

МЕТОДИКА ОБРАБОТКИ ПОТОКОВ ДАННЫХ

Белов А.Д.

Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь

Научный руководитель: Ролич О.Ч. – канд.техн.наук, доцент, доцент кафедры ПИКС

Аннотация. Представлены различные виды и способы организации обработки данных, выделены их преимущества и недостатки, приведены примеры использования.

Ключевые слова: обработка данных, ЭВМ, быстрое преобразование Фурье, быстрое преобразование Хартли

Введение. Обработка данных – процесс их приведения к форме, удобной для использования [1]. Независимо от вида информации, которая должна быть получена, и типа оборудования любая система обработки данных выполняет три основные группы операций: сбор данных, собственно их обработку и получение, и анализ результатов. В данной работе представлены основные методики обработки числовых потоков для разных исходных данных.

Основная часть. Для выбора методики необходимо определить режим и способ обработки данных. На рисунке 1 представлена классификационная схема с распространёнными режимами и способами их обработки.

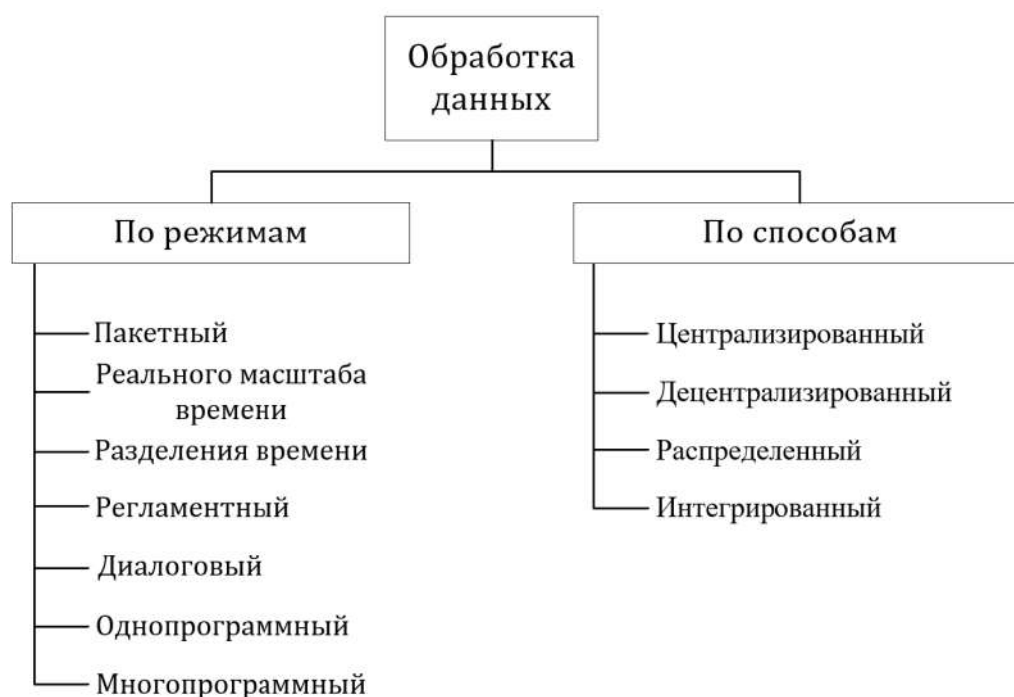


Рисунок 1 – Классификационная схема режимов и способов обработки данных

Основной принцип пакетного режима заключается в том, что пользователь подготавливает данные и формирует их в пакеты. Информация распределяется по пакетам несколькими способами: по времени создания, по типу данных, по источнику, по содержанию и так далее. После подготовки они загружаются в ЭВМ, где система обрабатывает данные по заданному алгоритму. Разделяют два вида выполнения операций: последовательный и параллельный. Этот режим часто используется для обучения искусственных нейронных сетей или при обработке изображений.

Режим реального масштаба времени – режим, при котором вычислительная система взаимодействует с контролируемыми или управляемыми процессами в темпе протекания

процессов. Например, используется для измерения изменений сигнала, поступающего от аналоговых или цифровых датчиков.

Для обработки аналоговых сигналов применяется аналого-цифровой преобразователь, преобразующий сигнал в набор данных, впоследствии используемых ЭВМ. С целью уменьшения погрешности проводится фильтрация потока данных, после чего осуществляется анализ сигнала с помощью быстрого преобразования Фурье или быстрого преобразования Хартли. Результирующий обработанный сигнал анализируется по частоте, амплитуде и фазе. Данный способ применяется в обработке аудиосигнала, сигналов цифровых датчиков, вольтметра и др. Ниже приведена структурная схема спектроанализатора, в основе работы которого лежит данный способ (рисунок 1(а)), и пример работы устройства (рисунок 1(б)).

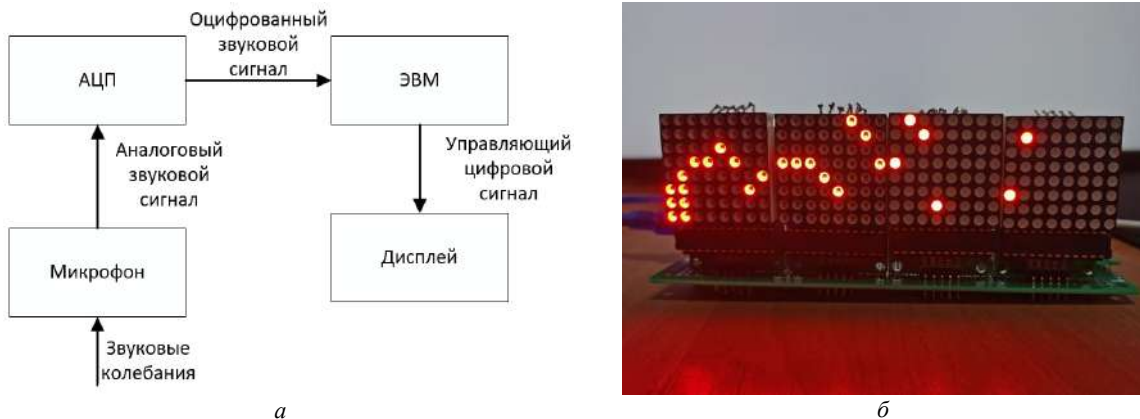


Рисунок 2 – Структурная схема *а* спектроанализатора и пример *б* работы устройства

Для работы с цифровыми датчиками при наличии у ЭВМ необходимых интерфейсов потребность в дополнительных устройствах предварительной подготовки сигнала и данных отсутствует.

В режиме разделения времени система работает с несколькими пользователями, выделяя для каждого определенное количество тактов на выполнения задачи. Если задача не успевает выполняться, она останавливается и переносится в очередь задач.

При работе в диалоговом режиме пользователь имеет возможность непосредственно взаимодействовать с вычислительной системой. Для организации потоковой обработки данных создается программное обеспечение с возможностью выбора разных вариантов алгоритмов работы, получения ответа от ЭВМ и отправки корректирующей или подтверждающей команды.

Регламентный режим характеризуется определенностью отдельных задач пользователя во времени. Например, получение результатных сводок по окончании месяца, расчет ведомостей начисления зарплаты к определенным датам. Сроки решения устанавливаются заранее по регламенту в противоположность к произвольным запросам.

При централизованном способе пользователь доставляет в вычислительный центр данные и получает результаты обработки в виде документов. Особенностью такого способа обработки являются сложность и трудоемкость налаживания быстрой бесперебойной связи, большая загруженность вычислительного центра информацией, организация безопасности системы от возможного несанкционированного доступа.

Децентрализованный способ связан с появлением персональных ЭВМ, дающих возможность автоматизировать конкретное рабочее место.

Распределенный способ обработки данных основан на распределении функций обработки между различными ЭВМ, включенными в сеть. Этот способ может быть реализован двумя путями: первый предполагает установку ЭВМ в каждом узле сети, при этом обработка данных осуществляется одной или несколькими ЭВМ в зависимости от реальных возможностей системы и ее потребностей на текущий момент времени. Второй путь – размещение большого числа различных процессоров внутри одной системы. Для организации

последнего используется система *Docker*. С помощью данной платформы создаются изолированные среды для каждого процесса, что позволяет защитить данные при обработке информации из разных источников, исключить конфликт процессов и распределить нагрузки между задачами.

Интегрированный способ обработки информации предусматривает создание информационной модели управляемого объекта, то есть создание распределенной базы данных. Такой способ обеспечивает максимальное удобство. С одной стороны, базы данных предусматривают коллективное пользование и централизованное управление. С другой стороны, объем информации, разнообразие решаемых задач требуют распределения базы данных. Технология интегрированной обработки информации позволяет улучшить качество, достоверность и скорость обработки, так как обработка производится на основе единого информационного массива, однократно введенного в ЭВМ. Особенностью данного способа является отделение процедуры обработки технологически и по времени от процедур сбора, подготовки и ввода данных.

Заключение. Рассмотрены различные режимы и способы обработки данных, перечислены их преимущества и недостатки, представлен пример использования способа обработки данных в режиме реального масштаба времени.

Список литературы

1. Келлехер, Д.Д. Наука о данных / Д.Д. Келлехер, Брендан Турни. – Москва : Альпина, . – 222 с.
2. Хабр [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/258967/>.
3. Пселтис, Эндрю Дж. Поточковая обработка данных. Конвейер реального времени / Эндрю Дж. Пселтис. – Москва : ДМКПресс, 2018. – 197 с.

UDC 004.042

METHODS OF PROCESSING DATA STREAMS

Belov A.D.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Rolich O.Ch. – PhD, assistant professor, associate professor of the department of ICSD

Annotation. Various types and methods of organizing data processing are presented, their advantages and disadvantages are highlighted, and the examples of use are given.

Keywords: Data processing, computer, fast Fourier transform, fast Hartley transform