

Estudio comparativo entre bases de datos temporales y bases de datos relacionales aplicado a historias clínicas electrónicas

Juan Esteban Díaz Montejo
Universidad Pedagógica y
Tecnológica de Colombia
juan.diaz@uptc.edu.co

Daniel Felipe Bellon Cely
Universidad Pedagógica y
Tecnológica de Colombia
daniel.bellon@uptc.edu.co

Juan Sebastián
González Sanabria
Universidad Pedagógica y Tecnológica
de Colombia
juansebastian.gonzalez@uptc.edu.co

(Tipo de Artículo: Artículo de Reflexión. Recibido el 05/03/2015. Aprobado el 12/07/2015)

Resumen. La elección de un paradigma de base de datos cuando se crea un sistema, es una decisión que establece el curso del proyecto, entender las ventajas y desventajas de estos paradigmas proporciona el punto de partida. Por lo anterior el trabajo reflexiona sobre la comparación entre las bases de datos relacionales y bases de datos temporales evaluando su desempeño en un caso práctico: Gestión de historias clínicas electrónicas. El proceso se inicia con la selección de características y sub-características deseables del DBMS, posteriormente el diseño del caso práctico, y una selección de los gestores que representan cada paradigma. Todo esto con el fin de proporcionar un marco para proyectos futuros que se encuentran dentro del área de Telemedicina, registros médicos electrónicos y los sistemas de salud en general que implique el almacenamiento de información temporal.

Palabras clave. Historias Clínicas Electrónicas, Bases de Datos Temporales, Bases de Datos Relacionales.

A comparative study between temporary databases and relational databases applied to Electronic Health Records

Abstract. Choosing a database model when a system is developed is a decision that establishes the course of a project, and understanding the advantages and disadvantages of these models provides the starting point. This article reflects on the comparison between relational and temporary databases by evaluating their performance in a case study: the management of electronic medical records. The process starts by selecting desirable features and sub-features of the DBMS and the subsequent design of the case study and the choice of representative managers of each model. All of this having as purpose to provide a framework for future projects within the field of telemedicine, electronic medical records and mostly health systems involving temporary data storage.

Keywords. Electronic Health Records (EHR's), Temporal Databases, Relational Databases.

1. INTRODUCCIÓN

La existencia de múltiples paradigmas de bases de datos abre el debate y la discusión respecto a las ventajas y desventajas que cada uno de ellos puede brindar a un caso de estudio particular. Bases de datos relacionales, bases de datos temporales, bases de datos no relacionales (NoSQL), entre otros; se convierten sin lugar a dudas en una amplia gama de opciones con las que cuentan desarrolladores y administradores de bases de datos. Es evidente que cada paradigma puede llegar a ser más eficiente que otro si se compara en un caso específico, por tal razón, es necesario definir un escenario de pruebas donde exista una aplicación concreta que pueda ser evaluada con la implementación de los paradigmas a comparar.

Las soluciones, plataformas y sistemas de información que soportan funcionalidades y aplicaciones relacionadas con telemedicina, necesitan día a día mayor gestión de información, con base en

lo anterior y teniendo como foco el rendimiento de los paradigmas de bases de datos orientados a un caso de estudio particular que se relaciona directamente con sistemas para telemedicina, se propone evaluar y comparar el rendimiento de bases de datos relacionales y temporales en plataformas para la gestión de Historias Clínicas Electrónicas (HCE). De esta manera se puede obtener un marco de diseño que exponga de manera clara y concisa, con resultados y sustento teórico, las ventajas y desventajas de cada uno de los paradigmas a comparar, para brindar un punto de partida que proporcione una visión objetiva respecto al rendimiento de una base de datos que almacene y gestione HCE.

2. REFERENTE TEÓRICO

A continuación se describen los conceptos y teorías involucrados en el desarrollo de la investigación, además de trabajos y experiencias anteriores que

sirven como fundamento y base para que los resultados obtenidos sean lo más fiables posibles.

2.1 Bases de Datos Relacionales (BDR)

Las bases de datos relacionales son las más populares para brindar soporte de datos en los diferentes sistemas de información, mostrándolos a manera de entidades y relaciones. Este modelo fue creado por Edgar Frank Codd a finales de los años 60's bajo los objetivos de independencia física, independencia lógica, flexibilidad, uniformidad y sencillez [1].

Los sistemas de gestión de bases de datos relacionales (SGBDR) brindan un soporte tanto a la definición de los datos como a la manipulación de los mismos, mediante la estructura de los datos y las operaciones del modelo relacional, respectivamente.

Para realizar consultas en las BDR es utilizada el álgebra relacional, la cual se basa en la teoría de conjuntos, el estudio de esta permite entender que servicios de consulta debe proveer un lenguaje relacional y a comprender la construcción del lenguaje de consulta estructurado (SQL, *Structured Query Language*). Los lenguajes de consulta basados en álgebra relacional son lenguajes procedimentales que tienen que seguir una serie de pasos para construir, mediante las operaciones de álgebra relacional, la consulta mediante las relaciones almacenadas [2].

El lenguaje más utilizado sin lugar a duda es SQL, este lenguaje combina el álgebra relacional y el cálculo relacional, predominando el último en la construcción de este, por lo que SQL se define como un lenguaje declarativo [2].

2.2 Bases de Datos Temporales (BDT)

Las bases de datos temporales son aquellas que administran los datos considerando la variación del tiempo en los mismos, partiendo del hecho de que el tiempo es una variable importante en la información, y que convencionalmente las bases de datos representan el estado de la información en un solo instante de tiempo, existen sectores dentro de los que se incluyen las finanzas, la medicina y el entorno gubernamental, que necesitan representar su información en un tiempo pasado; razón por la cual, durante los últimos veinte años se han presentado modelos de bases de datos temporales, con el fin de representar la evolución histórica de los datos [3].

Para el modelamiento de las BDT se puede hacer una extensión del modelo relacional, adicionando atributos temporales a cada relación. Así mismo, para

lo que refiere a la consulta de los datos existen lenguajes especializados para estas bases de datos, como lo son TQUEL y SQL3.

Se pueden ver dos estados importantes de los datos en las BDT, en el tiempo válido, que es el periodo de tiempo en el que el hecho es real y se permite una serie de operaciones como la eliminación, la inserción y la modificación en cualquier instante de tiempo; y el tiempo de transacción, el cual describe el punto en el tiempo en que fue ingresada la información, soportando operaciones como actualizaciones únicamente en el estado actual (no se puede modificar el pasado), volver a un estado anterior "*Rollback*", y eliminación lógica puesto que no se puede eliminar físicamente la información para conservar el histórico de los datos [4-5].

2.3 Historias Clínicas Electrónicas (HCE)

La información clínica, asociada a pacientes médicos, necesita día a día de mayor accesibilidad e interoperabilidad con el fin de lograr una gestión adecuada, además de comunicación entre equipos biomédicos [6]. Por otro lado, los pacientes requieren cada vez mayor acceso a su historia clínica con el fin de obtener más participación en lo que respecta a su atención médica [7].

Debido a lo anterior surge la necesidad de construir sistemas de información que brinden herramientas adecuadas para que el personal involucrado en el manejo y gestión de historias clínicas pueda cumplir su tarea de manera exitosa, es así como nace el concepto de HCE.

La información que se integra en este tipo de sistemas de información, necesita sin lugar a duda de estandarización logrando, que tanto las imágenes obtenidas a partir de dispositivos biomédicos, como la terminología usada en la descripción de patologías y procedimientos pueda ser interpretada por cualquier plataforma independientemente de la tecnología con la que esté construida. En este sentido, *Health Level Seven* (HL7) ha sido la organización que más ha ahondado en su esfuerzo para la consolidación y divulgación de estándares aplicables a las HCE.

3. METODOLOGÍA

Para el proceso de comparación, que se realiza teniendo como objeto de estudio la gestión de HCE, se ha definido un marco de trabajo que desglosa el flujo de actividades que se realizan con la finalidad de obtener unos resultados veraces y confiables. En primera instancia, se realiza un diseño concreto del caso de estudio teniendo como base fundamental el

estándar HL7/CDA para la construcción de las HCE's, para posteriormente implementarlo usando un SGBDR y un SGBDT.

El proceso inicia con la definición y caracterización de los criterios de evaluación, seguido de la especificación de escalas y ponderaciones de evaluación, para finalmente determinar el grado de cumplimiento que cada SGBD presenta con relación a los criterios de evaluación definidos en un trabajo de comparación de SGBD realizado previamente por Castro en [8].

3.1 Definición criterios de evaluación y subcaracterísticas

Se toman como base las características y atributos de calidad especificados por la norma ISO 25000, pero se hace una selección rigurosa haciendo énfasis a las características deseables en un SGBD y a las características que necesita un gestor de HCE.

3.2 Criterios de evaluación

Para la evaluación se definieron los siguientes criterios:

- Seguridad de la información: Integridad y Autenticidad.
- Compatibilidad: Coexistencia e Interoperabilidad.
- Eficacia: Flexibilidad de entrada de datos y Multilenguaje.
- Mantenibilidad: Facilidad de cambio y Variables e indicadores.
- Portabilidad: Instalabilidad, Multiplataforma y Adaptabilidad.
- Funcionalidad: Granularidad, Almacenamiento de datos históricos, Operaciones, Tiempo de respuesta a consultas y Complejidad del lenguaje de consulta.

3.3 Ponderaciones

Se establecieron una serie de categorías para determinar la relevancia de los criterios de evaluación, como se muestra en la Tabla 1.

TABLA 1
Nivel de relevancia [8]

Ponderación	Valor
Indispensable	4
Importante	3
Necesario	2
Deseable	1

3.4 Grado de cumplimiento

Con el objeto de dejar a un lado la subjetividad, se establece una escala que determina el grado de cumplimiento de cada SGBD respecto a cada característica o criterio de evaluación, la escala de calificación se muestra en la Tabla 2.

TABLA 2
Grado de cumplimiento [8]

Calificación	Grado de cumplimiento
4.7 - 5.0	Pleno
4.0 - 4.6	Alto grado
3.0 - 3.9	Aceptable
2.0 - 2.9	Insatisfactorio
1.0 - 1.9	Deficiente
0.0 - 0.9	No cumple

Para cada criterio de evaluación debe existir una calificación (C_i) con su determinada ponderación (P_i), con la sumatoria del producto de estos valores, dividido en la sumatoria de las ponderaciones, se obtiene la calificación ponderada, como lo muestra la Ecuación (1) [8].

$$C_{pon} = \frac{C_1P_1 + C_2P_2 + C_3P_3 + \Lambda + C_iP_i}{P_1 + P_2 + P_3 + \Lambda + P_i} = \frac{\sum_{i=1}^n C_iP_i}{\sum_{i=1}^n P_i} \quad (1)$$

3.5 Selección de Herramientas

Los resultados de la comparación dependen en gran medida de las tecnologías que se utilicen para la implementación del caso de estudio en los paradigmas definidos, debido a que sus características pueden ajustarse y presentar un desempeño más ventajoso respecto a los criterios de evaluación que se han establecido.

Por lo anterior, es necesario realizar una selección de un SGBD que represente a cada paradigma en la comparación, dicha selección de herramientas, para el caso del presente trabajo, obedece a un análisis comparativo que tiene como punto de partida un estudio de tendencias de búsqueda, que provee una perspectiva global del uso y la popularidad de las tecnologías, hecho que sirve como precedente para la búsqueda de estudios detallados que tengan como base las tecnología con mayor potencial de uso para seleccionar aquella cuyo rendimiento sea el indicado.

Selección SGBDR. Para esta selección en primera instancia se limitaron las SGBD a las dos con mayores tendencias de búsqueda en los años comprendidos entre el 2006 y 2013, arrojando como resultado Oracle y MySQL, como se muestra en la Figura 1.

Como última medida de selección se evaluaron las características funcionales de los gestores, obteniendo que MySQL 5.1 sobrepasa a Oracle 10g, en cuanto eficiencia, mantenibilidad, compatibilidad, portabilidad, y derechos de propiedad, lo anterior soportado en un estudio que evaluaba estas características 0 a 5, y en el cual MySQL alcanzó un puntaje total de 4.68, mientras que Oracle obtuvo 4.51 [8].

Selección SGBDT. PgFoundry es un proyecto que permite extender el lenguaje SQL a TSQL2 para la implementación de BDT, incluyendo tipos de datos, operadores y funciones propios de las BDT.

Teniendo en cuenta que los estudios sobre SGBD temporales son muy escasos, se acudió a las tendencias de búsqueda, como se ilustra en la Figura 2, obteniendo que PgFoundry (Extensión de PostgreSQL) en comparación con Time Series Cartridge y TimeDB es mucho más buscada.



Fig. 1. Tendencias de búsqueda de SGBDR. Fuente: Google Trends.

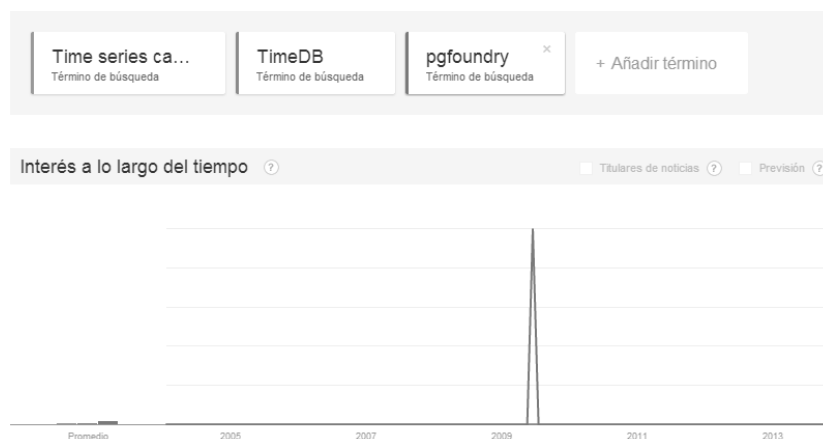


Fig. 2. Tendencias de búsqueda TimeDB y Time Series Cartridge. Fuente: Google Trends

3.6 Diseño del Caso de Estudio

La información que debe almacenarse para cada HCE, ha sido seleccionada tomando como base la arquitectura CDA definida por el estándar HL7 en su versión 3 [9-10], haciendo una depuración para trabajar con los datos más influyentes al momento de evaluar el performance de cada SGBD. De lo anterior se puede obtener que la información a almacenar es: Identificación de la HCE, Paciente (Identificación,

nombre, fecha de nacimiento, dirección y teléfono), Doctor (Identificación, nombre, especialidad), Diagnóstico, Fórmula y Observaciones.

Con base en la información anterior, se define el modelo entidad-relación, Figura 3, para su implementación en el SGBDR y para su posterior evaluación.

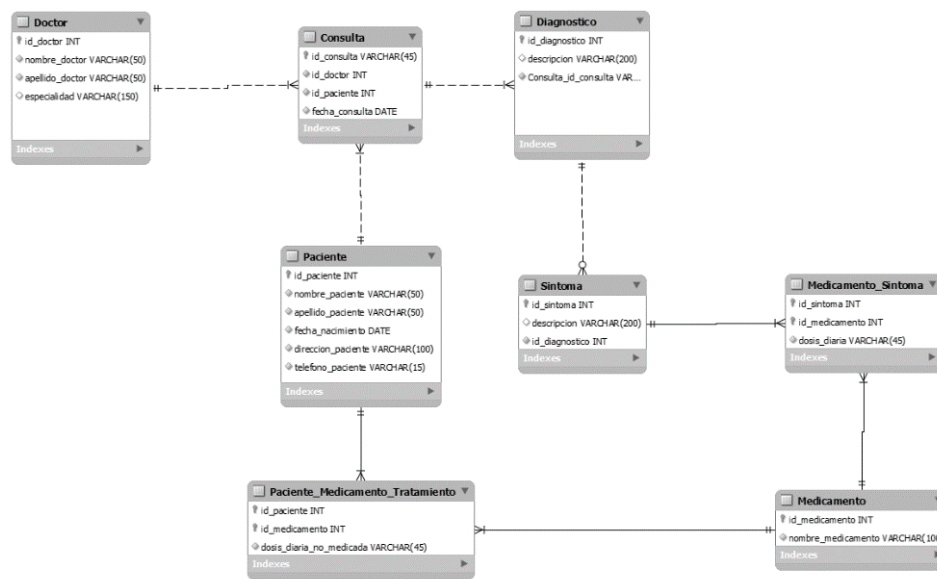


Fig. 3. Modelo Entidad-Relación de la base de datos relacional.

Mientras que para el caso de SGBDT se definió el modelo expuesto en la Figura 4, en donde una de las principales diferencias que cabe resaltar es el uso del tipo

de dato PERIOD, el cual permite un mejor manejo de intervalos de tiempo para el tratamiento de este tipo de información.

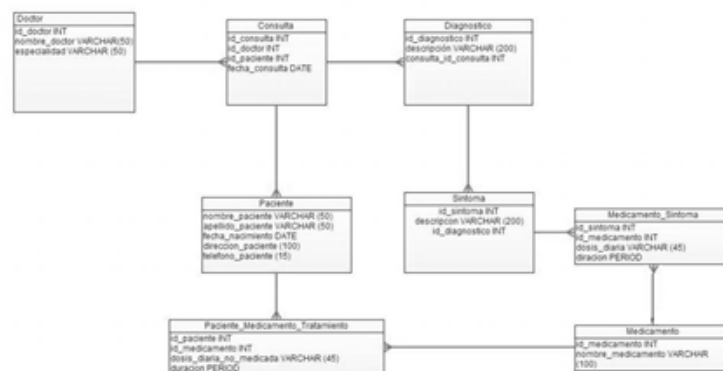


Fig. 4. Modelo de la base de datos temporal del Caso de Estudio

4. RESULTADOS

Luego del diseño del caso de estudio que se implementó en cada SGBD, se aplicaron una serie de pruebas en un conjunto de datos de aproximadamente 500 registros, con el fin de obtener la ponderación adecuada para cada uno de los criterios de evaluación que se seleccionaron para determinar el desempeño de cada paradigma de bases de datos en lo que respecta a la gestión de Historias Clínicas Electrónicas.

Como pruebas en los gestores de bases de datos, tanto relaciones como temporales, se realizaron principalmente consultas que se presentan en ámbitos médicos como por ejemplo “medicamentos que un paciente tiene que consumir en determinado periodo de tiempo”, para de esta manera medir los tiempos de respuesta y consistencia de los resultados retornados.

En cuanto a la calificación de la seguridad se tuvo en cuenta estudios previamente realizados, que brindaron información de las bondades y deficiencias que puedan poseer [11]. Para el caso de la compatibilidad y

portabilidad se tuvo en cuenta los instaladores para los diferentes sistemas operativos (Windows, Linux, MacOS), de esta forma se garantiza que puedan ejecutarse en dichas plataformas y coexistir en las mismas. En cuanto a mantenibilidad se basó en la facilidad para migración de la información, lo cual se facilita por el hecho de manejar SQL en los dos casos, contando con la única diferencia en cuanto las PgFoundry, maneja una estructura de datos diferente que dificulta la migración de información a otra SGBD.

Es de destacar que debido a la caracterización que se realizó para determinar los comportamientos deseables para el caso de estudio, se establecieron dichos criterios de evaluación en función de una serie de subcaracterísticas; en consecuencia, la calificación final de cada criterio dependió de la calificación final de cada subcaracterística obtenida de la evaluación de las pruebas realizadas a cada una. En la Tabla 3 se observa la relación entre las ponderación de cada subcaracterística con respecto al grado de cumplimiento.

TABLA 3
Ponderación y calificación de acuerdo a las subcaracterísticas (1/2)

Criterio	Subcaracterística	Ponderación	Grado de cumplimiento	
			MySQL	PgFoundry
Seguridad de la información	Integridad	3	4.5	4.8
	Autenticidad	3	4.9	4.9
Compatibilidad	Coexistencia	2	4.3	4.3
	Interoperabilidad	3	4.4	4.2
Eficacia	Flexibilidad de entrada de datos	2	4.6	4.5
	Multilinguaje	2	4.8	4.9
Mantenibilidad	Facilidad de cambio	3	4.8	4.0
	Variables e indicadores	3	4.7	4.9

TABLA 4
Ponderación y calificación de acuerdo a las subcaracterísticas (2/2)

Criterio	Subcaracterística	Ponderación	Grado de cumplimiento	
			MySQL	PgFoundry
Portabilidad	Instalabilidad	4	4.9	3.9
	Multiplataforma	3	5.0	5.0
	Adaptabilidad	3	4.9	4.5
	Granularidad	3	4.4	4.4
	Almacenamiento de datos históricos	4	4.2	4.8
Funcionalidad	Operaciones	4	4.7	4.9
	Tiempo de respuesta a consultas	3	4.5	4.5
	Facilidad del lenguaje de consulta	3	4.9	4.3

En la Tabla 4 se presenta la calificación ponderada de cada característica, y la calificación total de cada SGBD, la cual se obtuvo del promedio de la calificación de las características.

TABLA 5
Calificación final

Criterio	Resultado ponderado	
	MySQL	PgFoundry
Seguridad de la información	4.70	4.85
Compatibilidad	4.36	4.24
Eficacia	4.70	4.70
Mantenibilidad	4.75	4.45

Criterio	Resultado ponderado	
	MySQL	PgFoundry
Portabilidad	4.93	4.41
Funcionalidad	4.53	4.61
Total	4.66	4.54

Adicionalmente, y con fines de demostrar las pruebas realizadas, en las Figuras 5 y 6 se ilustran ejemplos de las consultas realizadas para medir los criterios como eficacia y funcionalidad. Es conveniente resaltar que no se presenta evidencia de cada calificación, pues no es el objeto del trabajo

```

1 select nombre_paciente, apellido_paciente, dosis_diaria_no_medicada, nombre_medimento
2 from paciente p, paciente_medimento_tratamiento pmt, medicamento m
3 where p.id_paciente=pmt.id_paciente
4 and pmt.id_medimento=m.id_medimento
5 and m.id_medimento=1;

```

Fig. 5. Consulta SGBDR.

```

1 select nombre_paciente, apellido_paciente, dosis_diaria_no_medicada, nombre_medimento, duracion
2 from pacientes p, pacientes_medimento_tratamiento pmt, medicamentos m
3 where p.id_paciente=pmt.id_paciente
4 and pmt.duracion>='01-06-2013'
5 and pmt.id_medimento=m.id_medimento
6 and m.id_medimento=1;

```

Fig. 6. Consulta SGBDT.

5. DISCUSIÓN

El proceso de comparación y los resultados obtenidos evidencia que existen características a favor de los SGBDR y otras que benefician al momento de usar un SGBDT. Las tablas 3 y 4 permiten dar al lector una visión que acorde a los criterios que requieran de mayor importancia dentro del proyecto a desarrollar. Para las Historias Clínicas Electrónicas se sugiere, sin ser restrictivo, el uso de MySQL como SGBD, debido principalmente a la variedad de tipos de datos que se usan a nivel Médico. Otro factor que impacta al recomendar el uso de MySQL, es que los SGBDT, pese a ofrecer un mayor conjunto de operaciones en cuanto al manejo del tiempo, aún se encuentran en desarrollo lo que hace, que en caso de requerir una migración, se convierta en un proceso de mayor tiempo.

6. CONCLUSIONES

En cuanto a integridad de la información se puede observar que el SGBDT tiene un mejor desempeño dado a que permiten una mayor serie de operación en cuanto a restricciones de temporales.

Tanto el SGBDT como el SGBDR, manejan los mismos cuidados en cuanto a autenticidad de la información, mediante la creación de usuarios y control de acceso.

El desempeño del SGBDR y el SGBDT en coexistencia con otros sistemas es el mismo, pero en cuanto a interoperabilidad de sistemas los SGBDR tienen un mejor rendimiento.

Se puede encontrar una mayor flexibilidad de entrada de datos en el SGBDR, puesto que existen más tipos de fuentes para poblar las bases de datos; así como una mayor facilidad de cambio por la diversidad de opciones para migrar la información a otras bases de datos.

Se observan más tipos de variables e indicadores en el SGBDT, debido a que se basa en TSQL2, el cual es una extensión de SQL estándar que brinda tipos de datos y operaciones propias para el manejo de datos temporales.

7. TRABAJOS FUTUROS

El presente trabajo es la base para el desarrollo del sistema de gestión de historias clínicas del proyecto "Plataforma de servicios médicos" desarrollado para la Maestría en Tecnología Informática.

8. REFERENCIAS

- [1] C. J. Date. "Introducción a los sistemas de bases de datos", séptima edición. Editorial Prentice-Hall, 2001.
- [2] D. Costal-Cost, "El modelo relacional y el álgebra relacional". Editorial UOC, 2009.
- [3] E. Nina, P. N. Hubler, M. M. Moro, G. Demartini, "A temporal database management system implemented on top of a conventional

- database," Computer Science Society, 2000. SCCC '00. Proceedings. XX International Conference of the Chilean, vol., no., pp.58,67, 2000.
- [4] K. Kulkarni, J. Michels. "Temporal features in SQL: 2011". ACM SIGMOD Record, 2012.
- [5] M. Al-Kateb, et al. "Temporal Query Processing in Teradata", Genoa, Italy. 2013.
- [6] R. Smith, "What clinical information do doctors need?", BMJ, vol. 313, pp.1062-1068, 1996.
- [7] D. Kalra, "Electronic Health Record Standards", IMIA Yearbook of Medical Informatics 2006.
- [8] J. D. Castro, Tesis de Grado "Modelos de calidad orientados a los sistemas gestores de bases de datos". Universidad Pedagógica y Tecnológica de Colombia. 2009.
- [9] HL7 en español, Documentos clínicos electrónicos HL7 CDA R2 - Estructura general. Disponible en: <http://hl7es.blogspot.com/2011/04/documentos-clinicos-electronicos-hl7.html>.
- [10] E. de la Cruz, D. M. López, "Guía de implementación HL7 del estándar CDA para el almacenamiento de historias clínicas de acuerdo con los requerimientos de la Fundación HL7 Colombia", Cali, Colombia. 2011.
- [11] K. Shashi, "A generalized model for a relational temporal database". Proceeding SIGMOD '88 Proceedings of the 1988 ACM SIGMOD international conference on Management of data.