

TP N° 4: XPATH

Introducción

Desde hace un tiempo viene tomando fuerza en el mundo de Internet, la idea de derribar los “silos de información” que representa cada sitio Web, en favor de la portabilidad de los datos entre todos los sitios Web. El objetivo que se busca es estandarizar de algún modo la forma en que se transmite y comparte la información entre sitios Web, la forma en que se definen los intereses de un usuario, un evento público en cierto lugar, la autenticación y autorización entre sitios web, etc.

Si hablamos de buscar estandarizar la forma en que se comparte de forma portable la información entre sitios Web, pensamos entonces en XML como tipo de documento. Así es como surgió hace algunos años el proyecto <http://dataportability.org/> que nuclea a diversas iniciativas tendientes a desarrollar estándares para cada uno de los puntos mencionados previamente.

Entre estos estándares, podemos mencionar:

- OpenID: intenta unificar la autenticación a un sitio web
- OAuth: protocolo para otorgar autorizaciones entre aplicaciones o sitios web
- RSS: ya visto en el TP 1 para el intercambio de noticias.
- APML (Attention Profiling Markup Language)
- OPML, Microformats, RDF, XMPP, etc.

A lo largo de este TP trabajaremos con APML, un estándar basado en XML que intenta homogeneizar la forma en que un usuario de Internet puede describir sus intereses personales. En un escenario ideal, los sitios Web podrían aceptar la entrada de un APML por parte del usuario y entregar de esta forma contenido personalizado según los intereses de éste.

Para poder hacer las consultas pedidas en esta guía, utilizar alguno de los siguientes motores de XPath online (algunos de ellos vistos durante la clase):

- XPathTester: <http://www.xpathtester.com/xpath>
- Free Formatter: <https://www.freeformatter.com/xpath-tester.html>
- CodeBeautify: <https://codebeautify.org/Xpath-Tester>
- Cualquier otro XPath engine online disponible

Ejercicio 1

Para el archivo apml1.xml, escribir las expresiones Xpath para poder obtener:

1. Los nombres de todos los perfiles de atención del usuario. En este caso la respuesta esperada sería: “Work, Home”.
2. El listado de los nombres (key) de los conceptos en el perfil de atención de intereses “Home”. En este caso la respuesta esperada sería: “Golf, Gadgets, Mobile, Política, Tecnología, Software”

3. El listado de los nombres (key) de los conceptos en todos los perfiles de atención . En este caso la respuesta esperada sería: “attention, content distribution, information, business, alerting, intelligent agents, development, services, direct attention, Golf, Gadgets, Mobile, Política, Tecnología, Software”
4. La fecha de actualización de cada una de las fuentes (sources) de los datos implícitos. En este caso la respuesta esperada sería: “2007-03-11T01:55:00Z”.
5. Los nombres de todos los perfiles de atención donde los datos implícitos asociados es un conjunto vacío. En este caso la respuesta sería: “Home”

Ejercicio 2

Para el archivo apml1.xml, escribir las expresiones Xpath para poder obtener:

1. Los conceptos (no importa en qué perfil de atención) que refieren al término "business" en alguna parte. En este caso la respuesta sería: “<Concept key="business" value="0.93" from="GatheringTool.com" updated="2009-03-11T01:55:00Z" />”
2. Los conceptos cuyo valor sea mayor o igual a 0.95 En este caso la respuesta sería: “<Concept key="attention" value="0.99" from="GatheringTool.com" updated="2007-03-11T01:55:00Z" />, <Concept key="content distribution" value="0.97" from="GatheringTool.com" updated="2008-03-11T01:55:00Z" />, <Concept key="information" value="0.95" from="GatheringTool.com" updated="2008-03-11T01:55:00Z" />, <Concept key="direct attention" value="0.99" />”
3. Los nombres de los perfiles de atención del usuario que tengan menos conceptos que el promedio de conceptos por perfil. En este caso la respuesta sería: “Home”
4. Los nombres (key) de los conceptos cuyo nombre (key) empiecen con una vocal. En este caso la respuesta sería: “attention, information, alerting, intelligent agents”

Ejercicio 3

Para el archivo apml1.xml, escribir las expresiones Xpath para poder obtener:

1. El peso promedio de los conceptos de su perfil de intereses personal, del trabajo y para todos los perfiles juntos.
2. El peso promedio de los conceptos para todos los perfiles juntos, agrupados por las fechas de actualización.
3. La cantidad de fuentes que conforman todos sus perfiles de atención. En este caso la respuesta sería 5.
4. El promedio de la cantidad de fuentes de cada uno de los perfiles.

Ejercicio 4

Para el archivo apml1.xml, escribir las expresiones Xpath para poder obtener:

1. El nombre del primer perfil de atención. En este caso la respuesta sería “Work”.
2. El nombre (key) del primer concepto del último perfil de atención . En este caso la respuesta sería “Golf”.
3. Los nombres (key) del tercer concepto (en caso de existir) en los datos implícitos de cada uno de los perfiles de atención. En este caso la respuesta sería: “information”

Ejercicio 5

Para el archivo apml1.xml, escribir las expresiones Xpath para poder obtener:

1. Todos los nodos ancestros de cada uno de los nodos del perfil de atención. En este caso la respuesta sería: “<Body defaultprofile=“Work”> ... </Body>, <APML xmlns=“http://www.apml.org/apml-0.6” version=“0.6” > ... </APML>”
2. Todos los nodos descendientes de los nodos fuentes (sources), incluyendo dicho nodo. En este caso la respuesta sería: “<Sources> <Source key=“http://feeds.feedburner.com/apmlspec” name=“APML.org” value=“1.00” type=“application/rss+xml” from=“GatheringTool.com” updated=“2007-03-11T01:55:00Z”> ... </Source> </Sources>, <Sources> <Source key=“http://feeds.feedburner.com/TechCrunch” name=“Techcrunch” type=“application/rss+xml” value=“0.4”> ... </Source> </Sources>, <Sources> <Source key=“http://feeds.feedburner.com/TechCrunch” name=“Techcrunch” type=“application/atom+xml” value=“0.4”> ... </Source> <Source key=“http://www.uberbin.net/feed” name=“Uberbin” type=“application/rss+xml” value=“0.2”> ... </Source> <Source key=“http://lanacion.com.ar/herramientas/rss/index.asp” name=“LaNacion.com” type=“application/rss+xml” value=“0.3”> ... </Source> </Sources>... </Source> </Sources><Sources...>...</Source>...<Author>...</Author>”

Ejercicio 6

Graficar el nodo contextual para cada uno de las sub-expresiones generadas en las consultas, para los ejercicios 2.1 y 4.3.

Ejercicio 7

Para el archivo apml1.xml, escribir las expresiones XPath para poder obtener:

1. Los nombres de los perfiles de atención distintos que tiene el usuario.
2. La cantidad de conceptos distintos que conforman el perfil de intereses personal (Home). En este caso la respuesta sería 6

Ejercicio 8

Para el archivo apml1.xml, escribir al menos dos expresiones Xpath distintas para poder obtener:

1. Los nombres de todos los perfiles de atención del usuario.
2. Los nombres de todos los perfiles de atención del usuario ordenados alfabéticamente.
3. Los nodos predecesores de todos los perfiles de atención del usuario.
4. Los perfiles de atención del usuario que tengan mas de tres conceptos como parte de los datos implícitos.

Ejercicio 9

Para el archivo asistencia.xml, armar los pares posibles de expresiones XPath equivalentes (conceptualmente iguales), indicando qué resultado arroja cada una de dichos pares. Indicar también el resultado obtenido para aquellas expresiones Xpath que no armen pares (si hubiera).

1. `/work-resources/work-resource[@name` =
`//work-resource/@name[count(..work-exceptions/absent) > 1]]/@name`
2. `//work-resource[count(..work-exceptions/absent) > 2]`
3. `/work-resources/work-resource[count(..work-exceptions/absent) > 2]`
4. `/work-resources/work-resource/@name[count(..work-exceptions/absent) > 1]`