ambulatoria

Claudio Torres Casanelli ^a, Hernán Navas ^b, Sonia Benítez ^b, Leandro Biaggini ^a, Gerardo Morales ^a,
Pelayo Navarro ^a, Daniel Luna ^b, Fernán González B. de Quirós ^b, Marcelo Maira ^c

^a Departamento de Informática Médica, Megasalud, Chile

^b Área de Informática Médica, Hospital Italiano de Buenos Aires, Argentina

^c Gerente General de Megasalud, Chile

Resumen

El almacenamiento de texto en forma de lenguaje natural, es poco útil al momento de recuperar y analizar la información en él contenida, a pesar de ser el medio más expresivo y sencillo de utilizar para ingresar la información. Por el contrario, el texto codificado ofrece una mayor capacidad para recuperar y analizar la información.

Megasalud posee una aplicación electrónica para el registro de la actividad médica llamada SIAP-Win, que utiliza desde hace 10 años. El texto se almacenó como texto libre.

La Nueva Ficha Clínica Electrónica que está desarrollando contempla la inclusión de codificación de las patologías del paciente en una lista de problemas.

Se decidió utilizar un Servidor de Terminología en español para procesar y conducir la migración de los datos existentes.

Se procesó una muestra de 14.120.752 términos que representó el 80% de los textos más ingresados por los médicos.

El Servidor de Terminología detectó el 79% de los textos de la muestra (11.099.274), pudiendo asignar automáticamente un código al 74% (10.373.276), el restante 5% (725.998) fue detectado como no válido.

Palabras Claves:

Snomed-CT, Servidor de terminología, lista de problemas.

Introducción

Siendo el lenguaje natural el medio más expresivo de la información médica, es también el que menor utilidad posee. Si el ingreso de los datos es completamente libre, la posibilidad de errores por omisión, inconsistencia, ambigüedad, etc. es elevada [1]. Como aspecto favorable, el ingreso de texto libre es el más sencillo e intuitivo. Estas características se encuentran tanto en un medio físico, como el papel, como en un medio electrónico. La información de las notas clínicas u otros reportes médicos electrónicos permanece inaccesible a la mayoría de los métodos de análisis y recuperación de datos basados en computadoras. De esa manera la utilidad clínica de un texto permanece limitada por el hecho que mucha información de gran valor se pierde por la dificultad de procesar la parte narrativa [2].

El texto estructurado tiene una lógica completamente inversa. Posee una mayor dificultad en el ingreso, pero permite una rápida recuperación y análisis posterior de la información. Es

por ello que la necesidad de vocabularios controlados en los sistemas médicos informatizados es ampliamente reconocida [3,4]. Aun los sistemas que tratan con textos narrativos e imágenes tienen la capacidad de codificar sus datos con vocabularios controlados [3].

Esta es una realidad a la que se enfrentan los sistemas actuales tales como los sistemas de Megasalud; red de salud de atención ambulatoria perteneciente a la Cámara Chilena de la Construcción y con cobertura en todo el territorio nacional chileno. Con 31 Centros médico-dentales distribuidos en todo el país, cuenta con alrededor de 1.900 profesionales médicos y 850 odontólogos, realizando 4 millones de prestaciones ambulatorias médicas y 2 millones de prestaciones dentales por año. Esta red posee un sistema Capitado denominado "Plan Mi Doctor" centrado en el primer nivel de atención [5].

Megasalud cuenta con un dos registros electrónicos, uno para el área médica llamado "SIAP-Win"; y otro para el área dental posee una aplicación diferente llamada "Atenea". El primero es una aplicación comercial, en cambio el segundo es un desarrollo propio.

La aplicación médica almacena todos los datos referentes a la consulta médica y a la salud del paciente, tomados a partir del ingreso mediante texto libre por parte del profesional que lo asiste.

En el año 2007, Megasalud, decidió dejar de utilizar el Registro Clínico SIAP-Win para construir su propia Ficha Clínica Médica con nuevas opciones que permitan el cuidado longitudinal de los pacientes atendidos en la red, integrando diferentes fuentes de información y proveyendo soporte para la toma de decisiones en el punto de atención. La construcción de la Nueva Ficha Clínica Electrónica (NFCE) fue bautizada como proyecto SIRIS (Sistema de Información por una Red Integral de Salud) integrando la información clínica histórica del registro de SIAP-Win, y adaptando los sistemas administrativos heredados al nuevo Sistema de Información en Salud.

El modelo para el desarrollo in-house del nuevo registro clínico es el de una Historia Clínica Orientada a Problemas, con un repositorio de datos clínicos, centrado en el paciente unívoca y correctamente identificado.

Dado que la aplicación "SIAP-Win" tiene aproximadamente 10 años de uso, el proyecto prevé la migración de toda la información médica en ella contenida. Esta aplicación posee varios apartados donde el médico puede volcar la información

relacionada a la consulta. Su ingreso es mediante la escritura en texto libre en la mayoría de los casos.

Por lo explicado anteriormente, de la utilidad del texto libre vs. el texto estructurado, se decidió utilizar un Servidor de Terminología, pieza clave en el manejo de vocabularios médicos [6], para procesar la información médica y conducir la migración de datos desde el registro clínico existente al nuevo repositorio de datos clínicos, orientado a problemas. Este proceso permitirá asignar un código único a los textos que sean reconocidos. Con lo cual, además de poder mantener el texto de la misma manera en que fue ingresado, el código permitirá realizar todo tipo de análisis automático posterior.

Objetivo

Describir el proceso de codificación automática a la que se sometió la información médica contenida en el SIAP-Win y analizar los resultados.

Materiales y Métodos

Dado que el nuevo modelo de registro se encuentra orientado a utilizar una lista de “problemas” por paciente, se analizaron las potenciales fuentes de información del registro actual. Los siguientes campos son los que contaban con la información más apropiada:

- **Motivo de consulta:** este campo contiene aquella razón que hizo que el paciente solicitara turno con el médico. Por ejemplo: Vómitos.
- **Diagnósticos:** este campo contiene los resultados de la conclusión a la que arribó el médico luego de la evaluación del paciente. Por ejemplo: Apendicitis.
- **Hábitos:** en este campo se encuentran aquellas conductas que se repiten sistemáticamente en el paciente. Por ejemplo: consumo de 1 litro vino diario.
- **Alergias:** este campo contiene información sobre aquellas sustancias que provocan reacciones de hipersensibilidad en el paciente. Por ejemplo: Alergia al polen.
- **Factores de riesgo:** contiene información sobre toda circunstancia o situación que aumenta las probabilidades de una persona de contraer una enfermedad. Ejemplo: Tabaquismo.
- **Antecedentes personales:** contiene información sobre cualquier enfermedad que el paciente haya tenido con anterioridad a la primera consulta médica. Fractura de tibia y peroné derecho a los 10 años.

Estos campos permitían al médico ingresar la información médica asociada a la consulta. En todos los casos se ingresaba

mediante texto libre, excepto el campo “Diagnósticos”, donde además existía la posibilidad de seleccionar una opción desde un menú previamente confeccionado.

De toda la información existente se procesó aquella que estaba destinada a formar parte de los “Problemas Activos” del paciente. Esta categoría representa la carga mórbida actual. Por ello se excluyó el campo: “Antecedentes personales”.

Los textos almacenados, fueron procesados por un Servidor de Terminología en español, desarrollado por el Hospital Italiano de Buenos Aires (STHIBA) [7, 8] que cuenta con:

- **Vocabulario de Interfase:** vocabulario controlado con el que se relaciona al usuario al momento de ingresar información al sistema. En este grupo se encuentran todas las variantes léxicas, acrónimos, abreviaturas y jerga locales utilizadas por los usuarios. Por ejemplo: “HTA”.
- **Vocabulario de Referencia:** vocabulario controlado que se utiliza para la representación más detallada de los datos en el sistema de información, es la forma en que los datos son finalmente almacenados en la base de datos. Todas las diferentes Terminologías de Interfase convergen en una única terminología de referencia, en este caso se utilizó SNOMED-CT. Por ejemplo, para el caso anterior: en SNOMED, el concepto se llamaría “Hipertensión Arterial” y le correspondería un código único: 38341003”.
- **Vocabulario de Salida:** Son las diferentes terminologías o clasificaciones que se utilizan para el análisis posterior de la información. Por ejemplo: ICD9-CM: 401; CIAP: K86; CIAP: K99.

Una representación gráfica de los componentes mencionados puede observarse en la clásica pirámide de la representación del conocimiento médico adaptada, Figura 1.

El STHIBA contiene más de 100.000 términos en español, agrupados en diferentes Dominios [9]. Los Dominios son palabras usadas o disponibles para su uso, por un individuo o grupo, o dentro de un particular tipo de trabajo o campo del conocimiento, como por ejemplo “Procedimientos”, “Problemas”, “Sitios de toma de muestra”, etc.

El STHIBA mediante Web Services desarrollados en Java y Stored Procedures desarrollados en PL/SQL en bases Oracle permite buscar una cadena de texto que se ingresa mediante texto libre. Primero realiza un proceso llamado “Normalización” que quita del texto ingresado todos aquellos caracteres que pueden deberse a un error de tipeo, como el guión del medio (-), guión bajo (_), barras (/), slashes (\), corchetes ([,]), comillas (“), etc. Si luego de realizar este proceso el texto queda vacío (por ejemplo si sólo se hubiera ingresado en un campo “-----”), el STHIBA devuelve el mensaje de “**Texto erróneo**”. En caso de quedar texto, realiza la búsqueda dentro de la base de datos.

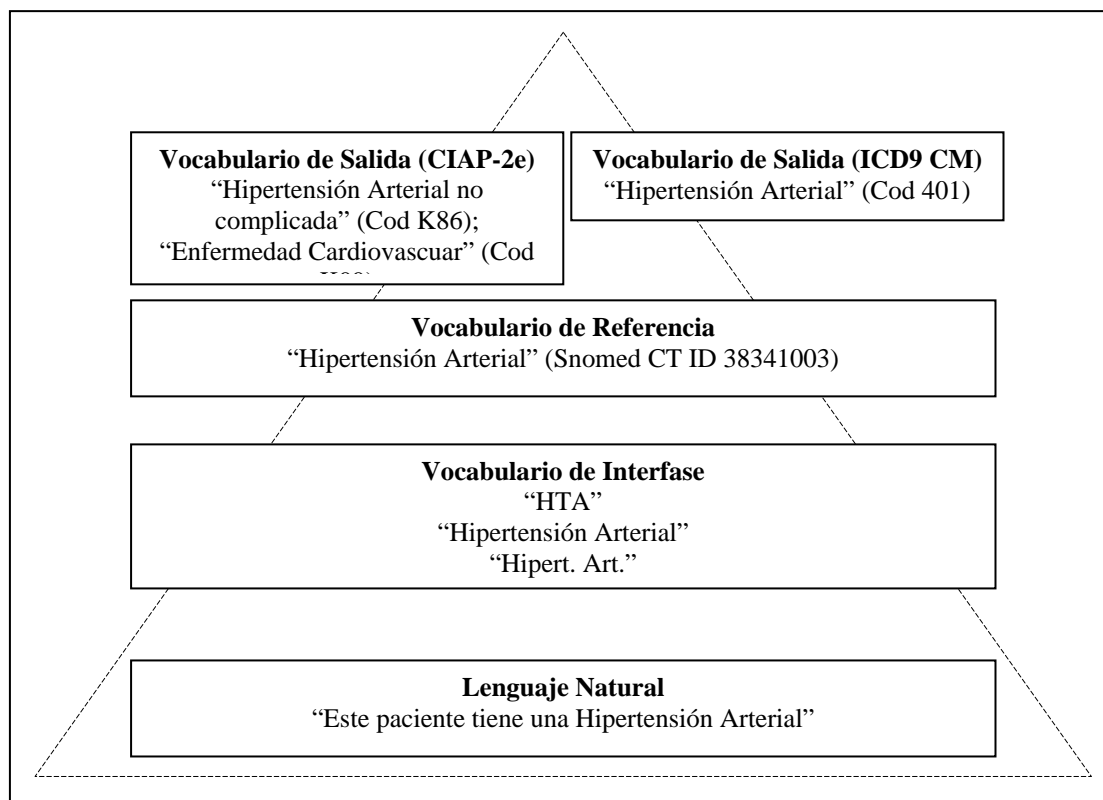


Figura 1 Modelo de relación entre los vocabularios (pirámide del conocimiento médico)

El texto ingresado podía encontrarse o no dicha la base de datos; esto determinaba si era reconocido o no, respectivamente. En caso de ser reconocido, podía ser válido o no. Por ello, las repuestas posibles eran:

- **Texto No Reconocido (NR)**. Por ejemplo: Hipertensión Arterial durante las noches de verano.
- **Texto Reconocido:**
 - El texto ingresado **no tiene opciones de refinamiento (V)**. Por ejemplo: Sarampión.
 - **Existen otras opciones de refinamiento** para el texto ingresado (V). Por ejemplo: Diabetes Mellitus. El profesional debería especificar el tipo de Diabetes Mellitus encontrada.
 - **No válido (NV)**, porque:
 - **No es una abreviatura válida**. Por ejemplo: HASL.
 - **Texto poco específico**. Por ejemplo: trauma.
 - **No es un problema**. Por ejemplo: apendicectomía.

- **Texto ambiguo**. Por ejemplo: angina.
- **Multi-término**. Por ejemplo: obesidad e hipertensión.
- **Debe indicar localización del problema o procedimiento**. Por ejemplo: verruga

Se decidió analizar aquellos textos que tuvieran 10 o más repeticiones y 100 o menos caracteres, que llamaremos regla 10/100. Esta decisión se tomó en base a poder tomar una muestra representativa de la cantidad de textos existentes. Los que cumplen con esta regla representan el 80% del total de los textos almacenados en la base de datos.

Resultados

Luego de extraer los textos desde los campos explicados anteriormente se encontró un total de 16.245.054 registros, y luego de correr la regla 10/100 quedaron 14.120.752. La cantidad de términos de acuerdo al campo extraído puede verse en la Tabla 1.

| Campo del SIAP-Win | Cantidad | % |
|-----------------------------|------------|------|
| Alergias | 47.898 | 0,3 |
| Diagnósticos no Codificados | 8.254.749 | 58,5 |
| Diagnósticos Codificados | 2.374.148 | 16,8 |
| Factores de riesgo | 132.374 | 0,9 |
| Hábitos | 43.526 | 0,3 |
| Motivos de consulta | 3.268.057 | 23,1 |
| | 14.120.752 | |

Tabla 1 Cantidad de textos por dominios

El resultado de procesar (normalización y búsqueda de coincidencia en la base de datos) los 14.120.752 textos encontrados, se muestra en la Tabla 2:

| | Válidos | No Válidos | No reconocidos | Texto erróneo |
|-----------------------------|------------|------------|----------------|---------------|
| Alergias | 36.539 | | 11.359 | |
| Diagnósticos no Codificados | 6.660.487 | 555.559 | 1.037.237 | 1.466 |
| Diagnósticos Codificados | 1.543.798 | 20.482 | 809.867 | |
| Factores de riesgo | 119.452 | 4.292 | 8.630 | |
| Hábitos | 577 | 8.986 | 15.943 | 18.020 |
| Motivos de consulta | 2.012.423 | 136.679 | 1.118.955 | |
| | 10.373.276 | 725.998 | 3.001.991 | 19.486 |

Tabla 2 Resultado separados por dominios y categorías

La cantidad global de textos reconocidos por el STHIBA fue de 79% (teniendo en cuenta los términos Válidos y los No Válidos), como muestra la Figura 2.

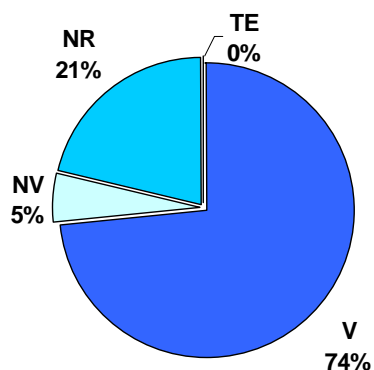


Figura 2 Problemas reconocidos y no reconocidos

Discusión

Este proceso permitió tener disponible una gran cantidad de la información existente actualmente en el SIAP-Win en la Nueva Ficha Clínica con su correspondiente codificación. Se podrán adoptar medidas de cuidados preventivos rápidamente en la población de pacientes propios de Megasalud, pertenecientes al plan "Mi Doctor". De la misma manera que planes para aquellos que presentan enfermedades crónicas como diabetes mellitus o hipertensión arterial.

Fue necesario realizar algunos ajustes en el proceso de Normalización del STHIBA debido a la diferencia en la forma de registrar la información médica que existe entre Argentina y Chile

Si bien podemos encontrar como debilidad el hecho de no haber procesado el total de la información disponible, el análisis realizado sobre los textos que quedaron fuera mostró que existía una cantidad de caracteres superior a los 300 y que la mayoría de los textos tenían muy poca cantidad de repeticiones, en su mayoría estaban ingresados sólo 1 vez. Se encontraron múltiples registros donde en lugar de haber una conceptualización médica, existía una evolución completa. Es probable que si se hubiera utilizado una función que truncara los textos largos se podría haber obtenido una codificación mayor. Otra alternativa hubiera sido la de buscar coincidencia exacta de determinadas cadenas de caracteres dentro de estos textos tan grandes, como por ejemplo: "hipertensión".

Se encontró una elevada cantidad de textos erróneos, sobre todo en el campo "Hábitos". Es probable que los médicos colocaran el símbolo "-" como una manera de representar que el paciente no posee hábitos tóxicos; pero este símbolo es eliminado previo al momento de realizar la consulta sobre si el término existe o no, dejando el texto a consultar en blanco.

Conclusión

El procesamiento de la información médica almacenada en texto libre y procesada mediante un Servidor de Terminología permite obtener una alta disponibilidad de la información existente, disminuyendo la tarea de la codificación manual.

Referencias

- [1] Johnson SB, Bakken S, Dine D, Hyun S, Mendonca E, Morrison F, et al. An electronic health record based on structured narrative. J Am Med Inform Assoc. 2008 Jan-Feb;15(1):54-64.
- [2] Panko WS, J, Lincoln, T, editor. Technologies for Extracting Full Value from the Electronic Patient Record. 32nd Hawaii International Conference on System Sciences; 1999; Hawaii. IEEE.
- [3] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine. 1998 Nov;37(4-5):394-403.
- [4] Bleeker SE, Derksen-Lubsen G, van Ginneken AM, van der Lei J, Moll HA. Structured data entry for narrative data in a broad specialty: patient history and physical

examination in pediatrics. BMC medical informatics and decision making. 2006;6:29.

- [5] <http://www.megasalud.cl/>. [cited]; Available from.
- [6] Osornio AL, Luna D, Gambarte ML, Gomez A, Reynoso G, de Quiros FG. Creation of a local interface terminology to SNOMED CT. Studies in health technology and informatics. 2007;129(Pt 1):765-9.
- [7] Lopez Osornio AL, D.; González Bernaldo de Quirós, F. Creación de un sistema para la codificación automática de una lista de problemas. 5to Simposio de Informática en Salud - 31 JAIIO; 2002. 2002.
- [8] Lopez Osornio A.; Gambarte L.; Otero C.; Gomez AMMSEL, D.; Otero P.; Clusella M.; Mitre M.; Reynoso G.; González Bernaldo de Quirós, F. Desarrollo de un servidor de terminología clínico. VIII Simposio Argentino de Informática y Salud – SIS 2005; 2005. 2005.
- [9] Lopez Osornio AM, S.; García Marti, S.; Toselli, L.; Otero, C.; Tavasci, I.; Luna, D.; Gomez, A.; González Bernaldo de Quirós, F. Codificación múltiple de una lista de problemas utilizando la CIAP-2, CIE-10 y SNOMED CT. 3er Virtual Congress of Medical Informatics; 2004. 2004

Dirección para correspondencia:

Claudio Torres Casanelli
claudio.torres@megasalud.net