

УДК 026.6

МОДЕЛИРОВАНИЕ ПРОЦЕССА ЗАГРУЗКИ ДАННЫХ В АВТОМАТИЗИРОВАННЫЕ БИБЛИОТЕЧНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ С ОТКРЫТЫМ КОДОМ

¹Стукалов В.А., ²Савотченко С.Е.

¹ГОУ ВПО «Белгородский государственный институт искусств и культуры»,
Белгород, e-mail: svadglo@gmail.com;

²ОГАОУ ДПО «Белгородский институт развития образования»,
Белгород, e-mail: savotchenko@hotmail.ru

На сегодняшний день многие учреждения используют различные информационные системы для управления и контроля своих библиотек. При этом наибольшую распространенность получили автоматизированные библиотечные информационные системы (АБИС) с закрытым исходным кодом. Это препятствует изменению функциональности этих систем, добавлению новых возможностей. В то же время существуют различные бесплатные АБИС с открытым кодом. В работе представлены результаты компьютерных экспериментов по измерению зависимости времени выполнения импорта данных в базу данных автоматизированной библиотечной информационной системы SLiMS от частоты используемого процессора и объема данных. Предложена математическая модель скорости загрузки данных. Установлено, что при сравнительно малых объемах записей время загрузки линейно зависит от количества записей (строк), загрузка происходит с постоянной скоростью при фиксированных параметрах вычислительной системы. Показано, что можно совершить переход на бесплатные библиотечные системы с открытым кодом даже на относительно слабом аппаратном обеспечении.

Ключевые слова: информационные системы, автоматизированные библиотечные информационные системы, математические модели, производительность информационных систем, базы данных, конвертация данных

SIMULATION OF DATA LOADING PROCESS INTO OPEN SOURCE AUTOMATED LIBRARY INFORMATION SYSTEM

¹Stukalov V.A., ²Savotchenko S.E.

¹Belgorod State Institute of Arts and Culture, Belgorod, e-mail: svadglo@gmail.com;

²Belgorod Institute of Education Development, Belgorod, e-mail: savotchenko@hotmail.ru

Today, many institutions use a variety of information systems for the management and control of their libraries. library information systems (lis) WITH closed sources are mostly spread. Closed source code – a significant obstacle to change the functionality of these systems, adding new features. The paper presents results of computer experiments of measuring dependence of importing data into the database of LiS SLiMS on the frequency of the processor used and the amount of data. We suggest a mathematical model of the speed of loading data. It was found, that whis relatively small amounts of entries, the load time depends linearly on the number of records (rows), loading occurs at a constant rate with fixed parameters of the computing system. It is shown that it is possible to make the transition to the free library system open source even at relatively low hardware.

Keywords: information system, library information systems, mathematical models, performance of information systems, databases, data conversion

Автоматизированная библиотечно-информационная система [4] SLiMS (Senayan Library Management System) представляет собой программное обеспечение для управления библиотекой с открытым исходным кодом.

С технической точки зрения, информационная система Senayan представляет из себя Web-приложение, доступ к которому осуществляется через браузер. Для развертывания АБИС потребуются Web-сервер, СУБД MySQL и интерпретатор языка PHP.

Сама система в свою очередь построена также на открытых программных инструментах и языках, что существенно упрощает изменение и доработку данной СУБД.

В связи с этими преимуществами, а также с общей тенденцией перехода от закрытого программного обеспечения к ПО с открытым кодом, вероятно стремление

пользователей АБИС обратить своё внимание на системы, подобные SLiMS. При этом перед ними встает задача переноса данных из ранее используемой системы в новую. В данной статье рассмотрим программное решение, позволяющее совершить перенос записей библиотечного фонда из популярной АБИС Марк-SQL[1] в SLiMS [2].

Описание процесса загрузки информации в АБИС SLiMS

В первую очередь необходимо экспортировать записи из Марк-SQL. Для этого в главном меню программы нужно выбрать пункт «Документ/Найти все», а затем нажать «Сервис/Выгрузка» в текстовый файл.

Файл экспорта представляет из себя массив строк, каждая строка содержит информацию об одном документе из фонда библиотеки. Информация закодирована

в следующем формате: Автор/Заглавие/Издательство/Дата издания/Объем/ISBN.

Для импорта данных из файла экспорта в базу данных автоматизированной библиотечной системы SLiMS была разработана программа на языке PHP. Выбор PHP как языка обусловлен тем, что сама система SLiMS написана на нем, следовательно у пользователя на сервере гарантированно будет установлен интерпретатор PHP. Приложение принимает в качестве параметра путь к файлу экспорта, устанавливает соединение с базой данных АБИС и вставляет в её соответствующие таблицы информацию из записей в файле.

Представляет интерес моделирование импорта информации из текстового файла в БД SLiMS, также изучение зависимости его времени от технических параметров вычислительной системы. В связи с тем, что не всегда в распоряжении оператора СУБД имеется достаточно производительный компьютер [3], программа импорта данных должна работать за приемлемое время на

том диапазоне количества документов, который соответствует возможным размерам фондов школьной библиотеки.

Проведение вычислительных экспериментов

Компьютерные эксперименты проводились на следующем оборудовании: процессор AMD Athlon 2 X2 250, такт. частота 3 ГГц, кэш первого уровня 64×2 КБ (по 64 кб. на ядро), кэш второго уровня 1024×2 КБ. Оперативная память 4086 МБ DDR3. Замеры производились в операционной системе Linux Mint с помощью утилиты `crufreq`, которая позволяет выставить частоту ядер процессора на определенное значение: 800 МГц, 1,3 ГГц, 1,8 ГГц, 2,3 ГГц, 3 ГГц. Остальные параметры вычислительной системы не менялись. Количество записей в импортируемом файле составляло от 5 до 50 тысяч.

Результаты компьютерных экспериментов представлены на рис. 1

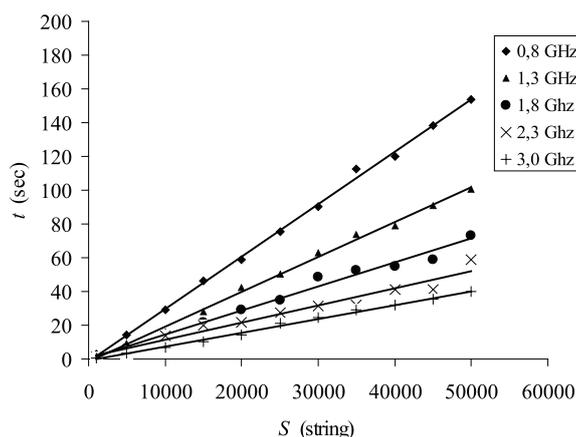


Рис. 1. Зависимость времени загрузки от объема записей: маркеры – экспериментальные данные, сплошные линии – графики аппроксимирующих прямых, полученных МНК

Результаты компьютерных экспериментов показали, что при сравнительно малых объемах записей время загрузки линейно зависит от количества записей (строк) в файле (рис. 1). Это означает, что загрузка происходит с постоянной скоростью при фиксированных параметрах вычислительной системы.

Математическое моделирование загрузки данных

Для формулировки математической модели, описывающей скорость процесса загрузки записей из файла в базу данных SLiMS, удобно рассматривать зависимость объема записей S (количества строк) от времени загрузки t .

Поскольку производная по времени от объема записей представляет собой скорость загрузки, то эта величина должна быть постоянной при неизменных параметрах вычислительной системы. Исходя из данного положения, можно сформулировать математическую модель рассматриваемого процесса:

$$\begin{cases} \frac{dS}{dt} = v, t > t_0 \\ S(t_0) = S_0 \end{cases} \quad (1)$$

где v – скорость загрузки; t_0 , S_0 – эмпирические параметры, зависящие от технических характеристик вычислительной системы. Математическая модель (1) представляет

собой задачу Коши для дифференциального уравнения первого порядка [5]. Эмпирические параметры, входящие в начальное условие, можно условно трактовать, например, как минимальное время загрузки t_0 наименьшего объема записей S_0 .

Решение задачи Коши (1) описывается выражением:

$$S(t) = v(t - t_0) - S_0. \quad (2)$$

Значения эмпирических параметров функции (2) определены методом наименьших квадратов [6] и приведены в таблице для различных частот процессора. Для каждого процессора было задано значение наименьшего файла $S_0 = 1000$, для которого получены значения времени загрузки и скорости.

Зависимость значений эмпирических параметров функции (2) от частоты процессора

f (GHz)	v (string/sec)	t_0 (sec)	R^2
0,8	328,12	2,663664	0,9985
1,3	484,23	0,633757	0,9963
1,8	691,13	0,932617	0,9853
2,3	940,07	1,018653	0,9547
3,0	1214,29	-0,081354	0,9953

Все рассчитанные значения коэффициентов детерминации (достоверности

аппроксимации) для уравнения (2) с параметрами из табл. 1 близки к единице, на основании чего можно утверждать, что предложенная модель согласуется с результатами компьютерных экспериментов [6].

Полученные результаты показывают, что скорость загрузки v линейно возрастает с увеличением тактовой частоты процессора (рис. 2) по закону:

$$v = 412,9 f - 28,2. \quad (3)$$

Установленная экспериментально закономерность роста скорости с увеличением частоты процессора и рассчитанной прямой (3) приведены на рис. 2.

Величина коэффициента детерминации (достоверности аппроксимации) для уравнения прямой (3) $R^2 = 0,9956$. Так как данное значение близко к единице, то это свидетельствует о том, что выражение модели (3) удовлетворительно описывает экспериментальные данные [6].

По результатам проведенной работы можно сделать вывод, что перенос информации о библиотечном фонде из АБИС «Марк-SQL» в SLiMS вполне реализуем даже на относительно слабом по современным меркам аппаратном обеспечении. Используемые в разработке средства (язык программирования PHP, СУБД MySQL, web-сервер apache) являются бесплатными, и доступны для установки и использования на большом числе современных операционных систем.

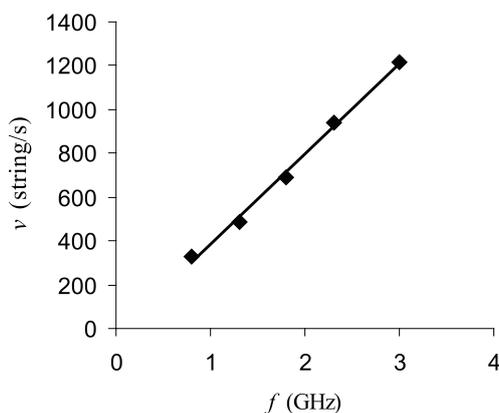


Рис. 2. Зависимость скорости загрузки от частоты процессора: маркеры – значения параметров из табл. 1, сплошная линия – график прямой (3)

Перенос данных из закрытых АБИС в свободные, такие как SLiMS, позволит с меньшими затратами усилий осуществить переход на свободные автоматизированные информационные системы. Разработанное программное решение де-

монстрирует хорошую масштабируемость относительно изменения частоты процессора, поскольку, как следует из результатов эксперимента, время работы механизма переноса данных почти линейно зависит от частоты ЦПУ.

Список литературы

1. Информ–система – программное обеспечение. URL: <http://www.informsystema.ru/> (дата обращения: 22.04.2013).
2. SLIMS | Open Source Library Management System. URL: <http://www.slims.web.id> (дата обращения: 22.04.2013).
3. Земсков А.И. Электронные библиотеки как элемент информационного общества // Научные и технические библиотеки. – 2002. – № 6. – С. 5–15.
4. Воройский Ф.С. Основы проектирования автоматизированных библиотечно-информационных систем. – М.: Физматлит, 2002. – 384 с
5. Тихонов А.Н. Дифференциальные уравнения // А.Н. Тихонов, А.Б. Васильева, А.Г. Свешников. – М.: Наука, 1999. – 232 с.
6. Математическая статистика: учеб. для вузов / В.Б. Горяинов, И.В. Павлов, Г.М. Цветкова и др.; под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. – 424 с.

References

1. Inform-system – software, available at: <http://www.informsystema.ru/> (accessed 22 April 2013).
2. SLIMS Open Source Library Management System, available at: <http://www.slims.web.id/> (accessed 22 April 2013).

3. Zemskov A. I. Jelektronnye biblioteki kak jelement informacionnogo obshhestva. *Nauchnye i tehniczeskie biblioteki*, 2002, no.6, pp. 5–15.

4. Vorojckij F.S. Osnovy proektirovanija avtomatizirovannyh bibliotечно-informacionnyh sistem [Bases for design of automated library information systems]. Moscow, Fizmatlit, 2002. 384 p.

5. Tihonov A.N., Vasil'eva A.B., Svешnikov A.G. Diferencialnye uravnenija [Differential equations]. Moscow, Nauka, 1999. 232 p.

6. Gorjainov V.B., Pavlov I.V., Cvetkova G.M. Matematicheskaja statistika: Ucheb. dlja vuzov [Mathematical Statistics: Textbook for high schools]. Moscow, MGTU im. N. Ja. Bauman, 2001. 424 p.

Рецензенты:

Блажевич С.В., д.ф.-м.н., заведующий кафедрой информатики и вычислительной техники, ФГАОУ «Белгородский государственный национальный исследовательский университет», г. Белгород;

Логачев К.И., д.т.н., профессор, ФГБОУ ВПО «Белгородский государственный технологический университет им. В.Г. Шухова», г. Белгород.

Работа поступила в редакцию 07.06.2013.