

Informe sobre la Codificación Automática

Objetivo del proyecto

En diciembre de 2007 tuvimos la oportunidad de interactuar con un entorno real de urgencias en la Fundación Puigvert ante la necesidad de establecer un código sobre los motivos y diagnósticos de este servicio.

Ello nos permitió poner a prueba diferentes conceptos sobre los que ya habíamos tratado anteriormente como base para el desarrollo de un sistema de codificación automático que era necesario ante las necesidades actuales de codificación.

Si bien existen diferentes proyectos y realizaciones de sistemas de codificación automática, estos se han basado por lo general en sistemas de coincidencia terminológica, lo que en un entorno real en qué el médico utiliza libremente la lengua, pone límites excesivamente restrictivos a los sistemas: tanto la terminología utilizada en los sistemas estándar (UMLS, SNOMED, ICD, etc.) como los errores de escritura, tanto de ortografía como tipográficos, son un elemento de dispersión muy amplio para permitir que pueda desarrollarse un sistema automático basado en esta tecnología.

Por otra parte, la abundante terminología médica existente y la ausencia de recursos reales tanto en español como en catalán, dificultan enormemente la posibilidad de desarrollar sistemas basados en ésta que puedan generalizarse de forma sencilla y práctica. Más aún en un entorno multilingüe.

Así, perseguíamos desarrollar un sistema de codificación que fuera independiente de recursos lingüísticos específicos, independiente también de la lengua, y adaptable en la práctica, a los diferentes estándares de codificación existentes en el sector de salud.

Por otra parte, las necesidades de producción de la F. Puigvert, que dispone de un equipo de Documentalistas Médicos para la codificación de los informes, era un entorno ideal para el desarrollo y prueba de nuestro sistema. Disponíamos de suficiente material para desarrollar un primer prototipo y probarlo en este entorno real con la ayuda de

profesionales que podían valorar la eficiencia del sistema, así como identificar los elementos que debían mejorarse para optimizar este resultado.

Solución propuesta

Nuestro proyecto se basa en la utilización de diferentes tecnologías que nos permiten acercarnos desde perspectivas que se complementan entre sí al problema de la codificación.

En esencia, lo que perseguimos es establecer un modelo semántico (arquetipo) de la información contenida en cada código, que nos permitiera identificar, por comparación, el motivo o diagnóstico escrito por el médico. Como es lógico, este modelo debe ser capaz de asumir tanto los acrónimos, las abreviaturas, los errores tipográficos y las faltas de ortografía, a la vez que los diferentes idiomas utilizados en la práctica. El sistema no plantea restricciones sobre el uso del lenguaje del médico y en la práctica debe ser capaz de adaptarse a cualquier entorno sin necesitar una fuerte inversión en adaptarse a la terminología específica de cada servicio, departamento, área etc. Tal solución, en la práctica, supondría un freno a las posibilidades reales de implementación y puesta en funcionamiento de un sistema general de codificación.

La ubicación del motor de codificación en el entorno de la F. Puigvert se ha realizado *off-line*, es decir, de acuerdo con los criterios de la Fundación, la codificación se realiza *a posteriori*, y el médico emisor del informe no tiene contacto con la codificación, ésta queda en manos de un equipo específico que realiza y verifica esta función.

Desde el punto de vista del código utilizado, hay que señalar que si bien el área está limitada a una especialidad, y esto supone una restricción en el número de códigos posibles, la Fundación Puigvert utiliza por sus propias necesidades de codificación una extensión propia del sistema ICD 9 CM (CIE 9 MC).

Índice de Confianza (Ic): Validación de los resultados

Un sistema automático requiere, además de una solución, una garantía en los resultados obtenidos. Si los resultados obtenidos son válidos de forma aleatoria, sin una cobertura del 100% en los aciertos, no puede hablarse de codificación automática, sino de sistema de ayuda a la codificación, o de sistemas semiautomáticos.

Bajo estas condiciones el sistema requiere de intervención humana, sean los médicos que confirman o los documentalistas médicos que supervisan las decisiones del sistema.

Desde esta perspectiva, y dado que los sistemas que trabajan sobre lenguaje natural no pueden dar un resultado del 100%, se planteó como elemento esencial del prototipo establecer un mecanismo que permitirá discriminar por lo menos aquellos resultados que sin ninguna duda fueran correctos, de aquellos que plantearan dudas sobre su corrección.

Para ello establecimos un *Índice de Confianza*, cuyos valores pueden ir de 25 a 100, que mide la seguridad del sistema en el resultado presentado.

Dado que el motor de codificación utiliza diferentes tecnologías, este *Índice de Confianza* es una suma vectorial de diferentes factores constitutivos y su representación es no lineal.

Resultados obtenidos

A primeros de mayo de 2008 pusimos en marcha nuestra primera versión del sistema de codificación.

Como se ha explicado, el sistema obtiene un informe médico de la Base de Datos del centro y para cada informe establece dos códigos, uno para el motivo y otro para el diagnóstico.

Adicionalmente, el sistema establece para cada uno de estos códigos un *Índice de Confianza* en el código obtenido.

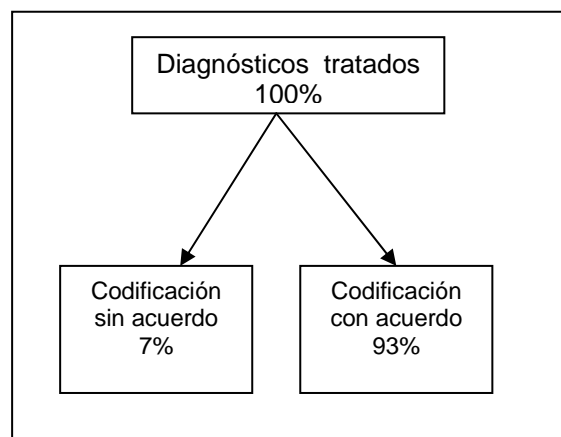
El volumen de urgencias tratado es de unas 125 diarias, de forma que el sistema establece aproximadamente 250 códigos al día. El sistema se ha programado para que diariamente (por la noche) capture los informes médicos de la Base de Datos del centro, establece los códigos y los deposita de nuevo en la Base de Datos. Para la realización

de estas funciones el sistema tarda apenas unos minutos, la mayor parte de este tiempo se destina a la localización e intercambio de información con el sistema informático del centro. Sólo en las tareas de codificación, apenas tarda unos segundos.

El sistema ha sido implementado para no interferir en el proceso de producción del centro. Pero en la actual fase de pruebas la codificación, realizada por especialistas, los médicos documentalistas supervisan la totalidad de los códigos establecidos por el motor de codificación, así como el comportamiento del *Índice de Confianza*.

Esta primera prueba de funcionamiento en un entorno real nos permitió por una parte, evaluar el funcionamiento del motor de codificación y del *Índice de Confianza*, así como valorar las diferentes situaciones que pueden inducir el mal funcionamiento del sistema.

El primer análisis lo centramos en la validación absoluta de los resultados, sin discriminar las causas que producían el error, y los resultados se analizaron por separado entre motivos y diagnósticos.

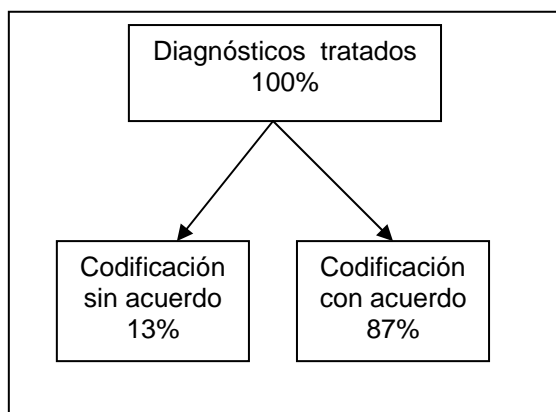


Los resultados de los motivos son muy positivos comparados con los obtenidos y publicados en otros trabajos realizados.

El estudio de los errores detectados nos permitió identificar una casuística variada, pero en la que predominaban un conjunto de motivos no clínicos que, en consecuencia, no eran codificables en ningún caso.

Junto a esta situación, se identificaron causas específicas de uso del lenguaje que nuestro sistema no era capaz de tratar.

Por lo que se refiere a los diagnósticos, los resultados obtenidos fueron los siguientes:



Del análisis de estos resultados, se identificaron también algunas de las causas de error más comunes: por una parte expresiones no clínicas y en consecuencia no codificables; y por otra parte determinados formatos de comunicación y expresiones utilizadas por los médicos que no habían sido previstas inicialmente.

Entre éstas, los ejemplos más frecuentes son la presencia de negaciones en el diagnóstico y situaciones de diagnóstico múltiple. El sistema informático existente en la Fundación Puigvert dispone de un único campo para el diagnóstico, y no existe un protocolo unificado para la expresión de estas situaciones por lo que el multid diagnóstico introduce confusión al sistema.

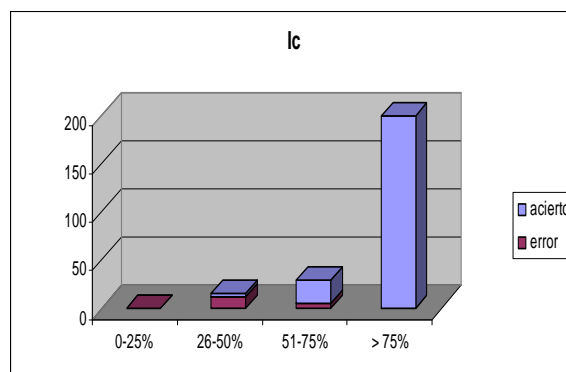
No obstante, y a la vista de los resultados, la valoración del sistema de codificación se ha considerado excelente.

El análisis del *Índice de Confianza* ha dado también unos resultados equivalentes y ha permitido establecer una frontera en el valor del 75% que segrega de forma aceptable el conjunto de motivos y diagnósticos correctamente aceptados de aquellos que deben ser revisados.

En la práctica, un 80% de los códigos presentan un *Índice de Confianza* superior al 75% y un 20% inferior. La distribución de los errores presenta una distribución correlacionada con la confianza por lo que es posible identificar el segmento donde se concentran los errores a partir de este índice.

Los resultados, representados en la siguiente cuadro, permiten considerar correctamente codificado un 80% de los motivos y diagnósticos tratados, y limitar a un 20% aquellos que deben ser revisados. De éstos,

aproximadamente en la mitad de los casos existe acuerdo en los códigos establecidos por el sistema y el resto debe modificarse.



Si bien estos resultados son buenos para el centro, ya que reducen la carga de trabajo del departamento de codificación en un 80% y en esa medida permiten extender la codificación a otras áreas y servicios, la identificación de las causas de mal funcionamiento nos ha permitido abordar una nueva versión del sistema de codificación que supere las limitaciones de la primera versión.

Mejoras incorporadas al sistema

De lo expuesto, se deduce que el primer objetivo es identificar aquellas entradas sin contenido clínico, para excluirlas del proceso. Efectivamente, si el médico no incluye en el motivo o en el diagnóstico información que permita su codificación, ésta no debe ser tratada por el sistema. De la casuística analizada, entre un 3% (diagnósticos) y un 5% (motivos) no pueden ser tratados por un sistema de codificación, por lo que o bien se modifica la entrada y los protocolos del centro, o bien deberá seguir codificándose por los expertos que actuarán de acuerdo a los protocolos y decisiones que la Fundación Puigvert establezca.

Aunque parezca simple, no es sencillo identificar de forma automática estas situaciones, fundamentalmente porque el lenguaje utilizado presenta una elevada variabilidad y su tratamiento no está relacionado con la terminología.

El segundo problema que nos planteamos resolver es el del multid diagnóstico. La Fundación Puigvert está poniendo en funcionamiento un nuevo HIS que contempla esta situación, por lo que en un futuro inmediato los diagnósticos múltiples vendrán identificados en campos diferentes y dejarán de suponer un problema para su codificación.

Entretanto, se ha acordado con el centro que en estas situaciones se codificará de acuerdo con el diagnóstico indicado en primera posición, ignorando el resto de diagnósticos.

El tercer elemento a incluir en el sistema es la negación, ya que su presencia en los diagnósticos es persistente, con un importante contenido semántico, y su resolución ayudará de forma significativa a mejorar las prestaciones del sistema de codificación.

Estas mejoras se desarrollaron durante el mes de junio de 2008 y se aplicaron en una nueva versión del motor de codificación en entorno de producción en julio de este mismo año.

Resultados y situación actual

Paralelamente a estas mejoras, para adaptar el motor de codificación al nuevo HIS que está implementando la Fundación Puigvert, se ha desarrollado un Webservice que responde con uno o varios códigos a las consultas formuladas desde el sistema. De esta forma el sistema puede operar tanto *on-line* como *off-line*, adaptándose de esta forma a las diferentes necesidades y sistemas existentes.

Por lo que se refiere a los resultados obtenidos en esta segunda versión, la siguiente tabla nos ofrece los resultados obtenidos:

	lc > 75	lc < 75
coincidencia	84,7%	9,7%
error	0,4%	5,1%

Los errores del sistema se han reducido a un 5,5% de los códigos establecidos y el volumen de códigos que no requiere supervisión sube a un 85%, quedando un 15% que necesita revisión para confirmar la codificación establecida por el sistema.

Desde un punto de vista tecnológico, aunque el nivel de acierto podrá mejorar en el futuro, el sistema puede considerarse maduro y con unos resultados muy buenos, las posibles mejoras que en cualquier caso serán marginales, solo podrán ser identificadas y abordadas a partir de la extensión del sistema de codificación a otros entornos y estándares de codificación.

Efectivamente, no hay que pecar de ingenuidad, las tecnologías de PLN (Procesamiento del Lenguaje Natural) se enfrentan a situaciones casi ilimitadas, dada la flexibilidad del lenguaje. Es por ello que

sistemas basados en estadística y terminología que pueden alcanzar sin demasiadas dificultades unos niveles de acierto del 50% ó 60%, y que se utilizan en los sistemas de ayuda a la codificación, encuentran dificultades para mejorar estos niveles de acierto. En la práctica, un sistema que codificara el 15% de los códigos más frecuentes, tendría una resolución superior al 60%.

Paralelamente, los errores tipográficos y ortográficos, con una presencia constante y aleatoria (un 30% de los textos presentan errores), son uno de los principales obstáculos para la correcta codificación.

La resolución de estas limitaciones con tecnologías de PLN es lo que permite a un sistema sobrepasar esta frontera para poder ser aplicado de forma automática.

Objetivos inmediatos

Así pues, los objetivos actuales son dos. Por una parte, la extensión del sistema a otros centros, tanto en Urgencias como en otros servicios, lo que ha de permitir identificar las limitaciones actuales de nuestro sistema que incidan en la mejora de la resolución. Por otra parte, su extensión a otros estándares de codificación, y en especial la codificación con SNOMED CT, han de permitir dotar al sistema de la potencia y resolución esperada.

Efectivamente, la aplicación del sistema a informes de alta, listas de espera, consultas externas, además de las urgencias, ha de permitir generalizar las soluciones desarrolladas hasta el momento y en consecuencia mejorar su resolución.

En la práctica, el objetivo de un sistema de codificación es hacer accesible la información acumulada en los informes médicos para su explotación como fuente de conocimiento. Mientras que sistemas de codificación como CIE 9 MC o CIE 10 tienen un uso en la gestión y organización, sistemas más extensos como SNOMED CT tienen como objetivo la utilización de esta información por los médicos y con un objetivo clínico. Este es el desafío en el momento actual.

Barcelona, 2 de septiembre de 2008
Enric Castellón
Thera, Centre de Llenguatges i Computació, SL