

# Статистика в DSpace 1.8.x

В этой статье рассмотрим подключение и настройку статистики для DSpace. Поскольку по умолчанию система имеет два пользовательских интерфейса (jsrui и xmlui), постараюсь рассказать как можно ввести статистику для каждого из них.

Поскольку DSpace – это не только средство хранения различных документов, но и платформа для предоставления доступа к документам и может выполнять функции электронной библиотеки, поэтому сбор статистики есть важным, если необходимым условием функционирования и существования системы.

Средства просмотра статистики в DSpace доступны из меню администратора, а также отдельно для пользователя на уровне документов, разделов и коллекций.

Система автоматически собирает информацию о содержании и использовании вашего архива из внутренних файлов отчета

*Отчет о статистике включает следующую информацию:*

- о Настраиваемый общий обзор о деятельности в архиве (по умолчанию)
  - количество добавленных документов в архив
  - количество просмотров прикрепленных документов
  - количество просмотров страницы документа
  - количество просмотров страницы коллекции
  - количество просмотров страницы раздела
  - количество входов зарегистрированных пользователей
  - количество поисковых запросов
  - количество непринятых лицензий использования
  - количество OAI-запросов
- о Статистическое распределение документов в архиве по типу (Archive Information)
- о Перечень наиболее просматриваемых документов (Items Viewed)
- о Полный перечень выполненных действий (All Actions Performed)
- о Перечень пользователей и количество их входа в архив (User Logins)
- о Наиболее популярные поисковые запросы (Words Searched)
- о Отчет об ошибках (Log Level Information)

Результаты статистического анализа могут быть представлены ежемесячно и в общем отчете, и доступны через пользовательский интерфейс. Отчеты могут также быть доступны каждому или доступны только администратору архива.

Также статистику использования можно посмотреть на отдельных страницах документа, раздела и коллекции.

Здесь можно получить следующую информацию:

- общее количество просмотров (за все время) (Total Visits)
- количество просмотров за месяц Total Visits per Month

- количество загрузок прикрепленного файла (за все время)\* (File Downloads)
- топ просмотров по странам (за все время) (Top Country Views)
- топ просмотров по городам (за все время) (Top City Views)
- Информация об общем количестве загрузок доступна только для страниц документов. Следует обратить внимание, что информация о загрузках доступна по каждому отдельному файлу. Система способна обработать и сохранить информацию о загрузке файла даже в том случае, если скачивание файла проводилось по внешней ссылке с внешних ресурсов.

Перед тем как рассказать о настройке статистики в DSpace, необходимо сперва рассказать как пересобирать систему после изменения конфигурационных файлов или файлов исходного кода. Все ниже рекомендации касаются исключительно для ОС семейства Windows.

*Обозначение директорий, которые будут использоваться далее:*

[dspace] – папка установки DSpace;  
 [dspace-source]/dspace/ – папка, где размещаются исходные файлы пакета DSpace;  
 [dspace-source]/dspace/target/dspace-[version]-build.dir/ – папка, где будут размещены новые исходные файлы;  
 [Tomcat]/webapps/ – папка web-интерфейсов.

В зависимости от выполняемых действий можно выделить 3 процесса:

**Быстрый Рестарт** (просто перезапустить web-сервер после изменения файлов конфигурации):

- Остановить web-сервер (использовать команду остановки из монитора управления сервером)
- Запустить web-сервер (использовать команду запуска из монитора управления сервером)

**Быстрая сборка** (применяется, как правило, после внесения изменения в исходный код jsrui и xmlui интерфейсов):

1. Зайти в систему под администратором.
2. Запустить командную строку (cmd) и выполнить команду "cd [dspace-source]\dspace". Например, cd c:\elib\dspace.
3. Выполнить команду "mvn package" (рекомпиляция исходного кода DSpace и подготовка инсталляционного пакета).
4. Выполнить команду "cd [dspace-source]\dspace\target\dspace-[version]-build.dir". Перемещает пользователя в каталог с подготовленными файлами исходного кода для сборки. Например, cd c:\elib\dspace\target\dspace-1.8.0-build.dir.

5. Выполнить команду "ant update". Обновление DSpace на основании нового контента. Если не нужно переустанавливать jar-файлы, достаточно обновить файлы web-интерфейсов путем выполнения команды "update\_webapps".
6. Копировать папки из каталога webapps в одноименную папку сервера Apache Tomcat.
7. Тестировать изменения в DSpace.

**Полное обновление/Пересборка** (полностью обновляет DSpace в целом)

1. Зайти в систему под администратором.
2. Запустить командную строку (cmd) и выполнить команду "cd [dspace-source]\dspace". Например, cd c:\elib\dspace.
3. Выполнить команду "mvn clean package". Удаляет все файлы скомпилированного кода и подготавливает новый пакет файлов для установки DSpace.
4. Выполнить команду "cd [dspace-source]\dspace\target\dspace-[version]-build.dir". Перемещает пользователя в каталог с подготовленными файлами исходного кода для сборки. Например, cd c:\elib\dspace\target\dspace-1.8.0-build.dir.
5. Выполнить команду "ant update". Обновление DSpace на основании нового контента. Если не нужно переустанавливать jar-файлы, достаточно обновить файлы web-интерфейсов путем выполнения команды "update\_webapps".
6. Копировать папки из каталога webapps в одноименную папку сервера Apache Tomcat. Перезапустить сервер Apache Tomcat.
7. Тестировать изменения в DSpace.

### *Настройка статистика для интерфейса JSPUI*

Статистика в DSpace формируется в процессе анализа лог-файлов, которые создаются в процессе выполнения скриптов (команд), начинающихся с "dspace stat-".

Но сперва необходимо настроить конфигурационный файл dspace.cfg, в котором необходимо определиться можно ли предоставлять статистическую информацию всем пользователям без исключения или только администратору архива (report.public), а также указать каталог, где будут располагаться файлы отчета в формате html (параметр report.dir). После изменений не забудьте перезапустить Apache Tomcat.

report.public = false/true

Параметр может принимать только одно значение true или false, true - статистическую информацию может просмотреть каждый, false - доступна только администратору архива.

report.dir = \${dspace.dir}/reports/

Путь к папке с файлами отчета.  
`dspace.dir = /dspace`

Путь установки DSpace.

Здесь необходимо дать некоторое пояснение. Указывание значений параметров можно указывать и в относительном виде. Например, `${dspace.dir}/reports/` = `c:\dspace\reports`, если `dspace.dir = c:\dspace`. Применяется именно правый слеш, поскольку пути необходимо указывать в linux-формате. Рекомендуются файлы отчета размещать в отдельной папке и проводить систематическое резервное копирование.

Далее необходимо перейти к настройке собственного конфигурационного файла статистики и изменить:

- `start.year` - начальный год сбора статистики (например, 2012)
- `start.month` - начальный месяц сбора статистики (например, 2, 1 = Январь, 12 =Декабрь)
- `general.summary` - перечень необходимых параметров общей статистики
  - `exclude.word` - перечень слов, которые необходимо исключить из отчета поисковых запросов (союзы, предлоги), т. н. стоп-слова для фильтрации поисковых запросов
  - `exclude.type` - перечень полей поиска из которых необходимо делать исключения слов
  - `exclude.character` - перечень символов, которые необходимо исключить, применяются для выделения терминов словаря
  - `item.type` - перечень типов документов, по которым необходимо подсчитывать количество документов в архиве (например, монография, автореферат, диссертация)
  - `item.floor` - минимальное число просмотров документа, чтобы он мог войти в отчет
  - `search.floor` - минимальное число одинаково выполненных поисковых запросов, чтобы запрос вошел в отчет
  - `item.lookup` - определяет максимальное количество элементов списка Автор/Заглавие в статистике (все остальные просматриваемые перечислены по URL)
  - `user.email` - признак раскрытия электронной почты пользователя и может иметь три значения `on` - показывает адрес электронной почты пользователя, `off` - не показывает адрес электронной почты, `alias` - показывает идентификатор пользователя из БД, отвечает за статистику входа пользователей (в целях обеспечения конфиденциальности по умолчанию `alias`).

Следует упомянуть о том, что в состав DSpace входит файл `dstat.map`, который позволяет отобразить действия из файла отчета (`dspace/log`) в человекочитаемую форму. Чтобы ввести новые показатели статистики или выполнить локализацию на другие языки, необходимо редактировать именно этот файл.

Последовательность запуска скриптов:

1. Сначала необходимо запустить скрипт инициализации (`dspace stat-initial`), чтобы собрать данные за последние месяцы (начиная с месяца, указанного в `dstat.cfg` через параметры

`$start.year` и `$start.month`). Этот скрипт потребуется запустить единожды, хотя можно его и перезапустить позже в любой момент, когда это потребуется (в этом случае все результаты будет перезаписаны). После запуска скрипта создаются файлы данных (\*.dat), на основе которых и будут формироваться html-файлы, которые есть по сути способом отображения статистики.

2. Далее необходимо запустить 2 скрипта сбора статистики:
  - за текущий месяц (`dspace stat-monthly`)
  - общая статистика (за все месяцы) (`dspace stat-general`)
3. Обязательным условием есть запуск 2 скриптов формирования отчета:
  - за текущий месяц (`dspace stat-report-monthly`)
  - общий отчет (`dspace stat-report-general`)

Если последовательность запуска скриптов для месячной статистики или общей не критична, то скрипты формирования отчета следует запускать только после запуска сбора отчета.

Чтобы статистика была актуальной и на следующий день, скрипты сбора и формирования отчета следует запускать ежедневно во время минимальной нагрузки сервера или посещаемости. Процесс запуска скриптов можно автоматизировать, применив "Планировщик заданий". Ниже я расскажу как это можно сделать.

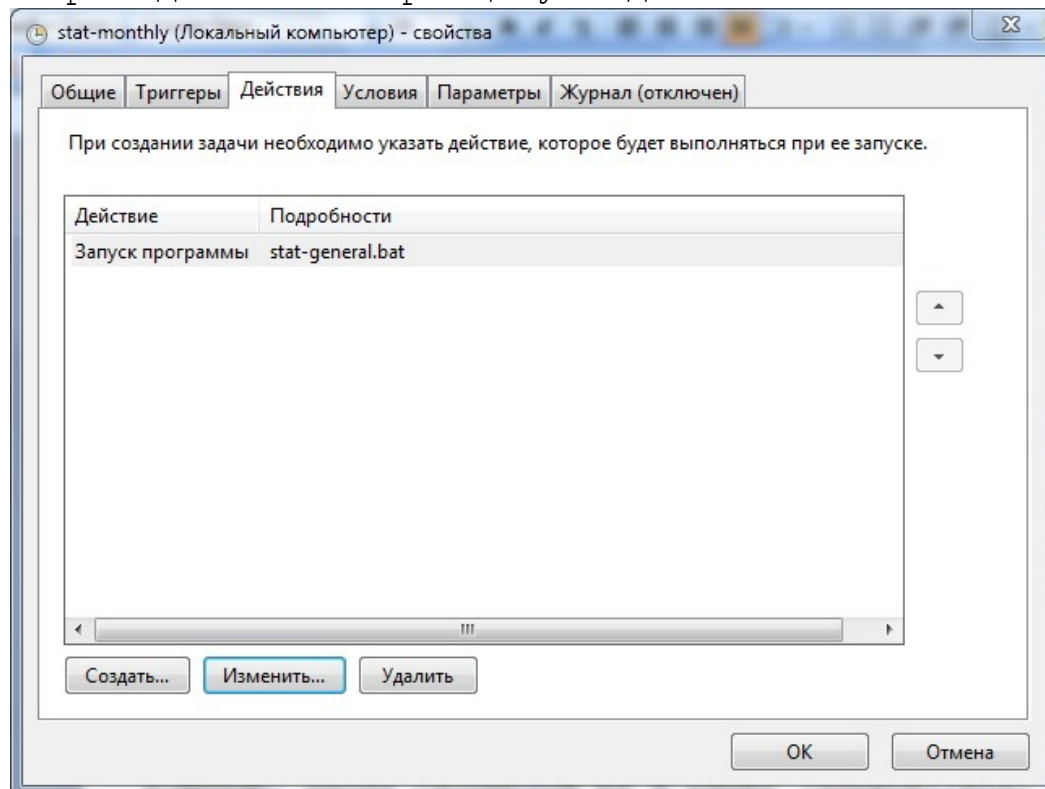
Допустим, что скрипты сбора статистики необходимо запускать в 00:10, а скрипты формирования отчета – в 00:15 каждый день.

Предварительно создадим 4 bat-файла с одной строкой:

- 1) `dspace stat-general`
- 2) `dspace stat-monthly`
- 3) `dspace stat-report-general`
- 4) `dspace stat-report-monthly`

Чтобы не запутаться имена пакетных файлов будут такими как команды. Далее скопируем их в папку `[dspace]/bin`.

Теперь переходим к Планировщику заданий Windows.



Здесь следует обратить внимание на вкладки Триггер (назначается время и день выполнения задачи) и Действия (указывается действие, программа и рабочая папка – обязательно).

Чтобы можно было просмотреть статистические данные по каждому документу, коллекции или разделу (количество посещений/просмотров всего и за месяц, загрузок, региональное посещение (по городу, по стране)), следует настроить так называемую Solr-статистику. Для этого необходимо:

1. Копировать папку из каталога `[dspace]/webapps/` в каталог `[Tomcat]/webapps/`.
2. В файле `solr-statistics.cfg` из каталога `[dspace]/config/modules/` сделать правку <http://localhost:8080/solr/statistics>
3. Обновить файл `GeoLiteCity.dat` из каталога `[dspace]/` по ссылке <http://geolite.maxmind.com/download/geoip/database/GeoLiteCity.dat.gz>
4. Перезапустить Tomcat.

## *Настройка статистика для интерфейса XMLUI*

Настройка статистики для интерфейса XMLUI сводится к получению GA-кода вида UA-XXXXXXXX-X, который необходимо вписать в значение параметра `$xmlui.google.analytics.key`, предварительно раскомментируя строку. Чтобы изменения вошли в силу, не забудьте перезапустить Tomcat.

Таким образом, система электронного архивирования DSpace обладает мощными средствами ведения статистики и позволяет предоставить как общую информацию о деятельности архива, так и показать востребованность ресурса через показатели посещения и загрузок полнотекстовых документов. Встроенные средства статистики можно разделить на два вида: базовая статистика и статистика Solr. Статистика ведется, собирается, формируется в автоматическом режиме путем однократного запуска инициализации статистики и запуска специальных скриптов сбора и формирования статистических показателей ежедневно.