

# Creación y Anotación de Páginas Web usando CREAM<sup>1</sup>

Sigfried Handschuh

*Instituto AIFB, Universidad de Karlsruhe, Alemania*

Steffen Staab

*Instituto AIFB, Universidad de Karlsruhe, Alemania*

*Learning Lab Lower Saxony, Hannover, Alemania*

*Ontoprise, Karlsruhe, Alemania*

Resumen redactado por:

**Cuauhtémoc Rivera Loaiza**

**crivera@dcc.uchile.cl**

27 de noviembre de 2002

## Abstract del artículo original

*“Richly interlinked, machine-understandable data constitute the basis for the Semantic Web. We provide a framework, CREAM, that allows for creation of metadata. While the annotation mode of CREAM allows to create metadata for existing web pages, the authoring mode lets authors create metadata— almost for free— while putting together the content of a page.*

*As a particularity of our framework, CREAM allows to create relational metadata, i.e. metadata that instantiate interrelated definitions of classes in a domain ontology rather than a comparatively rigid template-like schema as Dublin Core. We discuss some of the requirements one has to meet when developing such an ontology-based framework, e.g. the integration of a metadata crawler, inference services, document management and a meta-ontology, and describe its implementation, viz. Ont-O-Mat a component-based, ontology-driven Web page authoring and annotation tool”.*

## 1. ¿Qué es la Web Semántica?

Tim Berners-Lee (junto a otros colaboradores) en una entrevista a Scientific American<sup>2</sup> dio la siguiente definición de la web semántica:

---

<sup>1</sup> **Detalles del artículo:** Este artículo fue presentado en la 11a Conferencia Internacional del World Wide Web (Eleventh International World Wide Web Conference), en Hawaii, EEUU en el 2002. El título original es “Authoring and Annotation of Web Pages in CREAM”. El artículo fue publicado en los proceedings de la conferencia.

*“La Web Semántica es una extensión de la Web actual en la cual la información tiene un significado bien definido, permitiendo de una mejor manera que las personas y las computadoras trabajen en cooperación”.*

Una parte fundamental de la web semántica son los metadatos que contendrán las páginas web que integran la web semántica. Esencialmente los metadatos son datos sobre los datos que contiene cada página web, los cuales ayudan a describir los contenidos del documento.

La web semántica requiere que los metadatos sean relacionales. Esto es, metadatos que describen la manera en como las descripciones de un recurso instancian definiciones de clases y la manera en como están semánticamente ligadas por propiedades

### **1.1. ¿Qué es una ontología?**

Una ontología es una especificación de una conceptualización. Una ontología es una descripción (como una especificación formal de un programa) de los conceptos y relaciones que pueden existir para un agente o una comunidad de agentes. Específicamente en CREAM, una ontología especificación formal y explícita de una conceptualización compartida de un dominio de interés, y está constituida por enunciados que expresan definiciones de clases y propiedades de DAML+OIL.

## **2. Herramientas para anotaciones**

Por lo general, cuando se quiere hacer anotaciones a páginas web mediante alguna aplicación de software el autor primero tiene que crear el contenido de la página y luego crear los metadatos (anotaciones) en un paso posterior. Este es un problema, ya que implica una etapa extra a la que hay que dedicarle tiempo y recursos que muchas veces son limitados. Lo ideal sería que en el momento mismo de creación de las páginas web, se insertarían los metadatos. Los autores de este artículo proponen una estrategia para hacer esto posible.

## **3. CREAM**

CREAM es un acrónimo de CREAting Metadata (creando metadatos). Se trata de un esquema o framework para el agregado de metadatos relacionales, de una manera fácil y confortable, a páginas web. Para realizarlo se utiliza una herramienta, también desarrollada por los autores de CREAM, para hacer las anotaciones llamada Ont-O-Mat (<http://annotation.semanticweb.org/tools/ontomat/index.html>). La versión actualizada de esta herramienta es el OntoMat-Annotizer.

---

<sup>2</sup>Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001  
<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>

Una meta-ontología en CREAM va a ser definida como la manera en como los modos de anotación y de creación de páginas de Ont-O-Mat interfieren con las clases y las propiedades de la ontología

Debido a que no existía ninguna herramienta que implementará en su totalidad las propuestas de CREAM, se decidió implementar una nueva herramienta, el Ont-O-Mat. En CREAM se introducen dos elementos:

- Meta-ontología, la cual describe la manera en como los modos de anotación y autoría (creación de páginas web) del Ont-O-Mat interfieren con las clases y propiedades de una ontología, y
- Nuevos modos de interacción que permiten cambiar fácilmente entre los modos de autoría y de anotación.

### **3.1. Requisitos para la herramienta de anotaciones**

La herramienta que los autores de este artículo decidieron implementar, debería de tener las siguientes características:

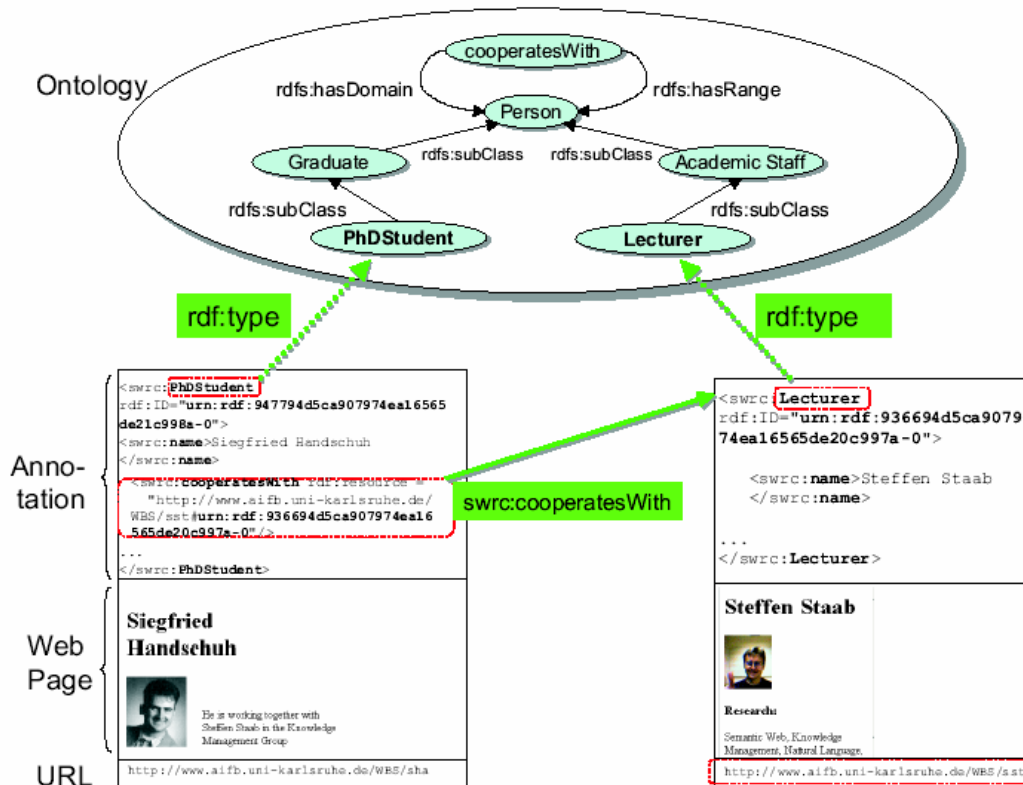
- Consistencia. Las estructuras semánticas deben de adherirse a una ontología dada para permitir un mejor intercambio de conocimiento.
- Referencias propias. Los identificadores de instancias (por ejemplo de personas, empresas, organizaciones, etc.) deben de ser únicos.
- Evitar redundancias. Debe de ser posible que los usuarios logren identificar si un documento ya tiene anotaciones, y en su caso permitir la reutilización de material previamente desarrollado.
- Metadatos relacionales. Al igual que la información en HTML, la cual está diseminada en la web, la marcación del conocimiento (knowledge markup) puede estar distribuida, pero debe de estar semánticamente relacionada.
- Mantenimiento. La herramienta debe de permitir labores de mantenimiento de una manera clara y simple.
- Facilidad de uso y eficiencia. Obviamente el ambiente de anotación debe de ser fácil de usar para ser realmente útil. La eficiencia en la creación de metadatos permitirá que los usuarios creen una mayor cantidad de metadatos.

En el contexto de este artículo, se consideran a las anotaciones como un conjunto de instancias incrustadas en documento en HTML. En CREAM se distinguen tres tipos de anotaciones:

- Instancias de clases de DAML+OIL
- Propiedades instanciadas de una clase a una instancia e un tipo de datos
- Propiedades instanciadas de una instancia de una clase a otra instancia de una clase

Las instancias de clases tienen URIs únicas.

La siguiente figura muestra la manera en que se emplean los términos “ontología”, “anotación” y “metadatos relacionales” dentro del esquema propuesto por CREAM. En ella se muestra parte de la ontología de SWRC (Semantic Web Research Community).



Se puede ver que se tienen dos páginas web con anotaciones dadas en una serialización de XML de hechos de RDF. Para las dos personas existen instancias, las cuales se denotan por sus correspondientes URI's (rdf:ID). Además, existe una instancia de relación entre las dos personas: cooperan entre ellas. Esta información se ve reflejada en ambas páginas.

#### 4. Diseño de CREAM, módulos

CREAM consiste de varios módulos. Estos son:

- Editor/visor de documentos. Permite la visualización del documento. El creador de metadatos permite la creación de nuevos metadatos al seleccionar trozos del texto y colocarla en la parte que le corresponda dentro de la ontología. El visor de documentos soporta distintos formatos (HTML, PDF, XML, etc.).
- Generación de contenido. El editor permite la creación convencional de documentos. Además de esto, las instancias ya existentes pueden ser tomadas de una visualización del contenido del servidor de inferencias de anotaciones y llevada hacia el documento.
- Navegador de hechos y guía de la ontología. CREAM necesita de un sistema para guiar al usuario dentro de la ontología. Para permitir el intercambio de conocimiento, las anotaciones recién creadas deben de ser consistentes con la ontología de la comunidad. Tanto el navegador de hechos y la guía de la ontología, y el editor/visor de documentos deben de ser fáciles de usar (permitir

acciones de tipo *Drag-and-Drop* para evitar en lo posible errores de sintaxis, por ejemplo).

- Crawler. La creación de los metadatos relacionales debe de tener lugar *dentro* de la web semántica. Durante la creación de los metadatos los usuarios deben de estar concientes de cuales entidades ya existen en su parte de la web semántica. Esto sólo es posible si un crawler hace relevantes de alguna manera las entidades que estén disponibles.
- Servidor de inferencias de anotaciones. Los metadatos relacionales, las referencias propias y la verificación de anotaciones redundantes requieren de consultas sobre instancias. Para este propósito, así como para asegurar la consistencia, CREAM provee de un servidor de inferencias de anotaciones. Este va a actuar sobre el crawler y las nuevas instancias creadas, y sobre la ontología.
- Administración de documentos. Un objetivo futuro del framework descrito en CREAM será el mantenimiento semi-automático de las anotaciones en páginas web externas. Cuando sólo cambian algunas partes del documento, el casamiento de patrones (pattern matching) podría ayudar a identificar las anotaciones viejas. En la implementación actual, Ont-O-Mat utiliza la URI para detectar la repetición de documentos anteriormente anotados y destaca las anotaciones en el documento antiguo del usuario. En ese momento el usuario decide si quiere ignorar o incluso borrar las viejas anotaciones y crear nuevos metadatos.
- Re-reconocimiento de metadatos y extracción de información. Aún con herramientas sofisticadas es laboriosa la creación de anotaciones semánticas. Uno de los objetivos principales de CREAM es permitir la creación semiautomática de metadatos utilizando técnicas de extracción de información para proponer anotaciones a los creadores de metadatos y así facilitar la tarea de creación de metadatos.
- Meta ontología. El propósito de la meta-ontología es la separación del diseño y el uso de una ontología.

Además, se pueden definir funciones que afectan a varios módulos, como:

- Almacenamiento. CREAM suporta dos maneras de almacenamiento. Las anotaciones pueden ser almacenadas dentro del documento que se encuentra dentro del componente de administración de documentos, pero también se puede almacenar en el servidor de inferencias de anotaciones.
- Replicación. Se cuenta con un sencillo mecanismo de replicación, el cual realiza un recorrido (*crawling*) de las anotaciones que se encuentran en el servidor de inferencias de anotaciones. De esta manera, es fácil reconocer las posibles inconsistencias.

## 5. Modos de Interacción del Ont-O-Mat

El Ont-O-Mat ofrece tres modos de interacción con la herramienta. Un modo de interacción permite cambiar fácilmente entre los ambientes de autoría de páginas web y las anotaciones. Los modos de interacción en Ont-O-Mat son:

- Anotación por escritura de enunciados (Annotation by Typing Statements)
- Anotación por marcado (Annotation by Markup)

- Anotación en la elaboración de páginas web (Annotation by Authoring Web Pages)

### **5.1. Anotación por escritura de enunciados**

Se basa únicamente en utilizar el navegador de guía de ontología/hechos y los templates generados. Por lo general, el usuario experimentado navega la ontología y los hechos; el menos experimentado utiliza los templates que proporciona la herramienta.

El usuario genera metadatos (instancias de clases, instancias de atributos, instancias de relaciones) que son completamente independientes de la página web que se este viendo en ese momento. Además, se puede hacer drag-and-drop de las instancias que ya están en la base de conocimientos para crear nuevas instancias de relaciones

### **5.2. Anotación por Marcado**

La idea básica de la anotación por marcado es usar el contenido en el editor/visor de documentos para generar instancias:

- De clases. Cuando el usuario lleva una pieza de contenido hacia un concepto de la ontología se genera una instancia.
- De atributos. Para poder generar una instancia de atributos el usuario simplemente lleva el contenido hacia al entrada correspondiente en la tabla.
- De relaciones. El usuario lleva el contenido deseado hacia la relación de la instancia pre-seleccionada.

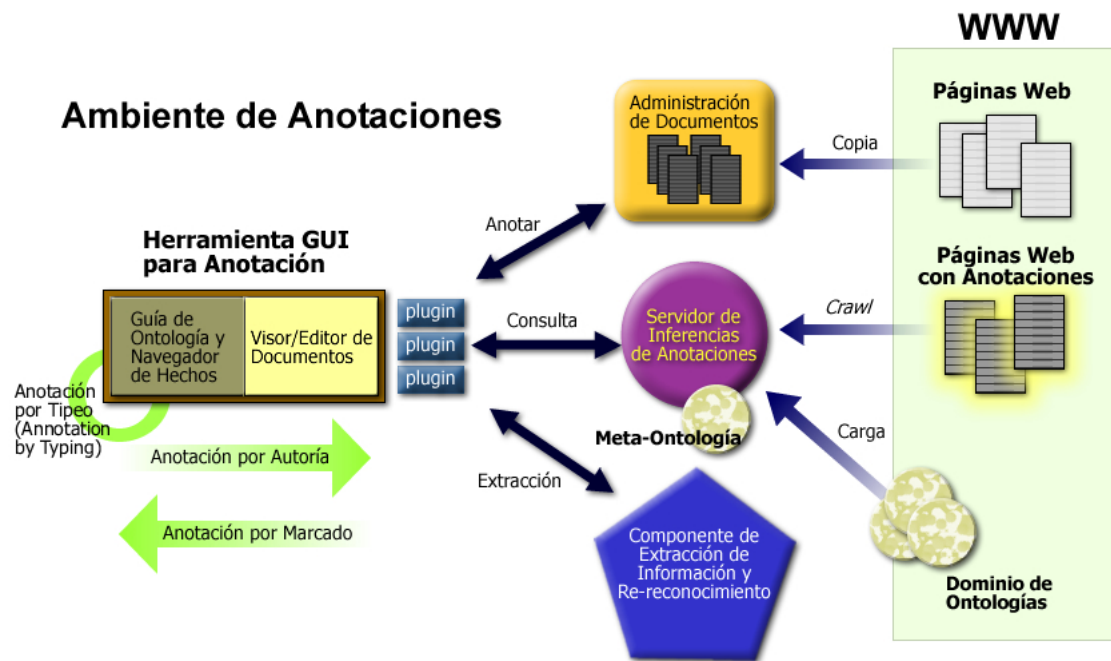
### **5.3. Anotación en la Elaboración de Páginas Web**

Se refiere a la creación de páginas web y de metadatos al mismo tiempo.

Se puede hacer de dos maneras:

- Utilizando la guía de ontología y el navegador de hechos para la generación de contenido
- Con la ayuda de el re-reconocimiento de metadatos o –de una manera más general- con la extracción de información

La siguiente figura muestra la arquitectura empleada en CREAM, y la cual muestra la manera en como se dan los modos de interacción.



## 6. Comparación con trabajo relacionado

Existen varias maneras de realizar anotaciones a páginas web, por lo que es de esperarse que CREAM (y el Ont-O-Mat) requieran de una comparación para identificar las ventajas y desventajas del framework propuesto.

CREAM se puede comparar en cuatro dimensiones con las otras propuestas:

- Es un framework para el marcado de la web semántica.
- Puede ser considerado como un framework particular para la adquisición de conocimiento
- Es un framework para anotaciones
- Es un framework para la creación de páginas web con énfasis en los metadatos

Los autores hicieron un estudio exhaustivo de las herramientas representativas de cada uno de los puntos anteriores para incorporar en Ont-O-Mat, dentro de lo posible, las características más destacadas de esas herramientas, y mejorar las desventajas de estas respecto al esquema propuesto en CREAM<sup>3</sup>.

## 7. Conclusiones

CREAM es un óptimo framework para la creación de anotaciones, especialmente de metadatos relacionales –el fundamento de la futura web semántica

<sup>3</sup> Las herramientas analizadas por los autores fueron Annotea Amaya, Protegè-2000, SHOE Knowledge Annotator, ThirdVoice, Yawas, CritLink y Briefing Associate.

Ont-O-Mat, como la implementación de referencia del framework CREAM, es una herramienta eficaz para el marcaje de documentos. En general, resulta fácil de utilizar e instalar, pero carece de una documentación en línea (se ofrece un breve tutorial al usuario en su lugar).

Actualmente, los autores están trabajando en el desarrollo de herramientas que permitan la creación de metadatos relacionales en PDF, SVG o SMIL, usando Ont-O-Mat. Los autores del artículo reconocen que Ont-O-Mat es un trabajo en progreso y falta un buen trecho para que esta herramienta implemente en su totalidad a CREAM.

## **Referencias**

Web semántica, <http://www.semanticweb.org/>

Anotaciones, <http://annotation.semanticweb.org/>

“What is an ontology?”, Tom Gruber <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>