Use DSpace Statistics (JSP)



Files:

- *dspace*/**bin/stat-** (i.e. all files beginning with "stat-")
- *dspace*/config/dspace.cfg
- *dspace*/config/dstat.cfg and dstat.map
- `http://web-address-to-my-dspace/statistics` (URL of DSpace statistics page)

Instructions:

1. DSpace comes with a very basic web statistics package which you can use to gather information on page views (not the same as file downloads) and general activities performed within DSpace (e.g. searches performed, logins, errors and warnings). This statistics package will analyze and parse all of your old DSpace logs (in *dspace*/log) to generate HTML statistical reports. It consists of a set of Perl scripts all beginning with "stat-" in your *dspace*/bin/ directory. (Note: In order to utilize these scripts you must have Perl installed!)
2. There are two basic options in dspace.cfg which allow you to specify whether or not your final statistics reports should be publicly accessible (`report.public`), and where the final statistical reports should be stored (`report.dir`). It's very possible you may not want to change either of these. However, if you change either option, don't forget to restart Tomcat (See Quick Restart in Rebuild DSpace)
    - report.public = false
    - report.dir = ${dspace.dir}/reports/
3. The DSpace Statistics package comes with its own configuration file, dstat.cfg. This file has some settings which you must change (marked with an asterisk below), and a few that you may wish to review:
4. * `dspace.log`*- the full path of your *dspace*/log directory
    - `general.summary`- actions listed in the DSpace log file which you want to list in the "Overview" section. You don't need to change these unless needed.
    - `exclude.word`- stopwords to filter out of search terms in statistics
    - `exclude.type`- Lucene search index terms to filter out of statistics (corresponds to the search indicies, see (Modify search fields)
    - `exclude.character`- Lucene special characters to filter out of statistics
    - `item.type`- Item types to find statistics for. Corresponds to the form values (see Change a form value) defined for your `dc.type` field, or any metadata field with an element named `type`.
    - `item.floor` and `search.floor`- specify the number of minimum accesses necessary before an item or search term is listed in statistics
    - `item.lookup`- specifies the maximum number of items to list Author/Title information for in statistics (all other viewed items are listed by URL)
    - `user.email`- specifies whether to display user email information in login statistics. For privacy, this defaults to `false` (i.e. do not display email)
5. * `host.name` and `host.url`* - The name and URL of your DSpace install, which will be displayed at the top of the statistics page.

6. In addition, the DSpace Statistics package comes with a mapping file (dstat.map), which maps DSpace "actions" which appear in the log files (in *dspace*/log) into human readable text. So, if you wish to update the text which appears in the statistics reports, or change the language, you can edit this file to do so.
7. First, you must modify each of these scripts slightly based on your own DSpace installation. Look for the following section in each of the stat-* files:
    - {{# Details used ## }}
8. In that section, you will likely need to modify any variable which specifies the full path of a directory or file location (these paths will all start with `/dspace/` by default). In particular keep an eye out for these variables:
    - `$dsrun`- the full path of the *dspace*/bin/dsrun script
9. * `$in_directory` - the full path of the input directory (for stat-report-* scripts)
    - `$out_directory` - the full path of the statistical reports output directory. This must correspond to the directory listed in the `report.dir` option in your dspace.cfg (see above)!
    - To keep things simple, you may wish to specify that both the `$in_directory` and `$out_directory` be the same location (e.g. *dspace*/reports/).
    - You may also wish to modify the `$start_year` and `$start_month` variables in `stat-initial` and `stat-report-initial`. These should correspond to the year and month which you wish to start tracking statistics from.
10. Make sure you updated every one of the stat-* scripts! There's six of them total: `stat-general`, `stat-initial`, `stat-monthly`, `stat-report-general`, `stat-report-initial`, `stat-report-monthly`
11. You will first need to run the initialization script to gather data about all of the past months (back to the month and year specified in `$start_month` and `$start_year`). This script only needs to be run once, though you can rerun it if you needed to (in which case it would overwrite its past results). Its output is a set of historic data files which can be used by stat-report-initial to generate historic HTML reports.
12. *dspace*`/bin/stat-initial` (Note: For Windows, you may need to run `perl` *dspace*`/bin/stat-initial`)
13. Next, generate a series of monthly reports for all of this past data by running the corresponding stat-report-* script. Again, this script only needs to be run once, though you can rerun it as needed. Its output is historic monthly HTML reports.
14. *dspace*`/bin/stat-report-initial` (Note: For Windows, `perl` *dspace*`/bin/stat-report-initial`)
15. The other four scripts should be run at least on a monthly basis, though you may even want to schedule them to run on a nightly basis in order to keep your statistics up-to-date at all times. These scripts will generate a report based on the activities this current month (stat-monthly and stat-report-monthly) and a general report aggregating all activities in the history of the repository (stat-general and stat-report-general). As in the examples above, the stat-report-* scripts should always be scheduled to run after their corresponding stat-* script.
16. For Linux or Mac OSX, you can schedule the scripts to run by adding a `cron` entry similar to the following to the crontab for the user who installed DSpace:

```
30 0 * * * [dspace]/bin/stat-monthly
```

```
35 0 * * * [dspace]/bin/stat-general
```

```
30 1 * * * [dspace]/bin/stat-report-monthly
```

```
35 1 * * * [dspace]/bin/stat-report-general
```

17. (The above cron entry would schedule these scripts to run nightly. stat-monthly and stat-general would be scheduled for 12:30am and 12:35am respectively. While their corresponding report scripts would run at 1:30am and 1:35am respectively. Note: You would need to change dspace to the full path of your DSpace installation directory.)
18. For Windows, you should use Windows Scheduled Tasks to schedule those same commands at an appropriate time of day. Just remember to schedule stat-monthly and stat-general to run before their corresponding report scripts (similar to the Linux/Mac OSX example above). In addition, you may need to call Perl explicitly (e.g. perl stat-monthly)
19. After running all statistics scripts, your DSpace site's statistical reports will be available immediately at `http://web-address-to-my-dspace/statistics`