

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра дискретной математики и алгоритмики

Аннотация к магистерской диссертации

**РАЗРАБОТКА РАСПРЕДЕЛЁННОГО ОТКАЗОУСТОЙЧИВОГО
ХРАНИЛИЩА ФАЙЛОВ ДЛЯ СИСТЕМЫ АВТОМАТИЧЕСКОГО
ТЕСТИРОВАНИЯ**

Соболь Сергей Александрович

Научный руководитель — кандидат. физико-математических наук,
доцент Е. П. Соболевская

2015

Реферат

Магистерская диссертация, 60 страниц, 22 рисунка, 11 источников.

СИСТЕМА АВТОМАТИЧЕСКОГО ТЕСТИРОВАНИЯ, БАЗА ДАННЫХ, ХРАНИЛИЩЕ, MYSQL, MONGODB, GRIDFS, РАСПРЕДЕЛЁННОСТЬ, ОТКАЗОУСТОЙЧИВОСТЬ, ХЕШ-ФУНКЦИЯ, MD5, SHA-1.

Объект исследования — системы, используемые для хранения данных в системах автоматического тестирования.

Цель работы — внедрить в действующую систему автоматического тестирования новое хранилище файлов, реализовать на его основе функциональность по сохранению и отображению результатов выполнения решений на тестах, настроить функционирование компонентов на производственных серверах.

Результат работы — была создана и введена в эксплуатацию система хранения, реализованы необходимые функции в модулях системы тестирования.

Дано общее описание проблем, связанных с хранением данных системы автоматического тестирования. Получены оценки объёма информации и распределения размеров файлов на основе опыта эксплуатации реальной системы. Особое внимание уделено вопросу хранения повторяющихся файлов: при тестировании разных решений одной задачи нередко получаются совпадающие результаты (одинаково правильные или совпадающие ошибочные). Исследован метод поиска дубликатов файлов на основе хеш-функций, оценена вероятность коллизии. Проанализирована эффективность метода в применении к данным малого размера и предложена его оптимизация. Затем изучены подходы к хранению данных: хранение в файловой системе в виде отдельных файлов, представление данных бинарными блоками в реляционной базе данных, хранение в специализированном NoSQL-хранилище. Описана современная система управления базами данных MongoDB и модель данных GridFS. Осуществлено поэтапное внедрение хранилища файлов на основе GridFS в действующую систему автоматического тестирования, решены проблемы отображения данных в веб-интерфейсе. Изучен вопрос обеспечения распределённости и отказоустойчивости системы, показано, как настроить репликацию данных на нескольких серверах.

Работа имеет большое практическое значение. Система автоматического тестирования применяется как в рамках учебного процесса, так и для проведения соревнований по программированию в Белорусском государственном университете. Сохранение результатов тестирования в хранилище файлов с возможностью обращения к ним через интерфейс является полезной функцией.

Abstract

Master thesis, 60 pages, 22 figures, 11 references.

AUTOMATIC TESTING SYSTEM, DATABASE, STORAGE, MYSQL, MONGODB, GRIDFS, DISTRIBUTABILITY, FAULT TOLERANCE, HASH FUNCTION, MD5, SHA-1.

Object of research is systems used for storing data of automatic testing systems.

Purpose is to integrate a new file storage into the working automatic testing system, to implement the functionality of saving and viewing results of solution execution on tests, to setup component operation on production servers.

Result of current work is storage system that was created and put into operation. Necessary functions were implemented in testing system modules.

Problems of saving data of automatic testing systems were described in general. Information volume and file size distribution were estimated using the operating experience of the real system. Special attention is paid to the problem of duplicate files storage: when different solutions of one problem are tested, similar results are often obtained (equally correct or erroneous). Method of finding duplicate files based on hash functions was introduced, probability of collision was evaluated. Effectiveness of the method applied to small size data was analyzed and optimization was proposed. Then several storage approaches were studied: storing data in a file system as individual files, binary data representation as blocks in a relational database, storing in a specialized NoSQL storage. Modern database management system MongoDB and GridFS data model are described. GridFS-based storage engine was implemented step-by-step in the current automatic testing system, the problem of displaying data in a web interface was solved. Question of making distributable and fault tolerant system was studied, configuration of data replication on multiple servers was presented.

The work has great practical importance. Automatic testing system is used in the educational process and during programming contests at Belarusian State University. Saving test results to file storage with an option of viewing them via the interface is a useful feature.