

Servidor de Terminología para Sistemas Hospitalarios

Evelyn Aranda, Cynthia Villalba

Facultad Politécnica

Universidad Nacional de Asunción

San Lorenzo, Paraguay

Email: evearandag@gmail.com, cvillalba@pol.una.py

Abstract—Representar de manera precisa el acto médico es una tarea particularmente compleja debido a la utilización de un lenguaje ambiguo, con gran dependencia del contexto y por la común utilización de jergas y acrónimos entre los médicos. La codificación del acto médico en estándares posibilita el registro médico de manera completa, detallada y precisa. La búsqueda manual de esta codificación conlleva mucho tiempo, dificultando así la codificación en estándares al momento del registro del acto médico. La propuesta de este trabajo consiste en el diseño y la implementación de un servidor de terminología médica ágil (rápido), que mapea el texto ingresado en lenguaje natural a una lista de terminologías codificadas en estándares médicos. La propuesta fue evaluada en cuanto al tiempo de respuesta y a la usabilidad del sistema. El tiempo de respuesta promedio fue de 893 milisegundos, el cual fue comparado con el tiempo de respuesta de otro buscador. El servidor de terminología propuesto resultó más veloz en el 86% de los casos. Además se evaluó la satisfacción del usuario en la utilización del sistema, a través de cuestionarios de usabilidad, SUS (System Usability Scale). El servidor de terminología resultó satisfactorio para el usuario, calificándolo de excelente según la escala de Bangor et. al.

I. INTRODUCCIÓN

La documentación de los registros médicos, usualmente, se realiza en texto libre, ya que es el medio más expresivo y sencillo de utilizar. Sin embargo, el texto libre no permite el almacenamiento estructurado de la información. Sin una estructura en la documentación, se dificulta el procesamiento de los datos. Para los programas informáticos resulta difícil procesar información no estructurada [2]. Por lo tanto, cuando se desea recuperar la información para el análisis de datos, el texto libre para la documentación del registro médico, no es quizás la mejor elección.

Las terminologías médicas codificadas en estándares son vocabularios controlados y estructurados que hacen posible el almacenamiento de información detallada, precisa y completa. Además, reducen el ingreso de términos vagos, ambiguos o redundantes, mejorando de esta manera el proceso de registro y reduciendo las posibilidades de error. Así también, las terminologías proveen un lenguaje común para el uso de la información o para el intercambio de la misma [3].

La codificación de los registros médicos en estándares es un procedimiento que se realiza a diario en los establecimientos médicos. Dicha información es utilizada para realizar informes estadísticos en el área de la salud. En algunos casos la búsqueda de la codificación de terminologías médicas se

realiza manualmente, consultando manuales de codificación o a través de internet utilizando teléfonos celulares. Debido al tiempo que toma esta búsqueda manual, alrededor de 25 segundos en promedio, la codificación no se realiza al momento de la consulta médica.

Con el fin de agilizar el proceso de búsqueda de codificaciones para que pueda ser llevado a cabo durante la consulta médica, y como objetivo general de este proyecto, se pretende diseñar e implementar un servidor de terminología médica ágil que a partir de texto ingresado en lenguaje natural, proporcione un vocabulario específico, detallado y codificado en estándares internacionales de terminología médica. Por lo tanto, un *servidor de terminología médica* es un sistema de software que mapea el texto ingresado en lenguaje natural, a una lista de terminologías codificadas en estándares médicos [1]. Se pretende disponibilizar el servidor de terminología en un sitio web para que otros sistemas y médicos interesados en la codificación, puedan tener acceso a él a través de una conexión a internet.

Los objetivos específicos del proyecto son los siguientes:

- Reducir el tiempo de búsqueda de terminologías médicas codificadas, respecto al proceso actual.
- Realizar búsquedas de terminologías complejas a través de términos amigables.
- Determinar la facilidad de uso del sistema.

A continuación se presentan las limitaciones y supuestos de este trabajo:

- No abarca las cuestiones de seguridad.
- No se cubren mecanismos de autenticación.

El resto del trabajo está organizado como sigue. En la sección 2 se presentan los antecedentes, incluyendo el marco conceptual y los trabajos relacionados. En la sección 3 se presenta la propuesta de solución, que abarca el diseño de la arquitectura propuesta y los detalles de la implementación. En la sección 4 se describe el diseño de la evaluación, el análisis e interpretación de los resultados obtenidos. Por último, en la sección 5 se presenta la conclusión y trabajos futuros.

II. ANTECEDENTES

En esta sección se definen los conceptos necesarios para comprender el problema y los trabajos relacionados.

A. Marco Conceptual

El *servidor de terminología* es un sistema que mapea un texto ingresado en lenguaje natural a una terminología de referencia, completa, detallada, formal y codificada en estándares [4].

Un estándar comprende un grupo de reglas y definiciones que especifican como llevar a cabo un proceso [19]. El estándar aplicado al ámbito médico se refiere a contar con un lenguaje estándar para comunicar la información del cuidado de la salud [20].

Para lograr su propósito, el servidor de terminología se vale de tres capas en las que se dividen las terminologías o vocabularios. En la Fig. 1. se observan las tres terminologías que componen el servidor. Estas terminologías son independientes entre sí, pero dentro del servidor de terminología están organizadas en tres niveles visibles en la siguiente pirámide.

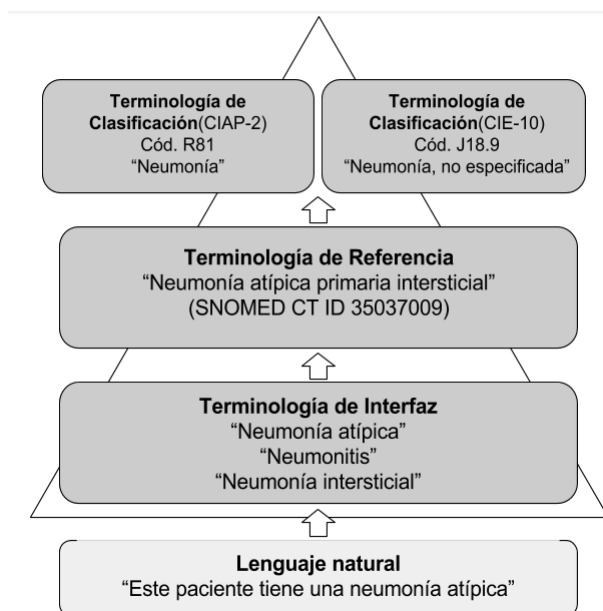


Fig. 1: División de los vocabularios de un servidor de terminología [5].

Se encuentran organizados de la siguiente manera porque partimos de términos sencillos, utilizados comúnmente en el lenguaje natural, hacia términos más complejos como los de la terminología de referencia. La terminología de interfaz (como su nombre lo indica) sirve de conexión entre el lenguaje común utilizado (natural) y terminologías complejas codificadas (de referencia), permitiendo así que varios términos comunes puedan hacer referencia a un mismo término de la terminología de referencia. La terminología de interfaz representa el dominio y la jerga local, es el lenguaje utilizado por los médicos en el registro, tiene la ventaja de utilizar términos médicos amigables y familiares [5]. Usualmente se puebla con términos utilizados en el lenguaje natural.

Las terminologías de referencia están compuestas por términos estandarizados que permiten a las computadoras "entender" los datos clínicos.

Las terminologías de clasificación, como su nombre lo indica, son terminologías que permiten clasificar los datos, como por ejemplo, diagnósticos, enfermedades y otros, para poder realizar un análisis estadístico sobre los datos. También están codificados en estándares [5].

La funcionalidad principal de los servidores de terminología es, entonces, el mapeo entre el vocabulario médico natural y códigos estandarizados.

B. Trabajos Relacionados

En varios países como en Estados Unidos [12, 16], Chile [8] y Argentina [5], se han implementado servidores de terminología.

En Argentina, el Hospital Italiano de Buenos Aires (HIBA) implementó su propio servidor de terminología con la idea de aplicarlo a su sistema hospitalario. El HIBA se basó en la estructura del estándar "Nomenclatura Sistemática de Medicina - Términos Clínicos" (SNOMED-CT, por sus siglas en Inglés), para el diseño de la estructura de su servidor. La población del servidor con datos se basa en la sugerencia directa de términos de parte de los usuarios. Estas sugerencias son evaluadas por especialistas en terminología antes de la incorporación formal al servidor. Además, al iniciar el servidor de terminología se analizaron dos millones de términos médicos incluidos en documentos del hospital para luego incluirlos en el servidor formalmente [4].

Hacia el año 1986, Estados Unidos desarrolló el Sistema Unificado de Lenguaje Médico (UMLS, por sus siglas en inglés), creado para reducir la barrera para la aplicación de sistemas informáticos en la medicina y para que sirva de base a desarrolladores de software que quieran adaptarlo a sus sistemas hospitalarios. UMLS se compone de tres herramientas entre las que se encuentra el Metatesauro, que es la integración de tesauros y ontologías biomédicas desarrolladas independientemente a lo largo de años. Integra cerca de 2.000.000 de nombres, unos 900.000 conceptos de más de 60 familias de vocabularios biomédicos [9], y entre las terminologías y clasificaciones más conocidas de las que se compone, se encuentran SNOMED CT, Clasificación Internacional de Enfermedades Versión 9 (CIE9) Y Clasificación Internacional de Enfermedades Versión 10 (CIE10), Asociación Norteamericana de Diagnóstico de Enfermería (NANDA, por sus siglas en Inglés), Nombres y Códigos de Identificadores de Observación Lógica (LOINC, por sus siglas en Inglés) y Clasificación Internacional de Atención Primaria (CIAP) [11].

MEDCIN es una terminología médica de tipo comercial, desarrollado por MediComp Systems INC. diseñado para su uso en sistemas de registros electrónicos de salud. Incluye cerca de 280.000 datos clínicos incluyendo síntomas, tests, examinación física, diagnósticos y terapias. Además, tiene la capacidad de mapearse a los estándares SNOMED CT, Clasificación Internacional de Enfermedades Versión 9, Modificación Clínica (CIE9-MC), Clasificación Internacional de Enfermedades Versión 10, Modificación Clínica (CIE10-MC), LOINC, entre otros [13].

Terminología Médica Convergente (CMT, por sus siglas en

inglés) es una terminología médica que sirve como fuente de definiciones de conceptos médicos. Su núcleo está comprendido por los estándares SNOMED-CT, LOINC, CIE9 y CIE10. Inicialmente desarrollada en Estados Unidos por la empresa Kaiser Permanente para su propio uso y para cubrir necesidades en sus registros electrónicos de salud. Tras 16 años, Kaiser Permanente donó CMT a la Organización Internacional para el Desarrollo de Normas de Terminología de la Salud (IHTSDO, por sus siglas en Inglés) para su distribución en los Estados Unidos a fin de que otros vendedores de sistemas pudieran utilizar esta terminología común en sus registros electrónicos de una manera más rápida [14].

Como Sistema Nacional de Salud de los Estados Unidos, la Administración de Salud de Veteranos (VHA, por sus siglas en inglés) creó una terminología de referencia denominada Terminología de Referencia Empresarial (ERT, por sus siglas en Inglés) para promover la estandarización de datos. ERT contiene terminología mapeada a estándares. Para ello adquirieron licencias de SNOMED-CT, Terminología Actual de Procedimiento (CPT, por sus siglas en Inglés), UMLS y LOINC [7].

Chile, a su vez, dió inicio a la implementación del servidor de terminología “Semantikos” que busca estandarizar el lenguaje utilizado en el ámbito de la salud. El servidor de terminología pretende proveer a los médicos de un buscador de terminologías rico en sinónimos, acrónimos y términos comunes o amigables utilizados en el ámbito médico. Se utilizarían los estándares NANDA, Clasificación de Resultados de Enfermería (NOC, por sus siglas en Inglés), Clasificación de Intervenciones de Enfermería (NIC, por sus siglas en Inglés), SNOMED-CT (que incluye mapeos oficiales a otras terminologías de clasificación como CIE9 y CIE10), entre otros [8].

III. SOLUCIÓN PROPUESTA

En esta sección se presentan las funcionalidades esperadas del sistema, los requisitos no funcionales, el diseño de la arquitectura propuesta y los detalles de la implementación. Se espera que el servidor de terminología posea las siguientes funcionalidades:

- Sugerencia de inclusión de términos (llamados términos amigables o familiares) a la terminología de interfaz, para asociarlos a términos (llamados formales) pertenecientes a la terminología de referencia.
- Validación de la asociación del término amigable al término formal de la terminología de referencia. Este proceso es llevado a cabo por un especialista en codificación. De esta manera se produce el alta de la sugerencia (ítem anterior).
- Búsqueda de codificaciones en estándares utilizando términos amigables o formales en dicha búsqueda.

Como requisitos no funcionales se espera un tiempo de respuesta aceptable para el usuario menor a 10 segundos, límite en el tiempo de respuesta según Jakob Nielsen [34], y la disposición del sistema para ser utilizado desde otros sistemas de software.

A. Diseño de la arquitectura propuesta

Se propone una arquitectura cliente - servidor de tres capas (también conocida como arquitectura multinivel), organizada de la siguiente manera: capa de presentación, de negocios y capa de datos.

Se eligió utilizar este patrón por la independencia entre las capas y la clara definición de cada una de ellas en cuanto al objetivo que persigue. De esta manera, es posible implementar cada capa de forma totalmente independiente a las otras.

En la Fig. 2. se observa la distribución de los componentes principales del sistema en estas tres capas.

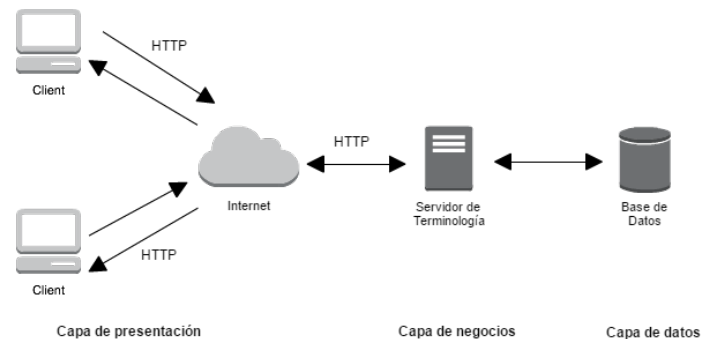


Fig. 2: Arquitectura cliente - servidor de tres niveles.

La capa de presentación es la que se expone del lado del cliente, está compuesta por interfaces que proveen acceso a la capa de negocios. La capa de negocios está formada por el servidor de terminología (conjunto de servicios web y otras herramientas) y la capa de datos está conformada por las bases de datos.

En la Fig. 3. se detalla los componentes de cada capa.

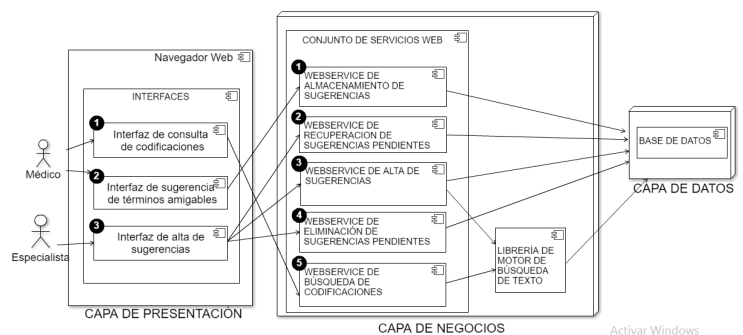


Fig. 3: Diagrama de componentes de la arquitectura propuesta.

1) *Capa de Presentación*: está formada por las siguientes interfaces:

- *Interfaz de consulta de codificaciones*, utilizada por los médicos, permite la búsqueda de codificaciones médicas en estándares, a partir de términos amigables o formales.
- *Interfaz de sugerencia de términos amigables*, utilizada por los médicos, permite sugerir la inclusión de un término amigable a la terminología de interfaz asociándolo a un término formal de la terminología de

referencia. Esta sugerencia de inclusión y asociación debe ser validada por un especialista en codificación médica para que se produzca el alta del término amigable. La asociación entre el término amigable y el término formal, determina la relación entre la terminología de interfaz y la terminología de referencia. Un término amigable puede estar asociado a solo un término formal de la terminología de referencia. Sin embargo, un término formal puede tener cero o varios términos amigables asociados a él. En la Fig. 1., por ejemplo, veíamos cómo los términos amigables “Neumonía atípica”, “Neumonitis” y “Neumonía intersticial”, de la terminología de interfaz, hacían referencia al mismo término “Neumonía atípica primaria intersticial” de SNOMED-CT en la terminología de referencia.

- *Interfaz de alta de sugerencias*, utilizada por un especialista en codificación médica, permite dar de alta una sugerencia a través de la validación o rechazar la sugerencia, invalidándola.

2) *Capa de Negocios*: está compuesta por un conjunto de servicios web y una librería de motor de búsqueda de texto. Este conjunto de servicios web son consumidos por las interfaces de la capa de presentación y también pueden ser consumidos por otros sistemas externos. Estos servicios se comunican con la capa de datos para la persistencia de la información. Los servicios web son los siguientes:

- *WebService de almacenamiento de sugerencias*, se encarga de persistir en base de datos la sugerencia de asociación del término amigable (a incluir en la terminología de interfaz) y el término formal de la terminología de referencia.
- *WebService de recuperación de sugerencias pendientes*, recupera de la capa de datos, las sugerencias realizadas por los médicos que no hayan sido validadas aún.
- *WebService de alta de sugerencias*, previa validación de la sugerencia por un especialista en codificación, este servicio web persiste en base de datos la asociación formal entre los términos sugeridos.
- *WebService de invalidación de sugerencias pendientes*, permite invalidar una sugerencia que todavía no haya sido validada.
- *WebService de búsqueda de codificaciones*, recibe un texto a partir del cual realiza la búsqueda de codificaciones en estándares médicos (estándares de referencia y de clasificación) asociados al texto recibido. A partir del texto ingresado en la interfaz de consulta de codificaciones, devuelve un listado de codificaciones en estándares médicos (estándares de referencia y de clasificación).

Como se visualiza en la Fig. 3., dos servicios web utilizan una librería de motor de búsqueda de texto para el almacenamiento y recuperación de información de la capa de datos. Los demás se comunican directamente con esta capa.

3) *Capa de Datos*: está formada por bases de datos donde almacenar y desde donde recuperar la información. Estas bases de datos se detallan en la sección de implementación.

B. Implementación

Para implementar el diseño, se procedió a la búsqueda de herramientas que cubrieran estos requerimientos. Se eligió a la librería de motor de búsqueda *Apache Lucene*, y en cuanto a los datos utilizados por el servidor de terminología, se empleó una base de datos de vocabulario médico muy grande denominado “Metatesauro”. A continuación se detallan ambas herramientas.

Herramientas Utilizadas:

Metatesauro de UMLS: se decidió utilizar el Metatesauro (mencionado en la sección de Antecedentes) porque es una fuente de información sobre conceptos biomédicos bastante completa y porque aporta un gran número de vocabularios controlados o estándares, por lo que supone una referencia prácticamente universal de conceptos.

Existen otras dos herramientas (también mencionadas en la sección de Antecedentes) que sirven como fuente de información, ellas son: CMT y ERT. De entre estas fuentes de información, se decide utilizar el Metatesauro de UMLS porque es el único que posee información en Español.

Para llevar a cabo la instalación del Metatesauro, se utiliza una herramienta llamada *Metamorphosys*. Esta herramienta permite almacenar en bases de datos MySQL u Oracle, la información contenida en el Metatesauro [21]. Además posee un buscador para navegación sobre los datos instalados [29]. Previamente a la implementación de este trabajo, y con miras a utilizar *Metamorphosys* como buscador de terminologías, se realizaron búsquedas preeliminares sobre él para determinar si era apto para su utilización como buscador de terminologías en el marco de este trabajo. Tras las búsquedas se pudo constatar que *Metamorphosys* presenta las siguientes características:

- Tiempos de respuesta muy diferentes y apartados ante la búsqueda de un mismo texto. La primera búsqueda tiene un tiempo de respuesta muy lento, llegando a tardar hasta 2,7 minutos en responder, tiempo considerado no aceptable para el usuario según Jakob Nielsen [37].
- A partir de la segunda búsqueda (búsquedas posteriores) de un término, los tiempos de respuesta son más rápidos debido a que *Metamorphosys* cachea datos en memoria realizando la lectura de datos desde este punto.
- En base al punto anterior, observamos que el tiempo de respuesta se ve influenciado, por lo tanto, por la cantidad de memoria disponible para el cacheo.
- *Metamorphosys* está orientado a la instalación y personalización de datos del Metatesauro, no precisamente como buscador de terminologías.
- Los términos se instalan ya clasificados en términos frecuentes y no frecuentes. La búsqueda de términos

frecuentes es un poco más rápida que la búsqueda de términos no frecuentes, debido a que los términos frecuentes se encuentran indexados.

Por estos motivos, se decidió buscar alguna herramienta que agilice el tiempo de respuesta en la recuperación de datos y que no dependa de la cantidad de memoria para cacheo de datos. La herramienta elegida es Apache Lucene.

Apache Lucene: es una librería de motor de búsqueda de texto, de alto rendimiento y escalable, enfocada en la indexación y en la búsqueda de texto completo (o *full-text search*, en inglés) sobre cualquier dato que pueda ser convertido a texto [6]. Si bien Lucene es una librería, se lo suele utilizar como motor de búsqueda ya que su núcleo está formado por tres conceptos interrelacionados y utilizados en los motores de búsqueda. Estos conceptos son: la indexación, el análisis y la búsqueda [12]. Así como Lucene, existen otras librerías de recuperación de información como Egothor y Xapian, ambas reúnen características interesantes para este trabajo. Una de las características decisivas a la hora de elegir una de ellas, es el lenguaje en el que está implementado. Debido a la decisión de implementar este trabajo en el lenguaje Java (por familiaridad), Xapian queda descartada por estar escrita en C++ y poseer apenas una implementación Beta en Java [17]. Las características entre Egothor y Lucene, son casi las mismas, pero se decide utilizar Apache Lucene por ser parte de una comunidad mucho más grande (la Apache Software Foundation) permitiendo así encontrar mayor soporte en caso necesario [22].

Implementación de capas:

En esta sección se detalla la implementación de cada una de las capas que conforman la arquitectura.

- **Capa de Presentación:** las interfaces se implementaron utilizando JavaServer Pages (JSP), que es una tecnología utilizada para crear páginas web dinámicas. Los archivos JSP se compilan como cualquier clase Java en un contenedor de aplicaciones y por lo tanto es una tecnología del lado servidor. Utilizan los tags de HTML y permiten incluir código Java en sus páginas [24].

Para cuestiones de diseño y funcionalidad de las interfaces se utilizó el framework de código abierto *Twitter Bootstrap* en su versión 3.3.7. El framework contiene plantillas de diseño con tipografía, formularios, botones y otros elementos comúnmente utilizados basados en HTML y CSS, además de extensiones JavaScript adicionales. La elección se debe además a que el framework es soportado en la mayoría de los navegadores existentes. Uno de los atractivos, además de los ya mencionados, es que a partir de la versión 2, Bootstrap contempla funcionalidad responsive por defecto, ajustándose perfectamente a un entorno mobile [15].

A continuación se exponen las interfaces implementadas.

En la Fig. 4. se visualiza la interfaz de sugerencia de términos amigables que permite sugerir una relación

Fig. 4: Interfaz de sugerencia de términos amigables.

entre un término amigable y un término formal, a través del ingreso de ambos términos en los campos *Término sugerido* y *Término formal*, respectivamente.

En la Fig. 5. se visualiza la interfaz de alta de sugerencias. Esta interfaz permite a un especialista en codificación, elegir una sugerencia pendiente de validación, validarla y darla de alta. También permite invalidar una sugerencia en caso de que no sea correcta.

Al seleccionar una sugerencia pendiente, mediante el combobox, se precargan los campos *Término sugerido* y *Asociado al término formal* con los datos de la sugerencia elegida. En este punto, el especialista puede editar estos dos campos en caso de que requieran una corrección. A partir de este punto, el especialista tiene dos opciones a seguir: uno, la invalidación de la sugerencia o dos, la búsqueda de la terminología de referencia a la cual asociar el término amigable. El botón “Buscar Terminología” permite realizar la búsqueda de la terminología de referencia asociada al término formal ingresado. Esta búsqueda devuelve un listado de codificaciones en estándares que se precargan en el combobox etiquetado “Despliegue”. El especialista selecciona la codificación de la terminología de referencia que se más se ajuste y presiona el botón “Aceptar” que se encarga de persistir la relación entre el término amigable de la terminología de interfaz y el término elegido de la terminología de referencia.

En la Fig. 6. se visualiza la interfaz de consulta de codificaciones. Esta interfaz permite al médico ingresar el texto de búsqueda, en el campo *Término a consultar*, y devuelve un listado de terminologías codificadas en estándares de referencia y clasificación, asociadas al término consultado. La estructura devuelta en las terminologías codificadas es la siguiente: descripción de la terminología, nombre del estándar y la codificación.

- **Capa de Negocios:** compuesta por servicios web del tipo “Transferencia de Estado Representacional” (REST) uti-

POBLACIÓN DE LA TERMINOLOGÍA DE INTERFAZ
ESTA INTERFAZ PERMITE DAR DE ALTA O ELIMINAR UNA SUGERENCIA PENDIENTE

1er. Paso: Elegir una sugerencia pendiente. Esta elección precarga los términos sugeridos y el término formal a asociar para edición en caso de necesidad

Seleccione una sugerencia pendiente

Termino sugerido:

Asociado al termino formal:

2do. Paso: da la posibilidad de eliminar la sugerencia hecha por el médico o buscar la terminología codificada en estándares asociada al termino formal. La búsqueda de terminologías devuelve una lista codificada en varios estándares.

Buscar Terminología

Despliegue

Fig. 5: Interfaz de alta de sugerencias.

CONSULTA DE TERMINOLOGIAS

Termino a consultar:

Patrón: Descripción - Estándar - Codificación

REFERENCIAS: ICD (International Classification of Disease o CIE, por sus siglas en español) SCTSPA (SNOMED CT version Español)

Fig. 6: Interfaz de consulta de codificaciones.

lizando el lenguaje de programación Java y consumidos por las interfaces de la capa de presentación a través de peticiones Xml y Javascript Asíncrono (AJAX, por sus siglas en Inglés). Se decide implementar servicios web de tipo REST por sus principios de diseño, ellos son:

- Utiliza los métodos HTTP de manera explícita para solicitar acciones sobre el servidor.
- No mantiene ningún estado acerca del cliente.
- Expone URIs con forma de directorios.
- Transfiere XML, Notación de Objetos JavaScript (JSON, por sus siglas en Inglés), o ambos. [16]

A continuación se describe la implementación de cada servicio web:

- **WEBSERVICE DE ALMACENAMIENTO DE SUGERENCIAS**, recibe el término amigable y el término formal a través de la petición AJAX realizada por la *interfaz de sugerencia de términos amigables*. Luego se comunica con la base de datos relacional e inserta la sugerencia con estado PENDIENTE en una tabla llamada "IFT" (InterFace Terminology) donde se registran todas las sugerencias realizadas. URI del servicio:

```
POST {urlbase}/sugerencias?sugerido={val1}&asociado={val2} HTTP/1.1
```

donde:

urlbase = ruta del contexto de la aplicación

val1 = término amigable

val2 = término formal

- **WEBSERVICE DE RECUPERACIÓN DE SUGERENCIAS PENDIENTES**, este servicio se comunica con la base de datos relacional para recuperar de la tabla "IFT", todas las sugerencias con estado pendiente. Este servicio es consumido por la *interfaz de alta de sugerencias* para que las mismas puedan ser validadas y dadas de alta por el especialista en codificación. URI del servicio:

```
GET {urlbase}/sugerencias HTTP/1.1
```

- **WEBSERVICE DE ALTA DE SUGERENCIAS**, recibe el término amigable, el término formal y dos códigos que identifican al término formal en la terminología de referencia. Se reciben dos códigos para identificar correctamente al término formal en el metatesauro. El servicio se utiliza para dar de alta una sugerencia con estado PENDIENTE. En primer lugar, el servicio se comunica con la base de datos relacional para la actualización del estado de la sugerencia al estado FINALIZADO en la tabla IFT. Posteriormente, apache lucene indexa los datos de la sugerencia y los almacena en "Base de Datos Lucene IFT". URI del servicio:

```
POST {urlbase}/sugerencias/{id}?sugerido={val1}&asociado={val2}&
aui={cod1}&cui={cod2} HTTP/1.1
```

donde:

val1 = término amigable

val2 = término formal

aui y cui = identificadores relacionados al término formal

- **WEBSERVICE DE INVALIDACIÓN DE SUGERENCIAS PENDIENTES**, este servicio invalida una sugerencia pendiente con identificador *id* (recibido como parámetro), actualizando la sugerencia al estado CANCELADO. Este servicio es consumido por la *interfaz de alta de sugerencias*.

URI del servicio:

```
DELETE {urlbase}/sugerencia/{id} HTTP/1.1
```

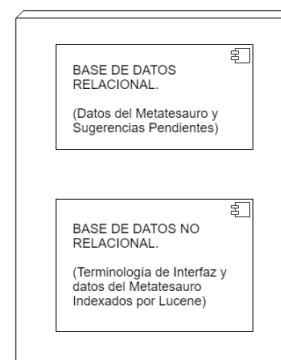
- **WEBSERVICE DE BÚSQUEDA DE CODIFICACIONES**, este servicio recibe un texto de búsqueda y devuelve una lista de terminologías (codificadas en estándares) de referencia y clasificación. El servicio utiliza la librería de motor de búsqueda de texto para la recuperación de información contenida en los repositorios de datos “Base de Datos Lucene IFT” y “Base de Datos Lucene MRCONSO”. La librería de motor de búsqueda hace uso del analizador implementado (Fig. 6) en la recuperación de datos.

URI del servicio:

```
GET {urlbase}/ift/search?diagnostico={term_búsqueda} HTTP/1.1
```

donde:

term_búsqueda = texto de búsqueda



CAPA DE DATOS

Fig. 7: Componentes de la capa de datos.

IV. DISEÑO DE LA EVALUACIÓN, ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

En esta sección se presenta el diseño de la evaluación del tiempo de respuesta del buscador del servidor de terminología propuesto y de la satisfacción del usuario final respecto a la utilización del sistema. Finalmente se exponen los resultados de la evaluación y la interpretación de los resultados obtenidos.

A. Diseño de la evaluación

En este trabajo se llevan a cabo dos tipos de evaluaciones: la encuesta (survey, en inglés) para la obtención de información en cuanto a la satisfacción del usuario en el uso del sistema (usabilidad) y casos de estudio (investigación empírica o evaluación) para comparar los tiempos de respuesta de los buscadores del Servidor de Terminología y de Metamorphosys. Esta evaluación no es considerada un experimento ya que, según Wohlin [36], en un experimento el objetivo es manipular una o más variables de entrada que afectan a los resultados o variables de salida. Se mide el efecto de la manipulación, y basándose en esto se puede realizar un análisis estadístico. Las evaluaciones, sin embargo, pueden aplicarse como una estrategia de investigación comparativa, comparando los resultados de utilizar un método, contra los resultados de utilizar otro enfoque.

A continuación, se expone y utiliza el proceso sugerido por Wohlin para las evaluaciones.

1) Alcance:

- **Objeto de estudio:** el servidor de terminología en cuanto al tiempo de respuesta del buscador y la satisfacción del usuario.
- **Propósito:** evaluar el tiempo de respuesta del buscador del Servidor de Terminología contra con el tiempo de respuesta del buscador de Metamorphosys, y conocer la satisfacción del usuario en la utilización del sistema a través de un cuestionario de usabilidad (SUS).
- **Focos de calidad:** productividad (en cuanto a la velocidad de completitud de una tarea) y satisfacción del usuario.
- **Perspectiva:** punto de vista del investigador.
- **Contexto:** la evaluación se desarrolla en el contexto de trece médicos residentes (categorías R1, R2 y R3)

- **Capa de Datos:** está formada por una base relacional MySQL versión 5.6 y un repositorio de datos (o base de datos no relacional) creado por Apache Lucene. Se decidió utilizar MySQL versión 5.6 debido a que es de código abierto y porque en el proceso de instalación de datos, Metamorphosys ofrece scripts de población de datos sólo para MySQL 5.5, 5.6 u Oracle.

Como se visualiza en la Fig. 7., la base de datos relacional está formada por los datos del Metatesauro y por las sugerencias pendientes. El repositorio de datos (no relacional), está formado por los datos que Lucene va indexando para luego realizar su búsqueda a partir de ahí. Lucene toma los datos a indexar desde la base de datos relacional, acotando los datos al tipo “Problemas o Enfermedades” de los estándares SNOMED-CT, CIE9 y CIE10. Se limitó el universo de datos a dichos estándares por ser los más utilizados a nivel mundial [38 - 39]. El repositorio de datos no relacional almacena, además, las sugerencias dadas de alta por el especialista, una vez indexados estos datos por Apache Lucene.

seleccionados randómicamente en el área de Pediatría del Hospital de Clínicas de Paraguay.

2) *Planeamiento*: En el área de Pediatría del Hospital de Clínicas del Paraguay, se dispondrán dos computadores personales con procesadores de 2.7 GHz., memoria RAM de 4 GB., y disco duro de 250 y 500 GB., en los trece médicos residentes pueden realizar búsquedas de problemas, codificados en estándares, utilizando el servidor de terminología médica. Los mismos realizarán búsquedas de al menos, tres problemas. Además, pueden realizar sugerencias de términos amigables para luego realizar búsquedas de terminologías codificadas en estándares, a partir de los términos amigables insertados. Al finalizar las búsquedas, los médicos residentes completan el cuestionario de usabilidad, SUS¹ (System Usability Scale), para la evaluación de la satisfacción del usuario y la usabilidad general del sistema.

Preguntas de investigación:

- ¿El buscador del servidor de terminología implementado, es más ágil que el buscador de Metamorphosys?.
- El servidor de terminología, ¿resulta eficiente, y satisfactorio en su uso?.

3) *Operación*: : el experimento se lleva a cabo de a dos participantes a la vez por la disponibilidad de computadoras. Al iniciar se le explica a cada médico residente de qué se trata el experimento y que con el buscador del servidor de terminología realicen búsquedas de al menos cinco problemas. El servidor de terminología responde con terminologías médicas codificadas en los estándares SNOMED-CT, CIE9 y CIE10, asociados al problema de búsqueda. En este punto cabe mencionar que dos médicos buscaron palabras que no eran precisamente “enfermedades” como por ejemplo “riesgo social”, “social”, “entorno” y “problemas sociales”. Además, otros dos médicos insertaron los siguientes términos amigables: “civ” y “hta”, haciendo alusión a comunicación ventricular e hipertensión arterial, respectivamente. Estas inserciones se realizaron con el fin de probar la funcionalidad de sugerencia y alta de términos amigables (o población de la terminología de interfaz) para realizar búsquedas de problemas codificados en estándares a través de estos términos amigables insertados. Ambos acrónimos fueron insertados exitosamente al servidor de terminología y las búsquedas a partir de estos acrónimos devolvieron los resultados esperados. En cuanto a los cuestionarios SUS completados al final de cada prueba, todos fueron llenados exitosamente.

B. Análisis e Interpretación de Resultados

Por cada palabra buscada por los médicos, se comparan los tiempos de respuesta del servidor de terminología con los tiempos de respuesta de Metamorphosys. Se analizan los peores y mejores tiempos de respuesta de cada uno. Cabe mencionar que Metamorphosys dispone de cuatro opciones de búsqueda (las cuales arrojan diferentes tiempos de

respuesta) y que las mismas fueron utilizadas en la comparación, éstas son: *Requiere coincidencia en la palabra con frecuencia más baja* (Opción 1), *Descartar coincidencias que sólo contienen palabras más frecuentes* (Opción 2), *Requiere coincidencia en al menos dos palabras* (Opción 3) y *Coincidencia básica* (Opción 4).

En las tablas comparativas siguientes se observan en las columnas tres tiempos que corresponden a Metamorphosys: el tiempo de respuesta de la primera búsqueda, de las búsquedas posteriores, y el tiempo promedio entre las dos primeras. El tiempo promedio es el tiempo comparado con el tiempo de respuesta del servidor de terminología.

En las Tablas I y II se exponen los tiempos máximos de respuesta (peor caso) de Metamorphosys y del servidor de terminología, respectivamente.

En el peor caso de Metamorphosys se observa que el servidor de terminología es hasta un 98% más rápido.

En el peor caso del Servidor de Terminología, Metamorphosys es más ágil en un 33% .

En las Tablas III y IV se exponen los tiempos mínimos de respuesta (mejor caso) de Metamorphosys y del servidor de terminología, respectivamente.

El mejor caso de Metamorphosys no presenta diferencias significativas en el tiempo de respuesta contra el Servidor de Terminología, la diferencia es de milisegundos. En este caso, se considera que ambos tienen prácticamente el mismo rendimiento.

En cuanto al mejor caso del Servidor de Terminología, se concluye que es 90% más rápido que Metamorphosys. En la Tabla IV se observa que la búsqueda de la palabra “Diarrea cronica” en Metamorphosys utilizando la opción 3, no retorna ningún resultado. Esto se debe a que la opción 3 requiere coincidencia en al menos dos palabras y en este caso no hubo ningún registro en Metamorphosys que incluyera a las dos palabras juntas en el registro.

Al finalizar la búsqueda de terminologías codificadas, los médicos completaron un cuestionario SUS de diez ítems para poder evaluar la usabilidad del sistema. Se eligió utilizar SUS porque es el cuestionario estándar más utilizado para medir percepciones de usabilidad, es independiente de la tecnología, es decir, puede utilizarse para evaluar hardware, aplicaciones software, sitios web, celulares, entre otros, y además porque está disponible de manera gratuita y se ha comprobado que es confiable y válido [31].

Se procedió a evaluar cada cuestionario. Como cada cuestionario por sí solo no puede determinar la usabilidad del sistema, se halló el promedio de los resultados arrojados por cada cuestionario y este promedio representa la usabilidad del servidor de terminología. Para la interpretación de resultados, Bangor et al. [30] elaboraron una escala adjetiva de los puntajes SUS. Esta escala adjetiva muestra el nivel de aceptabilidad de lo que fue evaluado, en donde un puntaje SUS menor a 60 corresponde a una “F”, entre 60 y 69 corresponde a una “D”, entre 70 y 79 corresponde a una “C”, entre 80 y 89 corresponde a una “B”, y un puntaje de 90 o más corresponde a una “A”, siendo “A” el adjetivo de mayor aceptabilidad y “F”

¹ SUS es una escala simple de diez ítems que da una visión global de la evaluación de usabilidad [31].

Peor caso en el tiempo de respuesta de Metamorphosys por opción de búsqueda (en segundos)					
Opción de Búsq.	Texto buscado	Metamorphosys			Servidor de Terminología
		Primera búsqueda	Búsquedas posteriores	Tiempo promedio	
1	Infeccion del tracto urinario	23	1	12	0,1
2	Infeccion del tracto urinario	26	1	14	0,1
3	Neumonía adquirida en la comunidad	122	21	71	0,2
4	Neumonía adquirida en la comunidad	166	32	99	0,2

Tabla I: Peor caso en el tiempo de respuesta de Metamorphosys por opción de búsqueda.

Peor caso en el tiempo de respuesta del Servidor de Terminología contra las opciones de búsqueda de Metamorphosys (en segundos)					
Opción de Búsq.	Texto buscado	Metamorphosys			Servidor de Terminología
		Primera búsqueda	Búsquedas posteriores	Tiempo promedio	
1	Inflamacion	10	0,7	5	6
2	Inflamacion	15	1	8	6
3	Inflamacion	9	1	5	6
4	Inflamacion	4	0,7	2	6

Tabla II: Peor caso en el tiempo de respuesta del servidor de terminología contra Metamorphosys por cada opción de búsqueda.

Mejor caso en el tiempo de respuesta de Metamorphosys por opción de búsqueda (en segundos)					
Opción de Búsq.	Texto buscado	Metamorphosys			Servidor de Terminología
		Primera búsqueda	Búsquedas posteriores	Tiempo promedio	
1	Pielonefritis aguda	0,2	0,1	0,2	1
2	Poliquistosis	0,3	0,1	0,2	0,2
3	Criptocardia	0,1	0,1	0,1	0,2
4	Poliquistosis	0,2	0,1	0,2	0,2

Tabla III: Mejor caso en el tiempo de respuesta de Metamorphosys por opción de búsqueda.

el de menor aceptabilidad. En la Fig. 10. se aprecia la escala adjetiva de los puntajes SUS de Bangor et al.

En la Fig. 9. se visualizan los resultados que arrojaron cada

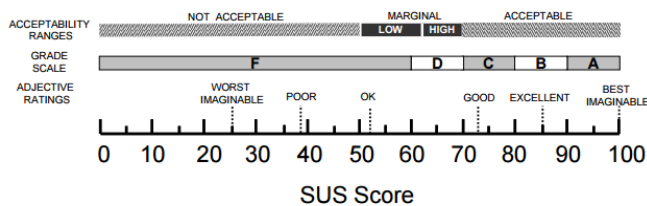


Fig. 8: Escala adjetiva de los puntajes SUS de Bangor et al. [30].

uno de los cuestionarios evaluados. El promedio de estos trece resultados, arroja un valor de $\frac{1085 \text{ puntuaciones SUS}}{13 \text{ cuestionarios}} = 83.46$ situando al servidor de terminología en una escala de “B” en la escala de Bangor et al., calificando al servidor de terminología como “Excelente”.

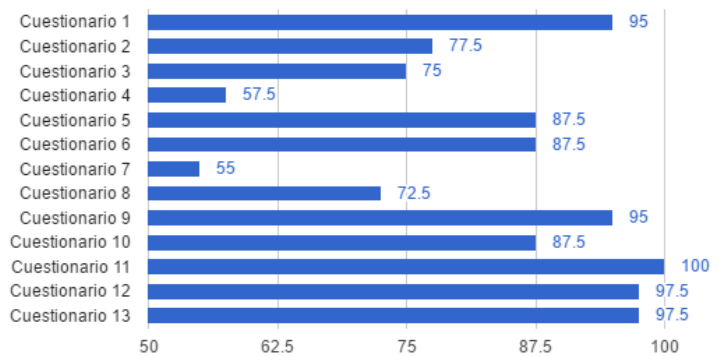


Fig. 9: Puntuaciones SUS de cada uno de los cuestionarios.

V. CONCLUSIÓN Y TRABAJOS FUTUROS

Contar con un servidor de terminología médica facilita el trabajo de codificación de problemas en estándares de salud. Debido a esto se decide implementarlo en el Paraguay para apoyar los registros médicos. El tiempo, durante este registro, es un factor clave y por ello es necesario contar con un sistema ágil que permita llevar a cabo la codificación del problema

Mejor caso en el tiempo de respuesta del Servidor de Terminología contra las opciones de búsqueda de Metamorphosys (en segundos)					
Opción de Búsq.	Texto buscado	Metamorphosys			Servidor de Terminología
		Primera búsqueda	Búsquedas posteriores	Tiempo promedio	
1	Diarrea cronica	7	0,4	3	0
2	Diarrea cronica	8	0,6	4	0
3	Diarrea cronica	-	-	-	0
4	Diarrea cronica	18	0,7	9	0

Tabla IV: Mejor caso en el tiempo de respuesta del servidor de terminología contra Metamorphosys por cada opción de búsqueda.

durante la consulta del paciente.

Este servidor de terminología fue diseñado como arquitectura multinivel para mantener la independencia de las capas. Se desarrollaron servicios web para que puedan ser consumidos por otros sistemas en caso de que se desee.

Con el fin de evaluar el tiempo de respuesta, se comparó el tiempo obtenido del buscador del servidor de terminología con el tiempo obtenido por el buscador de Metamorphosys. Se concluye que el servidor de terminología es más ágil que Metamorphosys salvo en algunos casos en el que Metamorphosys lee datos cacheados en memoria. Además, se logró reducir el tiempo de búsqueda de terminologías médicas codificadas a un tiempo de 6 segundos como máximo, contra los 25 segundos en promedio de búsqueda manual actual. Cumpliendo uno de los objetivos específicos, se logró además insertar términos amigables a la terminología de interfaz, que luego fueron utilizados para realizar las búsquedas correspondientes. Los casos de inserción y búsqueda fueron exitosos.

En base al análisis de los cuestionarios de usabilidad, se concluye que la utilización del Servidor de Terminología resultó satisfactorio para el usuario, calificándolo como “Excelente” según la escala de Bangor et. al.

Como trabajo futuro podría añadirse cuestiones de seguridad y autenticación que no abarca el sistema.

REFERENCES

- [1] de Quirós, F. G. B., Luna, D., Baum, A., Plazzotta, F., Otero, C., & Benítez, S. (2012). Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires. Santiago de Chile.
- [2] Johnson, S. B., Bakken, S., Dine, D., Hyun, S., Mendonça, E., Morrison, F., Bright, Tiffany., Van Vleck, Tielman., Wrenn, Jesse., & Stetson, P. (2008). An electronic health record based on structured narrative. *Journal of the American Medical Informatics Association*, 15(1), 54-64.
- [3] Terminologías Clínicas. Chile: Ministerio de Salud - Gobierno de Chile. <http://www.salud-e.cl/servicios-terminologicos/pagina-menu-int-prueba/>
- [4] Lopez Osornio, A., Luna, D., Gambarte, M. L., Gómez, A., Reynoso, G., & González Bernaldo de Quiros, F. (2007). Creation of a local interface terminology to SNOMED CT. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems* (p. 765). IOS Press.
- [5] Otero, Carlos. (2011). Hospital Italiano de Buenos Aires. Buenos Aires, Argentina: Equilibrio entre utilidad y usabilidad, el impacto de los Servicios Terminológicos. http://www.sueiidiss.org/images/stories/EVENTO_2011/Equilibrio_entre_utilidad_y_usabilidad.pdf
- [6] Sonawane, Amol. (18/08/2009). Using Apache Lucene to Search Text. Estados Unidos: IBM. <http://www.ibm.com/developerworks/library/os-apache-lucenesearch/>
- [7] Erlbaumc, M. S. T. (2004). US department of veterans affairs enterprise reference terminology strategic overview. Medinfo.
- [8] Plan de e-salud. (11/04/2016). Servidor de Terminologías “Semantikos” prepara su desarrollo e implementación tras reciente licitación y adjudicación. Chile: Ministerio de Salud - Gobierno de Chile. <http://www.salud-e.cl/prensa/servidor-de-terminologias-semantikos-prepara-su-desarrollo-e-implementacion-tras-reciente-licitacion-y-adjudicacion/>
- [9] U.S. National Library of Medicine. (20/04/2016). Unified Medical Language System (UMLS). Bethesda, Maryland, USA: U.S. National Library of Medicine. <https://www.nlm.nih.gov/research/umls/>
- [10] Humphreys, B. L., Lindberg, D. A., Schoolman, H. M., & Barnett, G. O. (1998). The unified medical language system. *Journal of the American Medical Informatics Association*, 5(1), 1-11.
- [11] U.S. National Library of Medicine. (22/04/2016). UMLS Metathesaurus Vocabulary Documentation. Bethesda, Maryland, USA: U.S. National Library of Medicine. <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html>
- [12] Feliú, Esteban J. (2011). Iniciación a la tecnología Lucene y aplicación.
- [13] U.S. National Library of Medicine. (14/01/2016). 2015AB UMLS MEDCIN Source Information. Bethesda, Maryland, USA: U.S. National Library of Medicine. <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/MEDCIN/>
- [14] Fieschi, M. (2004). Kaiser Permanent’s convergent medical terminology. In *Medinfo 2004* (Vol. 107, p. 346). IOS Press
- [15] <http://getbootstrap.com/>
- [16] Van der Vlist, E., Ayers, D., Bruchez, E., Fawcett, J., & Vernet, A. (2007). Programación Web 2.0.
- [17] <https://xapian.org/docs/bindings/java/>
- [18] (2008-2016). Source Vocabularies. Bethesda, Maryland, USA: U.S. National Library of Medicine. https://www.nlm.nih.gov/research/umls/new_users/online_learning/Meta_002.html
- [19] Cimino, J. J., & Shortliffe, E. H. (2006). *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*. Springer-Verlag New York, Inc..
- [20] Thomas-Brogan, T. (2009). *Health Information Technology Basics: A Concise Guide to Principles and Practice*. Jones & Bartlett Publishers.
- [21] U.S. National Library of Medicine. (06/03/2013). Fact Sheet Unified Medical Language System. Bethesda, Maryland, USA: U.S. National Library of Medicine. <https://www.nlm.nih.gov/pubs/factsheets/umls.html>
- [22] Josiane Gamgo. (2010). *Architecture and Implementation of Apache Lucene*. University of Applied Sciences Mittelhessen. Gießen, Alemania
- [23] Cambridge University Press. (2009). *Stemming and lemmatization*. Stanford, California, USA: Introduction to Information Retrieval. <http://nlp.stanford.edu/IR-book/html/htmledition/ stemming-and-lemmatization-1.html>
- [24] Oracle Corporation. (2013). *JavaServer(TM) Pages Implementation*. Redwood Shores, California, USA: JavaServer Pages Implementation. <https://jsp.java.net/get-started.html>
- [25] U.S. National Library of Medicine. (19/07/2016). About the UMLS. Bethesda, Maryland, USA: U.S. National Library of Medicine. https://www.nlm.nih.gov/research/umls/about_umls.html#Metathesaurus
- [26] U.S. National Library of Medicine. (15/10/2010). Kaiser Permanent Opens Access to CMT to Support HHS Health IT Goals Frequently Asked Questions. Bethesda, Maryland, USA: U.S. National Library of Medicine. https://www.nlm.nih.gov/research/umls/cmt/cmt_faq.html?#q10

- [27] Ministerio de Salud Pública y Bienestar Social. SISTEMA NACIONAL DE INFORMACION EN SALUD. Asunción, Paraguay: Ministerio de Salud Pública y Bienestar Social. http://www.mspbs.gov.py/digies/wp-content/uploads/2012/01/2011_avances-sinai-agosto.pdf
- [28] Ministerio de Salud Pública y Bienestar Social. (14/12/2014). Culmina curso de "Clasificación Internacional de Enfermedades CIE 10". Asunción, Paraguay: Ministerio de Salud Pública y Bienestar Social. <http://www.mspbs.gov.py/culmina-curso-de-clasificacion-internacional-de-enfermedades-cie-10/>
- [29] U.S. National Library of Medicine. (20/03/2017). Metamorphosys Help. Bethesda, Maryland, USA: U.S. National Library of Medicine. https://www.nlm.nih.gov/research/umls/implementation_resources/metamorphosys/help.html
- [30] Aaron Bangor; Philip Kortum; James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, 2009.
- [31] Jeff Sauro. (02/2011). Measuring Usability With The System Usability Scale (SUS). Denver, Colorado, USA: Measuring U. <http://www.measuringusability.com/sus.php>
- [32] U.S. National Library of Medicine. (22/02/2017). Appendix to the License Agreement for Use of the UMLS Metathesaurus. Bethesda, Maryland, USA: U.S. National Library of Medicine. https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/license_agreement_appendix.html
- [33] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.
- [34] Nielsen, J. (1994). *Usability engineering*. Elsevier. ISO 690
- [35] SNOMED CT The Global Language of Healthcare. London, United Kingdom: SNOMED CT The Global Language of Healthcare. <http://www.snomed.org/snomed-ct/>
- [36] (29/11/2016). Classifications. World Health Organization. <http://www.who.int/classifications/icd/en/>