

Ping The Semantic Web

By,

Shantanu Fauji

shan15@umbc.edu

CMSC 691S

Agenda

- What is Ping the Semantic Web
- About the creator
- How does it work
- Content negotiation
- How to deliver files
- How to get the latest pings
- API for Ping the semantic web
- Why should one use PTSW

What is Ping the Semantic web

- web service archiving the location of recently created/updated RDF documents.
- author can notify PTSW that the document has been created or updated by pinging the service with the URL of the document.
- PingtheSemanticWeb.com is used by crawlers or other types of software agents to know when and where the latest updated RDF documents can be found.
- Like a site map, but not limited to a particular website

About the creator

- Created by Fredrick Giasson, a computer scientist and entrepreneur.
- Is also the co founder of the Music Ontology, the Bibliographic Ontology.

Website: <http://fgiasson.com/>

Email: fred@fgiasson.com

- This is third version of pingthesemanticweb and was released in Aug 2007
- First version was released Aug 2006

How Does it work

- You can use the URL of an HTML or RDF document when pinging PingtheSemanticWeb.com web service.
- The file will be updated if it is found valid
- Else it will be discarded
- Also, for each URL, PTSW will try to do content negotiation with the remote web server to try to find RDF/XML, RDF/N3 or RDF/Turtle by this method.

Content Negotiation

- Content negotiation is a mechanism defined in the HTTP specification that makes it possible to serve different versions of a document (or more generally, a resource) at the same URI, so that user agents can specify which version fit their capabilities the best.
- the user agent uses an Accept HTTP header that lists acceptable media types. The server is then able to supply the version of the resource that best fits the user agent's needs.

Accept-Language: fr; q=1.0, en; q=0.5 Accept: text/html; q=1.0, text/*; q=0.8, image/gif; q=0.6, image/jpeg; q=0.6, image/*; q=0.5, */*; q=0.1

Content Negotiation Continued..

- The Accept header parameter sent by PTSW is the following one:

Accept: text/html, html/xml, application/rdf+xml;q=0.9,
text/rdf+n3;q=0.9, application/turtle;q=0.9,
application/rdf+n3;q=0.9, */*;q=0.8

- If the Web server returned a text/html document, then PTSW will send a second query to try to get a RDF document too with the following Accept header parameter:

Accept: application/rdf+xml, text/rdf+n3,
application/turtle, application/rdf+n3

Delivering a file to PTSW

- Linking a RDF document from a HTML file

RDF: `<link rel="meta" type="application/rdf+xml" title="name-of-the-ontology-used" href="[url]" />`

- Example

Getting the latest Pings

- First you have to register an account on PTSW.
- Once the account is created, and the IP registered, the server can start downloading the xml file describing all the latest pings here available at this URL:
<http://pingthesemanticweb.com/export/>
- The data is returned in XML format

Data Format

- All the export files conform to the following XML document structure:
- ```
<pingthesemanticwebUpdate version="1.4"
updated="2007-07-31 11:20:54">
 <rdfdocument
url="http://b4mad.net/datenbrei/index.php?sioc_type=p
ost&sioc_id=300" created="2006-08-11 11:21:00"
updated="2006-08-14 09:57:26" serialization="xml"
ns="http://www.w3.org/1999/02/22-rdf-syntax-ns#
http://xmlns.com/foaf/0.1/ ..." />
</pingthesemanticwebUpdate>
```

## Data format continued..

- The `<pingthesemanticwebUpdate>` element

Each export file contain a single `pingthesemanticwebUpdate` element. It has two attributes: `version`, `updated`; and any number of `rdfdocuments` sub-elements.

- `version` is a number. It is the version of the current export file format.
- `updated` is a string, it indicates when export file was requested/created.

## Continued..

- The `<rdfdocument>` element
- `rdfdocument` has five attributes: `url`, `created`, `updated`, `serialisation`, `namespace`:
- `url` is a string; it is the URL of an updated RDF document.
- `created` is a date; it is the time when this document as been pinged to `PingtheSemanticWeb.com` for the first time
- `updated` is a date; it is the time it have been updated for the last time.
- `serialization` is a string; it is a string that tell which serialization method is used to write the document; this variable can have the value `"xml"` or `"n3"`.
- `ns` is a string; it is a list of space (`%20`) separated namespace(s). If a namespace appears in this list, this mean that a resource has been typed (`rdf:type`) with a class defined in that ontology.

# Setting up a ping service interface

- Any web service can communicate with PingtheSemanticWeb.com via two interfaces: REST and XML-RPC.

- XML-RPC

It's a spec and a set of implementations that allow software running on disparate operating systems, running in different environments to make procedure calls over the Internet.

It's remote procedure calling using HTTP as the transport and XML as the encoding.

## Continued..

- REST

Representational state transfer (REST) is a style of software architecture for distributed hypermedia systems such as the World Wide Web.

The term is often used more loosely to describe any simple interface which transmits domain-specific data over HTTP without an additional messaging layer such as SOAP or session tracking via HTTP cookies.

# REST vs RPC

- A RESTful web application requires a different design approach from an RPC application.
- Before a client communicates with the application through RPC it must have knowledge of the object identity in order to locate it and must also have knowledge of the object type in order to communicate with it.
- REST design seeks to define a set of resources with which clients can interact uniformly, and to provide hyperlinks between resources which clients can navigate without requiring knowledge of the whole resource set.

# XML-RPC client

- The XML RPC request should be constructed based on:  
**RPC endpoint:** `http://rpc.pingthesemanticweb.com/`  
**Method name:** `weblogUpdates.ping`  
**Parameters:** (should be sent in the same order as listed below)

**name of site** (string, limited to 1024 characters [this field is not used by PingtheSemanticWeb for the moment so this string can be empty])

**URL of site or RDF document** (string, limited to 2048 characters)



# XML RPC response

- The XML RPC response will have a Struct with two members:

**flerror** (Boolean) which is true/ 1 if an error occurred

**message** (string) which contains "Thanks for the ping." (if successful) or an error message.

# Example of XML-RPC request

POST /ping/xmlrpc HTTP/1.0

User-Agent: Java

Host: rpc.pingthesemanticweb.com

Content-Type: text/xml

Content-length: 250

```
<?xml version="1.0"?>
```

```
<methodCall>
```

```
 <methodName>weblogUpdates.ping</methodName>
```

```
 <params>
```

```
 <param>
```

```
 <value>Some title page</value>
```

```
 </param>
```

```
 <param>
```

```
 <value>http://www.talkdigger.com/conversations/some_conversati
on.html</value>
```

```
 </param>
```

```
 </params>
```

```
</methodCall>
```

# XML RPC Response

- XML Response
- **Possible XML-RPC errors**

1. This document have already been updated by PingtheSemanticWeb.com in the last hour.
2. Ping the Semantic Web is not allowed to index this URL.
3. An unexpected error occurred.

# REST client

- PingtheSemanticWeb.com accepts form-based HTTP POST and GET requests . For example, the following is a valid HTTP GET ping request:

URL: `http://pingthesemanticweb.com/rest/?url=[url]`

- [Demo](#)

# Why use this service?

- Validated RDF resources

In the version 2.0, PTSW was doing a pseudo validation of RDF files. In the version 3.0, it fully validates RDF documents. This means that all pings the service export are valid RDF documents.

This is a major upgrade to the system since now all agents requesting pings from PTSW will know that each of them are valid RDF documents. That way, they will save time and bandwidth since they won't try to process bad RDF documents.

## Continued..

- Simplified pings list export system

Now all ping consumers need to be registered to the PTSW web service. This simple registration greatly helps consuming pings coming from PTSW. Additionally, he has to setup his [pings retrieval preferences in the user account section](#).

PTSW is archiving the time of the latest request of the user.

## Continued..

- Faster Pinging Interfaces

The web service is now hosted on a much bigger server.

Switched from MySQL to Virtuoso. These changes result in a more powerful service

can handle up to 5 million pings per day

## Continued..

- New statistics

[All stats about Namespaces](#). This is the list of namespaces used to describe entities in RDF. There is currently **702 used namespaces** know by PTSW.

[All stats about Types](#). This is the number of typed entities defined in each RDF document know by PTSW. If the same entity (URI) is defined in two different RDF documents, the type of the entity will be calculated twice. So take these numbers as a good approximation, but not as an absolute truth. There is currently **9534 types** know by PTSW.



# Questions



*Thank You*