

# Creación de un modelo semántico para modelar los fenómenos meteorológicos de los siglos XIX y XX en Latinoamérica

Diana Comesaña\*  
Bryan Barreiro\*\*

*Recibido 22 de marzo de 2018, aceptado 18 de agosto de 2018*

## Resumen

Los fenómenos meteorológicos son eventos que se desarrollan en la atmósfera y están asociados a un lugar específico. El objetivo del proyecto “Compartiendo la historia escondida del cambio climático en América Latina a través de las TIC”, es recuperar y georreferenciar las noticias sobre eventos meteorológicos registrados en la prensa escrita de los siglos XIX y primer mitad del XX, de los países involucrados en el proyecto: Colombia, Ecuador, México y Uruguay, y desarrollar modelos semánticos, transformando la información original recopilada para incrementar la interoperabilidad conforme a los principios de Linked Data y estándares del W3C. La terminología a modelar se obtiene por un análisis del dominio basado en el paradigma de Análisis de Dominio de Birger Hjørland. Se estudia la prensa digitalizada por las Bibliotecas Nacionales de los países intervinientes y el lenguaje empleado por sus instituciones meteorológicas nacionales y por la Organización Meteorológica Mundial. El análisis muestra un lenguaje uniforme, tanto respecto a la distribución geográfica, como temporal, con poca sinonimia y escasa polisemia. Se aplica el método METHONTOLOGY para desarrollar una ontología de Eventos\_Meteorológicos en el lenguaje OWL. Posibilita recuperar información no visible de las Bibliotecas Nacionales con técnicas de recuperación de información. Sus desarrollos son aplicables a otros documentos digitalizados y otras disciplinas.

\* Universidad de la República – FIC, San Salvador 1944, 11200 Montevideo, correo electrónico: [diana.comesana@fic.edu.uy](mailto:diana.comesana@fic.edu.uy)

\*\* Centro Regional de Profesores del Centro, Independencia y 24 de Abril, Ciudad de Florencio Varela, correo electrónico: [bryanalexander03@gmail.com](mailto:bryanalexander03@gmail.com)

Palabras clave: *Web Semántica, Semántica Geoespacial, recuperación de la información, ontología, meteorología, cambio climático.*

## Resumo

Os fenômenos meteorológicos são eventos que se desenvolvem na atmosfera e estão associados a um lugar específico. O objetivo do projeto “Compartilhando a história escondida da mudança climática na América Latina através das TIC”, consiste em recuperar e georreferenciar as notícias sobre eventos meteorológicos registrados na imprensa escrita dos séculos XIX e primeira metade do século XX, dos países envolvidos no projeto: Colômbia, Equador, México e Uruguai. Trata de desenvolver modelos semânticos, transformando a informação original recompilada para incrementar a interoperabilidade conforme os princípios de Linked Data e padrões do W3C. A terminologia a ser elaborada é obtida através de uma análise do domínio baseado no paradigma de Análise de Domínio de Birger Hjørland. Se estuda a imprensa digitalizada das Bibliotecas Nacionais dos países intervenientes e a linguagem empregada por suas instituições meteorológicas nacionais e Organização Meteorológica Mundial. A análise mostra uma linguagem uniforme, tanto no que diz respeito a distribuição geográfica, como temporal, com pouca sinonímia e baixa polissemia. Se aplica o método METHONTOLOGY para desenvolver uma ontologia de Eventos\_Meteorológicos na linguagem OWL. Possibilita recuperar informação não visível das Bibliotecas Nacionais com técnicas de recuperação de informação. Seus desenvolvimentos são aplicáveis a outros documentos digitalizados e outras disciplinas.

Palavras chave: *Web Semântica, Semântica Geoespacial, recuperação da informação, ontologia, meteorologia, mudança climática.*

## Abstract

The meteorological phenomena are events that develop in the atmosphere and are associated to a specific place. The objective of the project “Sharing the hidden history of climate change in Latin America through ICT” is to retrieve and georeference the news about meteorological events recorded in the written press of the 19th through the first half of the 20th century, of the countries involved in the project: Colombia, Ecuador, Mexico and Uruguay and develop semantic models, transforming the original information collected to increase interoperability according to the principles of Linked Data and W3C standards. The terminology to be modeled is obtained by an analysis of the domain based on the paradigm of Domain Analysis of Birger Hjørland. The digitized press by the National Libraries of the countries involved and the language used by their national meteorological institutions and the World Meteorological Organization is examined. The analysis shows a uniform use of language, in terms

of geographical distribution, as temporary, with little synonymy and only one case of apparent polysemy. The METHONTOLOGY method is applied to develop the ontology of Eventos\_Meteorologicos. It makes it possible to retrieve non-visible information from National Libraries with information retrieval techniques and natural language processing. Its developments are applicable to other digitized documents from other disciplines.

Key words: *Semantic Web, Geospatial Semantics, Information retrieval, ontology, meteorology, climate change.*

## Introducción

En la década de los sesenta, Salton estableció las bases de lo que hoy conocemos como Recuperación de la Información (RI). El surgimiento de la web en los años 90, le aportó una dimensión nunca imaginada, afectando a millones de personas. La información que se encuentra disponible es inagotable, pero no siempre es fácil llegar a ella. El problema central de la RI clásica era poder determinar la relevancia de un documento ante cualquier consulta.

La web actual es lo que se denomina la web de documentos, es decir una web integrada por páginas web o documentos creados para ser interpretados por los seres humanos y es una web en transición a lo que se denomina Web Semántica.

Según el World Wide Web Consortium (W3C. 2017), “la Web Semántica es una Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante”.

Aspira a la recuperación automática de datos, a la interoperabilidad de entre comunidades y dominios y al razonamiento automático.

La web semántica está relacionada con la inteligencia artificial ya que busca que las computadoras u ordenadores “comprendan” la consulta que hace un usuario, con el fin de devolver un conjunto de resultados acordes a la misma.

Como dice Regillo (2018):

La web semántica es una extensión de la Web a través de una serie de estándares definidos por el W3C con el fin de dotar de significado a los datos. Se pretende que estos datos resulten procesables y entendibles por un sistema mediante una serie de anotaciones semánticas en formato XML, fundamentado en el concepto de tripla (sujeto, predicado, objeto) como estructura predicativa para exponer un hecho o propiedad sobre un objeto.

La web semántica aspira a combatir la *infoxicación* o sobrecarga informativa y la frustración de los usuarios frente a la avalancha de información que se encuentra en

Internet, ahorrando tiempo al usuario y devolviendo resultados de búsquedas acordes a sus necesidades.

Un gran porcentaje de las preguntas habituales que aparecen en una consulta web presentan una componente espacial: ¿Dónde se ubica...?, ¿dónde trabajas?, ¿dónde está el hospital más cercano al lugar en que me encuentro?, ¿cuál es el límite Este del Uruguay?, ¿qué clima nos espera hoy en Montevideo? Son información geográfica: datos censales, división político-administrativa, redes de transportes, fotografías aéreas, imágenes satelitales, informes meteorológicos, etc.

Este tipo de información, con los datos distribuidos, servicios compartidos, tecnologías y arquitecturas a los que se agrega el reconocimiento de información geográfica, de asociaciones espaciales y de razonamiento espacial, constituye la Web Geoespacial. Si esta información geográfica se halla enriquecida con la incorporación de metadatos semánticos y el desarrollo de ontologías que faciliten la comprensión e interoperabilidad de información, estaremos frente a una web geoespacial semántica, donde los metadatos son estándares para describir la información geográfica y las ontologías, formas explícitas y lógicas de expresar el lenguaje.

Es el escenario que estudia el proyecto IPGH N° HIST 02 2017 “Compartiendo la historia escondida del cambio climático en Latinoamérica a través de las TIC”, cuyo objetivo es obtener modelos semánticos asociados a los eventos meteorológicos recogidos en la prensa escrita de los siglos XIX y primera mitad del XX. Este proyecto fue auspiciado por el Instituto Panamericano de Geografía e Historia (IPGH), enmarcado en el contexto de los Proyectos Panamericanos de Asistencia Técnica (PAT), 2017. En él participaron investigadores de Colombia, Ecuador, México y Uruguay y se alinea con los objetivos de la Agenda 2030 de Desarrollo Sostenible del Programa de las Naciones Unidas para el Desarrollo, que en su punto N°13 “Acciones por el clima”, destaca: “Las pérdidas anuales promedio causadas solo por tsunamis, ciclones tropicales e inundaciones alcanzan los cientos de miles de millones de dólares y exigen inversiones de unos US\$ 6.000 millones anuales solo en gestión de los riesgos de desastres” (PNUD, 2018).

Todos los países sufren los efectos del cambio climático, y Latinoamérica constituye un caso particular: mientras que los países del llamado “Primer Mundo” generan la mayor parte de emisiones de gases de efecto invernadero y sufren sus consecuencias en cuanto al aumento de intensidad y frecuencia de los eventos meteorológicos extremos, la región de América Latina contribuye poco y sufre de manera desproporcionada.

Es una de las regiones más vulnerables al localizarse dentro de la franja de huracanes y tener numerosos Estados insulares y zonas costeras bajas. Depende de los deshielos andinos para el suministro de agua a sectores urbanos y agrícolas y está expuesta a inundaciones e incendios forestales.

Estudiar el comportamiento de los eventos meteorológicos del pasado, lleva a entender mejor el presente, con el fin de poder diseñar acciones tendientes a minimizar las consecuencias del cambio en el comportamiento del clima de la región.

### **Los fenómenos meteorológicos**

Los fenómenos meteorológicos son eventos que se desarrollan en la atmósfera y están asociados a un lugar específico, puede asignarles coordenadas y constituyen por ello información geográfica.

Según el Programa Regional de Meteorología (s.f.) se pueden definir como “aque- llos procesos permanentes de movimientos y de transformaciones que sufre la natu- raleza y que pueden influir en la vida humana” Comprenden: lluvias, vientos, fríos extremos, calor, nevadas, granizo, niebla y otros que, estudiados en su conjunto y en su variabilidad en un lapso de tiempo determinado, nos muestran indicios del clima y sus variaciones.

Es decir, son cambios que suceden en la naturaleza por sí mismos. Puede tratarse de un fenómeno tan habitual e inofensivo como una llovizna o de un fenómeno poco usual y catastrófico como un huracán o tornado. La ocurrencia o no de estos fenóme- nos está relacionada a ciertas zonas geográficas y épocas del año.

Los eventos o fenómenos meteorológicos están fuertemente relacionados con la información geoespacial, ya que los mismos son la materia prima para la construcción de mapas o atlas climáticos. Dichos fenómenos son las variables representadas en estos documentos cartográficos. Se pueden elaborar, por ejemplo, mapas de riesgo de inundaciones, mapas de huracanes y mapas de sequías entre otros.

Estos mapas climáticos sirven de apoyo para la investigación y el estudio en la prevención de catástrofes naturales o para la gestión de alternativas cuando los fenó- menos son inevitables.

El surgimiento de los Sistemas de Información Geográfica (SIG) ha facilitado enormemente la construcción de este tipo de cartografía. Según la *National Center for Geographic Information and Analysis*, citada por Feito y Segura (2010) un SIG es “un sistema de hardware, procedimientos y software elaborado para facilitar la ob- tención, gestión, manipulación, análisis, modelado y representación de datos espa- cialmente referenciados y para la resolución de problemas complejos que impliquen la manipulación y gestión de dichos datos”, por ende, capaz de producir cartografía y permitir avanzar en el estudio y análisis de los fenómenos geográficos. La climato- logía es una de las disciplinas que se beneficia del uso de estas técnicas. “Un SIG puede utilizar mapas analógicos, mapas digitales, fotografías aéreas, modelos digita- les de elevaciones (MDE), imágenes satelitales, registros de GPS para, posterior- mente, elaborar cartografía de gestión de riesgos” (Ribera Masgrau, 2004).

Por lo tanto, los fenómenos naturales o meteorológicos, tienen estrecha relación con la información geográfica y con la cartografía producida, especialmente con aquella destinada a la toma de decisiones para la gestión y prevención de desastres naturales.

Para Gallegos (2011)

Todo trabajo de investigación de tipo histórico se vincula con una multiplicidad de fuentes que conforman un eje desde el cual el historiador trae hacia el presente los hechos que se analizan. La labor del investigador radica en la selección de las metodologías adecuadas y la recopilación de bibliografía que dé cuenta, por un lado, del estado de la cuestión, y por otro del contexto de estudio.

La Web se ha transformado en un banco infinito de datos, y el desafío lo constituye el poder descubrir patrones que permitan la exploración automática de los mismos. Es lo que se denomina “minería de datos”, una disciplina que reúne la estadística, la informática y la documentación, empleando métodos de inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. Si se aplican esos métodos y herramientas a los fondos de una biblioteca, entonces nos hallamos frente a una *bibliominería*.

El éxito o fracaso de la bibliominería depende de la pertinencia de las fuentes de información seleccionadas. Las características de la prensa hispanoamericana del siglo XIX, la transforman en una inmejorable fuente de información para el estudio de los eventos de este siglo.

Como dicen Mejías y Arias (1998), para América Latina, “El siglo XIX es, por excelencia, el siglo de la prensa escrita”. Muchas personalidades de la cultura y de diferentes ámbitos científicos y profesionales, estimaban de prestigio colaborar en publicaciones periódicas.

Emilio Castelar, citado por Mejías y Arias (1998), dijo: “Cuando tomo en mis manos un periódico, cuando recorro sus columnas, cuando considero la diversidad de sus materias y la riqueza de sus noticias, no puedo menos de sentir un raptó de orgullo por mi siglo...” y es que la prensa hispanoamericana estaba totalmente comprometida con su labor educadora y de difusión, de allí la variedad e importancia de los datos que recoge.

La ventaja que ofrece el análisis de las publicaciones periódicas, en particular los periódicos, es que los mismos documentan hechos cotidianos del día a día y permite rescatarlos del olvido para las generaciones futuras.

Los periódicos registran los distintos fenómenos meteorológicos ocurridos y sus consecuencias: lluvias, heladas, sequías, etc., que conforman fuentes documentales para el estudio del cambio climático en determinada región o país.

El límite cronológico se estableció en 1950, pues en 1951 se creó la Organización Meteorológica Mundial y se asume que, a partir de este momento, es factible hallar datos cuantificados y análisis de los fenómenos meteorológicos.

### **Análisis del dominio y de las fuentes documentales**

Para llegar a un lenguaje inicial, que permitiera la recuperación automática o semi-automática de los datos buscados, se recurrió a un análisis del dominio basado en el paradigma de Análisis de Dominio de Birger Hjørland, donde, según sus palabras: “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” Birger Hjørland (2002).

Este paradigma plantea estudiar los dominios del conocimiento como comunidades discursivas. Se toma en consideración el contexto psicosocial y socio lingüístico de la sociología del conocimiento y de la ciencia. Implica que un dominio del conocimiento se estudia a través de sus representantes, inmersos en la comunidad de la que forman parte.

Así se consultó a expertos en meteorología y se estudiaron los documentos por ellos sugeridos. Se analizaron los glosarios meteorológicos empleados por las instituciones meteorológicas de cada país y de la Organización Meteorológica Mundial. Uno de los problemas que causaba más inquietud era poder cuantificar fenómenos que la prensa describía sin dimensionar, para poder llegar a una clasificación jerárquica de los fenómenos meteorológicos registrados. Por consejo de los expertos asesores, se emplearon las escalas Anemométrica de Beauford, que 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, y la Escala de Douglas que proporciona la altura de las mismas.

Los documentos analizados definen terminología que está en uso y es común a todos los países, permiten establecer un orden jerárquico en los términos y lo interesante es que surgen en la época que se estudia, por lo que no hay contradicciones con el lenguaje reflejado por los periódicos analizados.

Una de las observaciones de este análisis permitió determinar que, en esta disciplina, no existe gran variación terminológica, ni por los factores geográficos, ni por la coordenada temporal. Se identificaron unos pocos sinónimos y un único caso de aparente polisemia, aunque un estudio más profundo, determinó que uno de los significados era un caso específico del otro, el cual correspondía a un concepto algo más general.

El lenguaje natural en que se redactaron las noticias, aportaron sinónimos y términos relacionados con la terminología específica del área de conocimiento tratada, principalmente en cuanto a los efectos detectables de un fenómeno meteorológico.

## Desarrollo de una ontología de eventos meteorológicos

Desde el punto de vista semántico, el principal problema es que la mayoría de la información que se comparte es entregada en documentos en texto plano, lo cual implica que no haya metadatos o información adicional que ayude a los buscadores web a procesar la información y sugerir a los usuarios el contenido más relevante (Caldón *et al.*, 2010).

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

¿Qué significa esto? Es necesario hallar la forma de proporcionar la semántica más fuerte, para que los equipos puedan realizar sus inferencias.

El estudio de las ontologías disponibles mostró la falta de una que presentara suficiente especificidad para abarcar los términos correspondientes a los fenómenos meteorológicos en toda su complejidad.

Entre otras ontologías existentes se revisaron las ontologías *Falcons search*, especialmente donde la clase *meteorological\_event* es sub-clase de *weather\_events* y está disponible en <http://umbel.org/reference-concept/?uri=ImmediateWeatherProcess>.

Se verificó que no desarrolla los términos más allá de presentar la clase *meteorological\_event*. Se estudió el conjunto de ontologías “objetos\_geográficos”, desarrollados en la tesis de maestría “Modelo Conceptual de Información Geográfica para la IDE —Uruguay”, destinadas a conformar un meta-modelo mediante la extensión de lógica descriptiva ALCQM, y que prevé el desarrollo de una clase información\_meteorológica, en la ontología “Usos\_especiales”, (Comesaña, 2015).

Se decidió crear una ontología nueva, que, partiendo del término “eventos\_meteorológicos”, se adaptara a los dos desarrollos anteriores e incluyera la terminología contenida en los glosarios de los institutos meteorológicos de Colombia, Ecuador, México, Uruguay, la Organización Meteorológica Mundial y las tablas Anemométrica de Beauford y de Douglas.

La ontología de eventos meteorológicos, se desarrolló en lenguaje OWL, a través del método de desarrollo de ontologías METHONTOLOGY y se implementó con el software Protege 5.2.0.

Si bien se conocen nuevos métodos como NeOn, para construir ontologías, atendiendo a la interdisciplinariedad del grupo, donde trabajaron Documentalistas, se optó por el empleo de METHONTOLOGY, por hallarse éstos más familiarizados con su uso.

En lugar de crear el meta-modelo con la extensión ALCQM, se empleó la metodología NeOn para construir una red de ontologías sumando a la desarrollada de



“eventos\_meteorológicos”, aquellas que cubren otras necesidades del proyecto, por ejemplo, en cuanto a nombres geográficos y fechas.

“Eventos\_meteorológicos”, está disponible registrándose en la comunidad GNOSS, en el grupo “Interoperabilidad Semántica”, donde fue publicada bajo licencia Creative Commons Attribution 3.0, en el sitio: <<https://my.gnoss.com/comunidad/mygnoss/recurso/eventosmeteorologicos/cff52099-003f-4395-be54-4a528fc3553f>>.

Este desarrollo comprendió la construcción de un glosario. La Tabla 1, muestra un sector del mismo donde se indica nombre del término, sinónimos, descripción en lenguaje natural y tipo (clase, concepto, propiedad o atributo).

**Tabla 1**  
**Construcción del Glosario**

<i>Nombre</i>	<i>Sinónimo</i>	<i>Descripción</i>	<i>Tipo</i>
<i>Agua Nieve</i>	Cellisca	Tipo de precipitación en la que el agua presenta dos estados, teniéndose una mezcla de agua congelada y agua líquida	Concepto
<i>Eventos meteorológicos</i>	Fenómenos Meteorológicos	Los <i>fenómenos meteorológicos</i> incluye todos los acontecimientos que se pueden producir ligados a la dinámica de la atmósfera.	Clase
<i>Meteoro</i>		Fenómeno además de las nubes, que es observado en la atmósfera o en la superficie del globo terrestre. Los meteoros, teniendo en cuenta la naturaleza de sus partículas constitutivas o los procesos físicos que intervienen en su formación, se han clasificado en cuatro grupos principales: hidrometeoro, litometeoro, fotometeoro y electrometeoro	Subclase
<i>Velocidad del viento</i>		Aire en movimiento, medido en Km/h	Atributo

En esta Tabla 1, la diferencia entre “clase” y “concepto”, está dada porque “clase” es un término que reúne jerárquicamente un conjunto de términos-concepto simples

relacionados entre sí por cierta característica. Por ejemplo, la Figura 1, se muestra una clase “meteoros”, subdividida en Subclases “electro-meteoros”, “fotometeoros”, “hidrometeoros”, etc., según el tipo de evento meteorológico de que se trate. La subclase “electrometeoros” reúne los conceptos: “trueno”, “fuego\_de\_San\_Telmo”, “centella”, “relámpago”, “rayo” y “aurora\_polar”, términos simples o elementales.

Con base en este glosario se construye una taxonomía o árbol terminológico del dominio, en que se establece la jerarquía de los conceptos (Figura 1).



**Figura 1 .** Sector del árbol terminológico, tal como se desarrolla a través del software Protégé 5.0.0.

La jerarquía mostrada en el árbol terminológico surge del estudio de los glosarios y documentos sugeridos por los expertos en el análisis del dominio. Se separaron aquellos términos que corresponden a la climatología, de los de meteorología, y dentro de éstos, los que corresponden a los diferentes conceptos generales. Los fenómenos meteorológicos se clasificaron de acuerdo al agente que los produce.

Es necesario crear una clase nominal “efectos\_detectables”, que constituye parte importante del cuerpo de las noticias que se analizan y que permite identificar y cuantificar los eventos meteorológicos documentados. También se crea la clase “registro\_documental”, cuyas instancias recogen el corpus de documentos con que se trabaja.

Se establecen las relaciones binarias “documenta” y su inversa “documentado\_en” para relacionar los “efectos\_detectables” con el “registro\_documental”; “produce” y “producido\_por” relacionan los eventos meteorológicos con sus efectos, “registra” y “registrado\_en” indica en que registro documental se recoge cierto evento meteorológico, y “publica” y “publicado\_en” une los “avisos\_meteorológicos” con su “registro\_documental”. Las relaciones “localiza” y “ubicado\_en” dejan la “puerta abierta” para enlazar el evento a un lugar geográfico determinado y “ocurre” y “sucede\_el” con su fecha.

El árbol terminológico, lleva a obtener un Diccionario de conceptos que muestra la Tabla 2, donde se registra el nombre del concepto, las instancias correspondientes, propios de los atributos de clase, de las instancias y las relaciones que los unen a los documentos en que están registrados y los efectos que producen.

**Tabla 2**  
**Diccionario de Conceptos**

<i>Nombre del concepto</i>	<i>Instancias</i>	<i>Atributos de clase</i>	<i>Atributos de instancia</i>	<i>Relaciones</i>
<i>Viento</i>	Pampero	Velocidad del viento	Efectos detectables	Produce
	Sudestada			Documentado_en
<i>Niebla</i>	—	Visibilidad	—	Produce
				Documentado_en

Para los Atributos de Clase e Instancia, se construyen tablas que nos muestran el nombre del atributo, concepto a que se relaciona, tipo de valor, rango y cardinalidad (Tablas 3 y 4).

Los atributos de clase surgen de la aplicación de las definiciones de los glosarios y de las tablas de Beauford y Douglas mencionadas.

**Tabla 3**  
**Atributos de clase**

<i>Atributo de clase</i>	<i>Concepto</i>	<i>Tipo de valor</i>	<i>Rango</i>	<i>Cardinalidad</i>
<i>velocidad_del_viento</i>	Viento	No negativo (en km/h)	{0,n}	(1,n)
<i>Altura_de_la_ola</i>	Tormenta	No negativo (en mts)	{0,n}	(1,n)

**Tabla 4**  
**Atributos de Instancia**

<i>Atributo de Instancia</i>	<i>Concepto</i>	<i>Tipo de valor</i>	<i>Rango</i>	<i>Cardinalidad</i>
<i>efecto_detectado</i>	Tornado	cadena_de_caracteres	—	(1,n)

Las instancias se obtienen al rescatar las noticias de la prensa digitalizada en forma automática o semiautomática, y aún manual cuando el estado de la digitalización no lo permita de otra manera. Se obtiene el periódico que la registra el evento, la fecha en que aconteció, el lugar y los efectos que detectados.

Las definiciones lógicas de los conceptos se construyeron empleando la sintaxis Mánchester, que emplea el software Protégé.

Así puede decirse que existe una clase llamada eventos meteorológicos, la precipitación es una sub-clase de los hidrometeoros, que son una sub-clase de los meteoros y a su vez son una sub-clase de los eventos meteorológicos. La precipitación tiene asociada exactamente una medida decimal y produce cierto efecto detectable:

$\text{precipitación} \subseteq \text{hidrometeoro}$   
 $\text{hidrometeoro} \subseteq \text{meteoro}$   
 $\text{meteoro} \subseteq \text{eventos\_meteorologicos}$   
 $\text{precipitación} = n.R \text{ (decimal)}$   
 $\exists (\text{precipitación Produce. efecto\_detectable})$

La idea de aplicar estos modelos semánticos es poder transformar la información para incrementar los niveles de interoperabilidad según los principios de Linked Data y los estándares sugeridos por el World Wide Web Consortium (W3C).

Se desarrolló la plataforma LOD-GF que es un *framework* (plataforma de trabajo) basado en Pentaho Data Integration para brindar un entorno unificado de soporte en cada una de las fases de la metodología de publicación de Linked Open Data.

En este caso el desarrollo emplea plugins o complementos para el procesamiento especializado, tanto nativos del entorno, como desarrollados y que ofrecen un marco de trabajo gráfico flexible, de forma que los usuarios puedan convertir sus datos originarios de un amplio dominio de fuentes y formatos, a datos enlazados de calidad.

Comprende cinco fases: especificación, modelado, generación y publicación de los datos.

En el modelado debe identificarse, seleccionarse o generar vocabularios que permitan describir semánticamente los datos de las fuentes disponibles. En este punto interviene la ontología anteriormente desarrollada.

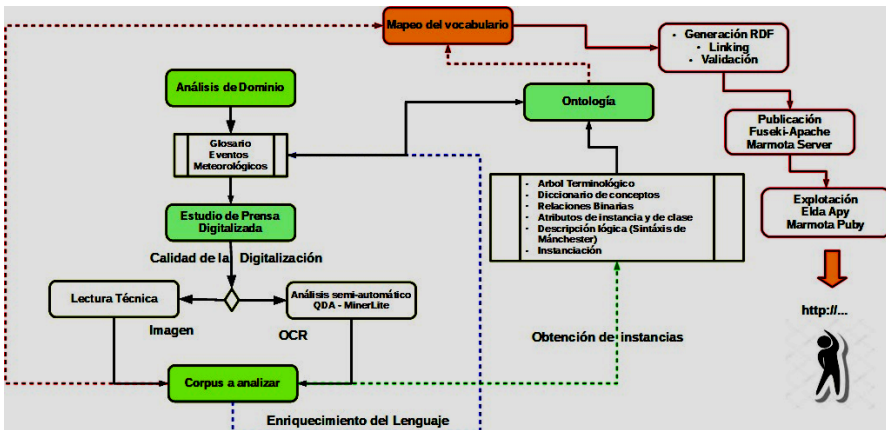


Figura 2. Procesos de desarrollo en el proyecto.

La Figura 2 muestra el proceso de trabajo del proyecto a partir del Análisis de Dominio. Es la etapa inicial del estudio de la prensa digitalizada con la obtención del vocabulario enriquecido para el desarrollo de la ontología, sus pasos y como ésta se inserta en el entorno de trabajo definido por la plataforma LOD-GF, hasta que el usuario recibe una respuesta del sistema a su consulta.

### Visualizando resultados

Para la publicación de los datos, se almacenan en un repositorio RDF o base de datos especialmente diseñada para el almacenamiento y recuperación de tripletas a través de consultas semánticas. Estas tripletas son entidades de datos compuestas de sujeto-predicado-objeto y disponen de un medio de acceso a los datos (SPARQL Endpoint).

En este caso es un *plugin* especializado conocido como Fuseki Loader, el cual permite configurar los parámetros básicos para el despliegue de las tripletas, es decir la información se rescata a través de consultas SPARQL en la plataforma de trabajo.

Los resultados pueden visualizarse también desde la propia ontología desarrollada en Protégé. Al desarrollar la clase “registro\_documental”, cuyas instancias las conforman el corpus de prensa digitalizada con que se trabaja, establecido las relaciones binarias “documentado\_en” y su relación inversa “documenta”, es posible consultar, empleando consultas SPARQL, por ejemplo, que documentos que registran eventos meteorológicos a partir de sus efectos detectables:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
SELECT ?efectos_detectables ?registro_documental
WHERE { ?efectos_detectables
<http://www.semanticweb.org/diana/ontologies/2017/9/untitled-ontology-197#
OWLObjectProperty_6e8e9440_5505_4ee8_8a60_22cee5297c9d>
?registro_documental }
```

La respuesta del sistema corresponde a las instancias ingresadas, y su resultado se observa en la copia de la descarga de pantalla de la figura 3. En ella se observa que en los documentos identificados por los números “23888” y “23887”, se registran daños en los árboles. Si recurrimos a la Escala de Bauford, esto corresponde a un temporal fuerte con vientos de 75 a 88 km/h.

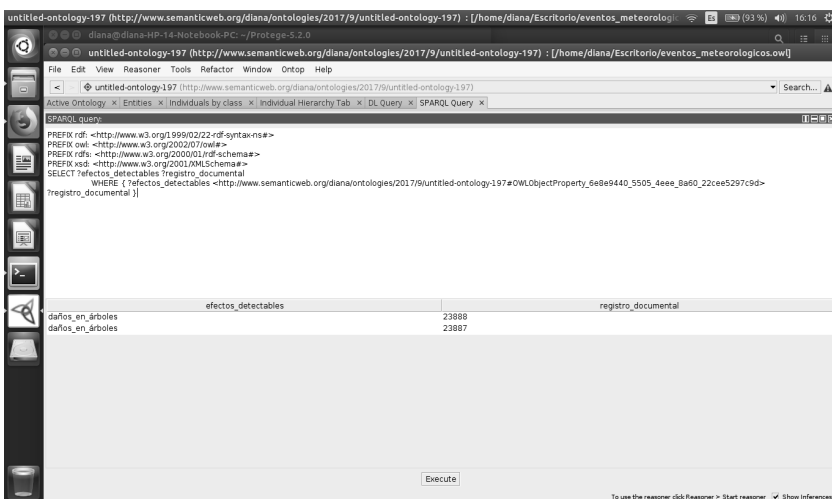


Figura 3. Resultado de la consulta SPARQL a la ontología.

## Conclusiones

El proyecto de investigación cuyo caso se presenta, mostró la fortaleza del trabajo multidisciplinar, al reunir especialistas de diferentes disciplinas: informática, ciencias de la información y geografía.

Rescatar información de la prensa digitalizada del siglo XIX, significó un gran desafío. Es necesario enfrentarse a documentos digitalizados a partir de originales con diferentes grados de deterioro físico y diversas condiciones de digitalización, como imagen, con software de reconocimiento de caracteres (OCR), baja resolución, lo que condujo a que no siempre se pudieran procesar los documentos en forma automática y fuera necesario recurrir al procesamiento automático y aún a la lectura técnica de los mismos.

Sería importante que las instituciones, al enfrentar un proyecto de esta envergadura, tuviesen en cuenta una estandarización en la digitalización, de forma que la calidad de su producto, permitiera un posterior procesamiento automático.

Otra dificultad se presentó con la evolución de los nombres geográficos, para solucionarlo se consultó diccionarios de nombres geográficos y *gazetteers* disponibles.

Los eventos meteorológicos que permite rescatar constituyen antecedentes históricos del cambio climático en América Latina. Nos es posible recuperar información que no estaba visible de las Bibliotecas Nacionales utilizando los beneficios de las Tecnologías de la Información y las Comunicaciones (TIC), con técnicas de recuperación de información.

La localización de la información se logra a través de consultas SPARQL en el entorno de trabajo desarrollado. El prototipo desarrollado para la consulta de artículos sobre eventos meteorológicos documentados en la prensa escrita digitalizada, es aplicable a otros documentos digitalizados y, desarrollando los modelos semánticos correspondientes, a infinidad de disciplinas, cuyos datos, por haber sido considerados menores en su momento, aún aguardan ser descubiertos en el volumen siempre creciente de datos volcados a la web.

## Bibliografía

- Caldón, E.F.; Uribe, G.; López, D.M.; de Oliveira, J.P.M. y Krug Wives, L. (2010). *Mecanismos de anotación semántica de contenidos en plataformas de redes sociales*, CI. 2010 5(1), 90-99.
- Comesaña, D. (2015). "Modelo conceptual de información geográfica para la IDE – Uruguay", tesis de maestría, Universidad de la República (Uruguay), Facultad de Información y Comunicación, disponible en: <<https://www.colibri.udelar.edu.uy/jspui/handle/123456789/5230>>.

- Fenómenos Meteorológicos (s.f.). Mendoza, Argentina: Programa Regional de Meteorología, disponible en: <<http://www.prmarg.org/fenomenos-meteorologicos>>.
- Feito Higuera, F. y Segura Sánchez, R. (2010). *Herramientas SIG 3D*. Virtual Archaeology Review, 1 (1), disponible en: <<https://polipapers.upv.es/index.php/var/issue/view/500>>.
- Gallegos, C. (2011). “Abordaje metodológico de prensa escrita: el semanario Cuba Libre”, *Latinoamérica*, núm. 53, julio-diciembre, México, disponible en: <[http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S1665-85742011000200006](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-85742011000200006)>.
- Hjørland, B. (2002). *Epistemology and Sociocognitive Perspective in Information Science*, JASIS, 53(4), 257-70.
- Mejías, A. y Arias, A. (1998). “La prensa del siglo XIX como medio de difusión de la literatura Hispanoamericana”, *Revista general de Información y Documentación*, 8(2), 241-257
- PNUD (2018). Objetivos del Desarrollo Sostenible. Objetivo 13: Acciones por el Clima. Disponible en: <http://www.undp.org/content/undp/es/home/sustainable-development-goals/goal-13-climate-action.html>
- Regillo, R. (2018). *Un vistazo al estado actual de la Web Semántica*. Disponible en: <https://www.beeva.com/beeva-view/innovacion/un-vistazo-al-estado-actual-de-la-web-semantica/>.
- Ribera Masgrau, L. (2004). *Los mapas de riesgos de inundaciones: representación de la vulnerabilidad y aportación de las innovaciones tecnológicas*, *Documents D'Analisi Geografica*, 43, 153-171. Disponible en <<http://www.raco.cat/index.php/DocumentsAnalisi/article/viewFile/31812/31647>>.
- World Wide Web Consortium (2017). *Guía breve de Web Semántica*, disponible en <<https://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>>.