

Auto-Codificador CIAP en español a través de Internet

Lopez Osornio Alejandro ^a, Schpillberg Mónica ^a, Soriano Enrique ^b, Luna Daniel ^b, González Bernaldo de Quirós Fernán ^b

^a Residencia de Informática Médica, Hospital Italiano de Buenos Aires

^b Departamento de Información Hospitalaria, Hospital Italiano de Buenos Aires
Buenos Aires, Argentina

Resumen

Ante la importancia del control del vocabulario, y la codificación de diagnósticos médicos con clasificaciones o nomenclaturas, es necesario contar con procesos de codificación válidos con un máximo de confiabilidad y un mínimo de variabilidad.

Tomando como base la experiencia de más de 500.000 diagnósticos médicos codificados por un equipo de codificadores validados científicamente, se construyó una base de conocimientos que permite sugerir códigos CIAP a textos narrativos libres ingresados por el usuario. El sistema de selección de códigos se realiza por repetición y normalización del texto libre.

Se construyó una página Web con el primer auto-codificador CIAP en español, para demostrar la utilidad de este método y compartir esta base de conocimientos con otros centros de codificación.

El sitio Web permite el ingreso de un texto, este texto es normalizado y comparado con la base de conocimientos y da como resultado el código CIAP con el que nuestro equipo de codificadores ha codificado este término.

Introducción

La Clasificación Internacional de Atención Primaria (CIAP), creada y mantenida por la Organización Mundial de Médicos de Familia (WOMCA)[1], es una de las herramientas más ampliamente utilizadas en el mundo para la codificación de problemas de atención primaria. Sus cerca de mil códigos representan apropiadamente la realidad del proceso de atención ambulatoria, con un nivel de detalle (granularidad) suficiente para poder realizar análisis epidemiológicos y de gestión.

Los principios básicos de la Informática médica sostienen la importancia de mantener vocabularios controlados[2], en el caso del control del vocabulario de los diagnósticos ambulatorios se presentan diferentes estrategias para su implementación. Una vez elegido la clasificación o nomenclatura a utilizar, existe la posibilidad de entrenar a los profesionales para que elijan los códigos apropiados para registrar los datos de sus pacientes, esto se denomina codificación primaria. Es común encontrar resistencia por parte de los profesionales médicos, que encuentran más rápido y claro el registro tradicional con descripciones mediante textos narrativos no controlados de los problemas. Una solución a esta dificultad es la codificación secundaria mediante un equipo entrenado, que lee el texto ingresado por los médicos y elige el código correspondiente de la clasificación.

En cualquiera de estas estrategias muy importante un profundo entrenamiento del encargado de la codificación para asegurar el mismo criterio de codificación entre diferentes personas y a través del tiempo. En este sentido es de gran utilidad contar con una herramienta inteligente que sugiera los códigos apropiados a partir de una base de conocimientos validada y creada por codificadores altamente capacitados. Hasta el momento no existe ninguna herramienta de este tipo que analice texto en español y sugiera el código CIAP apropiado.

Objetivo

Crear una herramienta basada en la Web que sugiera códigos CIAP para textos libres redactados en español, para su utilización durante el proceso de codificación por parte de codificadores de habla hispana.

Materiales y Métodos

Entre las diferentes opciones para crear un sistema de codificación automática utilizamos la repetición de textos. Creamos una base de conocimientos a partir de un proceso de agrupación y normalización de textos similares, verificando la predominancia de la asignación de un código determinado para cada texto libre. Este proceso está ampliamente descrito en un estudio previo [3] y aunque tiene menos confiabilidad que el análisis de lenguaje natural, es mucho más sencillo de implementar ahorrando en recursos de desarrollo. Cuanto mayor es el número de problemas a partir del cual se genere la base de conocimientos, mayor validez tiene la estrategia de repetición de textos libres.

El codificador se generó a partir de una base de datos de 524.248 problemas médicos, recopilados a lo largo de 4 años, ingresados en forma de texto libre por médicos en la Historia Clínica Electrónica del Hospital Italiano de Buenos Aires. Estos problemas fueron codificados manualmente por un equipo de codificadores entrenados y ya sometidos a un proceso científico de validación. Mediante el proceso de normalización, agrupación y selección de códigos CIAP, generamos un Tesoro de más de 14.000 descripciones con su correspondiente código CIAP asignado.

Se desarrolló una página Web con la posibilidad de recorrer la clasificación CIAP ordenada por aparatos o realizar búsquedas en la misma, para permitir que el usuario se familiarice con la misma. En la sección de auto-codificación el usuario del sitio Web ingresa un texto de hasta 40 caracteres y luego de un proceso de análisis el sistema le sugiere el código CIAP más adecuado y le presenta la descripción para que el usuario pueda comprobar su exactitud.

El proceso de elección del código sugerido consiste en el procesamiento del texto ingresado en el formulario Web con el mismo proceso de normalización que se utilizó durante la creación del tesoro y la búsqueda de la coincidencia con alguna entrada del tesoro. La búsqueda de coincidencias con el tesoro consiste en verificar si existe un término exactamente (opción 'Búsqueda Exacta' en la Web) igual al ingresado por el usuario luego de la normalización, o si existe algún otro que lo contenga (opción 'Búsqueda General' en la Web). Si no se encuentran resultados luego de este paso el texto ingresado por el usuario será considerado *desconocido* y el sistema no ofrecerá ninguna sugerencia, y lo almacenará en una base de datos para su posterior codificación manual y retroalimentación del sistema.

La programación del sitio Web se realizó en lenguaje PHP, utilizando una base de datos MySQL y un servidor Web apache, bajo sistema operativo Linux.

Resultados

El sitio Web está disponible en la dirección <http://200.10.104.201/codificador-ciap>. El acceso es gratuito.

Discusión

La herramienta de codificación automática se basa en la experiencia de un equipo de codificadores entrenados, con aproximadamente un 85% de confiabilidad contra un Gold Standard y de acuerdo inter e intra-codificadores, según se demostró en un estudio anterior [4].

Aunque la estrategia de repetición de texto posee peor versatilidad y características operativas, con respecto a un servidor de terminología con análisis de lenguaje natural [5, 6], que es la estrategia más en boga actualmente, la estrategia propuesta es de muy simple implementación y bajo costo. Su utilidad mejora ampliamente al contar con una gran base de conocimientos y al utilizar una clasificación con pocos códigos como la CIAP. Hay experiencia en la utilización de este tipo de estrategia en otros idiomas [7].

La herramienta codificó automáticamente más del 70% de los diagnósticos ingresados por los médicos que estaban almacenados en una base de datos ambulatoria, suponemos que en este caso, al utilizar un ingreso interactivo, en el que se puede modificar el diagnóstico si no se obtienen resultados, o se puede elegir el código más pertinente de una lista de más de un resultado, el porcentaje de cobertura aumentaría en gran medida.

Sería interesante realizar en el futuro un estudio que evalúe el impacto, medido en disminución de la variabilidad o aumento de la confiabilidad, del uso de una herramienta de ayuda a la codificación como ésta en un centro de codificación secundaria o sistemas con codificación primaria.

Conclusiones

El codificador automático CIAP basado en la Web es un intento de compartir la experiencia y confiabilidad de un equipo de codificadores, alcanzada luego de 4 años de trabajo, permitiendo que grupos pequeños o descentralizados de codificación tengan una herramienta de apoyo en el proceso de codificación de los diagnósticos médicos.

Referencias

1. WONCA, *Wonca Online - Website of the World Organization of Family Doctors*. 2002.
2. Shortliffe, E., *Medical Informatics: Computer Applications in Health Care and Biomedicine*. 2nd ed. 2000: Springer Verlag.
3. Lopez Osornio A, L.D., Bernaldo de Quiros FG. *Creación de un sistema para la codificación automática de una lista de problemas*. in *5to Simposio de Informática en Salud - 31 JAIIO; 2002*. 2002. Santa Fe, Argentina: SADIO.
4. Luna, D., F.G. Bernaldo de Quiros, L. Garfí, E. Soriano and M. O'Flaherty. *Reliability of secondary central coding of medical problems in primary care by non medical coders, using the International Classification of Primary Care (ICPC)*. in *Medinfo, 2001*. 2001. London, England.
5. Chute, C.G., P.L. Elkin, D.D. Sherertz and M.S. Tuttle. *Desiderata for a clinical terminology server*. in *Proc AMIA Symp, 1999*. 1999.
6. Lussier, Y.A., L. Shagina and C. Friedman. *Automating SNOMED coding using medical language understanding: a feasibility study*. in *Proc AMIA Symp, 2001*. 2001.
7. Letrilliart, L., C. Viboud, P.Y. Boelle and A. Flahault. *Automatic coding of reasons for hospital referral from general medicine free-text reports*. in *Proc AMIA Symp, 2000*. 2000.

Dirección para correspondencia

Dr. Alejandro Lopez Osornio
Residencia de Informática Médica
Hospital Italiano de Buenos Aires
Gascón 450
(1181) Buenos Aires – Argentina
alejandro.lopez@hospitalitaliano.org.ar