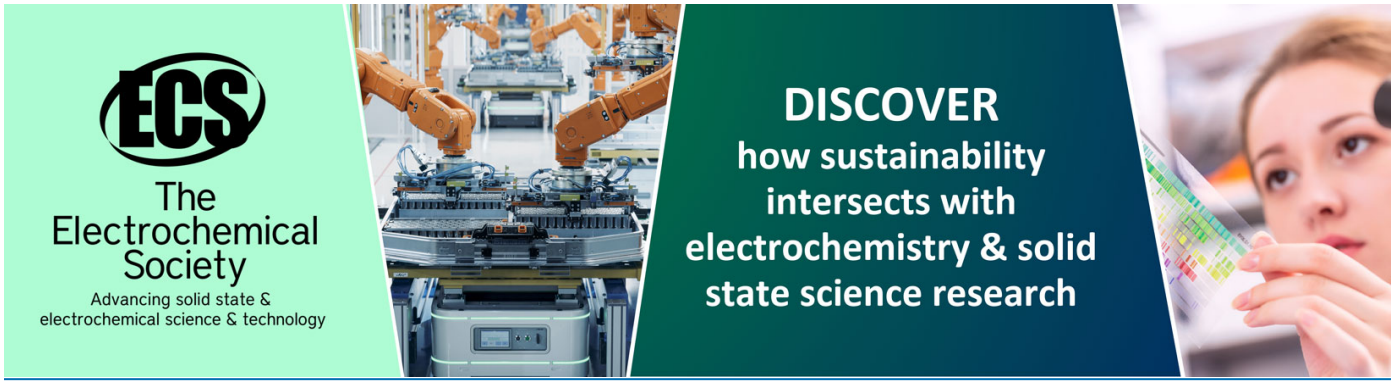
On providing the fault-tolerant operation of information systems based on open content management systems



IOP Conf. Series: Journal of Physics: Conf. Series **944** (2018) 012067 doi:10.1088/1742-6596/944/1/012067

On providing the fault-tolerant operation of information systems based on open content management systems

Sergey Kratov

Foundation of Algorithms and Programs SB RAS, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

Abstract. Modern information systems designed to service a wide range of users, regardless of their subject area, are increasingly based on Web technologies and are available to users via Internet. The article discusses the issues of providing the fault-tolerant operation of such information systems, based on free and open source content management systems. The toolkit available to administrators of similar systems is shown; the scenarios for using these tools are described. Options for organizing backups and restoring the operability of systems after failures are suggested. Application of the proposed methods and approaches allows providing continuous monitoring of the state of systems, timely response to the emergence of possible problems and their prompt solution.

1. Introduction

A growing number of information systems designed to serve a wide range of users are based on Web technologies nowadays. Users of such systems get access to them through Internet. One of the most optimal ways of developing such Web-oriented information systems in terms of time and money is to use content management systems or, in more complex cases, frameworks for designing CMS as their basis [1]. Using CMS or frameworks allows significantly reducing the time to develop basic (common for all) subsystems, such as subsystems for authorizing users, granting rights, structuring content, connecting external libraries, etc. Today there are a large number of free and open CMS/frameworks, additional modules and libraries, significantly expanding their functionality, which in turn allows significantly reducing the financial costs of development [2, 3].

At the same time, the use of open CMS is associated with the risk, since the source code of such systems is publicly available and, in case of errors detection in it, can be exploited by intruders [4-7]. That in turn can lead to their compromising and hacking [8-10] in the case of the general availability information systems based on such CMS.

In this article, the author would like to describe the process of supporting the fault-tolerant operation of an information system based on an open and free CMS, using the example of the Foundation of Algorithms and Programs of the SB RAS (http://fap.sbras.ru) [11]. The author would like to show the set of tools provided to the system administrator by the developers of its kernel and modules, through which it will be able to monitor its state and provide trouble-free and uninterrupted operation.

The Foundation's information system (figure 1) [12, 13] is based on the free CMS Drupal, which means it has all the advantages as well as potentially being exposed to all the problems described above.

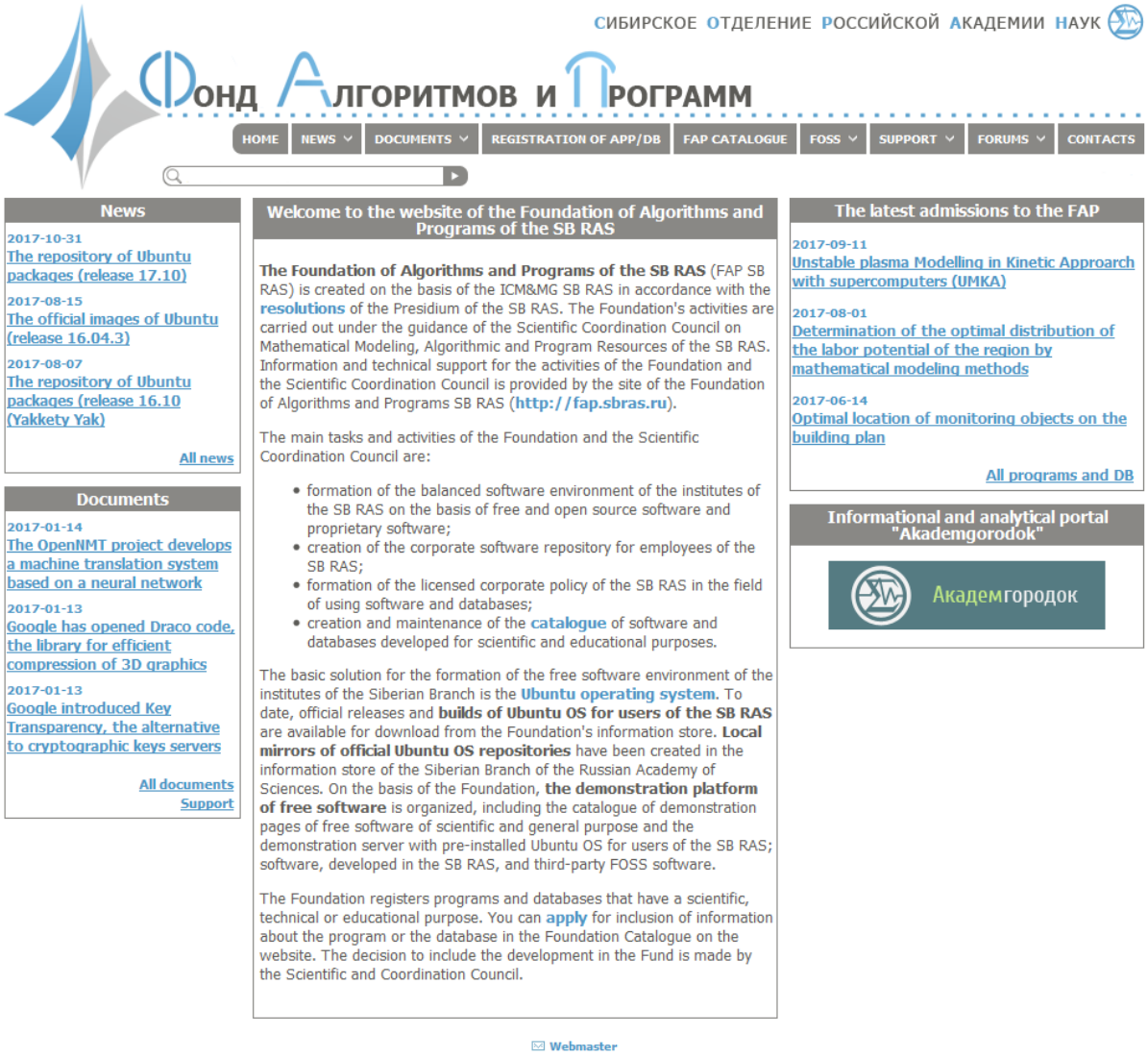


Figure 1. FAP SB RAS.

2. The general information panel

The basic administrative tool is the status report. It is the brief display of the main parameters of the information system state, critical for its overall functioning, and detected problems. For example, in figure 2 are displayed:

- the kernel version of the system and its update status. It can be seen that a kernel update of the system is available, non-critical in terms of security,
- the status of the file system storage, the protection of the main CMS configuration files,
- the time of the last run of scheduled tasks (cron),
- type and version of DBMS, its support for text encodings, system database state.

Figure 3 lists:

- PHP interpreter version, PHP modules, required to run the information system kernel, RAM limits for executing PHP scripts, etc.,
- availability of updates for modules, themes and translations of interfaces,
- availability and versions of system graphical (GD) and javascript libraries (jQuery),
- Web server version.

Drupal	
Access to update.php	Protected
CAPTCHA	Already 17031 blocked form submissions
CKEditor	
Colorbox plugin	
Configuration file	Protected
Cron maintenance tasks	Last run 17 min 11 sec ago You can run cron manually. To run cron from outside the site, go to
CTools CSS Cache	Exists
Database 4 byte UTF-8 support	Enabled 4 byte UTF-8 for mysql is enabled.
Database system	MySQL, MariaDB, or equivalent
Database system version	
Database updates	Up to date
Date API	System date settings The timezone has been set to Asia/Novosibirsk. The first day of the week has been set to Monday. The medium date format type has been set to Thu, 08/06/2017 - 16:04. You may find it helpful to add new format types like Date, Time, Month, or Year, with appropriate formats, at Date and time settings.
⚠ Drupal core update status	Out of date (version available) There are updates available for your version of Drupal. To ensure the proper functioning of your site, you should update as soon as possible. See the available updates page for more information and to install your missing updates.
File system	Writable (*public* download method)
GD library PNG support	
GD library rotate and desaturate effects	

Figure 2. Part 1 of the status report.

Using this information panel, the system administrator can evaluate the overall state of the system at any given time. In case of any problems, he can go down to the level below, going to more specialized reports.

jQuery Update	jQuery (configure) and jQuery UI
Module and theme update status	Up to date
Node Access Permissions	Disabled
If the site is experiencing problems with permissions to content, you may have to rebuild the permissions cache. Rebuilding will remove all privileges to content and replace them with permissions based on the current modules and settings. Rebuilding may take some time if there is a lot of content or complex permission settings. After rebuilding has completed, content will automatically use the new permissions. Rebuild permissions	
PHP	(more information)
PHP extensions	Enabled
PHP memory limit	128M
PHP register globals	Disabled
⚠ Translation update status	There are available updates
There are new or updated translations available for currently installed modules and themes. To check for updates, you can visit the translation update page.	
Unicode library	PHP Mbstring Extension
Update notifications	Enabled
Upload progress	Enabled (PECL uploadprogress)
Web server	Apache/ (Ubuntu)
XML sitemap	Last attempted generation on Thu, 08/06/2017 – 14:45 (1 hour 18 min ago).
XML sitemap cache directory	Writable

Figure 3. Part 2 of the status report.

3. The «available updates» report

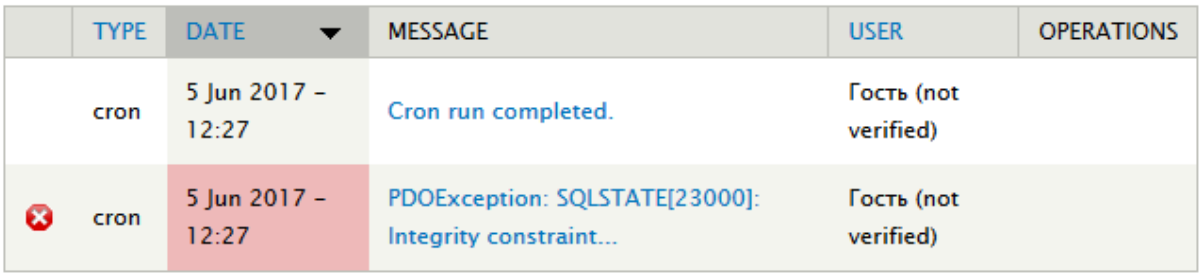
CMS Drupal consists of the kernel that provides its basic functions, and has an API for connecting additional modules that extend its functionality. Modules are developed by third-party developers and are publicly available to all web developers (https://www.drupal.org/project/project_module). The authors of the modules release regular updates, that expanding their functionality, adding support for new versions of browsers, the implementation of new standards, and correcting errors found in the modules. Due to the fact that after correcting the errors, information about them becomes available to the public after a while, it is recommended to immediately update the modules used in the information system after the release of their new versions. There is the tool to track updates in the system, which monitors the release of new versions of modules and notifies the system administrator about it (both through the interface of the system itself and via e-mail). Tracking occurs at the start of cron, which allows system administrators to set any frequency of it. Administrator can also check for updates at

any time manually. The list of modules with their update status and the time of the last update check is shown in figure 4.

Figure 4. Available updates.

4. The recent log messages

The information system keeps a living log of events. Messages that fall into the log are saved in it, followed by rotation for the period selected by the system administrator. A user with the appropriate rights can view the system event log at any time. The log is the list containing information messages about usage data, performance data, errors, warnings and operational information. It is possible to filter messages in the log according to their source (kernel or module) and the significance level (emergency, alert, critical, error, warning, notice, info, debug). In figure 5 the part of the log is shown in which there are displayed as the ordinary message about the cron start and the critical error of one of the system components functioning.



	TYPE	DATE	MESSAGE	USER	OPERATIONS
	cron	5 Jun 2017 - 12:27	Cron run completed.	Гость (not verified)	
⊗	cron	5 Jun 2017 - 12:27	PDOException: SQLSTATE[23000]: Integrity constraint...	Гость (not verified)	

Figure 5. The recent log messages.

Figure 6 shows detailed information about this critical error, including the maximum possible description, the address of the page on which it occurred, the user, the time moment, etc.

Regular viewing of the system log is critically important, as it is often the only way to early notice and fix a system failure.

5. The external libraries

To extend the functions of the information system in addition to installed modules, web developers can connect third-party libraries that will be shared by different modules. Figure 7 lists the libraries used in the system with their versions. Versions of libraries are also recommended to keep up to date - first, old versions can contain security-critical errors; secondly, support for new versions of browsers, standards, etc. are implemented in new versions of libraries.

TYPE	cron
DATE	Monday, 5 June, 2017 - 15:28
USER	Гость (not verified)
LOCATION	[redacted]
REFERRER	
MESSAGE	*PDOException*: SQLSTATE[23000]: Integrity constraint violation: 1062 Duplicate entry 'ynews-ru' for key 'PRIMARY': INSERT INTO {l10n_update_file} (project, language, filename, version, status, last_checked) VALUES (:db_insert_placeholder_0, :db_insert_placeholder_1, :db_insert_placeholder_2, :db_insert_placeholder_3, :db_insert_placeholder_4, :db_insert_placeholder_5); Array ([:db_insert_placeholder_0] => ynews [:db_insert_placeholder_1] => ru [:db_insert_placeholder_2] => ynews-7.x-1.x-dev.ru.po [:db_insert_placeholder_3] => 7.x-1.x-dev [:db_insert_placeholder_4] => 1 [:db_insert_placeholder_5] => 1496651288) in *drupal_write_record()* (line *7383* of [redacted]).
SEVERITY	error
HOSTNAME	141.8.142.16
OPERATIONS	

Figure 6. Details about an error.

NAME	STATUS	INSTALLED VERSION	PROVIDER	LINKS
Colorbox plugin	OK	[redacted]	*Colorbox* module	Homepage | Download
MailChimp API	OK	[redacted]	*MailChimp* module	Homepage | Download

Figure 7. The libraries list.

Figure 8 provides detailed information about one of the libraries connected to the system, indicating its current version, the installation path in the system, and the download page, where it is possible to check the version's relevance.

6. Top 'access denied' and 'page not found' errors

Pages with 'access denied' (figure 9) and 'page not found' (figure 10) errors also carry important information about the health of the system. They allow administrators to track erroneous links on the system pages, as well as attempts to hack it using automated scanners for common vulnerabilities. That in turn allows administrators to quickly track them and take actions to protect of the system.

GENERAL INFORMATION	
Name	Colorbox plugin
Machine name	colorbox
Vendor URL	http://www.jacklmoore.com/colorbox
Download URL	https://github.com/jackmoore/colorbox/archive/1.x.zip
Provider	*Colorbox* module
Library path	sites/all/libraries/colorbox
Version	[illegible]
Variants	minified, source

Figure 8. Status report for the library.

COUNT	MESSAGE
2	admin/reports/event/406724
2	admin/config.php
1	user/1645

Figure 9. Top 'access denied' errors.

COUNT	MESSAGE
6	vremenno-aktualnye-rubriki/80-let-sibniirs
3	wp-login.php
2	content/v-novosibirske-do-sih-por-doedayut-sovetskuyu-energetiku
1	sites/default/files/js/js_FvM1alErE930ZiQ6dh2VnUkPAFS739apuHmc11fKTYw.js
1	administrator/index.php

Figure 10. Top 'page not found' errors.

7. Top pages

The main purpose of this report (shown in figure 11) for the administrator can be to track the average time of generation of individual pages, as well as the total time spent by the server on the generation of the particular page. Pages with an abnormally long generation time potentially contain some errors in the page code, database queries, connected external scripts, etc. By correcting such errors or by optimizing queries it is possible to significantly reduce the load of the information system server(s).

Pages with an abnormally large number of hits can also indicate the presence of vulnerabilities used by bots.

HITS	PAGE	AVERAGE PAGE GENERATION TIME	TOTAL PAGE GENERATION TIME
15592	News home	*99* ms	25 min 46 sec
2860	The list of employees content/spisok-sotrudnikov	*153* ms	7 min 18 sec
1865	404 node/4385	*76* ms	2 min 21 sec
939	Web-mail node/2720	*37* ms	34 sec
769	The news about conferences conf_archive	*71* ms	54 sec
571	All taxonomy/term/all/all	*141* ms	1 min 20 sec

Figure 11. Top pages in the past 2 weeks.

8. Top visitors

List of registered visitors to the information system, or (for anonymous visitors) IP-addresses that viewed the largest number of pages during the reporting period is shown in figure 12. It can indicate to both a search bot indexing the system's content, and the presence of an open vulnerability that using by bots. In the case of making heavy load on the system server, it is possible to block an individual IP address or an entire subnet, which can be useful in organizing the reflection of attacks aimed at denial of system service.

HITS	VISITOR	TOTAL PAGE GENERATION TIME	OPERATIONS
21681	141.8.132.81	27 min 4 sec	block IP address
3706	88.198.158.233	5 min 35 sec	block IP address
3391	34.253.95.150	5 min 9 sec	block IP address

Figure 12. Top visitors in the past 4 weeks.

9. Backup

Despite all the measures taken to organize a fault-tolerant functioning of the information system, the system administrator also needs to back up both the program code of the system and the user information contained therein. One of the simplest methods is writing shell scripts that dump the database and create an archive of the file system storage. Under the condition that such scripts are launched by cron, this will ensure the creation of regular snapshots of the system state with the possibility of their rapid deployment in case of failures. Examples of such scripts for creating a database dump and an archive of file system storage can be seen in figure 13 and figure 14.

```
#!/bin/sh

USER=username
PASSWD=password
DB_PROD=databasename

if [ -z "$1" ]; then
  echo "USAGE: $0 <file_name>";
  exit 0;
fi

echo "Dumping $DB_PROD to $1"
mysqldump -u $USER -p$PASSWD $DB_PROD > $1
```

Figure 13. Shell script for dumping system database.

```
tar -czf `date +%Y%m%d`.tar.gz /path/to/files/directory

gzip DB-dumpname.sql `date +%Y%m%d`.sql.gz
```

Figure 14. Shell scripts for archiving DB-dump and file directory.

10. Conclusion

The article raised the problem of supporting the functioning of information systems developed on the basis of publicly available free and open CMS. A number of tools provided to the administrator of the system kernel were described, variants and scenarios for their use are shown. Also, options were offered for organizing a regular backup of the database and file storage of the system. The system administrator can use the proposed tools to keep the system up-to-date, to detect and correct the error in time, and to ensure its quickest recovery in case of failure.

References

[1] Drucker J and Svensson P B O 2016 The why and how of Middleware *Digital Humanities Quarterly* iss. 2 **10**

[2] Atayero A A, Chijioke-Keme and Ogunjobi B 2014 FOSS implementation of an educational virtual office suite *EDULEARN14: 6th Int. Conf. on Education and New Learning Technologies* 5018–27

[3] Tramullas J 2013 Content management with Drupal: a review of modules specific to libraries, archives and museums *Profesional De La Informacion* iss. 5 **22** 425–31

[4] Mainka C, Mladenov V and Schwenk J 2016 Do not trust me: using malicious IdPs for analyzing and attacking single sign-on *1st IEEE European Symp. on Security and Privacy* 321–6

[5] Zhang Y, Lo D, Xia X et al. 2015 Combining software metrics and text features for vulnerable file prediction *2015 20th Int. Conf. on Engineering of Complex Computer Systems (ICECCS)* 40–9

[6] Eshkevari L, Dos Santos F, Cordy J R et al. Are PHP applications ready for hack? *2015 22nd Int. Conf. on Software Analysis, Evolution, and Reengineering (SANER)* 63–72

[7] Javed A and Schwenk J 2014 Towards elimination of cross-site scripting on mobile versions of Web applications *Information Security Applications, WISA 2013* **8267** 103–23

[8] Sanchez A B, Segura S, Parejo J A et al. 2017 Variability testing in the wild: the Drupal case study *Software And Systems Modeling* iss. 1 **16** 173–94

[9] Chimos K and Loumpardas P 2014 Identifying security issues in a higher education's institute CMS lab site *EDULEARN14: 6th Int. Conf. on Education and New Learning Technologies Proc.* 3349–54

[10] Patel S K, Rathod V R and Prajapati J B 2013 Comparative analysis of Web security in open source content management system *2013 Int. Conf. on Intelligent Systems and Signal Processing (ISSP)* 344–9
[11] Sokolova O D and Kratov S V 2016 Information systems for popularization of scientific and knowledge-based software *2016 13th Int. Scientific-Technical Conf. on Actual Problems of Electronic Instrument Engineering (APEIE) – 39281 Proc.* **1** 519–22
[12] Kratov S and Bukhtiarov I 2013 The technological platform for software development in the SB RAS *8th Int. Forum on Strategic Technology 2013, IFOST 2013 – Proc.* 353–5
[13] Zybarev Y and Kratov S 2012 The information support system of corporate software resources development *Proc. - 2012 7th Int. Forum on Strategic Technology, IFOST 2012* 683–5