

Modelo conceptual de eventos meteorológicos para el proyecto

“Compartiendo la historia escondida del cambio climático en Latinoamérica a través de las TIC”.

GT3- Humanidades digitales

FIC - Instituto de Información

Diana Comesaña

diana.comesana@fic.edu.uy

Resumen

En el marco del proyecto “Compartiendo la historia escondida del cambio climático en Latinoamérica a través de las TIC”, se analizan más de 5.000 documentos digitalizados de las colecciones “Primeros Impresos Periódicos”, “Diarios” y “Semanarios”, correspondientes a los siglos XIX y XX.

Este proyecto, integrado por Colombia, Ecuador, México y Uruguay, tiene por objetivo compartir la historia del cambio climático en Latinoamérica a través de la interrelación y la interoperabilidad entre los sistemas de la región. La obtención de datos se realiza a través de un proceso de bibliominería.

Se realiza un análisis de dominio estudiando el lenguaje natural expresado en la prensa analizada y en los Glosarios de la Organización Meteorológica Mundial y de cada uno de los Institutos Meteorológicos Nacionales de los países intervinientes.

Con el fin de optimizar la recuperación de datos se desarrollan modelos semánticos asociados a los eventos meteorológicos recogidos en los periódicos.

Uruguay desarrolla su modelo ontológico a través del método methontology de desarrollo de ontologías y se implementa la misma con el software Protergé que actualmente soporta el lenguaje OWL 2.

La información recopilada se transformará para fomentar e incrementar los niveles de interoperabilidad conforme a los principios de Linked Data y estándares del W3C.

El producto final será plasmado por georreferenciación de los periódicos digitalizados.

Palabras-clave: *fenómenos meteorológicos, bibliominería, modelo ontológico, web semántica*

Introducción

El objetivo general de este proyecto es compartir la mirada social del cambio climático en Latinoamérica a través de la interrelación de las Tecnologías de la Información y las Comunicaciones (TIC) con la historia, geografía y cartografía.

Es un proyecto internacional e interdisciplinario, donde intervienen la Pontificia Universidad Javeriana (Bogotá - Colombia), la Universidad de Cuenca (Ecuador), la

Universidad de las Américas (Puebla - México) y la Universidad de la República de Uruguay, a través de las Facultades de Ingeniería e Información y Comunicación.

Se trata de recopilar y analizar la información disponible en las hemerotecas de las Bibliotecas Nacionales asociadas al proyecto para optimizar la recuperación y para la georreferenciación de estos datos, desarrollando modelos semánticos asociados con los eventos meteorológicos documentados en los periódicos, conforme a los estándares y buenas prácticas de las Infraestructuras de Datos Espaciales y de la Web Semántica.

El Proyecto

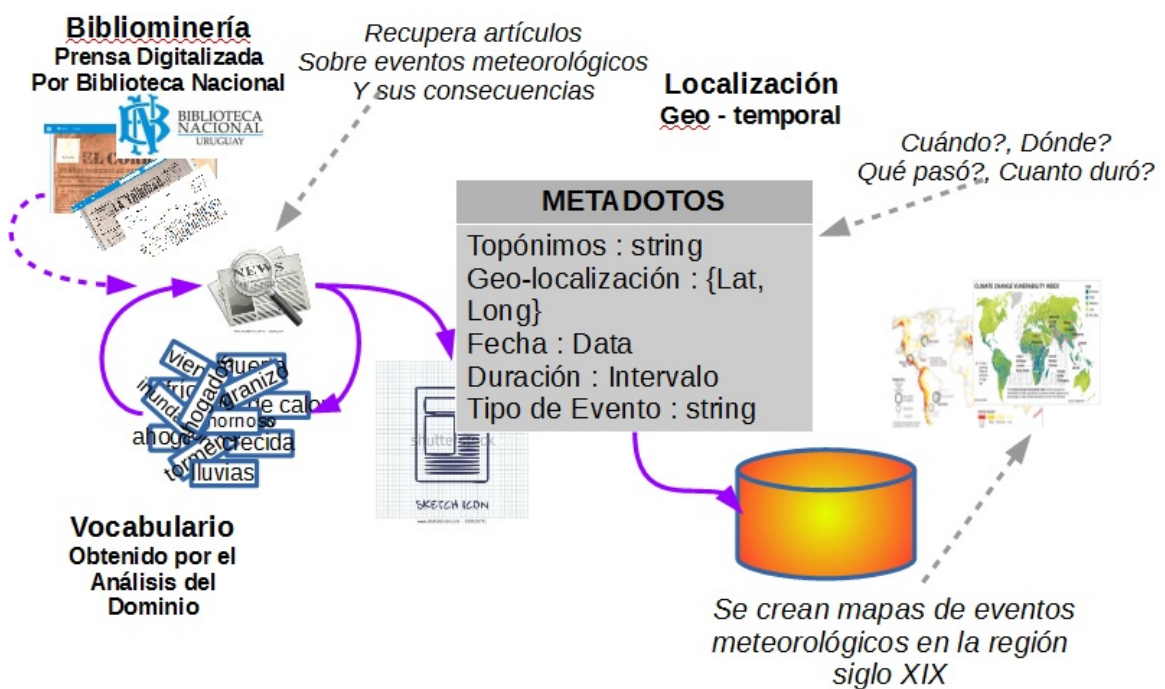


Figura 1: Diagrama del Proyecto

La Figura 1 nos muestra un diagrama de las etapas que comprende el proyecto.

La Obtención de los datos

La mayoría de nosotros hemos oído hablar de minería de datos o Data Mining como el conjunto de técnicas que consisten en la extracción no trivial de información que existe de manera implícita en grandes volúmenes de datos, información que hasta el momento era desconocida y que puede sernos útil para algún proceso.

Es una actividad que está a medio camino entre la Informática, la Estadística y la Documentación.

En bibliotecas, su práctica está poco extendida a pesar las múltiples aplicaciones que tiene, pero cuando la minería se aplica en las bibliotecas, se denomina Bibliominería.

Este término que deriva del inglés, *bibliomining* como procedente de *bibliometrics* (*bibliometría*) y *data mining* (*minería de datos*).

Es un término relativamente reciente. Si bien se utiliza desde la década de 1990, es citado como tal por primera vez por (Nicholson & Stanton, 2003) y nace por la necesidad de diferenciar o unificar los términos librería y minería de datos, ya que estos en su conjunto hacen referencia a un grupo de herramientas de software para la minería de datos.

El bibliomining se refiere al uso de estas técnicas que permiten sondear las enormes cantidades de datos generados por las bibliotecas automatizadas (Nicholson, 2003) y con estas técnicas se han efectuado estudios en los que se combinan el lenguaje natural y la minería de datos textual, es decir la minería aplicada a textos, para descubrir potenciales asociaciones desconocidas entre los documentos de colecciones digitales.

El proceso fundamental es el de la selección de los datos. Si el conjunto a utilizar no es el adecuado para el tipo de análisis que se pretende llevar a cabo, el éxito del proceso de minería, se verá frustrado.

En nuestro caso, al indagar, con una mirada social, los efectos de los eventos meteorológicos en la sociedad, seleccionamos tres de las colecciones digitales de la prensa periódica de la Biblioteca Nacional.

Se analizaron más de 5.500 documentos digitalizados de las colecciones “Primeros Impresos Periódicos”, “Diarios” y “Semanarios”, correspondientes a los siglos XIX y XX.

Se marcó el límite en la década de 1950 porque en 1951 se creó la Organización Meteorológica Mundial y se asumió que, a partir de este momento era factible hallar datos cuantificados y análisis de los fenómenos meteorológicos.

Estudio del Dominio

Para llegar a un lenguaje que nos permitiera una mejor recuperación de datos, recurrimos a un análisis del dominio, basándonos en el paradigma de Análisis de Dominio de Birger Hjørland.

Este paradigma plantea estudiar los dominios del conocimientos como comunidades discursivas. Se toma en consideración el contexto psicosocial y sociolingüístico, de la sociología del conocimiento y de la ciencia.

Según Hjørland (2003), en cuanto al objeto de estudio de la ciencia, “*se estudian de las relaciones entre documentos, áreas de conocimiento y discursos en relación con las posibles perspectivas de acceso de las distintas comunidades de usuarios*”

Se consultó a expertos en el tema y se analizaron los Glosarios de la Organización Meteorológica Mundial y de los Institutos Nacionales Meteorológicos de cada uno de los países intervinientes, junto con el lenguaje natural reflejado en la prensa de la época estudiada.

Uno de los problemas a resolver fue que algunos centros poblados y parajes, habían variado su denominación en el transcurso de los ciento cincuenta años analizados, lo que

llevó a bucear en la historia de los pueblos para lograr su ubicación geográfica. En el Uruguay, recurrimos al Diccionario Geográfico de Orestes Araujo, en su versión de 1912.

ESTUDIO DEL DOMINIO

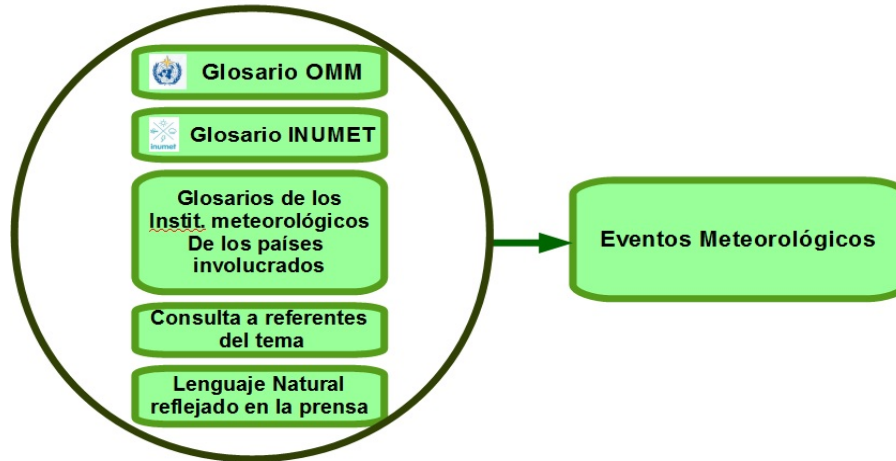


Figura 2: Fuentes para el Análisis del Dominio

La figura 2 nos muestra el conjunto de fuentes consultadas para el análisis del dominio.

Los modelos semánticos y el modelo ontológico uruguayo

Las nuevas tecnologías y conceptos que conforman la web semántica nos permiten describir formalmente los conceptos y sus relaciones mediante lenguajes especializados evitando ambigüedades conceptuales.

Para Edwin F. Caldón (Caldón y otros, 2014) *“Desde el punto de vista semántico, el principal problema es que la mayoría de la información que se comparte en plataformas de redes sociales es entregada en documentos en texto plano, lo cual implica que no hayan metadatos o información adicional que ayude a los buscadores web a procesar la información y sugerir a los usuarios el contenido más relevante”*. Esto es aplicable a la prensa digitalizada.

Los desarrollos ontológicos son los modelos por los que se ha optado en el desarrollo del proyecto, con el fin de agregar semántica, desde el punto de vista informático, al sector.

Se desarrollan en lenguaje owl, y, en nuestro caso se utilizó el método de desarrollo de ontologías Methontology y el software Protege 5.0.0. para la implementación.

Este método comprende el desarrollo de un glosario, donde se reúnen términos, sinónimos, descripción y tipo a que pertenecen.

La siguiente tabla muestra un sector de éste documento, donde se da el nombre del término estudiado, los sinónimos hallados, su descripción en lenguaje natural y a que tipo pertenecen (concepto, clase, atributo o propiedad)

TERMINO PREFERENTE	Sinónimos en los documentos de base	DESCRIPCION	TIPO
BRISA DE VALLE		Es el viento húmedo y cálido que se eleva por una ladera y que a su paso se condensa provocando la formación de nubes de tipo lenticular en la cima	Concepto
BRISA DEBIL	FLOJO	El mar tiene pequeñas olas con crestas rompientes, en tierra se agitan las hojas de los árboles y ondean las banderas	Concepto
BRISA FRESCA	FRESQUITO	Olas medianas y alargadas, borreguillos muy abundantes. En tierra hay pequeños movimientos en los árboles y se ondula la superficie de los lagos	Concepto

Tabla 1: Construcción del Glosario

Dice Lizalde (s. d.) *“En la idea de organizar el conocimiento en estructuras, subyace la de explotar las regularidades presentes para crear abstracciones. Las abstracciones son una colección de propiedades compartidas por los miembros de una categoría, un conjunto o de una clase”...* *“Cuando las abstracciones están organizadas empleando relaciones de inclusión se crea una estructura de árbol y el resultado es una taxonomía jerárquica”*

Una vez confeccionado el glosario construimos la taxonomía o árbol terminológico del dominio, donde se establece la jerarquía de los conceptos, seleccionándolos del glosario previamente construido.

Luego se construye el Diccionario de conceptos y se incluye la clase nominal “efectos_detectados”, que nos permite clasificar los eventos meteorológicos según las transformaciones que producen en tierra o en el mar.

La Figura 3 nos muestra una representación gráfica del modelo obtenido:

Representación Gráfica de un sector del Modelo

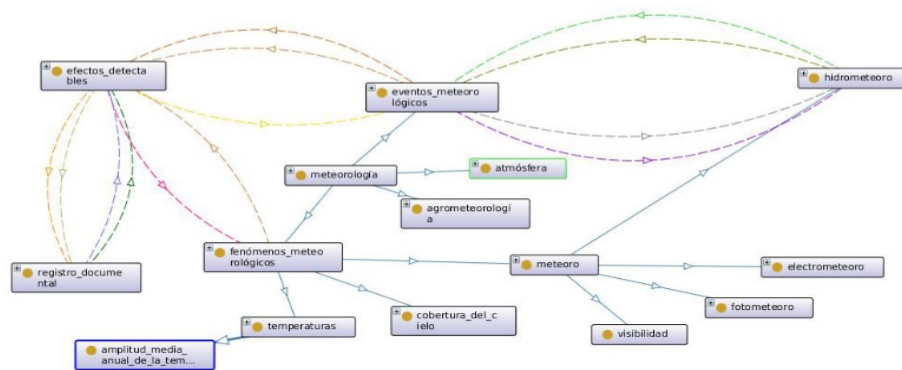


Figura 3: Representación gráfica del modelo semántico para los eventos meteorológicos

A instancias de las recomendaciones de expertos se usó la Escala Anemométrica de Beauford que nos indica los efectos detectados tanto en tierra como en el mar según el rango de velocidad del viento y empleamos especialmente la escala publicada por el Servicio de Oceanografía Hidrografía y Meteorología de la Armada de Uruguay (SOHMA), que combina ésta con la Escala de Douglas.

(Disponible en: <http://www.armada.gub.uy/Pagina/institucion/dimat/sohma.html#1>)

La escala de Beauford constituye una medida empírica para la velocidad del viento y está basada principalmente en la fuerza del mismo y el estado del mar y sus olas. Fue creada por Sir Francis Beaufort, oficial naval e hidrógrafo irlandés alrededor de 1805 y es ampliamente utilizada por los institutos meteorológicos de los países involucrados en el proyecto.

Por otra parte la Escala de Douglas es una escala que clasifica los diferentes estados del mar en 10 grados tomando como referencia el tamaño de las olas y fue creada por el vicealmirante inglés Henry Percy Douglas en 1917 cuando dirigía el Servicio Meteorológico de la Armada Británica.

Se establecieron las relaciones binarias “registrado_en” cierto documento digital y “produce”, determinado efecto en el ambiente, junto con sus inversas: “documenta” y “producido_por”

Las tablas de Atributos de Clase de atributos de Instancia describirán todos los atributos de las clases e instancias incluidos en el diccionario de conceptos. Para cada atributo se indicarán: nombre del atributo, concepto donde el atributo se define y tipo de valor, rango y cardinalidad.

Reflexión final

Una de las ventajas del desarrollo de ontologías es que son reutilizables y las hay que son estándares internacionales, asociadas con Open Geospatial Consortium y World Wide Web

Consortium (W3C). Estas ontologías se toman para conformar la red que permite una explotación eficiente de la información contenida en los documentos digitalizados.

En el grupo de trabajo se acordó por una parte, extender el glosario de términos construido para recoger todos los glosarios y correspondencias en los países que forman parte de este proyecto.

Una de las observaciones, al estudiar los glosarios de los diferentes institutos meteorológicos, es que no existe, en este campo grandes diferencias terminológicas ni por factores geográficos, ni por el lapso de tiempo transcurrido.

Era de gran importancia obtener datos cuantitativos de los fenómenos registrados, por lo que se recurrió a las Escala Anemométrica de Beauford y Douglas.

El tratamiento dado a la información original recopilada para mejorar la interoperabilidad y la generación de servicios geospaciales permitirá una visión de conjunto para el estudio de los efectos climáticos en la población.

Bibliografía

Hjorland, B. *Epistemology and Sociocognitive Perspective in Information Science*. JASIS. 2002;53(4):257-70.

Lizalde, E. (s. d.). *Representación del conocimiento*. Disponible en:
<http://rodrigorodriguez.files.wordpress.com/2010/03/conocimiento-intro.pdf>

Nicholson, S. (2003) *The Bibliomining Process: Data Warehousing and Data Mining for Library Decision-Making*. Information Technology and Libraries 22 (4)

Nicholson, S. & Stanton, J. (2003). *Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries*. Nemati, H. & Barko, C. (Eds.). Organizational data mining: Leveraging enterprise data resources for optimal performance. Hershey, PA: Idea Group Publishing.