SCIDIP-ES

SCIENCE DATA INFRASTRUCTURE FOR PRESERVATION - EARTH SCIENCE

Webserver Registry

Overview

Current Version: M30 Release - Based on SVN ver. 6665
Last Updated: 10 June 2014
Original Author: Simon Berriman

Table of Contents

Document Control

Date	Version	Description	Author
22/09/2013	M24 Draft - Based on SVN ver. 5970	Initial version	Simon Berriman
31/10/2013	M24 Release - Based on SVN ver. 6014	Updated for frst release	Simon Berriman
23/02/2014	M30 Release - Based on SVN ver. 6665	Updated for second release	Simon Berriman

Introduction

The Webserver Registry is a simple implementation of the concept of a SCIDIP-ES Registry. Its name is a bit a misleading; it started out life as a mere collection of fles on a webserver (from where it got its name), however it has since been properly rewritten and continues to be actively developed.

The processing is provided by a small set of PHP fles, which operate in conjunction with a custom Apache HTTPd confguration in order to provide a valid RESTful API. It also incorporates a Python script, run via a cron job, to build the search indexes.

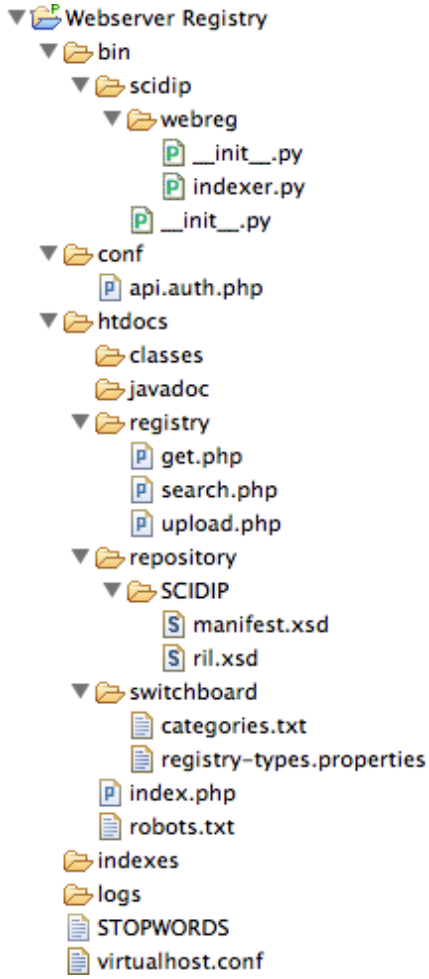
The implementation is deliberately kept minimalistic, and its stateless design means that it is inherently scalable; the simplest upscaling solution being to use a load balancer to distribute requests to multiple HTTPd servers.

Structure

The base installation is available in SVN from the following path:

```
/SCIDIP-ES/software/common/webserver-registry
```

Under this directory you will find the following file structure:



Prerequisites

To run the Webserver Registry, the following software needs to be installed and confgured in advance:

- Apache HTTPd version 2.2+ (i.e. the Apache webserver)
  - sudo apt-get install apache2
- PHP version 5.3+
  - sudo apt-get install php5
- Python 2.7+
  - sudo apt-get install python

Installation

The registry itself is designed to run as an Apache HTTPd named vitualhost website. That is not to say it cannot run on any other webserver, only that to do so would require a customised connection devised by the user. The 'virtualhost.conf' file is intended to be referenced as a named virtualhost using an `include` directive. It will firstly need to be modified to reference the new installation's actually filing system location and intended domain name. It should then be linked into the HTTPd configuration with:

/var/scidipwebreg

```
<VirtualHost *>
    Include /path/to/virtualhost.conf
</VirtualHost>
```

with the file path adjusted for is actual location.

The root of the website is the `htdocs` directory. The subdirectory `registry` contains the three PHP fles, and RepinfoLabels and Manifests will be placed in a sub-hierarchy from here. The `repository` subdirectory is not necessary unless the same virtualhost is intended to be used as a fle store. Also, the `javadoc` and `switchboard` directories can be disregarded under normal circumstances.

T h e `conf` directory contains a PHP file holding the user credentials for the HTTP security. At the moment this holds the password unencrypted, however this will change to a digest in due course.

The `bin` directory contains the Python indexer. This needs to be installed as a cron job. The following example command could be used to run the indexer every 15 minutes:

```
*/15 * * * * cd /path/to/bin && /usr/bin/python -m scidip.webreg.indexer
```

ISSUE: no /javadoc/framework

ISSUE: username/password required - apifull/scidip123

ISSUE: indexer error: hard coded /opt/sites/scidip/logs/indexer.log [Had to create the directory - not clear where it is defined]

The crontab is edited from a Linux command line with:

```
# crontab -e
```

For an introduction to cron, a good tutorial can be found at:

https://help.ubuntu.com/community/CronHowto

The indexer places its generated fles in the `indexes` directory. The `STOPWORDS` fle contains a newline separated list of words which the index should ignore; this should be edited as needed.

Finally, output and log files are placed in the `logs` directory.