ИСПОЛЬЗОВАНИЕ ПОНЯТИЯ «ДИАПАЗОН» ПРИ РАБОТЕ С БОЛЬШИМИ НАБОРАМИ ДАННЫХ В ЭКСПЕРИМЕНТАЛЬНЫХ РАСЧЕТАХ

Л.В. Яблокова

Самарский государственный аэрокосмический университет им. академика С.П. Королёва (национально исследовательский университет)

В статье рассматривается понятие диапазона, как основного средства проектирования и реализации алгоритмов обработки больших наборов данных, используемых в экспериментальных расчетах математической физики.

Разумное использование современной вычислительной техники немыслимо без умелого применения приближенного и численного анализа. Этим объясняется чрезвычайно возросший интерес к методам приближенных вычислений и анализа экспериментальных данных содержащих большой объем информации. Как никакая другая научная ветвь, численные методы, как основной инструмент математической физики, самым тесным образом переплелись с многочисленными программными приложениями, являясь либо средством, либо предметом исследования. При этом задачи, решаемые современной наукой и современными технологиями, значительно усложнились, и способ их решения не ограничивается только общими вопросами. Вычислительные методы и алгоритмы обработки больших наборов данных, как средство успешного решения новых сложных задач, задействуются в таких областях как численное решение уравнений Маквелла и д'Аламбера. Как правило, для создания сложной математической компьютерной модели, от специалиста требуется не только безупречная математическая подготовка, но и знание основ современных, высокопроизводительных алгоритмов, а также эффективные способы их применения для данных большого размера. В настоящее время основная масса библиотек для научных расчетов содержит реализации множества базовых алгоритмов. Однако ручная реализация простых вариантов основных алгоритмов позволяет лучше понять их и, следовательно, эффективнее использовать и настраивать более совершенные библиотечные версии. Также важным обстоятельством, для повторной реализации основных алгоритмов, является тот факт, что мы достаточно часто сталкиваемся с новыми вычислительными средами, с их новыми свойствами, которые не могут быть наилучшим образом задействованы в старых реализациях. Реализуя базовые версии алгоритмов, более приспособленными к конкретным задачам, а, не основываясь на системных подпрограммах, специалисты-исследователи могут добиться большей переносимости и дольше сохранить актуальность, применяемых в процессе решения, алгоритмов и структур данных. Кроме того, несмотря на усовершенствования, встроенные в программные библиотеки для научных расчетов, механизмы, применяемые для совместного использования программ, не всегда достаточно мощны, чтобы библиотечные функции можно было бы легко приспособить к эффективному выполнению в рамках конкретной задачи. Обеспечение наиболее эффективной реализации конкретного алгоритма, ориентированного на работу с большим набором данных, позволит использовать его для решения сложных задач вычислительного характера и, причем многократно.

Предположим, что имеется последовательность, элементы которой должны быть обработаны каким-либо алгоритмом. В процессе такой обработки происходит обращение ко всем элементам последовательности, начиная от первого и заканчивая последним. Подобная ситуация встречается в процессе реализации алгоритмов так часто, что для обрабатываемых элементов последовательности существует специальное обозначение – диапазон. Например, диапазон $[first, last)$ состоит из всех элементов от $first$ до $last$, не включая последний. Такая ассиметричная форма записи используется для того, чтобы акцентировать внимание на том, что $[first, last)$ является полуоткрытым интервалом, который включает в себя все элементы, начиная с $first$, но не $last$.

Диапазон $[first, last)$ является допустимым, если ко всем его элементам, исключая $last$ можно получить доступ и если элемент, стоящий перед $last$ достижим из $first$, т.е. если последовательно перемещаясь, начиная с $first$, по всем позициям диапазона конечное число раз можно попасть в позицию предшествовавшую $last$. Таким образом, например, допустимым является диапазон $[0, N)$. Допустимым так же является и пустой диапазон $[0,0)$, а вот диапазон $[N-1,0)$ является недопустимым, так как элемент в позиции $N-1$ появляется позднее элемента в позиции 0. Поэтому не имеет смысла говорить об элементах в позициях от $N-1$ до 0 как о допустимом диапазоне.

Диапазоны должны удовлетворять следующим свойствам:

1. Для любого элемента в позиции i, диапазон $[i, i)$ является допустимым диапазоном.
2. Если $[i, N)$ – допустимый диапазон, то $[i+1, N)$ – также допустимый диапазон.
3. Если $[i, N)$ – допустимый диапазон, а элемент в позиции k достижим из элемента в позиции i, а элемент в позиции $N-1$ достижим из элемента в позиции k, то диапазоны $[i, k)$ и $[k, N)$ являются допустимыми диапазонами.
4. Если оба диапазона $[i, k)$ и $[k, N)$ являются допустимыми, то диапазон $[i, N)$ – также допустимый диапазон.

С фундаментальной точки зрения диапазоны определены таким образом потому, что их ассиметричная форма помогает избежать ошибок, так называемой, потерянной единицы, т.е. количество элементов в диапазоне $[i, N)$ равно $N-i$, а именно ровно столько, сколько и ожидается.

Используя концепцию диапазона можно перейти к другому более общему понятию интервальной функции. Понятие интервальной функции обеспечивает возможность создания алгоритмов способных обрабатывать любые структуры данных поддерживающих понятие одномерной последовательности элементов, например одномерные массивы в языках Fortran и др. Кроме того, написание кода программы с использованием интервальных функций требует меньших усилий, а решения с их применением обычно выглядят более наглядно и логично.

Литература

1. Артемов И.Л. Fortran: основы программирования. – М.: Диалог-МИФИ, 2007. – 304с.
2. Голуб Дж., Ван Лоун Ч. Матричные вычисления: Пер. с англ. – М.: Мир, 1999. – 548с., ил.
3. Деммель Дж. Вычислительная линейная алгебра. Теория и приложения. Пер. с англ. – М.: Мир, 2001. – 430с., ил.
4. Подбельский В.В., Фомин С.С. Курс программирование на языке Си: Учеб. Пособие. – М.: ДМК Пресс, 2013. – 384с.