



## Especificación del proyecto final del curso de Lógica y Matemática

### Un experimento en la Web Semántica con Programación Lógica

Jacinto Dávila

Estamos en procura de una aplicación que persuada a usuarios regulares de la World Wide Web de que es posible que los programas manipulen los contenidos electrónicos con criterios semánticos. Es decir, que los programas “entiendan” los contenidos, más allá del simple análisis sintáctico y lexicográfico, al punto que puedan usar relaciones lógicas y ontológicas para recuperar y presentar la información.

En este experimento vamos a construir un sistema de almacenamiento, indizado y búsqueda semántica. Los componentes del sistema son:

**El comentarista:** Es un programa que interactúa con la persona que diseñó una página web y le ayuda a elaborar un registro semántico-ontológico del contenido de esa página. El contenido en cuestión es agregado a la página y puede ser usado por los programas que consultan la página en búsqueda de información. En particular, esos registros son usados por los robots que revisan la WEB y crean los índices para ubicar la información.

**El robot semántico:** Los robots son programas que rastrean la WEB descargando páginas que luego son procesadas para asociar sus contenidos a índices de búsqueda. Para hacer este trabajo, los robots visitan sistemáticamente una gran cantidad de web-sites, revisando las páginas web colocadas en directorios autorizados (Existe un protocolo para autorizar a los robots a visitar páginas en un website). Dada la magnitud de la tarea, los robots suelen concentrar su revisión en ciertos campos predefinidos de cada página web (los campos meta, normalmente).

En nuestro caso, el robot creará un índice para asociar términos selectos con cada página WEB. A diferencia de los robots conocidos, los términos serán seleccionados entre aquellos suministrados por el creador de la página al interactuar con el programa comentaristas. Es decir, el índice será creado con un criterio semántico.

**El buscador semántico:** Este programa recibe las solicitudes de búsqueda. Los usuarios siempre desean ubicar las páginas web que hablan sobre ciertos términos y conceptos. Este programa interactuará con el usuario, posiblemente en lenguaje natural.

Con la información del usuario y con sus propias reglas sobre sinónimos y relaciones ontológicas, el buscador definirá la consulta al índice. Con las referencias obtenidas en el índice, recuperará las páginas correspondientes y analizará el registro ontológico de cada una (si existe). Esta última revisión le permitirá al programa decidir si la página efectivamente responde a la pregunta del usuario.

Para realizar el proyecto, cada uno de Uds escogerá uno de estos programas y lo implementará en (SWI) PROLOG. Para garantizar la interoperabilidad, respetarán los siguientes detalles de implementación:

El comentarista generará registros como:

```
<meta http-equiv=Content-Type content="text/html; charset=windows-1252">
<meta name=isa object=jacinto class=profesor>
<meta name=partof object=profesor class="Universidad de Los Andes">
<meta name=belong this=pagina to="Facultad de Ingeniería">
<meta name=about concept=profesor>
```

que añadirá a una página HTML sin alterar el resto de su contenido.

El robot semántico creará índices en PROLOG con registros como los siguientes:

```
% i( Termino, URL, Comentario ): al Termino se le explica en URL
i( jacinto, 'http://cesimo.ing.ula.ve/~jacinto', 'Pagina Hogar' ).
...

% s( Termino1, Termino2 ): El Termino1 es sinónimo del Termino2
s(jacinto, 'Jacinto Davila').
s(jacinto, jacinto@ula.ve ).
...
```

y **cualquier otro** que pueda ser necesario.

El buscador semántico operará sobre esas representaciones y usará los recursos y paquetes del SWI para descargar y procesar las páginas WEB. Para los análisis empleará reglas y representaciones de las ontologías como las que hemos usado durante el curso.

**Fin de la especificación.**