DSpace 1.6 usage statistics: How does it work?

Ben Bosman, C.T.O. @MIRE, <ben@atmire.com>

INTRODUCTION

DSpace 1.6 has been extended with Apache Solr based statistics. This contribution to DSpace is the open-source version of @mire's commercial "Content and Usage Analysis" DSpace module. The DSpace 1.6 statistics offer logging of usage data including bitstream downloads, item display page visits, collection and community homepage visits, ...

TECHNICAL OVERVIEW

This presentation will dive into the implementation details of this new statistics solution by outlining the following topics:

- the storage of the statistics data;
- various potential usage scenarios of the different statistics;
- some of the opportunities for improving the statistics in future DSpace releases or customizing local requirements.

The statistics contribution present in DSpace 1.6 is generic by default and includes configuration options to determine whether the usage data is displayed in the user interface, and whether the statistics are available for all users or only for administrators. The default display only offers a limited view on the real data being stored. In order to extend this, it is important to understand how the statistics are being stored and what data is available.

The standard solution is configured by default to be lightweight and easily installable. If your DSpace installation has a large amount of visitors, this default configuration might not be ideal for your installation. To address this situation, a number of simple and more advanced solutions for performance optimization will be discussed.

USER INTERFACE ADDITIONS

The most common and easiest additions to the DSpace statistics are changes to the displayed usage data in the various pages of the DSpace repository. A

@mire

Technologielaan 9
B-3001 Heverlee
Belgium
☎ +32 2 888 2956

533 2nd St.
Encinitas, CA 92024
U.S.A.
☎ +1 858 368 4681

info@atmire.com
www.atmire.com

few typical examples of these are the display of repository wide statistics, a differentiation between file downloads and page views in the existing display, the ability to specify the timespan for which the statistics should be displayed, …. An example of such an addition will be displayed during the presentation, including all of the changes required to implement this for the DSpace XMLUI.

ADVANCED USE CASES

This presentation will also include some example use-cases for which the full potential of the statistics can be used to offer advanced solutions for situations requiring more than just user interface extensions. These are of course too elaborate to display all implications.

One example is an OAI crosswalk offering access to a view on the usage data which is important for e.g. NEEO harvesting (http://homepages.ulb.ac.be/~bpauwels/NEEO/WP5/WP5%20Technical%20guidelines%20-%20Annex3.pdf).

Another example would be the extension of the data being stored for each usage event by storing and displaying e.g. the referrer URL retrieved from the browser. This referrer can be used to link back to pages being frequently used for accessing a record, or for generating an overview of the domain names from which most visits originate. It can also provide a view on the typical internal navigation within the repository.

A third example would be the inclusion of usage metadata other than page views and file downloads. One can e.g. store an overview of performed searches within the repository to create an overview of the most popular searches or a tag cloud of keywords frequently used in combination with the searches currently executed (an example of such a functionality can be found at http://www.getcloudlet.com).

HIGH LOAD INSTALLATIONS

A final subject of this presentation handles a number of configuration improvements to address high load installations. In case your DSpace installation generates a large amount of statistics data, statistics query execution might become slower. A number of optimizations based on the auto-commit and the query warmup system offered by the Apache Solr server will be explained in more detail.

@mire

Technologielaan 9
B-3001 Heverlee
Belgium
☎ +32 2 888 2956

533 2nd St.
Encinitas, CA 92024
U.S.A.
☎ +1 858 368 4681

info@atmire.com
www.atmire.com