

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 002.383

Бобков
Артём Андреевич

Методы получения и обработки данных в системах печати на основе
переменных данных

АВТОРЕФЕРАТ
на соискание академической степени
магистра

по специальности 1-40 80 05 – Программная инженерия

Научный руководитель
Смолякова О.Г.
к.т.н.

Минск 2021

КРАТКОЕ 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. Таким образом 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.

Особый интерес для анализа представляют ситуации, когда необходимо обработать огромный массив данных для последующей обработки и печати документов. Стоит отметить, что в этом случае особую значимость приобретает время обработки входящих данных, так как от этого будет напрямую зависеть эффективность функционирования организации, которая предоставляет соответствующие услуги.

Обработка большого количества и на ее основе печать документов являются весьма трудоемкими задачами, так как для это требуется большое количество времени, а также ресурсов компьютера. Таким образом, наиболее привлекательным становится реализация соответствующего программного средства, использование которого позволяет получить наиболее оптимальные показатели производительности обработки и печати входящих данных.

Исходя из личных интересов автора диссертации, было принято решение исследовать методы эффективной обработки больших данных на примере системы печати VDP.

В ходе исследования были проанализированы используемые системы в этой области, а также программные средства для печати документов. Были рассмотрены документы, которые регламентируют требования к подобным системам и программным средствам на различных уровнях: от локального до глобального.

После анализа и выбора типа модели для исследования были сформулированы идеи по улучшению соответствующих моделей. Улучшенная модель была положена в основу реализации программного средства для получения и обработки данных в системах печати на основе переменных данных.

Разработанное в рамках исследования программное средство может быть интегрировано в систему печати на основе переменных данных для получения и обработки данных.

Общая характеристика работы

Цель и задачи исследования

Целью диссертационной работы является проведение анализа применяемых моделей и алгоритмов, которые используются для получения и обработки данных в системах печати документов на основе переменных данных. На основе проведенного анализа предложить варианты улучшения существующих алгоритмов, а также увеличения производительности обработки больших массивов данных.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Ознакомиться с другими системами печати документов на основе переменных данных.
2. Провести сравнительный анализ различных алгоритмов обработки данных в системах печати документов.
3. На основе сравнительного анализа предложить обоснованные улучшения существующих алгоритмов.
4. Реализовать предложенные алгоритмы и модели, а также провести экспериментальные исследования на основе полученных данных, а также выполнить сравнение работы и показателей производительности с существующими системами печати.

Объектом исследования выступает система печати документов на основе переменных данных.

Предметом исследования являются модели и алгоритмы, которые используются в системах печати для получения и обработки данных.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования программной реализации улучшенных алгоритмов получения и обработки пользовательских данных в системах печати на основе переменных данных.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий по теме «Разработка моделей, методов, алгоритмов, повышающих показатели проектирования, внедрения и эксплуатации программных средств для перспективных платформ обработки информации, решения интеллектуальных задач, работы с большими массивами данных и внедрение в современные производственные комплексы».

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя О.Г. Смоляковой заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Разработанная система печати используется в производстве, а результаты исследований, полученные в ходе выполнения магистерской диссертации интегрированы в разработку и активно используются в данный момент.

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатные работы в сборниках материалов международных научных конференций. Из 1 работа в сборнике материалов международной научной конференции ИТС-2020 БГУИР, другая работа в сборнике материалов IX Республиканской научно-практической конференции «Вычислительные методы, модели и образовательные технологии» 2020 г..

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена описанию и анализу модели печати данных. В третьей главе предложена разработка программной реализации системы печати на основе переменных данных. В четвертой главе проведены экспериментальные исследования, а также проведено сравнение с аналогами.

Общий объем работы составляет 86 страниц, из которых основного текста – 54 страницы, 17 рисунков, 5 таблиц, список использованных источников из 20 наименований и 2 приложения.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В первой главе проведен анализ существующих систем печати на основе переменных данных. Проанализированы достоинства и недостатки существующих систем, а также предложены решения по улучшению данных систем, которые необходимо реализовать в разрабатываемой системе.

Вторая глава посвящена описанию и анализу модели печати данных, а также проектированию разрабатываемой системы печати на основе переменных данных.

В третьей главе предложена разработка программной реализации системы печати на основе переменных данных. Было проведено описание хранилища данных. Был подробно описан процесс создания движка для печати данных, а также модуля импорта данных и модуля парсинга данных.

Четвертая глава посвящена результатам экспериментальных исследований предложенной методики. Проведено сравнение результатов производительности с предыдущими методами обработки данных. Были представлены итоговые показатели производительности, а также описаны методы, которые привели к повышению данных показателей.

ЗАКЛЮЧЕНИЕ

В ходе работы над магистерской диссертацией была глубоко изучена проблематика исследования: различные факторы, которые влияют на производительность запросов и работу самой базы данных, а также проектирование системы печати.

В результате была построена система печати на основе переменных данных, которая способна на основе пользовательских данных генерировать документы формата PDF. Были изучены альтернативные системы печати, на основании этого были предложены и реализованы методы, позволяющие решить проблемы, с которыми столкнулись аналоги. Был разработан новый анализатор данных, который за более быстрое время получает данные в форматированном виде.

При работе над диссертацией и сборе информации были изучены методы работы с базой данных, а также методы обработки больших объемов данных. На основе полученной информации были проведены исследования, которые позволили увеличить производительность обработки данных почти в два раза.

Результаты, полученные в ходе исследования, могут быть использованы не только в системах печати, но и в абсолютно любой системе, основным вопросом для которой является оперативная обработка большого объема данных.

Большое внимание уделялось технической стороне реализации системы. В частности, был применен модульный подход при реализации некоторых подсистем, что позволило получить большую гибкость при выборе источников информации, а также выделить разработанные модули в самостоятельные средства, которые решают конкретные практические задачи.

Данная работа имеет большой потенциал для дальнейших исследований в проблемной области и для расширения функциональности созданной системы.

Возможны улучшения в части скорости импорта данных, использования других моделей и оптимизации реализованных. В будущем планируется добавление импорта данных на основе различных форматов (например, TXT, XLSX, CSV и т.д.), а не только DAT. Также планируется увеличение массива обрабатываемых данных для обеспечения эффективной обработки всё большего и большего объема данных.

Библиотека БГУИР

список ОПУБЛИКОВАННЫХ РАБОТ

1. Бобков, А. А. Методы получения и обработки данных в системах печати на основе переменных данных / Бобков А.А. // Информационные технологии и системы 2020 (ИТС 2020) = Information Technologies and Systems 2020 (ITS 2020) : материалы международной научной конференции, Минск, 18 ноября 2020 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2020. – С. 155–156.

2. Бобков А.А. Методы получения и обработки данных в системах печати на основе переменных данных/ Бобков А.А. // IX Республиканская научно-практическая конференция «Вычислительные методы, модели и образовательные технологии»: сборник материалов. – Брест: БрГУ, 2020.