Intelligent Information Management
Targeted Competition Framework
ICT-2011.4.4(d)

Project **FP7-318652 / BioASQ**

Deliverable **D1.1**

Distribution **Public**

http://www.bioasq.org

Project Wiki

Sergios Petridis, Aris Kosmopoulos and Konstantinos Stamatakis

Status: Final (Version 2.0)

January 2013

Project

Project ref.no.	FP7-318652
Project acronym	BioASQ
Project full title	A challenge on large-scale biomedical semantic indexing and question answering
Porject site	http://www.bioasq.org
Project start	October 2012
Project duration	2 years
EC Project Officer	Ms Martina Eydner

Deliverable

Deliverabe type	Report
Distribution level	Public
Deliverable Number	D1.1
Deliverable title	Project Wiki
Contractual date of delivery	M3 (December 2012)
Actual date of delivery	January 2013
Relevant Task(s)	WP1
Partner Responsible	NCSR “D”
Other contributors	
Number of pages	15
Author(s)	Sergios Petridis, Aris Kosmopoulos and Konstantinos Stamatakis
Internal Reviewers	-
Status & version	Final
Keywords	document management system, project collaboration tools, project site

Contents

List of Figures

1

Introduction

According to the the BioASQ DoW:

> *The project Wiki will form the document repository and project portal, allowing access to all documents produced by the consortium by common Web browsers.*

The BioASQ Project Wiki has been implemented accordingly and has two main components:

- the BioASQ site, accessed publicly at `http://www.bioasq.org/` and
- the BioASQ documents' repository, accessed by BioASQ members at `https://idefix.iit.demokritos.gr/BioASQ/` (main URL) and `https://lshtc.iit.demokritos.gr:8443/BioASQ/` (web front end).

Note that, though the system is referred as a *Wiki*, the term is used in its broader sense, i.e. as a project collaboration tool for creating, co-editing and sharing content. Though deploying a Wiki system nowadays is straightforward from a technical point of view (e.g. a Drupal module on the project site), early conversations between the project members suggested that the proposed set of tools are more appropriate for collaboration within the project rather than the standard Wiki interface.

This deliverable serves mainly as a documentation reference of the BioASQ Project Wiki. Chapter 2 describes the components of the Project Wiki and their interactions, while Chapter 3 describes how the user can interact with the Project Wiki to have access or contribute information.

2

System Components

2.1 Document Management System

The aim of the Document Management System is to provide a unique resource for all the documentation and reporting created within the project. It consists of a revision control system and a complete web front end interface, described in the Sections 2.1.1 and 2.1.2 respectively.

2.1.1 Revision Control

The BioASQ Document Management System supports revision control. Revision control, also known as version control, refers to the management of changes to documents or other collections of information. Changes are usually identified by the *revision number*. Each revision is associated with a time stamp and the person making the change. Revisions can be compared, restored, and with some types of files, merged. Several open source revision control system exist nowadays, operating in either the client-server model (e.g. CVS, Subversion) or in the distributed model (e.g. Git, Mercurial, Bazaar, Darcs).

We have opted to work with the Subversion revision control system. Our choice has been made by considering:

- the experience of most members of the consortium with the given system,
- reliability resulting from software maturity,
- clients existing in both Windows and Linux operating systems,
- the existence of a convenient web-client for read/write access.

Figure 2.1 shows the top-level view of a working copy, using the RabbitVCS Subversion client, integrated into the Nautilus file viewer (Linux). A similar interface is provided for Windows using the TortoiseSVN client.

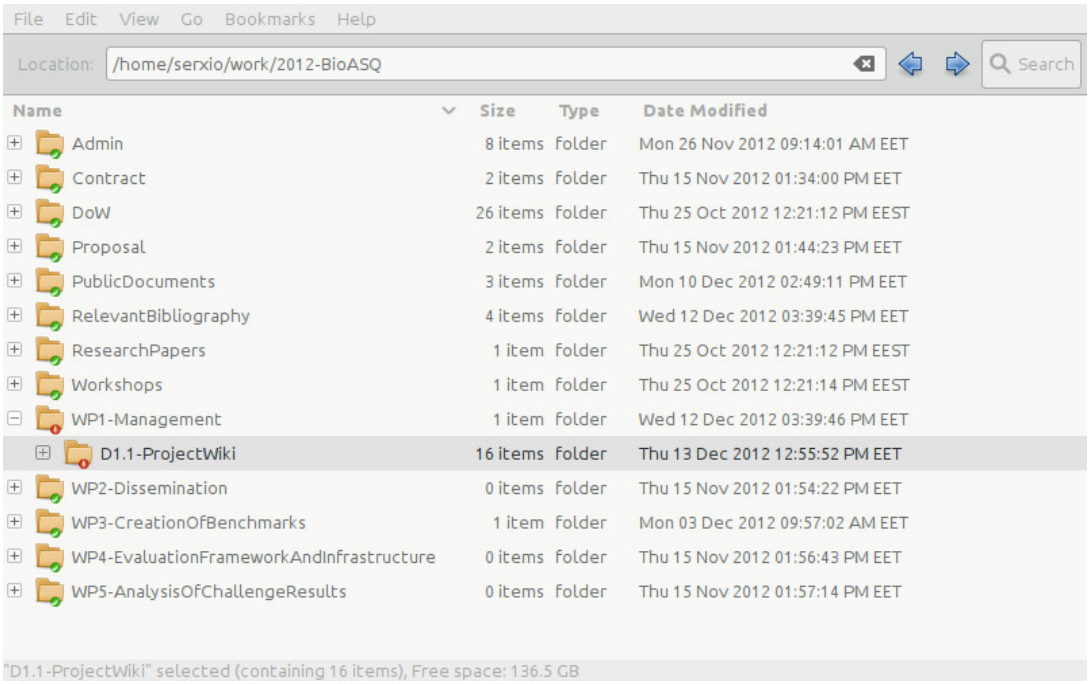

Figure 2.1: Working copy of the Subversion Repository. Items with a question mark indicate changes on the PC that are not yet committed to the BioASQ server.

2.1.2 Web front end

The BioASQ consortium includes computer specialists and others that are not. For users that are not computer specialists, working with version control systems, such as Subversion, would be a non-trivial task. Therefore, we have opted for the solution of letting these members interact with the version control via a web browser, hiding as much as possible the underlying complexity of a versioning system, and allowing them to interact in the same way as with any other document management tool. To that end, we have considered the following web-interface options:

WebSVN[1] offers a view onto subversion repositories, reflecting the Subversion methodology. One can view the log of any file or directory and see a list of all the files changed, added or deleted in any given revision. The differences between two versions of a file are also visible, allowing one to see exactly what was changed in a particular revision.

However, WebSVN is limited to browsing and therefore it does not allow for uploading or modifying the repository content.

SCMManager[2] offers similar features as WebSVN while additionally supporting interaction with other versioning systems, such as Git and Mercurial. Similarly with WebSVN, it does not support uploading or modifying the repository content.

Polarion Free Subversion Tool[3] is a handy SVN client that enables Subversion users to work with SVN repositories using a web browser. When installed on your Subversion server, this free SVN web client provides a convenient way to browse a Subversion repository (content and history), plus perform simple SVN write operations. It can connect to any Subversion repository that provides HTTP access. Its features include:

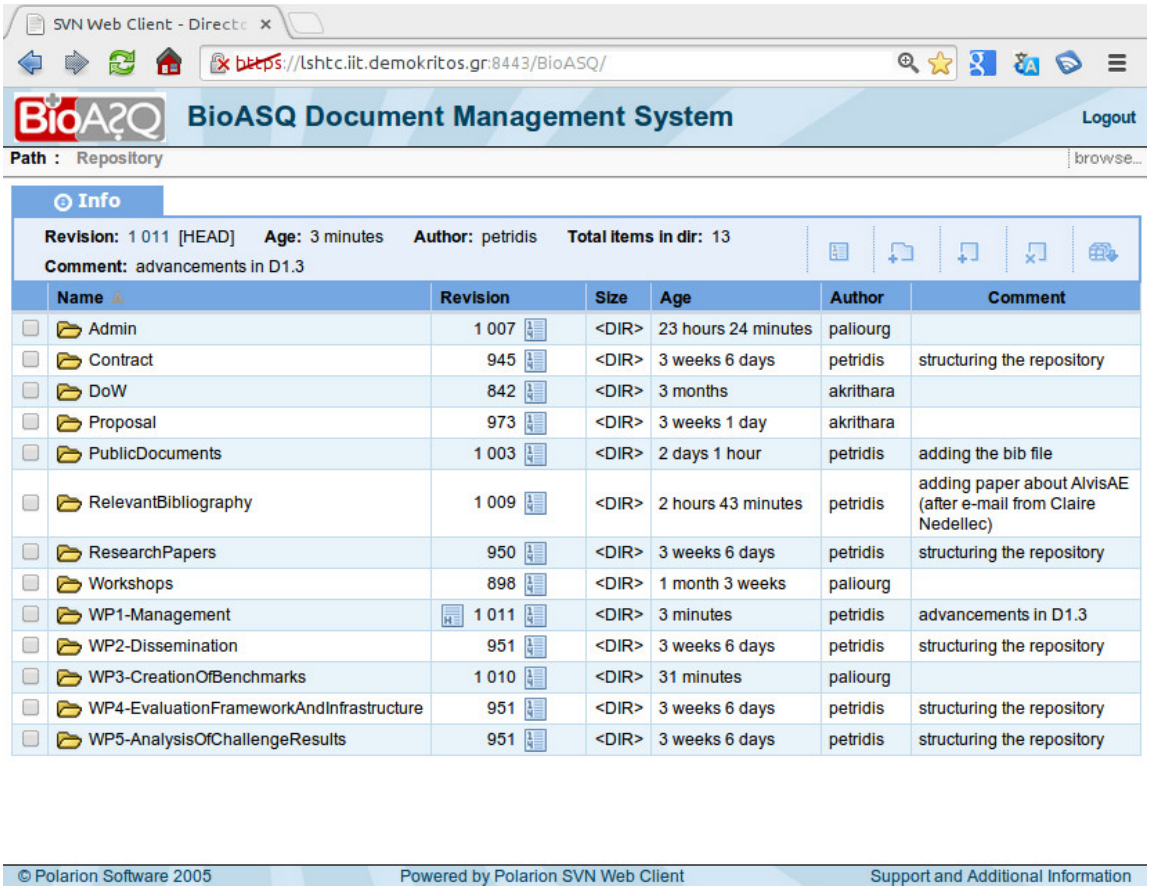

Figure 2.2: Read/write Web front-end Document Repository. Top level view

- Browsing Subversion repository folders and files
- Browsing and comparing revisions of SVN folders and files
- Viewing revision details and comparing SVN revisions
- Creating/deleting/modifying files in a Subversion repository
- Creating/deleting SVN repository folders
- Easy browsing of SVN branches/tags

Given that the Polarion Free Subversion Tool has been the only tool that supports uploading, removing or modifying existing content within the repository directly through the web browser, we have opted for this solution. Installing the software required some extra care, since in its current version, it only supports http connection, thereby demanding extra security steps to be taken. We also modified its layout to use project information, such as the BioASQ logo.

Figure 2.2 shows the top level view of the web front end subversion repository.

2.2 The Project's Web-site

The project site (`bioasq.org`) has been designed to be used as a portal to provide information for both the project itself and the challenge organisation. The website is made with the Content Management System

(CMS) Drupal[4]. Here are some reasons for which Drupal has been chosen.

- It is open source .
- It's a mature platform, used by thousands of organizations and businesses and supported by a world-wide development community. It is also continuously being monitored for security issues by a large community of developers.
- It's easy to use, it allows every-day users to add and update web pages without technical assistance.
- It is dynamic and customizable, it can be extended by adding new or by modifying existing features.

An important number of Drupal modules has been used incuding: admin_menu, biblio, calendar, ctools, date, emogrifier, faq, feeds, filetree, freelinking, google_analytics htmlmail, insertFrame, job_scheduler, libraries, linkimagefield, mailmime, mailsysteml, pathologic, path_access, popup, realname, simplenews, tagadelic, token, views_slideshow

2.3 Security Issues

The project documents may contain critical information about the challenge. Therefore, particular cautions need to be taken in order to guarantee that project information is only accessible by the authorized users. In particular

- Network access to any document content is given after a user has been authorised using his credentials. As an exception, public project documents can be accessed anonymously through the project web-site
- Network communication project restricted (non-public) documents and for user authentication is done using encryption mechanisms, in particular the Secure Sockets Layer (`ssl`) protocol.

Access to the documents from outside the NCSR "D" local network is provided only to authorised members through `ssl` encryption. Plain `http` access to the repository is given only to the server that hosts its web interface. the reason for this exception is that `http` is the only working port for the web-interface-to-repository communication. However, since both the subversion repository and its web-interface servers are protected from a further firewall within the NCSR "D" premises, this exception is not a threat (see Figure 2.3).

[4] http://drupal.org/

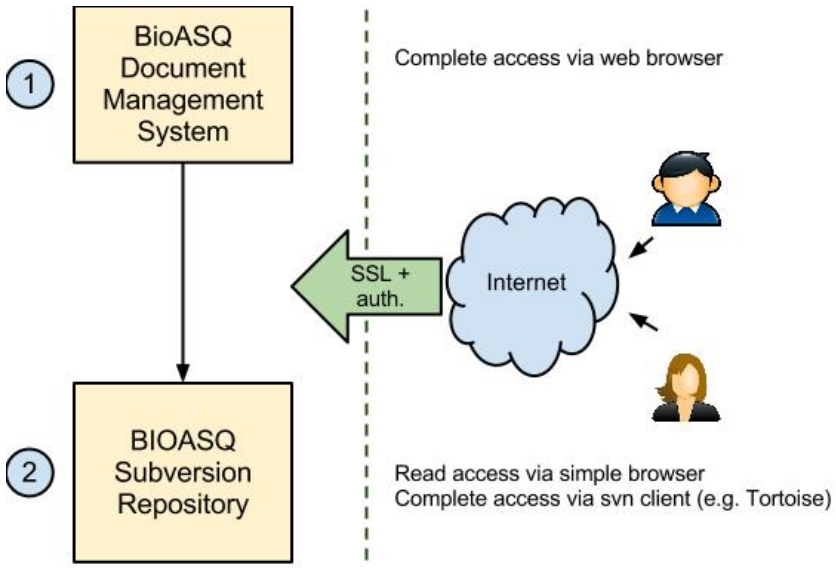

Figure 2.3: Accessing BioASQ document repository system.

3

Interacting with the system

3.1 Interacting with the Document Management System

3.1.1 Authorization

To access any project-restricted content, the user should use his credentials. i.e. his user name and password. The same credentials apply to all interfaces (subversion client, repository web-interface, project web site).

3.1.2 Content organisation

All documents are represented by files, organised into a directory tree-structure. Figure 3.1 shows the first two levels of the current directory structure.

In general, access to the contents of the DMS is restricted to the consortium. As an exception, all documents that can be accessed publicly are placed into the `PublicDocuments` directory. The contents of this folder are publicly accessible through the BioASQ web site. Contents of the `RelevantBibliography` folder are also publicly accessible in the same way.

Note that challenge-specific data, such as biomedical articles and/or abstracts in their original or preprocessed form are not stored in the Document Management System.

3.1.3 Adding, Removing and Modifying content

Uploading content can be done either through a subversion client or through the web interface. For experienced users, the subversion client is recommended.

Figure 3.1: Directory Tree Structure of BioASQ DMS

Using a subversion client

The first time one wishes to upload content, some preparatory steps are required to set up the environment.

1. First, the Subversion client of one's choice needs to be installed. For Windows users, a popular choice is Tortoise SVN. For Linux users, beside the command line tools, one may try RabbitVCS.
2. Using the Subversion client, one should check out a working copy of the BioASQ document repository. The BioASQ repository URL is `https://idefix.iit.demokritos.gr/BioASQ/`. Sufficient space should be available on the hard disk to hold the entire repository.
3. To complete this operation, one needs to provide his credentials, provided by the BioASQ administration office.

Each time one wishes to interact with the repository, he needs to make sure his working copy reflects the latest. revision of the BioASQ repository. To do so, the client needs to "update" the working copy to the "head" Subsequently, changes to the files of the repository (add/remove/modify) can be made and committed.

Using the web interface

For those who do not wish or cannot use an SVN client for full read/write repository access, the web frontend (based on the Polarion Free Subversion Tool), can be accessed at `https://lshtc.iit.demokritos.gr:8443/BioASQ/`. For authentication, one needs to use the same user name and password as for a Subversion client. Once this is done, will be gained access to exactly the same files, through a more user friendly interface. One will also be allowed to add, delete and modify files. Checking differences between revisions is also possible. Using this interface should be straightforward. Modifying the contents of a file is perhaps the most tricky part, and here are some instructions that may help:

- Click on the file that you wish to modify.
- The Screen `Info` appears - hover over the three icons at the right to see their labels `Revision List`, `Commit` and `Download`.
- Click `Download`, to download the file you wish to modify on your computer,
- Update the file using your editor, word processor etc.
- Click `Commit`, to commit the file back to the repository.

Note that after adding,deleting or modifying a file through the web interface, a new SVN revision is committed. It is expected that the web interface is somewhat slower than the direct `https` link.

3.1.4 Making a document public

To make a document public, one should do the following

- go to an updated working copy of the `PublicDocuments` folder of the repository. For a command line subversion client, this looks like:

```
cd PublicDoccuments; svn up
```

- Place a copy of the document in the `PublicDocuments` folder, in PDF format.

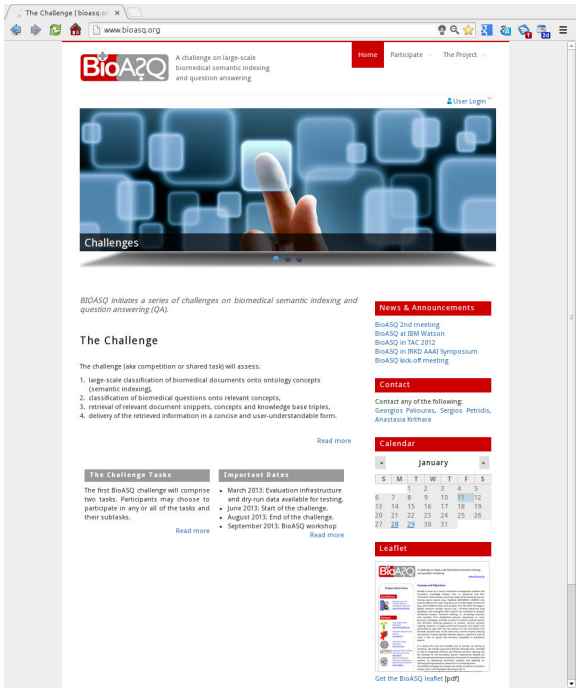

Figure 3.2: The home page of the BioASQ web site

- Add it to the repository

  ```
  svn add 2012-Paliouras-BioASQ.pdf ;
  ```

- Add the corresponding bib file with a single bib entry describing the file. Take care to include the field `URL`, in order to create a clickable link to the full text.

  ```
  echo "@conference{2012-Paliouras-BioASQ,
  title={BioASQ: A challenge on large-scale biomedical semantic indexing and
  question answering},
  author={Georgios Paliouras},
  note={Presented at the Text Analysis Conference (TAC), NIST, Gaithersburg
  Maryland USA},
  year={2012},
  month={November},
  URL={2012-Paliouras-BioASQ.pdf}
  }" > 2012-Paliouras-BioASQ.bib
  ; svn add 2012-Paliouras-BioASQ.bib ;
  ```

- commit the changes

  ```
  svn commit
  ```

- check that the uploaded document is accessible through the project web-site (see also Section 3.2.1).

Figure 3.3: A challenge-related information page of the BioASQ web site

3.1.5 Adding relevant bibliography

To add a relevant bibliography item, one should follow the same steps as those needed to make a document public, with the following differences

- The repository path to add the bib file is `RelevantBibliography`
- Adding the full text of the reference to the repository is not mandatory, though recommended. However, the corresponding file should not be accessible through the web site.
- The URL field of the bib file should link to an external source (e.g. the article publisher's web site).

3.2 Browsing the BioASQ Web-Site

The BioASQ web-site url can be found `http://www.bioasq.org/`. It has been designed to serve both as a reference web-site for the BioASQ project and as the web-site for the challenge itself. It supports anonymous browsing, as well as restricted-content browsing after authentication, for specific groups such as BioASQ project members, the BioASQ advisory board and the BioASQ biomedical experts.

3.2.1 Site's structure

The home page

Figure 3.2 shows the home page of the BioASQ web site. The home page gives emphasis to the challenge itself, rather than the project. The idea is that a reader not directly familiar with BioASQ, but interested

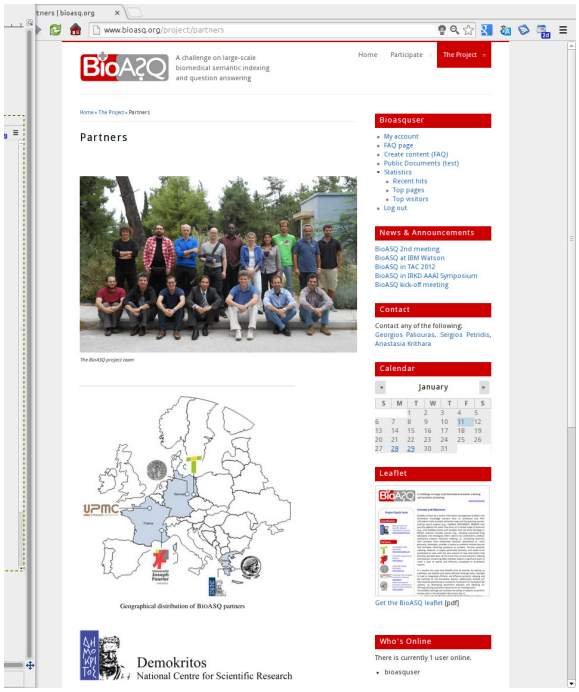

Figure 3.4: A project-related information of the BioASQ web site

in the challenge, should directly get information about the challenge, while also be able to navigate to the project information later. Therefore, the user is given a quick overview of the BioASQ challenge, including its aim, the tasks, important dates and contact information. News and announcements as well as a project-related calendar are accessible from a right-side column. Their information is regularly updated by the BioASQ administration team.

Furthermore, links at the top of page allow the user to navigate to either to the more detailed challenge-related pages (“partipate” menu) or the project-related pages. At any time, the user may give his credential in order to login as project-related user.

The “participate” menu

Figure 3.3 shows a view of the web-site accessible from the ”participate” menu. This menu is addressed to a potential BioASQ contestant. It gives all information related to the challenges, the data and the schedule. These pages will contain direct links to allow registration of interested users to the challenge, once registration options are available.

The “project” menu

Figure 3.4 shows a view of the web-site accessible from the ”project” menu. This menu is addressed to the visitor that wishes to know more about the BioASQ project. It provides information for its objectives and the consortium. It also gives link to all public documents the project releases as a well a ”press room” link with dissemination material, such as leaflets and posters.

Finally, it also features a ”related content” page that links the web-site to other relevant challenges, projects and events sites of interest.

Figure 3.5: The BioASQ web site view for an authorised user

3.2.2 Logging and Adding content

Users and User groups

Besides anonymous browsing, the BioASQ web site supports user login that allows access or editing specific content, such as forums. In particular, three user groups exist: the BioASQ project members, the BioASQ advisory board and the BioASQ biomedical experts. Once logged in, a special bar appears at the top of the browsing window that provides the user with the extra functionality (see Figure 3.5). Notably, some of the users from the above groups are also allowed to directly add content to the site, such as announcements, articles, FAQ topics and forum topics.

Forums

The BioASQ site enables the possibility of hosting forums in order to facilitate the discussion among users of the same group (e.g. the advisory board,) or, people from different groups. In the BioASQ forum (www.bioasq.org/forum), a user will be able to initiate a discussion by creating a discussion topic, post comments as well as answer to other users comments.

A

Installation Notes

In this section, we document the steps to been taken in order to set up the BioASQ document management system. In our installation, both the `svn host` and `web-svn host` servers run on recent versions of Ubuntu (Linux kernels 3.2.0 and 2.6.32).

- Set up the SVN repository at `svn host`. The proposed path to use is `/var/svn/BioASQ`. Read/write permissions should be given to the group `svnusers` that includes `www-data` (to allow Apache) and `svnowner` (to allow `svnserve`). Use something like

  ```
  chown -R svnowner /var/svn/BioASQ  ; chgroup -R svnusers /var/svn/BioASQ.
  ```

  if you need to relocate the repository from elsewhere, you may use the `svnadmin dump/load` utility. Clients should then use `svn switch –relocate` or `svn relocate` depending on the SVN version.

- set up SVN access protocols at `svn host`
  - set up the SVN protocol at the `/etc/init/` directory, add the fie `svnserve.conf` with the following content

    ```
    description ”Subversion server”
    start on (local-filesystems and net-device-up IFACE!=lo)
    stop on runlevel [016]
    exec /bin/su -c ’exec /usr/bin/svnserve -d -r /var/svn/’ svnowner
    ```

    The SVN client should then be able to access the repository through `svn://svn host/BioASQ`.
  - configure `http/https` access at the `/etc/apache2/mods-enabled` directory of `svn host`, edit the `dav_svn.conf` file.

    ```
    ##################### BioASQ --- BEGIN
    #### BioASQ with SSL
    <Location /BioASQ>
    DAV svn
    SVNPath /var/svn/BioASQ
    SSLRequireSSL
    AuthType Basic
    AuthName ”BioASQ Subversion Repository”
    ```

```
AuthUserFile /etc/apache2/BioASQ.passwd
Require valid-user
</Location>
#### BioASQ without SSL
<Location /svn/BioASQ>
DAV svn
SVNPath /var/svn/BioASQ
AuthType Basic
AuthName ”BioASQ Subversion Repository”
AuthUserFile /etc/apache2/BioASQ.passwd
Require valid-user
</Location>
##################### BioASQ --- END
```

to put SVN users passwords, use the command

```
sudo htpasswd -m /etc/apache2/BioASQ.passwd username
```

The web browser or SVN client should then be able to use both `http` and `https` access.

- configure port access at `svn host`
  - Within a local network, the repository should be accessed using the SVN protocol (port 3690)
  - Also the `web-svn host` should have http access - The Polarion Free Subversion Tool software needs it (port 80)
  - Anyone should have `https` access (port 443)
  - `ssh` connection should also be enabled (ports 222 or 22)

You may use the ufw tool to create the firewall - do not forget to enable it by default on start up. The gufw tool is a nice graphical interface to ufw. In the end, you should test that the following command gives a similar result

```
serxio@\svnHost:~$ sudo ufw status
[sudo] password for serxio:
Status: active

To                         Action      From
--                         ------      ----
222                        ALLOW       Anywhere
22                         ALLOW       Anywhere
443                        ALLOW       Anywhere
80                         ALLOW       143.233.226.26
3690/tcp                   ALLOW       Anywhere
3690/tcp                   ALLOW       Anywhere (v6)

443                        ALLOW OUT   Anywhere
```

- install the Polarion Free Subversion Tool software
  - fetch the the Polarion Free Subversion Tool software
  - modify the `web.xml` file so that it links to the `svn host http` URL with a valid user name and password. Other elements such as title can also be modified.

  - check http://forums.polarion.com/viewtopic.php?f=6&t=3092 to correct the bug for uploading the file
  - deploy war in the appropriate path, depending on the installation (e.g. /var/lib/tomcat6/webapps)

- configure port access at web-svn host.

  make sure that the tomcat ssl port (in our installation, 8443) is accessible from the outside (if not, search for possible responsible firewalls)