PuertoTex: un *software* de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías

PuertoTex: a data mining software based on ontologies for automatic summarization on port and coastal engineering domain

Amed LEIVA-MEDEROS[1]
Sandor DOMÍNGUEZ-VELASCO[1]
Jose Antonio SENSO[2]

Resumen

El objeto de este trabajo es elaborar y evaluar los resultados de la implementación de un *software* basado en ontologías, capaz de generar resúmenes automáticos en el campo de la Ingeniería de Puertos y Costas. Para el desarrollo de la herramienta se emplean diversas técnicas emanadas del análisis de discurso así como técnicas cognitivas, que permiten generar reglas para el tratamiento de los textos. También se apela a la construcción de una ontología que facilite los procesos de etiquetado a partir de las potencialidades de *Resource Description Framework* y *Extensible Makup Language*. Se construye un conjunto de agentes que actúa sobre la ontología, de la cual se declaran sus principales elementos. Como producto generado se presenta Puertotex, un *software* para la construcción de resúmenes automáticos basado en ontologías. La evaluación de los resúmenes generados refleja la calidad del sistema, que tiene como única limitación su capacidad para trabajar con el dominio objeto de investigación.

Palabras-clave: Autómatas. Desambiguación de textos científicos. Minería de texto. Ontologías. Resumen automático.

Abstract

The aim of this word to develop and evaluate the results of the implementation of an ontology-based software capable of generating automatic summaries in the field of Ports and Coastal Engineering. For the development of the tool has been used several techniques emanating from the analysis of discourse and cognitive techniques for generating rules for the treatment of the texts. It also calls for the construction of an ontology that facilitates the processes of labeling from the potential of Resource Description Framework and Extensible Makup Language. It has built a set of agents acting on the ontology, which declared its main elements. Are presented as a product generated, Puertotex, software for the construction of ontology-based automatic summaries. The evaluation of the summaries generated reflects the quality of the system, which is limited only by your ability to work with the domain under investigation.

Keywords: *Automata. Disambiguation of scientific texts. Data mining. Ontologies. Automatic summarization.*

---

[1] Profesores Doutores, Universidad Central Marta Abreu de las Villas, Facultad de Ciencias de la Información. Santa Clara, Cuba.

[2] Profesor Titular, Universidad de Granada, Departamento de Biblioteconomía y Documentación. *Campus* de Cartuja, 18071, Granada, España. Correspondencia a nombre de/*Correspondence to*: J.A. SENSO. *E-mail*: <jsenso@ugr.es>.

Recibido el día 19/12/2011, re-apresentado el 4/5/2012 y aceptado para su publicación el 23/5/2012.

Introducción

A diferencia de la indización y la clasificación, los resúmenes son una reelaboración textual, una propuesta por reducción significativa de los originales que parte de algunas cuestiones previas, como las características del texto a resumir, el destinatario del resumen, los objetivos que se planteen y el proceso que le da lugar, y datan, como servicio organizado, desde mediados del siglo XIX (Hernández, 2007a).

Los resúmenes se erigen como metadatos, pues son la expresión que enuncia el contenido de los documentos textuales (Herrera, 2007). En la literatura de la Ciencia de la Información constantemente se hace referencia a sistemas de metadatos como *Dublin Core*, *Encoded Archival Description* (EAD), Norma Internacional General de Descripción Archivística (ISAD-G), *Machine Readable Cataloging* (MARC 21) etc. Aunque muchos de estos sistemas incluyen etiquetas específicas para incluir un breve resumen del objeto referenciado, en pocos casos se habla del resumen documental como metadato que representa de forma resumida el contenido de los textos, pues es poco su reconocimiento como estructura representacional de datos no estructurados. Esta forma de procesamiento de la información ha sido tratada tradicionalmente como un proceso netamente lingüístico, computacional y semántico.

La producción automática de resúmenes se ha basado en métodos estadísticos desde las investigaciones de Lunh (1958) y continúa hoy con los trabajos del Procesamiento del Lenguaje Natural, las aplicaciones basadas en analizadores automáticos y en modelos socio--pragmáticos del discurso basados en técnicas de "dominio de personalidad propia", subyugados por la extracción de palabras o cadenas significativas, que ya en los años 1990 recortan las "estructuras retóricas" (Mann; Thompson, 1990; Endres-Niggemeyer *et al*., 1995; Endres-Niggemeyer, 2005), muy relacionadas con el contenido de los documentos, al aportar indicios sobre el esquema conceptual y organizador de las distintas unidades que conforman un texto para reproducir y ordenar determinadas frases (Ono *et al*., 1994).

Este desarrollo ha sustentado la aparición de múltiples herramientas para la extracción de conocimiento en materias específicas. Sin embargo, existen disciplinas que no poseen *software* que permitan especificar y detallar el conocimiento que poseen, tal es el caso de La Ingeniería de Puertos y Costas. Para Sánchez--Cuadrado *et al.* (2007) parten de la premisa, compartida en otros trabajos similares, de que la mayor parte del conocimiento está explicitado en los documentos de un dominio mediante términos y relaciones y que sólo el conocimiento que no esté expresado en los documentos tendrá que ser aportado por los expertos del dominio.

La Ingeniería de Puertos y Costas es una disciplina que difiere del resto de las ingenierías. La inexistencia de diccionarios específicos, la ausencia de teorías específicas para este campo multidisciplinar y la necesidad que tienen los profesionales que en ella trabajan de poseer herramientas terminológicas y léxico semánticas para realizar su trabajo le confieren un cierto halo de área discriminada, al menos hablando desde el punto de vista de la documentación y la terminología (Leiva *et al*., 2009). Esta situación da pie a los objetivos del presente texto. Por un lado, explicar los pasos seguidos para elaborar un *software* basado en ontologías capaz de generar resúmenes automáticos en el campo de la Ingeniería de Puertos y Costas. Por otro, evaluar las cualidades de este *software* dentro del campo de la Minería de Texto.

Métodos

Para la realización de este trabajo se ha utilizado una muestra aleatoria del corpus del proyecto Puertoterm (http://lexicon.ugr.es/projects) de la Universidad de Granada (2 890 documentos, de los cuales se seleccionan 1 390) a la que se aplicó técnicas de análisis de discurso para determinar reglas cohesivas amparadas en la teoría de Mann y Thompson (1990). Se ha utilizado el concepto de discurso asumido por Dijk, quien lo denota como una forma específica de uso del lenguaje, un suceso de comunicación (quién utiliza el lenguaje, cómo lo utiliza, por qué y cuándo lo hace) en el que se transmiten ideas, visto también como una forma específica de interacción social. Así, el discurso se interpreta como un evento comunicativo real (completo) en una situación social dada.

La función fundamental del lenguaje realizado como discurso es la constitución de una realidad semiótica. Ésta se estructura de tal manera que podamos comprenderla y apoderarnos de ella (Metzeltin, 2003). De esta forma, el análisis del discurso, básicamente, se ocupa

de la dimensión interactiva e intersubjetiva del uso del lenguaje mediante la investigación y análisis de datos reales (Pilleux, 2001).

El estudio del discurso encuentra razón en los fenómenos detrás de la oración, aunque no deja de observarlo como un factor dinámico de las interacciones sociales que permiten encontrar modelos para su interpretación y análisis, teniendo en cuenta que los discursos pueden describirse en términos de "acciones sociales que llevan a cabo los usuarios del lenguaje cuando se comunican entre sí en situaciones sociales y dentro de la sociedad y la cultura en general" (Dijk; Kintsch, 1983, p.38).

Unido a estas técnicas se aplican métodos de estudio de usuarios a 23 expertos en el terreno de la Ingeniería de Puertos y Costas conjuntamente con la modelación de Sistema. Para la realización del *software* se utilizó el lenguaje de programación Python.

Para evaluar el sistema se usaron varias metodologías. Por un lado se ha analizado la usabilidad del sistema por medio de un estudio heurístico. Luego se ha analizado la capacidad del sistema para recuperar información determinada. Y para finalizar se analiza la calidad del producto resultante, el resumen automático, empleando dos métodos de evaluación.

PuertoTex: componentes del sistema

PuertoTex es un sistema de Minería de Textos basado en reglas de discurso emanadas de los textos que consumen y generan diversos grupos de investigadores dentro del campo de la Ingeniería de Puertos y Costas. Las prestaciones que facilita este *software* están en consonancia con las necesidades de la comunidad de usuarios a los que va dirigido este tipo de información. Los agentes de *software* que forman este sistema emulan las estrategias resumidoras de 12 expertos en Ingeniería de Puertos y Costas. A continuación analizaremos sus componentes y fases de trabajo.

Corpus de trabajo

Los corpus que han servido de sustento para esta investigación poseen características muy disímiles en lo referente a nivel discursivo, prestigio de sus fuentes y tipología documental. Esta situación hace que las estrategias de trabajo y análisis sean más complejas, debido a la disparidad en la concepción del corpus (en inglés y español). Si se analiza el corpus desde el punto de vista de estructura discursiva puede llegarse a la conclusión de que el inglés es superior, pues el discurso que genera es eminentemente especializado. Más de 4 000 registros están en esta categoría, de los cuales 818 son semi-especializados y solo 43 son artículos divulgativos. El corpus español es, en esencia, divulgativo y en menor grado especializado.

La ontología

A partir de los documentos de ese corpus se pudo crear Ontosatcol, ontología diseñada para la desambiguación, extracción y búsqueda de información y empleada por la herramienta PuertoTex. Esta ontología (Figura 1) cuenta con 3 005 conceptos, 3 006 subconceptos y 28 900 anotaciones. Posee también 12 propiedades individuales para cada objeto y 8 propiedades especiales para los datos. La herramienta dispone, además, de 890 instancias.

La Ontología del Sistema es generada a partir de los términos que han sido etiquetados en el Corpus, y para ello el sistema incluye un marcador o etiquetador que facilita dicho proceso. Una vez marcados los vocablos dentro del texto, estos comienzan a organizarse teniendo en cuenta la Semántica de Marcos, un modelo de organización del conocimiento que permite que los términos sean descritos de acuerdo a diversas formas de actuación dentro de un contexto, formando un concepto único (Faber *et al*., 2005). Cada término representa una disciplina, por tanto debe poseer uno o varios eventos (o marcos conceptuales) que describen procesos o acciones que tienen lugar dentro de la disciplina representada por el término. Dichos eventos poseen una estructura que facilita la creación y organización de los diferentes conceptos que los integran. La elaboración del marco conceptual es el culmen del análisis de las definiciones lexicográficas y de corpus extraídas previamente (Senso, 2009).

Una vez finalizada la fase de construcción de la semántica de marcos comienzan a estructurarse los conceptos y se construyen todas las estructuras

lingüísticas de las clases a partir de diversos elementos o relaciones de diversos tipos, como merónimos, hiperónimos o hipónimos. Para estructurar las mencionadas relaciones se contrastan los términos contra obras especializadas en Ingeniería de Puertos y Costas, así como en diccionarios, glosarios, enciclopedias y listas de folksonomías. También se declaran las propiedades de los datos y las de los objetos.

Etiquetado

Para desarrollar los procesos de resumen se llevó a cabo el etiquetado en *Extensive Markup Language* (XML) de los elementos del texto de cada uno de los documentos seleccionados del corpus, con el fin de organizar las unidades textuales en segmentos. Dichos segmentos aúnan elementos de diversos órdenes como el discursivo, el sintáctico y el comunicativo que, a su vez, son referentes internos de las cargas semánticas y estructurales de cada oración. El desarrollo de estas etiquetas permite la obtención de resúmenes de un texto único.

Los textos se han etiquetado tanto a nivel oracional como discursivo. En este punto los autores encontraron que no existe ninguna herramienta para el marcado de relaciones discursivas en español o que facilite la marcación de núcleos, satélites, temas y remas en los textos. Con el objetivo de emplear un sistema semiautomático de etiquetado se ha decidido desarrollar una interfaz de anotación utilizando como herramienta el programa *XMLMarker* (http://symbolclick.com/), editor XML que, entre otras cosas, permite administrar bases de datos en XML mediante interfaces *Web*, además de ejecutar sentencias en XML y en Python.

Los pasos que se han seguido para la concepción de las etiquetas obedecen a tres criterios fundamentales: la secuencia en que se realiza el etiquetado, la simbología (el nombre de los elementos XML) utilizada para declarar

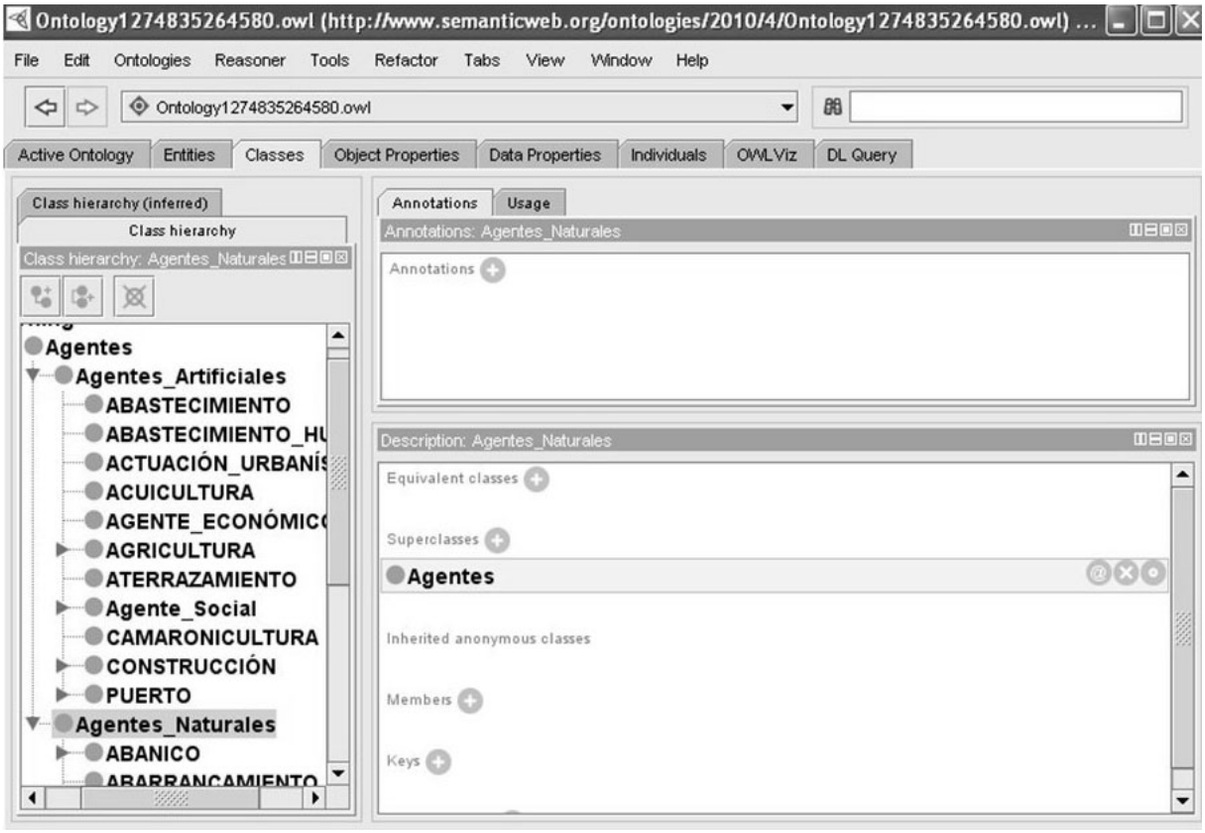

Figura 1. Aspecto de la ontología desde el editor Protégé.

Fuente: Elaboración propia.

los elementos cohesivos del texto y la presencia en el texto de elementos geográficos, verbos, datos estadísticos, fórmulas y procesos.

Lo primero que se marca es la oración que no estará prefijada por un punto, sino por la presencia de formas verbales y sus conectores. A continuación se etiqueta la estructura discursiva de cada sentencia de forma independiente; en otras palabras, se delimitan los núcleos y los satélites.

Como puede intuirse, todas las etiquetas que explican relaciones de tipo núcleo-satélite poseen una estructura similar, a excepción de los casos en que existe un satélite discursivo en relaciones de secuencia, lista o unión. En el caso de la presencia de satélites en la relación de elaboración, se enuncian en las marcas de discurso los elementos tema y rema de su núcleo. A igual que en los trabajos de D'Cunha (2006), esta construcción semántica se ejecuta a través del atributo <tem>, cuyo valor será el tema o el rema de su núcleo (marcados como "t" y "r", respectivamente, junto con el número con el que aparecen en su etiqueta), dependiendo de a cuál de los dos se refiera el satélite (por ejemplo, <item="t">, <item="r">). Es imposible encontrar satélites en relaciones multinucleares, lo cual obliga a declarar en la marcación solo los diversos núcleos, que poseen igual nivel de significado para el resumen.

En este sistema de etiquetado se han declarado elementos de la superestructura del texto, lo que permite que el mismo sea reconocido por estándares como el *Dublin Core, Resource Description Framework* (RDF) y *Friend of a Friend* (FOAF).

Agentes

El objetivo fundamental de esta aplicación es que sea capaz de reproducir e imitar el proceso que sigue un ser humano a la hora de realizar un resumen. Para ello se ha diseñado un conjunto de agentes encargados de realizar diversas tareas, con la meta de conseguir dicho objetivo.

Agente de Lectura: se encarga de realizar la primera estrategia cognitiva que hacen los resumidores. Lee el documento y comprueba que esté acorde con el dominio de la ingeniería de puertos y costas. Para ello localiza el texto marcado en XML y contrasta el conocimiento contra la ontología y la base léxica, determinando que el texto es relevante para el dominio ya que posee términos que son relevantes para él. Envía al agente de resumen el texto para que este lo resuma.

Agente de Resumen: este agente toma el texto que le envía el agente de lectura y realiza dos procesos. Primero extrae todas las oraciones candidatas para el resumen de acuerdo a las calificaciones declaradas en el análisis del discurso; o sea, un resumen por extracción. En un segundo paso hace un resumen por abstracción, es decir, reescribe el texto utilizando reglas heurísticas que tiene implementadas. Mediante estas reglas logra la cohesión, la coherencia y el balance textual. En el proceso de extracción, el agente utiliza diversas reglas predefinidas para gestionar la información en dos sentidos. Dichas reglas son de ponderación (permiten elegir las oraciones de acuerdo a la presencia en ellas de términos que hagan referencia a: elementos estadísticos, elementos químicos, nombre geográficos y términos de la lista del discurso), sintáctico-comunicativas (encargadas de la coherencia del texto, manejando criterios que permitan detectar y eliminar los núcleos y satélites en las oraciones y suprimir y eliminar elementos de tema y rema) y de desambiguación (para darle sentido al texto. Se centran en la presencia de pronombres anafóricos para detectar los antecedentes de los grupos oracionales y la asignación de signos de puntuación en las oraciones). El conjunto de autómatas con los que se trabaja en esta fase están basados en una modificación del algoritmo *Multivariate Adaptive Regression Splines* (MARS) (Sidorov; Zazueta, 2006; Domínguez, 2011).

Agente de Desambiguación: una vez realizado el resumen, este agente recibe el texto y reconoce en él signos de puntuación erróneos y palabras fuera de contexto utilizando el algoritmo de Aguirre (Aguirre, 1998; Rigau, 2002) y verificando los vocablos por medio de una comparación con las clases de la ontología.

Agente de Normalización: se encarga de determinar la estructura del resumen. Recibe el texto del agente de desambiguación y le asigna al texto el segmento de referencia del artículo, así como su *Uniform Resource Locaton* (URL), para que la presentación del resumen permita determinar dónde está el texto fuente y su corpus de referencia.

Agente de Búsqueda: es el agente de más trabajo en el sistema, hace búsquedas a través de la ontología y

a través de la base de datos léxica desarrollada. En la ontología se encarga de buscar por autor, título, materia y tipo de documento utilizando las bondades del sistema de metadatos *Dublin Core*. También permite buscar en redes sociales desarrolladas a partir de los datos de los autores que se encuentran en la estructura FOAF. Friend of a Friend facilita la búsqueda y recuperación de los siguientes elementos: institución del autor y título del proyecto, así como el correo electrónico del autor y los autores que colaboran con él.

Agente de Redundancia: conecta la solicitud del usuario con los resúmenes que ya tiene almacenados en la base de datos. En el caso de que el documento solicitado ya se encuentre resumido, el agente le proporcionará dicho resumen. Cuando se realice la solicitud de un texto que no está en el corpus, el agente le pedirá al usuario que indique dónde se encuentra, para localizarlo (generalmente por medio de la URI), incluirlo en el corpus e iniciar el proceso con el agente de lectura.

Una vez que un documento entra al sistema se obtiene su contenido semántico y se ubica en la ontología. A continuación se realiza un procesamiento léxico--semántico a partir de la base de datos léxica del sistema (ontología). El siguiente paso consiste en realizar las marcas de discurso en el texto, de acuerdo con la heurística (estrategia cognitiva) tomada de los resumidores humanos que participan en este estudio, generando un documento etiquetado en XML. De ahí el documento se envía al agente de lectura que se encarga de leer el documento emulando el comportamiento de un lector humano. Cuando este agente termina la lectura y comprueba que el texto está etiquetado correctamente, envía el texto al agente de resumen. Éste utiliza varias estrategias de resumen - tal y como se explica en el punto 3.4 - para determinar las oraciones relevantes, asignar pesos a los términos y extraer las oraciones (hace el resumen por abstracción), buscando luego la cohesión mediante el tratamiento de núcleos y satélites de acuerdo con las reglas sintáctico-comunicativas. El resumen resultante de este paso aún no es legible, por tanto el agente de desambiguación tiene que darle sentido a los términos utilizando la ontología. Después de la desambiguación se pasa a la normalización del texto; es decir, se colocan los puntos, las comas y las unidades necesarias para que el texto sea legible. Resumido el texto, este se almacena y está listo para que el agente de búsqueda localice informaciones (teniendo en cuenta los valores que posibilita la ontología) e interrogue al agente de redundancia para cuando se realice una solicitud de información que ya ha sido procesada.

Las operaciones llevadas a cabo en PuertoTex

La aplicación posee diversas opciones que facilitan el trabajo por medio de un entorno basado en formularios *Web*. Así, es posible el registro, la búsqueda y recuperación de información y la posibilidad de resumir una fuente concreta.

El módulo principal carga la herramienta con las opciones para el ingreso en el sistema, el registro y la ayuda para que llegue a completar todas las acciones que se realizan desde la ventana principal.

El módulo de búsqueda permite una consulta basada en diversos criterios: materia, autor, año y personas. La consulta por materias usa los conceptos expuestos en la ontología, permitiendo el uso de operadores booleanos, y permite visualizar los conceptos y sus dimensiones semánticas de forma dinámica. La búsqueda de autor devuelve el título o lista de títulos que tenga asociados, accediendo al texto completo pinchando dos veces sobre el título escogido.

Como ya se comentó anteriormente, en el proceso de etiquetado se ha empleado FOAF. Así, los nombres de los autores y de las personas o instituciones que se mencionan en los textos están conectados entre si, formado una red social. De esa manera es fácil obtener información sobre los expertos de un sub-área de conocimiento concreto dentro de la Ingeniería de Puertos y Costas.

Es en este módulo donde también se encuentra la función principal del sistema: resumir una fuente. Una vez que el usuario ha seleccionado un título concreto de la lista de resultados tras su consulta, el agente de búsqueda (apoyado por diversas reglas de inferencia y de las capacidades semánticas de la ontología) filtra la información y localiza el documento en la base de datos, mostrando el resumen del texto, la superestructura del resumen y una URI para que el usuario consulte o

descargue el documento original. Antes de que se resuma la fuente, cada documento pasa por diversos procesos.

Las reglas empleadas para la realización de resúmenes en PuertoTex se han llevado a cabo teniendo en cuenta tres rasgos esenciales, al igual que en los trabajos de D´Cunha (2006) y Paneca (2009): 1) Estructura textual; 2) Unidades léxicas y 3) Estructura discursiva y sintáctico-comunicativa.

Desde el punto de vista de la implementación, se ha intentado hacer que las reglas coincidan con los procesos de realización del resumen, con sus operaciones y la estrategia de construcción del mismo. Estas reglas se unen a las posturas de lectura y son implementadas por agentes de *software* cuyo nivel de especialización en la lectura y construcción del texto ofrece confiabilidad en la calidad del proceso. De esta forma, el agente de lectura y el de resumen asumen diversos niveles textuales para construir un extracto: 1) Nivel textual: consiste en la implementación de reglas basadas en la estructura del texto (denominadas reglas textuales); 2) Nivel léxico: detección, eliminación y selección de unidades léxicas de la base de conocimientos y de la ontología del sistema y 3) Nivel discursivo y sintáctico-comunicativo: formalización de normativas de las reglas discursivo-sintáctico--comunicativas (que hemos denominado reglas PuertoRules).

El análisis de los dos corpus con los que se ha trabajado nos ha mostrado la existencia de una estructura retórica: Introducción, Objetivos, Materiales y Métodos, Resultados y Conclusiones (IOMRC), al igual que lo expuesto en los estudios de Pinto (2001, 2004) y Hernández (2006), y que da sentido a la existencia del agente de lectura de PuertoTex. IOMRC es una extensión de la estructura Introducción, Metodología, Resultados y Conclusiones (IMRC), que facilita la construcción de resúmenes informativos e indicativos. Los primeros se basan exclusivamente en la estructura retórica del texto científico IOMRC y se obtienen mediante métodos de abstracción. Los resúmenes que se realizan sobre textos divulgativos (caso que sucede principalmente con el corpus en español) solo son de categoría indicativa, ya que los textos divulgativos no poseen estructura IOMRC y se construyen exclusivamente a partir de métodos de extracción.

Esta regla, que vamos a denominar *modelo de reconocimiento de texto*, se ha creado con la finalidad de localizar en el texto las unidades retóricas específicas, es decir los apartados introducción, metodología, resultados y conclusiones clásicos de los trabajos científicos.

Partiendo de los criterios de Lunh (1958) y las modificaciones de Mathis *et al*. (1973) se ha constatado la posición estratégica de determinadas oraciones en los segmentos de los textos analizados, las cuales evidencian mayor relevancia que otras en el momento de la selección del resumen.

A modo de táctica operacional, para este *software* se siguen los mismos supuestos teóricos de Lunh (1958) y Salton (1996), donde los modelos de resumen utilizan una aproximación por extracción donde ocurren procesos de transformación de corpus en los cuales se convierte el texto en tokens y se eliminan los signos de puntuación existentes en su volumen. Las acciones propuestas para PuertoTex son las siguientes:

Eliminar oraciones o fragmentos del texto: 1) Asignar un peso ponderado a cada oración del texto según la presencia en ella de vocablos, y su función o posición en el texto y 2) Suprimir oraciones de acuerdo al poco peso en sus niveles semánticos.

A partir de esta forma ya clásica en el tratamiento del resumen se ha decidido desarrollar diversos tipos de reglas para operar con un texto científico: a) Reglas que suprimen secuencias del texto fuente, oraciones y/o pasajes textuales con determinado nivel léxico-semántico; b) Reglas que delimitan aquellas oraciones que en el texto original han de ser eliminadas y almacenadas para otros tipos de construcción; c) Reglas que obligan a buscar en otros niveles de descripción si los candidatos oracionales no son correctos, es decir buscar en la descripción de otros niveles de texto (imágenes y gráficos); d) Reglas que aportan relevancia a las oraciones en dependencia del lugar donde se encuentren y de la presencia en ellas de vocablos o términos de las listas de prioridad y e) Reglas que permiten desarrollar estrategias de desambiguación léxica.

De esta manera, el modelo de extracción de texto asume dos fases con las siguientes particularidades:

Fase I: Constituye la supresión de oraciones y pasajes textuales de bajo nivel de relevancia en el texto (usar reglas del tipo a).

Fase II: Se proponen oraciones o fragmentos de texto candidatos a ser eliminados (mediante las reglas del tipo b); Si no existen oraciones que cumplan este criterio, buscar otros textos asociados al texto que permitan mejorar la descripción. Reglas de tipo c; Descartar aquellas oraciones que poseen menos nivel de puntuación y faciliten la proyección de aquellas que poseen más alto nivel de relevancia; Aplicar reglas de sintaxis y comunicación para lograr la cohesión del texto y Desambigüar el texto para brindar mayor nivel de claridad en la lectura de la información (regla d).

La aplicación exacta de estos elementos en el sistema de desambiguación y extracción quedan de la siguiente forma:

Nivel 1: Es utilizado y visualizado como un nivel de segmentación de texto donde, al igual que en varios estudios como los de Endres-Niggemeyer (2005), se delimita el texto en 5 apartados con los contenidos que poseen mayor nivel de preponderancia. En ocasiones estos apartados no están delimitados, lo que obliga a que el agente de lectura reconozca los elementos por los que tradicionalmente se marcan estos segmentos del texto y confronte la base de sinónimos donde aparecen sinónimos de estos apartados. Esto permite asignar la estructura IOMRC.

Nivel léxico: utiliza normas basadas en unidades léxicas que facilitan la asignación de relevancias a determinados segmentos textuales y/u oraciones. En este nivel también se aplican reglas basadas en unidades léxicas, que asignan los signos de puntuación a las oraciones.

Nivel discursivo y sintáctico-comunicativo: consistente en la aplicación de reglas (PuertoRules) donde, dependiendo de cada segmento textual, se aplicará un determinado elemento de cada relación discursiva. Estas reglas determinan qué elementos discursivos se mantienen o se eliminan, dependiendo de las marcas de cohesión que tenga el texto en función a la relevancia previamente otorgada por los expertos a ciertos grupos textuales y oracionales y almacenadas en la base de datos.

La segunda generación de reglas PuertoRules trabajan en consonancia con los modelos de cohesión de Mann y Thompson (1990). En el modelo de D´Cunha (2006) se tiene en cuenta la longitud de las oraciones deseadas por el usuario; sin embargo, no existe constatación de que ése sea un criterio fundamental para determinar la calidad de un resumen. En el modelo extractivo que aquí se describe los autores han creído pertinente desarrollar una fase de desambiguación (Nivel 4) a nivel oracional, utilizando mejoras en las oraciones resultantes de los niveles anteriores.

Fase final: constituida por la unión de las oraciones de todos los apartados anteriores. La parte superior (Anexo 1) muestra un documento original almacenado en el repositorio del sistema. En la parte inferior podemos encontrar el resumen generado automáticamente a partir de dicho original.

Evaluación del sistema

Existen multitud de trabajos centrados en analizar diversos aspectos relacionados con la evaluación de la calidad de los sistemas de recuperación de información. Pero dado que no existe un único método que sea capaz de evaluar este tipo de sistemas de manera integral, los autores se han decidido por realizar tres evaluaciones diferentes del mismo, con la idea de analizar la mayor cantidad posible de vertientes de PuertoTex. Por un lado se ha analizado la usabilidad del sistema por medio de un estudio heurístico. Luego se ha analizado la capacidad del sistema para recuperar información determinada. Para finalizar se analiza la calidad del producto resultante - el resumen automático -, empleando Rouge y un sistema de evaluación por jueces.

La selección de la muestra ha sido de forma aleatoria, escogiéndose 1 200 artículos en inglés y otros tantos en español de entre los almacenados en el sistema. Para las diferentes evaluaciones se han empleado a 94 usuarios en total. Dependiendo de cada fase, estos usuarios eran expertos en un área concreta: documentalistas, ingenieros, biólogos etc.

Evaluación de la usabilidad

Para esta primera fase se diseñó un experimento donde expertos y usuarios finales intervienen en la evaluación del mismo. En esta fase se han empleado a 59 personas: 5 expertos en Ciencia de la Información, 42 expertos en Medio Ambiente y 12 usuarios de especialidades relacionadas con el tema de la herramienta. En esta investigación se ha determinado emplear un estudio de usabilidad denominado usabilidad de "inspección". Se trata de una forma de evaluación que

tiene nexos directos con la evaluación heurística, actividad que consiste en el concurso de un grupo de evaluadores a los que se encarga la valoración de la interfaz del sistema mediante principios heurísticos de usabilidad. Cada evaluador realiza la valoración de forma individual, como si fuese el usuario del sistema. La revisión se realiza de manera individual y asumiendo el papel de usuario.

En este análisis se han tomado como referencia algunas de las obras más destacadas sobre el diseño centrado en el usuario y, en particular, los principios heurísticos de Nielsen (1994), Machón (2002) y Nielsen (2002a, 2002b). Teniendo en cuenta el tipo de sistema que estamos evaluando, se ha creado una plantilla que recoge indicadores de análisis relativos a varios aspectos. Los parámetros que se evalúan tienen como centro la heurística propuesta por Nielsen (1994), centrada en 10 aspectos claves relacionados con la usabilidad: la navegación, la funcionalidad, el control por parte del usuario, el uso de la lengua, la ayuda en línea y guía del usuario, la información proporcionada por el sistema, la accesibilidad, la coherencia, la prevención y corrección de errores, y la claridad arquitectónica y visual del sistema.

Para lograr una uniformidad en la evaluación del sistema se desarrolló un cuestionario donde se evalúan las variables que se utilizan en la investigación heurística. En el referido cuestionario se les proporcionó a los evaluadores una serie de preguntas con tres opciones de respuesta: "no, nunca", "a veces" y "sí, siempre", a las que se le dio el valor, respectivamente, de 0, 1 y 2, siendo 2 el valor más alto. Con este proceder las respuestas de los evaluadores obtienen cualidades numéricas que se utilizan para valorar los elementos del sistema.

En la Tabla 1 se muestran los resultados de la aplicación del test heurístico de uso del usuario. La navegación y la claridad arquitectónica son los elementos de menor nivel, posiblemente porque los usuarios no están adaptados a trabajar con sistemas que utilizan ontologías como herramientas de consulta y, por lo tanto, no logran identificar a qué lugares del sistema dirigir su búsqueda. El sistema es funcional, pues un 84% de los encuestados identifica la presencia en él de diversas informaciones relativas a idioma, lenguas, esquema de usuarios, algoritmos de trabajo etc.

Por las respuestas obtenidas, se observa que PuertoTex logra que sus usuarios se sientan seguros y con control sobre las acciones, aunque hay que señalar a un grupo de usuarios que no poseen un entrenamiento con estos sistemas y que, en algunas ocasiones, dicen sentirse espiados por él anotando observaciones del tipo "este sistema piensa por mí".

Tabla 1. Usabilidad del sistema.

Variables de uso	Sí	A veces	No
Navegación	62,0	18,0	10,0
Funcionalidad	81,4	10,6	18,0
Control del usuario	89,0	11,0	0,0
Lengua y contenido	100,0	0,0	0,0
Ayuda en línea	100,0	0,0	0,0
Información del sistema	100,0	0,0	0,0
Accesibilidad	85,0	15,0	0,0
Coherencia	90,0	10,0	0,0
Prevención errores	60,0	20,0	20,0
Claridad arquitectónica	65,5	25,0	10,8

Evaluación de la recuperación de información

Para llevar a cabo esta evaluación se proporcionó un test a los usuarios y los expertos a partir de un conjunto de tareas específicas con PuertoTex. Se preparó una serie de 8 tareas y se realizó la prueba con 10 usuarios con conocimientos en ingeniería y 5 expertos, de manera individual, durante 40 minutos, puntuando cada una de las tareas hasta un máximo de cien. La media de cada uno de los resultados fue: 1) Búsqueda de un término en dos idiomas: 99,55; 2) Búsqueda de información no linguística: 96,85; 3) Búsqueda de lema: 100; 4) Claridad de presentación de la información (idiomas): 100; 5) Búsqueda de un término: 98; 6) *Feedback*: 95; 7) Claridad de resultados: 86 y 8) Búsqueda de Resúmenes: 100.

La mayoría de los parámetros que se analizan arrojan resultados bastante satisfactorios. La búsqueda de un término en dos idiomas arrojó un 99,5% de efectividad. La recuperación de información no lingüística un 96,5%. La recuperación y búsqueda de información lematizada fue efectiva en un 100%, así como la claridad de la Representación de la Información. La búsqueda de un término es efectiva en un 100%, al igual que la búsqueda de resúmenes.

Evaluación de los resúmenes automáticos

A pesar de contar con evaluaciones muy favorables en los apartados relativos a la usabilidad del sistema y a

su capacidad para recuperar información, los autores son conscientes de que el punto fuerte de PuertoTex es el de ser un sistema capaz de generar resúmenes automáticos. Por eso es importante evaluar este aspecto de manera independiente.

El primer análisis se realizó por medio de Rouge, herramienta para la evaluación de la sumarización de documentos (Lin; Hovy, 2003). Se ha convertido en una de las pautas esenciales para determinar la calidad de los resúmenes realizados mediante método automáticos. Rouge posee varias métricas para realizar la valoración de los resúmenes; esto facilita la selección de registros adecuados para su adaptación a cada proyecto. En este caso se ha utilizado *Rouge N (N-Gram Co-Ocurrence Statistic)*. Esta media estudia la cantidad de n-gramas que ocurren en un resumen candidato y en un resumen de referencia. La ecuación declara que *Rouge N* es el cociente de la suma de las referencias del resumen base entre las referencias de un resumen obtenido mediante medidas automáticas. En ella $Count_{match}\ (gram_n)$ es el número máximo de referencias recibidas por un resumen candidato y un resumen de referencia. Los n-gramas son medidas resultantes de n, que significa la cantidad de n-gramas que ocurren en ambos textos.

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{Referemce Summaries}\}} \sum_{gram_n \in S} \text{Count}_{match}\ (gram_n)}{\sum_{S \in \{\text{Referemce Summaries}\}} \sum_{gram_n \in S} \text{Count}\ (gram_n)}$$

Cuando se trabaja con Rouge es vital poseer resúmenes que sirvan de patrón o referencia para ser comparados con los resúmenes base o candidatos. En este experimento las comparaciones se realizan atendiendo a los siguientes tipos de resumen:

- Resúmenes por extracción generados por el sistema (Puertot_1a y 1b).

- Resúmenes por abstracción generados por el sistema (Puertot_2 a y 2b).

- Resúmenes por extracción realizados por 10 especialistas de la Comunidad Ciencias Biológicas.

- Resúmenes por abstracción realizados por 10 especialistas de la Comunidad Ciencias Biológicas.

- Resúmenes realizados mediante el sistema *Microsoft Word*.

- Resúmenes Baseline desarrollados por expertos en el tema.

La mayoría de los autores prefieren evaluar con Rouge utilizando como materia de análisis solo los resúmenes de los autores. La experiencia internacional en la evaluación con esta métrica aclara que no es fiable evaluar un resumen automático usando un solo tipo de resumen obtenido bajo un solo modelo, por tanto en esta evaluación se introducen otros resúmenes para ser comparados (Tabla 2).

Es evidente que todos los resúmenes hechos mediante PuertoTex son muy similares al corpus baseline; sin embargo, éstos no se asemejan al corpus construido por medio de otros sistemas (para este experimento se optó por usar la opción de resumen automático del procesador de textos *Word*).

Las medidas que se obtienen con Rouge solo son proporciones estadísticas, y es por ello que autores como Pinto (2001) y Hernández (2007b), enfatizan en que los resúmenes no se deben evaluar de acuerdo a métricas de similitud establecidas sobre patrones físicos, en este caso

Tabla 2. Resultado de aplicar Rouge N sobre un conjunto de textos generados por PuertoTex.

Rouge-N	Tex 1	Tex 2	Tex 3	Tex 4	Tex 5
Puertot_1a	0.6891	0.6734	0.6791	0.6732	0.6791
Puertot_1b	0.6781	0.6234	0.6451	0.6831	0.6321
Puertot_2a	0.6893	0.6129	0.6234	0.6612	0.6523
Puertot_2b	0.6321	0.6321	0.6247	0.6231	0.6643
Baseline	0.1345	0.1245	0.1891	0.1789	0.6745
Word	0.4567	0.4321	0.4831	0.4890	0.6267
Rouge-N	Tex 6	Tex 7	Tex 8	Tex 9	Tex 10
Puertot_1a	0.6589	0.6128	0.6238	0.6345	0.6538
Puertot_1b	0.6635	0.6238	0.6363	0.6136	0.6156
Puertot_2a	0.6123	0.6341	0.6721	0.6721	0.6825
Puertot_2b	0.6621	0.6458	0.6821	0.6239	0.6523
Baseline	0.1980	0.1213	0.1890	0.1992	0.1389
Word	0.4467	0.4125	0.4712	0.4529	0.4578

Nota: Puertot_1a y Puertot_1b son resúmenes hechos por extracción, mientras que Puertot_2a y Puertot_2b son resúmenes hechos por abstracción. Para obtener datos más representativos se emplean otros resúmenes realizados sobre el mismo corpus documental. Dichos resúmenes han sido creados automáticamente por el procesador de textos (Word) y un corpus de contraste (baseline) constituido por 10 resúmenes realizados por expertos en la materia.

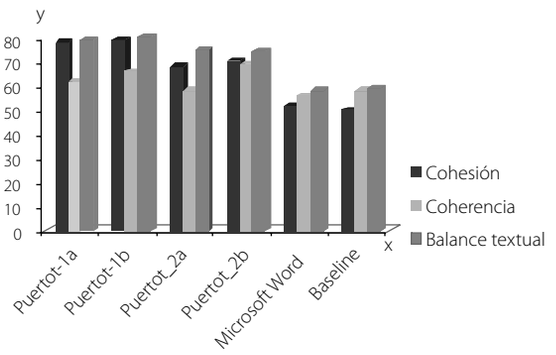

Figura 2. Evaluación de la cohesión y la coherencia textual.

Fuente: Elaboración propia.

Nota: En el eje X se encuentran los diferentes tipos de resúmenes evaluados. En el eje Y la puntuación otorgada por los observadores en base a los criterios de cohesión, coherencia y balance textual.

los bigramas. Por ello, en esta investigación se insiste en la evaluación de la cohesión y la coherencia del texto como única vía para saber si los textos resumidos cumplen con su objetivo pragmático (Pinto, 2001). Es decir, que sean resúmenes legibles y entendibles por los humanos. Para realizar estos análisis sometimos a todos los textos que se analizaron en los experimentos a la valoración de 2 jueces, que determinaron, puntuando de 0 a 100, el nivel de cohesión, coherencia y balance textual de los textos resultantes de la aplicación de los diferentes modelos de resumen que se utilizan en la evaluación (Figura 2).

Conclusiones

Si bien es cierto que existen otros métodos de desambiguación de probada calidad (entre ellos los que trabajan con resolución de anáforas), tenemos claro que la mayoría de estos métodos no contaban con la posibilidad de emplear una ontología. En el caso de PuertoTex se ha optado por crear y emplear una ontología propia, ya que defendemos la teoría de que los resultados obtenidos con un sistema supervisado nunca serán peores que los de un sistema no supervisado. Aquí se ha partido de la idea de que el texto, como unidad íntegra, debe estar desambiguado. Por eso un agente debe ejecutar la desambiguación empleando para ello la ontología, asignar reglas de puntuación para facilitar la lectura y ayudar en el etiquetado de las palabras en el texto. La desambiguación culmina cuando el agente desarrolla el proceso de resolución de anáfora pronominal para dar cohesión al texto.

Existen muchos sistemas de resumen desarrollados desde la óptica de la Ciencia de la Computación, sin embargo en PuertoTex se mezclan criterios lingüísticos, comunicacionales, sociocognitivos y matemáticos.

Las prestaciones de PuertoTex son válidas para sus usuarios debido a que en su concepción, además de utilizar los procesos de ingeniería de *software*, se utilizaron técnicas de estudio de usuarios que facilitan que un investigador de la Rama de Ingeniería de Puertos y Costas no solo resuma un documento, sino que también obtenga información con grupos de personas que traban el mismo tema, o localice información en *Internet*, constituyendo una red social que realmente representa información y conocimiento.

Los resultados de la explotación del *software* son halagüeños si se tiene en cuenta que los resultados de los test de usabilidad son de elevada calidad y se corresponden con las prestaciones que ofrece el sistema.

La herramienta facilita la construcción de resúmenes tanto en español como en inglés, siempre y cuando sean textos que pertenezcan al dominio de la Ingeniería de Puertos y Costas. Los textos seleccionados para ambos idiomas facilitan la construcción de resúmenes indicativos e informativos.

La principal limitación de este estudio se centra en las propias estrategias cognitivas y reglas discursivas desarrolladas e implementadas a través de la herramienta, ya que solo operan en el dominio en cuestión y con artículos científicos. Otros textos, como, la poesía o la prosa, serían muy complejos de tratar mediante estas técnicas debido a que la retórica y el estilo del texto suelen variar mucho y no pueden determinarse regularidades para construir reglas de discurso.

Agradecimientos

Esta investigación se ha realizado dentro del proyecto RECORD: Representación del Conocimiento en Redes Dinámicas FFI2011-22397, financiado por el Ministerio de Ciencia e Innovación de España.

Referencias

AGUIRRE, E. *Formalization of concept-relatedness using ontologies*: applications in the construction of lexical knowledge bases, word sense disambiguation and automatic spelling correction. 1998. Tesis (Doctoral) - Universidad de País Vasco, San Sebastián, 1998.

D'CUNHA, I. *Hacia un modelo lingüístico de resumen automático de artículos médicos en español*. 2006. Tesis (Doctoral) - Universidad Pompeu Fabra, Barcelona, 2006.

DIJK, T.; KINTSCH, W. *Strategies of discourse comprehension*. Orlando: Academic, 1983.

DOMÍNGUEZ, S. *Calculus copora*. Santa Clara: Universidad Central Marta Abreu de las Villas, 2011.

ENDRES-NIGGEMEYER, B. Simsum: an empirically founded simulation of summarizing. *Information Processing and Management*, v.36, n.4, p.659-682, 2005.

ENDRES-NIGGEMEYER, B.; MAIRE, E.; SIGEL, A. How to implement a naturalistic model of abstracting: four core working steps of an expert abstractor. *Information Processing & Management*, v.31, n.5, p.631-674, 1995.

FABER, P.; MÁRQUEZ, C.; VEGA, M. Framing terminology: a process-oriented approach. *Meta*, v.50, n.4, p.189-213, 2005.

HERNÁNDEZ, A. *Indización y resumen*. La Habana: Universidad de la Habana, 2006.

HERNÁNDEZ, A. *Organización y representación del conocimiento*: paradigmas, hipertextos y fundamentación metamodélica. 2007. Tesis (Doctoral Bibliotecología y Ciencia de la Información) - Universidad de la Habana, La Habana, 2007a.

HERNÁNDEZ, A *Paradigmas dominantes de la representación de la información y el conocimiento*. La Habana: Universidad de la Habana, 2007b.

HERRERA, R. *Formatos de comunicación*. 2007. Disponible en: <http://fcom.uh.edu.cu>. Acceso en: 16 jul. 2011.

LEIVA, A. *et al*. An automat for the semantic processing of structured information. In: INTERNATIONAL CONFERENCE OF DESING OF SOFTWARE AND APLICACTIÓN, 9., 2009, Pisa, Italy. *Proceedings*… Pisa: IEEE, 2009. p.85-89.

LIN, C.; HOVY, E. Automatic evaluation of summaries using n--gram co-ocurrence statistic. In: HUMAN TECHNOLOGY-CONFERENCE, 2003, Edmonton, Canada. *Proceedings*… Edmonton, Canada: HLT-NAACL, 2003. p.71-78.

LUNH, H. The automatic creation of literature abstracts. *Journal of Research of Development*, v.2, n.2, p.59-165, 1958.

MANCHÓN, E. *Evaluación por criterios o heurística*. 2002. Disponible en: <http://www.ainda.info/evaluacion_heuristica.html>. Acceso en: 1 enero 2011.

MANN, W.; THOMPSON, S. *Rhetorical structure theory*: a theory of text organization. Los Angeles: Information Sciences Institute, 1990.

MATHIS, B.; RUSH, J.; YOUNG, C. Improvement of automatic abstracts by the use of structural analysis. *Jounal of the American Sciety for Information Science*, v.24, n.2, p.101-109, 1973.

METZELTIN, M. De la retórica al análisis del discurso. *Revista Electrónica de Estudios Filológicos*, n.6, 2003. Disponible en: <http://www.tonosdigital.es/ojs /index.php/tonos/article/view/506>. Acceso en: 1 marzo 2012.

NIELSEN, J. Heuristic evaluation. In: NIELSEN, J.; MACK, R. (Ed.). *Usability inspection methods*. New York: John Wiley 1994. p.25-52.

NIELSEN, J. *How to conduct a heuristic evaluation*. 2002a. Available from: <http://www.useit.com/papers/heuristic/heuristic_evaluation.html>. Cited: 26 Jan. 2011.

NIELSEN, J. *Ten usability heuristics*. 2002b. Available from: <http://www.useit.com/papers/heuristic/heuristic_list.html>. Cited: 21 Jan. 2011.

ONO, K.; SUMITA, K.; MIIKE, S. Abstract generation based on rhetorical structure extraction. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 15., 1994, Kyoto, Japan. *Proceedings*… Kyoto, Japan: DBLP, 1994.

PANECA, F. *La prensa remediana del siglo XIX*: reflejo de la cultura popular tradicional de la localidad. 2009. Tesis (Doctoral) - Facultad de Humanidades, Universidad Central Marta Abreu de las Villas, Santa Clara, 2009.

PILLEUX, M. Competencia comunicativa y análisis del discurso. *Estudios Filológicos*, n.36, p.143-152, 2001. Disponible en: <http://www.scielo.cl>. Acceso en: 1 marzo 2012.

PINTO, M. *El resumen documental*: principios y métodos. Madrid: Fundación Germán Sánchez Ruipérez, 2001.

PINTO, M. Interdisciplinary approaches to the concept and practice of written text documentary content analysis. *Journal of Documentation*, v.50, n.2, p.405-418, 2004.

RIGAU, G. *Resolución automática de la ambigüedad semántica de palabras*. Soria, España: Fundación Duques de Soria, 2002.

SALTON, G. On the application of syntactic methodologies in automatic text analysis. *Information Processing and Management*, v.26, p.73-92, 1996.

SÁNCHEZ-CUADRADO, S. *et al*. Definición de una metodología para la construcción de sistemas de organización del conocimiento a partir de un corpus documental en lenguaje natural. *Procesamiento del Lenguaje Natural*, n.39, p.213-220, 2007.

SENSO, J. *Representación del conocimiento en la ingeniería de puertos y costas*: proyecto investigador. Granada, España: Universidad de Granada, 2009.

SIDOROV, G.; ZAZUETA, O.O. Resolución de anáfora pronominal para el español usando el método de conocimiento limitado. In: ENCUENTRO NACIONAL DE CIENCIAS DE LA COMPUTACIÓN, 7., 2006, San Luís Potosi, Mexico. *Anales*… San Luís Potosi, Mexico: SMCC, 2006. p.276-281.

ANEXO

Documento original (parte superior) y resumen automático obtenido a partir de los datos procesados del documento original (parte inferior).

Documento original

Estudio de corrientes y dispersión en el puerto de Barcelona

M. Díez1, J.M. Redondo1, J. Vila[2]

1 Universidad Politécnica de Cataluña. Departamento de Física Aplicada, Jordi Girona 1-3 Barcelona 08034

2 Autoridad Portuaria de Barcelona, APB Medio Ambiente C/Circunvalación tramo VI Barcelona 08039

1. Introducción

Uno de los problemas medioambientales de los puertos es el vertido de contaminantes, en especial de hidrocarburos. Dado que la dispersión es un mecanismo que hace crecer exponencialmente el tamaño de las manchas y que el coste de su limpieza y daños también es función del tamaño es imprescindible un buen conocimiento del fenómeno que junto con un buen sistema de alerta y acción permita una rápida intervención.

2. Experimentos

En el Puerto de Barcelona se están llevando a cabo una serie de experimentos con el fin de estudiar el campo de corrientes y la capacidad dispersiva del medio. Están previstas campañas para cada estación climática, así como campañas especiales para analizar fenómenos hidrodinámicos como la difracción o estratificación. La duración de cada campaña es de dos semanas debido al gran dominio de trabajo.

Los experimentos consisten en el vertido de trazadores lagrangianos y la filmación de su comportamiento. Asimismo con un GPS se registran los puntos de vertido y recogida tomándose como puntos fiduciales en la restitución digital de las imágenes. Los trazadores son de dos tipos lagrangianos para medir corrientes, flotadores adecuadamente lastrados para evitar la interferencia del viento, y manchas de leche con fluoresceína por su elevado contraste y persistencia.

3. Resultados

A través del análisis digital de imágenes (Rodriguez *et al* 1997) se obtienen las trayectorias, direcciones y velocidades de la corriente y los coeficientes de dispersión. Por otro lado se analizan los agentes impulsores como el viento, marea, oleaje, corrientes etc.

Los primeros resultados muestran que la influencia de la marea o la pluma del Llobregat pueden desviar la dirección de la corriente respecto del viento significativamente, llegando incluso a oponerse a él. A continuación se muestra un ejemplo en el que se detecta un cambio en el régimen de la corriente de entrante a saliente (figura 2) que no responde estrictamente a una variación del viento. Si se observa la evolución de la marea (figura 3) parece coincidir con un cambio de vaciante a llenante.

Sin embargo, la escasa carrera de marea no hacía esperar un cambio de tendencia.

6. Referencias

Rodriguez A., Bahia E., Díez M., Sanchez-Arcilla A., Redondo J.M., Mestres M. Estudio

Experimental de Procesos de Mezcla en Aguas Costeras (Mayo 1997). Proc. IV Jornadas españolas de Ingeniería de costas y Puertos, Ed. J.C. Serra, pp. 643-656. ISBN 84-7721-573-1

Resumen

Uno de los problemas medioambientales de los puertos es el vertido de contaminantes, en especial de hidrocarburos. Dado que la dispersión es un mecanismo que hace crecer exponencialmente el tamaño de las manchas y que el coste de su limpieza y daños también es función del tamaño es imprescindible un buen conocimiento del fenómeno que junto con un buen sistema de alerta y acción permita una rápida intervención. En el Puerto de Barcelona se están llevando a cabo una serie de experimentos con el fin de estudiar el campo de corrientes y la capacidad dispersiva del medio. Están previstas campañas para cada estación climática, así como campañas especiales para analizar fenómenos hidrodinámicos como la difracción o estratificación. Los primeros resultados muestran que la influencia de la marea o la pluma del Llobregat pueden desviar la dirección de la corriente respecto del viento significativamente, llegando incluso a oponerse a él.