PEDECIBA Informática
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay

Reporte Técnico RT 06-07

Algunas experiencias en el uso de ontologías para la clasificación de documentos

Mónica Martínez Amarante

2006

Algunas experiencias en el uso de ontologías para la clasificación de documentos
Martínez Amarante, Mónica
ISSN 0797-6410
Reporte Técnico RT 06-07
PEDECIBA
Instituto de Computación – Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay, 2006

Algunas experiencias en el uso de ontologías para la clasificación de documentos

Mónica Martínez
Instituto de Computación (InCo), Facultad de Ingeniería, Universidad de la Republica, Montevideo, Uruguay

mmartine@fing.edu.uy

Resumen: En el presente trabajo se presenta el desarrollo de una herramienta basada en el uso de ontologías que permite la clasificación de documentos a partir de un conjunto de criterios de calidad. En un primer paso se definen los criterios a utilizar, posteriormente se muestran los pasos seguidos en la construcción de la ontología para cada uno de ellos.
Finalmente se obtiene una clasificación resumen de los documentos considerando los criterios definidos en forma global.

Palabras claves: Ontologías, Clasificación de documentos, Protégé.

1. Introducción

Hoy en día existe una gran cantidad de documentos disponibles (fundamentalmente por la web) cuya clasificación resulta difícil, por lo que sería deseable contar con algún mecanismo que permita clasificarlos.
Existen documentos que por el hecho de haber sido publicados en una revista o conferencia ya dan ciertas garantías de calidad. Sin embargo existen muchos documentos que no se encuentran en estas situaciones pero no por eso dejan de tener valor, por ejemplo tesis de doctorado, reportes internos, etc., pero si es necesario distinguirlos de otros no tan buenos.

El presente trabajo tiene como objetivo definir algún mecanismo que permita la clasificación de los documentos.
Para ello se define una ontología ya que tiene funcionalidades que permiten modelar las características de los documentos y también cuenta con mecanismos que permiten inferir el tipo de un documento a partir de las características modeladas.

En una primera sección se presentan criterios que podrían utilizarse para la clasificación de los documentos.
Luego se seleccionan algunos de ellos que se toman como ejemplo para realizar la ontología de clasificación. Con ellos se muestran las ventajas y limitaciones que se presentan al momento del modelado y clasificación.
Finalmente se muestra cómo se debe utilizar la ontología para efectivamente realizar la clasificación.

Las herramientas utilizadas en el desarrollo son: Protégé 3.0 [Horr *et al* 04] como editor de ontologías y racer 1.7.23 y pellet 1.3. como razonadores.

2. Criterios de clasificación

Existen muchos criterios posibles para la clasificación de los documentos y dependiendo del ámbito y la actividad que se esté realizando algunos de ellos toman una relevancia mayor con respecto a otros.
A continuación se enumeran algunos posibles criterios pero en cada caso se deberá analizar si estos criterios planteados son los de interés y en todo caso modificarlos.

1. Con respecto a la bibliografía, a las referencias del documento
    a. Considerando las fechas de las referencias
       Se consideran mejores los documentos que tienen referencias a artículos recientes con respecto a la fecha del documento.
    b. Considerando las referencias a grupos externos al grupo del autor del documento.
       Se consideran mejores los documentos que tienen referencias a grupos de trabajos distintos a los autores
    c. Considerando las referencias al autor
       Se puede considerar un trabajo inicial en el tema por parte del autor si no tiene referencias a sí mismo, de lo contrario es un trabajo que ya se viene desarrollando desde hace un tiempo, por lo tanto se da mayor valor al documento
    d. Con respecto al tipo de las referencias.
       Si corresponden a publicaciones con valor (capítulos de libro, revistas, conferencias, etc.).

2. Con respecto a la fecha del documento
    a. Se considera mejor si es un trabajo reciente (Aunque esto depende un poco de la temática del articulo, por ejemplo en artículos sobre fundamentos puede no ser tan importante como si es de aspectos tecnológicos)

3. Con respecto a la estructura
    a. Resumen (abstract)
       Se valora la existencia ya que permite identificar el artículo y valorar si tiene sentido leerlo
    b. Conclusiones
       Se valora su existencia, aunque la aparición del título conclusiones no es mucha garantía
4. Con respecto al autor
    a. Se valora si el autor tiene alguna publicación anterior y el tipo de la misma.
    b. Se valora el grado académico del autor
    c. Se valora si el autor ha sido evaluador en conferencias, revistas.

3. Desarrollo de la herramienta

Como fue mencionado anteriormente se construyó una ontología que permite la clasificación de los documentos.

De los criterios antes mencionados se consideraron (a modo de ejemplo) los siguientes criterios para la creación de la ontología.

Criterio 1
: Según las fechas de los documentos referenciados relativas a la fecha del documento considerado. (1.a)

Criterio 2
: Según las referencias a instituciones externas a la institución del autor del documento. (1.b)

Criterio 3
: Según el tipo de las publicaciones anteriores que tiene el autor del documento. (4.a)

Criterio 4
: Según las características del autor del documento desde el punto de vista de su grado académico (4.b), si ha sido evaluador (4.c) y si se esta iniciando en el tema del documento (1.c).

Cada documento es clasificado según estos criterios y finalmente se realiza una clasificación a modo de resumen 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.

3.1 Criterio 1

En general se considera como una buena característica que las referencias utilizadas sean relativamente nuevas al momento de referenciarlas (por lo menos en una alta proporción).

El primer punto es definir qué se entiende por nuevas. Considerando los tiempos que se manejan (por lo menos en computación) para la presentación de un artículo, su aceptación y su publicación en una revista, por ejemplo, se optó por manejar un rango de 0 a 5 años como recientes.

Lo que se busca en este caso es, a partir de la fecha del documento a analizar y las fechas de cada uno de los documentos referenciados, compararlos y determinar si son relativamente recientes.

Para lograr esto es necesario una vez asociado a cada documento su fecha de creación y las fechas de las referencias, poder contar con predicados que permitan comparar valores, contar cuántos correspondientes tienen determinada característica (diferencia de fechas), para poder calcular porcentajes.
Este tipo de cosas no son posibles en OWL ya que no lo son en Description Logic (marco formal de OWL). 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 sea posible tener predicados definidos sobre los tipos concretos [Baad *et al.* 03].
Por lo tanto esta forma de trabajar con este criterio debe ser descartada por no contar con las propiedades requeridas.

Como alternativa se pensó en tener conceptos que representen distintos rangos para las diferencias de fechas relativas.

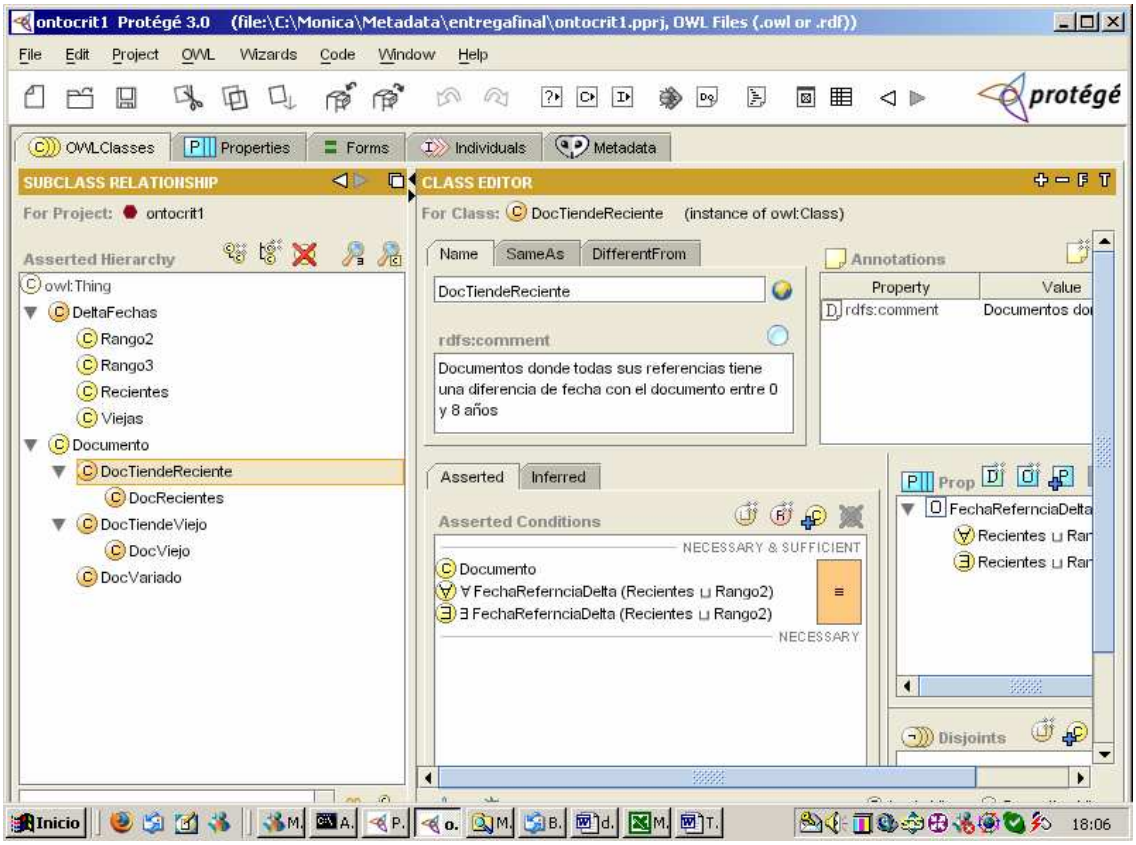
Recientes (0 a 5 años)
Rango 2 (6 a 8 años)
Rango 3 (9 a 10 años)
Viejas (más de 10 años)

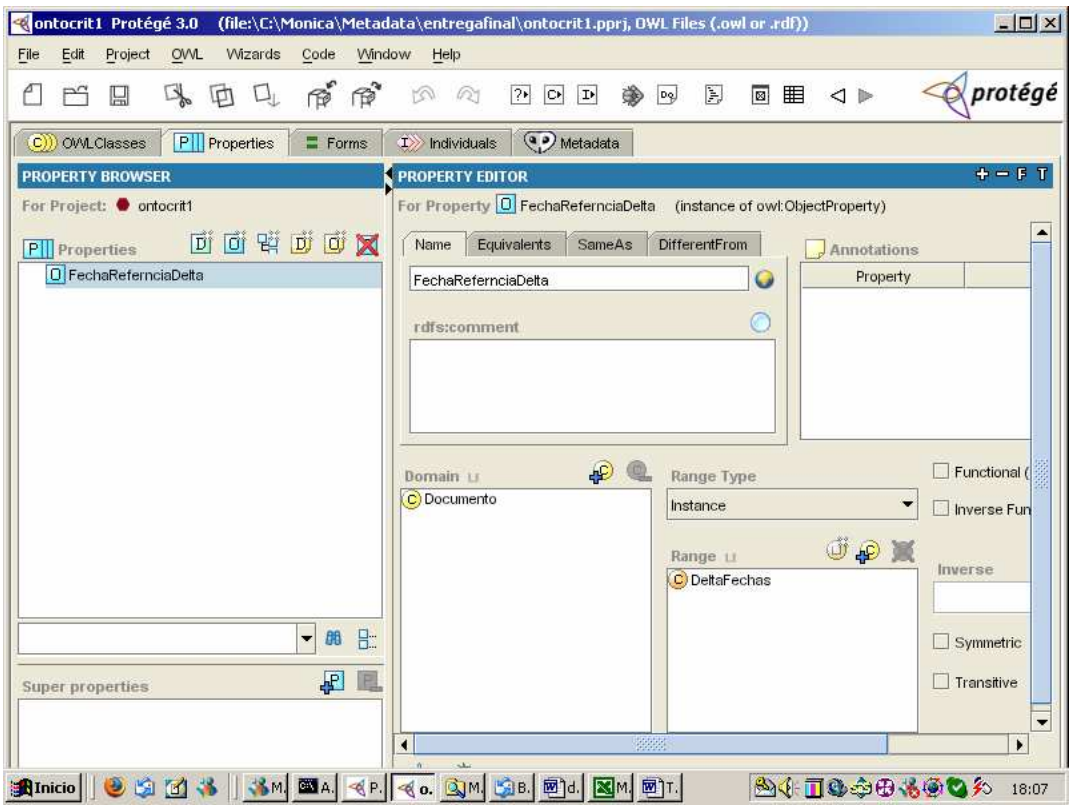
Y que cada documento pertenezca a un rango con un distinto valor de pertenencia, lo que lleva a la búsqueda de ontologías difusas.
Según lo analizado en la bibliografía, las extensiones que trabajan con lógicas difusas lo realizan en la jerarquía de las clases pero no en la relación de pertenencia de un individuo a una clase [QHC 04].
Por lo tanto esta opción de modelado también debe ser descartada ya que aunque se contara con la extensión correspondiente igual no refleja lo que se necesita.

Al no contar con una relación de pertenencia difusa se decidió simular distintos grados de pertenencia a partir de la pertenencia a una clase según si todas las referencias del documento se encuentran en determinados rangos.
De esta forma se definió:

Documentos Recientes: aquellos documentos donde todas sus referencias tienen una diferencia de fecha con el documento considerado en el rango recientes.
Documentos TiendeReciente: aquellos documentos donde todas sus referencias tienen una diferencia de fecha con el documento considerado en el rango recientes o rango 2.
Documentos Viejo: aquellos documentos donde todas sus referencias tienen una diferencia de fecha con el documento considerado en el rango viejas.
Documentos TiendeViejo: aquellos documentos donde todas sus referencias tienen una diferencia de fecha con el documento considerado en el rango viejas o rango 3.
Documentos Variado: aquellos documentos que no pertenecen a los tipos anteriores. Sus referencias se encuentran distribuidas en los diferentes rangos.

Esto fue modelado en una ontología llamada ontocrit1.
A continuación se muestran las partes más significativas de esta ontología.

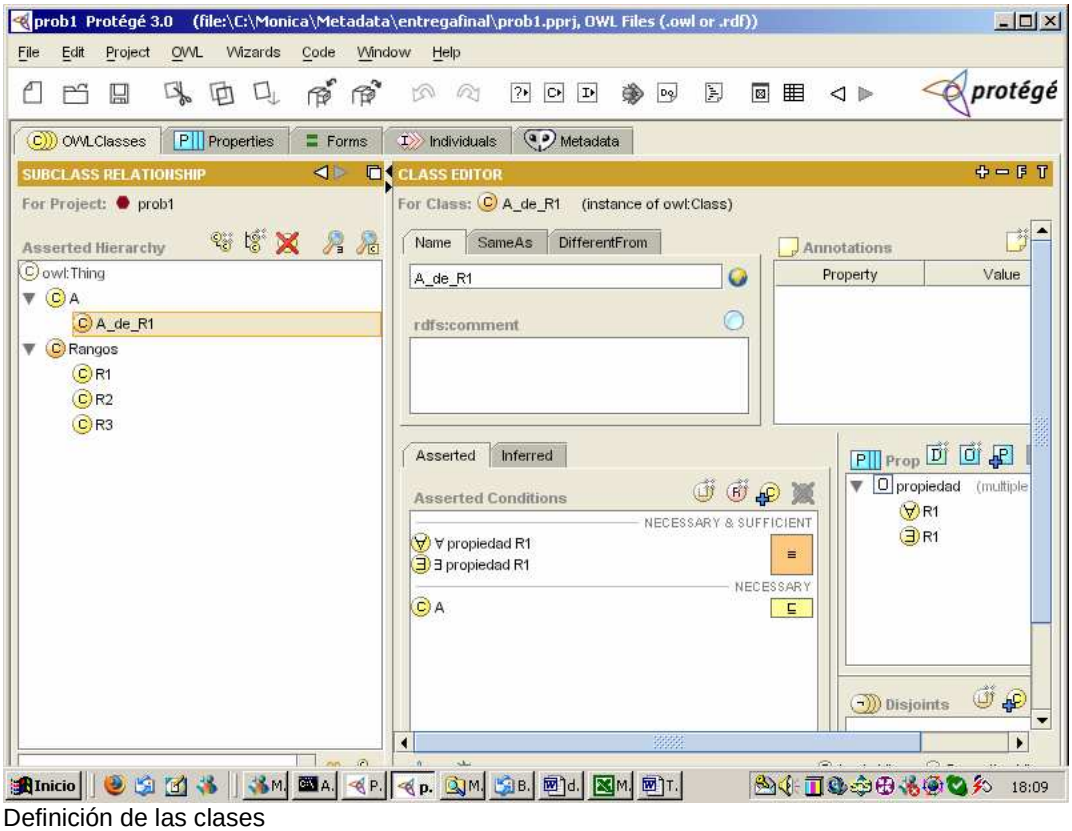

ontocrit1 Protégé 3.0 (file:\C:\Monica\Metadata\entregafinal\ontocrit1.pprj, OWL Files (.owl or .rdf))
File Edit Project OWL Wizards Code Window Help
protégé
OWLClasses Properties Forms Individuals Metadata
SUBCLASS RELATIONSHIP
For Project: ontocrit1
Asserted Hierarchy
owl:Thing
DeltaFechas
Rango2
Rango3
Recientes
Viejas
Documento
DocTiendeReciente
DocRecientes
DocTiendeViejo
DocViejo
DocVariado
CLASS EDITOR
For Class: DocTiendeReciente (instance of owl:Class)
Name SameAs DifferentFrom
DocTiendeReciente
rdfs:comment
Documentos donde todas sus referencias tiene una diferencia de fecha con el documento entre 0 y 8 años
Annotations
Property Value
rdfs:comment
Asserted Inferred
Asserted Conditions
NECESSARY & SUFFICIENT
Documento
∀ FechaReferenciaDelta (Recientes ⊔ Rango2)
∃ FechaReferenciaDelta (Recientes ⊔ Rango2)
NECESSARY
Prop
FechaReferenciaDelta
Disjoints
Inicio
18:06


ontocrit1 Protégé 3.0 (file:\C:\Monica\Metadata\entregafinal\ontocrit1.pprj, OWL Files (.owl or .rdf))
File Edit Project OWL Wizards Code Window Help
protégé
OWLClasses Properties Forms Individuals Metadata
PROPERTY BROWSER
For Project: ontocrit1
Properties
FechaReferenciaDelta
Super properties
PROPERTY EDITOR
For Property FechaReferenciaDelta (instance of owl:ObjectProperty)
Name Equivalents SameAs DifferentFrom
FechaReferenciaDelta
rdfs:comment
Annotations
Property
Domain
Documento
Range Type
Instance
Range
DeltaFechas
Functional
Inverse Fun
Inverse
Symmetric
Transitive
Inicio
18:07

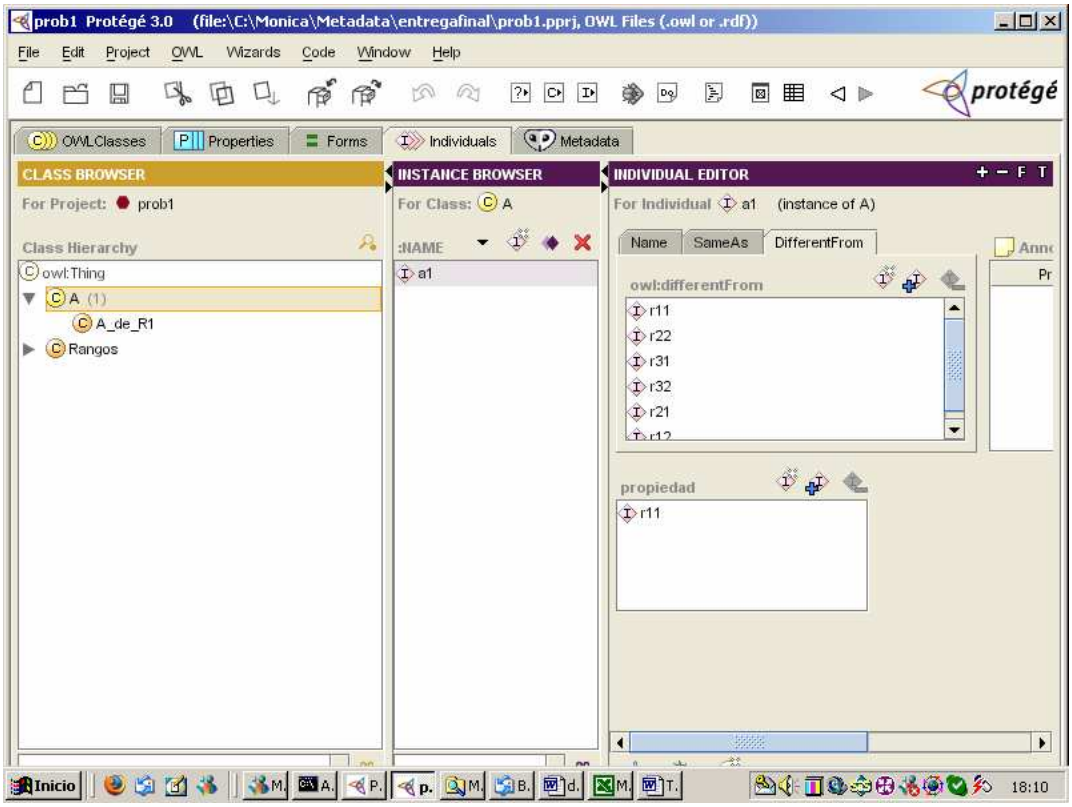
Limitaciones de las herramientas de desarrollo

Al comenzar a trabajar en la clasificación de documentos considerando únicamente este criterio se detectaron diferencias de funcionamiento entre las herramientas que se estaban utilizando y la idea intuitiva. Concretamente no se infieren de la forma esperada los individuos de una clase si ésta está definida con condiciones que involucran cuantificadores universales.
Ante esta situación se simuló la situación en un caso totalmente restringido, donde sólo se considera el problema en cuestión para determinar si ese es efectivamente el origen de las diferencias o resultaba de una combinación de situaciones más complejas.
Se diseña la ontología prob1 donde se tiene una clase A y una clase Rangos (que es la unión de las subclases disjuntas R1,R2,R3), una propiedad con dominio A y range Rangos.
También se define una clase A_de_R1 (subclase de A) definida como aquellos A que todos sus correspondientes en la propiedad pertenecen a la clase R1 y existe por lo menos uno de ellos.
A continuación se crearon individuos en las clases R1 (r11,r12), R2 (r21,r22), R3(r31,r32) y en A (a1), todos diferentes entre si. Se definió correspondiente de a1 en la propiedad el individuo r11.
Por las definiciones realizadas se espera que el individuo a1 pertenezca a la clase A_de_R1, ya que su único correspondiente en la propiedad es un individuo de R1..
Sin embargo al chequear el tipo del individuo a1 el razonador solo encuentra A y al verificar los individuos que pertenecen a la clase A_de_R1 el razonador determina que se encuentra vacía.

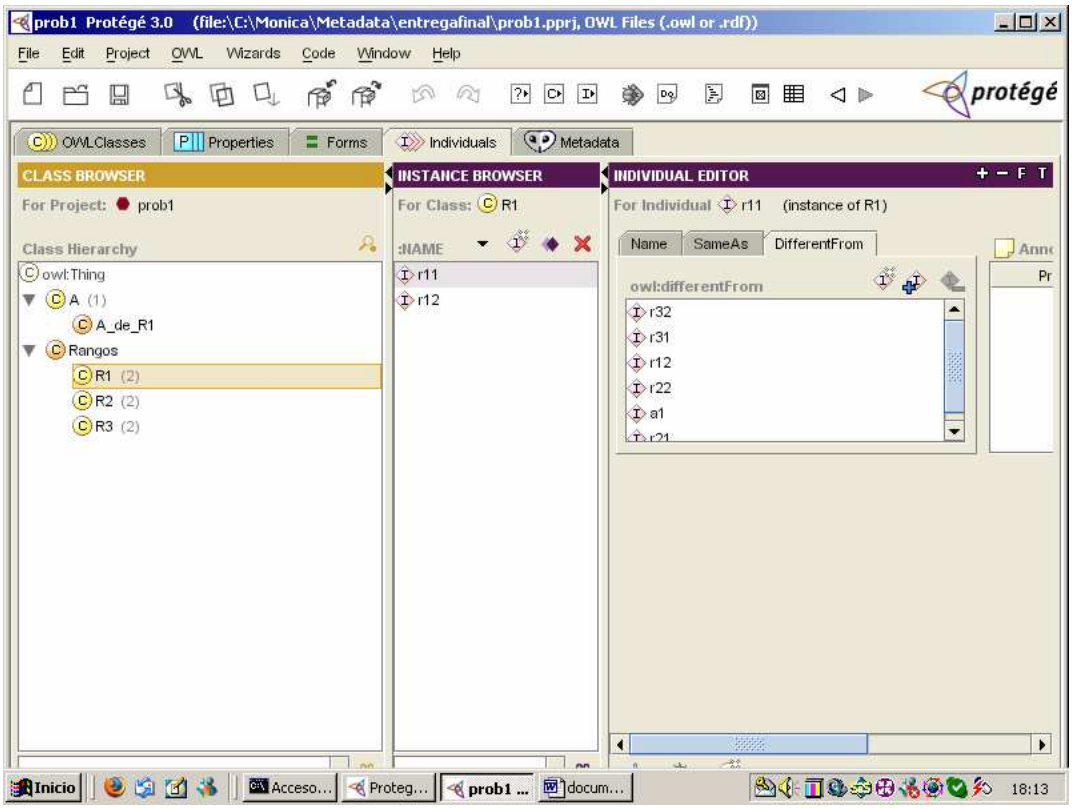
A continuación se muestra la definición de esta ontología y el resultado del razonador:



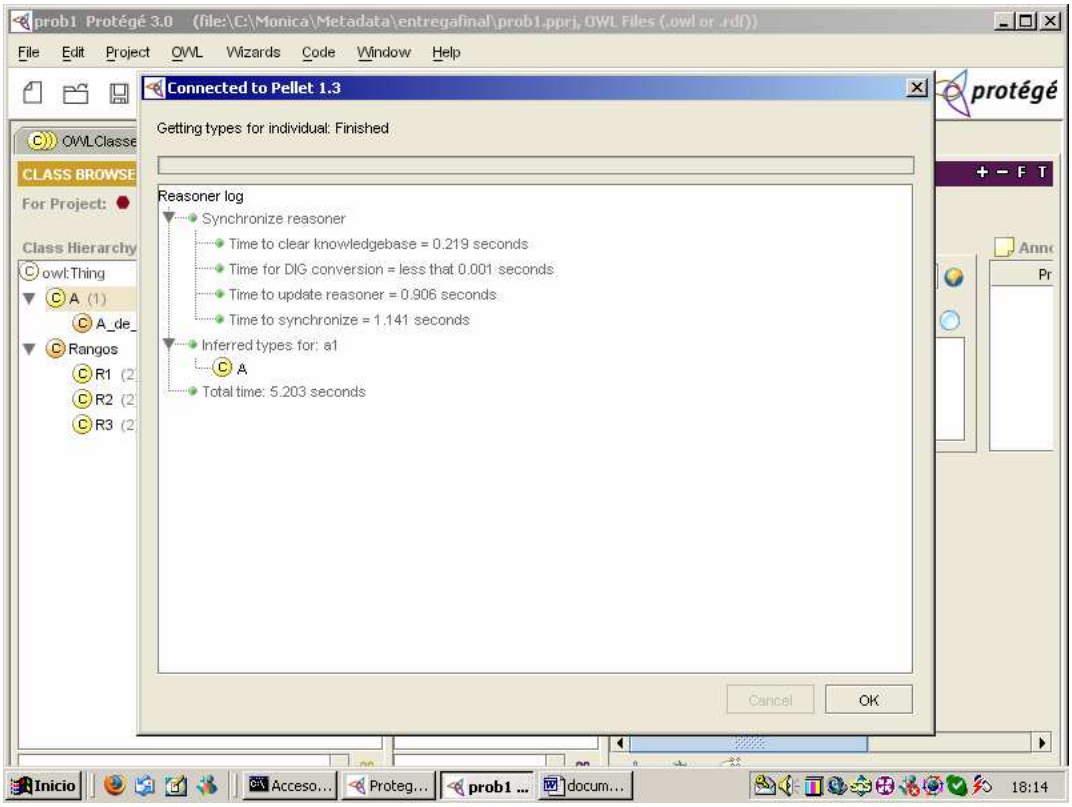
Definición de las clases



Definición del individuo a1



Muestra la pertenencia del correspondiente de a1 en la propiedad a la clase R1



Muestra el resultado del razonador.

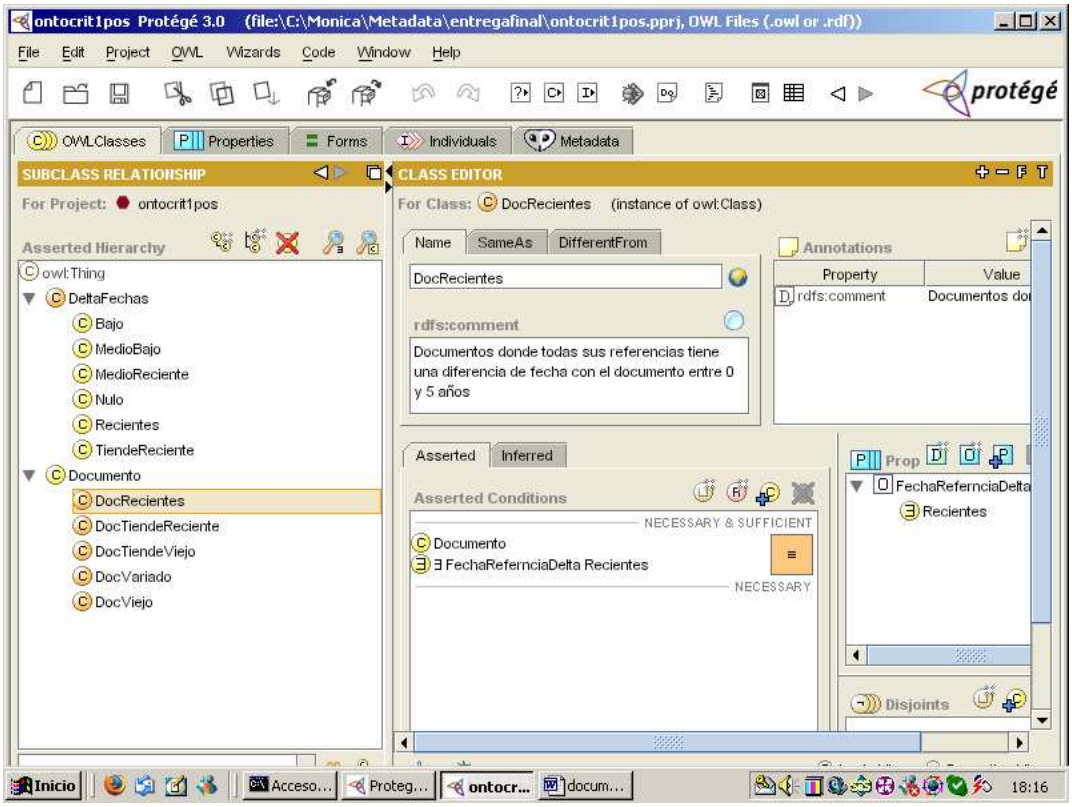
La diferencia encontrada se origina en el hecho de que en la realidad considerada se trabaja en un hipótesis de mundo cerrado pero las herramientas utilizadas trabajan bajo la hipótesis de mundo abierto. Por lo tanto no es posible inferir que todos los correspondientes de un individuo pertenecen a una determinada clase a partir de un correspondiente. Que tenga un correspondiente que cumple con lo definido no dice nada con respecto al resto del universo. Sería necesario poder expresar que el único correspondiente del individuo es el dado y que no se relaciona con ningún otro individuo, lo que se llama cerrar la relación.
Este tipo de especificaciones no son posibles de realizar en OWL, por lo menos en una forma directa[1] por lo tanto se buscó una solución alternativa que evitara este tipo de especificaciones.

Con estas nuevas restricciones se replanteó el modelado de forma tal de poder trabajar con restricciones existenciales (para las cuales las herramientas se observó que se comportan de la forma esperada). Por lo tanto se vio necesario trabajar con propiedades funcionales y pasar la determinación de si un documento tiene todas sus referencias con una diferencia de fechas en un determinado rango a funciones auxiliares, externas a la ontología.

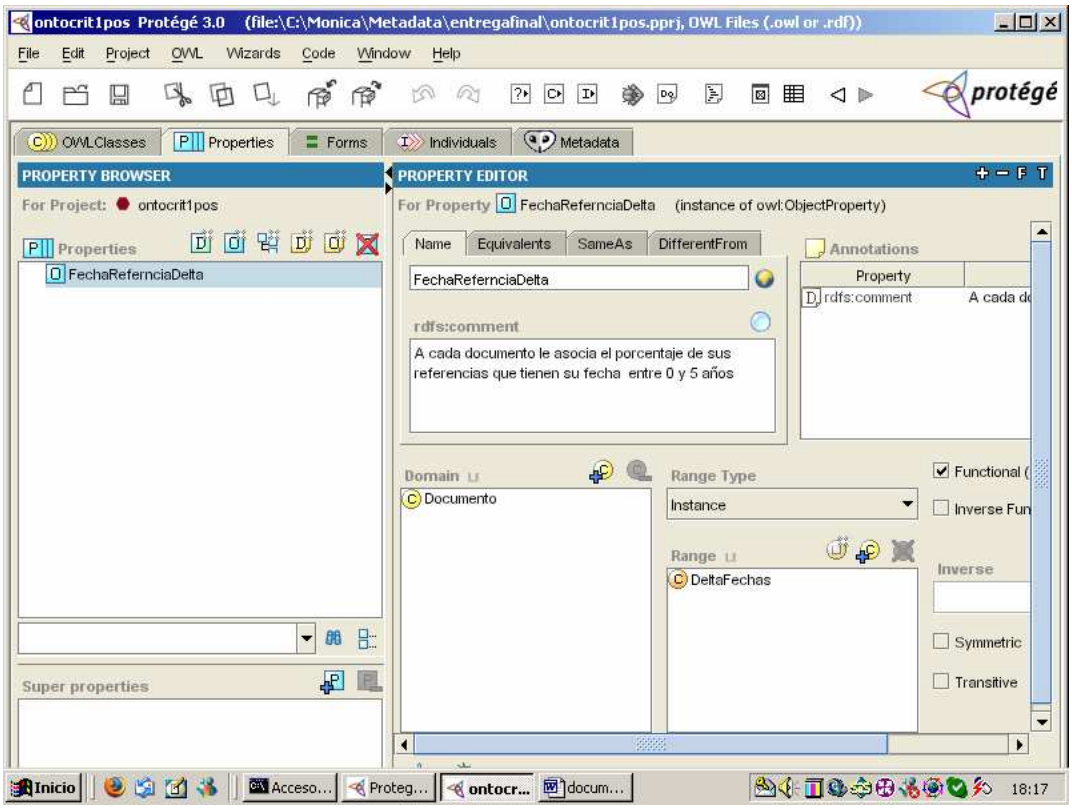
Al pasar la responsabilidad de la determinación a qué rango corresponde cada documento a funciones externas a la ontología se retomó la idea original de trabajar con porcentajes de referencias recientes, considerando que la herramienta que se utilice para el cálculo de estas funciones sí cuenta con la posibilidad de trabajar con los enteros, sus funciones y predicados.

Esta nueva situación se modeló en la ontología ontocrit1pos.

[1] Se considera que una posible solución sería reificar la propiedad y de esta forma poder expresar mayor cantidad de restricciones sobre la misma, pero esto queda fuera del alcance de este trabajo.



Definición de la clase Decrecientes



Finalmente se observó que esta ontología permite realizar la clasificación de los documentos según este criterio, con todas las limitantes que se realizaron durante la evolución del modelado de este criterio provocado por las diferentes restricciones encontradas.

3.2 Criterio 2

En este criterio se valora que los documentos tengan referencias a documentos correspondientes a grupos de trabajos externos al del autor como forma de considerar que es un trabajo, o área temática que esta siendo estudiada por distintos grupos y eso marcaría una mayor importancia.
El que existan distintos grupos de trabajo en el tema se evalúa considerando que en las referencias se encuentren documentos realizados por instituciones externas a la institución que pertenece el autor.

En un primer momento se realizó una ontología (ontocrit2) en donde se clasifican los documentos en dos categorías: aquellos que entre sus referencias tienen por lo menos una referencia a una institución externa y aquellos que no tienen ninguna referencia externa.
Para esto se definieron dos relaciones con dominio en los documentos y rango en una clase correspondiente a las instituciones:

AfiliacionAutor: asocia a cada documento con las instituciones a las que pertenece el autor.

AfiliacionReferencia: asocia a cada documento con las instituciones a las que pertenecen los autores de los documentos referenciados.

La clase instituciones se clasificó en dos subclases:

Inst_Autor: instituciones del autor del documento

Inst_Referencia: instituciones de los autores de las referencias.

Y se difinieron las clases:

Inst_Autor $\cap$ Inst_Referencia (Inst_Inter) y

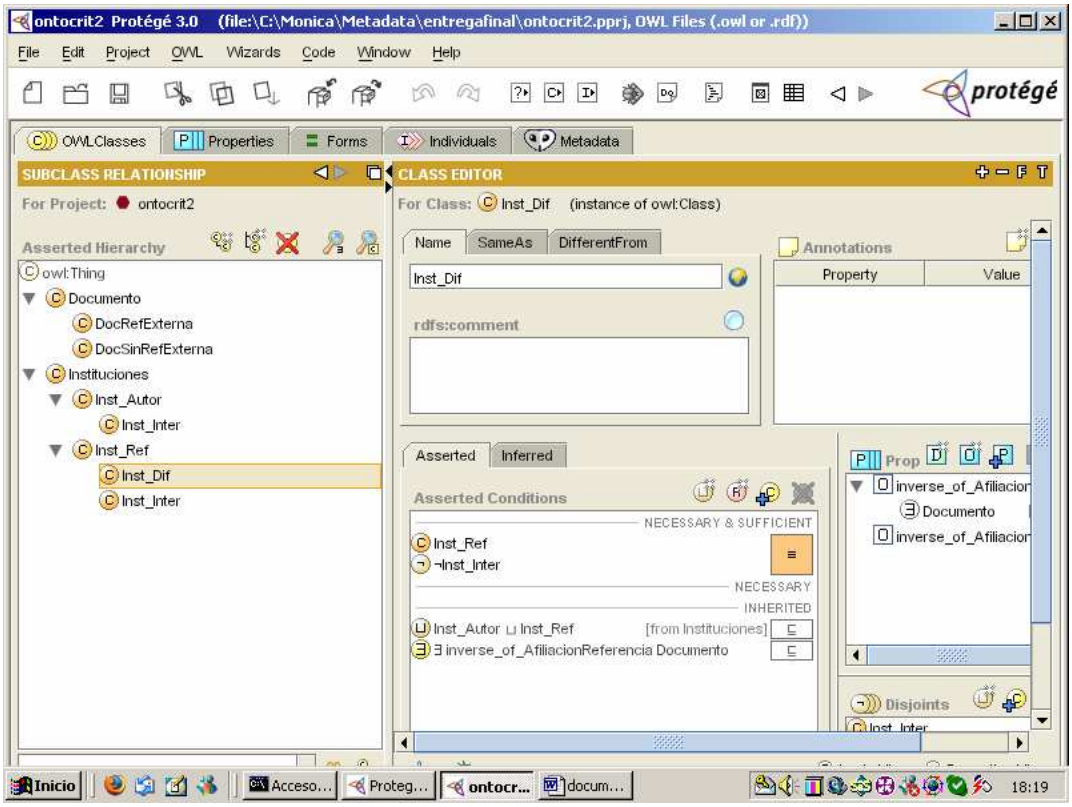
Inst_Referencia – Inst_Autor (Inst_Dif)

A partir de estas clases se definen los documentos:

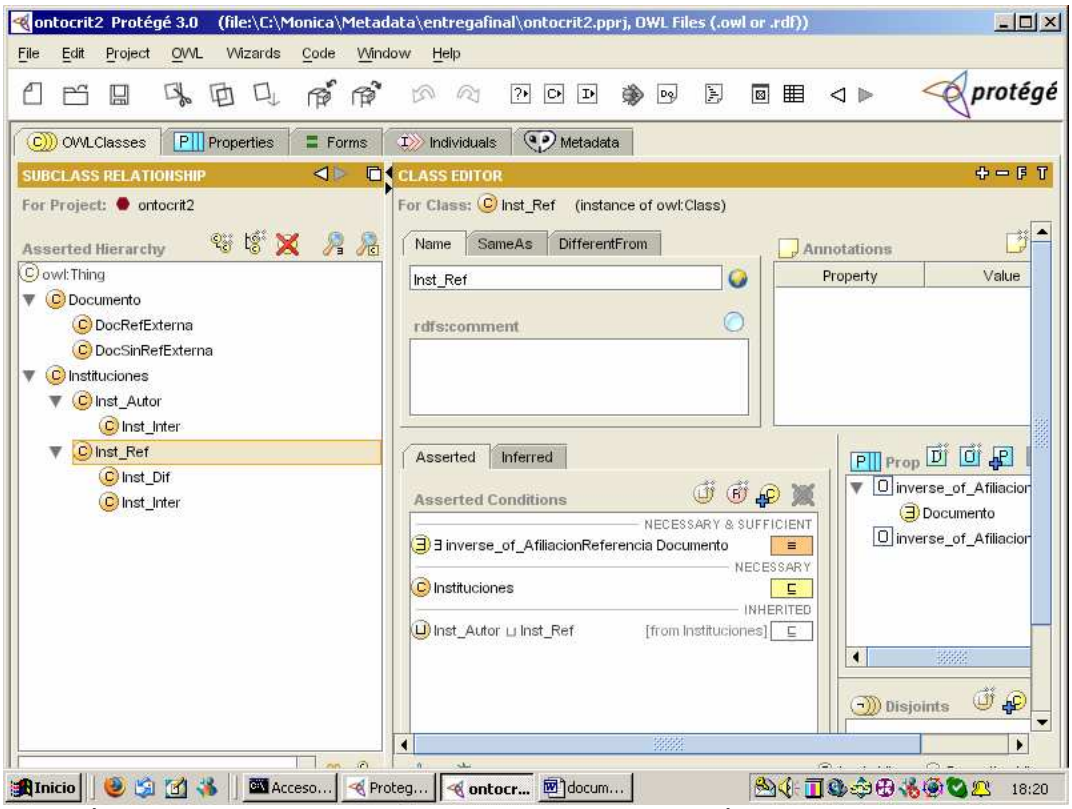
DocRefExterna: Aquellos documentos donde por lo menos una de sus referencias proviene de una institución externa a la del autor. (Existe por lo menos un correspondientes en Inst_Referencia perteneciente a la clase Inst_Dif).

DocSinExterna: Aquellos documentos sin referencias a instituciones externas (Todos sus correspondientes en Inst_Referencia pertenecen a la clase Inst_Inter.

A continuación se muestran las características más relevantes de la ontología definida.



Definición de la clase diferencia



Definición de la clase Inst_Ref (Inst_Autor se define en forma análoga)

Si bien esta forma de trabajo permite una clasificación de los documentos aceptable se considera que la opción tomada para el criterio 1 es más conveniente ya que permite clasificar los documentos según distintos grados de cumplimiento con el criterio.
Por lo tanto se opta por una solución análoga a la planteada para el criterio 1, donde se definen niveles (alta, tiendeAlta, mediaAlta,medioBajo,Bajo,Nulo) para indicar qué porcentaje del total de las referencias del documento corresponden a referencias externas.
La relación entre el documento y estos valores nuevamente es calculada por una función externa a la ontología.

De esta forma es posible realizar la clasificación de los documentos según este criterio.

3.3 Criterio 3

En este criterio se valora que el autor del documento tenga publicaciones realizadas anteriormente y también se valoran el tipo de la publicación considerando el siguiente orden de importancia: capítulo de libro, revista, conferencia, otro.

A partir del modelado de este criterio se tomó en cuenta todas las restricciones encontradas en el modelado de los criterios anteriores y no se consideraron posibilidades que provocaran caer en los problemas ya identificados (sean de OWL o de las herramientas).
Por lo tanto se definen niveles a los que pertenecen los documentos y los valores correspondientes a la función que determina el nivel son calculados por funciones externas a la ontología.

Para esta clasificación se utiliza una función que asocia a cada documento el tipo de mayor nivel de las publicaciones anteriores del autor.(TipoPublicacionAntAutor)

De esta forma los documentos se clasifican en las siguientes categorías:

> **DocAutoresCapLibro:** documentos tales que la publicación anterior de mayor nivel que tiene corresponde a un capítulo de libro. (existe por lo menos un correspondiente del documento en TipoPublicacionAntAutor con el valor capLibro)
> **DocAutoresRevistas**: documentos tales que la publicación anterior de mayor nivel que tiene corresponde a una publicación en una revista. (existe por lo menos un correspondiente del documento en TipoPublicacionAntAutor con el valor revista)
> **DocAutoresConferencia**: documentos tales que la publicación anterior de mayor nivel que tiene corresponde a una publicación en una conferencia. (existe por lo menos un correspondiente del documento en TipoPublicacionAntAutor con el valor conferencia)
> **DocAutoresOtro**: documentos tales que el autor tiene publicaciones anteriores pero no corresponden a las antes mencionadas. (existe por lo menos un correspondiente del documento en TipoPublicacionAntAutor con el valor otro)
> **DocAutoresIniciales:** documentos tales que el autor no tiene publicaciones anteriores. (existe por lo menos un correspondiente del documento en TipoPublicacionAntAutor con el valor noPub)

3.4 Criterio 4

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).

Se realiza el modelado de la misma forma que en los criterios anteriores, definiendo niveles según los valores de las funciones consideradas.

Para la clasificación según este criterio se definen tres funciones:

> **GradoAutor**: relaciona el documento con el mayor grado académico del autor, considerando los posibles valores ordenados en forma decreciente de importancia: posgrado, grado, estudiante, nada.

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 de la forma descripta a continuación para definir los niveles de los autores, de nivel 1 a nivel 5 (en orden decreciente de importancia):

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

Cabe destacar que es una clasificación totalmente personal donde se le da mayor peso al grado, luego a si existen referencias al mismo autor y finalmente se considera si el autor ha sido evaluador. Ante el uso de la herramienta desarrollada debería considerarse si esta clasificación es adecuada o es necesario modificarla.

3.5 Criterio Resumen

Una vez que se tiene la clasificación del documento según los distintos criterios es deseable tener una clasificación tipo resumen, global.

Con este objetivo se clasificaron los distintos niveles de cada criterio en 3 grupos:
Alta: corresponde a las 2 categorías de mayor nivel del criterio
Baja: corresponde a las 2 categorías de menor nivel del criterio
Medio: corresponde a las restantes categorías.

A modo de resumen se optó por clasificar a los documentos en las siguientes categorías:

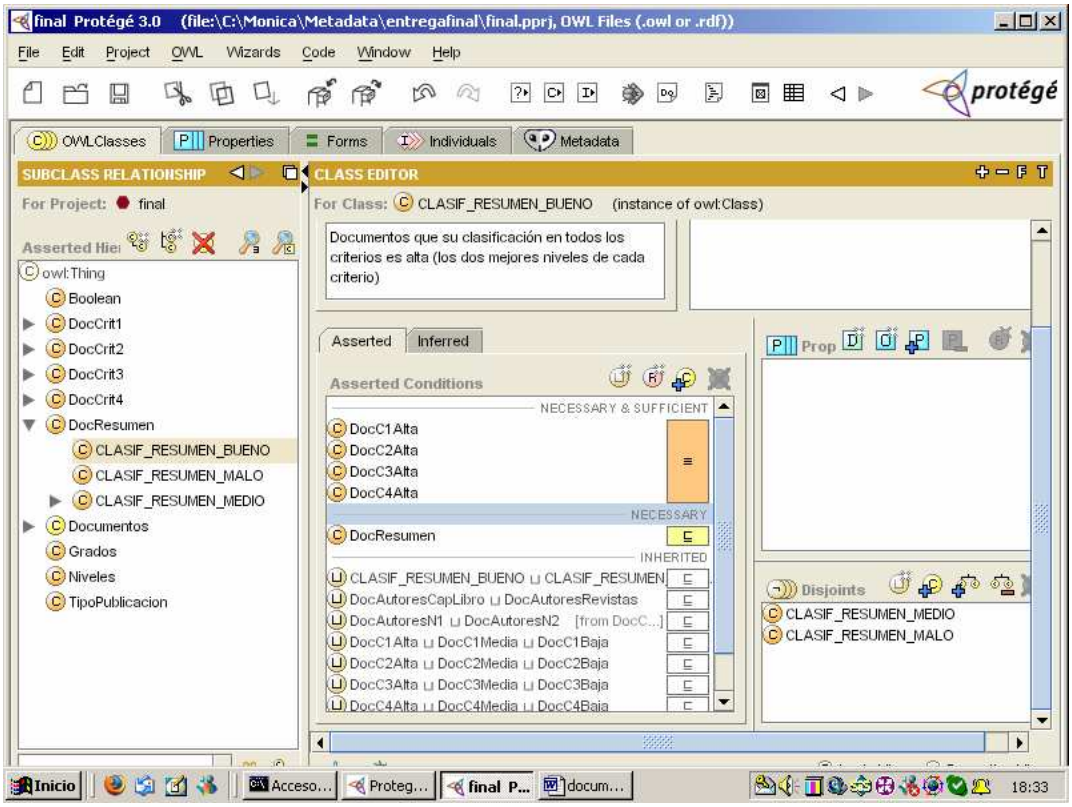
Buenos: aquellos documentos que en todos los criterios clasifican en niveles correspondientes al grupo alto.

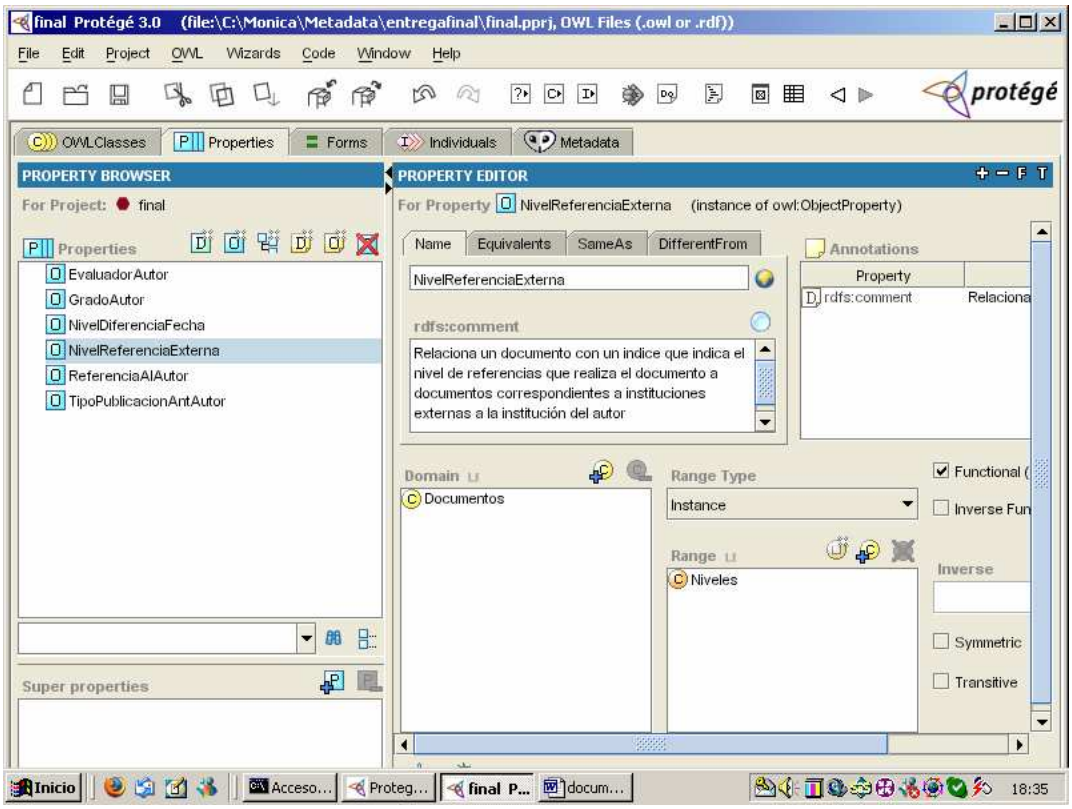
Malos: aquellos documentos que en todos los criterios clasifican en niveles correspondientes al grupo baja.

Medio: los restantes documentos.

Considerando las definiciones dadas para cada una de estas categorías la forma directa de especificar las clases que las representan sería utilizando restricciones del tipo cuantificador universal y restricciones con complemento, pero por las limitaciones encontradas fue necesario especificarlas por enumeración.

A continuación se muestra a modo de ejemplo el primer nivel de clases, la definición de la clase correspondiente a los documentos buenos, y las propiedades definidas en la ontología final, la herramienta desarrollada, cuyos archivos acompañan a este documento.





4. Forma de uso de la herramienta

Para clasificar un documento en particular se deberá crear un individuo en la clase Documento, asignarle los valores correspondientes a todas las funciones definidas para esta clase y finalmente inferir el tipo del individuo.

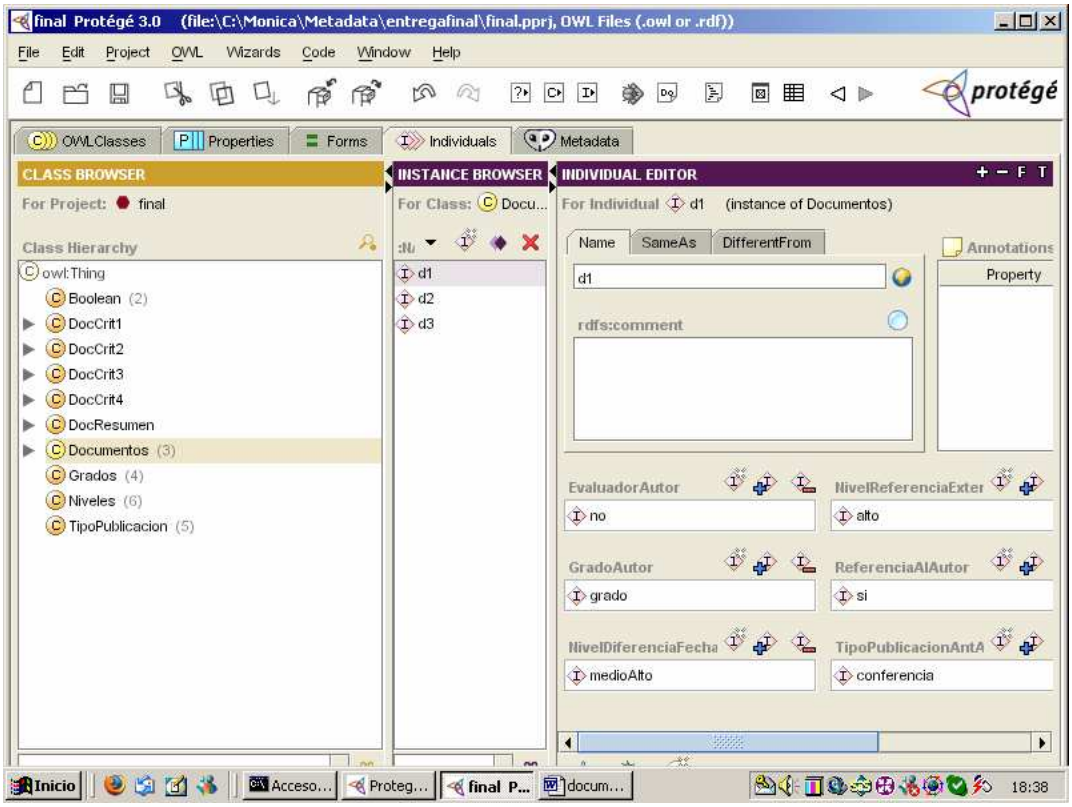
El razonador listará los nombres de las clases a las cuales pertenece el documento. En la clase CLASIF_RESUMEN_xxxx se observa la clasificación resumen para el documento y en las clases DocCNxxxxx se observa la clasificación obtenida por el documento en el criterio N.

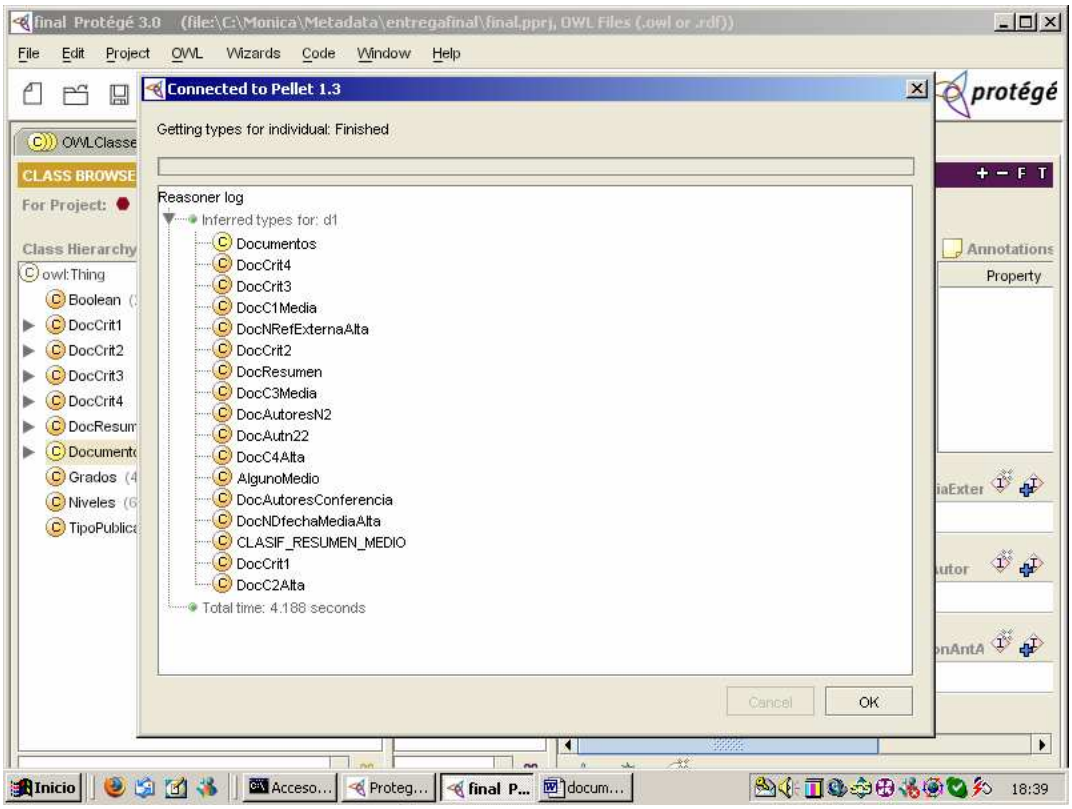
Otra forma de usar la herramienta consiste en ingresar todos los documentos que se deseen en la clase documentos con todas sus correspondientes en las funciones definidas para la clase y luego dirigirse a la clase que se desea y determinar todos los elementos que pertenecen a la clase.

Con respecto a cómo asignar los correspondientes en cada función para cada documento creado es importante marcar que al momento de seleccionar los criterios para la clasificación se consideró que se estuviera hablando de características de los documentos que correspondieran a elementos que se pueden encontrar en la codificación MARC 21[MARC 21] para documentos digitales.
En los criterios seleccionados se habla de las referencias del documento, información que se encuentra en el campo 504 de la codificación MARC 21.
También se habla del autor del documento, campo 100 y de la institución del autor, campo 110.
Asimismo 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 que se puedan consultar.

Por lo tanto para la asignación de correspondiente en las funciones NivelDiferenciaFecha, NivelReferenciaExterna, ReferenciaAlAutor se debe inspeccionar las referencias del documento y luego manipularlas, para las funciones EvaluadorAutor, TipoPublicacionAntAutor se debe consultar bd del estilo de Scholar, para la función GradoAutor (o incluso las anteriores) se debe consultar bd del estilo de cvlatex.

A continuación se muestra a modo de ejemplo la definición de un documento y el resultado de su clasificación


final Protégé 3.0 (file:\C:\Monica\Metadata\entregafinal\final.pprj, OWL Files (.owl or .rdf))
File Edit Project OWL Wizards Code Window Help
protégé
OWLClasses Properties Forms Individuals Metadata
CLASS BROWSER
For Project: final
Class Hierarchy
owl:Thing
Boolean (2)
DocCrit1
DocCrit2
DocCrit3
DocCrit4
DocResumen
Documentos (3)
Grados (4)
Niveles (6)
TipoPublicacion (5)
INSTANCE BROWSER
For Class: Docu...
d1
d2
d3
INDIVIDUAL EDITOR
For Individual d1 (instance of Documentos)
Name SameAs DifferentFrom
d1
rdfs:comment
Annotations
Property
EvaluadorAutor
no
NivelReferenciaExter
alto
GradoAutor
grado
ReferenciaAlAutor
si
NivelDiferenciaFecha
medioAlto
TipoPublicacionAntA
conferencia
Inicio
Acceso...
Proteg...
final P...
docum...
18:38


final Protégé 3.0 (file:\C:\Monica\Metadata\entregafinal\final.pprj, OWL Files (.owl or .rdf))
File Edit Project OWL Wizards Code Window Help
protégé
Connected to Pellet 1.3
Getting types for individual: Finished
Reasoner log
Inferred types for: d1
Documentos
DocCrit4
DocCrit3
DocC1Media
DocNRefExternaAlta
DocCrit2
DocResumen
DocC3Media
DocAutoresN2
DocAutn22
DocC4Alta
AlgunoMedio
DocAutoresConferencia
DocNDfechaMediaAlta
CLASIF_RESUMEN_MEDIO
DocCrit1
DocC2Alta
Total time: 4.188 seconds
Cancel
OK
Annotations
Property
Inicio
Acceso...
Proteg...
final P...
docum...
18:39

5. Conclusiones

Como primer punto es importante marcar que se cumplió con el objetivo de definir una herramienta que permita clasificar documentos según ciertos criterios a definirse.
A modo de ejemplo sólo se trabajo con 4 criterios pero es posible aumentar la cantidad de los mismos en una forma similar a la realizada. Y de esta forma sería posible valorar mejor el aporte de la herramienta ya que su utilidad se aprecia mejor a medida que aumentan las variables que se consideran en la definición de los criterios.

Un aspecto negativo en la herramienta definida es que se traslada gran parte del trabajo a funciones externas a la herramienta que permitan calcular los correspondientes para cada una de las funciones. Pero como se marcó en cada uno de los criterios esto en la mayoría de los casos está provocado por las limitaciones de las herramientas que se están utilizando en el desarrollo. Considerando que son herramientas que se encuentran en desarrollo actualmente es de esperar que en poco tiempo se superen los inconvenientes que tienen en este momento.

Finalmente con este trabajo se deja en evidencia que en caso de contar con las herramientas que funcionen correctamente la clasificación de documentos es una aplicación donde el uso de una ontología es de gran ayuda.

6. Referencias

[Baad *et al.* 03] F.Baader, D.Calvanese, D.McGuinness, D.Nardi, P.Patel-Schneider :The Description Logic Handbook Theory, Implementation and Applications, Cambridge University Press 2003, ISBN 0-521-78176-0

[Horr *et al* 04] M. Horridge, H.Knublauch, A. Rector, R. Stevens, C. Wroe: A Practical Guide to Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools tutorial en Cooperative Ontologies Programme (CO-ODE)
http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf
(último acceso : 20 de marzo de 2006)

[MARC 21] Documentación de MARC 21
(http://www.loc.gov/marc/)
(último acceso : 20 de marzo de 2006)

[QHC 04] T.T.Quan, S.C.Hui, T.H.Cao : FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web, Workshop 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Pisa, Italia, 24 de setiembre de 2004
http://olp.dfki.de/pkdd04/quan-final.pdf
(último acceso : 20 de marzo de 2006)