

Ontología para Criterios de Clasificación *

Mónica Martínez¹, Regina Motz¹

¹Instituto de Computación – Universidad de la República(UdelaR)
Montevideo – Uruguay

{mmartine, rmotz}@fing.edu.uy

***Abstract.** Almost in any area of interest, it eventually becomes interesting to classify the elements of a certain domain according to a set of criteria, either to obtain how each element behaves in terms of a criterion or to compare the behaviour of different elements. The use of an ontology for modelling the classification criterion allows to verify the definition criterion consistency and moreover, to classify instances of the domain. This work provides some guidelines for the development of such kind of ontology that can be used for classification purposes. The work discusses ontology design alternatives and presents the application of the classification ontology to determine document quality.*

***Resumen.** Casi en cualquier ámbito, en algún momento resulta de interés clasificar los elementos del dominio en que se está trabajando según un conjunto de criterios, tanto para obtener cómo se comporta cada elemento frente a un criterio cómo para poder comparar el comportamiento de distintos elementos entre sí. La utilización de una ontología para modelar los criterios de clasificación permite verificar la consistencia de la definición de los criterios y además realizar la clasificación de las instancias del dominio. Este trabajo presenta guías para el diseño de una ontología capaz de ser explotada en estos sentidos para tareas de clasificación. Se discuten alternativas de diseño y se presenta la aplicación de la utilización de la ontología en la determinación del grado de calidad de un documento.*

1. Introducción

La clasificación es una tarea muy común en la mayoría de los sistemas de información actuales. En especial se puede observar su utilización para determinar perfiles de usuarios en sistemas de personalización, donde se clasifican las características personales; para evaluar calidad de datos, donde se clasifican los valores que determinan la calidad; o en general para determinar pertenencia a un dominio.

La utilización de ontologías en sistemas de información generalmente están orientadas a especificar y comunicar el conocimiento del dominio de una manera genérica y para estructurar y definir el significado de los términos. Sin embargo, raramente se ven explotados los servicios de razonamiento de las ontologías en los sistemas de información.

Este trabajo presenta pautas para modelar una ontología utilizable para explotar su servicio de razonamiento en la clasificación de elementos de un dominio. El aporte del trabajo es brindar guías de diseño genéricas para el modelado de la ontología. La

*Parcialmente financiado por el Proyecto DocWebQuali Prosul Edital CNPq No.040/2005

construcción de la ontología se basa en los trabajos de [Noy and McGuinness 2001] y de [Óscar Corcho et al. 2003] pero se caracteriza por incorporar criterios de clasificación del área de ingeniería de software directamente en el modelado. Se muestra además su aplicación en la determinación de la calidad de documentos.

El resto del trabajo está organizado de la siguiente forma. En la Sección 2 se presentan criterios genéricos para la clasificación. En la Sección 3 se dan las guías de diseño de la ontología para modelar los criterios de clasificación y se argumenta sobre el alcance y limitaciones del marco teórico. En la Sección 4 se muestra la aplicación directa de las guías de diseño para la clasificación de documentos digitales en la determinación de su calidad. Finalmente, en la Sección 5 se presentan algunas conclusiones y líneas de trabajos futuros.

2. Criterios de Clasificación

El primer paso para realizar una clasificación consiste en establecer cuáles son las variables que se desean medir y de qué forma se realizará la medición.

Basándonos en la teoría de la medida propuesta por [Fenton 1991] encontramos cuatro formas de medida útiles para la clasificación:

- **Forma Nominal:** da un nombre o una etiqueta como valor a un atributo repartiéndolo a los individuos en *categorías*. Estos valores permiten identificar a los individuos como pertenecientes a una y una sola de las categorías. Por ejemplo para una persona tienen medida nominal los atributos color de cabellos (*morocho, rubio, castaño*) y estado civil (*soltero, casado, viudo, divorciado*).
- **Forma Ordinal:** reparte a los individuos también en categorías, pero a diferencia de la forma nominal, las categorías son comparables, permitiendo establecer un orden creciente o decreciente. Por ejemplo, las categorías *muy bajo, bajo, medio, alto, muy alto*.
- **Forma de Intervalos:** implica la obtención de medidas numéricas y la diferencia entre dos valores de la variable. En este caso es de interés determinar para cada elemento del dominio el grado de cumplimiento de una regla o condición. Por ejemplo en el caso de estar evaluando documentos es interesante determinar el porcentaje de sus referencias bibliográficas correspondientes a documentos recientes.
- **Forma Funcional:** en este caso es de interés determinar para cada elemento del dominio un valor que es función de los valores que tiene el elemento en un conjunto de propiedades. Ciertas combinaciones de valores en determinadas propiedades son más relevantes que las propiedades consideradas en forma individual.

A cada una de estas formas de medidas le asociamos un criterio de clasificación. En particular para las formas nominal y ordinal utilizamos un criterio de clasificación que llamamos *Criterio de Valor*, ya que en ambos casos es de interés determinar para cada elemento del dominio el valor que tiene con respecto a una propiedad. Por ejemplo el grado académico del autor de un documento (nominal) o el tipo de reconocimiento que tiene una institución (ordinal) lo trataremos en forma general como clasificación por criterio de valor. A la forma de medida de intervalos le asociamos el criterio que llamamos *Criterio de Afinidad* y a la forma de medida funcional le asociamos un criterio de clasificación que

llamamos *Criterio de Funcionalidad*. Además de tener los elementos clasificados según cada criterio en particular en la mayoría de los casos resulta muy útil tener una clasificación globalizadora de todos los criterios, que llamamos *Criterio de Globalización*. Para esta clasificación se considera la clasificación en cada uno de los criterios individuales y se los reclasifica según el resultado obtenido en cada uno de ellos y el peso que se le otorgue a cada uno.

3. Definición de la Ontología

Para diseñar una ontología útil a la clasificación según los criterios presentados en la sección anterior sugerimos realizar las siguientes definiciones:

- Una clase (llamada DOMINIO) que representa al conjunto de los elementos a clasificar.
- Propiedades con dominio en esta clase para representar las características de estos elementos según sean necesarias para cada tipo de criterio.
- Clases que representan a cada uno de los criterios de clasificación considerados(DomCritN) y subclases de estas que representan la partición de los elementos según cada criterio.
- Una clase que representa la clasificación global (DomGlobal) con subclases que representan cada uno de los posibles resultados.

A continuación presentamos para cada tipo de criterio las definiciones particulares que corresponde realizar. En todo momento se prioriza el trabajar con el lenguaje de definición de ontologías OWL-DL, de forma de garantizar la computabilidad al utilizar los razonadores disponibles hoy en día.

3.1. Modelado del Criterio de Afinidad

En este tipo de criterio se apunta a la clasificación de los elementos según el grado de cumplimiento de una regla o condición. Por lo tanto será deseable poder definir una clase que represente la condición y luego especificar la pertenencia de un elemento a este concepto indicando un grado de pertenencia. Pensar en el grado de pertenencia lleva a la búsqueda de ontologías difusas. Sin embargo, las extensiones encontradas que trabajan con lógicas difusas lo realizan en la jerarquía de las clases pero no en la relación de pertenencia de un individuo a una clase [T.T.Quan et al. 2004]. Por lo tanto, en un diseño sencillo la opción de utilizar ontologías difusas las descartamos ya que no es posible fácilmente reflejar el criterio sin realizar modificaciones al marco teórico para el tratamiento de pertenencia difusa a una clase. En su lugar optamos por la definición de rangos. Estos rangos surgen del análisis de que en la gran mayoría de los casos el grado de pertenencia puede ser medido en porcentajes a partir de cálculos matemáticos sobre valores en determinadas propiedades bases que son conocidas en la realidad que se está trabajando. Por ejemplo el grado de las referencias bibliográficas recientes de un documento surge a partir de la cantidad total de referencias del documento y la cantidad de las mismas consideradas recientes. Una opción que aparece entonces como válida para modelar este tipo de criterio es la de representar estas propiedades bases y operar sobre ellas. Sin embargo, este tipo de manipulaciones no son posibles en OWL-DL. Para poder realizarlo sería necesario trabajar con extensiones de OWL basadas en las extensiones de DL tipo *Concrete Domains*, donde más allá de tener *data type properties* es posible tener predicados definidos sobre los tipos concretos [Baader et al. 2003]. Por lo tanto en

un diseño sencillo esta forma de modelar el criterio debe ser descartada. Optamos entonces, por modelar el criterio a través de la definición de rangos, conjunto de valores de las propiedades bases y clasificar los elementos según si todas sus componentes tienen sus correspondientes en las propiedades bases en un determinado rango. Definiendo los rangos de una forma inteligente se puede obtener una aproximación interesante a lo que se pretendía originalmente. De esta forma se obtiene una clase de la ontología que representa a todos los elementos del dominio: DOMINIO y subclases de esta (DOM_RANGO1 ... DOM_RANGO_N) formada por los elementos de DOMINIO que tienen todos sus correspondientes de la propiedad base (PROP_BASE) en un rango determinado. Se deben definir además de estas subclases las clases que representan estos rangos y la propiedad PROP_BASE con dominio en DOMINIO y recorrido en la unión de todos los RANGOS definidos. La especificación de cada una de las subclases de DOMINIO queda entonces de la siguiente forma:

$$\text{DOM_RANGO}_i \equiv \text{DOMINIO} \sqcap \forall \text{PROP_BASE } \text{RANGO}_i \sqcap \exists \text{PROP_BASE } \text{RANGO}_i$$

Para realizar ahora la clasificación de un elemento utilizando este modelo de ontología es necesario seguir los siguientes pasos:

- Definir un individuo de la clase DOMINIO: e_1 , que representa al elemento que se pretende clasificar.
- Definir para e_1 sus correspondientes en la PROP_BASE
- Determinar el tipo del elemento e_1

Observar que en el caso de definir un elemento donde todas los correspondientes que se le especifican en la propiedad base pertenezcan a un único rango y por lo tanto deba ser un elemento de una subclase particular, al momento de calcular el tipo del elemento con algún razonador como Racer [rac] o Pellet [pel] se obtendrá DOMINIO y no la subclase esperada. El motivo de esta diferencia radica en que el esperar la pertenencia del elemento a la subclase esta causado por estar trabajando en una hipótesis de mundo cerrado y por lo contrario los razonadores trabajan con el supuesto de mundo abierto. Entonces a pesar que las definiciones dadas cumplen que todos los correspondientes de un elemento pertenecen a un rango determinado no está especificado que no existan otros correspondientes que eventualmente no pertenezcan al rango, por lo tanto no es posible inferir que el elemento pertenece a la subclase. Para lograr esto sería necesario cerrar la definición de la propiedad, especificar que los únicos correspondientes de un elemento son los dados explícitamente, lo que se llama cerrar la relación. Esto no es posible realizarlo en una forma directa y sencilla en OWL. Una posible forma de especificar este tipo de restricciones requiere de la reificación de la propiedad sobre la que se están realizando cuantificaciones universales, pero el estudio de esta posibilidad no se encuentra dentro de los alcances de este trabajo. Por lo tanto es necesario buscar otra forma de modelarlo y es en ese sentido que optamos por simplificar la ontología modelándola sin restricciones del tipo cuantificador universal, lo que trae como consecuencia el modelar utilizando únicamente propiedades funcionales. Para mantener el objetivo fundamental de este tipo de criterios se pasa la determinación del porcentaje de cumplimiento de una condición a funcionalidades externas a la ontología. En la ontología únicamente se modela para cada elemento un valor que representa el rango dentro del cual se encuentra su porcentaje de cumplimiento del criterio (de esta forma se evita el trabajar con valores continuos).

Se define entonces una clase RANGOS con individuos: *alto*, *tiendeAlto*,

medioAlto, *medioBajo*, *Bajo* y *Nulo*, donde cada uno de ellos representan los siguientes criterios de rangos:¹ Alto corresponde a más del 90%, *tiendeAlto* del 70 a 90%, *medioAlto* del 50 a 70%, *medioBajo* del 25 a 50%, *Bajo* del 0 a 25%, *Nulo* corresponde a que no hay correspondientes para la determinación de los porcentajes.

Se define la propiedad *CAProp*: $\text{DOMINIO} \rightarrow \text{RANGOS}$, funcional que representa la propiedad sobre la cual se está evaluando el porcentaje de cumplimiento. Se define una clase *DomCA* con subclases para representar al conjunto de los elementos que tienen como correspondiente en la propiedad cada uno de estos valores. De esta forma se obtienen las clases:

- $\text{ValorCA-alto} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{Alto}$
- $\text{ValorCA-Talto} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{tiendeAlto}$
- $\text{ValorCA-mAlto} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{medioAlto}$
- $\text{ValorCA-mBajo} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{medioBajo}$
- $\text{ValorCA-bajo} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{Bajo}$
- $\text{ValorCA-nulo} \equiv \text{DOMINIO} \sqcap \text{CAProp} \ni \text{Nulo}$

3.2. Modelado del Criterio de Valor

Con este tipo de criterio interesa clasificar los elementos según los valores que tienen en una determinada propiedad. En este caso se define una clase que representa el recorrido de la propiedad (*DISCRIMINADOR*), en la cual se definen individuos que representan a cada uno de los elementos de este conjunto (individuos distintos entre sí). Como solo se puede representar un conjunto finito de individuos, en caso de que el recorrido de la propiedad no lo sea se deberá buscar una forma de representarlo mediante un conjunto finito (por ejemplo definiendo valores que representan rangos de valores).

Se define entonces una propiedad *CVProp* con dominio en la clase *DOMINIO* y recorrido en la clase *DISCRIMINADOR*. Teniendo en cuenta todas las consideraciones realizadas en el análisis del criterio de afinidad esta propiedad debe ser funcional. Por lo tanto si la propiedad que se esta representando no es funcional se deberá buscar una forma alternativa de representarla para que cumpla con esta característica. Por último se define una clase *DomCV* con tantas subclases como la cardinalidad del recorrido de la propiedad (o del conjunto que lo representa), cada una de las cuales corresponde a los elementos del dominio que tienen como correspondiente el valor que representa la subclase en la propiedad considerada. A modo de ejemplo consideremos una propiedad con tres posibles valores (*valor1*, *valor2*, *valor3*). Se define la propiedad *CVProp* como se describió y se definen las clases:

- $\text{ValorCV-v1} \equiv \text{DOMINIO} \sqcap \text{CVProp} \ni \text{valor1}$
- $\text{ValorCV-v2} \equiv \text{DOMINIO} \sqcap \text{CVProp} \ni \text{valor2}$
- $\text{ValorCV-v3} \equiv \text{DOMINIO} \sqcap \text{CVProp} \ni \text{valor3}$

3.3. Modelado del Criterio de Funcionalidad

Con este tipo de criterio se busca clasificar los elementos del dominio en función de la combinación de valores que le corresponde a cada elemento en un conjunto de

¹En todos los casos se considera que el valor alto corresponde al mejor nivel en el criterio, en caso de que no sea lo mejor cumplir en un alto porcentaje, para mantener la uniformidad será necesario plantear la propiedad en forma inversa.

propiedades. En este caso se realizan definiciones análogas a las realizadas en el criterio de valor para representar cada una de las propiedades. A cada combinación de valores se le asocia un valor o ponderación, que llamamos nivel (el conjunto de estos valores debe cumplir con las mismas características que el conjunto recorrido de la propiedad en el criterio de valor). Al igual entonces que en el criterio de valor se define una subclase de DOMINIO para representar a cada subconjunto de elementos del dominio que tienen cada uno de los niveles asociados según su combinación de valores en las propiedades involucradas. A modo de ejemplo consideremos un criterio en el cual interesa ponderar las combinaciones de valores de dos propiedades. Se definen entonces dos clases RecorridoProp1CF y RecorridoProp2CF que representan los conjuntos recorridos de las propiedades involucradas. En cada una de estas clases se definen los individuos que representan a los elementos de estos conjuntos. Suponiendo que los elementos de RecorridoProp1CF son: val1prop1 y val2prop1; y los valores de RecorridoProp2CF son: val1prop2, val2prop2 y val3prop2, definimos las siguientes propiedades, ambas funcionales: Prop1CF con dominio en DOMINIO y con recorrido en RecorridoProp1CF y Prop2CF con dominio en DOMINIO y con recorrido en RecorridoProp2CF. Se establece luego la relación entre las distintas combinaciones de valores y el conjunto que representa los niveles. Las combinaciones para este ejemplo se presentan en la tabla 1.

Table 1. Definición de niveles

Prop1CF	Prop2CF	Nivel	Prop1CF	Prop2CF	Nivel
val1prop1	val1prop2	1	val2prop1	val1prop2	4
val1prop1	val2prop2	2	val2prop1	val2prop2	5
val1prop1	val3prop2	3	val2prop1	val3prop2	5

A partir de estos valores se define una clase DomCF con subclases que representan a cada uno de los subconjuntos del dominio que tienen una combinación de valores en las propiedades que lo hacen pertenecer a cada nivel.

- CF-Nivel1 \equiv DOMINIO \sqcap Prop1CF \ni val1prop1 \sqcap Prop2CF \ni val1prop2
- CF-Nivel2 \equiv DOMINIO \sqcap Prop1CF \ni val1prop1 \sqcap Prop2CF \ni val2prop2
- CF-Nivel3 \equiv DOMINIO \sqcap Prop1CF \ni val1prop1 \sqcap Prop2CF \ni val3prop2
- CF-Nivel4 \equiv DOMINIO \sqcap Prop1CF \ni val2prop1 \sqcap Prop2CF \ni val1prop2
- CF-Nivel5 \equiv DOMINIO \sqcap Prop1CF \ni val2prop1 \sqcap (Prop2CF \ni val2prop2 \sqcup Prop2CF \ni val3prop2)

3.4. Modelado del Criterio de Globalidad

Para clasificar utilizando el criterio de globalidad proponemos agrupar los posibles resultados de cada criterio en diferentes categorías disjuntas, una propuesta es utilizar las siguientes tres categorías: ALTA, MEDIA, BAJA, de tal forma que pertenecen a la categoría ALTA aquellos resultados que se consideran de mayor valor en el criterio, a la categoría BAJA los de menor valor y a la MEDIA los restantes.² Luego que se tienen todas las categorías definidas para cada criterio se debe definir aquellos elementos que se consideran BUENOS, MALOS Y MEDIOS. Una opción es considerar como BUENOS a

²Se deben considerar en forma particular aquellos criterios que no tienen por lo menos tres posibles resultados, por ej. los criterios de valor donde el recorrido de las propiedades es booleano.

aquellos elementos tales que clasifican en todos los criterios en niveles correspondientes a Alto, MALOS a aquellos que clasifican en todos los criterios en nivel Bajo y MEDIOS los restantes. Estas definiciones pueden variarse para cada caso particular sin mayores dificultades de forma de reflejar adecuadamente la ponderación entre los diferentes criterios que se considera relevante en cada caso particular. En la ontología es necesario definir subclases de las clases definidas para cada criterio que representan a las categorías definidas. Por ejemplo para el criterio de afinidad se definen las clases:

- DomCA-Alta \equiv ValorCA-alto \sqcup ValorCA-Talto
- DomCA-Medio \equiv ValorCA-mAlto \sqcup ValorCA-mBajo
- DomCA-Bajo \equiv ValorCA-bajo \sqcup ValorCA-nulo
- DomCA \equiv DomCA-Alta \sqcup DomCA-Medio \sqcup DomCA-Bajo

Finalmente se define una clase DominioGlobal que tiene como subclases las clases que representan a los elementos clasificados como Buenos, Malos y Medio.

DominioGlobal \equiv ClasifGlobalBUENO \sqcup ClasifGlobalMEDIO \sqcup ClasifGlobalMALO

ClasifGlobalBUENO \equiv DomCA-Alta \sqcap DomCV-Alta \sqcap DomCF-Alta

ClasifGlobalMALO \equiv DomCA-Bajo \sqcap DomCV-Bajo \sqcap DomCF-Bajo

La definición de la clase correspondiente a ClasifGlobalMEDIO es posible realizarla como el complemento de *ClasifGlobalBUENO* \sqcup *ClasifGlobalMALO* pero se optó por definir esta clase por enumeración de todos los posibles casos ya que de esta forma en algunos casos es posible obtener la clasificación de elementos del dominio para los cuales no se conocen los datos suficientes para clasificarlos según todos los criterios definidos.

3.5. Uso de la Ontología para Clasificación

Para clasificar un elemento en particular se debe crear un individuo en la clase DOMINIO, asignarle los valores correspondientes a todas las funciones definidas para esta clase e inferir el tipo del individuo. El razonador listará los nombres de las clases a las cuales pertenece el elemento. En la clase ClasifGlobalX se observa la clasificación global del elemento y en las clases DomCAX, DomCVX y DomCFX, se observa la clasificación obtenida por el elemento en cada criterio. En caso de no disponer de la información requerida para alguno de los criterios no se clasifica al elemento en ese criterio y se obtiene una clasificación global si con la información disponible es posible inferirla. Otra forma de usar clasificar consiste en ingresar todos los elementos que se deseen en la clase DOMINIO con todas sus correspondientes en las funciones definidas para la clase y luego dirigirse a la clase que se desea y determinar todos los elementos que pertenecen a la clase.

4. Ejemplo de clasificación para determinar la calidad de un documento

La clasificación de elementos de un dominio según un determinado conjunto de criterios tiene diversas aplicaciones en muy variadas áreas. A modo de ejemplo se muestran los pasos anteriormente descritos para el caso de aplicar clasificación de documentos para determinar un nivel de calidad de los mismos.

Es posible considerar una amplia variedad de criterios para la clasificación de documentos según los aspectos de los mismos que se deseen resaltar. Para este ejemplo

seleccionamos los siguientes criterios: (Criterio 1) el tipo de las publicaciones anteriores que tiene el autor y (Criterio 2) el nivel académico del autor del documento, considerando su formación, la existencia de publicaciones anteriores, etc.

Cada documento es clasificado según estos criterios y finalmente se realiza una clasificación global donde se obtiene una clasificación en Bueno, Medio o Malo basada en las clasificaciones en cada criterio ponderándolos de igual forma. A continuación se detalla el proceso seguido en el modelado/clasificación en cada uno de los criterios.

Modelado del Criterio 1: En este criterio se valora que el autor del documento tenga publicaciones realizadas anteriormente y el tipo de la misma. Para esta clasificación se utiliza una función que asocia a cada documento el tipo de mayor nivel de las publicaciones anteriores del autor (TipoPublicacionAntAutor). Se considera que son valores de clasificación alto capítulo de libro y revista; como bajo el valor otro y noPub quedando revista como valor medio. Este criterio corresponde a un criterio de valor, por lo tanto se realizan las siguientes definiciones en la ontología:

- TipoPublicacion $\equiv \{\text{capLibro, revista, conf, otro, noPub}\}$
- La propiedad TipoPublicacionAntAutor funcional, con dominio en Documento y recorrido en TipoPublicacion.
- Las clases correspondiente a cada valor posible del recorrido de la propiedad. Al agrupar en las categorías se consideran valores de clasificación altos capítulo de libro y revista; como bajo el valor otro y noPub, quedando conferencia como valor medio.

DocCrit1 $\equiv \text{Doc1Media} \sqcup \text{Doc1Alta} \sqcup \text{Doc1Baja}$

Doc1Alta $\equiv \text{DocAutoresCapLibro} \sqcup \text{DocAutoresRevistas}$

DocAutoresCapLibro $\equiv \text{Documentos} \sqcap \text{TipoPublicacionAntAutor} \ni \text{capLibro}$

DocAutoresRevistas $\equiv \text{Documentos} \sqcap \text{TipoPublicacionAntAutor} \ni \text{revista}$

Doc1Media $\equiv \text{DocAutoresConferencia}$

DocAutoresConferencia $\equiv \text{Documentos} \sqcap \text{TipoPublicacionAntAutor} \ni \text{conf}$

Doc1Baja $\equiv \text{DocAutoresIniciales} \sqcup \text{DocAutoresOtro}$

DocAutoresIniciales $\equiv \text{Documentos} \sqcap \text{TipoPublicacionAntAutor} \ni \text{noPub}$

DocAutoresOtro $\equiv \text{Documentos} \sqcap \text{TipoPublicacionAntAutor} \ni \text{otro}$

Modelado del Criterio 2: En este criterio se valora al autor del documento desde el punto de vista de su grado académico, si ha sido evaluador y si el documento considerado es el primero en el tema (no hay referencias en el documento a otros documentos del mismo autor). Para la clasificación según este criterio de funcionalidad se definen tres funciones:

GradoAutor relaciona el documento con el mayor grado académico del autor.

ReferenciaAlAutor relaciona el documento con la indicación de si existen referencias a documentos del mismo autor.

EvaluadorAutor relaciona el documento con la indicación de si el autor ha sido evaluador anteriormente.

Los valores de estas funciones se combinan según la tabla 2 para definir los niveles de los autores. Para el modelado de la ontología se realizan las siguientes definiciones:

- Las clases que representan los recorridos de las propiedades consideradas:
 - Grados** $\equiv \{\text{grado, posgrado, estudiante, nada}\}$
 - Boolean** $\equiv \{\text{si,no}\}$

Table 2. Definición de niveles

Grado	RefAutor	Evaluador	Nivel	Grado	RefAutor	Evaluador	Nivel
grado	si	si	1	grado	no	si	3
posgrado	si	no	1	posgrado	no	no	3
posgrado	si	si	1	nada	si	no	4
estudiante	si	si	2	estudiante	no	si	4
grado	si	no	2	grado	no	no	4
posgrado	no	si	2	nada	no	no	5
nada	si	si	3	nada	no	si	5
estudiante	si	no	3	estudiante	no	no	5

- Las propiedades que representan las propiedades base del criterio:
GradoAutor funcional, con dominio en Documento y recorrido en Grados.
ReferenciaAlAutor funcional, con dominio en Documento y recorrido en Boolean.
EvaluadorAutor funcional, con dominio en Documento y recorrido en Boolean.
- Las clases correspondientes a cada uno de los posibles niveles de documento. Al agrupar en las categorías se consideran como valores de clasificación ALTO los niveles 1 y 2; BAJO a los niveles 4 y 5; correspondiendo MEDIO al nivel 3.

DocCrit2 \equiv Doc2Media \sqcup Doc2Alta \sqcup Doc2Baja

Doc2Alta \equiv DocAutoresN1 \sqcup DocAutoresN2

DocAutoresN1 \equiv DocAutN11 \sqcup DocAutN12

DocAutN11 \equiv Documentos \sqcap EvaluadorAutor \ni si \sqcap GradoAutor \ni grado \sqcap ReferenciaAlAutor \ni si

DocAutN12 \equiv Documentos \sqcap GradoAutor \ni posgrado \sqcap ReferenciaAlAutor \ni si

DocAutoresN2 \equiv DocAutN21 \sqcup DocAutN22 \sqcup DocAutN23

Doc2Media \equiv DocAutoresN3

DocAutoresN3 \equiv DocAutN31 \sqcup DocAutN32 \sqcup DocAutN33 \sqcup DocAutN34

Doc2Baja \equiv DocAutoresN4 \sqcup DocAutoresN5

DocAutoresN4 \equiv DocAutN41 \sqcup DocAutN42 \sqcup DocAutN43

DocAutoresN5 \equiv DocAutN51 \sqcup DocAutN52 \sqcup DocAutN53

Criterio de Globalización: Por último se realizan las definiciones correspondientes al criterio de globalización que en este caso se consideran de la siguiente forma: **Buenos** aquellos documentos que en todos los criterios clasifican en niveles correspondientes al grupo alto; **Malos** aquellos documentos que en todos los criterios clasifican en niveles correspondientes al grupo bajo, **Medio** los restantes documentos.

La clasificación de los documentos Una vez definida la ontología se procede a la clasificación de los elementos. Como se mencionó anteriormente se crean individuos en la clase Documento que representan los documentos a clasificar y se procede a inferir el tipo de cada uno de ellos o a determinar los elementos de alguna clase de interés (por ej. Clasif-GlobalBUENO). Al definir las funciones que se utilizan en cada uno de los criterios se pensó en funciones que se pudieran calcular en función de la información disponible en el documento en sí mismo o que existan base de datos públicas que se puedan consultar.

Una versión extendida de este ejemplo se encuentra en el reporte [Martínez 2006]

5. Conclusiones y Trabajos Futuros

En este trabajo presentamos pautas para el desarrollo de una ontología que modela los criterios de clasificación de elementos de un dominio. Basándonos en la teoría de las medidas definimos tres criterios básicos de clasificación (Criterio de Afinidad, de Valor y de Funcionalidad) y un cuarto criterio de globalidad para la combinación de los tres anteriores.

Las soluciones adoptadas para modelar cada uno de los criterios de clasificación en la ontología es ampliamente argumentada y discutida. De este análisis surgen las siguientes líneas de trabajos futuros: (i) Realizar una validación experimental de la propuesta. Esto se realizará en marco del proyecto DocWebQuali Prosul Edital CNPq No.040/2005 (ii) Estudiar la posibilidad de representar el mundo cerrado de una forma más adecuada. Como se mencionó anteriormente se considera que un posible camino en esta dirección puede ser el reificar las propiedades de forma de tener mayor poder de expresión sobre ellas, (iii) Considerar extensiones de OWL que en estos momentos se encuentran en estudio y analizar cómo impactaría su uso en la representación de los criterios y las propiedades de los elementos.

References

- Pellet OWL Reasoner*. <http://www.mindswap.org/2003/pellet/>(último acceso : diciembre 2005).
- Racer OWL Reasoner*. <http://www.racer-systems.com> (último acceso : diciembre 2005).
- Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F. (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.
- Fenton, N. E. (1991). *Software Metrics: A Rigorous Approach*. Chapman and Hall, London.
- Martínez, M. (2006). Algunas experiencias en el uso de ontologías para la clasificación de documentos. Reporte Técnico RT 06-07, Facultad de Ingeniería Instituto de Computación - PEDECIBA Informática.
- Noy, N. F. and McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory.
- Óscar Corcho, Fernández-Lopez, M., and Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies: Where is their meeting point? *Data Knowl.Eng.*, 46(1):41–64.
- T.T.Quan, S.C.Hui, and T.H.Cao (2004). Foga: A fuzzy ontology generation framework for scholarly semantic web. *Workshop 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*. <http://olp.dfki.de/pkdd04/quan-final.pdf>. (último acceso : diciembre 2005).