

Homework 1: URI and URL specifications.

Morales Flores Víctor Leonel

15 de Febrero de 2019

1 Introducción

Hoy en día se usan muchos protocolos para buscar y recuperar documentos. Los sistemas tienen como objetivo lograr la búsqueda global y la lectura de documentos en diferentes plataformas informáticas. El concepto de conjunto universal de todos los objetos y conjunto universal de todos los nombres y direcciones se vuelve importante debido a la necesidad de acceder y mezclar objetos que forman parte de diferentes sistemas.

2 URI

El Identificador Universal de Recursos (URI) es un miembro codificado y etiquetado para encapsular un nombre en cualquier espacio de nombre registrado y poder etiquetarlo para que forme parte del conjunto universal.

La sintaxis de un URI está pensada para ser extensible, poder codificar cualquier esquema y que pueda ser imprimible. Un URI completo debe tener un especificador de esquema de nombres que es seguido por una cadena que se obtiene en función del esquema usado.

El esquema del nombre del objeto es el primero en aparecer en un URI, que se separa del resto del objeto con el uso de dos puntos(:). Posterior a los dos puntos va el “path” que se generará e interpretará dependiendo del esquema que se use. Independiente al esquema, la aparición de una barra inclinada sugerirá una estructura jerárquica.

Dentro de los URIs se encuentran caracteres reservados a los que se les asignó un uso especial.

- Porcentaje(%): Se usa como un carácter de escape en el esquema de codificación y su uso no permite nada más. Existen caracteres que o son permitidos por los URIs y para representarlos en los URIs se hace uno del % seguido de 2 dígitos hexadecimales que hacen referencia al código ISO Latin 1.
- Formas jerárquicas(/): Se usa para delimitar subcadenas que tienen una relación jerárquica. La similitud de esto con los nombres de archivos de los UNIX es una coincidencia y no la debemos confundir ya que los URIs no deben interpretarse como nombres de archivos.
- Hash(#): Se usa para separar el URI de un objeto de un identificador de fragmento.
- Cadena de consulta(?): Su uso es para separar el URI de un objeto y un conjunto de palabras que se utilizarán para hacer consultas a ese objeto.

3 URL

Los URIs que se hacen referencia a objetos a los que se puede acceder a través de protocolos que ya existen se les conoce como “Localizador Uniforme de Recursos”. Direcciones físicas de objetos que son recuperables usando protocolos que ya están desplegados en la red.

Una URL completa contiene un especificador de esquema de nombres seguido de una cadena cuyo formato es una función del esquema de nombres. Para localizadores de información en internet:

- Prefijo: Para que sea un localizador de recursos uniformes como fue definido por el grupo de trabajo de URI la cadena debe comenzar con un prefijo constante “URL:”

- Esquema: El primer elemento dentro de la URL es el nombre del esquema, separado por el resto del objeto por dos puntos. El resto de la URL que va después de los 2 puntos dependerá del esquema.

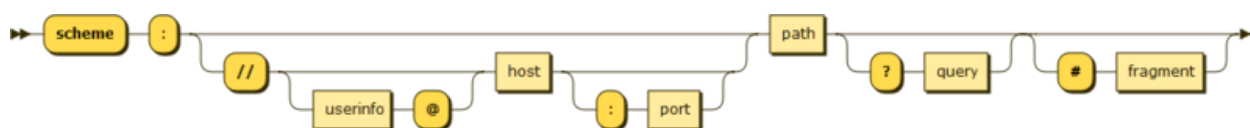
Los esquemas que se refieren a protocolos de Internet suelen tener una sintaxis común. Esta sintaxis comienza con “//” para indicar su presencia y continúa hasta la siguiente “/”. Dentro de esta sección se puede encontrar:

Nombre de usuario opcional	Esto solo cuando es necesario como con pocos servidores FTP. La contraseña va seguida del nombre de usuario separados por dos puntos. El nombre de usuario y contraseña van seguidos de que son anunciados con el signo “@”
Nombre de dominio de internet	Este debe ser dado recomendablemente en formato RFC1037. Otra forma es a través de la dirección IP con un conjunto de 4 dígitos decimales pero es menos aconsejable
Número de puerto	Esto en el caso de que el número de puerto utilizado no sea el predefinido por el protocolo. Es en formato decimal después de dos puntos
Path	Es el resto del localizador. Puede definir cómo se deben comunicar el cliente y el servidor. En general el “/” denota un nivel en una estructura jerárquica aunque esto depende del esquema.

Table 1: Sintaxis común para protocolos de Internet

La estructura final de un URL es:

- URL: scheme: [//[user[:password]@] host[:port]][/path][?query][fragment][%ISO-Latin1-Code]



Algunos ejemplos:

- URL: <https://www.w3.org/Addressing/URL/url-spec.txt>
- <http://www.w3.org/albert/bertram%2Fmarie-claude>