

**CS6795 Semantic Web Techniques**

**Project proposal**

Schema.org Combined with Rules for the Semantic Annotation of

HTML5 Pages

**Advisor:** Dr. Harold Boley **Project Members:**

Manisha Gupta

Sharath Chandra

Swapnil Sule

**INTRODUCTION**

* **HTML5:**

Web designing has come a long way since the inception of internet. When the WWW was developed, HTML was the only code for web design and there weren't many design choices. Web designers were challenged with the task of making their web pages more appealing. That is how designers starting adding their modifications to the HTML and this gave rise to newer versions of the mark-up language. With several improvements over their respective previous versions HTML2, HTML3 and HTML4 and HTML5 came into being. HTML5 and it has several advantages over its predecessors.

* **Schema.RDFS.org:**

The three big search engines Bing, Google and Yahoo! introduced [Schema.org](http://schema.org/), a collection of terms that webmasters can use to mark-up their pages to improve the display of search results. Various *tools* that use or produce Schema.org terms are already available. Schema.RDFS.org provides the Schema.org Type Hierarchy in various formats including RDF/XML, RDF/Turtle , RDF/Triples and JSON.

There are Form based online tools which allows us to generate HTML snippets decorated with Schema.org terms in microdata.

* [Schema Creator](http://schema-creator.org/)
* The [Microdata Generator](http://www.microdatagenerator.com/)

The Microdata Generator uses mark-up structure and vocabulary found on Schema.org

**OBJECTIVE**

The main objective of our project is to implement the subClassOf definitions of the RDFS-formatted Type Hierarchy in OO jDREW, analyze them using OO jDREW’s taxonomy querying, and define RuleML1.0 facts.

**Approach:**

To attain the above stated objectives and annotate them in a clear and concise way, we plan to make a webpage that contains variety of information about few movies. This webpage shall be made using HTML5 Microdata and schema.org. We believe this approach will make it easy to express and showcase the benefits of HTML5 Microdata and schema.org in the area of web programming, especially that related to search engines while employing the SubClassOf definitions of the RDFS-formatted Type Hierarchy which is the primary objective.

DataType

Boolean

Date

DateTime

Number

Float

Integer

Text

URL

Time

MOVIES

ACTOR (PERSON TYPE)

DIRECTORS (PERSON TYPE)

MUSIC DIRECTOR

CINEMATOGRAPHER

CHOREOGRAPHER

GRAPHICS EXPERT

MAKE-UP

MUSIC

MusicPlaylist: numTracks, track, tracks

MusicAlbum: byArtist

MusicRecording: byArtist, duration, inAlbum, inPlaylist

DURATION (TIME)

PRODUCERS (PERSON)

PRODUCTION COMPANY (ORGANIZATION)

TRAILER (URL)

How to analyze using OO jDREW’s Taxonomy querying system

MOVIES

MUSICS

ACTOR

PRODUCER

DIRECTOR

TRAILERR

MusicAlbum

MusicPlaylist

CHOREOGRAPHER

CINEMATOGRAPHER

MUSIC DIRECTOR

MusicRecording

When we have RDFS/XML schema, we can query the taxonomy :-

There are 10 different ways to query the Taxonomy

* To test if two classes are in a direct super/sub class relationship then the following query is used:

*In POSL Syntax:*  
  
subsumes(superClass,subClass).

In RuleML Syntax:

<Subsumes>

<Rel>superClass</Rel>

<Rel>subClass</Rel>

</Subsumes>

* To test if two classes are in indirect (transitive closure) super/sub class relationship then the following query is used:  
    
  In POSL Syntax:  
    
  subsumesPlus(superClass,subClass).   
    
  In RuleML Syntax:

<SubsumesPlus>

<Rel>superClass</Rel>

<Rel>subClass</Rel>

</SubsumesPlus>

*And so on...*

### The next step is to Input the taxonomy into OO jDREW

We can do this by two formats, one is POSL and other is RDFS, once we have selected the format, we can copy and paste into the text area.

Once the taxonomy information has been parsed, switch to the "Type Query" to query the taxonomy.

The query will give us various results like True or False and also various variables and bindings as per our query.

At the end these definitions should refer to a coherent selection of Web pages containing HTML5 Microdata.

**TOOLS USED**

* Google's [Rich Snippets](http://www.google.com/webmasters/tools/richsnippets) Testing Tool
* Henri Sivonen's [Validator.nu](http://validator.nu/)

For validating and testing the result.

**REFRENCES**

[1] Schema.org (Type Hierarchy): http://schema.org/docs/full.html

[2] Schema.RDFS.org: <http://schema.rdfs.org/>

[3] OO jDREW (Taxonomy Querying System)

[4] http://www.jdrew.org/oojdrew/docs/TaxonomyQuerying.html