

**EL MODELO SOCIAL DE LA TRADUCCIÓN Y SU
APLICACIÓN A LOS BUSCADORES ESPECIALIZADOS
EN SALUD EN INTERNET: EL BUSCADOR DE
ENFERMEDADES NEUROMUSCULARES DE ASEM
(MARCO TEÓRICO E IMPLEMENTACIÓN
COMPUTACIONAL)¹**

Joan Miquel-Vergés
Universidade de Vigo
jmv@uvigo.es

Elena Sánchez-Trigo
Universidade de Vigo
etrigo@uvigo.es

Resumen: el uso de Internet como fuente de información sanitaria tanto por parte de profesionales como de usuarios o pacientes es cada vez más importante; sin embargo, identificar la información relevante y válida puede ser problemático. Nuestro objetivo es, en primer lugar, investigar la eficacia de los motores de búsqueda; y, después, determinar la calidad de la información en línea referente a un subdominio médico específico: el de las enfermedades neuromusculares. En este artículo presentaremos el marco teórico y su posterior implementación computacional para la elaboración y explotación de un corpus electrónico textual y bilingüe (francés-español), relacionado con las enfermedades neuromusculares.

Palavras-chave: información sanitaria, enfermedad neuromuscular, corpus textual, buscador especializado, web semántica.

Abstract: The Internet provides users with access to online health information; however, identifying relevant and valid information can be problematic. Our objectives were firstly to investigate the efficiency of search-en-

gines, and then to assess the quality of online information pertaining to the specialized medical area of neuromuscular diseases. In this paper we present the theoretical framework and its computational implementation for the elaboration and exploitation of an electronic bilingual corpus (French-Spanish) related to neuromuscular diseases.

Keywords: health information, neuromuscular disease, textual corpus, specialized search-engine, semantic web.

1. El uso de Internet como fuente de información sanitaria

1.1. Ámbito global: estado del arte

El uso de Internet como fuente de información sanitaria tanto por parte de profesionales como de usuarios o pacientes es cada vez más importante. Millones de personas en todo el mundo buscan información sanitaria en la red, sobre enfermedades y sus tratamientos, acerca de medicamentos y medidas diversas de prevención y diagnóstico e incluso consejos médicos a través de la consulta virtual en el chat o en el correo electrónico. La calidad de esta información es extraordinariamente variable; desde información científica y basada en la evidencia hasta remedios caseros o de muy dudoso origen, incluso de efectos nocivos o perjudiciales. La preocupación de gobiernos, sociedades e instituciones profesionales y científicas así como de los usuarios es creciente y requiere algún tipo de intervención.

Así, por ejemplo, según un reciente estudio de Nancy et al. [9] en el que se compara el uso que médicos de primaria y especializada hacen de Internet en los Estados Unidos, tanto los unos como los otros, consideran que la red es una buena fuente de conocimiento y se sienten capaces de emplear la tecnología para acceder a ella. La diferencia fundamental entre ambos grupos estriba en el hecho de que los médicos de familia acceden a Internet fundamentalmente para profundizar en su conocimiento de las patologías que ven

diariamente en su consulta, mientras que sus homólogos de hospital se muestran más inclinados a leer revistas científicas en línea.

De esta forma, al igual que los sitios webs con información biomédica han proliferado enormemente, siendo de los más visitados por los navegantes, las iniciativas para garantizar y mejorar la calidad de la información ofrecida por estos portales ha seguido un camino paralelo, existiendo actualmente un importante número de instrumentos para el control, supervisión y evaluación de los contenidos biomédicos en la red.

El objetivo básico de la mayoría de estos instrumentos que intentan avalar la información biomédica, ha sido garantizar al navegante que un determinado sitio web cumple con una serie de requisitos de calidad, principalmente referidos a ofrecer información transparente sobre los autores, los patrocinadores y la existencia de conflictos de interés. Las iniciativas propuestas están orientadas fundamentalmente hacia la acreditación, certificación, autorregulación, sistemas de calificación y concesión de sellos de calidad indicativos del cumplimiento de criterios concretos [2].

Existen destacados estudios publicados sobre la calidad de la información biomédica disponible en Internet en inglés y español, pero es difícil encontrar trabajos sobre las iniciativas hispanas que velan por la calidad de esta información mediante métodos de acreditación-certificación. Uno de ellos es el trabajo de García et al. [6], que muestra la experiencia de cada uno de los proyectos encontrados, comparando los métodos aplicados y valorando de esta forma la convivencia en la red Internet de los sellos de calidad otorgados por cada iniciativa.

Pero los usuarios, principalmente los pacientes, difícilmente conocen todos los sistemas de acreditación de los sitios web y más difícil es que entiendan el valor que pueden tener. Por ello, hemos de llegar a la conclusión de que la valoración de la calidad de la información biomédica en Internet no debe reducirse a saber si los sitios web cumplen con unos criterios estáticos. Deberíamos conocer en qué medida la calidad de esta informa-

ción influye sobre los usuarios, tanto profesionales como público general.

La propuesta de realizar una plataforma, reuniendo la experiencia de todas las iniciativas que trabajan sobre la calidad de la información biosanitaria en Internet, implicando en ella a los productores de la información y a sus usuarios finales, puede mejorar la información del ciudadano, paciente o profesional y evitar su confusión, además de mejorar el asesoramiento sobre los contenidos biomédicos de Internet, evitando que información de baja calidad llegue a los usuarios.

En esta labor se encuentra la Unión Europea, que estudia el desarrollo de una acción conjunta entre los países miembros para establecer un sistema comunitario reconocible de sellos de calidad para sitios web que dispongan de información sobre salud, siendo deseable una plataforma multilingüe que unifique experiencias. Este es el trabajo conjunto del proyecto MedCircle y HON.

Lo que queda ya fuera de toda duda o interpretación es el hecho de que, en poco más de una década desde su aparición, Internet se ha convertido en un instrumento de uso cotidiano en nuestra sociedad, comparable a otros medios tan importantes como la radio, la televisión o el teléfono, a los cuales aventaja en muchos aspectos. Paralelamente al crecimiento espectacular de la web, las tecnologías que la hacen posible experimentaron una rápida evolución, permitiendo una web mejor, más amplia, más potente, más flexible, o más fácil de mantener. Estos cambios influyen y son al mismo tiempo influidos por la propia transformación de lo que entendemos por WWW. La generación dinámica de páginas, el acoplamiento con bases de datos, la mayor interactividad con el usuario, la concepción de la web como plataforma universal para el despliegue de aplicaciones, la adaptación al usuario, son algunas de las tendencias evolutivas más marcadas en los últimos años.

Con todo, son diversos los aspectos susceptibles de mejora. Entre estas últimas tendencias que pueden repercutir en el futuro de la web a medio plazo aparece la "web semántica", cuyo último fin

es lograr que las máquinas puedan entender, y por lo tanto utilizar, lo que contiene la web. Esta nueva web estaría poblada por agentes o representantes software capaces de navegar y de realizar operaciones por nosotros para ahorrarnos trabajo y optimizar los resultados. Para conseguir esta meta, la web semántica propone describir los recursos de la web con representaciones procesables (es decir, entendibles) no solamente por personas, sino también por programas que puedan asistir, representar, o sustituir a las personas en tareas rutinarias o inabarcables para un humano. Las tecnologías de la web semántica buscan desarrollar una web más cohesionada, donde todavía sea más fácil localizar, compartir e integrar información y servicios, para sacar un partido todavía mayor de los recursos disponibles en la web.

Esta concepción de web semántica se corresponde con una nueva visión del concepto de salud introducida por la OMS (Organización Mundial de la Salud) en el año 2001, basada en la integración de dos modelos antagónicos: el “modelo médico” y el “modelo social” [10]. El primero se corresponde con una visión más tradicional y concibe la salud como un problema individual, relacionado de forma directa con una enfermedad que requiere un tratamiento médico concreto. El modelo social se inscribe, sin embargo, en un contexto más amplio, que necesariamente remite a situaciones que, al menos parcialmente, están originadas por el entorno social y que afectan por tanto a toda la sociedad. Esta nueva orientación hace énfasis en la importancia de considerar la implicación de la salud en todos los foros de debate social, lo que ha dado lugar a un cambio importante en la proyección que tradicionalmente se atribuía a la salud.

1.2. Ámbito específico: el subdominio médico de las enfermedades neuromusculares

El punto de partida de nuestro proyecto de investigación específico lo hallamos en la solicitud de colaboración externa que en el

año 2004 la Asociación Gallega contra las Enfermedades Neuromusculares (ASEM-Galicia) hizo a un equipo de investigadores de la Universidade de Vigo y del Hospital del Meixoeiro, integrados respectivamente por especialistas en traducción e informática y en el tratamiento e investigación sobre enfermedades neuromusculares. La finalidad era aunar los conocimientos de los diferentes equipos implicados para crear una base de representación del conocimiento que fuese de referencia en el ámbito de las enfermedades neuromusculares y de utilidad e interés para todos los implicados de una manera directa (enfermos y entorno próximo) o indirecta (profesionales del ámbito sanitario, profesionales e investigadores de otras áreas como laboratorios, traductores, etc. o administraciones públicas).

Como ya hemos visto en el apartado anterior, partimos de la base que:

- La calidad de la información médica existente en Internet es extraordinariamente variable; desde información científica y basada en la evidencia hasta remedios caseros o de muy dudoso origen, incluso de efectos nocivos o perjudiciales. La preocupación de gobiernos, sociedades e instituciones profesionales y científicas así como de los usuarios es creciente y requiere algún tipo de intervención. Las iniciativas propuestas están orientadas fundamentalmente hacia la acreditación, certificación, autorregulación, sistemas de calificación y concesión de sellos de calidad indicativos del cumplimiento de criterios concretos. Pero los usuarios, principalmente los pacientes, difícilmente conocen todos los sistemas de acreditación de los sitios web y más difícil es que entiendan el valor que puedan tener.
- Esta valoración global es también aplicable al subdominio médico de las enfermedades neuromusculares; con el “añadido” que este tipo de enfermedades son crónicas, generalmente progresivas y de gravedad muy diversa, encuadradas dentro de lo que se conoce

como “enfermedades raras”. Las variadas consecuencias que pueden tener estas enfermedades hacen que en este ámbito confluya un buen número de especialidades médicas (como la neurología, la miología, la cardiología, etc.). Se trata, además, de un ámbito en el que están presentes muchas otras áreas relacionadas con la salud en su sentido más amplio (como la fisioterapia, la ortopedia, la psicología, la nutrición, aspectos relacionados con la calidad de vida de los enfermos, etc.).

- Por ello, en una primera fase de nuestro proyecto, nuestra propuesta se concretó en la realización de una plataforma, reuniendo la experiencia de todas las iniciativas que trabajan (directa o indirectamente) en dicho ámbito médico, e implicando en ella a los productores de la información y a sus usuarios finales. Con ello creímos poder mejorar la información del ciudadano, paciente o profesional y evitar su confusión, además de mejorar el asesoramiento sobre los contenidos médicos, evitando que información de baja calidad llegase a los usuarios.
- Además de la creación del portal de ASEM [1] (un sitio web especializado en el tratamiento de todo tipo de información relacionada con las enfermedades neuromusculares), la concreción de nuestro modelo teórico se basó, en primer lugar, en la creación de un corpus electrónico textual bilingüe (español-francés), a partir de un volumen reducido de textos de alta calidad relacionados con las enfermedades neuromusculares, creados por especialistas (en francés) y traducidos al español (por traductores profesionales especializados en la traducción médica).
- En segundo lugar, procedimos a la elaboración de un motor de búsqueda documental para la extracción de información de dicho corpus. Aunque vimos que, desde un punto de vista práctico, parecía factible la construcción de un motor de búsqueda

semántica de enfermedades neuromusculares; sin embargo, debido al condicionante temporal de nuestro proyecto de investigación y a las necesidades de resultados inminentes por parte de ASEM-Galicia, nos decidimos empezar con la implementación de un motor de búsqueda por palabras clave.

Así, pues, la implementación computacional de nuestro modelo teórico hubo de hacerse patente mediante la creación de un corpus bilingüe (francés y español) con textos sobre enfermedades neuromusculares; y, una herramienta de consulta documental, que realizase una búsqueda directa sobre el corpus, tanto en castellano como en francés.

Probamos el sistema en un entorno acotado de experimentación para, con posterioridad, y después de requerir de los profesionales (de todos los ámbitos implicados) su parecer; en el caso de que los resultados sean satisfactorios, poner la herramienta en Internet, a disposición de toda la comunidad.

2. Creación de un corpus textual de enfermedades neuromusculares y elaboración de una herramienta de consulta documental

2.1. Creación de un corpus textual sobre enfermedades neuromusculares

La primera parte del proyecto tiene como finalidad la creación de un corpus bilingüe (francés-español) con textos sobre enfermedades neuromusculares (al cual denominaremos MYOCOR [12] [13]). Dicho corpus debía ser lo suficientemente representativo desde el punto de vista de la actualidad científica y social, y tener en consideración una gran variedad de géneros textuales.

Para la obtención de los textos iniciales en francés acudimos a la *Association Française contre les Myopathies* (AFM), ya que la

Asociación Gallega contra las Enfermedades Neuromusculares (ASEM-Galicia) disfruta de un convenio gracias al cual ponen a su disposición todas sus publicaciones para realizar una utilización de las mismas en beneficio de los enfermos y sin ánimo de lucro.

Estas publicaciones se caracterizan por ser textos producidos por diferentes tipos de sujetos e instituciones implicadas en el ámbito de las enfermedades neuromusculares. Se incluyen, por tanto, textos elaborados por médicos o profesionales sanitarios con diferentes niveles de especialización (biólogos, enfermeros, fisioterapeutas, etc.), por investigadores y por diversas instituciones. De forma análoga, los receptores de los textos también se corresponden con un amplio abanico que incluye no sólo a los profesionales que acabamos de citar y a los enfermos, sino a la sociedad en su conjunto. Se trata de textos de gran calidad desde el punto de vista médico y lo suficientemente representativos desde un punto de vista de la variedad de géneros textuales y de su actualidad.

Tomando como punto de partida los textos producidos por la AFM y traducidos por personal docente e investigador de la Facultade de Filología e Traducción de la Universidade de Vigo (especializados en la traducción de textos médicos), elaboramos una base de recursos documentales bilingüe (en francés y en español) para ponerlos, a través de una página web especializada en enfermedades neuromusculares, a disposición de la sociedad gallega, del resto del estado y de los posibles interesados de otros países (como las asociaciones de enfermos Iberoamericanas que ya han mostrado su interés en el mismo).

En este aspecto, enlazaría con otros corpus y proyectos de reciente elaboración, de carácter nacional e internacional, como son: MEDICOR (*Medical Copus*), LIQUID (*Language Independent QUerying for Information Discovery*), u ONCOTERM (Sistema Bilingüe de Información y Recursos Oncológicos [5] [7]), especializados en diferentes ámbitos médicos.

Hemos de tener en cuenta el hecho que el usuario final tenga acceso no sólo al texto en su formato original (ya sea en francés o

en español), sino también a información parcial del mismo, según sean sus preferencias. Si bien en un primer momento valoramos la posibilidad de trabajar con ontologías, dada la complejidad del proceso y la urgencia “social” del mismo, al final, optamos por estructurar *sólo* los documentos en “parcelas de información”, mediante la conversión de los documentos originales (independientemente de cual fuese su formato) en documentos XML estructurados.

Para ello [3] [4], empleamos un DTD (siglas de *Document Type Definition* o “Definición de Tipo de Documento”) para determinar la estructura de dichos documentos XML. El DTD describe cada elemento admisible dentro del documento, los atributos posibles y (opcionalmente) los valores de atributo permitidos para cada elemento. Además, describe los anidamientos y ocurrencias de elementos. Se compone, como la mayoría de los DTD, de definiciones de ELEMENT (para definir los elementos que componen el documento) y definiciones de ATTLIST (declaración de atributos de un elemento), que proporcionan información adicional al procesador XML.

Definimos los elementos del *documento* XML, que son la *cabecera* del documento y el *contenido* del documento. Como atributos del *documento* consideraremos el *tipo* de documento, que podrá ser: una *comunicación*, una *comunicación breve*, un *artículo especializado*, un *artículo divulgativo*, una *ficha*, un *folleto especializado*, un *folleto divulgativo*, un *testimonio*, una *monografía especializada*, una *monografía divulgativa*, un *manual*, una *conferencia*, una *jurisprudencia*; u *otro* tipo de documento.

La cabecera está formada por el *título* del documento, el *autor* del documento, la *fecha* de creación del documento, la *editorial* de publicación del documento, el autor de la *traducción*, el responsable de la *inserción* de las marcas, la *ruta* al documento original en PDF, *más* información a destacar, información de *contacto*; y, finalmente, la *referencia* bibliográfica del documento .

El *contenido* del documento, por otra parte, está formado por el *inicio* del documento, el *cuerpo* del documento, y un *apéndice*. El

inicio del documento está constituido por la portada, el resumen o *abstract*, la *introducción*, la *dedicatoria*, el *prólogo*, y el *epígrafe*.

El *cuerpo* del documento está formado por la *enfermedad* a la cual hace referencia y el *texto* que aparece en él. Por su parte, la *enfermedad* tiene como atributo asociado su nombre o acrónimo, y está formada por su *descripción*, las *causas* que la producen, sus *síntomas*, la *evolución* de la enfermedad en el tiempo, su *transmisión*, su *diagnóstico*, su *tratamiento*, la *investigación* que sobre la misma se ha realizado, aspectos relacionados con la *vida* diaria del paciente, aspectos relacionados con su *estudio*; y *otros* aspectos remarcables. En el apartado de *enfermedad* debemos incluir todos los nombres de las enfermedades que queremos que después aparezcan en el menú desplegable de la “búsqueda por enfermedad”.

Por último, el *apéndice* del texto está formado por la *conclusión*, la *bibliografía*, el *glosario*, y el *índice*.

Los elementos existentes en el DTD deben estar pensados para que satisfagan las diferentes expectativas de los diferentes tipos de usuarios potenciales. Así, los elementos “introducción, enfermedad, descripción de la enfermedad, causas, síntomas, evolución, transmisión, diagnóstico, tratamiento, investigación, vida diaria, estudio y contacto”, están pensados para satisfacer las necesidades de los pacientes, familiares de los pacientes, o médicos (que buscan información de primera mano de la enfermedad); mientras que, elementos como “glosario”, pueden ser de mayor interés para terminólogos o traductores; o, elementos como “ruta”, serán de uso interno para buscador (le indicarán donde están guardados los textos originales en PDF).

El proceso de conversión [13] que tiene lugar entre el documento original (en francés, en un formato indeterminado) y el documento final en XML “bien formado” (es decir, que cumple con todas las definiciones básicas de formato y pueden, por lo tanto, ser analizados correctamente por cualquier analizador o *parser* que cumpla con la norma) y “válido” (es decir, que su estructura se corresponde con la definida en el DTD) es el siguiente (figura 1):

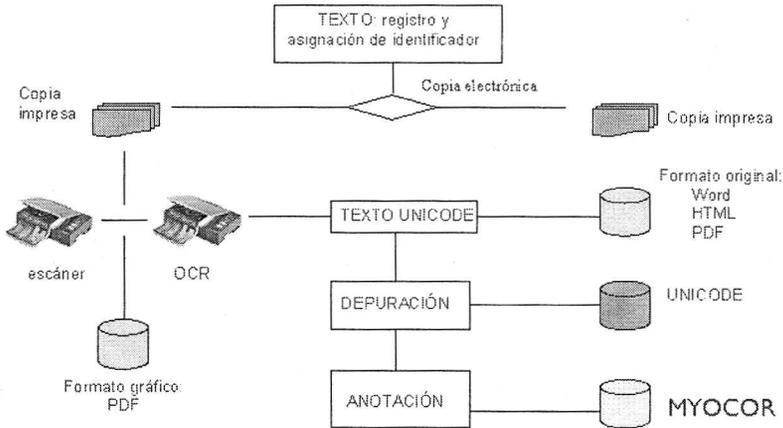


Figura 1: “procesos en la creación del corpus MYOCOR”

Para facilitar el cambio de formato de UNICODE a XML creamos diversos “scripts” para intentar automatizar la inserción de las marcas XML. En algunos casos, como son los elementos de la *cabecera*, el proceso puede semiautomatizarse (por ejemplo, pueden insertarse las marcas “vacías”) o, incluso, automatizarse totalmente (como es el caso de la *traducción* o la *inserción*). En otros casos, sin embargo, el proceso de inserción de los elementos (marcas) de la enfermedad debe hacerse de manera totalmente manual (es el caso, por ejemplo, de los elementos de la enfermedad).

2.2. Creación de una herramienta de consulta documental para MYOCOR

2.2.1. El buscador de enfermedades neuronales de ASEM: marco teórico

Muchos modelos de búsqueda de textos en las páginas web de Internet dependen de un emparejamiento de palabras “al pie de la

letra” entre las palabras que busca el usuario y las que existen en las páginas web. Se plantean dos situaciones antitéticas: por un lado, el buscador encuentra términos homónimos y, por otro los sinónimos le pasan desapercibidos. En el primer caso, son recuperadas páginas cuyo tema no nos interesa; en el segundo, son ignoradas páginas en las que estaríamos interesados.

A estos dos factores distorsionantes hemos de sumarle todavía otro: el multilingüismo. Tiene lugar cuando realizamos una consulta en el buscador introduciendo términos en una lengua determinada y encontramos páginas web relacionadas con lo que nos interesa, pero escritas en otra lengua diferente a la empleada para los términos.

En todos estos casos, el problema es que la consulta no identifica de forma precisa aquello en lo que estamos interesados sino que solamente identifica un término que en cierto idioma tiene entre sus significados aquello que estamos buscando. Este procedimiento de búsqueda en el que el resultado de la búsqueda es el conjunto de documentos que contengan los términos introducidos en la consulta se denomina “búsqueda sintáctica”.

Para poder ir más allá de esta búsqueda sintáctica (o lineal), numerosos autores han manifestado sus puntos de vista al respecto, proponiendo múltiples soluciones (unas más teóricas que otras) desde diferentes ámbitos de investigación.

Algunos autores² [11] proponen que si las páginas web tuviesen asociadas anotaciones formales que identificasen inequívocamente los conceptos y entidades principales contenidos en las páginas, un buscador (semántico) podría evitar cometer los tres tipos de errores descritos antes y encontrar sólo los documentos relacionados exactamente con aquello en lo que realmente estábamos interesados al realizar la consulta.

Sin embargo, como ellos mismos reconocen, disponer de anotaciones semánticas puede no ser suficiente para las aplicaciones que utilicen la web semántica. La mayoría de estas aplicaciones precisan de un modelo del dominio en el que operen que incluya el

vocabulario de los conceptos relevantes a ese dominio; y, probablemente, las propiedades que relacionan los diferentes conceptos, así como las reglas que gobiernan ese dominio. Partiendo de ese modelo, el sistema sería capaz de obtener conclusiones y/o tomar decisiones procesando las anotaciones extraídas de páginas web. Tales modelos son definidos por medio de las “ontologías³”.

Así, una forma de entender un motor de búsqueda semántica sería como la de una herramienta que recibe consultas basadas en ontologías, las ejecuta contra una base de conocimientos, y devuelve tuplas que satisfacen la consulta. Estas técnicas típicamente utilizan modelos booleanos de búsqueda y se basan en una visión “ideal” del espacio de información, consistente en piezas formales de conocimiento ontológico sin ambigüedad ni redundancia.

En la práctica [14], sin embargo, es bien conocido que existen límites respecto al punto hasta donde el conocimiento se puede formalizar usando ontologías.

- En primer lugar, debido al enorme volumen de información disponible hoy en día en forma de texto y contenidos multimedia no estructurados, convertir esta cantidad de información en conocimiento ontológico con un coste viable es un problema sin resolver en general.
- En segundo lugar, los documentos tienen un valor por sí mismos, y no son equivalentes a la suma de sus partes. Aunque es útil descomponer documentos en unidades de información menores que puedan ser reutilizadas y ensambladas para diferentes propósitos, a menudo es apropiado mantener los documentos originales en el sistema.
- En tercer y último lugar, si no se prevé ningún criterio claro de graduación, el sistema de búsqueda podría llegar a ser inútil si el espacio de búsqueda fuese demasiado grande.

Para la implementación de nuestro buscador documental sobre enfermedades neuromusculares, sin embargo, los límites respecto al punto hasta donde el conocimiento se puede formalizar usando ontologías son menos problemáticos:

- Por una parte, el trabajar con un volumen inicial limitado de textos hace que sea mucho más factible convertir dicha información en conocimiento ontológico, estructurando los documentos de texto y contenidos multimedia (mayoritariamente imágenes). A medida que se vayan incorporando nuevos textos, sin embargo, el problema se hará cada vez más evidente.
- Por otro lado, para respetar el principio de que los documentos tienen un valor por si mismos y no son equivalentes a la suma de sus partes, trabajaremos siempre con dos copias del mismo original. Una de las copias permanecerá “inalterable” (en formato PDF); la otra copia será convenientemente estructurada (en formato XML).
- En tercer lugar, y para prevenir restringir el abaste y volumen de las búsquedas, tomaremos como punto de partida el “nombre de las enfermedades neuromusculares”. El número de enfermedades neuromusculares, aún siendo elevado (estamos hablando del orden de 150-200 enfermedades conocidas), es, de por si, limitado.

Así pues, desde un punto de vista práctico, parece factible la construcción de un motor de búsqueda semántica de enfermedades neuromusculares. Sin embargo, y tal como apuntan Mayfield y Finin [8]: “es mejor que la búsqueda semántica sea un complemento de la búsqueda por palabra clave, mientras no haya suficientes ontologías y metadatos disponibles”.

Por ello, y debido al condicionante temporal de nuestro proyecto de investigación y a las necesidades de resultados inminentes por parte de ASEM-Galicia, decidimos empezar con la implementación de un motor de búsqueda, por palabras clave, de enfermedades neuromusculares. Para después, en una segunda fase, proceder a la construcción de las ontologías y metadatos e implementarlos en un motor de búsqueda, semántico, de enfermedades neuromusculares.

Otro condicionante de nuestro trabajo de investigación tiene una estrecha relación con el “modelo social” de la salud del que hemos hablado antes y del fin último de nuestro proyecto. Esta “socialización” del entorno había de reflejarla también en las herramientas informáticas empleadas. Por ello decidimos trabajar con software de código abierto. Los factores que nos animaron a ello son muchos y de naturaleza muy diversa⁴, pero podrían resumirse en las conclusiones del Informe Wheeler [15]:

“El código abierto tiene una cuota de mercado significativa, es quizá el más fiable, y en muchos casos, tiene el mejor rendimiento. El código abierto escala, tanto en tamaño del problema como en tamaño del proyecto. El código abierto generalmente tiene mejor seguridad, particularmente cuando se compara con Windows. El coste total de propiedad para el código abierto es normalmente menor que para el código cerrado, particularmente conforme crece el número de plataformas. Estas aseveraciones no son meramente opiniones. Estos efectos se pueden demostrar cuantitativamente, utilizando una amplia variedad de medidas. Esto incluso no considera otros aspectos difíciles de medir, como la libertad frente al control por una única fuente, la libertad frente a la gestión de licencias con su correspondientes pleitos y la creciente flexibilidad. Creo que las opciones abiertas deberían considerarse cuidadosamente en cualquier momento en el que se necesite programas o hardware para ordenadores.”

Por último, siempre tuvimos en mente que las enfermedades neuromusculares están asociadas a individuos y entornos sociales. El garantizar y mejorar la calidad de la información ofrecida en nuestro portal fue siempre nuestro fin último y primordial. Pero también lo fue el conocer en qué medida la calidad de esta información podía influir sobre los usuarios, tanto profesionales como público general. Por ello la información que se muestra al usuario al final de su búsqueda se muestra en su “versión completa” (original) para dar respuesta a los profesionales e investigadores; mientras que, para aquellas personas (usuarios) que tengan unas necesidades más puntuales (por ejemplo, saber cuales son las causas, los síntomas, el tratamiento, etc. de una determinada enfermedad), hay una “selección de la información” mostrada. Esto lo conseguimos trabajando con dos versiones del texto original; mientras una se muestra inalterable y con todos sus elementos gráficos, la otra es estructurada (siguiendo un “patrón” común en todos los textos) y tratada convenientemente (usando metadatos) para que el buscador pueda acceder fácilmente a dicha subinformación.

Antes de poner la herramienta en línea, accesible para todo el mundo, es necesario también establecer un “período de prueba” para evaluar su eficacia (en términos humanos y computacionales”) y la calidad de la información mostrada.

2.2.2. El buscador de enfermedades neuromusculares de ASEM: implementación

Nuestro motor de búsqueda, inicialmente, está basado en la búsqueda por palabras clave. A fin de satisfacer todas las necesidades puntuales de sus diferentes usuarios, ha de ser capaz de analizar los textos estructurados y mostrar los textos completos que cumplan las condiciones reclamadas por los usuarios en su búsqueda; pero también ha de ser capaz de realizar búsquedas parciales en los textos sobre aquellos ámbitos o parcelas del conocimiento que el usuario quiera conocer.

Mediante la notación UML⁵, de los casos de uso, intentaremos entender y mostrar lo que hace el programa desarrollado. Se trata básicamente de analizar el sistema y mostrar, de una manera formal el modelo teórico presentado, para después implementarlo (computacionalmente). También intentamos explicar y concretar otros aspectos como, por ejemplo, los requisitos no funcionales que no pueden ser mostrados con los casos de uso.

▪ **Planteamiento general y dominio del problema**

A grandes rasgos, como ya hemos visto, el problema radica en la explotación de un corpus textual anotado relacionado con las enfermedades neuromusculares, de gran calidad y bilingüe (francés y español). Para ello era fundamental, en primer lugar, la creación de un corpus de textos especializados en el subdominio médico de las enfermedades neuromusculares, bilingüe (francés-español) y representativo (que incluya gran variedad de géneros textuales y sean de actualidad).

La creación de dicho corpus permitirá la elaboración de una herramienta de consulta en línea para obtener información sobre dichas enfermedades neuromusculares. Dicha herramienta realizará una búsqueda directa sobre el corpus, tanto en castellano como en francés. Ya hemos visto las bases teóricas de dicho corpus textual, pero no es nuestro propósito entrar en detalle en cuanto al análisis de su creación. Nos centraremos sólo en su explotación mediante un motor de búsqueda.

▪ **Usuarios**

El número de usuarios de la aplicación a desarrollar será de uno sólo, ya que este tipo de aplicaciones son claramente monousuario, debido a la funcionalidad que tienen.

▪ Diagrama de casos de uso

Ningún sistema se encuentra aislado. Cualquier sistema interactúa con actores humanos o mecánicos que lo utilizan con algún objetivo. Un “caso de uso” especifica el comportamiento de un sistema o una parte del mismo, y es una descripción de un conjunto de secuencias de acciones, donde cada secuencia representa la interacción de los elementos externos del sistema (sus actores) con el propio sistema. Un caso de uso representa un requerimiento funcional del sistema.

Los casos de uso se emplean para capturar el comportamiento deseado del sistema en desarrollo, sin tener que especificar cómo se implementa ese comportamiento. Proporcionan un medio para que los desarrolladores, los usuarios finales del sistema y los expertos del dominio lleguen a una comprensión común del sistema. Además ayudan a validar la arquitectura y a verificar el sistema mientras evoluciona a lo largo del desarrollo.

El diagrama de casos de uso del buscador de ASEM tiene un solo actor o “Usuario”, y dos casos de uso que son “Consulta por enfermedad” y “Consulta por palabra clave” (figura 2).

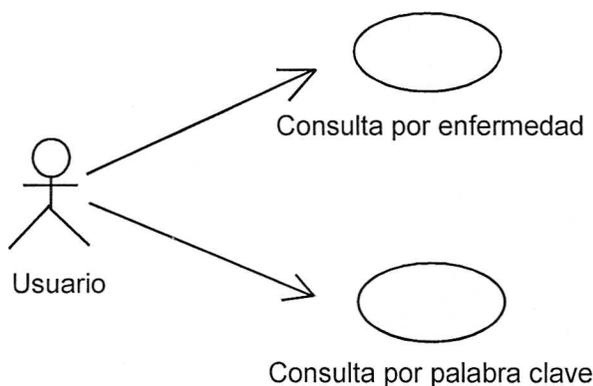


Figura 2: “diagrama de casos de la aplicación”

▪ Diagrama de clases

Las clases se documentan con una descripción de lo que hacen, sus métodos y sus atributos. Las relaciones entre clases se documentan con una descripción de su propósito, su cardinalidad (cuantos objetos intervienen en la relación) y su opcionalidad (cuando un objeto es opcional el que intervenga en una relación).

Nuestra aplicación está compuesta por cinco clases (figura 3): a) la clase “buscador”, encargada de recibir las consultas del usuario y devolver los resultados de la ejecución de la consulta; b) la clase “documento”, que permite la recuperación de información de los documentos XML almacenados; c) la clase “configuración”, que contiene información de configuración de la herramienta y, finalmente; d) la clase “operaciones”, que realiza operaciones auxiliares de la aplicación.

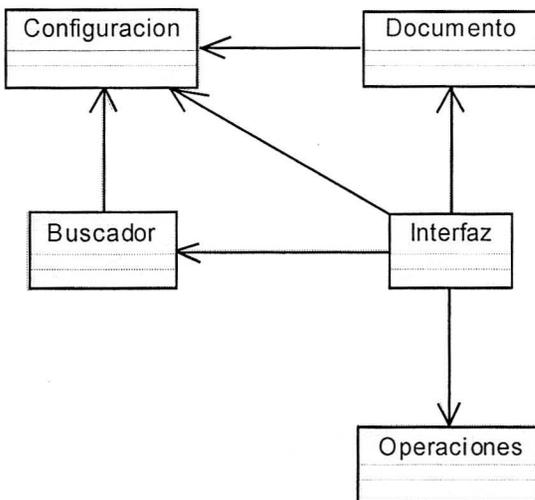


Figura 3: “diagrama de clases de la aplicación”

▪ Diagramas de secuencia

Los diagramas de secuencia muestran las interacciones entre objetos, ordenadas en secuencia temporal (figuras 4 y 5). Muestran los objetos que se encuentran en el escenario y la secuencia de mensajes intercambiados entre los objetos para llevar a cabo la funcionalidad descrita por el escenario. Nuestra aplicación se ha diseñado de manera que ofrezca la posibilidad de realizar dos tipos de búsqueda o consulta. Es por ello que distinguiremos dos diagramas de secuencia: a) la consulta por palabra clave (figura 4); y, b) la consulta por enfermedad (figura 5).

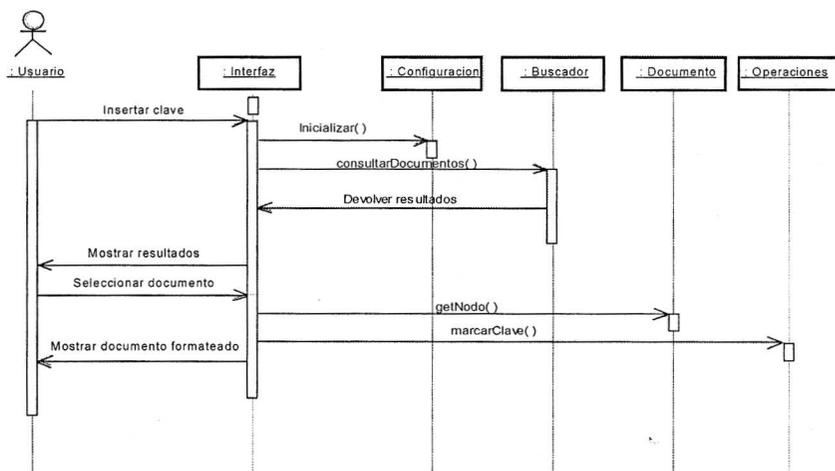


Figura 4: "Consulta por palabra clave"

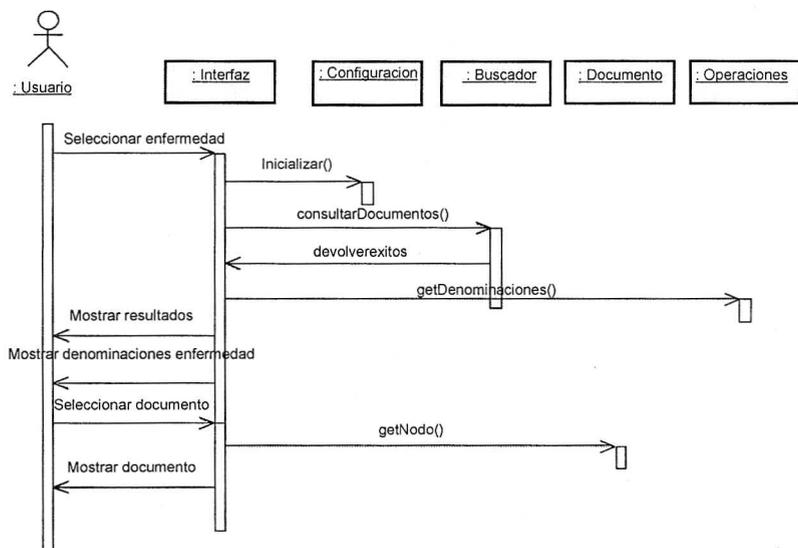


Figura 5: "Consulta por enfermedad"

La primera de estas búsquedas (consulta por palabras clave), permite al usuario acceder a la información a través de palabras clave, con la finalidad de obtener una información organizada. El usuario introduce la denominación de una enfermedad y realiza la selección de la información que quiere obtener en relación con la totalidad o parte de aspectos específicos como: descripción, causas, síntomas, evolución, transmisión, diagnóstico, tratamiento, investigación y vida diaria.

Por ejemplo, si un usuario desea obtener información sobre el diagnóstico de la enfermedad "miopatía miotubular", en el caso de la búsqueda por palabras clave, tras introducir el nombre de la enfermedad en *palabras clave* (para buscar textualmente varias palabras deberá introducir dichas palabras entre comillas), deberá seleccionamos "diagnóstico" en *opciones* (figura 6).

Figura 6: “Ejemplo de búsqueda por palabras clave”

El resultado obtenido será el siguiente (figura 7):

Figura 7: “Resultado de la búsqueda por palabras clave”

mento (figura 9).

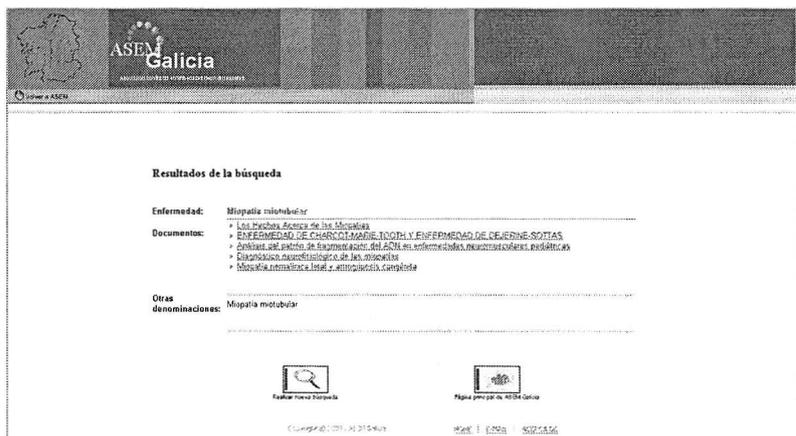


Figura 9: “Resultado de búsqueda por nombre de enfermedad”

Observamos que, si realizamos una búsqueda textual (entre comillas) de la enfermedad, en ambos casos (búsqueda por palabras clave y búsqueda por nombre), los documentos hallados por el sistema son los mismos.

Sin embargo, en el caso de búsquedas por palabra no textuales (sin comillas) el sistema realiza una búsqueda exhaustiva de cada una de las palabras clave (juntas o por separado, apareciendo marcadas en los textos en color rojo⁶); y, por ello, el número de documentos obtenidos será mucho mayor y algunos de ellos no tendrán relación alguna con los resultados esperados.

2.2.3 El buscador de enfermedades neuromusculares de ASEM: diseño de la aplicación

En este apartado explicaremos cómo se realizó la aplicación [2] [3]. Nos centraremos en la arquitectura hardware y software del sistema y resaltaremos algunas características remarcables del proceso de instalación (que entendemos son importantes por el he-

cho de trabajar con software de código abierto).

Esta aplicación utiliza una arquitectura Cliente-Servidor, proporcionando servicio a los usuarios a través de Internet, mediante el protocolo TCP/IP. Un usuario se conecta desde un ordenador que funciona como cliente y ejecuta su explorador de Internet para acceder a la aplicación.

El equipo servidor donde reside el programa ejecuta el servidor web *Apache Tomcat*, que procesa las peticiones del cliente e interactúa con *Lucene* y *Xerces*, a través de los cuales podrá acceder a los índices de búsqueda y a los documentos del corpus.

La aplicación se ha desarrollado a partir del uso de software *open source* (libre y gratuito) y en estos momentos ha finalizado su fase de prueba en castellano.

Se ha recurrido, con adaptaciones, al sistema Apache. El esquema de la aplicación es el siguiente (figura 10): *Lucene* indexa los documentos y realiza las búsquedas solicitadas por los usuarios; *Xerces* recupera la información de los documentos; *Tomcat* y las clases que contiene la lógica de la aplicación. Opcionalmente se puede utilizar Apache conjuntamente con *Tomcat* como servidor de aplicaciones para optimizar el trabajo del servidor web. Un navegador web permite el acceso a la aplicación y la interacción del usuario con la aplicación para realizar las búsquedas que desee.

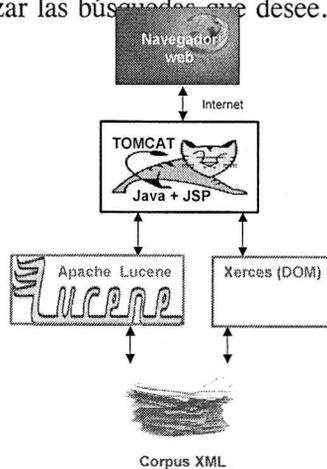


Figura 10: "Diseño de la aplicación"

▪ **Arquitectura hardware**

La aplicación debe ser instalada en un servidor para permitir que los clientes accedan a ella a través de Internet. El servidor donde se instale la aplicación necesita los siguientes requisitos mínimos tanto para instalar como para almacenar los datos y ejecutar la aplicación y todos los programas necesarios para su ejecución:

- Procesador a 1GHz o superior.
- Memoria RAM de 512 Mb.
- Disco duro de 40 Gb o superior (el tamaño puede aumentar en función del tamaño del corpus).
- Lector de CD-ROM 32X, para poder instalar tanto los programas como la aplicación.
- Tarjeta gráfica VGA o monitor con mayor resolución.
- Conexión a Internet.

El equipo cliente que acceda a la aplicación solamente necesitará una conexión a Internet para poder trabajar con el programa.

▪ **Arquitectura software**

La aplicación necesita una serie de programas para su correcto funcionamiento, además del software propio del programa. La apli-

cación se desarrolló básicamente sobre Java. Por tanto, la mayor parte de los requisitos software en el servidor tienen relación con la instalación del soporte Java necesario para que funcione el programa. Todo el software utilizado es *open source* (libre y gratuito) y, por ello, puede descargarse libremente de la red y todo aquel que lo utilice tiene acceso al código fuente y puede sugerir a sus creadores la instalación de nuevas utilidades.

- *J2SE*. Es una máquina virtual de Java, cuya versión para Windows puede descargarse de la página de Sun: <http://java.sun.com/j2se/>.
- *Apache*. Es un servidor de páginas Web distribuido por Apache Group. Es un software rápido y fiable; que puede descargarse de la página de Apache en: <http://www.apache.org>.
- *Jakarta-Tomcat*. Es un contenedor de *servlets* (objetos que corren dentro del contexto de un servidor web como es Tomcat, y extienden su funcionalidad) con un entorno JSP (un caso especial de *servlet*). Constituye la capa que en tiempo de ejecución se encarga de manejar e invocar los *servlets* de parte de los usuarios. Tomcat puede ser usado como servidor (contenedor independiente), principalmente para el desarrollo y eliminación de fallos; o como un módulo integrado en un servidor web (soporta Apache, IIS...). En nuestra aplicación lo usaremos junto con Apache para permitir la consulta de las páginas por parte de los clientes. Tomcat necesita un compilador Java instalado en el equipo donde va a trabajar. En nuestro caso éste ya no es un problema pues ya tenemos el J2SE, que contiene un compilador de Java. Para ejecutar Tomcat conjuntamente con Apache necesitamos instalar un conector, Jakarta, que nos permita ejecutar Tomcat como un módulo de Apache. Una vez instalado dicho conector, configuraremos el servidor para que permita acceder a las aplicaciones de Tomcat vía Apache en el puerto 80. Tanto Tomcat como de Jakarta

(en sus versiones para Windows) puede descargarse del sitio web de Jakarta en: <http://jakarta.apache.org>.

- *Lucene*. Es una herramienta (una Interfaz de Programación de Aplicaciones o API) compuesta por un conjunto de especificaciones de comunicación entre componentes software; que permite el indexado y la búsqueda de información en los documentos.

En el equipo cliente solo será necesario tener instalado un navegador. Las páginas de la aplicación se diseñaron para un óptimo funcionamiento con el navegador Microsoft Internet Explorer 5.0 o superior. La resolución mínima para la correcta visualización de las páginas es de 800x600, aunque la resolución óptima es de 1024x768.

Por supuesto, tanto en el equipo cliente como en el servidor necesitaremos tener instalado un sistema operativo. El sistema operativo deberá ser Windows en el servidor (NT, 2000 Server o XP), mientras que en el cliente se podrá usar tanto un entorno Windows, como UNIX/Linux.

▪ **Instalación**

No entraremos en detalles sobre como se realiza la instalación del programa completo, ya que ello está perfectamente especificado en el Manual de instalación del programa [3]. Sólo destacaremos que, una vez ya tenemos la aplicación instalada en el servidor y lista para funcionar, hemos de prever la posibilidad de incorporar nuevos documentos. Para ello deberemos tener en cuenta los siguientes aspectos:

- *Ubicación de los documentos*. Los archivos de la aplicación (estructurados, en formato XML; y en su formato gráfico, en

formato PDF) deben estar depositados en unos directorios concretos para que puedan ser localizados por la aplicación.

- *Creación de los índices:* Una vez hemos depositado los documentos en su emplazamiento correcto, aún tenemos que crear los índices que permitirán a la herramienta realizar las búsquedas en los documentos. El proceso de creación de índices está automatizado y es muy sencillo. Esta operación habrá que realizarla cada vez que insertemos nuevos documentos en el corpus, para hacer que la herramienta inserte los nuevos documentos en el árbol de índices que maneja.

Una vez hemos realizado dichos procesos, ya podremos comenzar a utilizar la herramienta de búsqueda en el corpus de enfermedades neuromusculares.

▪ **Otros aspectos a considerar: la creación de nuevas enfermedades**

La posible aparición o catalogación de nuevas enfermedades neuromusculares motivó la creación de un DTD flexible que permitiese adaptarse a esta particularidad. Para crear una nueva enfermedad habrá que darla de alta en los siguientes lugares:

- Será necesario crear una entrada para la nueva enfermedad en el DTD. En el atributo nombre del nodo enfermedad habrá añadir la nueva enfermedad:
- En la página de inicio habrá que crear también una nueva entrada en el menú de enfermedades para la nueva enfermedad:
- En la clase “Operaciones” habrá que indicar el nombre que se mostrará cuando se seleccione esa enfermedad:

3 Conclusiones

La calidad de la información médica existente en Internet es extraordinariamente variable; desde información científica y basada en la evidencia hasta remedios caseros o de muy dudoso origen, incluso de efectos nocivos o perjudiciales. La preocupación de gobiernos, sociedades e instituciones profesionales y científicas así como de los usuarios es creciente y requiere algún tipo de intervención. Las iniciativas propuestas están orientadas fundamentalmente hacia la acreditación, certificación, autorregulación, sistemas de calificación y concesión de sellos de calidad indicativos del cumplimiento de criterios concretos. Pero los usuarios, principalmente los pacientes, difícilmente conocen todos los sistemas de acreditación de los sitios web y más difícil es que entiendan el valor que pueden tener.

Esta valoración global es también aplicable al subdominio médico de las enfermedades neuromusculares; con el “añadido” que este tipo de enfermedades son crónicas, generalmente progresivas y de gravedad muy diversa, encuadradas dentro de lo que se conoce como “enfermedades raras”. Las variadas consecuencias que pueden tener estas enfermedades hacen que en este ámbito confluya un buen número de especialidades médicas (como la neurología, la miología, la cardiología, etc.). Se trata, además, de un ámbito en el que están presentes muchas otras áreas relacionadas con la salud en su sentido más amplio (como la fisioterapia, la ortopedia, la psicología, la nutrición, aspectos relacionados con la calidad de vida de los enfermos, etc.).

Por ello, en una primera fase de nuestro proyecto, nuestra propuesta se concreta en la realización de una plataforma, reuniendo la experiencia de todas las iniciativas que trabajan (directa o indirectamente) en dicho ámbito médico, e implicando en ella a los productores de la información y a sus usuarios finales. Con ello creemos se puede mejorar la información del ciudadano, paciente o profesional y evitar su confusión, además de mejorar el asesora-

miento sobre los contenidos médicos, evitando que información de baja calidad llegue a los usuarios.

Además de la creación del portal de ASEM (un sitio web especializado en el tratamiento de todo tipo de información relacionada con las enfermedades neuromusculares), la concreción de nuestro modelo teórico se basa, en primer lugar, en la creación de un corpus electrónico textual bilingüe (español- francés), a partir de un volumen reducido de textos de alta calidad relacionados con las enfermedades neuromusculares, creados por especialistas (en francés) y traducidos al español (por traductores profesionales especializados en la traducción médica).

En segundo lugar, procedimos a la elaboración de un motor de búsqueda documental para la extracción de información de dicho corpus. Aunque hemos visto que, desde un punto de vista práctico, parece factible la construcción de un motor de búsqueda semántica de enfermedades neuromusculares; sin embargo, debido al condicionante temporal de nuestro proyecto de investigación y a las necesidades de resultados inminentes por parte de ASEM-Galicia, decidimos empezar con la implementación de un motor de búsqueda por palabras clave. El buscador de enfermedades neuromusculares de ASEM (MYOCOR), en su versión de búsqueda sintáctica o por palabras clave, se encuentra ya implementado. En estos momentos está en fase de revisión y actualización de contenidos (textos). Puede accederse a él directamente, a través del enlace <<http://asemgalicia.dnip.net/ASEM>>; o bien, indirectamente, a través de la página web de ASEM-Galicia <<http://www.asemgalicia.com>> .

En una segunda fase del proyecto (pendiente todavía de aprobación), nuestra intención es construir las ontologías y metadatos necesarios e implementarlos en un motor de búsqueda semántico de enfermedades neuromusculares.

Para poder llevar a cabo dicho proceso, una vez implementado el motor de búsqueda por palabras clave, el siguiente paso sería evaluar la información y los servicios que el servidor web actual

de ASEM-Galicia ofrece en la actualidad a los profesionales de la salud, pacientes y familiares, así como a traductores y redactores de textos especializados y a desarrolladores de software. Por último, y no por ello de menor importancia, deberíamos ver como encaja la web semántica en la web actual; es decir, como accederá el usuario a la web semántica y, sobre todo, como hacer la transición de la web actual (sintáctica o lineal) a la web semántica.

Quizás deberíamos plantearnos el combinar una pequeña parte del trabajo manual que ello conlleva, con la automatización del resto del proceso. Lo deseable sería poder automatizar, como mínimo, el componente terminológico mediante la aplicación de un sistema gestor de bases de datos terminológico (basado en el conocimiento) ya existente.

Notas

1. Este trabajo se integra en el marco de los proyectos de investigación PGIDIGT04SIN065E (Xunta de Galicia), TIN2004-07246-C03-01 (Ministerio de Educación y Ciencia) y PGIDIGT05PXIC30501PN (Xunta de Galicia). Esta herramienta es uno de los resultados del proyecto de investigación titulado: "Creación y explotación de recursos documentales sobre enfermedades neuromusculares" (PGIDIT04SIN065E). El proyecto está financiado por la Consellería de Innovación e Industria (Xunta de Galicia) dentro del Plan Galego de Investigación, Desenvolvemento e Innovación Tecnolóxica. Forma parte de la Convocatoria de Proyectos de I+D y de Innovación Tecnológica de empresas, agrupaciones de empresas y asociaciones sin ánimo de lucro y la entidad solicitadora fue la Asociación Gallega contra las Enfermedades Neuromusculares (ASEM Galicia). Su duración es del 2004 al 2007. Como colaboradores externos han participado: la Universidade de Vigo (equipo de investigación dirigido por: Elena Sánchez Trigo, Catedrática de Traducción e Interpretación, del que forma parte Joan Miquel Vergés, junto con miembros del área de Informática) y el Hospital do Meixoeiro (equipo de investigación dirigido por: Carmen Navarro Fernández-Balbuena, Jefa de Servicio de Anatomía

Patológica y Neuropatología). Alexandre Dapena Mora, Ingeniero en Informática, ha sido el encargado del desarrollo de la aplicación informática.

2. Para un estudio exhaustivo de lo que es la *web semántica*, cuál es su “estado del arte” y cuáles son sus posibles aplicaciones, recomendamos al lector consulte la monografía que sobre dicho tema publicó la revista *Novática* en su número 178 (noviembre - diciembre 2005).

3. Existen muchas definiciones de qué es una ontología. Así, por ejemplo, según Studer et al.: “Una ontología es una especialización formal y explícita de una conceptualización compartida” [Studer, R.; Benjamins, V.R.; Fensel D. (1998). “Knowledge engineering: principles and methods”. En: *Data and Knowledge Engineering*, 25(1-2), 161-197]; donde “conceptualización” se refiere a un modelo abstracto de un cierto fenómeno en el mundo, identificando conceptos relevantes de ese fenómeno; “explícita” quiere decir que los conceptos y las propiedades usadas, y las leyes que gobiernan el modelo, se definen explícitamente; “formal” se refiere al hecho que la ontología debe ser procesable por un ordenador; y, finalmente, “compartida” refleja la noción de que la ontología captura conocimiento consensuado, es decir, no es la visión de un individuo sino algo aceptado por un grupo.

4. No nos extenderemos en mostrar las “excelencias” del software de código abierto o de las licencias de software libre. El lector puede consultar las diferentes monografías que al respecto le ha dedicado la revista *Novática* (por ejemplo, los números 181, 175 y 154).

5. UML (*Unified Modeling Language*) es un lenguaje gráfico para visualizar, especificar, construir y documentar los artefactos de un sistema con gran cantidad de software. Cubre tanto aspectos conceptuales (procesos de negocio, funciones del sistema) como cosas concretas (clases, esquemas de bases de datos, componentes reutilizables).

6. Éste es uno de los aspectos que resta todavía por depurar de la aplicación y que esperamos mejorarlo en próximas versiones.

Bibliografía

- [1] ASEM-Galicia (*Asociación Gallega contra las Enfermedades Neuromusculares*) <<http://www.asemgalicia.com>> [24-07-07].
- [2] Ávila, J.F.; Portillo, B.E.; Pajares J.M. (2001) "Calidad en la información biomédica existente en Internet". *Aten Primaria* 2001;28(10), 674-679.
- [3] Dapena, A. (2006) *Manual de Instalación del buscador de ASEM*. Manuscrito.
- [4] Dapena, A. (2006) *Manual Técnico del buscador de ASEM*. Manuscrito.
- [5] Faber, Pamela (2002). "Oncoterm: sistema bilingüe de información y recursos oncológicos. Amparo Alsina (ed.) *La Traducción Científico-Técnica y la Terminología en la Sociedad de la Información*. Castellón: Univ. Jaume I: 177-188.
- [6] García, S.; Montesinos, E.; Boyer, C. (2004) "Iniciativas en Español para la Evaluación de Contenidos Biomédicos en Internet". *Informédica 2004. Information & Communication Technologies in Healthcare Development. 3rd Virtual Congress in Internet* (1-30 marzo 2004) <http://www.informaticamedica.org/I04/papers/garcia_70.pdf> [24-07-07].
- [7] Grupo de Investigación OncoTerm. *OncoTerm: Sistema Bilingüe de Información y Recursos Oncológicos* <<http://www.ugr.es/~oncoterm/>> [24-07-07].
- [8] Mayfield, J.; Finin, T. (2003) "Information retrieval on the Semantic Web: Integrating inference and retrieval". Workshop on the Semantic Web at the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003). Toronto, Canadá, 2003 <<http://www.cs.umbc.edu/~finin/papers/sigir03.pdf>> [24-07-07].
- [9] Nancy, L.; Bennett, N.L.; Casebeer, L.L.; Kristofco, R.; Collins, B.C. (2005) "Family physicians' information seeking behaviors: A survey comparison with other specialties". En: *BMC Medical Informatics and Decision Making* 2005, 5:9

(22 marzo 2005) <<http://www.biomedcentral.com/content/pdf/1472-6947-5-9.pdf>> [24-07-07].

[10] OMS (2001) *CIF. Introduction* <<http://www3.who.int/icf/intros/ICF-Eng-Intro.pdf>> [24-07-07].

[11] Sánchez-Fernández, L.; Fernández-García, N. (2005) “La Web Semántica: fundamentos y breve «estado del arte»”. *Novática*, 178 (noviembre-diciembre), 6-11 <<http://www.ati.es/novatica/2005/178/178-6.pdf>> [24-07-07].

[12] Sánchez-Trigo, E. (2005) “Creación y explotación de recursos documentales sobre enfermedades neuromusculares”. *Actas de ponencias del XXII Congreso nacional de ASEM sobre enfermedades neuromusculares* (7 y 8 de octubre de 2005), 85-94 <<http://www.asemgalicia.com/biblioteca/XXII Congreso ASEM 2005 Libro de PONENCIAS-1.pdf>> [24-07-07].

[13] Sánchez-Trigo, E. (2006) “MYOCOR: creación y explotación de un corpus bilingüe (fr-es) sobre enfermedades neuromusculares”. *Confluências: 4* (mayo 2006), 67-83 <http://www.confluencias.net/n4/n4_sanchez-trigo.pdf> [24-07-07].

[14] Vallet, D.; Fernández-Sánchez, M.; Castells, P. (2005) “The Quest for Information Retrieval on The Semantic Web”. En: *Upgrade 6 (6), Monograph: The Semantic Web*. (diciembre 2005), 19-23 <<http://www.ati.es/novatica/UPGRADE/issues/2005/6/up6-6Vallet.pdf>> [24-07-07] (Publicado también, en español, como “Recuperación de información en la Web Semántica”. *Novática*, 178, noviembre – diciembre 2005, 12-15. <<http://www.ati.es/novatica/2005/178/178-12.pdf>> [24-07-07]).

[15] Wheeler, D.A. (2005) *Why Open Source Software / Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers!* <http://www.dwheeler.com/oss_fs_why.html>. Existe también una traducción del mismo al español: *¿Por qué usar programas abiertos? ¡Atención a las cifras!* <<http://www.hispalinux.es/informes/wheeler/index.html>> [24-07-06].