

SLD265 PROPUESTA DE SERVIDOR DE TERMINOLOGÍAS MÉDICAS PARA LAS APLICACIONES DEL CENTRO DE INFORMÁTICA MÉDICA

SLD265 MEDICAL TERMINOLOGIES SERVER PROPOSAL FOR MEDICAL INFORMATICS CENTER'S SOFTWARES

Alberto Bradshaw González

Universidad de las Ciencias Informáticas, Cuba, bradsg@uci.cu, Carretera a San Antonio de los Baños, Km. 2 ½, Municipio Boyeros, Ciudad de la Habana

RESUMEN: *En el Centro de Informática Médica (CESIM) de la Universidad de las Ciencias Informáticas, se desarrollan aplicaciones de la salud que son desplegadas en una misma institución, esta puede ser un consultorio médico, un hospital o un centro de investigación biomédico. Estas aplicaciones se comunican entre sí e intercambian información pero no son capaces de interpretar o entender los datos provenientes de terminologías médicas diferentes a las que usan. Por lo que surge la necesidad de crear un servidor de terminologías que reúna los vocabularios médicos usados por estos sistemas y que brinde servicios de consultas y traducción entre estos. Con el presente trabajo cuyo objetivo es garantizar la interpretación de mensajes entre aplicaciones médicas, se propone construir un servidor de terminologías médicas que sea utilizado en el CESIM. Para alcanzar este objetivo se estudiaron sistemas y tecnologías principalmente de código abierto que se relacionan con servidores terminológicos o que componen uno en sí. A partir del estudio realizado se escogió un conjunto de herramientas que integradas son de gran utilidad para la construcción del sistema. También se seleccionaron los servicios necesarios que debería implementar y ofrecer este software. Se propone implementar además, una interfaz de usuario que brindará la posibilidad de probar cada una de las funciones.*

Palabras Clave: Interfaz, servicios web, servidor de terminologías, terminologías, vocabulario.

ABSTRACT: *At the Medical Informatics Center (CESIM in Spanish) of the University of Informatic Sciences, health applications are developed that are deployed in a single institution. A health application may be a clinic physician, a hospital or a biomedical research center. These applications communicate with each other, and they exchanged information. But they are not able to interpret or understand terminologies data from medical different to those used. Then, arise the necessity of create a server of terminologies that gather medical vocabularies used by these systems and to provide consultation and translation between these service. With the present work whose objective is to ensure the interpretation of messages between medical applications, intends to build a terminology medical server to be used in the CESIM. To achieve this objective were studied systems and main technologies of open source that relate to servers or that compose one itself. From the study it was chosen a set of integrated tools that are very useful for the construction of the system. Also, were selected the necessary services should implement and offer this software. It also proposes to implement a user interface that will provide the possibility to prove each one of the functions.*

Keywords: Interface, terminologies server, terminology.

1. INTRODUCCIÓN

La descentralización de la atención médica en redes asistenciales ha generado la necesidad de conectar múltiples sistemas informáticos. Los registros clínicos del paciente se encuentran dispersos entre los diferentes niveles de atención. Como consecuencia los recursos de datos están distribuidos y son heterogéneos, creando un gran desfase entre el valor potencial y el valor real de la información presente en la historia clínica de la persona. Ya no basta con mantener el registro único en los sistemas internos de una organización de salud, sino que es necesario ir más allá de los muros de una institución, permitiendo un fluido intercambio de datos clínicos.

Conseguir un historial clínico electrónico que incluya todos los datos existentes sobre un paciente con independencia del lugar donde haya sido generada o dónde se encuentre almacenada y mantenida, requiere la interoperabilidad semántica de los sistemas de información.

La interoperabilidad semántica es la capacidad de los sistemas informáticos de comunicar, incorporar y usar información generada por sistemas externos [1]. Para que esa información tenga un significado claro, no ambiguo, es necesario describirla mediante anotaciones terminológicas que sean accesibles desde ambos sistemas, el que generó los datos y el que los recibe.

Existen sistemas diferentes no sólo entre centros de salud de países distintos, sino incluso entre los que son de una misma zona, y departamentos de la misma institución. La existencia de sistemas diseñados sin coordinación implica un modelo diferente para la representación de la información. Además, la enorme amplitud del dominio médico ha dado lugar a multitud de terminologías diferentes diseñadas para su aplicación en sub-dominios concretos.

Cada terminología abarca un dominio y está diseñada para ser utilizada en un ámbito distinto. No existe una que sea universal, como tampoco existe una estructura que permita representar completamente todo el conocimiento, incluso si se limita a tratar únicamente con el vocabulario médico. Además, una terminología de tal magnitud resultaría inmanejable, debido al volumen de datos necesario para abarcar todo el dominio completo el cual no deja de crecer.

El problema de la interoperabilidad semántica tiene como solución emplear una terminología de referencia común con la que todos los sistemas puedan comunicarse, manteniendo al mismo tiempo sus propias terminologías locales intactas y generando mapeos (correspondencias) entre éstas y aquella. Además, el uso de una interfaz que proporcione un

conjunto de métodos comunes de acceso a estas terminologías para ser usados por diferentes sistemas de información [2].

En el Centro de Informática Médica (CESIM) de la Universidad de las Ciencias Informáticas (UCI) se desarrollan aplicaciones especializadas en las diferentes áreas de atención sanitaria. Estas difieren en cuanto al lenguaje de programación en la que son desarrolladas y servidor de aplicaciones; aunque pueden ser ejecutadas tanto en la plataforma GNU/Linux como en Windows. Manejan distintas terminologías como bases de conocimiento para cubrir las necesidades del área en la que son desplegadas, generando diversidad de recursos terminológicos en caso de ser utilizadas en el mismo lugar.

2. ANÁLISIS DE LAS SOLUCIONES EXISTENTES

Después de haber realizado el estudio para conocer los sistemas existentes relacionados con los estándares internacionales para la interoperabilidad entre sistemas de salud, se encontraron varias terminologías médicas utilizadas en diferentes ámbitos de la medicina. Se identificaron también herramientas y especificaciones de acceso a terminologías, además de servidores que definen métodos para realizar consultas a estas.

A partir del estudio realizado se propone el desarrollo del servidor de terminologías para que sea usado por las distintas aplicaciones del CESIM. Se definen las distintas terminologías médicas que albergará el sistema así como las herramientas y tecnologías que se usarán para su implementación. Se especifican también los servicios necesarios que brindará y las funciones de administración que se ejecutarán a través de una interfaz de usuario propuesta.

2.1 Terminologías médicas

2.1.1 Clasificación Internacional de Enfermedades

La Clasificación Internacional de Enfermedades (CIE) se creó por la Organización Mundial de la Salud (OMS) en 1893 y recibe mantenimiento desde esa fecha. Se originó como una clasificación de causas de muerte (mortalidad) pero se utiliza también para representar morbilidad.

La CIE está diseñada para promover la comparabilidad internacional en la recopilación, procesamiento, clasificación y la presentación de las estadísticas de mortalidad. La OMS propone normas y convenciones que el usuario ha de aplicar, para que mediante análisis estadístico puedan obtenerse datos

coherentes y comparables a nivel local, nacional e internacional. Entre ellas figura un formato para comunicar las causas de muerte en el certificado de defunción. Las afecciones notificadas se traducen luego a códigos. Se mejora así la utilidad de las estadísticas de mortalidad, dando preferencia a ciertas afecciones, consolidándolas y seleccionando sistemáticamente una única causa de muerte entre una secuencia de varias afecciones.

En algunos países se adaptó para ajustarla mejor a las necesidades locales, en los Estados Unidos de América (EE.UU) por ejemplo, se agregaron códigos para permitir su utilización para la facturación, y crearon la Modificación Clínica (CM). La versión 10 (CIE-10) no tiene procedimientos, sólo diagnósticos, y sus categorías no son compatibles con versiones anteriores, ya que se pasó de una organización numérica de los capítulos a una alfanumérica. En gran parte del mundo CIE-10 es el estándar para reportes de Mortalidad y Morbilidad [3].

2.1.2 Clasificación Internacional de Atención Primaria

Es una clasificación creada por la Sociedad Mundial de Médicos de Familia (WONCA por sus siglas en inglés) exclusivamente para ser utilizada en atención primaria. Es la evolución de más de 20 años de experiencia de WONCA en clasificaciones de este nivel de atención.

Es una referencia obligada no sólo para la investigación, sino también en el proceso de informatización de la historia clínica, al permitir codificar las actividades y elementos del episodio de atención al paciente. Está compuesta por alrededor de 700 códigos, con un nivel de detalle muy apropiado para su uso en el área ambulatoria por médicos generalistas, ya que permite la recolección y análisis de tres importantes componentes de la consulta médico-paciente: la razón de la consulta, el problema atendido y el proceso de atención. Posee un sólido mapeo (relación - asociación) con CIE-10 que permite utilizar CIAP como un método de acceso a CIE-10 [4].

2.1.3 SNOMED CT

La Nomenclatura Sistematizada de Medicina – Términos Clínicos (SNOMED – CT por sus siglas en inglés) es una nomenclatura médica que provee un idioma común y consistente para recolectar, almacenar, recuperar y analizar datos clínicos de diferentes especialidades y regiones geográficas. Desarrollada inicialmente por el Colegio Americano de Patólogos (CAP) en el año 2002, incluye términos de todo el espectro de la medicina, incluso de veterinaria.

SNOMED CT es el vocabulario más rico que existe para describir hallazgos clínicos, enfermedades, procedimientos, sustancias, etc. Sus características principales son:

- Su enfoque composicional.
- Sus funcionalidades de vocabulario de interfaz y de referencia.

SNOMED CT es una excelente nomenclatura para utilizar como vocabulario de referencia, por su nivel de detalle y calidad de las relaciones semánticas. Incluye funcionalidades de interface, es decir que para cada concepto tiene varias descripciones posibles, que pueden ser utilizadas como vocabulario de usuarios (también llamados de interface) en una implementación.

SNOMED CT posee:

- 311.000 conceptos (edición del 2011 en idioma inglés).
- 724.000 descripciones (edición del 2011 en idioma español).
- Más de 920.000 relaciones.

SNOMED CT ofrece un estándar para la información clínica. Las aplicaciones de software pueden utilizar los conceptos, jerarquías y relaciones como un punto de referencia en común para el análisis de datos. Es útil para médicos, pacientes, administradores y desarrolladores de software. Puede brindar a los proveedores en el área de salud información más accesible y completa relacionada con el proceso asistencial (historias clínicas, enfermedades, tratamientos, resultados de laboratorio, etc.), lo que redundará en mejores resultados para los pacientes. Para los proveedores de atención médica, una terminología clínica como SNOMED CT posibilita la identificación de pacientes de acuerdo con cierta información codificada en sus registros y, en consecuencia, facilita su seguimiento y tratamiento [5].

2.1.4 UMLS

El Sistema de Lenguajes Médicos Unificados (UMLS, por sus siglas en inglés) es un proyecto creado por la Biblioteca Nacional de Medicina (NLM, por sus siglas en inglés) de los Estados Unidos de América en 1989. Es usado para unificar las diferentes terminologías médicas existentes. Posee tres componentes:

- El MetaThesaurus.
- Una taxonomía o red semántica de tamaño reducido (poco más de un centenar de clases).
- Un conjunto de herramientas léxicas sólo disponibles en inglés.

El MetaThesaurus es un gran compendio de diccionarios donde las palabras son términos clínicos y

los diferentes idiomas/diccionarios son las diferentes terminologías que las contienen (SNOMED CT, ICD-9-CM, MeSh). La propia NLM se encarga de integrar las terminologías más relevantes en el MetaThesaurus (unas 150 en la última versión del 2009), así como de su actualización y mantenimiento [6].

La red semántica brinda información sobre el significado de los conceptos y las herramientas léxicas. El conjunto de herramientas léxicas está destinado a facilitar el procesamiento del lenguaje natural, alimentado con las palabras incluidas en el MetaThesaurus.

El objetivo del NLM al crear UMLS era facilitar el acceso homogéneo a terminologías y, además, facilitar la creación de herramientas semánticas para aprovechar estos recursos. Por ello, además del identificador de concepto, a los términos importados se les asigna otro atributo importante, el Tipo Semántico (o clase semántica), que los enlaza con la Red Semántica de UMLS. Esta asociación permite diferenciar entre significados de un mismo término según el contexto en el que aparece.

Las terminologías médicas poseen como desventaja que no disponen por sí solas de herramientas de búsqueda sobre sus términos, por lo que el uso de estas implica la implementación o reutilización de aplicaciones de acceso desarrolladas por terceros. Se construyen sin coordinación entre desarrolladores de otros vocabularios provocando diferencias en cuanto a estructura y formatos. Todas excepto SNOMED-CT están dirigidas a registrar los conceptos relacionados con un área específica de la medicina, siendo necesario el uso de esta para establecer relaciones de equivalencia cuando se integran sistemas que utilizan diferentes terminologías.

2.2 Herramientas de acceso a terminologías

2.2.1 Bioportal

Bioportal es una herramienta del Centro Nacional de Ontologías Biomédicas (NCBO, por sus siglas en inglés) de EE.UU creada en el 2008. Este buscador permite tanto la carga de terminologías y ontologías por red a una base de datos como el acceso posterior a las mismas mediante una página web, o mediante funciones de servicio web. Esta característica hace que puedan programarse herramientas que utilicen estos servicios, de forma gratuita, aunque para poder añadir aportes propios a la base de datos se precisa crear una cuenta en la comunidad NCBO.

Las funciones de servicio web de este buscador son genéricas para permitir el acceso a distintas terminologías y mapeos entre ellas. Posee como desventaja que crear una herramienta basándose

en él, requiere modificar las funciones que se implementarían para hacer búsquedas en las terminologías a través de un servicio web. La herramienta accedería al servicio web de Bioportal, sobre el que no se tendría control alguno [7].

2.2.2 ClineClueXplorer

Fue desarrollado por ClinicaInformationConsultancyLtd de los Estados Unidos de América durante el año 2009. Es uno de los buscadores gratuitos descargables vía web más conocidos. Requiere registro en línea, pero también ofrece una versión de prueba de 10 días sin registro. El programa obtiene la terminología SNOMED mediante descarga desde un servidor de actualizaciones propio, y permite que el usuario seleccione la distribución que le resulte más conveniente. No sólo permite búsquedas, también permite la navegación de SNOMED mediante un visor en el que se puede, interactivamente, saltar a través de sus términos siguiendo las relaciones que aplican a cada uno de los términos [8].

Ofrece también servicios web de búsqueda sobre su servidor de terminología. Además de las búsquedas por palabras, que ya ofrecen los buscadores en red, permite también filtrar los resultados por jerarquías. Tiene como desventajas que no permite el uso de otras terminologías, ni el acceso de herramientas externas a sus funciones.

2.3 Especificaciones de servicios de terminologías

2.3.1 CTS

Los Servicios de Terminologías Comunes (CTS por sus siglas en inglés) es una iniciativa desarrollada por la organización de estándares HL7 con sede principal en EE.UU, en el año 2003. Su versión 2 se publicó en el 2009. Define una Interfaz de Programación de Aplicaciones que puede ser utilizada por un software basado en la versión 3 de HL7 cuando accede a contenido terminológico [9]. Fue desarrollado como una alternativa a una estructura de datos común. En lugar de especificar a lo que una terminología externa debe parecerse, CTS identifica las características funcionales comunes que este tipo de vocabulario debe ser capaz de proporcionar. Por ejemplo, un servicio compatible con HL7 tendrá que ser capaz de determinar si un código de concepto dado es válido dentro de un recurso particular. Sus beneficios son:

- El software que use CTS se puede implementar sin la necesidad de comprender la base de datos de la terminología de cada proveedor.
- Oculta gran parte de la complejidad inherente a los sistemas modernos de la terminología.

- Los requerimientos funcionales de los servicios de terminología se especifican claramente.
- La implementación puede estar basada en bases de datos y software existentes.
- Es punto de entrada común a los contenidos de la terminología.
- Es usado para traducir conceptos de código de un sistema a otro.

CTS posee como desventaja que solo define las funcionalidades básicas que un servidor de terminologías debería proporcionar y no especifica cómo estas deben ser implementadas. Funciones más avanzadas son especificadas en un servidor más sofisticado como LQS.

2.3.2 TQS/LQS

Los Servicios de Consultas de Léxicas (LQS, por sus siglas en inglés), también conocido como Servicios de Consultas Terminológicas (TQS, por sus siglas en inglés), es un completo servidor de terminología publicado por el Grupo de Gestión de Objetos (OMG, por sus siglas en inglés) en el año 1997.

LQS se encuentra en su versión 1.0, está compuesto por 13 módulos y especifica un conjunto de métodos comunes y de solo lectura para acceder a contenido de terminologías médicas. Los métodos se encuentran especificados en Lenguaje de Descripción de Interfaces (IDL por sus siglas en inglés). Esta especificación define cómo las funciones deberían ser invocadas desde una amplia variedad de lenguajes de programación y describe el comportamiento esperado de estas.

La especificación de LQS describe un modelo de referencia para sus servicios. El propósito de este modelo es definir varias entidades que aparecen en la definición IDL de los servicios y no describir cuál debería ser la estructura de una terminología o como debería ser construida. Representa una unión común de los aspectos de diferentes sistemas terminológicos. Aborda un amplio espectro de vocabularios médicos representados como una simple tabla con los atributos código-valor o como una terminología sofisticada, ejemplo, SNOMED-CT [10].

Desventajas:

- El soporte del proveedor es mínimo.
- Su especificación es considerada demasiado densa y compleja.
- Solo permite acceso a las terminologías y no su modificación. Ha sido diseñado para la navegación y no para la inserción de los conceptos.
- No se conoce implementación de este servidor.

2.4 Servidores de Terminologías

2.4.1 LexGrid y LexBIG

LexGrid es un modelo de datos y una implementación de la Clínica Mayo de EE.UU para almacenar de forma estándar vocabularios controlados y ontologías. Define un formato y una representación para vocabularios y representa con exactitud una gran variedad de terminologías y otros recursos léxicos. Este modelo, además, define varios mecanismos de almacenamiento (bases de datos relacionales, LDAP) y formato XML.

LexGrid permite caracterizar las partes fundamentales de las terminologías de manera que todas pueden descomponerse y conceptualizarse para ser traducidas al modelo de datos. Este modelo común es la parte fundamental del proyecto. Una vez la información de vocabularios dispares se encuentre representada bajo la misma representación estándar, es posible construir repositorios compartidos para contenido terminológico e interfaces comunes y herramientas para acceder y gestionar el contenido.

A partir de este modelo de datos, la comunidad CancerBiomedicalInformaticsGrid (caBIG) refinó la implementación de la Clínica Mayo para dar lugar a la herramienta LexBIG. LexBIG es un servidor diseñado para almacenar y acceder a terminologías. Se trata de un proyecto específico que aplica la tecnología diseñada en el modelo LexGrid. Su objetivo es lograr que los vocabularios puedan ser accedidos mediante una API bien estructurada, capaz de acceder y distribuir recursos. El servidor se crea a partir de tecnologías estándares bien conocidas.

La herramienta LexBig permite generar una base de datos cargada con contenido terminológico para el uso de diferentes API cuyas funciones tiene ya implementadas (el CTS de HL7 y la propia API de LexBIG), aparte de tener sus propias herramientas de búsqueda.

Basado en el modelo de información de LexGRID para representar y compartir recursos terminológicos a gran escala, LexBIG agrupa también una serie de programas para cargar, indexar, publicar y editar el contenido de las terminologías. Posee una interfaz gráfica que permite tanto la carga de terminologías como la búsqueda de conceptos empleando sus propias API. Se pueden emplear las funciones implementadas de CTS con LexBIG, para esto es necesario descargar un paquete aparte proporcionado en la página de descargas del proyecto LexBIG, denominado "CTS deploy" [11].

LexBIG presenta como desventaja que al centrarse en funciones de búsqueda estándares para todas las terminologías que pretende soportar, carece de funciones para aprovechar las características espe-

cíficas de cada terminología.

2.4.2 Cazador de Términos Hunter

El Cazador de Términos Hunter (C.T Hunter) es un servidor de términos de SNOMED desarrollado por la Escuela Técnica Superior de Ingeniería informática de Valencia, en el año 2010. Constituye una herramienta para acceder a esta terminología y realizarle consultas. C.T Hunter usa tecnologías de probada utilidad (UMLS, CTS y LexGrid), accede a la terminología SNOMED CT extraída a partir de la distribución disponible en UMLS y realiza consultas avanzadas sobre la misma aprovechando las características propias de este vocabulario. Las funciones que posee implementadas son de consulta, las cuales se publican como servicios web para ser accedidos y utilizados por otras herramientas.

Posee implementados métodos de búsqueda tanto básicos como avanzados adaptados a las características propias de SNOMED (atributos, relaciones, jerarquías de conceptos). Están basados en las funciones de CTS usadas por el proyecto LexBIG. Las funciones avanzadas permiten realizar búsquedas semánticas complejas, por tipos de términos y atributos.

La búsqueda por palabras permite listar conceptos cuyas definiciones o propiedades contengan una cierta cadena. Sus funciones propias permiten introducir restricciones, ejemplo, buscar las palabras "fractura de hueso" en la propiedad nombre de un concepto y que tenga como atributo "sitio de hallazgo = torso".

Tiene una Interfaz Gráfica de Usuario que permite el uso de sus funcionalidades, la cual está basada en SWING, una librería gráfica para java. Está dividida en dos partes, una para los menús y las búsquedas y la otra para mostrar los resultados. Su aspecto está diseñado para parecerse a un navegador web, con la posibilidad de navegar por las pestañas, cerrar aquellas búsquedas que ya no sean útiles, y acceder fácilmente a los datos de un concepto referenciado en una de las pestañas.

C.T Hunter tiene como desventajas que las funciones propias que implementa no están disponibles para descarga y son solo de búsquedas, no de edición. El proyecto está destinado solamente a trabajar con una sola terminología, SNOMED-CT, por lo que sus métodos son específicos para realizar consultas sobre este vocabulario aprovechando sus características [12].

Luego de haber realizado un análisis sobre las terminologías, especificaciones e implementaciones de servidores y herramientas de acceso desarrolladas para la salud, como parte de los resultados de la investigación, se puede concluir que:

- Las terminologías médicas no ofrecen herramientas de búsqueda.
- Los buscadores no comparten sus funciones con otras aplicaciones.
- Existen servicios de terminologías incompletos ya que solo brindan las funcionalidades básicas que los vocabularios médicos deberían proporcionar.
- Las implementaciones de servidores terminológicos son escasas.
- Se requiere la integración de estas tecnologías para obtener un servidor de terminologías completo.

3. PROPUESTA DEL SISTEMA

3.1 Propuesta de Servidor de Terminologías

Se propone implementar un servidor terminológico que garantice la interoperabilidad semántica entre los diferentes sistemas del CESIM, después de haberse investigado acerca de las diferentes herramientas libres de acceso a terminologías. Se tomará como guía el modelo de datos propuesto por la especificación CTS para caracterizar las partes fundamentales de los sistemas de codificación que se utilizarán, de manera que todos puedan descomponerse y conceptualizarse en un mismo modelo. Esta representación también se empleará para facilitar la implementación de los métodos de consultas a estos vocabularios.

El servidor almacenará todos los sistemas de codificación usados en el Centro incluyendo la terminología más extensa que existe, SNOMED-CT, de la cual se aprovechará todo el potencial que ofrece con su organización jerárquica. SNOMED-CT se empleará como referencia entre el resto de los vocabularios médicos. Estas terminologías serán extraídas de la distribución más reciente de UMLS disponible para descarga desde internet, donde se encuentran enlazadas y presentan mapeos (correspondencias) entre ellas. Posteriormente se cargarán a un sistema gestor de base de datos relacional mediante la herramienta LexBIG, la cual también debe ser descargada.

Se implementarán funciones que permitirán realizar consultas y operaciones de administración sobre la distribución terminológica. Deben implementarse inicialmente las funciones de búsqueda básica que garantizarán el funcionamiento del servidor y luego las de búsqueda avanzada que permitirán realizar consultas semánticas complejas, ejemplo, mediante el filtrado por atributos.

Una búsqueda avanzada resulta listar sólo los conceptos en la jerarquía "Procedimiento" que tengan un atributo "Con uso de dispositivo" cuyo valor sea

"Endoscopio"). Las funciones estarán basadas en las API de CTS y LQS. Se tomará como referencia la implementación de CTS que posee el proyecto LexBIG. Estas funcionalidades serán publicadas como servicios web para que sean consumidos por las diferentes aplicaciones del CESIM. Se construirá una interfaz de usuario para realizar las operaciones de administración y probar las funcionalidades de los métodos a implementar.

3.1.1 Funciones de Consultas

Se propone descargar la última distribución de UMLS correspondiente al año 2009 y extraer de esta las terminologías que se usan en el CESIM (CIE-9, CIAP-2, LOINC), además de SNOMED-CT, la cual se usará como referencia. Para la extracción se usará la herramienta MetamorphoSys obteniendo los sistemas de codificación en formato RichReleaseFormat (RRF), propio de UMLS.

Se utilizará UMLS por ser un proyecto de libre uso y disponible para descarga, cuyo objetivo es unificar e integrar las diferentes terminologías médicas existentes y que establece relaciones de equivalencias entre sus términos. Esta característica permitirá implementar las funciones de mapeo del servidor terminológico.

SNOMED-CT se usará como referencia por su elevado nivel de detalle y calidad de las relaciones semánticas, además de ser el vocabulario más amplio que existe para describir hallazgos clínicos, enfermedades, procedimientos y sustancias. La carga de los datos se realizará en el sistema de base de datos relacional PostgreSQL utilizando la herramienta LexBIG. Se escoge PostgreSQL por ser el sistema gestor de base de datos que forma parte de la línea de desarrollo del CESIM. El proceso de carga es lento y el consumo de los recursos de hardware del ordenador es alto, por lo que se requiere una computadora con elevado nivel de procesamiento de información para realizar esta operación.

3.1.2 Base de Datos

Se deben implementar inicialmente las funciones de consultas básicas que garantizarán el funcionamiento del servidor, permitiendo a las aplicaciones realizar búsquedas sencillas sobre él. Posteriormente se implementarán los métodos de búsquedas avanzadas y de difícil implementación que garantizarán el entendimiento de los datos intercambiados entre las aplicaciones. Se utilizarán como referencia las funciones de CTS implementadas por LexBIG para así adquirir y aprovechar la lógica de estos métodos. Los datos devueltos por estas funciones deberán codificarse en XML, para el futuro

acceso de los sistemas de salud a estas funcionalidades mediante servicios web.

Para poder emplear las funciones implementadas de CTS con LexBIG es necesario descargar un paquete proporcionado en la página de descargas del proyecto LexBIG, denominado "CTS deploy", y descomprimirlo sobre el directorio de instalación de esta herramienta.

Se definen las siguientes funciones básicas que serán fáciles de implementar y que garantizarán el funcionamiento inicial del servidor:

Consultas básicas

ObtenerNombreServicio:

Entrada:

Salida: Nombre del servicio.

Descripción: Devuelve el nombre del servicio asignado.

ObtenerVersionServicio:

Entrada:

Salida: Versión del servicio.

Descripción: Devuelve la versión del servicio.

ObtenerDescripcionServicio:

Entrada:

Salida: Descripción del servicio.

Descripción: Devuelve la descripción del servicio.

ObtenerAlgoritmosBusquedasSoportados:

Entrada:

Salida: Lista de algoritmos de búsqueda soportados.

Descripción: Devuelve la lista de nombres de algoritmos de búsqueda implementados en el servicio.

ObtenerSistemasCodigos:

Entrada:

Salida: Lista de terminologías soportadas y sus descripciones.

Descripción: Devuelve la lista de terminologías disponibles en el servidor.

Consultas avanzadas

LlenarDetalles:

Entrada: código del concepto y código del lenguaje objetivo.

Salida: Nombre del concepto, sistema de codificación y versión del sistema de codificación.

Descripción: Completa los datos del concepto suministrado.

BuscarDesignacion:

Entrada: Identificador del sistema de código, código de concepto y código del lenguaje destino.

Salida: Texto definitorio.

Descripción: Devuelve la designación preferida del código de concepto en el lenguaje suministrado.

BuscarCodigoDeConceptoPorDesignacion:

Entrada: Texto a buscar y algoritmo, id de la terminología, id del lenguaje, flag indicando si los conceptos inactivos son válidos o no.

Salida: Lista de códigos de conceptos encontrados (emparejados con el identificador de su terminología).

Descripción: Devuelve la lista de códigos de conceptos cuyas designaciones se correspondan con la cadena de texto buscada, en el lenguaje introducido (si los códigos existen).

BuscarDesignaciones:

Entrada: Identificador de la terminología, código de concepto, texto a buscar, algoritmo, lenguaje objetivo.

Salida: Lista de nombres.

Descripción: Devuelve todos los nombres asociados al código.

CoincidenciaDeFrase:

Entrada: cadena de texto, nombre del sistema de codificación.

Salida: Lista de conceptos.

Descripción: Devuelve todos los conceptos que contengan la cadena de texto suministrada ordenados por grado de coincidencia.

TraducirCodigo:

Entrada: Código de concepto a ser traducido, sistema de código destino.

Salida: Lista de sinónimos.

Descripción: Traduce el concepto codificado a una forma que usa el sistema de código destino.

TraducirCodigos:

Entrada: Lista de códigos de conceptos a ser traducidos, sistema de código destino.

Salida: Lista de sinónimos.

Descripción: Traduce los conceptos codificados a formas que se usan en el sistema de código destino.

ListarConcepto:

Entrada: Identificador del sistema de códigos.

Salida: Lista de todos los conceptos.

Descripción: Lista todos los códigos de conceptos registrados en un sistema de códigos.

EsConceptoValido:

Entrada: Identificador de la terminología, código de concepto, flag que indica si los conceptos inactivos se consideran válidos.

Salida: Verdadero o falso.

Descripción: Determinar si el concepto es válido en una terminología específica.

SonCodigosRelacionados:

Entrada: Identificador del sistema de códigos, identificador del código fuente, identificador del código objetivo, código de relación, calificador de relación y flag que indica si usar solo relaciones directas o también relaciones por transitividad.

Salida: Verdadero o falso.

Descripción: Determina si la relación especificada existe entre los códigos fuente y objetivo.

BuscarExpansionCodigo:

Entrada: Identificador de la terminología, código de concepto, código de relación, indicador de la dirección de la relación, código de lenguaje objetivo, tiempo límite y tamaño límite.

Salida: Lista de expansión jerárquica del código.

Descripción: Lista recursiva de los códigos de concepto relacionados con el concepto proporcionado, incluyendo los nombres preferidos para los códigos.

BuscarAntecedentes:

Entrada: Código de concepto, número de generaciones, máximo de resultados a devolver.

Salida: Lista de códigos de conceptos.

Descripción: Devuelve los antecesores del concepto indicado, según la cantidad de generaciones que se le pida consultar.

BuscarDescendientes:

Entrada: Código de concepto, número de generaciones, máximo de resultados a devolver.

Salida: Lista de códigos de conceptos.

Descripción: Devuelve los descendientes del concepto indicado, según la cantidad de generaciones que se le pida consultar.

SonHermanos:

Entrada: Dos códigos de conceptos.

Salida: Verdadero o falso.

Descripción: Comprueba si los dos conceptos introducidos son hijos del mismo padre.

ObtenerGeneracion:

Entrada: Código de concepto, número de generación, flag indicando si se buscan antecesores o descendientes.

Salida: Lista de conceptos.

Descripción: Devuelve una generación concreta de padres o hijos del concepto indicado.

3.1.3 Interfaz Gráfica de Usuario

Se propone construir una interfaz gráfica de usuario usando el patrón de arquitectura Modelo Vista Con-

trolador para realizar las operaciones de administración y probar las funcionalidades de los métodos a implementar. Esta interfaz será una aplicación web basada en el lenguaje de programación Java. Para su construcción se utilizarán el servidor JBoss y el framework Seam, los cuales son libres, de código abierto y se usan para albergar e implementar respectivamente aplicaciones web basadas en Java.

SeamJBoss incluye otros frameworks que se deberán utilizar para este proyecto. Estos son la biblioteca de clases Richfaces para la vista de la interfaz e Hibernate para el mapeo entre la base de datos relacional y el modelo de objetos de la aplicación.

Funciones de administración:

AdicionarUsuario

Descripción: Permitirá añadir usuarios a la aplicación.

ModificarUsuario

Descripción: Se usará para modificar los datos de un usuario.

EliminarUsuario

Descripción: Se usará para eliminar un usuario del sistema.

IncluirTerminología

Descripción: Permitirá incluir una nueva terminología en la lista de terminología disponibles en el servidor, una vez que esta sea descargada y convertida mediante las herramientas UMLS y LexBIG respectivamente.

EliminarTerminología

Descripción: Permitirá eliminar una terminología de la lista de terminología disponibles en el servidor si se determina que esta no será usada por ningún sistema ni por el propio servidor.

CambiarEstadoTerminología

Descripción: Permitirá activar o desactivar una terminología y de esta forma restringir o no el acceso a las funciones que brindará el servidor para esta.

ObtenerTerminología

Descripción: Se usará para obtener la información detallada de una terminología.

ObtenerTerminologías

Descripción: Se usará para obtener la lista de terminologías soportadas por el servidor.

BuscarCoincidencia

Descripción: Se usará para obtener la lista de conceptos que coincidan con una cadena de texto suministrada.

ListarConceptos

Descripción: Devolverá todos los términos de una terminología específica.

TraducirConcepto

Descripción: Traducirá un concepto codificado a una forma que usará una terminología destino.

ValidarConcepto

Descripción: Determinará si un concepto es válido en una terminología específica.

El sistema propuesto en esta investigación y la integración de las tecnologías que lo conforman se representan en la siguiente figura:

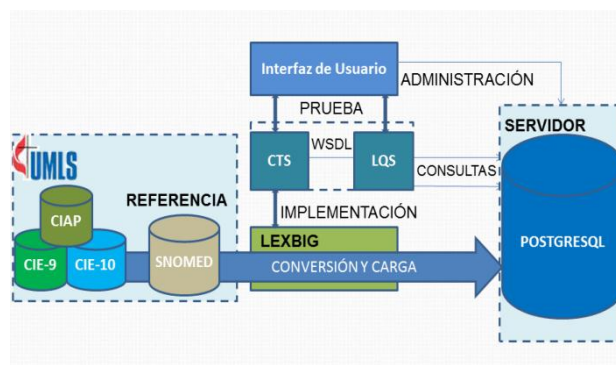


Figura. 1: Sistema propuesto

4. CONCLUSIONES

- El análisis del estado de los sistemas de salud del CESIM permitió definir la necesidad de entendimiento durante el intercambio de mensajes entre ellos.
- Se seleccionó UMLS como grupo de terminologías que estará disponible en el servidor y sobre el cual se realizarán las operaciones de consultas.
- Se desarrolló la propuesta de sistema que permita la interpretación de mensajes entre las aplicaciones del CESIM que utilizan distintas terminologías.
- La propuesta de sistema incluye funcionalidades que permitirán a los usuarios hacer solicitudes de consultas a través de una interfaz gráfica.

5. REFERENCIAS BIBLIOGRÁFICAS

1. **Curso Universitario Sistema de Información en los Sistemas de Salud. Unidad 3:** Interoperabilidad y estándares, Instituto Universitario del Hospital Italiano, Buenos Aires, Argentina, pp. 12, 2009.
2. **Portolés, Jesús:** Búsqueda Semántica en Repositorios de Conceptos Biomédicos Estandarizados, Capítulo 1: Introducción, Escuela Técnica Superior de Ingeniería informática, Valencia, España, pp.5, 2010.

3. Libicki, M. and I. Brahmakulam: The Costs and Benefits of Moving to the ICD-10 Code Sets, RAND Corporation, California, USA, 2004.

4. Ronald Corneta, Nicolette F. de Keizera: Overcoming Barriers to Evaluation of Terminological Systems, IMIA, 2004.

5. Portolés, Jesús: Búsqueda Semántica en Repositorios de Conceptos Biomédicos Estandarizados, Capítulo 4: C.T Hunter, Escuela Técnica Superior de Ingeniería informática, Valencia, España, pp. 16-23.

6. ANSI/AGMA ISO 17485-A08: Bevel Gears -- ISO System of Accuracy, American Gear Manufacturers Association (AGMA), Alexandria, Virginia, USA, 2008.

6. Portolés, Jesús: Búsqueda Semántica en Repositorios de Conceptos Biomédicos Estandarizados, Capítulo 4: C.T Hunter, Escuela Técnica Superior de Ingeniería informática, Valencia, España, pp. 24-27, 2010.

7. Bioportal de NCBO: [En línea] <http://bioportal.bioontology.org>.

8. CliniClueXplorer: [En línea] <http://www.cliniclue.com>.

9. Especificación de CTS, 2004.

10. Object Management Group, Inc.: Lexicon Query Service Specification, Object Management Group, Inc., 2000.

11. Common Terminologies Services (HL7/ISO CTS):[En línea] <http://informatics.mayo.edu/LexGrid/index.php?page=ctsspec>.

12. Portolés, Jesús: Búsqueda Semántica en Repositorios de Conceptos Biomédicos Estandarizados, Capítulo 4: C.T Hunter. Escuela Técnica Superior de Ingeniería informática, Valencia, España, pp. 48-62, 2010.

6. SÍNTESIS CURRICULAR DEL AUTOR

Alberto Bradshaw González: Nació el 9 de diciembre de 1987. En el año 2011 cumplió misión internacionalista en Venezuela como cooperante de la Misión Médica Cubana. Graduado en Ingeniería en Ciencias Informáticas en la sede central de la Universidad de las Ciencias Informáticas (UCI) en el año 2012. Como estudiante perteneció al proyecto Tecnologías, Integración y Estándares (TIE) del Centro de Informática Médica (CESIM) de la facultad 7 de esta universidad. Perteneció al Centro Telemática de la facultad 2 de la UCI, donde se desempeña como programador en el proyecto Sistema Integral de Análisis de Información (SIAI). Dirección postal: Carretera a San Antonio de los Baños, Km. 2 ½, Municipio Boyeros, Ciudad de la Habana, Universidad de las Ciencias Informáticas, edif. 75, apto. 303, tel.: 837-2237. Correo electrónico: bradsg@uci.cu.