

# AMPLIACIÓN DE LAS CAPACIDADES DE RECUPERACION DE INFORMACION DE LA WEB A PARTIR DE UNA HISTORIA CLINICA ELECTRÓNICA CONFORME AL PRE-ESTÁNDAR ENV13606.

*Jorge Plüss (1)(3), Francisco del Pozo, M. Elena Hernando, Santiago Rodríguez,  
Enrique J. Gómez y Paula de Toledo (1), Carlos H. Salvador (2), Cristina Bender, Claudia Deco,  
Alejandro Hernández y Néstor Pozzoli (3)*

(1)Grupo de Bioingeniería y Telemedicina. Universidad Politécnica de Madrid.

(2) Grupo de Bioelectrónica, Telemedicina y Biomecánica.  
Hospital Universitario Puerta de Hierro, Madrid

(3) Facultad de Ingeniería, Universidad Nacional de Rosario, Argentina

## RESUMEN

En este trabajo se propone una arquitectura para la recuperación de información bibliográfica a partir de datos obtenidos de una Historia Clínica Electrónica conforme a la pre-norma 13606 del CEN [1].

La recuperación de literatura médica se realiza en distintos tipos de fuentes de datos, brindándole al usuario una única respuesta integrada. Además se describen las tecnologías y herramientas que se utilizan en ella, como ser tesauros y ontologías.

Esta arquitectura facilita el acceso de los profesionales clínicos a los contenidos de conocimiento en los repositorios de literatura científica de acceso electrónico proporcionándoles una búsqueda más precisa.

## INTRODUCCIÓN

El concepto de medicina basada en la evidencia (MBE) fue elaborado por un grupo de la Universidad McMaster de Ontario, Canadá, liderado por David Sackett (médico internista), quién observó que sólo una minoría de las decisiones tomadas por los médicos, se basan en estudios científicos serios [2][3]. Las consecuencias fundamentales de este hecho son, por una parte cuestionar la validez de la información proveniente de la experiencia personal de los médicos y de los libros de texto, y por otra la necesidad de estimular a los profesionales a tener contacto directo con la literatura médica, capacitándolos para realizar un análisis crítico de la misma con herramientas para evaluar la validez y la utilidad de los estudios científicos publicados y descartar aquellos con deficiencias metodológicas.

Los médicos deben consultar rutinariamente la literatura científica más reciente referente a los problemas concretos que están abordando, perfeccionando su capacidad para realizar un análisis crítico de la misma. Esta permanente conexión con las fuentes originales sería la única forma de mantener actualizados los conocimientos.

La metodología que emplea la medicina basada en la evidencia puede resumirse en cuatro pasos:

1. La información que se quiere obtener para solucionar un problema clínico debe expresarse como una pregunta concreta y que sea plausible de ser contestada también concretamente.
2. Buscar evidencias científicas sobre el tema empleando las nuevas herramientas que ofrece la informática: páginas web especializadas, publicaciones on line, bases de datos de ensayos clínicos (Cochrane), sitios de publicación de síntesis de evidencias clínicas (Bandolier, Best Evidence), CD ROMS, etc.

1. Evaluación crítica de la información obtenida. Este punto implica tener la capacidad de analizar la corrección de la metodología empleada en cada trabajo. (Planteamiento de hipótesis, selección de muestras, procedimientos de obtención de datos, tratamiento estadístico de los mismos, inferencias extraídas de los resultados).
2. Integración de la nueva información con los conocimientos previos para tomar la decisión adecuada sobre el caso en estudio

La MBE ha ganado una significativa aceptación en la comunidad médica como una alternativa que trata de aportar soluciones a las carencias reales en el enfoque actual del proceso de toma de decisiones clínicas.

Un médico con experiencia tendrá un número limitado de cuestiones concretas para consultar, y si además tiene la capacidad de saber elegir la estrategia de búsqueda más adecuada, podrá solucionar el problema con el mínimo costo en tiempo, recursos y riesgos, y con las máximas probabilidades de éxito. En este sentido, el trabajo propuesto tiene dos ejes fundamentales.

Por un lado la posibilidad de disponer del conocimiento relativo a las estrategias de búsqueda. Por otro lado, un acceso eficiente a la historia clínica como un elemento esencial que proporciona el escenario real de la búsqueda a partir de la información disponible de un paciente o de un grupo de pacientes perfectamente individualizados.

Las estrategias de búsqueda suponen inicialmente el desarrollo de un proceso de adquisición de conocimiento y su empleo en situaciones concretas, como, así también, la posibilidad de establecer mecanismos de actualización de dichas estrategias en la medida que los usuarios sean capaces de discriminar la información relevante de la que no es relevante. O lo que es lo mismo, que el sistema “aprenda” a mostrar la información que puede ser relevante para una búsqueda, basándose en un posible estudio estadístico de la relevancia de los campos en búsquedas calificadas como “éxito” por los usuarios. El proceso de aprendizaje de las estrategias de búsqueda tendrá un importante componente subjetivo, y en principio conven- drá circunscribir el espacio de búsqueda al ámbito de publicaciones reconocidamente calificadas.

## **FUENTES DE INFORMACION Y ADQUISICIÓN DEL CONOCIMIENTO PARA LA BÚSQUEDA DE LITERATURA**

El conjunto de documentos resultante del Proceso de Búsqueda puede provenir de distintas fuentes de información. Estas fuentes pueden ser bases de datos bibliográficas disponibles en la Web, páginas Web fijas y páginas Web variables obtenidas a través de un buscador. Este proceso de recuperación, extracción e integración de la información responde a una estrategia de búsqueda validada por el usuario, y es enviada a estos distintos tipos de fuentes y las respuestas de cada una de ellas, se presentada en una única respuesta al usuario.

Entre las fuentes de información referentes a bases de datos bibliográficas, como ser Medline, Excerpta Medica, Lilacs, entre otras, el sistema de búsqueda le permitirá al usuario seleccionar la fuente de información conforme a su decisión sobre la naturaleza del problema y el repositorio que el considere más adecuado a sus propósitos.

La flexibilidad y variedad del lenguaje natural crea serias dificultades para el manejo automatizado de la información. Para solucionar este problema se utilizan tesauros que permiten traducir información del lenguaje natural de los documentos y usuarios a un lenguaje controlado del sistema. En Medline es posible utilizar el índice Mesh (Medical Subject Subheadings) el cual brinda un listado de términos controlados ó palabras clave utilizados en el campo de descriptores de la base de datos. Mesh es el vocabulario controlado creado por la Biblioteca Nacional de Medicina de Estados Unidos y la utilización de sus términos, permite realizar una búsqueda más precisa. El uso de descriptores aumenta la precisión pero confiar sólo en ellos puede reducir la recuperación, por lo tanto se aumenta la recuperación por usar tanto el descriptor como la frase de uso común ó término no controlado.

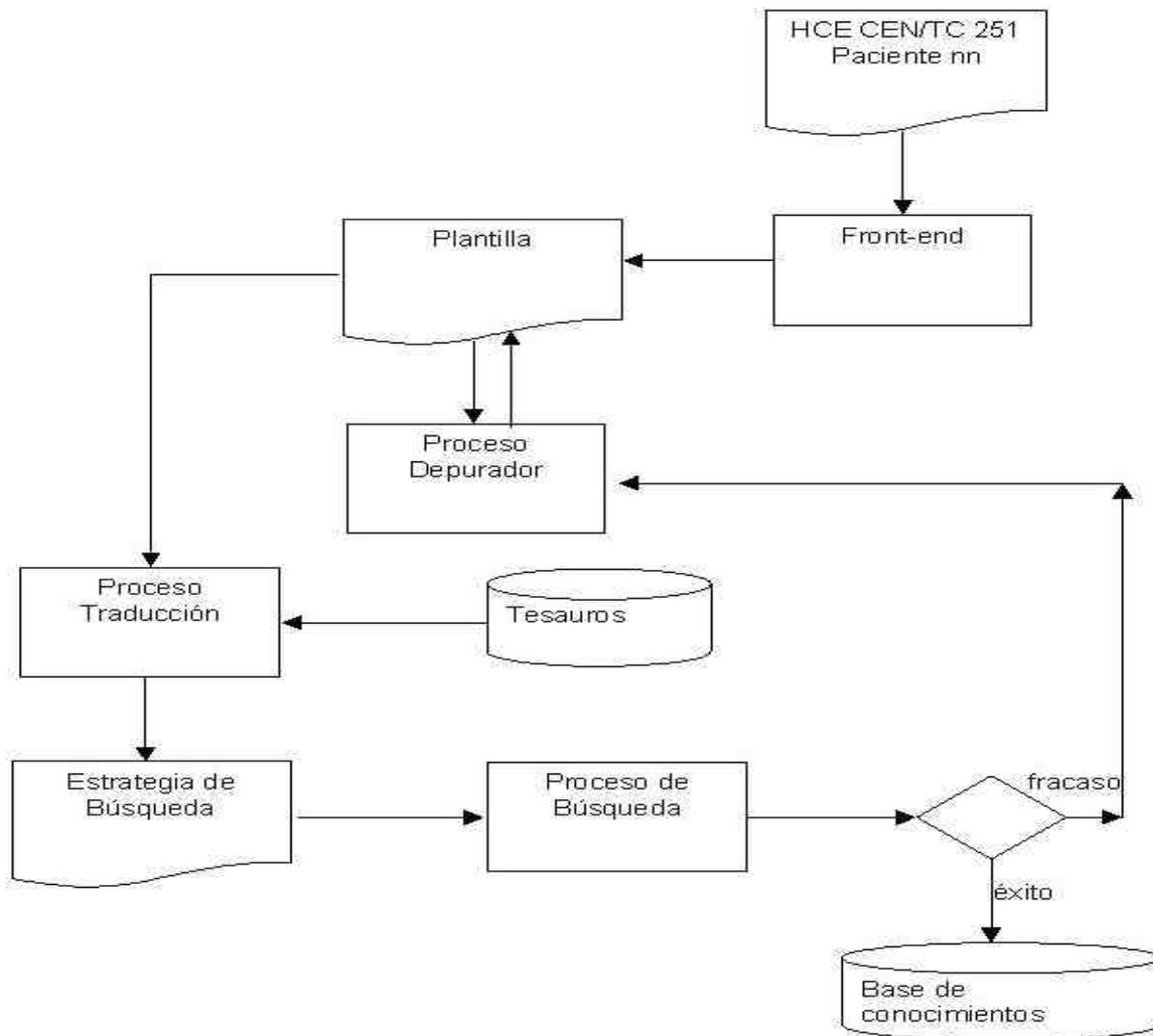
El proceso de adquisición del conocimiento en este sistema de búsqueda, tiene dos directrices, una inherente a la propia historia clínica electrónica (HCE) y otra relativa al front-end. Ambas están asociadas, pero en lo específico, la que involucra la HCE, se refiere a tratar de especificar una organización de dicha historia clínica que contribuya a facilitar la tarea de producir unas plantillas de búsqueda de literatura científica. La amplitud de criterio que supone la organización de la HCE según la pre-norma 13606 de la CEN251, representa una oportunidad importante para conseguir este objetivo. En este sentido, una parte significativa de este trabajo se ha dirigido a relacionar el conocimiento requerido para realizar búsquedas en la web, compatibles con una organización de la HCE que simplifique la tarea de definir los descriptores de la búsqueda. La otra parte de la adquisición del conocimiento está orientada a proporcionar un front-end que le permita al usuario construir una plantilla de búsqueda conforme a las expectativas de dicho usuario. Los elementos considerados relevantes por un determinado usuario pueden variar, pero un front-end que guíe la estrategia de búsqueda del profesional puede ser de utilidad si está relacionado con las preguntas que se hace el usuario a la hora de producir la plantilla de búsqueda, es decir, el foco de interés en términos de establecer el diagnóstico, el tratamiento y la evaluación.

El modelo de historia clínica utilizado en este trabajo, intenta cumplir una doble finalidad. Por una parte, permite ingresar la información en lenguaje natural para posibilitar al médico la descripción del caso clínico en sus propias palabras y con el adecuado grado de fidelidad; pero también permite incorporar los ítems que emplean lenguaje normalizado para el procesamiento automatizado de la información con diversos propósitos: elaboración de estadísticas, selección de casos para trabajos experimentales, alarmas para toma de decisiones, y para el objetivo del presente trabajo: la búsqueda automatizada de bibliografía referente al caso según los cánones de la Medicina Basada en la Evidencia. Esta historia clínica contiene por lo tanto, ítems cuyos valores posibles están predeterminados en listas de términos estandarizados; en algunos casos se emplean códigos aceptados internacionalmente (medicamentos, diagnósticos, síntomas, etc).

Los datos normalizados pueden ser datos «simples», ya sean síntomas (*cefalea, dolor precordial*), hallazgos del examen físico (*soplo cardíaco sistólico, rales crepitantes*), una característica aislada del trazado electrocardiográfico (*ritmo de base sinusal, localización de extrasístole supraventricular*), etc. ; ó datos “globalizadores”, que son los que contienen conceptos de mayor complejidad; son conclusiones a las que el médico arriba luego de analizar los datos simples mencionados anteriormente.

## ARQUITECTURA DEL SISTEMA DE BUSQUEDA

El perfil clínico o estado del paciente y la pertinencia del conocimiento que debe proporcionar la evidencia buscada, conlleva la necesidad de que los sistemas de búsqueda deban operar asociados a un contexto clínico específico. Esta característica es esencial para la funcionalidad adecuada de los sistemas de búsqueda, ya que implica una interoperabilidad con la historia clínica de cada paciente. La HCE parece el nexo adecuado a esta circunstancia [4][5]. En este trabajo se ha considerado oportuna la transformación de historias clínicas existentes en formato electrónico al modelo de la historia clínica electrónica establecido por la pre-norma 13606 del grupo de trabajo CEN/TC 251.



Arquitectura del sistema de búsqueda

En una sesión de trabajo del sistema, el primer paso es la generación de la plantilla básica de búsqueda. Este paso lo realiza el usuario mediante un front-end donde define en primer término el tipo de documentos que le interesa recuperar, como ser Guidelines, consensus, MBE reviews, clinical trials; el ó los lenguajes de interés para la consulta; el tipo de fuente a la que desea consultar: motores de búsqueda (Google, Yahoo, Altavista), bases de datos (Medline, Lilacs, Cochrane), journals (JACC, J. Am. Coll. Cardiol., Circulation, Revista Española de Cardiología, European Heart Journal). A continuación se le presenta al usuario una lista de los datos globalizadores presentes en la HCE para que éste elija los que considere relevantes. Si el usuario lo cree necesario puede también examinar y seleccionar datos simples de la HC.

A partir de los términos incluidos en la plantilla básica inicial, el sistema prepara la estrategia de búsqueda. Esto lo realiza a través del módulo Proceso Traducción, el cual mediante el uso de tesauros y ontologías expande semántica y multilingualmente cada concepto.

Supongamos por ejemplo que en la HCE de un paciente el médico selecciona el dato globalizador: «angina inestable», e incluye este término en la plantilla de búsqueda. El módulo Proceso de Traducción tomará esta frase y la ampliará multilingual y semánticamente.

Para la expansión multilingual se pueden utilizar diccionarios y/o glosarios multilinguales que traducen el concepto al ó los idiomas de interés. Supongamos en este caso que además del español, se desean trabajos escritos en inglés, obteniendo así “unstable angina”.

Para la expansión semántica se pueden utilizar tesauros y/u ontologías. El término obtenido para esta consulta, utilizando por ejemplo el tesoro de la Biblioteca Nacional de Medicina de Estados Unidos, es “Angina, unstable”.

Estos tres conceptos: “angina inestable”, “unstable angina” y “Angina, unstable” se combinan con el operador lógico O para ampliar la posibilidad de búsqueda. Por lo tanto, la estrategia de búsqueda resultante para este ejemplo será:

“angina inestable” O “unstable angina” O “angina, unstable”

La información de interés será cualquiera que contenga al menos uno de estos tres conceptos.

Esta ampliación multilingual y semántica se realiza para cada uno de los conceptos elegidos por el usuario. El módulo Proceso Traducción combina con el operador lógico Y las expansiones de todos los conceptos a una estrategia de búsqueda genérica, la cual es luego transformada, en forma transparente para el usuario, a la sintaxis de búsqueda de cada uno de los distintos tipos de fuentes.

Para este ejemplo, en el caso de bases de datos podría accederse a: Medline, que es la base de datos bibliográfica de la Biblioteca Nacional de Medicina de Estados Unidos; Excerpta Medica, producida por la editorial Elsevier; ó los Current Contents de Ciencias de la Salud, producido por el Institute for Scientific Information. En el caso de páginas web fijas, una posibilidad es predeterminedar una página como la de Amazon (www.amazon.com), para luego hacer búsquedas en ella. En el caso de páginas Web variables, una posibilidad es utilizar un buscador, como ser Google (www.google.com).

La estrategia traducida a Medline será:

angina inestable OR unstable angina OR angina, unstable

Y la estrategia traducida a la sintaxis propia de Google será:

“angina inestable” “unstable angina” “angina, unstable”

Cada una de estas fuentes ejecutará la consulta y devolverá un resultado.

El módulo Proceso de Búsqueda envía esta estrategia a la ó las fuentes determinadas al inicio de la sesión y se obtiene como respuesta un conjunto de referencias bibliográficas ordenadas según su importancia, que son presentadas al usuario, el cual evaluará si el resultado obtenido se adecúa a su necesidad de información. Si esta respuesta no es satisfactoria, el usuario tiene la posibilidad de volver a realizar el Proceso de Depuración de conceptos a partir de la última plantilla utilizada. El usuario podrá repetir el procedimiento, quitando o agregando nuevos conceptos, y ejecutando la nueva estrategia de búsqueda, hasta obtener un resultado satisfactorio. Cuando el usuario obtiene un conjunto de documentos que satisface su consulta, el sistema almacena la estrategia original construida automáticamente por el sistema y la estrategia de búsqueda asociada modificada por el usuario que produjo el resultado satisfactorio.

Ante el mismo requerimiento de información, distintos usuarios en distintos momentos realizarán distintas modificaciones a la plantilla presentada por el sistema. Todos estos cambios asociados a una plantilla ó estrategia, son almacenados para su posterior análisis. Es de interés determinar qué estrategia es la mayormente utilizada para una determinada necesidad de información; por ejemplo detectando qué conceptos tienen en común las distintas estrategias asociadas a un problema. Este análisis de ponderación de los conceptos, permitirá modificar la plantilla básica que en futuras utilizaciones el sistema presentará a los usuarios, adaptando el sistema a la constante evolución natural en que se ve inmersa la ciencia.

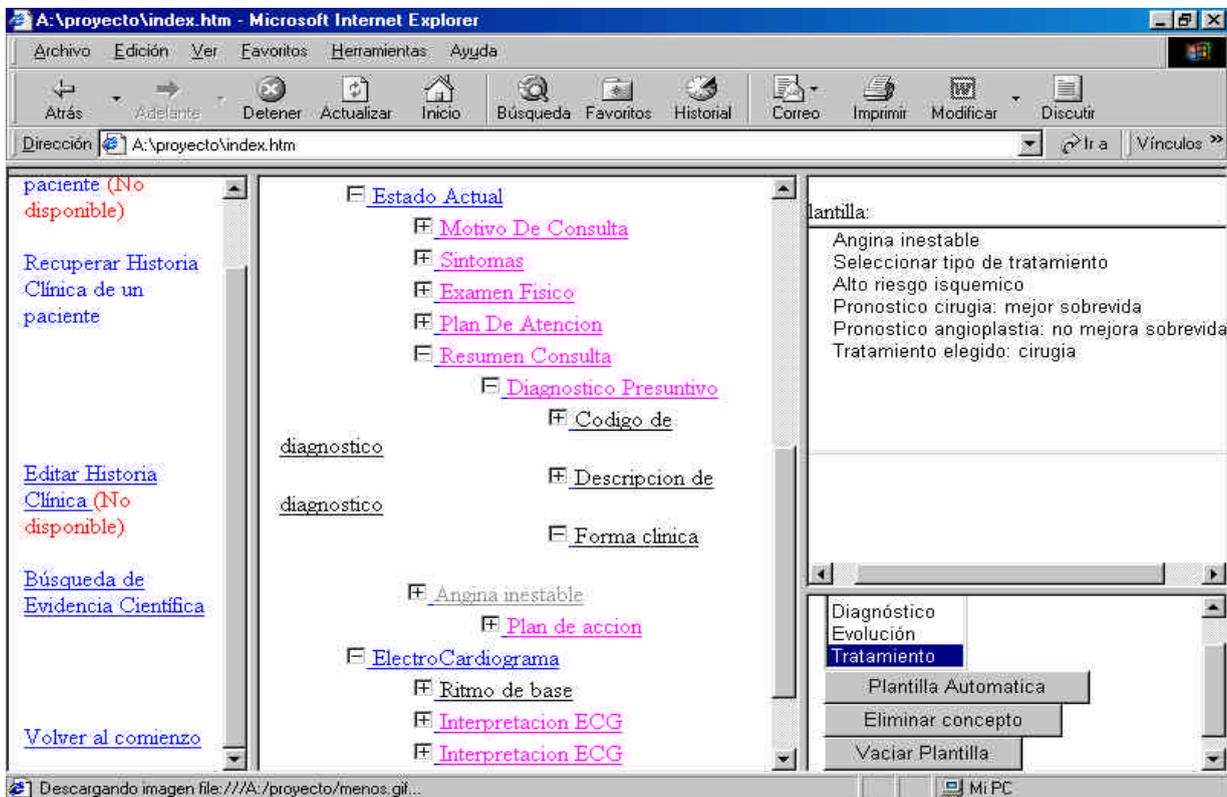
## LA INTERFASE CON EL USUARIO Y EL MECANISMO DE GENERACIÓN DE CONSULTAS

Uno de los puntos a considerar en la arquitectura propuesta, es el diseño de la interfase con el usuario, lo cual plantea un desafío desde el siguiente punto de vista: la utilización de operadores lógicos aumenta la precisión en la búsqueda, pero la mayoría de los usuarios no los conoce o no sabe utilizarlos adecuadamente, y por ende no los aplica, dando como resultado búsquedas ineficientes.

El diseño del front-end debe proveer mecanismos fáciles de ser comprendidos y utilizados por el usuario y que a su vez permitan al software generar las operaciones lógicas necesarias.

Por otro lado, la sintaxis y la estructura de la consulta dependen de la fuente a consultar, por esto es necesario encontrar una manera normalizada y uniforme de describir las consultas, que a su vez sea fácilmente transformable en función de la fuente de información.

A todo esto hay que agregar el hecho de que la consulta que llega a la fuente será previamente procesada de alguna manera, por ejemplo agregando o modificando las palabras a consultar debido a la consulta previa de un diccionario o un tesoro. Esto de por sí implica una primera transformación de la consulta.



## CONCLUSIONES

Actualmente el modelo propuesto se está implementando con un conjunto de historias clínicas disponible en las patologías de cardiopatía isquémica, diabetes y enfermedades respiratorias (EPOC). El proceso de conversión de estas historias clínicas a la preforma CEN 251 se realiza de forma simplificada y ajustando los arquetipos a los requerimientos que permitan obtener las plantillas de búsqueda. A tal efecto, se ha desarrollado una plantilla en XML Schema organizado por consultas o visitas y con la capacidad para contener los atributos que puedan ser seleccionados para desarrollar las plantillas de búsqueda. En el futuro se plantea la realización de esta conversión a través de servicios *middleware* que realicen las operaciones de forma dinámica.

Esta arquitectura, además, resuelve los problemas vinculados con búsqueda de información en la web. Estos problemas están relacionados con la ubicación distribuida, la calidad, la redundancia, la falta de estructura y la heterogeneidad semántica y/o estructural de los datos. La resolución del problema de la ubicación distribuida se realiza contemplando la consulta a los distintos tipos de fuentes posibles, la calidad de la información proveniente de páginas web a través de la clasificación de las mismas, la redundancia eliminando los duplicados en la integración, la falta de estructura estructurando los documentos al estándar XML, la heterogeneidad estructural integrando las estructuras mediante XML y la heterogeneidad semántica integrando la información mediante ontologías.

Mediante la resolución de estos problemas, la arquitectura presentada permite aumentar la Sensibilidad de la búsqueda, indicador que da la proporción de los documentos relevantes que son recuperados y la Precisión, que indica el ratio de documentos relevantes sobre el número total de documentos recuperados; que son los dos indicadores de la recuperación de información.

## BIBLIOGRAFÍA

- [1] Salvador H. C. “*Normas CEN para la comunicación de la Historia Clínica Electrónica. ENV 13606-I Arquitectura de la HCE. Componentes de la Arquitectura*”. Informática y salud, N° 30, pp 1547-1553. 2001
- [2] Evidence-Based Medicine Working Group. “*Evidence-Based Medicine. A new approach to teaching the practice of medicine*”. JAMA1992;268:2420-5, 1992.
- [3] Sackett DL, Rosenberg W, Muir JA, Haynes RB, Richardson WS. “*Evidence based medicine: what it is and what it isn't*”. BMJ; 312: 71-72, 1996.
- [4] Sim I et al. Clinical Decision Support Systems for the Practice of Evidence-based Medicine, JAMIA, Vol 8, N 6, pg 527-534, Dec 2001.
- [5] Yamazaki S & Satomura Y, “*Standard Method for Describing an Electronic Patient Record Template: Application of XML to Share Domain Knowledge*”, Method Inform Med, 2000.
- [6] Plüss J et al, Búsqueda de literatura clínica a partir de una historia clínica electrónica conforme al pre-estándar ENV 13606, Actas del XX Congreso Anual de la Sociedad Española de Ingeniería Biomédica, Zaragoza (España), 27-29 noviembre, 2002.