

Google Summer of Code 2009 Report Generation Tool

Title: Report Generation Tool for DSpace

Student: Ashly Markose, National University of Singapore

Mentor: Jayan C Kurian, RMIT International University, Vietnam
email : jayan.kurian@rmit.edu.vn

Abstract:

"Report Generation" in general brings added value to any Information Management System with no exception to Institutional Repositories. Taken from an academic perspective, one of the main advantages is to generate reports based on individual authors and contribution period (e.g. Esther, 01-01-2006 to 01-12-2008). In-addition, if publications (Journal e.g. JASIST (Wiley), IPM (Elsevier) or Conference e.g. JC DL, ECDL) can be segregated based on ranking it would add much value from management perspective. A summarized report in this form from various academic disciplines in Institutes of higher learning would definitely drive strong interest from the stakeholders of Institutional Repositories. Our plan is to achieve this report generation based on data extracted from the Dspace IR and present this grouped by various custom filter options such as author, contribution period, publication ranking, and summarized report for each academic discipline etc. The motivation to this proposal is based on feedback received from the academic community during a presentation that was meant for encouraging contribution into repositories. We believe that such a feature would definitely drive academic contribution to repositories keeping in mind its long-term benefits.

Project Plan:

1. Conduct a survey to receive feedback from DSpace users and administrators regarding report generation for DSpace

Development Progress:

- Survey Result & Analysis
 - Status : Completed
 - No of participants submitted : 64
 - Deliverables: Survey Report
 - Software used : SurveyMonkey
- Report Generation
 - Report on item title and authors
 - Report on titles filtered by author
 - Report on listing communities and collections in your dspace instance
 - Report on listing items in a given collection
 - Report on listing items based on a given date range
 - Report on item title and authors (Web Application)
 - Report on Items Withdrawn
 - Report on Items Internal ID and Handler
 - Report on Item's Metadata
 - Report on Item Type
 - Status : Completed
 - Deliverables: Source code with individual documentation
 - Environment: Tested on Windows Vista and XP

Project Deliverables:

- Survey Report

The survey was conducted to find out how DSpace serves the need of various organizations and features facilitating stakeholders and contributors interest. One of the features that would facilitate contribution was general report generation which came up as a feedback from the academic community. In-addition to this, for receiving feedback from potential DSpace users and administrators, we have designed this survey with 6 potential questions. The survey had 2 general questions and the remaining 4 questions were related with report generation features. 64 participants registered with the DSpace mailing list gave their response through the survey. The on-line survey was built using the open source tool - "SurveyMonkey.com".

Majority (56%) of respondents agreed that DSpace is used as a platform to support their organization's Academic needs. 50% of respondents agreed that they use DSpace for their organization's Research and Library requirements. Few used DSpace as a platform to archive cultural materials and subsequent digital preservation. Majority (84%) of respondents have not used reporting tools such as BIRT or Pentaho for generating reports. Few (2%) have used JasperReports/iReport, and DataVision for report generation. Some respondents (15%) use Google Analytics and others have designed in-house product for reporting statistical information. Few installations having report generation features deal with statistical information (e.g. Bitstream download statistics, page views etc) rather than general information (e.g. Author, Title, Handler information etc). All participants (100%) agreed that they are interested in using an open source reporting tool for DSpace.

Several suggestions were received for generating reports based on statistical information and general information. Relevant suggestions for statistical information include generating reports based on Numbers of views and downloads, Items in Community/Collection level, Full text items in Community /Collection levels, Item accumulation in a chosen time span, No of hits and deposits per day, No of documents in the Workflow for approval, No of items based on type (Journal articles, conference papers), Users by country, Access information per item and collection, and AWWStats features.

Relevant suggestions for general information include list of all Publications based on Date of Publication, Author, and Collection; grouping records based on Item type, COUNTER compliance, Cross-reference to multiple Collections, Item Metadata values, Items with empty Metadata values, Item Bitstream size, and Access Authorization. Few have suggested that the tool should be easy to implement and customize due to their lack of IT support. It was also indicated that the tool should have detailed documentation with the generated report being able to be exported to document formats such as Excel. The respondents were globally spread with 31% from North America followed by Europe with 29%, Asia with 20% and 5% from Africa.

- Report generation workflow

1. Create a JasperReport's report template either by hand-coding or using a graphical report designer
2. Compile the report template file into JasperReport's native binary template
3. Use the compiled template to generate the final report by providing it with required data
4. Export the report and get it displayed in required document format.

- The following reports are generated and supported with documentation for Windows Environment.

1. Report on item title and authors
2. Report on titles filtered by author
3. Report on listing communities and collections in your DSpace instance
4. Report on listing items in a given collection
5. Report on listing items based on a given date range
6. Report on item title and authors (Web Application)
7. Report on Items Withdrawn
8. Report on Items Internal ID and Handler
9. Report on Item's Metadata
10. Report on Item Type

Source Code with documentation available at <http://code.google.com/p/google-summer-of-code-2009-dspace/downloads/list>

- The above reports are yet to be tested on non-windows environment but should work well if the following changes are made to the Java source file.
- Change the windows path to suit your unix based environment in the following statements in Java source file.
 - `jR = JasperCompileManager.compileReport("F:/GSOC2009DSpaceReport/CollectionTitles.jrxml");`
 - `JasperExportManager.exportReportToPdfFile(jP,"F:/GSOC2009DSpaceReport/CollectionTitles.pdf");`
- Remove the following statement from the Java source file.
 - `Runtime.getRuntime().exec("rundll32 url.dll,FileProtocolHandler "+ "F:/GSOC2009DSpaceReport/CollectionTitles.pdf");`

Documentation:

Documentation for DSpace Report Generation Tool (i.e. Title Author Report) on Windows Platform

Software Requirements: DSpace 1.5.2, PostgreSQL 8.2, JDK version 1.6, JCreator LE v4.50 for Windows, and JasperReports 3.5.0

1. Create a folder called "GSOC2009DSpaceReport". Here we have the path of this folder as F:/ GSOC2009DSpaceReport. You may choose any folder of your choice, but make sure that you have made appropriate modifications in the Java source code as explained in the Notes section.
2. Download JasperReports 3.5.0 (i.e. jasperreports-3.5.0-project.zip) at <http://jasperforge.org/projects/jasperreports> into the folder "GSOC2009DSpaceReport". Extract the ZIP file into the same directory and this creates a folder named "jasperreports-3.5.0" automatically in the "GSOC2009DSpaceReport" folder.
3. Download jasperreports-3.5.0.jar file available at http://sourceforge.net/project/showfiles.php?group_id=36382&package_id=28579 into F:\GSOC2009DSpaceReport\jasperreports-3.5.0\lib. If this jar file is not used there would be a build error while compiling the project indicating that JasperReports engine is not available.
4. Copy the postgresql-8.1-408.jdbc3 jar file from dspace-1.5.2-release\dspace\target\dspace-1.5.2-build.dir\lib into F:\GSOC2009DSpaceReport\jasperreports-3.5.0\lib
5. Install Java SE 6 for Windows if your machine doesn't have one installed. <http://java.sun.com/javase/downloads/index.jsp>
6. Install the Java IDE, JCreator LE v4.50 for Windows <http://www.jcreator.com/download.htm>
7. Open the JCreator IDE
8. Select "File" - "New Project" and select "Basic Java Application". Click the "Next" button
9. In the project wizard give Name (i.e. project name) as DSpaceReport
10. Click "Finish" on the Project Wizard interface. This creates the "DSpaceReport" project.
11. On the JCreator IDE - Menu bar, Click "Project" - "Project Settings" and select "Required Libraries" tab.
12. Click the "New" tab and give Name as "DSpaceReportLibrary"
13. Click the "Add" button and select the "Add Archive" button. Browse to F:\GSOC2009DSpaceReport\jasperreports-3.5.0\lib. Using the CTRL + A computer keyboard combination select all jar files there and select "Open".
14. Click on the "OK" button.
15. Check the "DSpaceReportLibrary" and click "OK".
16. Download the given TitleAuthorReport.jrxml file into the folder "GSOC2009DSpaceReport".
17. Download the java file DSpaceReport.java into the folder "GSOC2009DSpaceReport". Open this downloaded java file and copy its java source code into the DSpaceReport.java file created using the JCreator IDE after completely clearing the default java code available in that file.
18. Download the given file Functions.sql into the folder "GSOC2009DSpaceReport".
19. Start PostgreSQL 8.2 and select "pgAdmin".
20. Connect to PostgreSQL using the "postgres" user account and password.
21. Select the DSpace database. Click on the SQL icon (Blue in color) and select "File" - "Open". Browse to the folder GSOC2009DSpaceReport and select the file "Functions.sql". Click Open.
22. Click on the "Execute Query" button to execute the functions. Exit the PostgreSQL interface using "File" - "Exit".
23. Build the java file using "Build" - "Build File".
24. Run the file using "Run" - "Run File".
25. The Report is generated and displayed in PDF file format.

Notes:

- Check the file name (i.e. TitleAuthorReport.jrxml) and path of JRXML file (i.e. "F:/GSOC2009DSpaceReport/TitleAuthorReport.jrxml") in the DSpaceReport.java source file incase you have modified the file name and path.
- If you have changed, modify the following statements in the Java source code file accordingly. Also note that forward slash (i.e. /) is used in indicating the file path.

```
jR = JasperCompileManager.compileReport("F:/GSOC2009DSpaceReport/TitleAuthorReport.jrxml");
JasperExportManager.exportReportToPdfFile(jP,"F:/GSOC2009DSpaceReport/TitleAuthorReport.pdf");
Runtime.getRuntime().exec("rundll32 url.dll,FileProtocolHandler "+ "F:/GSOC2009DSpaceReport/TitleAuthorReport.pdf");
```

- If you encounter error while compiling/running the java source code, make sure that you have the jasperreports-3.5.0.jar, postgresql-8.1-408.jdbc3 jar along with other jar files at F:\GSOC2009DSpaceReport\jasperreports-3.5.0\lib. This should also be available at "DSpaceReportLibrary" in "Project - Project settings" of the JCreator IDE and is checked.
- The source code is tested only on DSpace 1.5.2 using PostgreSQL 8.2, JDK version 1.6, JCreator LE v4.50 for Windows, JasperReports 3.5.0, PostgreSQL-8.1-408.jdbc3 jar, jasperreports-3.5.0.jar and iReport-nb-3.5.0. The JRXML file is developed using the iReport-nb-3.5.0 tool.

Future Work:

- Generating reports based on remote dspace instances
- Fully integrated web application

My University, School, and My Division:

[I'm here](#)

[National University of Singapore](#)

[School of Computing, NUS](#)

[Division of Information Systems](#)

References:

- Online survey using <http://www.surveymonkey.com/>
- Programming reference
 - JasperReports for Java Developers by David Heffelfinger
- JCreator - Java IDE, <http://www.jcreator.com/>
- NetBeans IDE 6.1, <http://java.sun.com/javase/downloads/netbeans.html>