

Metodología para la estructuración del conocimiento de una disciplina: el caso de *PuertoTerm*¹

Por Jose A. Senso, Pedro-Javier Magaña-Redondo, Pamela Faber-Benitez y Amparo Vila-Miranda



Jose A. Senso es doctor en documentación y profesor del Departamento de Biblioteconomía y Documentación de la Universidad de Granada. Imparte su docencia en la Facultad de Traducción e Interpretación donde además es vicedecano de nuevas tecnologías. Es miembro del grupo de investigación Lexicon y de la Red Temática de Documentación.



Pedro Javier Magaña-Redondo es ingeniero informático por la Universidad de Granada, y diplomado en estudios avanzados en diseño, análisis y aplicaciones de sistemas inteligentes. En la actualidad centra su investigación en el estudio y desarrollo de tecnologías que faciliten la adopción de la web semántica por parte de un mayor número de usuarios.



Pamela Faber-Benitez enseña terminología y lingüística aplicada en la niv. de Granada (UGR), donde es catedrática de traducción e interpretación. Es licenciada en CC de la información por la Univ. de Carolina del Norte (EUA) y en filología inglesa por la UGR. Realizó su tesis doctoral en lingüística en la Univ. de Paris IV (La Sorbona) y la UGR. Ha dirigido tres proyectos de investigación en gestión de terminología multilingüe en medicina e ingeniería. Es autora de varios libros y artículos sobre semántica léxica, traducción y terminología.



María-Amparo Vila Miranda es catedrática de universidad desde 1992. Obtuvo el grado de doctor en matemáticas por la Universidad de Granada, donde actualmente dirige el grupo de investigación Idbis (Intelligent Databases and Information System). Es autora de numerosas publicaciones en libros y revistas científicas internacionales, y ha participado como ponente en un gran número de congresos. La mayoría de sus trabajos están relacionados con el diseño de bases de datos, la recuperación de información y datamining.

Resumen: En este trabajo se describen los pasos llevados a cabo dentro del proyecto de investigación *PuertoTerm* para formalizar y estructurar el conocimiento desarrollado dentro de la Ingeniería de Puertos y Costas. Se analiza el proceso seguido para la obtención de la terminología asociada, cómo se ha estructurado y cómo se ha gestionado. Por último, se describe el desarrollo e implementación, prestando especial atención al diseño de la interfaz bidimensional de visualización de la información.

Palabras clave: Estructuración del conocimiento, Visualización de la información, Ontologías, Ingeniería de puertos y costas, Modelo lexemático-funcional, Semántica de marcos.

Title: Methodology for knowledge structuring in a discipline: the *PuertoTerm* case

Abstract: The *PuertoTerm* research project reports the steps taken to formalize and structure the knowledge developed within the field of coastal and ports engineering. The process used to obtain the associated terminology is described and the way in which the terminology was structured and processed is analysed. Finally, development and implementation are described, with special attention to the design of a two-dimensional information visualization interface.

Keywords: Knowledge structure, Information visualization, Ontologies, Coastal and ports engineering, Lexematic-functional model, Semantic frameworks.

Senso, Jose A.; Magaña-Redondo, Pedro Javier; Faber-Benitez, Pamela; Vila-Miranda, Amparo. "Metodología para la estructuración del conocimiento de una disciplina: el caso de *PuertoTerm*". En: *El profesional de la información*, 2007, noviembre-diciembre, v. 16, n. 6, pp. 591-604.

DOI: 10.3145/epi.2007.nov.06

1. Introducción

El crecimiento de la información en determinadas áreas dentro de la web ha generado un aumento desigual en lo que se refiere a la aparición de herramientas terminológicas que faciliten la recuperación de información (taxonomías, tesauros y ontologías). Así, dependiendo del nivel de asentamiento que tenga una disciplina determinada, será más o menos complejo localizar productos que permitan estructurar el conocimiento asociado a dicha rama del saber.

Un ejemplo claro lo encontramos en la ingeniería de puertos y costas. Aunque es relativamente nueva, la notable evolución científica que ha experimentado debida al desarrollo de nuevas técnicas de construcción, explotación y gestión de estructuras y recursos marinos no ha sido suficiente como para que exista una formalización de las estructuras de conocimiento que representen los procedimientos, definiciones, objetos y metaconocimiento que en ella se llevan a cabo.

Este trabajo refleja el esfuerzo realizado por el proyecto de investigación *PuertoTerm* en la representación de la estructura conceptual del dominio de la ingeniería de puertos y costas, con la finalidad principal de apuntar los posibles pasos para realizar futuras representaciones de la estructura del conocimiento de una disciplina concreta y ayudar de esta manera a la generación de nuevas herramientas para una mejor recuperación de la información.

La mayoría de miembros de dicho proyecto han formado parte de otros dos anteriores que han servido como base teórica y especialmente metodológica para el que se explica a continuación. El objetivo del primer proyecto fue desarrollar una lógica léxica para la traducción asistida por ordenador a partir de una base de datos y el segundo, de gran impacto dentro de la comunidad médica, permitió la elaboración de un sistema de información integrado en internet y dedicado al subdominio de la oncología. Su nombre fue *Oncoterm*.

<http://www.ugr.es/~oncoterm>

1.1. Objetivos del proyecto *PuertoTerm*

Como ya se ha mencionado anteriormente, la ingeniería de puertos y costas tiene ciertas peculiaridades que la hacen “diferente” al resto de las ingenierías. El hecho de no contar con diccionarios específicos, la ausencia de teoría concreta y la necesidad de los profesionales que trabajan en ella de poseer ambas herramientas para realizar su trabajo, le confiere un cierto halo de área discriminada (al menos hablando desde el punto de vista de la documentación y la terminología). Con *PuertoTerm* se pretendió solucionar esto partiendo del hecho de que para transmitir conocimiento especializado es necesario emplear un sistema que permita

especificar el componente semántico inherente a cualquier área del saber.

“Para transmitir conocimiento especializado es necesario emplear un sistema que permita especificar el componente semántico inherente a cualquier área del saber”

El proyecto de investigación se basó en el programa del *Ministerio de Obras Públicas y Urbanismo* “Recomendaciones para obras marítimas”, que se inició en 1987 con la constitución de la Comisión Técnica encargada de redactar una serie de recomendaciones para guiar a organismos estatales y empresas privadas en el proyecto, construcción, mantenimiento y explotación de las construcciones marinas. Este programa también dio lugar a la publicación *ROM 0.0 (Puertos del Estado, 2001)*, que proporciona un conjunto de normas y criterios técnicos para la gestión de obras portuarias y marítimas, independientemente de su clase y materiales empleados. La traducción de esta norma al inglés por parte de la investigadora principal del proyecto puso de manifiesto que esta rama de la ingeniería emplea una serie de conceptos nuevos y neologismos con una designación original en español. Esto difiere notablemente de lo que sucede en otros campos de investigación, caso de la medicina por ejemplo, donde el término original procede principalmente del inglés. Por lo tanto, poseer un mayor número de recursos terminológicos en esta área facilitaría la adquisición y transmisión de conocimientos entre expertos de diferentes países.

Los objetivos específicos sobre los que sustentó el proyecto fueron:

- Crear un corpus de textos específicos de esa materia en español, inglés y alemán, así como desarrollar un protocolo de gestión de la información de dichos textos.

- Definir cuáles serían los conceptos y los términos que desarrollan la arquitectura semántica de esa disciplina y establecer las relaciones conceptuales específicas.

- Diseñar y alimentar una base de conocimiento terminológico articulado en torno a la estructura hallada en la definición de los términos. En ella se deberían poder almacenar además las relaciones semánticas existentes entre los términos, de forma que posteriormente se pueda producir una conversión o vinculación a formas de representación más expresivas.

- Crear una aplicación informática que permitiera la recuperación de la información sin que se perdiera la estructura y relaciones formadas entre los conceptos y los términos.

- Crear un banco de imágenes para complementar y enriquecer las representaciones lingüísticas de los conceptos pertenecientes al campo especializado que, por sus características específicas, necesita una representación conceptual más visual y dinámica.

Como se puede observar, la consecución de estos puntos debería plasmarse en la construcción de la estructura conceptual de ingeniería de puertos y costas. De sus términos y conceptos con sus respectivas definiciones, pero también de sus procesos, aplicaciones y herramientas. El presente texto explica los pasos llevados a cabo para proponer una metodología para futuros proyectos similares.

1.2. Fundamentos teóricos

Realizar la estructuración conceptual de un dominio determinado se consigue principalmente por medio de la elaboración de jerarquías terminográficas que se obtienen a partir de la selección y extracción de la información conceptual de textos especializados, diccionarios, tesauros, etc. En realidad, estos principios sirven de base metodológica para varias disciplinas relacionadas con la lingüística (terminología, lexicografía, traducción especializada) y con la documentación (diseño y gestión de tesauros y ontologías).

En el caso de la terminología, desde finales de 1980 se ha generalizado el uso de textos especializados como fuente de información y localización de vocablos técnicos para alimentar bases de datos terminológicas. En la actualidad, y gracias al avance de las tecnologías de la información, estos corpus suelen encontrarse en formato electrónico (McEnery; Wilson, 1996). La documentación, por su parte, emplea esta misma técnica para la realización de herramientas propias de los lenguajes documentales, como los tesauros y, gracias al desarrollo de conceptos como la web semántica, las ontologías. Una vez más queda patente la estrecha relación que existe entre ambas áreas y que ha sido defendida por muchos autores (Irazzábal, 1996; Pérez Álvarez-Ossorio, 1988; Pinto; Cordón, 1999; Sales, 2006). Si bien es cierto que la finalidad de ambas ha sido diferente (la terminología establece relaciones entre conceptos, y la documentación se centra en la recuperación de la información), en el proyecto *PuertoTerm* se ha intentado (y creemos que logrado) aunar ambas.

Para llevar a cabo la estructuración del conocimiento de la ingeniería de puertos y costas se han empleado dos teorías lingüísticas con base semántica: el modelo lexemático-funcional de Martín-Mingorance (1989; 1995; Faber; Mairal, 1999) y la semántica de marcos *-frame semantics-* (Fillmore, 1982; Fillmore; Atkins, 1998; Gahl, 1998).

El modelo lexemático-funcional facilita la representación de relaciones conceptuales y su posterior

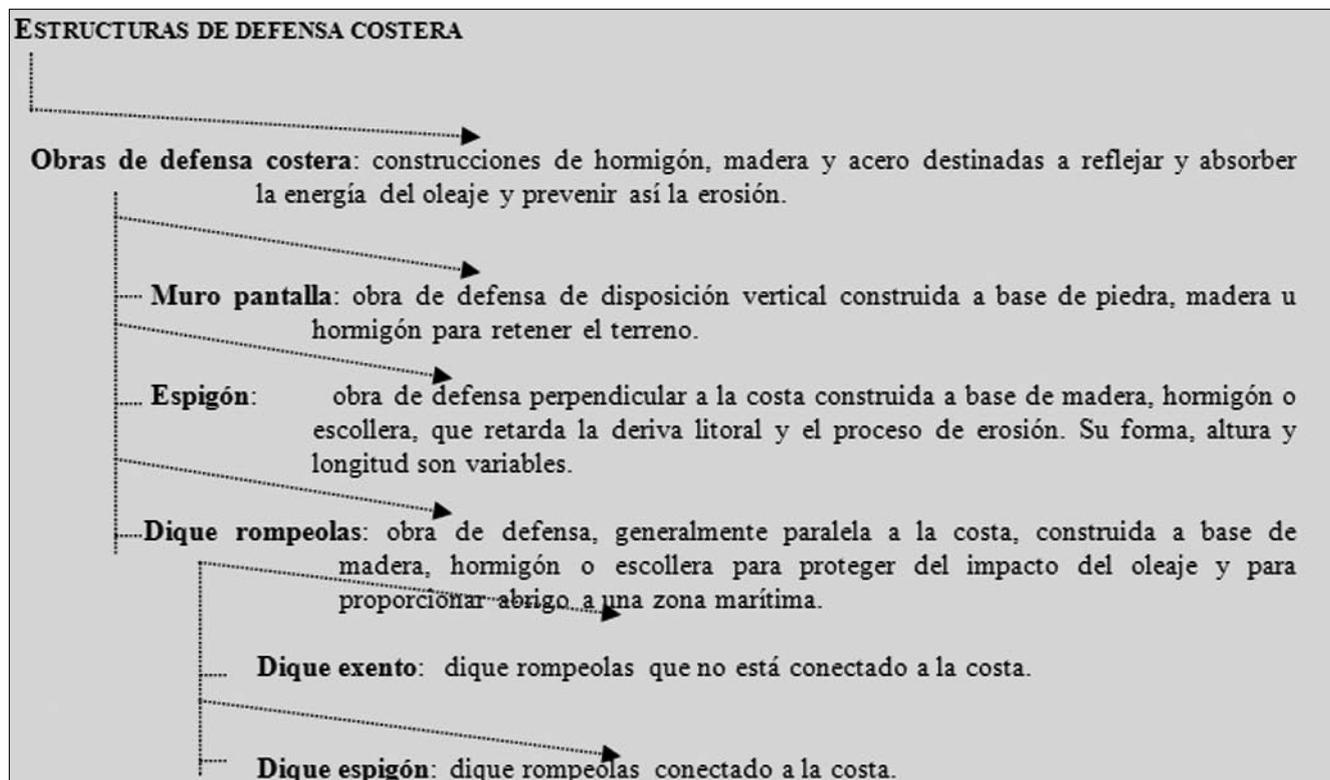


Figura 1. "Estructuras de defensa costera" tiene como subdominio "Obras de defensa costera", que muestra el siguiente diseño, explícito en las definiciones de los términos

clasificación dentro de un lenguaje general y/o especializado. Básicamente propone una distinción entre relaciones sintagmáticas y paradigmáticas –basada en los principios complementarios de combinación y selección (Lyons, 1977)- que permite una organización conceptual independiente del sistema lingüístico. Ofrece una metodología lingüística para la organización de conceptos. De hecho, concibe nuestro propio lexicón mental como una compleja red en la que cada nodo es un concepto y estos están interconectados por diferentes tipos de relaciones (López-Rodríguez; Tercedor-Sánchez; Faber-Benitez, 2006). Por ejemplo, en la figura 1 se puede observar la estructuración de “Obras de defensa costera” como subdominio de “Estructuras de defensa costera”.

La semántica de marcos (Fillmore; Johnson; Petruck, 2003), por su parte, entiende el concepto de marco como una representación esquemática de un conjunto de conceptos relacionados entre sí. El proyecto *FrameNet*, de la *Universidad de Berkeley* se basa en dicha teoría. Su gran ventaja radica en que la utilización de un único concepto activa todo el sistema conceptual permitiendo, gracias al empleo de enlaces web en su fase de implementación “física”, conectar todos los conceptos subordinados y superordinados, facilitando la representación genérico-específica de los términos. Un desarrollo más profundo de esta fase del proyecto se puede encontrar en Faber (2006).

<http://framenet.icsi.berkeley.edu/>

2. Metodología

Para realizar una verdadera estructuración del conocimiento dentro de una disciplina no basta con obtener unas listas de términos y delimitar sus posibles relaciones jerárquicas, ya que son demasiado estáticas y no ofrecen la verdadera visión dinámica que tiene cualquier área del saber. La mayoría de proyectos de este tipo que se basan en lenguajes documentales intentan recopilar términos o plasmar el conocimiento realizando traducciones de los utilizados en los vocabularios de la disciplina que se pretende representar. Esto elimina cierto nivel de comprensión a la hora de entender tareas y procesos, es decir, plasmar algo más que una simple jerarquía. Para evitarlo, el enfoque global que se pretende dar es el que denominamos gestión de terminología orientada a procesos (Faber et al., 2005).

La primera fase se inicia empleando técnicas de recuperación de información desarrolladas en el ámbito de la lingüística de corpus, disciplina que ha mostrado su utilidad en proyectos a gran escala como las dos fases de *Acquilex* o, a menor escala dentro del mundo de los lenguajes documentales, los que pretenden generar clasificaciones, tesauros u ontologías (ISO-5964, 1985).

<http://www.cl.cam.ac.uk/research/nl/acquilex/>

Estas técnicas son muy útiles, ya que desde el punto de vista práctico suponen una forma fácil y asequible de obtener gran cantidad de terminología asociada a un área. Desde el punto de vista de uso real de los vocablos, el investigador se asegura que la información extraída es fiel reflejo tanto de los contenidos reales, en nuestro caso de la ingeniería de costas, como del sublenguaje especializado empleado en el mismo. En el caso particular de este proyecto, después de estudiar los recursos terminológicos, se han creado otros nuevos que no existían antes en esta disciplina al ser un dominio muy joven y dinámico y con una terminología en proceso de fijación y estandarización. En otros casos con campos más formados este proceso no tiene por qué ser necesario. Como es habitual, las fuentes de información empleadas para este análisis de corpus han sido obras de referencia especializadas (enciclopedias y diccionarios técnicos mayoritariamente), monografías y artículos de publicaciones científicas. Todas ellas se almacenaron en ficheros de texto en diferentes directorios, dependiendo del idioma en el que estuviese expresado el texto, para facilitar su procesamiento por el resto de herramientas.

“Después de estudiar los recursos terminológicos existentes sobre esta materia, se han creado otros nuevos”

En esta fase, los términos recogidos forman una lista plana, donde ninguno tiene más valor semántico que otro. El siguiente paso se centró precisamente en elaborar un compendio de los conceptos más importantes. Para ello, y con la ayuda de especialistas y de la herramienta de análisis léxico *Wordsmith Tools*, distribuida por *Oxford University Press* y que permite explorar grandes conjuntos de textos mediante búsquedas basadas en parámetros contextuales o estadísticos, se realizó una lista de frecuencias que permitió inferir el conocimiento especializado de la ingeniería de puertos y costas. La identificación de las palabras clave permitió el modelado conceptual de la ontología que articularía todo el conocimiento de esta disciplina. Los lemas más frecuentes permiten identificar las categorías conceptuales sobre las que se fundamenta la definición de los términos del texto. La figura 2 muestra el resultado del análisis del corpus en español realizado por *Wordsmith Tools* a partir de los documentos extraídos tras la primera fase del proyecto.

<http://www.lexically.net/wordsmith/>

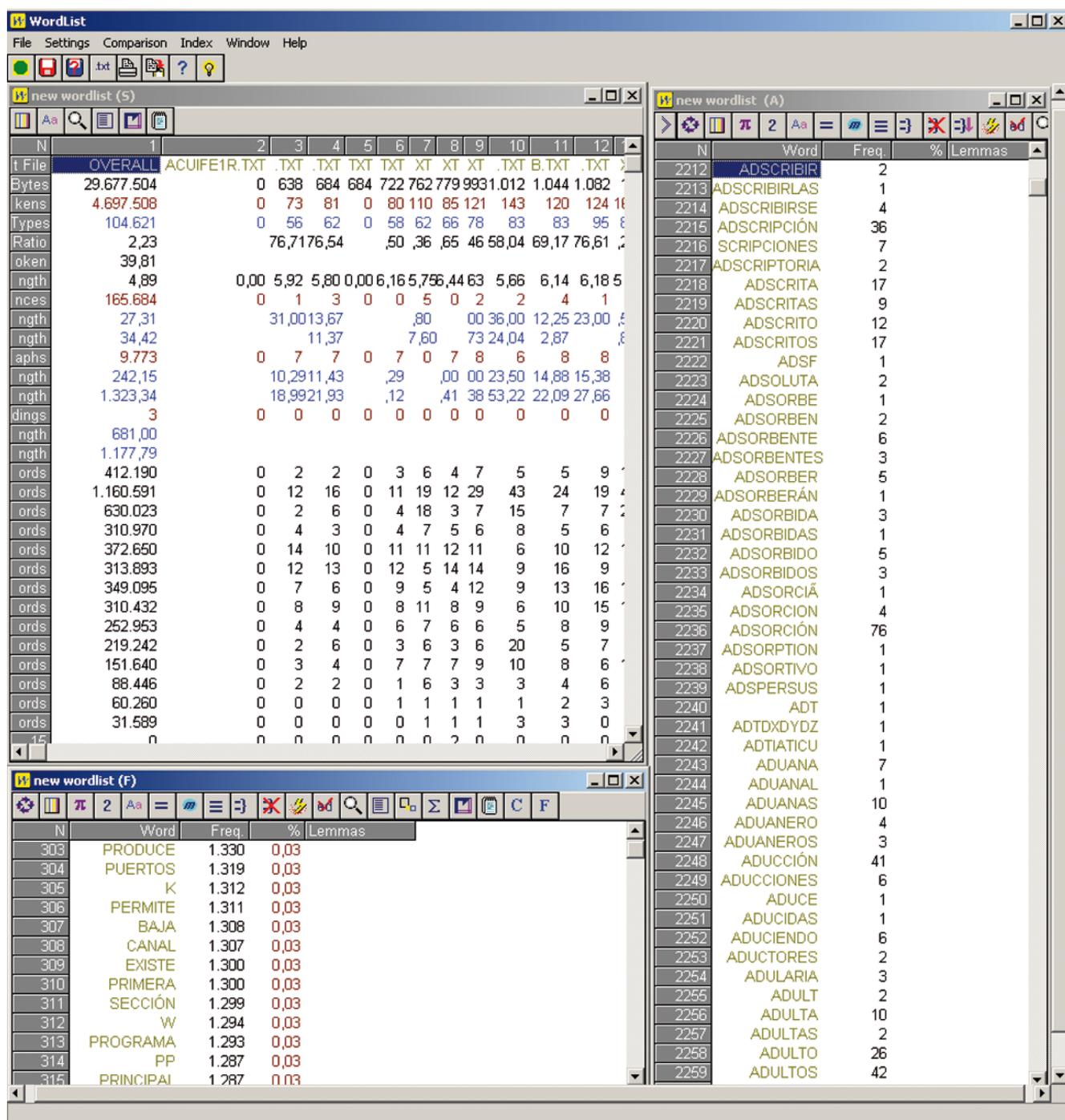


Figura 2. Resultado mostrado por Wordsmith Tools del análisis de los ficheros de texto localizados tras el análisis del corpus. Se puede observar la estadística general (izquierda arriba), la lista de frecuencia de los términos (izquierda abajo) y la lista alfabética de términos (derecha)

Con el fin de conseguir una mayor adecuación en el empleo de cada término, y poder enmarcarlo dentro de su uso correcto, no sólo se utilizaron técnicas estadísticas. Otro elemento clave fue el uso de concordancias, es decir, la presentación de las ocurrencias de una palabra en su contexto lingüístico, de tal forma que la palabra en cuestión aparece en el centro de una línea y a ambos lados se muestran otras con las que suele aparecer en los textos. Se trata de una técnica muy similar a la empleada en los índices *KWIC* (*keyword in context*) pero con una finalidad diferente, ya que aquí se emplea menos para recuperar y más para contextualizar.

Hasta ahora los pasos seguidos nos han permitido adquirir toda la información, en forma de palabras que pueden expresar conceptos o términos, relacionadas con la ingeniería de puertos y costas que nos interesa, y además tenemos cierta consciencia sobre el valor que tienen esas palabras en función de su peso conceptual.

Cada disciplina está caracterizada por poseer uno o varios eventos (o marcos conceptuales) que describen procesos o acciones que tienen lugar dentro de ella. Dichos eventos poseen una estructura que facilita la creación y organización de los diferentes conceptos

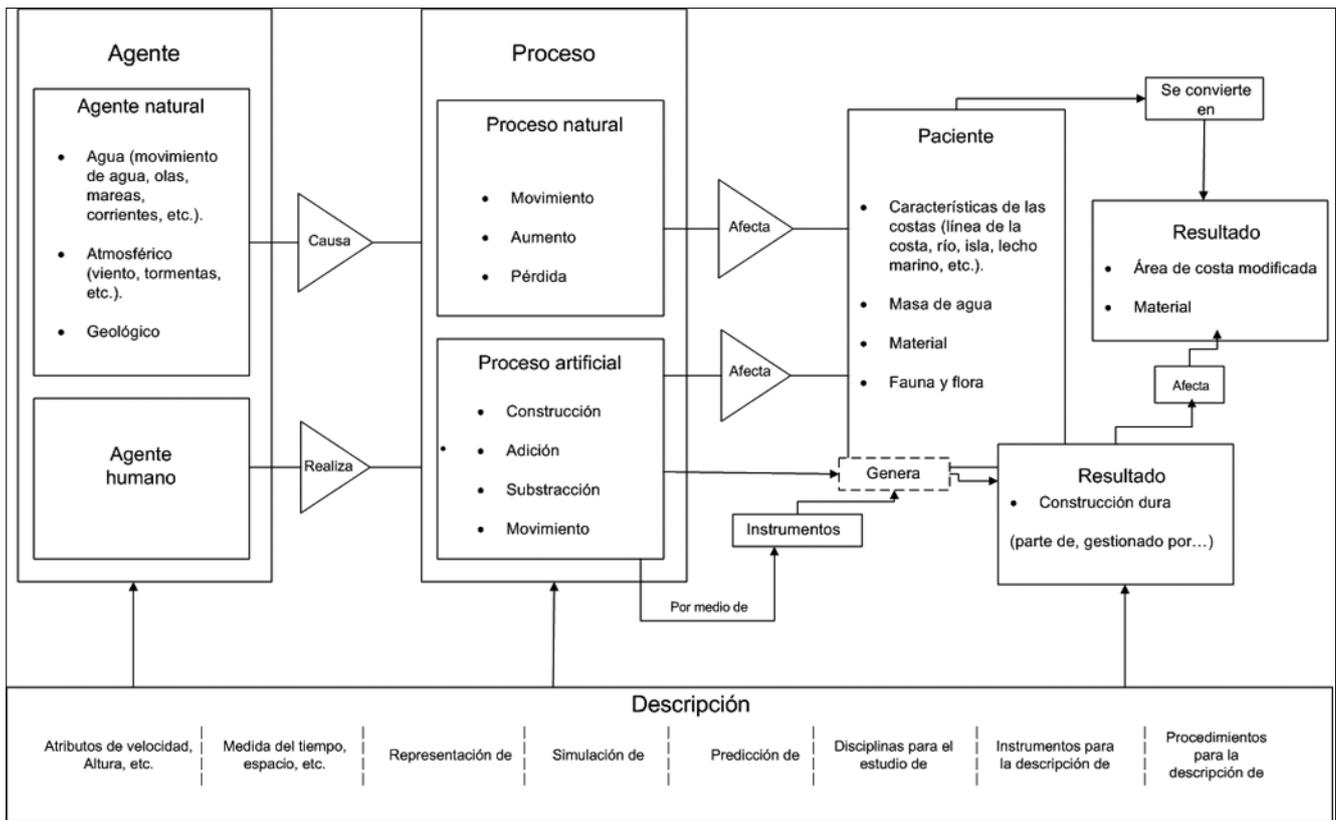


Figura 3. Descripción de los procesos llevados a cabo en la ingeniería de puertos y costas. El análisis del corpus y su posterior categorización permiten conocer cuáles son los eventos que se producen dentro de una disciplina

que los integran. La elaboración del marco conceptual es el culmen del análisis de las definiciones lexicográficas y de corpus extraídas previamente. En nuestro caso, dicho análisis se plasma en la figura 3, que muestra el marco genérico de la ingeniería de puertos y costas.

A partir de ella podemos deducir que todos los procesos llevados a cabo dentro de la ingeniería de puertos y costas son iniciados por un agente que, por medio de una serie de acciones (procesos naturales o artificiales), afecta a un paciente (la costa, la flora y fauna, etc.) y produce un resultado determinado. Estas macrocategorías (agente, proceso y paciente/resultado) constituyen los roles principales del dominio y sirven para proporcionar un modelo que permita representar las relaciones. Como suele ocurrir en muchas áreas, los agentes se pueden valer de instrumentos para generar determinados resultados. De esa forma contemplamos la posibilidad de añadir las herramientas que se suelen emplear en esta materia para, por ejemplo, realizar mediciones (anemómetro, anemoclimómetro, anemógrafo, etc.).

A partir de esta base se puede elaborar una jerarquía en donde cada concepto se relacionará de manera semántica con otros y así de forma continuada. De esa manera cada categoría conceptual se encuentra representada por un conjunto de tipos de información que tiene en cuenta las características de un concepto y las

entidades de esa categoría en el mundo real. Con estas categorías conceptuales (y sus respectivos esquemas) se modeló una base de conocimiento, estableciendo relaciones entre conceptos de “tipo de” (para relaciones de especialización), “parte de” (para las estructurales), “resultado de” (relaciones causa-efecto) o “atributo de” (asociativas). Así por ejemplo, el concepto “abanico fluvial” tiene una relación “tipo de”, que es “abanico”, una relación “atributo de” con “aluvial” y es “resultado de” del concepto “corriente”.

Toda esta información semántica es almacenada en la propia base de conocimiento, a pesar de que la implementación actual de la misma haga un uso limitado de ella. No obstante, resulta especialmente relevante este hecho, principalmente por dos razones. En primer lugar debido a que es frecuente en la actualidad que numerosos proyectos almacenen su información en sistemas heredados (*legacy systems*) en los cuales cualquier cambio en su diseño podría resultar dramático. La posibilidad de completar y añadir información semántica en dichos sistemas permitiría acoplarlos (*mapping*) con formas de representación que sí hagan uso de dicha información, sin necesidad de realizar ninguna modificación en los sistemas actuales, y accediendo a su información de forma semántica. Por otra parte, también es reseñable el hecho de que el diseño de sis-

temas basados en formas de representación más ricas semánticamente, como las ontologías, es más complejo y por tanto requiere más tiempo para su elaboración.

“El diseño de sistemas basados en formas de representación más ricas semánticamente, como las ontologías, es más complejo y requiere más tiempo para su elaboración”

En la actualidad nos encontramos trabajando en la creación de una ontología que permita el acceso a la base de conocimiento utilizando la información semántica almacenada. No obstante, lo más interesante de esta propuesta radica en el hecho de que la información no se encuentre replicada en ambas representaciones. Para ello se están utilizando herramientas que permiten vincular ontologías a otras formas de representación, como *D2RQ* de la *Freie Universität Berlin*, consiguiendo una ontología que carga los datos necesarios en el momento de ser solicitados (por ejemplo por parte de un razonador), pero que no los almacena.

<http://sites.wiwiw.fu-berlin.de/suhl/bizer/D2RQ/>

Por último, y una vez organizados todos los conceptos vinculándolos a varios idiomas, se elaboraron las entradas terminológicas de manera consensuada y supervisada por especialistas. La figura 4 muestra un ejemplo de entrada terminológica.

Como puede observarse en la entrada terminológica de ejemplo, ésta cuenta con un enlace a un banco de imágenes. Desde el comienzo del proyecto se pensó que, con fines didácticos y para lograr un mayor alcance en la explicación de los términos, sería interesante añadir contenido multimedia a la información asociada a cada término. Junto a la base de datos con la información referida al concepto (definición, equivalente en otros idiomas, relaciones con otros conceptos, categoría gramatical, contextos representativos, concordancias, etc.) se reservó un campo para albergar información multimedia. Dichos ficheros formarían un fondo de imágenes, vídeos, etc., que permitirían la creación de un tesoro visual de conceptos costeros. Además, esto posibilitaría crear una tipología de información gráfica que relacione ilustración y texto científico-técnico según criterios lingüístico-cognitivos.

Para poder gestionar toda esta gran cantidad de información y emplear un sistema de consulta eficaz fue necesario el uso de las diversas herramientas informáticas que se explican a continuación.

2.1. Desarrollo e implementación

Desde el punto de vista informático y de implementación del sistema de información, la experiencia previa fue fundamental a la hora de consolidar la arquitectura actual de *PuertoTerm*. Inicialmente el proyecto surgió como una memoria sobre términos especializados relacionados con el río Guadalfeo, información que era almacenada en una única base de datos centralizada. Entre sus funcionalidades destacaba un módulo constituido por varios formularios para su alimentación, y otro para la generación de informes que servía para la representación de la información contenida de un modo adecuado.

Sin embargo, a medida que la ambición del proyecto crecía, esta estructuración iba quedando obsoleta. El modelo centralizado obligaba a cada usuario a descargar la totalidad de la base de datos para realizar las modificaciones deseadas de forma local y posteriormente sobrescribir la versión disponible en el servidor.

Esto daba lugar a problemas de concurrencia. El sistema no facilitaba que varios usuarios pudiesen cooperar para realizar modificaciones de forma conjunta. Ni siquiera se garantizaba que cambios realizados de manera simultánea por diferentes usuarios se viesen reflejados finalmente en la base de datos. Eran los propios usuarios los encargados de comprobar que ningún otro usuario estuviera realizando modificaciones en un momento dado.

Otro aspecto fundamental que empezaba a verse afectado era el que concernía a la seguridad. Un usuario malintencionado que accediera a la base de datos podía eliminar o realizar cualquier modificación sin control alguno. Es más, era necesario realizar copias de seguridad de forma periódica para poder recuperar el sistema. La solución dada a este problema se basó en una configuración extraordinariamente restrictiva por parte del cortafuegos instalado en el servidor, que limitaba las posibilidades de acceso a los usuarios. Por ejemplo, sólo se podía acceder a la base de datos desde los ordenadores ubicados en los despachos de cada uno de los miembros del grupo, impidiéndoles trabajar en cualquier otro lugar.

Otro de los grandes problemas generados mediante esta forma de trabajo era que el único registro que se podía obtener del uso del servidor venía dado por el cortafuegos. Éste, sin embargo, no proporciona datos acerca del uso concreto realizado en el servidor, así que no se podían obtener estadísticas de uso u otro tipo de información que mejorase el aprovechamiento del servidor.

Por tanto, era necesario pasar de una arquitectura en donde el sistema de gestión de base de datos estaba orientado a un entorno personal, a un nuevo escenario

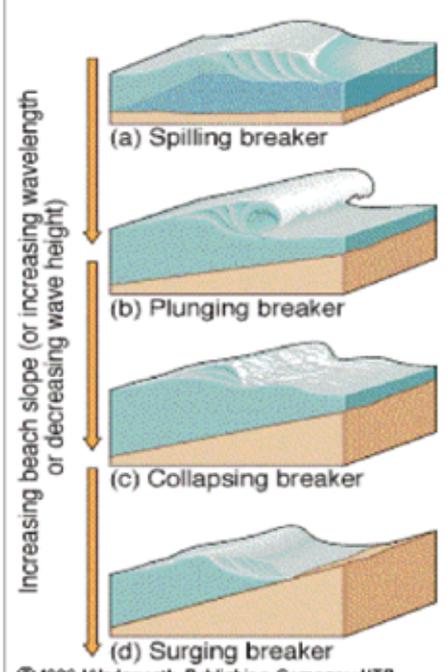
<p>Rompiente de derrame</p>	<p>Ola que presenta un movimiento progresivo de atenuación y poca fuerza de ascenso en una onda que rompe dando lugar a la aparición de espuma, extendiéndose por una amplia zona de barrido o surf.</p>	<p>-Ola -Rompiente</p>
<p>Contexto</p>	<p>Entre las posibles razones de recuperación de la misma, puede postularse como factor importante la baja altura del oleaje cotidiano que caracteriza el tipo de rompiente de derrame ("spilling breaker") [vínculo a archivos multimedia], el cual tiene la característica de expandirse por una amplia zona de barrido o "surf", predominando la acreción hacia tierra por la caída libre de las partículas debido a que la ola tiene poca fuerza de ascenso, lo cual aunado al relativo abrigo topográfico podría explicar la recuperación de la playa. Fuente: http://acta.ivic.ve/52-3/articulo5.pdf</p>	<p>-Rompiente en voluta -Rompiente de colapso -Rompiente ondulada</p> <p>(Palabras en cursiva en azul indica hipervínculos)</p>
<p>Spilling breaker (inglés)</p>	<p>Breaker in which the bubbles and turbulent water spill down front face of a wave. The upper 25 percent of the front face may become vertical before breaking. Breaking generally occurs over quite a distance.</p>	<p>-Wave -Breaker</p>
<p>Contexto</p>	<p><i>Breaker type</i> refers to the form of the wave at breaking. Wave breaking may be classified in four types: as spilling, plunging, collapsing, and surging In <i>spilling breakers</i> [vínculo a archivos multimedia], the wave crest becomes unstable and cascades down the shoreward face of the wave producing a foamy water surface. In <i>plunging breakers</i> [vínculo a archivos multimedia], the crest curls over the shoreward face of the wave and falls into the base of the wave, resulting in a high splash. In <i>collapsing breakers</i> [vínculo a archivos multimedia] the crest remains unbroken while the lower part of the shoreward face steepens and then falls, producing an irregular turbulent water surface. In <i>surging breakers</i> [vínculo a archivos multimedia], the crest remains unbroken and the front face of the wave advances up the beach with minor breaking. Fuente: http://www.vulcanhammer.net/marine/Part-II-Chap_4entire.pdf</p>	<p>-Plunging breaker -Collapsing Breaker -Surging breaker</p> <p>(Palabras en cursiva en azul indica hipervínculos)</p>
<p>Schwallbrecher (alemán)</p>	<p>Der Schwallbrecher tritt bei relativ großen Wellenstärken an flachen Stränden auf. Unter Beibehaltung der annähernd symmetrischen Form läuft das Wasser an der Oberfläche mit zunehmender Steilheit über den Wellenberg. Eine Brecherzunge löst sich aus dem Wellenberg, gleitet als Schwall den Leehang herunter und erzeugt zunehmend Wirbel und Turbulenzen.</p>	<p>-Welle -Brandungswelle</p>
<p>Contexto</p>	<p>Dabei wird die Wellenenergie über eine Strecke von mehreren Wellenlängen dissipiert.</p> <p>Beim Sturzbrecher, der für steilere Strände charakteristisch ist, wird der Wellenberg >übersteilt< und kippt vornüber in das Wellental, wobei auch verfestigte Sedimente erodiert und grobe Kiese und Wackeln bewegt werden. An flacheren Stränden überwiegen Schwallbrecher, bei denen der Wellenberg schäumend in das Wellental bricht. Letzlich wird in beiden Fällen Wellenenergie >vernichtet< (Dissipation), indem sie auf Sedimentpartikel übertragen wird, im Falle des Sturzbrechers allerdings nur auf kurzer Strecke mit hoher morphogenetischer Wirksamkeit, beim Schwallbrecher auf einem breiten Streifen mit geringerer Wirksamkeit. Fuente: http://www.uni-konstanz.de/limnologie/wetlands/PDF-Dokumente/Ostendorp-2004-SVGB-Bd122.pdf</p>	<p>-Sturzbrecher -Schwallbrecher -Reflexionsbrecher</p> <p>(Palabras en cursiva en azul indica hipervínculos)</p>
<p>Banco de Imágenes</p>	<p>archivo de audio spilling breaker archivo de video spilling breaker archivo de imagen SPILLING BREAKER:</p>  <p>© 1998 Wadsworth Publishing Company/ITP</p>	

Figura 4. Entrada terminológica del término rompiente de derrame

con varios usuarios actuando a la vez. En el caso de *PuertoTerm*, este punto resulta clave debido a la distribución espacial de sus integrantes. Para este cambio se aprovecharon las facilidades que ofrecen internet y sus numerosos marcos de trabajo disponibles.

Así, de un modelo centralizado y monolítico que hacía un uso ineficiente del servidor, se pasó a una arquitectura cliente-servidor, mucho más adecuada para un ambiente web. Un usuario ya no tendría que descargar la totalidad de dicha base de datos a su ordenador, aún cuando fuese a utilizar únicamente una parte, y provocar un escaso aprovechamiento del ancho de banda.

2.1.1. Portal de trabajo colaborativo

La arquitectura (figura 5) que se creó *ad hoc* está formada por diversos módulos:

a. Entrada de información: constituido por una serie de formularios web que se emplean para introducir la información relativa a los conceptos, términos, recursos y sus interrelaciones. Contiene varias tecnologías que hacen su utilización más amigable y eficaz, tales como atajos de teclado o mecanismos de búsqueda mientras se escribe, independiente del navegador web empleado.

b. Validación: revisa y corrige la información almacenada en la base de conocimiento. Incluye una funcionalidad de búsqueda que permite acceder de forma directa a cualquier elemento y que se puede realizar en base a cadenas completas o subcadenas, lo que amplía sus posibilidades. Debido al gran volumen de información introducido, no se requiere que genere una consulta sensible al caso. Por último, éste módulo facilita diferentes opciones para la presentación de la información. Así por ejemplo, se puede elegir si se valida una serie de registros mediante una navegación secuencial por orden de introducción, y otra ordenada mediante criterios alfabéticos.

c. Gestión de las comunicaciones: permite asociar discusiones a un concepto o término concreto. Otra posibilidad es la de generar informes de forma totalmente personalizada, para distribuir contenido selectivo a personas ajenas al sistema. De esta forma se facilita la interacción con usuarios a las que no se les desea dar acceso a la totalidad de la información. Finalmente, también se puede distribuir cualquier tipo de información mediante un simple gestor de contenidos integrado a cualquier visitante de la plataforma, que puede interactuar proporcionando comentarios acerca de la misma.

d. Usuarios: hace posible gestionar los accesos al sistema y la base de conocimiento. De esta manera se almacena cierta información asociada al usuario cada vez que accede y con el objeto de localizar posibles

problemas. Cada usuario tiene restringida la información que puede manipular según su perfil.

e. Seguridad: proporciona mecanismos de recuperación en caso de aparición de problemas. Un agente actúa realizando copias de seguridad tanto interna como externamente en el servidor. Debido al volumen de información, tiene en cuenta la actividad de los usuarios para, por ejemplo no actuar si no han realizado ningún cambio. Al mismo tiempo, dos sistemas independientes se encargan de monitorizar de forma continua el correcto estado del servidor que contiene todo el sistema.

2.1.2. Base de datos *versus* ontología

El cambio de naturaleza de la información residente en *PuertoTerm*, pasando de meramente organizar la estructura e integridad de sus datos, a tratar de especificar el significado del dominio de la ingeniería de costas mediante conceptos y sus interrelaciones, también condicionó la elección del nuevo modelo. Esta nueva forma de representación se acerca mucho más a las bases de conocimiento y ontologías, que a las bases de datos. Las enormes posibilidades que abre la representación de información en este tipo de modelos, como la bien conocida web semántica, hacía esta elección muy atractiva.

Sin embargo, la mayoría de las herramientas disponibles para el desarrollo de ontologías están pensadas aún de forma monolítica. La idea subyacente es realizar pequeñas ontologías que puedan integrarse y consensuarse entre individuos. No obstante, la naturaleza de una ontología de las características de *PuertoTerm*, que necesita ser alimentada por muchos usuarios, no se ve reflejada en ese estereotipo. En la actualidad, diversos proyectos pretenden cubrir este hueco, como por ejemplo *Collaborative Protégé*, un añadido (*plugin*) para el conocido editor de ontologías *Protégé* que permite la edición colaborativa de las mismas. Sin embargo, como sus propios desarrolladores apuntan, en la actualidad se trata de un prototipo bajo desarrollo.

<http://protege.stanford.edu/>

<http://smi-protege.stanford.edu/collab-protege/collab-protege.html>

Otro de los puntos en contra para la elección de este tipo de herramientas fue la escasa usabilidad que ofrecen. Están orientadas hacia usuarios experimentados, y su proceso de aprendizaje consume gran cantidad de tiempo, que no podía ser asumida por parte del proyecto.

Por tanto, la única solución viable pasaba por implementar una interfaz propia que accediese de forma remota a la información residente en la ontología. Siendo una opción que, aunque cumpliendo todos los requerimientos mencionados, también consumiría una gran cantidad de tiempo.



Figura 5. Aspecto de la interfaz creada para la introducción de datos y comunicación entre los miembros del grupo

Por todas estas razones, se decidió mantener un sistema de gestión de base de datos para almacenar la base de conocimiento. La amplia utilización de las bases de datos ha permitido una gran optimización en su rendimiento, algo que aún no ha sucedido con las herramientas ontológicas, que se encuentran en un grado incipiente.

La información semántica que no puede ser empleada por la utilización de un modelo más restrictivo en lo que respecta al nivel de representación que es capaz de expresar (es decir, la base de datos), es también almacenada con el objeto de utilizarla para establecer una correspondencia posterior con modelos más ricos en su semántica (en nuestro caso, una ontología). De esta manera, se alimenta un sistema por medio de un mecanismo ampliamente probado, sin que eso menoscabe su capacidad de representación ni implique duplicación de la información. La figura 6 muestra las relaciones existentes dentro de la base de datos.

2.1.3. Recuperación y visualización de la información

Desde el principio se pensó que el acceso a la información almacenada en la base de datos se debería realizar por medio de consultas (*querying*) y búsqueda analítica (*browsing*). Teniendo en cuenta que toda la información se encuentra centralizada en varias tablas almacenadas en *MySQL*, el acceso a la misma no debe-

ría ser excesivamente complejo y, además, posibilitaría fáciles migraciones con vistas a integrar este trabajo en otros proyectos de parecidas características. Además, se planteó la posibilidad de mostrar la información de una manera más visual y menos lineal. Por ese motivo, se optó por la utilización de tecnologías de visualización de la información en 2 y/o 3 dimensiones que faciliten la rápida comprensión de las estructuras arbóreas inherentes dentro de cualquier sistema clasificatorio.

“La utilización de tecnologías visuales para mostrar la información recuperada y las relaciones existentes resulta de especial trascendencia en *PuertoTerm*”

La utilización de tecnologías visuales para mostrar la información recuperada y las relaciones existentes resulta de especial trascendencia en *PuertoTerm* ya que posibilita una mejor y más profunda comprensión del contenido almacenado, proporcionando mecanismos para acceder a información relevante que podría no ser accesible mediante otros sistemas de representación. Además de permitir inferir conocimiento de manera más sencilla que en otros sistemas tradicionales.

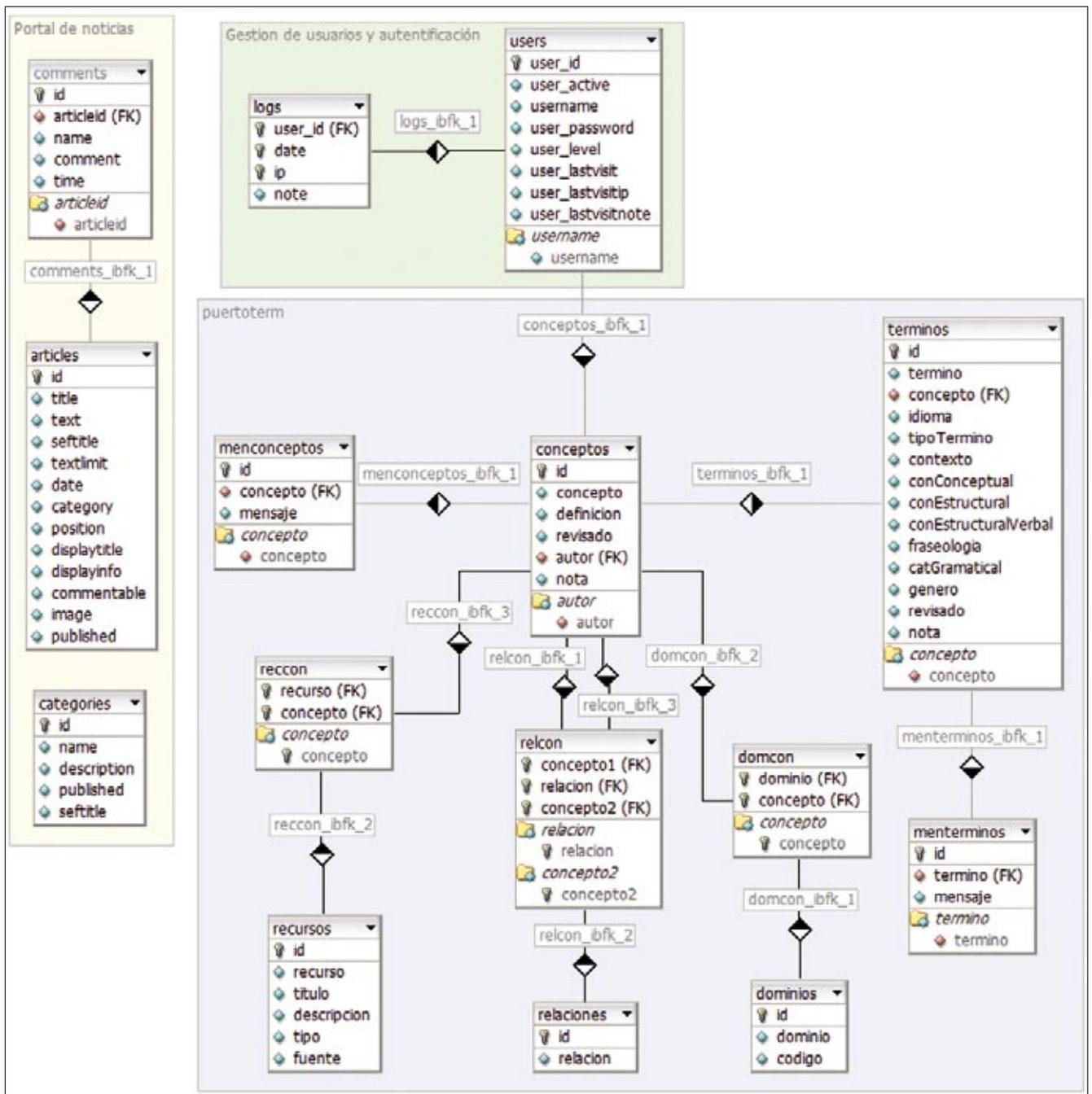


Figura 6. Relaciones existentes dentro de la arquitectura de PuertoTerm

No obstante, la presentación de información a través de la web cuenta con numerosos problemas, debido a las limitaciones del lenguaje en el que está implementada. Html limita el acceso a la información a modelos bidimensionales con escasa información estructurada. De ahí surge la necesidad de introducir algún complemento para que los navegadores extiendan su funcionalidad visual.

De entre todas las opciones disponibles la más difundida y con más posibilidades son los *applets*, ya que están implementados en un lenguaje muy potente como es *Java*. Las opciones de un lenguaje de programación tan versátil son prácticamente ilimitadas, presentando características para interactuar con otros

sistemas, la optimización de cualquiera de sus componentes, o el desarrollo de cualquier funcionalidad a medida.

Además, su amplia portabilidad es una característica fundamental para el desarrollo de cualquier sistema que quiera ser usable. El amplio rango de usuarios a los que *PuertoTerm* pretende abarcar hace de ésta una de sus principales características. Los usuarios están divididos fundamentalmente en tres grupos. El primero corresponde a los lingüistas que alimentan y desarrollan la base de conocimiento. El grupo de expertos en ingeniería de costas conforman el segundo de ellos. Finalmente, los usuarios finales del sistema podrían abarcar desde la administración pública hasta particulares que

deseen ampliar sus conocimientos mediante la lectura de textos en lenguas foráneas.

Las técnicas de visualización existentes para bases de datos no tienen en cuenta la información semántica considerada en esta aplicación, por lo que no son válidas para nuestro propósito. Por otra parte, las técnicas empleadas dentro de ese campo para las ontologías son más ricas en expresividad, pero aún son muy limitadas, y generalmente están integradas en el propio editor donde se desarrolla la ontología (*Jambalaya*, *Knowledge Tree*, *Ontoviz*, *TGViz*, etc.).

Por tanto, era necesario utilizar alguna técnica propia para la visualización del conocimiento. Una primera consideración podría llevar a pensar en el uso de *Mindmaps*, *Concept Maps*, y otra serie de representaciones de la información. Sin embargo, cada una de ellas está pensada para un tipo de aplicaciones concretas. Así, los *Mindmaps* están fundamentalmente orientados a caracterizar el conocimiento que un individuo tiene acerca de un tema principal, estructurando dicho contenido en forma de árbol y jerarquías radiales, no teniendo las conexiones entre sus elementos ningún tipo de restricción asociada. Los *Concept Maps* están enfocados a visualizar las relaciones entre conceptos (Novak, 1991). No tiene por qué haber un elemento principal en ellos, y la información acerca de sus conexiones es mucho más expresiva que en los *Mindmaps*, pero seguía sin permitimos mostrar todo el potencial de las relaciones de *PuertoTerm*.

De esta forma se optó por un solución intermedia, el software *ThinkMap*, que combina lo mejor de ambas representaciones de una manera eficaz para el desarrollo de nuestra aplicación (Kroeker, 2004). Es empleado con notable éxito en productos comerciales de similares características a *PuertoTerm*, como *Visual Thesaurus*, *Sony Music Licensing* o la *National Oceanic and Atmospheric Administration* de los EUA. Ayuda a organizar la información dotándola de sentido, ya que se presenta de forma contextual, tomando un elemento como central, y proporcionando toda la información necesaria para caracterizar sus relaciones.

<http://www.thinkmap.com>

<http://www.visualthesaurus.com>

<http://www.sonymusic.com/licensing>

<http://www.noaa.gov/>

El programa facilita una infraestructura (*framework*) que permite desarrollar aplicaciones muy flexibles (con un innovador entorno) extremadamente eficientes, fácilmente escalables y que puedan interactuar con diferentes fuentes de información. Éste último hecho adquiere especial relevancia en *PuertoTerm*, ya que aunque en estos momentos los *Think*

Maps implementados obtengan toda la información a partir de la base de datos (incluida la semántica), en nuestros futuros trabajos está la idea de que se alimente directamente de la ontología, algo que sería mucho más natural, siendo una opción no disponible en otros sistemas.

ThinkMap emplea un conjunto de algoritmos que permiten mostrar grandes cantidades de información de manera eficiente por medio de una interfaz bidimensional. De esa forma es posible comprender fácilmente la relación existente entre cientos de nodos de información. Además, la representación mostrada no es estática, ya que el usuario puede en todo momento realizar continuas búsquedas a partir de los resultados de una primera. Conforme se van realizando, los gráficos se regeneran, creciendo o decreciendo la cantidad de nodos mostrados en función del número de relaciones que existan entre los nuevos términos localizados. La figura 7 muestra las diversas opciones de búsqueda, historial, y representación de la información que se han implementado en *PuertoTerm* empleando esta herramienta.

Otro aspecto de especial relevancia para nuestros propósitos radica en la facilidad que tiene el programa para integrar contenidos multimedia. Este nuevo tipo de información visual complementa y amplía el contenido textual, como queda atestiguado en numerosos ejemplos (*Le grand dictionnaire terminologique*) o la inclusión de opciones para la gestión de contenido multimedia en la mayoría de las herramientas para la gestión terminológica.

<http://www.granddictionnaire.com>

3. Conclusiones

La gestión eficiente del conocimiento dentro de un dominio a partir de la utilización de herramientas y metodologías adecuadas resulta requisito imprescindible en una sociedad en donde la información juega un lugar tan destacado. Sin embargo, para que el conocimiento sea efectivo debe no sólo poder estructurarse para un fin particular, sino que debe poder combinarse e integrarse con otros sistemas de similares propósitos, para que dicho conocimiento sea enriquecido.

Éste es uno de los principales objetivos de la denominada web semántica, poder proporcionar una infraestructura común para la representación del conocimiento y, de esta manera, que la información no sea propietaria de las aplicaciones que la utilizan. En la actualidad la web está orientada al intercambio de documentos entre humanos. Al utilizar para la gestión del conocimiento tecnologías más expresivas en su semántica, se facilitaría además el intercambio y la inferencia de nuevo conocimiento por parte de las propias máquinas.

tology web language) (Smith, 2004). De esta manera el conocimiento de *PuertoTerm* se podría relacionar con aplicaciones llevadas a cabo dentro de su área.

Aunque la ingeniería de costas y puertos es una disciplina con ciertas peculiaridades (como ya hemos comentado) la metodología seguida en este trabajo se puede extender a otras áreas en donde las aplicaciones que se desean desarrollar compartan los objetivos, adaptando las necesidades específicas. La reutilización y combinación de muchas de las herramientas disponibles en la actualidad para la creación de aplicaciones ha propiciado la construcción de una infraestructura completa sin que ello haya supuesto hacerlo desde cero, algo que es fácilmente extrapolable.

En la actualidad nos encontramos completando y mejorando la calidad de la información contenida en *PuertoTerm*. Esta tarea se verá rematada por el diseño de una herramienta visual de consulta semántica que aproveche las características de *PuertoTerm* directamente desde la ontología y no desde la semántica almacenada en la base de datos. Como hemos comentado anteriormente, este hecho no supondría cambiar de herramienta de visualización, con el consiguiente ahorro de aprendizaje asociado para los usuarios. Por último también se plantea la posibilidad de incorporar nuevas tecnologías emergentes en la web semántica, tales como los microformatos, *RDFa* y *Grddl* (*gleaning resource descriptions from dialects of languages*). Todas estas herramientas están orientadas a la inclusión de la semántica en las propias páginas web, permitiendo que la base de conocimiento no permanezca oculta en un segundo plano.

4. Nota

1. Esta investigación forma parte del proyecto "Marcos de conocimiento multilingües en la gestión integrada de zonas costeras" [*MarcoCosta*](P06-HUM-01489), subvencionado por la Junta de Andalucía.

5. Bibliografía

Berners-Lee, T.; Hendler, J.; Lassila, O. "The semantic web". En: *Scientific American magazine*, 2001, v. 284, n. 5, pp. 34-43.

Faber, P., et al. "Process-oriented terminology management in the domain of coastal engineering". En: *Terminology*, 2006, v. 12, n. 2, pp. 189-213.

Faber, P.; Mairal-Usón, R. *Constructing a lexicon of English verbs*. Berlin: Mouton de Gruyter, 1999.

Faber, P.; Márquez, C.; Vega, M. "Framing terminology: a process-oriented approach". En: *Meta*, 2005, v. 50, n. 4. Consultado en: 12-03-07. <http://www.erudit.org/livre/meta/2005/000255co.pdf>

Fillmore, C. "Frame semantics". En: *The Linguistic Society of Korea* (ed.). *Linguistics in the morning calm*. Seoul: Hanshin, 1982, pp. 111-137.

Fillmore, C. F.; Johnson C. R.; Petruck, M. R. L. "Background to FrameNet". En: *Journal of lexicography*, 2003, v. 16, n. 3, pp. 235-250.

Fillmore, C.; Atkins, S. "FrameNet and lexicographic relevance". En: *Proceedings of the Granada conference on linguistics*.

Gahl, S. "Automatic extraction of subcategorization frames for corpus-based dictionary making". En: *Euralex'98 proceedings*, 1998, pp. 445-452.

Irazazábal, Amelia. "Terminología y documentación". En: *Jornada panlatina de terminología: perspectivas i camps d'aplicació*. 1996. Consultado en: 08-03-07. <http://www.realiter.net/jorb/irazazabal.htm>

Kroeker, Kirk L. "Seeing data: new methods for understanding information". En: *IEEE computer graphics and applications*, 2004, May/Jun, v. 24, n. 3, pp. 6-12.

López-Rodríguez, Clara-Inés; Tercedor-Sánchez, Maribel; Faber-Benitez, Pamela. "Gestión terminológica basada en el conocimiento y generación de recursos de información sobre el cáncer: el proyecto Oncoterm". En: *RevistaSalud.com*, v. 2, n. 8, 2006. Consultado en: 07-03-07. <http://www.revistaesalud.com/index.php/revistaesalud/article/view/127/322>

Lyons, J. *Semantics*. London: Cambridge University Press, 1977.

Martín-Mingorance, L. "Functional grammar and lexematics". En: Tomaszczyk, J.; Lewandowska, B. (eds.). *Meaning and lexicography*. Amsterdam/Philadelphia: John Benjamins, 1989, pp. 227-253.

Martín-Mingorance, L. "Lexical logic and structural semantics: methodological underpinnings in the structuring of a lexical database for natural language processing". En: Hoinkes (eds.). *Panorama der Lexikalischen Semantik*. Tübinga: Gunter Narr, 1995, pp. 461-474.

McEnery, A.; Wilson, A. *Corpus linguistics*. Edimburgo: Edinburgh University Press, 1996.

Novak, J. D. "Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning". En: *Instructional science*, 1991, n. 19, pp. 1-25.

Smith, M. K.; Welty, C.; McGuinness, Deborah L. *OWL web ontology language guide. W3C Recommendation*, 2004. <http://www.w3.org/TR/owl-guide/>

Pérez-Álvarez-Ossorio, J. R. *Introducción a la información y documentación científica*. Madrid: Alhambra/Universidad, 1998.

Pinto, María; Cerdón, José-María (eds.). *Técnicas documentales aplicadas a la traducción*. Madrid: Síntesis, 1999.

Puertos del Estado. Recomendaciones para obras marítimas. ROM. 0.0 Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias. Madrid: Ministerio de Obras Públicas y Urbanismo, 2001, 220 págs. ISBN 84-88975-30-9.

Sales-Salvador, Dora. *Documentación aplicada a la traducción: presente y futuro de una disciplina*. Gijón: Trea, 2006.

ISO-5964. *Guidelines for the establishment and development of multilingual thesauri*. Geneva: International Organization for Standardization, 1985.

Jose A. Senso, Departamento de Biblioteconomía y Documentación, Universidad de Granada. jsenso@ugr.es

Pedro-Javier Magaña-Redondo, Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada. pedro@correo.ugr.es

Pamela Faber-Benitez, Departamento de Traducción e Interpretación, Universidad de Granada. pfaber@ugr.es

María-Amparo Vila-Miranda, Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada. vila@decsai.ugr.es