

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

УДК 004.04-047.44

Войтович
Алексей Сергеевич

АЛГОРИТМ ДИНАМИЧЕСКОЙ ИДЕНТИФИКАЦИИ ТИПА ДАННЫХ

АВТОРЕФЕРАТ

на соискание степени магистра

по специальности 1 - 40 80 02

«Системный анализ, управление и обработка информации»

Научный руководитель

Ломако Александр Викторович,
кандидат технических наук, доцент

Минск 2024

КРАТКОЕ ВВЕДЕНИЕ

Проектирование надежных и оптимальных систем – одна из наиболее сложных задач современной системотехники. Поскольку масштаб технологий и способы обработки информации становятся критически значимыми, надежность в высокопроизводительных системах играет жизненно важную роль в обеспечении продолжительного времени безотказной работы и достоверности выходных данных.

В современном мире большинство приложений взаимодействуют с внешними сервисами и получают доступ к данным в структурированных форматах, таких как *XML* и *JSON*. Однако системы статических типов не предусматривают работу с такими форматами, что часто усложняет доступ к данным. Это приводит к необходимости использования динамических типов данных или проверке данных на этапе выполнения. Соответствующие методы сложны в разработке и сопровождении, что стимулировало поиск более простых и удобных альтернативных путей решения проблемы типизации.

Типизация данных является ключевым аспектом при проектировании систем, так как от неё зависит корректность обработки информации и предотвращение ошибок, связанных с неправильной интерпретацией данных. В связи с этим, разработка методов, позволяющих эффективно интегрировать структурированные данные в системы со строгой типизацией, является актуальной задачей.

Существующие решения, такие как использование библиотек для работы с *JSON* и *XML*, зачастую требуют написания большого количества вспомогательного кода для преобразования данных в структуры, соответствующие строгим типам. Это не только увеличивает объем кода, но и снижает его читабельность и сопровождаемость. Более того, такие подходы могут вводить дополнительные ошибки при преобразованиях, что негативно сказывается на надежности системы.

В магистерской диссертации была четко сформулирована и проанализирована указанная проблема, поставлена задача исследования и, как результат её решения, предложен новый алгоритм определения типа данных и привязки их к строго типизированной схеме. Показано, что этот алгоритм существенно сокращает объем кода доступа к данным, обеспечивает большую производительность и большую безопасность обращения к данным в отличие от широко используемых слабо типизированных методов.

Основной идеей предложенного подхода является автоматическое определение типов данных на основе анализа их структуры и последующая привязка этих данных к строгой типизированной схеме. Это позволяет

минимизировать вмешательство разработчика в процесс преобразования данных, тем самым снижая вероятность ошибок и упрощая процесс разработки[3].

Алгоритм включает несколько ключевых этапов: анализ структуры входных данных, определение их типов, создание выражений для этих данных и их компиляция. После определения типов данных создается схема, соответствующая этим типам, и данные привязываются к этой схеме. Надежность системы также возрастает за счет уменьшения количества ошибок, связанных с неправильной интерпретацией данных.

Таким образом, предложенный алгоритм является эффективным и надежным инструментом для работы с данными в системах со строгой типизацией. Его использование позволяет значительно упростить процесс разработки и сопровождения систем, повысить их производительность и надежность. В будущем возможны дальнейшие исследования и улучшения данного подхода, что откроет новые возможности для разработки высоконадежных и высокопроизводительных систем.

Проектирование систем, способных эффективно обрабатывать структурированные данные, остается важной задачей в области системотехники. Предложенный алгоритм определения типов данных и создания выражений для их обработки является значительным шагом в решении этой задачи, обеспечивая высокую производительность и надежность систем.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования

Типизация данных является важным аспектом при проектировании систем, так как от неё зависит корректность обработки информации и предотвращение ошибок, связанных с неправильной интерпретацией данных. В связи с этим, разработка методов, позволяющих эффективно интегрировать структурированные данные в системы со строгой типизацией, является актуальной задачей. Оптимальное определение типов данных также имеет большое значение для обеспечения качества программных продуктов. При этом в настоящее время не существует метода, который позволял бы обеспечить высокую производительность при гибком доступе к данным различных типов. Очевидна актуальность разработки такого метода.

Цель исследования

Целью данного исследования явилось разработка алгоритма оптимизации определения типов данных с использованием выражений с целью оптимизации доступа и повышения эффективности их обработки.

Задачи исследования

Изучить механизмы определения типов данных, построения выражений и их компиляции на этапе выполнения.

Разработать алгоритм оптимизации определения типов данных с использованием выражений.

Обосновать на практике работоспособность и эффективность предложенного алгоритма оптимизации определения типов данных с использованием выражений.

Область исследования

Информационные технологии в системе высшего образования, используемые для автоматического определения типов данных на основе анализа их структуры и последующей привязки этих данных к строгой типизированной схеме. Это позволяет минимизировать вмешательство разработчика в процесс преобразования данных, тем самым снижая вероятность ошибок и упрощая процесс разработки.

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-40 80 02-2020 специальности 1-40 80 02 «Системный анализ, управление и обработка информации».

Теоретическая и методологическая основа исследования

В основу диссертации легли работы отечественных и зарубежных исследователей в области построения деревьев выражений, кодогенерации, компиляции на этапе выполнения, а также технические нормативные правовые акты по тематике диссертационной работы.

Информационная база исследования сформирована на основе технической литературы, открытой информации, технических нормативно-правовых актов, сведений из электронных ресурсов, а также материалов научных конференций и семинаров.

Новизна полученных результатов

Научная новизна заключается в том, что был предложен алгоритм оптимизации определения типов данных с использованием выражений, учитывающий структуру обрабатываемых данных и рационализирующий операции получения и обработки этих данных. Этот метод, во-первых, обеспечивает снижение трудоемкости разработки вычислительных систем, ориентированных на информационный обмен, а во-вторых, обеспечивает производительность при работе с данными, сопоставимую с заранее скомпилированным исходным кодом.

Личный вклад соискателя

Соискателем выполнены все описанные в диссертации разработки и исследования. Постановка задач и обсуждение результатов проводились совместно с научным руководителем и сотрудниками кафедры информационных технологий автоматизированных систем Белорусского государственного университета информатики и радиоэлектроники. Обработка и интерпретация данных, а также выводы сделаны автором самостоятельно.

Опубликованность результатов диссертации

Изложенные в диссертации основные положения и выводы опубликованы в сборнике *Information Technologies and Systems 2023*. Общий объем публикаций по теме диссертации составляет 2 страницы.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В части, посвященной общей характеристике работы были представлены объект и предмет исследования, сформулированы главные цели и задачи работы, отражены методы исследования, представлены положения, выносимые на защиту, приведены сведения об апробации результатов исследования и их опубликованность, также отражена структура и объем диссертации.

Во введении было произведено обоснование актуальности проблемы идентификации типа данных. Рассмотрены проблемы интеграции в современных вычислительных системах, влияние типизации на корректность работы программного обеспечения.

В первой главе произведен анализ цели и объекта исследования, обзор существующих решений, выполнен анализ и эффективности существующих решений.

Во второй главе рассматриваются актуальные методы исследования производительности. Определен набор исследуемых параметров.

Третья глава предлагает алгоритм оптимизации определения типов данных. Рассматриваются причины, вызывающие дополнительные затраты времени и ресурсов при работе с типами данных на платформе .NET. В ответ на эти вызовы предлагаются методы оптимизации вычислений типов, включая построение дерева выражений и процесс компиляции выражений. Особое внимание уделяется разработке алгоритма, способного эффективно оптимизировать определение типов.

Четвёртая глава посвящена практическому анализу производительности предложенных механизмов и алгоритмов. Описывается подготовка к экспериментам, включая настройку тестового окружения и выбор критериев оценки. В ходе экспериментов исследуется производительность механизма определения типов, а также механизма построения выражений. Анализируется генерируемый код с точки зрения его эффективности и соответствия ожиданиям. Итогом является детальный анализ эффективности разработанного алгоритма определения типов данных, подкреплённый экспериментальными данными.

В приложениях представлена блок-схема алгоритма оптимизации определения типов данных и программный код библиотеки.

ЗАКЛЮЧЕНИЕ

В проведенном исследовании был разработан алгоритм динамической идентификации типов данных и произведен анализ его производительности. Также был проведен сравнительный анализ производительности предложенного алгоритма с наиболее распространённым алгоритмом на основе кеширования.

Разработанный алгоритм оптимизации определения типа данных позволяет достичь производительности получения данных, сравнимой с производительностью получения данных из типа, известного во время компиляции. Экспериментальные результаты показали, что потребление памяти на доступ к данным с использованием представленного алгоритма соответствует потреблению памяти при прямом доступе. Это означает, что алгоритм не требует дополнительных ресурсов для хранения данных, что важно для систем с ограниченными ресурсами. Время выполнения операций с использованием разработанного алгоритма сопоставимо с временем выполнения заранее скомпилированных методов. Это подтверждает, что предложенный алгоритм не только эффективен, но и быстр, что является критическим фактором для высокопроизводительных систем.

Алгоритм показывает линейный рост времени выполнения в зависимости от размера коллекции обрабатываемых данных. Линейная зависимость времени выполнения от размера коллекции подтверждает, что алгоритм хорошо масштабируется и может эффективно работать с большими объемами данных.

Несмотря на значительное преимущество в производительности при использовании алгоритма, его инициализация требует значительных ресурсов. Однако эта инициализация может быть произведена отдельно, что значительно расширяет область применения алгоритма. Время инициализации алгоритма для тестового набора составило всего 1 мс, что обеспечивает прямое преимущество по производительности при использовании с коллекциями более 224 элементов. Это делает алгоритм особенно эффективным для больших объемов данных, где предварительная инициализация может быть оправдана.

Сравнительный анализ показал, что разработанный алгоритм превосходит алгоритм на основе кеширования по ряду параметров. В частности, предложенный алгоритм обеспечивает более стабильное и предсказуемое время выполнения, а также лучше справляется с увеличением объема данных. Это делает его предпочтительным выбором для систем, требующих высокой производительности и надежности.

СПИСОК ПУБЛИКАЦИЙ АВТОРА

[1-А] Войтович, А. С. Метапрограммирование в С# с использованием деревьев выражений / Войтович А. С., Ломако А. В // *Information Technologies and Systems 2023 (ITS 2023)* : материалы Международной научной конференции – Минск, 2023 – С. 179–180.