

## Hacia un Enfoque Semántico para la Catalogación de Métricas

M<sup>a</sup> de los A. Martín<sup>1</sup>, M. F. Bertoa<sup>2</sup>, A. Vallecillo<sup>2</sup>, L. Olsina<sup>1</sup>

<sup>1</sup>*GIDIS, Departamento de Informática, UNLPam*

*Calle 9 y 110, (6360) General Pico, La Pampa. Argentina*

*E-mail [olsinal.martinma@ing.unlpam.edu.ar](mailto:olsinal.martinma@ing.unlpam.edu.ar)*

<sup>2</sup>*Departamento de Lenguajes y Ciencias de la Computación,*

*Universidad de Málaga, España*

*E-mail [bertoa.av@lcc.uma.es](mailto:bertoa.av@lcc.uma.es)*

**Resumen.** En los últimos lustros han surgido numerosas propuestas que definen conjuntos de métricas de calidad para la evaluación de productos y procesos de software en distintos dominios de aplicación. Un trabajo reciente sobre métricas [3, 7] propone un marco conceptual para la definición, documentación y explotación de dichas métricas desde una perspectiva integradora, y define los requisitos y la arquitectura que debe tener un repositorio de métricas con fines de consulta y reuso. En este paper introducimos el diseño de una capa semántica para tal repositorio que facilitará la explotación e intercambio consistente de información de métricas entre usuarios a partir de repositorios, como así también la explotación y uso por parte de aplicaciones, agentes y herramientas automáticas en la Web.

**Palabras Claves:** *Web Semántica, Repositorio de Métricas, XML, RDF.*

### 1. Introducción

El aseguramiento de calidad es una actividad no muy tenida en cuenta en los proyectos de software la cual abarca diversos aspectos de procesos, productos y recursos, entre otros, las métricas asociadas a esos entes, las cuales son de utilidad para evaluar, predecir, controlar y en definitiva mejorar la calidad. Afortunadamente, en los últimos años varios trabajos de investigación teórica y empírica han definido conjuntos de métricas de calidad para distintos dominios de aplicación ([2, 4, 6, 11], entre otros) los que pueden ser reusados por desarrolladores y evaluadores en pos de mejorar la calidad de productos y procesos. Además, han surgido herramientas de recolección de datos y evaluación automáticas que han sido desarrolladas para dar soporte a propuestas metodológicas y hacer más efectivas las actividades de aseguramiento de calidad.

Observando el volumen cada vez mayor y heterogéneo de información relacionada a métricas de calidad urge el disponer repositorios genéricos y herramientas de catalogación que faciliten de un modo extensible y consensado, el almacenamiento, consulta, explotación y reuso de toda la información relacionada a métricas que sirva de soporte a las actividades de aseguramiento de calidad. No menos importante es brindar una estructura común y consistente de dicha información (metadatos de las métricas), de manera que pueda ser usada como base para la especificación formal del tipo de entidad, de la definición del atributo, del objetivo y motivación, fórmula, criterios, procedimientos de recolección, unidad, tipo de escala, entre otros elementos.

Es importante destacar los esfuerzos recientes realizados por Olsina et al. [7] y Bertoa et al. [3], en la propuesta de un marco conceptual para la definición y explotación de métricas de calidad con las características antes citadas. Dichos trabajos especifican además los requisitos del repositorio de métricas y describen una propuesta concreta de arquitectura. Sin embargo, aspectos de diseño e implementación fueron parcialmente abordados.

Por otra parte, teniendo en cuenta la evidente revolución que ha causado el uso cada vez más generalizado de la Web en el intercambio de información personal, académica, científica y

comercial, y la evolución que va a implicar la denominada Web Semántica [1, 8], nos ha parecido importante que el entorno de catalogación de métricas cuente con una capa semántica convenientemente definida e implementada, de manera que: a) haya un vocabulario común (ontología) que defina los conceptos relacionados con métricas para distintos dominios; y b) la información relacionada con métricas pueda ser interpretada por computadoras sin necesidad de intervención humana. Esto potenciará la capacidad de consulta y procesamiento no sólo para los usuarios finales, sino también para agentes y herramientas.

El aporte del presente trabajo consiste en definir y especificar una parte de la capa semántica del repositorio genérico de métricas a partir del marco conceptual propuesto en [3, 7] con el fin de agregar interoperabilidad, generalidad y potencia de procesamiento, siendo estas características una de las fortalezas del enfoque de Web Semántica.

El resto del artículo se organiza de la siguiente manera. En la sección 2 se describen brevemente los principales conceptos de Web Semántica útiles para una mejor comprensión de este trabajo. En la sección 3 se describe el marco conceptual para la definición y explotación de métricas, el cual sirve de punto de partida para nuestra definición semántica del catálogo de métricas. Seguidamente, en la sección 4, definimos para el mismo la capa RDF (*Resource Description Framework*) y RDFS (*RDF Schema*) del modelo semántico. En la sección 5 se muestra mediante ejemplos la documentación de métricas usando los esquemas RDF definidos y se analiza la potencia del modelo semántico. Finalmente en la sección 6 se dan las conclusiones y líneas de avance.

## 2. Panorama del Modelo de Web Semántica

### 2.1 Estado del Arte.

Según Berners-Lee [1], la Web Semántica es un espacio donde la información tiene un significado bien definido, de manera que pueda ser interpretado tanto por agentes humanos como computarizados. En la actualidad, la web es un medio de comunicación diseñado para ser entendido por humanos, donde los contenidos (aún cuando sean generados a partir datos altamente estructurados, como una base de datos) no están suficientemente estructurados ni tienen un significado formal, como para que un agente computacional pueda “entenderlos” (procesarlos).

La Web Semántica está surgiendo como una evolución de la web actual a la que se agrega una estructura para captar el significado de los contenidos de las páginas y proporcionar un ambiente donde las aplicaciones puedan procesar y relacionar contenidos provenientes de distintas fuentes.

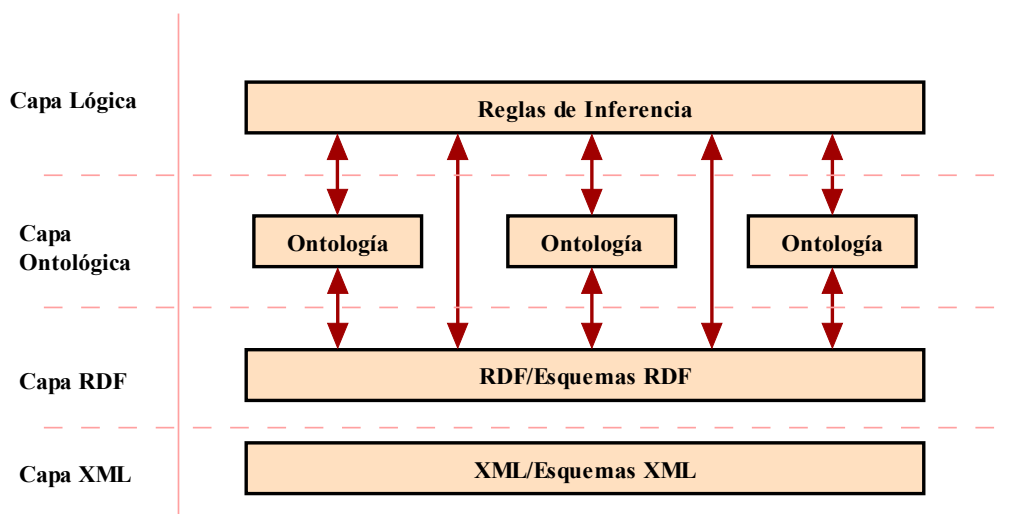


Figura 1. Arquitectura básica para el enfoque semántico.

En la propuesta de desarrollo de la Web Semántica del consorcio W3C [8] se sugiere una arquitectura en capas (ver Fig.1), comenzando por la capa de nivel inferior XML (*eXtended Markup Language*) que permite estructurar los datos, siguiendo por una capa RDF que define la semántica de dichos datos, siguiendo por la capa de ontología, que define relaciones entre dichos datos, y por último, la capa lógica, que define las reglas y mecanismos para hacer inferencias sobre dichos datos. A continuación explicamos brevemente la función de cada una de las capas.

## 2.2 La capa XML.

Las páginas web tradicionales contienen datos no estructurados que si bien son fáciles de interpretar por los seres humanos no pueden ser procesados por máquinas debido a la falta de estructura sintáctica. El lenguaje XML fue creado en 1998 por el W3C y soluciona este problema permitiendo describir la estructura de los contenidos de datos que se encuentran en las páginas Web. Utilizando este lenguaje es posible definir arbitrariamente una estructura para un documento, aún cuando no exista ninguna información sobre el significado de dicha estructura. El siguiente ejemplo, muestra un documento XML que define parcialmente los datos de una métrica de calidad en forma estructurada.

```
<?xml version="1.0"?>

<metrics
  xmlns="http://www.cotsmetric.com/Metrics.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:dc="http://purl.org/dc#"

  repository="http://www.theMetricsRepository.com">
  <name>Orphan Page Count</name>
  <acceptabilityRange>10</acceptabilityRange>
  .....
  <dataGathering>
    < procedure mode="automatic"
      xlink:href="http://www.yourtool.com?countOrphanPages()" />
  </dataGathering>

  <reference xlink:href="www.useit.com/alertbox/9605.html" >
    <dc:Title>Top Ten Mistakes in Web Design< /dc:Title >
    <dc:Creator> J. Nielsen < /dc:Creator >
    <dc:Date>1996-05-01< /dc:Date >
  </reference>
</metrics>
```

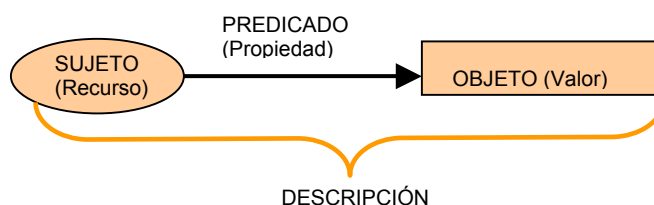
**Figura 2.** Ejemplo de documentación de una métrica en XML.

XML es un metalenguaje ya que nos permite definir lenguajes de marcado adecuados para usos en distintos dominios de una manera flexible y extensible. Esto se puede observar claramente en el ejemplo anterior donde las etiquetas correspondientes a los distintos dominios, en este caso el de métricas, de documentos bibliográficos y enlaces, están definidos en archivos XML separados, “http://www.cotsmetric.com/Metrics.xsd”, “http://purl.org/dc#” y “http://www.w3.org/1999/xlink” respectivamente.

## 2.3 La capa RDF.

Mientras que XML permite definir datos de una forma estructurada, RDF permite definir la semántica de esos datos. El modelo RDF se compone de tres conceptos principales: Los *recursos*, que representan cualquier cosa (lugares, personas, objetos) del mundo real y están identificados por un URI (*Universal Resource Identifier*); las *propiedades* que son atributos o relaciones para describir un recurso, y las *sentencias* que es una combinación de los conceptos anteriores y se compone de un sujeto, un predicado y un objeto. El significado de los datos se expresa mediante un conjunto de dichas sentencias que son representadas por tri-uplas (sujeto, predicado, objeto) y que se escriben

usando marcas XML. El sujeto es un recurso identificado por un URI, los recursos pueden representar personas, lugares o cosas en el mundo real. El predicado describe una propiedad o atributo del recurso, y permite la definición de un nuevo concepto a través de un puntero al documento (URI) donde está definido dicho concepto. Es decir, una propiedad es a su vez un recurso. El objeto es el valor asignado a dicha propiedad, y puede ser un literal, u otro recurso. Otra forma de representar las declaraciones RDF es a través de un grafo rotulado como se muestra en la figura 3.



**Figura 3.** Representación gráfica de las declaraciones RDF.

La figura 4 muestra un ejemplo de documento RDF que describe parcialmente a la métrica *OrphanPagesCount* (ver [7] para detalles de la métrica). Por ejemplo, se observa el siguiente triplete (*OrphanPagesCount*, <http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1#AutomatedBy>, <http://gidis.ing.unlpam.edu.ar/WebSiteMA>), en donde el sujeto es la métrica, el predicado, es la propiedad *AutomatedBy*, y el valor es el URI de la herramienta que la automatiza. En la figura se pueden identificar otros tripletes. Por otra parte, para la propiedad *Quantifies* el URI es <http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1#Quantifies>.

```
<?XML version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:metricas="http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1"
xmlns:qualityMetrics="http://www.cotsmetric.com/Metrics.xsd">

  <rdf:Description ID="OrphanPagesCount">
    <rdfs:Type rdf:resource="http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1#Metric"/>
    <metricas:Quantifies rdf:resource="http://gidis.ing.unlpam.edu.ar/Atributos#OrphanPages"/>
    <metricas:AppliesTo rdf:resource="http://gidis.ing.unlpam.edu.ar/Entidades#Webpages"/>
    <qualityMetrics:acceptabilityRange>10</qualityMetrics:acceptabilityRange>
    <metricas:AutomatedBy rdf:resource="http://gidis.ing.unlpam.edu.ar/WebSiteMA"/>
    <metricas:AutomatedBy rdf:resource="http://www.yourtool.com?countOrphanPages()"/>
  </rdf:Description>
```

**Figura 4.** Ejemplo de descripción de una métrica en RDF.

### 2.3 La capa Ontológica.

La utilización de URIs para describir las propiedades de los recursos en RDF garantiza la definición única de los conceptos que representan los datos de documentos en la web. Sin embargo, un mismo concepto se puede expresar con términos distintos en distintos sitios o aplicaciones Web. Por ejemplo, otro repositorio de métricas podría usar el término *Cuantifica* en vez de *Quantifies*. Para que la información existente en distintos documentos dispersos en distintos sitios se puedan procesar y relacionar conjuntamente, es necesario que exista una definición consensuada de los conceptos contenidos en distintos documentos y las relaciones entre dichos conceptos. La solución propuesta para la web semántica son las *Ontologías*.

Para los investigadores en web semántica, una Ontología es un documento o archivo que define los términos utilizados en un dominio y la relación entre estos términos [8].

## 2.4 La capa Lógica.

La capa lógica está compuesta por un conjunto de reglas de inferencia que los agentes (computacionales o humanos) podrán utilizar para relacionar y procesar la información. Estas reglas ofrecen el poder de deducir nuevas sentencias a partir de los datos y estructuras que están descritos en las capas XML y RDF, usando además las relaciones entre esos datos y estructuras definidas en la capa Ontológica.

## 3. Aspectos del Modelo Conceptual del Catálogo de Métricas y de la Arquitectura.

### 3.1. El Modelo Conceptual.

A continuación describimos el modelo para repositorios de métricas, en el cual representamos todos los elementos y relaciones necesarios para definir y explotar distintas métricas de diferentes dominios de software. A partir de este marco general, el comportamiento del modelo podrá instanciarse para proyectos de dominios de aplicación específicos, lo que permitirá, por un lado, guardar los meta-datos de los elementos intervinientes en un proceso de catalogación y, por otro, guardar instancias de métricas con sus valores para entes en proyectos de software específicos (este último objetivo es un trabajo futuro).

En la Figura 5 representamos (mediante un diagrama de clases y relaciones) los principales elementos del modelo conceptual [3, 7] que describimos seguidamente:

**3.1.1 Las Clases Entidad y Atributo.** Una entidad (o ente) representa a un objeto, tangible o intangible, que exhibe un comportamiento observable en el mundo real (o dominio empírico). Desde el punto de vista del dominio del software, entes de interés a un alto nivel de abstracción son: *Proyecto*, *Proceso*, *Recurso*, *Producto* y *Producto en Uso*. A su vez, los entes se pueden subdividir en sub-entes de diferentes tipos.

Los entes no se pueden medir directamente, sino a través de las propiedades atribuidas. De esta manera, el modelo representa tanto a los atributos relacionados con las entidades, como a la relación jerárquica de los entes entre sí, que desde la perspectiva del repositorio nos podrá permitir una navegación que facilite la exploración de atributos (y métricas) por jerarquías.

Por otra parte, en la Figura 5 se observan las clases *Atributo Directo* e *Indirecto*. Los atributos directos son atómicos, no se descomponen en nuevos atributos, en tanto que un atributo indirecto se puede representar por una relación de atributos.

Por último, dentro de la clase *Atributo* especificamos campos<sup>1</sup> como nombre, palabras clave, definición, objetivo/motivación para utilizar un atributo dado, entre otros.

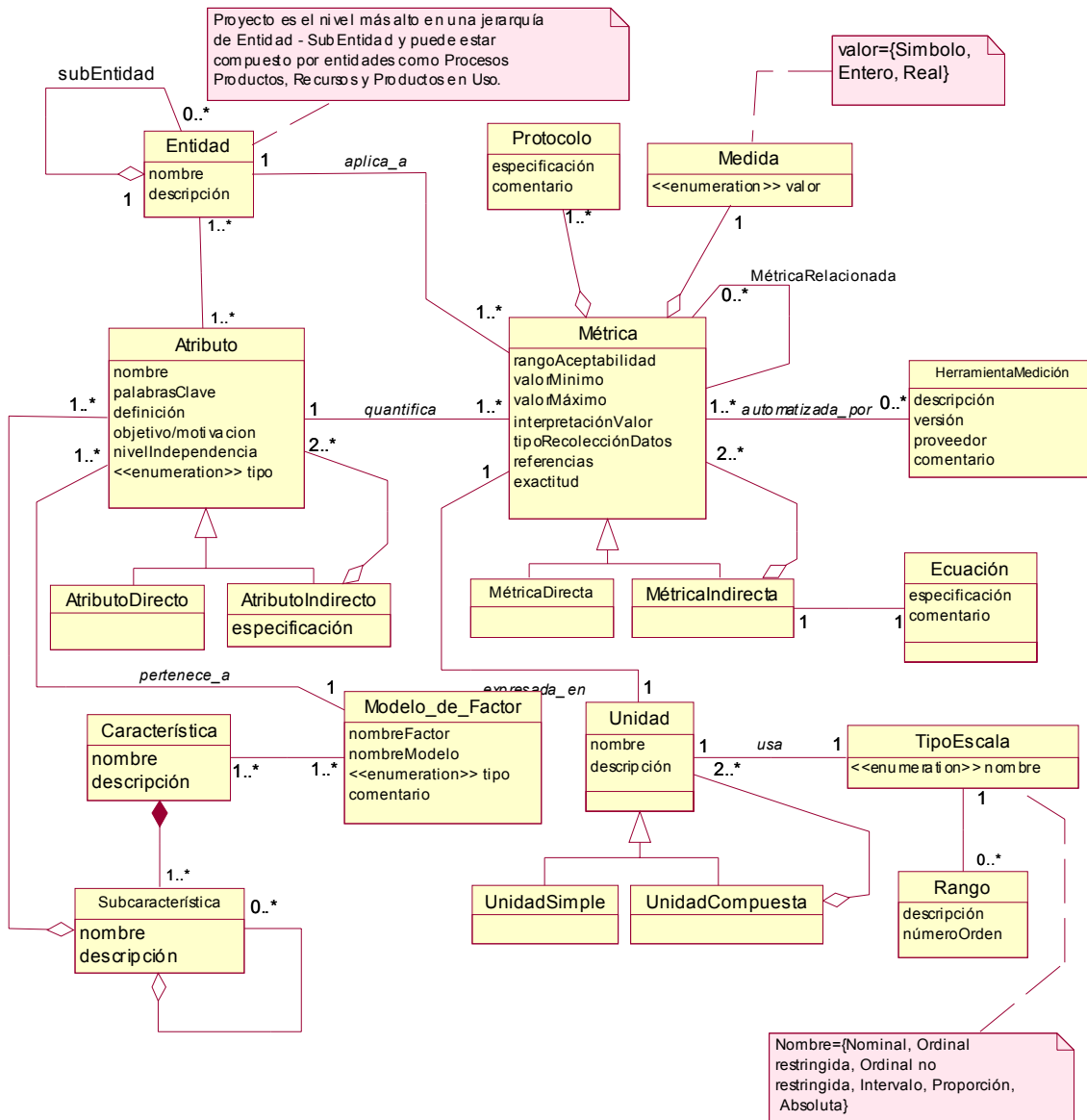
**3.1.2 La Clase Métrica y Clases Relacionadas.** Un atributo puede ser cuantificado por una o varias métricas. Por ejemplo, el atributo *Tamaño del Texto Visible de una Página* puede ser medido por una métrica  $m_1$  que cuantifique el tamaño por la cantidad de palabras, o por una métrica  $m_2$  que la cuantifique en bytes.

La idea de métrica no es un concepto simple de desarrollar, si no se comprenden y analizan sus componentes y relaciones. Se la debe comprender en consideración de los atributos a los que cuantifica y a los entes a los que se asocia. Asimismo, es preciso identificar el tipo de valores que se obtienen, la unidad en la que se expresa, y el tipo de escala que se usa, con el fin de poder realizar una apropiada interpretación y un análisis matemático y/o estadístico [4]. Además, un aspecto fundamental para garantizar la repetitividad y replicabilidad del proceso de medición, es conocer las reglas de conteo y procedimientos (que denominaremos en este ámbito como “protocolo”), y determinar si el proceso de recolección de datos y cálculo se puede automatizar mediante un instrumento de medición.

---

<sup>1</sup> Utilizamos la palabra campo, en reemplazo de atributo de una clase, para no sobrecargar el concepto de Atributo utilizado en el ámbito de métricas

Como se observa en la Figura 5, las clases *Métrica Directa* e *Indirecta* son un tipo de *Métrica*. El concepto de métrica directa e indirecta (en el dominio formal) está estrechamente relacionado a la cuantificación de atributos directos e indirectos respectivamente (en el dominio empírico). En el caso de una métrica indirecta, es necesario saber cual es la *Ecuación* que formaliza y especifica al atributo indirecto.



**Figura 5.** Modelo Conceptual para el dominio de métricas.

Una *Medida* (cuyo valor puede ser de distintos tipos) es parte de una métrica, y no puede ser interpretada sin una *Unidad* de medida y un *Tipo de Escala*. A su vez, podemos modelizar unidades simples y compuestas como se observa en la Figura 5.

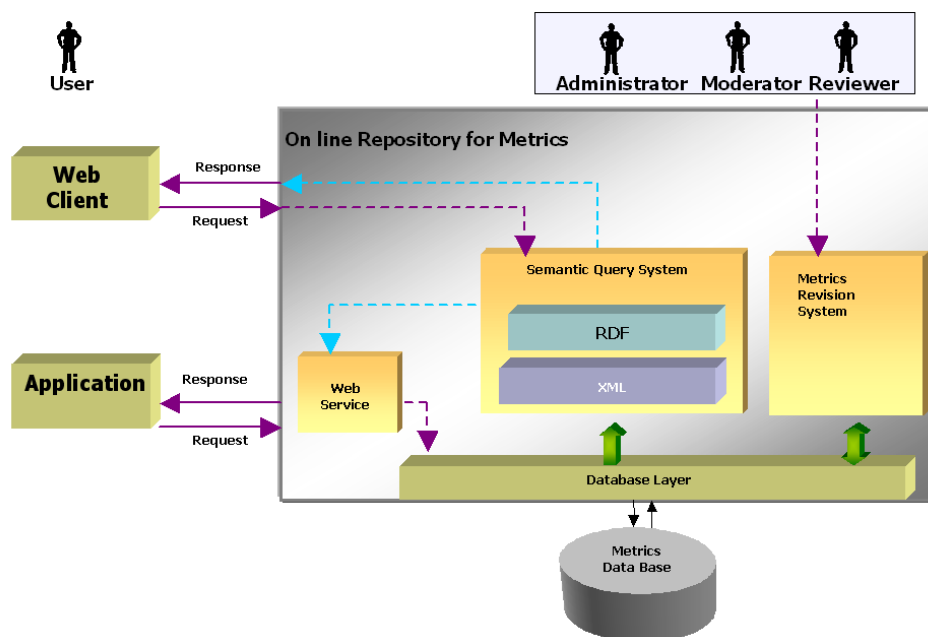
Por otra parte, es importante remarcar que una métrica puede tener uno o más protocolos de medición. La clase *Protocolo* especifica los mecanismos y procedimientos a ser tenidos en cuenta al momento de recolectar los datos con el fin de obtener el valor de una métrica. Así, las reglas de conteo para las dos métricas del ejemplo anterior varían según el caso. Además, si la métrica es

automatizable (o semiautomatizable) desde el punto de vista de la recolección de datos, se puede indicar en el campo “especificación” el algoritmo correspondiente (por ejemplo, en pseudo código u otro formalismo) para que sirva como guía en el proceso con el fin de garantizar repetitividad y replicabilidad. Para este tipo de métricas se puede usar herramientas de medición tal como se ilustra en el modelo. Por último, desde el punto de vista de consulta al repositorio de métricas catalogadas, puede ser útil navegar por métricas relacionadas como se aprecia en la Figura 5.

### 3.2 La Arquitectura del Repositorio de Métricas.

La Figura 6 muestra el esquema de lo que constituye la arquitectura objetivo del framework. En la parte superior se muestran los distintos tipos (roles) de usuarios. Los servicios del framework se ofrecen fundamentalmente a través de servicios web (*Web Services*), siendo XML la *lingua-franca* en la que se expresa toda la información. Esto va a ser el punto de partida para estructurar la información y permitir satisfacer los requisitos de interoperabilidad y generalidad que comentamos en la Introducción.

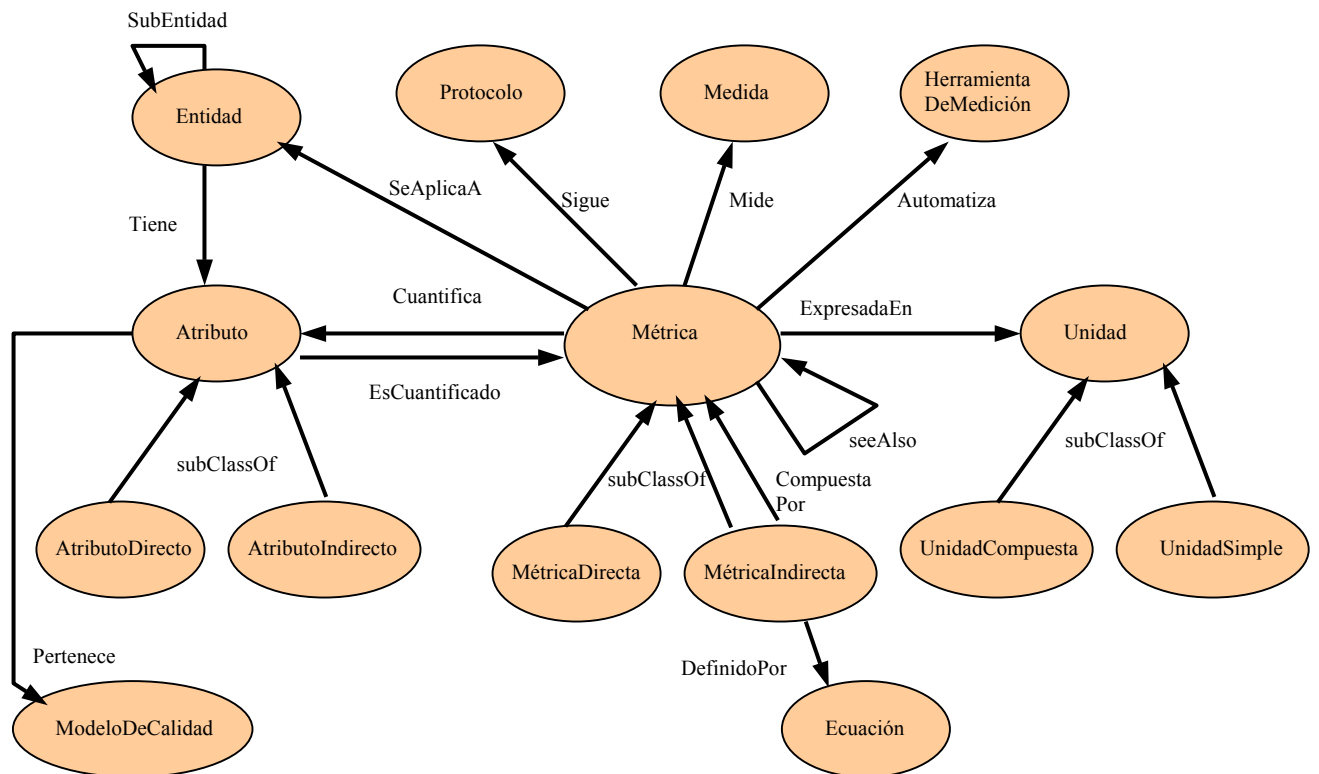
La definición de la arquitectura se ha realizado de modo tal que favorezca un camino de crecimiento hacia los aspectos semánticos de las búsquedas y el tratamiento de información mediante RDF y la Web Semántica.



**Figura 6.** Arquitectura general del framework de catalogación y consulta.

### 4. Definición del Esquema RDF para el Ámbito de Métricas

A continuación presentamos la definición de parte de la capa semántica del repositorio de métricas descrito anteriormente. Para la definición usamos el esquema RDF que nos permite construir un modelo abstracto de las relaciones entre los recursos, además de la explotación de las propiedades de orientación a objetos como reuso y herencia. Si observamos el modelo conceptual de la figura 5, se define una clase en RDF para cada clase de dicho modelo (ver Fig. 7), y se define una propiedad para cada relación (las propiedades *subClass* y *seeAlso* están predefinidas en el estándar de facto W3C y se decidió no redefinirlas para obtener mayor interoperabilidad). La descripción RDF está representada por un grafo (figura 7) que muestra las relaciones entre los distintos recursos que se definirán detalladamente en el documento RDF de la figura 8.



**Figura 7.** Vista del grafo RDF para el dominio de métricas.

En la figura 7 se muestran las clases de recursos definidos para el modelo del repositorio de métricas y las relaciones más importantes que facilitarían su exploración y análisis semántico.

El tipo de recurso principal para el dominio de nuestro interés es la clase métrica, que conceptualiza toda la información disponible sobre una métrica determinada, y que por lo tanto se relaciona con los demás tipos de recursos. A saber: el tipo de entidad a la que se aplica, el atributo de dicha entidad que mide, las reglas de conteo y procedimientos (representada por el recurso Protocolo), las herramientas de medición automáticas existentes (si la métrica fuera automatizable), la unidad en la que se expresa, y el conjunto de mediciones efectuadas para esa métrica (si las hubiera).

Se definen también otras propiedades (relaciones) para facilitar la exploración de las métricas a través de la web semántica.

### 5. La Potencia del Modelo Semántico

En esta sección analizamos mediante un ejemplo la utilidad de RDF para estructurar y potenciar la capacidad de procesamiento, en nuestro caso, la información de métricas de calidad en la web. En la figura 9 se muestra una posible descripción de dos métricas para páginas web [7], *OrphanPagesCount* y *PercentageOfOrphanPages*, usando el esquema RDF para el dominio de métricas definido en la sección anterior.

Como se puede observar, la información de una métrica puede estar relacionada con métricas definidas en otros repositorios distribuidos en la web, como es el caso de la métrica *PercentageOfOrphanPages*. Esto nos muestra la potencia del esquema RDF, que no sólo nos permite describir la información de métricas en forma estructurada, consensuada y consistente tanto desde el punto de vista sintáctico como semántico (a través de la definición de los metadatos), sino



que también (y no menos importante) facilita el reusar y compartir toda la información relacionada al dominio.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdfs:Class rdf:ID="Metric">
    <rdfs:label xml:lang="en">Metric</rdfs:label>
    <rdfs:label xml:lang="es">Métrica</rdfs:label>
    <rdfs:comment>Información sobre Métricas de Calidad </rdfs:comment>
  </rdfs:Class>

  <rdfs:Class rdf:ID="DirectMetric">
    <rdfs:label xml:lang="en">DirectMetric</rdfs:label>
    <rdfs:label xml:lang="es">MétricaDirecta</rdfs:label>
    <rdfs:comment>Cuantifica atributos directos </rdfs:comment>
    <rdfs:subClassOf rdf:resource="#Metric"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="IndirectMetric">
    <rdfs:label xml:lang="en">IndirectMetric</rdfs:label>
    <rdfs:label xml:lang="es">MétricaIndirecta</rdfs:label>
    <rdfs:comment>Cuantifica atributos indirectos </rdfs:comment>
    <rdfs:subClassOf rdf:resource="#Metric"/>
  </rdfs:Class>

  <rdfs:Class rdf:ID="Entity">
    <rdfs:label xml:lang="en">Entity</rdfs:label>
    <rdfs:label xml:lang="es">Entidad</rdfs:label>
    <rdfs:comment>Entidades a Evaluar </rdfs:comment>
  </rdfs:Class>

  <rdfs:Class rdf:ID="Attribute">
    <rdfs:label xml:lang="en">Attribute</rdfs:label>
    <rdfs:label xml:lang="es">Atributo</rdfs:label>
    <rdfs:comment>Atributo de Calidad </rdfs:comment>
  </rdfs:Class>

  <rdf:Property ID="SubEntity">
    <rdfs:label xml:lang="en">SubEntity</rdfs:label>
    <rdfs:label xml:lang="es">SubEntidad</rdfs:label>
    <rdfs:domain rdf:resource="#Entity"/>
    <rdfs:range rdf:resource="#Entity"/>
    <rdfs:comment>Usado para jerarquía de Entidades</rdfs:comment>
  </rdf:Property>

  <rdf:Property ID="Has">
    <rdfs:label xml:lang="en">Has</rdfs:label>
    <rdfs:label xml:lang="es">Tiene</rdfs:label>
    <rdfs:domain rdf:resource="#Entity"/>
    <rdfs:range rdf:resource="#Attribute"/>
    <rdfs:comment>Atributos que tiene una Entidad</rdfs:comment>
  </rdf:Property>

  <rdf:Property ID="QualityModel">
    <rdfs:label xml:lang="en">QualityModel</rdfs:label>
    <rdfs:label xml:lang="es">ModeloDeCalidad</rdfs:label>
    <rdfs:domain rdf:resource="#Attribute"/>
    <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#resource"/>
    <rdfs:comment>Definición del Modelo de Calidad al que pertenece el Atributo </rdfs:comment>
  </rdf:Property>
</rdf:RDF>
```

**Figura 8.** Vista parcial del esquema RDF para el dominio de Métricas (Ver definición completa en <http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1>)

```

<?XML version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:metricas="http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1"
xmlns:qualityMetrics="http://www.cotsmetric.com/Metrics.xsd">

  <rdf:Description ID="OrphanPagesCount">
    <rdfs:Type rdf:resource="http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1#DirectMetric"/>
    <metricas:Quantifies rdf:resource="http://gidis.ing.unlpam.edu.ar/Atributos#OrphanPages"/>
    <metricas:AppliesTo rdf:resource="http://gidis.ing.unlpam.edu.ar/Entidades#Webpages"/>
    <qualityMetrics:acceptabilityRange>10</qualityMetrics:acceptabilityRange>
    <metricas:AutomatedBy rdf:resource="http://gidis.ing.unlpam.edu.ar/WebSiteMA"/>
    <metricas:AutomatedBy rdf:resource="http://www.yourtool.com?countOrphanPages ()"/>
  </rdf:Description>

  <rdf:Description ID="PercentageOfOrphanPages">
    <rdfs:Type rdf:resource="http://gidis.ing.unlpam.edu.ar/rdf/RDF-Metricas1#IndirectMetric"/>
    <metricas:Quantifies
rdf:resource="http://gidis.ing.unlpam.edu.ar/Atributos#PercentageOfOrphanPages"/>
    <metricas:AppliesTo rdf:resource="http://gidis.ing.unlpam.edu.ar/Entidades#Webpages"/>
    <qualityMetrics:acceptabilityRange>5</qualityMetrics:acceptabilityRange>
    <metricas:AutomatedBy rdf:resource="http://gidis.ing.unlpam.edu.ar/WebSiteMA"/>
    <metricas:CompoundOf rdf:resource="#OrphanPagesCount"/>
    <metricas:CompoundOf rdf:resource="http://www.infometrics/myrepository#PagesCount"/>
  <metricas:DefinedBy
rdf:resource="http://gidis.ing.unlpam.edu.ar/Ecuaciones#PercentageOfOrphanPages"/>
  </rdf:Description>

```

**Figura 9.** Ejemplo de Descripción de las Métricas *OrphanPagesCount* y *PercentageOfOrphanPages* en RDF, y algunas de sus propiedades.

Además es posible relacionar otros conceptos como son: qué atributo es cuantificado por qué métricas, qué protocolos usa una métrica dada, en qué unidad está expresada, qué valores de mediciones se han tomado, e incluso, es posible obtener la información y URL (*Universal Resource Locator*) de herramientas de medición que las implementen, para ser ejecutadas automáticamente, y de forma remota, entre otras relaciones.

Otro beneficio importante de la capa semántica, es que la información bien definida semánticamente puede ser entendida no sólo por seres humanos, sino también por agentes, motores de búsqueda y herramientas automáticas y ser aprovechado para facilitar las tareas de aseguramiento de calidad.

Por ejemplo, se podría programar un agente para que recorra la web (semántica) buscando información de métricas que midan atributos de calidad de páginas web. Dicho agente podría analizar los contenidos de cada sitio ejecutando la consulta (query) que mostramos en la figura 10. La consulta está escrita en RQL (*RDF Query Language* -ver por ejemplo la siguiente URL <http://139.91.183.30:9090/RDF/RQL/>) y devuelve todas las instancias de la clase *Metric* que tiene la propiedad *AppliesTo* igual a "WebPages".

```

select x,y
from http://gidis.unlpam.edu.ar/ rdf/RDF-Metricas1#Metric {x}.
http://gidis.unlpam.edu.ar/ rdf/RDF-Metricas1#AppliesTo {y}
where y = "WebPages"

```

**Figura 10.** Consulta RQL para encontrar la definición de las métricas de calidad que se aplican a paginas web.

Cabe aclarar que RQL permite consultar descripciones RDF semi-estructuradas usando la taxonomía de clases y propiedades descriptas en un esquema RDF (por ej., como el de la figura 8).

Además esta basado en el modelo de datos en forma de grafo permitiendo considerar instancias RDF como datos semi-estructurados que pueden ser interpretados por varios esquemas RDF.

Finalmente, como mencionado previamente, una herramienta automática de medición puede aprovechar los beneficios de la web semántica, por ejemplo, una aplicación podría basarse en la información proporcionada por la instancia RDF de una métrica indirecta, calcular las métricas que la componen recorriendo los enlaces proporcionados por las propiedades *CompoundOf* y luego aplicar la ecuación definida (tal vez remotamente) en el URI proporcionado por la propiedad *DefinedBy* (Ver definición en RDF de la métrica *PercentageOfOrphanPages* en la Fig 9).

## 6. Conclusiones y Trabajos Futuros.

Actualmente, la Web es un espacio diseñado para el intercambio de información para ser entendida principalmente por el ser humano. Si bien en los últimos años algunas empresas están publicando datos que siguen algún esquema de estructuración (normalmente basado en XML), no existe un formato y semántica común para mostrar y procesar la información. Así, los desarrolladores de páginas web crean sus páginas dependiendo de los potenciales usuarios que van a visitarlas sin pensar en la interoperabilidad y procesamiento estructurado de los datos. Con los actuales algoritmos de los buscadores de información web, al ingresar una palabra clave (o varias), con más o menos éxito se recupera información que puede ser o no de nuestro interés. Esto es debido a que los agentes actuales de búsqueda no se diseñan para “comprender” (procesar) la información que reside en la web, precisamente porque es prácticamente imposible conocer la semántica de los datos ubicados en las diferentes páginas.

Por otra parte, para el dominio de métricas de software discutido, observamos la falta de mecanismos y servicios que brinden la posibilidad de consultar, relacionar y, en definitiva, explotar y reusar información relativa a las mismas, tan importante como soporte a las tareas de aseguramiento de calidad.

Como un aporte tecnológico tendiente a solucionar dichos problemas, propusimos en este trabajo la definición de una capa semántica para un repositorio de métricas genérico, y analizamos su potencia a la hora de ser usada por agentes y herramientas automatizadas.

Nuestro avance es intentar obtener un consenso sobre la terminología (vocabulario) definida en la capa semántica para el dominio de métricas, y relacionarla con otras posibles terminologías del ámbito, con el propósito de obtener una ontología robusta. Además nos proponemos definir reglas de inferencia (la capa lógica) para potenciar aún más el uso del repositorio por parte de agentes y herramientas automáticas.

**Reconocimientos.** Este trabajo y línea de investigación están soportados por el proyecto UNLPam-09/F022, y el proyecto VII.18. WEST (*Web-based Software Technology*, <http://www.dsic.upv.es/~west/>), CYTED (*Programa Iberoamericano de Ciencia Y Tecnología para El Desarrollo*, <http://www.cytcd.org/Nueva.asp>).

## Referencias

1. Berners-Lee, T.; Hendler, J.; Lassila, O., 2001, *The Semantic Web*. Scientific American, <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>.
2. Bertoa M. F.; Vallecillo A., 2002, “*Quality Attributes for COTS Components*”. In Proc. of the Sixth ECOOP International Workshop on Quantitative Approaches in Object-Oriented Software Engineering (QAOOSE 2002), pp. 54-66. Malaga, Spain.
3. Bertoa M. F.; Olsina, L., Lafuente G. J., Martín M. A., Katrib M., Vallecillo A. 2002, “*Un Marco Conceptual para la Definición y Explotación de Métricas de Calidad*”, Submitted paper.

4. Fenton, N.E.; Pfleeger, S.L., 1997, "*Software Metrics: a Rigorous and Practical Approach*", 2<sup>nd</sup> Ed., PWS Publishing Company.
5. Lima F.; Schwabe D., 2002, "*Exploring Semantic Web Modeling Approaches for Web Application Design*". In Proc. of the 2<sup>nd</sup> Int'l Workshop on Web-Oriented Software Technology" (IWWOST), in the framework of ECOOP 2002, pp. 120-133, Malaga, Spain.
6. Olsina, L.; González Rodríguez, J.; Lafuente, G.J.; Pastor, O., 2001, "*Towards Automated Web Metrics*", VIII Quality Brazilian Workshop, RJ-Br, pp. 74-86
7. Olsina, L.; Lafuente, G.J.; Pastor, O., 2002, "*Designing a Catalogue for Metrics*", 2nd Ibero-American Conference on Web Engineering (ICWE'02) in the framework of 31 JAIIO, Sta Fe, Argentina, pp. 108-122, ISSN 1666-6526.
8. W3C, WWW Consortium, 2001, *Semantic Web*, <http://www.w3.org/2001/sw/> (in this page there are different resources about RDF and Semantic Web)
9. W3C, WWW Consortium, 2002, *RDF Primer*, W3C Working Draft 19 March 2002 <http://www.w3.org/TR/2002/WD-rdf-primer-20020319/>
10. W3C, WWW Consortium, 2002, *RDF Model Theory*, W3C Working Draft 29 April 2002 <http://www.w3.org/TR/rdf-nt/>
11. Zuse, H., 1998, *A Framework of Software Measurement*, Walter de Gruyter, Berlín-NY.