

### Práctica. Marcado semántico y Recuperación de Información

#### Objetivos docentes:

- Aprender a editar ontologías con la herramienta Protege utilizando el lenguaje OWL-DL
- Entender las dificultades y la utilidad de las ontologías como mecanismo de anotación semántica y recuperación de información (imágenes).
- Entender las diferencias entre **recuperación sintáctica/semántica** para información textual.
- Cargar y manejar ontologías desde Java usando el puente OntoBridge<sup>1</sup> (JENA)/ DIG. Uso de razonadores de DLs.

#### Apartado A. Diseño y formalización en OWL de una ontología para representar conocimiento sobre escenas de películas y series.

La ontología incluirá conceptos primitivos y definidos sobre películas, series y actores y personajes que intervienen en ellas. Como el dominio es muy extenso se fijarán unos **individuos** concretos actores, personajes, y títulos de películas/series que se darán en clase. Se debe hacer especial hincapié en la definición de la *jerarquía conceptual* que clasificará los **individuos** anteriores que se introducirán posteriormente.

- Es obligatorio incluir *conceptos definidos* en la ontología.
- Introducir las instancias de prueba dadas en la ontología (actores, personajes, películas y/o series) y relaciones entre una película o capítulo, o serie y sus actores.
- Hacer una interfaz simple en java (por ejemplo, un botón para cada una de las consultas predefinidas siguientes) y conectar (usando **OntoBridge**<sup>1</sup>) para hacer algunas recuperaciones interesantes en las que se use inferencia (no tienen por qué ser exactamente estas):
  - Actores que hayan participado en más de 3 películas.
  - Parejas de actores que no han aparecido juntos en ninguna película pero tienen compañeros comunes.
  - Personajes que aparecen alguna serie o película española.

#### Apartado B. Anotación y recuperación semántica de escenas de películas

El objetivo de esta práctica es tener una herramienta que permita organizar imágenes de escenas de películas y recuperarlas usando consultas sobre su contenido. Por ejemplo, “escenas en las que salga Bart Simpson”, “Escena de la serie Los Simpson”, “imágenes de la película “Mr y Mrs. Smyth” “Escenas románticas”.

Para eso se deben seguir los siguientes pasos:

1. **Recopilar un conjunto de imágenes** correspondientes a escenas de nuestro dominio (según las series, películas, o actores del apartado A).
2. Desarrollar la **herramienta de anotación**.

Se pide hacer una herramienta similar a la de la figura que muestre la ontología en un lado y la imagen de la escena en otro lado para permitir marcarla semánticamente. Sólo se pueden marcar imágenes completas (no partes de una foto).

---

<sup>1</sup> Desarrollado por el grupo GAIA ([gaia.fdi.ucm.es](http://gaia.fdi.ucm.es)) y accesible a través de la página WEB <http://gaia.fdi.ucm.es/grupo/projects/ontobridge/>



Cada imagen se representará en la ontología mediante un individuo  $f_i$  (instancia de un concepto Imagen). El nombre del archivo que contiene la foto (es decir, el archivo jpg, gif,...) está en el datatype urlfoto.

Hay dos tipos de anotaciones posibles sobre una imagen  $f_i$ :

1. Asertarlo bajo alguno de los conceptos primitivos de la ontología (ej. 'escena1' es una 'Escena Familiar')
2. Relacionarlo con otro individuo de la base de conocimiento a través de un rol (ej. En 'escena1' *aparece* 'Bart Simpson', donde *aparece* es una relación de la ontología)

La pertenencia de  $f_i$  a los conceptos definidos no se debe asertar, sino que deberá ser el sistema quien lo infiera. Por ejemplo, los conceptos: Películas de acción, Películas de Kieffer Sutherland pueden ser *conceptos definidos* con nombre.

3. Desarrollar una herramienta de recuperación semántica
  - a. Identificar patrones de consultas en lenguaje natural y cómo serían las consultas a las ontología (a través de OntoBridge) para conseguir la recuperación.
  - b. “Escenas de Spiderman” Patrón LN: “escenas de #Película”  
recuperar las  $f_i$  tal que  $f_i$  *esun* Spiderman
  - c. “Fotos de Kieffer Sutherland en 24” Patrón LN: “Fotos de #Actor en #Película”  
recuperar las  $f_i$  tal que  $f_i$  *esun* Escena y  $f_i$  *aparece* Kieffer Sutherland y  $f_i$  *pertenece* 24
  - c. “Fotos de Jack Bauer en 24” Patrón LN: “Fotos de #Personaje en #Película”

### Desarrollo de la práctica y entregas:

- Se entregarán distintas versiones de la ontología. Para cada versión tiene que documentarse el conjunto de cambios relativo a la versión anterior.
- Entrega 1.
  - Ontología Versión 1. Versión realizada para el Apartado A.
  - Interfaz simple Java para las consultas predefinidas.
- Entrega 2.
  - Ontología Versión 2. Incluirá las modificaciones necesarias para marcar el conjunto de imágenes seleccionadas por los alumnos.
  - Ontología Versión 3. Incluirá las modificaciones necesarias para marcar el conjunto de imágenes dado por los profesores.
- Entrega 3.
  - Versión final de las herramientas de anotación y recuperación. Pueden ser herramientas separadas o integradas en una misma aplicación.
- Documentación de la práctica siguiente una plantilla que se os proporcionará a través del CV.