

МЕТОДЫ ПОЛУЧЕНИЯ И ОБРАБОТКИ ДАННЫХ В СИСТЕМАХ ПЕЧАТИ ДОКУМЕНТОВ НА ОСНОВЕ ПЕРЕМЕННЫХ ДАННЫХ

Бобков А. А.

Кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: artez1997@gmail.com

В статье рассматривается система печати документов на основе переменных данных, а также методы получения и обработки этих данных. Анализируется предыдущий опыт обработки большого объема данных и улучшения, которые были внедрены на данный момент. Исследуются технические аспекты использования численных моделей.

ВВЕДЕНИЕ

В современном развивающемся мире количество используемых документов стремительно растет. Эффективное функционирование каждой организации, независимо от её профиля, находится в прямой зависимости от уровня оперативной обработки документации и информации, скорости взаимодействия между структурными подразделениями организации [1]. В современных условиях сокращение времени работы с данными возможно, прежде всего, из-за стремительного роста технологий. Поэтому автоматизация документооборота является одной из первостепенных задач современных организаций. Продуктами развития технологий являются различные системы, в частности, системы печати на основе переменных данных, которые были созданы для решения проблем, описанных выше.

I. СИСТЕМА VDP

Система VDP (Variable Data Printing) – это система, предназначенная для печати различного рода документов (контракты, финансовые отчеты и т.д.) на основе переменных данных [2]. Печать переменных данных объединяет набор основных элементов, которые являются общими для копий документа, с набором переменных элементов, которые изменяются от копии к копии. VDP использует технологию цифровой печати, которая связывает базы данных, которые содержат информацию для печати, с печатными устройствами. Данная коммуникация включает правила, которые определяют выбор информации из базы данных и размещения этой информации в документе.

Данная система включает в себя инструменты, которые позволяют пользователю вносить все необходимые изменения в документы без программирования, включая схемы данных, сопоставление данных из внешних файлов и сайтов, а также программирование бизнес-объектов и создание шаблонов. Система также включает в себя инструменты, которые обрабатывают любые данные, предоставленные пользователем, а

затем выдают результат в виде документов различных форматов (например, pdf, docx, pptx, xlsx и т.д.) [3]. Таким образом, решение состоит из двух частей:

1. Проектирование. Определение пользователем того, что должно быть напечатано и каким образом это должно быть отображено.
2. Обработка. Обработка определенных ранее данных и получение результирующего документа.

Схема работы системы представлена ниже (см. рис. 1).

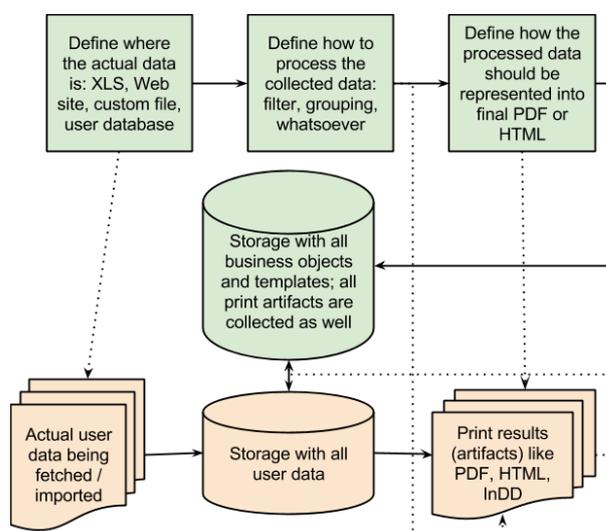


Рис. 1 – Схема работы системы печати VDP

На рисунке выше зеленый цвет означает первую стадию, то есть стадию проектирования будущего документа. Красный цвет означает стадию обработки, то есть стадию опеределения данных и получение финального документа.

Однако в такой системе очень важен вопрос получения и обработки данных для документов [4, 5]. Вопрос скорости импорта данных является ключевым. Таким образом, целью данного исследования является получение наилучшей производительности при получении и обработке дан-

ных, которые используются для печати документов.

II. ОБРАБОТКА ДАННЫХ

Исследование проводится на данных в объеме два гигабайта, которые содержатся в специальном файле формата .DAT. Затем данные файлы подвергаются обработке и приводятся в упорядоченный вид, после чего поступают в базу данных. Первоначальное время импорта данных в базу данных составляло 6 часов.

Далее было решено увеличить скорость, в первую очередь, за счет изменения метода анализа данных. Был разработан новый анализатор данных, который на основе содержимого файла получает множество xml-документов. Каждая строка файла представляет собой xml-документ, который затем подвергается обработке. Полученные в результате обработки данные представляют собой специально созданные структуры данных, которые отображают таблицы базы данных. Был проведен анализ обращения к полям таблиц базы данных и выявлены более оптимальные пути вставки и обновления новых данных [6]. Результаты работы нового алгоритма в сравнении с предыдущим представлены ниже (см. таблица 1).

Таблица 1 – Показатели производительности

Скорость импорта	Средняя	Максимальная	Минимальная
Алгоритм без использования нового метода обработки	21941.55 сек.	21231.6 сек.	22209.88 сек.
Новый алгоритм обработки данных	12843.16 сек.	13103.2 сек.	13935.77 сек.

Таким образом, в результате проведения исследования средняя скорость импорта данных в базу данных сократилась с 21231 секунд (около шести часов) до 13103 секунд (около трех с половиной часов), выигрыш по скорости импорта составил два с половиной часа, то есть обработка данных была ускорена почти в два раза.

На данный момент исследования в данном вопросе продолжаются, поэтому у этого анализа

есть потенциал для развития. В будущем планируется добавление импорта данных на основе различных форматов, например TXT, XLSX и т.д., а не только DAT. Также будет более подробно проанализировано количество обращений к определенным полям базы данных: получены дисперсия и математическое ожидание и другие статистические величины, составлена тепловая карта обращения к полям базы данных. В будущем массив обрабатываемых данных будет постоянно расширяться для обеспечения эффективной обработки всё большего и большего объема данных.

III. ЗАКЛЮЧЕНИЕ

Данная работа позволяет выявить наиболее оптимальные пути получения и обработки большого объема данных, которые в последствии используются для печати документов. В результате данного исследования был разработан новый анализатор данных, который за более быстрое время получает данные в форматированном виде, также были исследованы различные методы по оптимизации работы базы данных.

IV. СПИСОК ЛИТЕРАТУРЫ

1. Внедрение систем электронного документооборота: проблемы и решения [Электронный ресурс] – Режим доступа: <https://iteam.ru/publications/it/> – Дата доступа: 11.10.2020.
2. ABC's of VDP – A Variable Data Printing Basics Guide. [Electronic resource] – Mode of access: <http://help.efi.com/ref/vdp/en-us/VDP.pdf>. – Date of access: 11.10.2020.
3. Variable-Data Printing: Personalized Delivery of Information for the 21st Century [Electronic resource] – Mode of access: <https://web.archive.org/web/20111001035330/> – Date of access: 09.10.2020.
4. 3. Managing Variable Data Full-Color Printing at High Speeds [Electronic resource] /Hohensee, Reinhard. –InfoPrint Solutions, 2011. – Mode of access: <http://www.infoprintsolutionscompany.com/>. – Date of access: 12.10.2020.
5. 4. A New Generation of Variable Data Publishing Solutions.[Electronic resource] /Hohensee, Reinhard. – InfoPrint Solutions, 2011. – Mode of access: <http://www.infoprintsolutionscompany.com/>. – Date of access: 13.10.2020.
6. 5. Молинаро, К. SQL. Сборник рецептов. – Пер. с англ. – СПб: Символ-Плюс, 2009. – 672 с.