

*Посвящается моим родителям
Гизелле и Хельмуту Брассам*

Оглавление

От авторов перевода	8
Предисловие	10
Основные понятия.....	12
Примеры кода.....	14
Глава 1. Элементарные структуры	16
1.1. Стек.....	16
1.2. Очередь.....	23
1.3. Двусторонняя очередь.....	30
1.4. Динамическое выделение памяти для элементов.....	30
1.5. Теневые копии структур в массиве.....	32
Глава 2. Деревья поиска	37
2.1. Две модели деревьев поиска.....	37
2.2. Общие свойства и преобразования.....	40
2.3. Высота дерева поиска.....	43
2.4. Основные операции: поиск, добавление, удаление.....	44
2.5. Возврат от листа к корню.....	49
2.6. Повторяющиеся ключи.....	50
2.7. Интервалы ключей.....	51
2.8. Построение оптимальных деревьев поиска.....	53
2.9. Преобразование деревьев в списки.....	59
2.10. Удаление деревьев.....	60
Глава 3. Выровненные деревья поиска	62
3.1. Выровненные по высоте деревья.....	62
3.2. Выровненные по весу деревья.....	72
3.3. (a, b)- и B-деревья.....	82
3.4. Красно-черные деревья и деревья почти оптимальной высоты.....	96
3.5. Нисходящее выравнивание красно-черных деревьев.....	107
3.6. Деревья с постоянным временем обновления в заранее известном месте.....	116
3.7. Пальцевые деревья и поуровневое связывание.....	118
3.8. Деревья с частичной перестройкой: средняя оценка сложности.....	123
3.9. Дерево с всплывающими узлами: изменяемая структура данных.....	126
3.10. Списки с пропуском элементов: вероятностные структуры данных.....	137
3.11. Соединение и разделение выровненных деревьев поиска.....	144
Глава 4. Древовидные структуры на множестве интервалов	149
4.1. Деревья пересекающихся отрезков.....	149
4.2. Деревья полуоткрытых интервалов.....	154
4.3. Деревья объединения интервалов.....	161
4.4. Деревья сумм взвешенных интервалов.....	168
4.5. Деревья поиска интервалов с ограниченной максимальной суммой весов.....	173
4.6. Деревья прямоугольных областей.....	178
4.7. Деревья многомерных интервалов.....	190
4.8. Блочные структуры данных.....	193

4.9. Подсчет точек и модель полугруппы.....	195
4.10. <i>kd</i> -деревья и связанные с ними структуры.....	197
Глава 5. Кучи	202
5.1. Куча как выровненное дерево	203
5.2. Куча в массиве.....	207
5.3. Куча как упорядоченное и полуупорядоченное дерево	213
5.4. Левосторонние кучи.....	218
5.5. Косые кучи	225
5.6. Двоичные кучи	228
5.7. Изменение ключей в кучах	236
5.8. Кучи Фибоначчи	238
5.9. Кучи оптимальной сложности	248
5.10. Двусторонние и многомерные кучи	253
5.11. Связанные с кучей и постоянным временем обновления структуры	257
Глава 6. Системы непересекающихся множеств и связанные с ними структуры..	263
6.1. Непересекающиеся множества: объединение классов разделов.....	264
6.2. Система непересекающихся множеств с поддержкой копирования и динамическими деревьями интервалов	277
6.3. Разделение списка.....	287
6.4. Проблемы ориентированных корневых деревьев	291
6.5. Поддержание линейного порядка	301
Глава 7. Преобразования структуры данных.....	305
7.1. Создание динамических структур.....	305
7.2. Динамические структуры с сохранением истории	314
Глава 8. Структуры данных для строк	319
8.1. Проверки и сжимаемые проверки	320
8.2. Словари, допускающие появление ошибок в запросах	337
8.3. Суффиксные деревья.....	341
8.4. Суффиксные массивы.....	347
Глава 9. Хеш-таблицы.....	355
9.1. Основные хеш-таблицы и разрешение конфликтов.....	355
9.2. Универсальные семейства хеш-функций	360
9.3. Идеальные хеш-функции.....	370
9.4. Хеш-деревья.....	375
9.5. Расширяемое хеширование	376
9.6. Проверка принадлежности ключей и фильтры Блума	380
Глава 10. Приложение	383
10.1. Ссылочная машина и другие модели вычислений.....	383
10.2. Модели внешней памяти и алгоритмы без учета кеш-памяти.....	385
10.3. Названия структур данных	386
10.4. Решение линейных рекуррентных соотношений	386
10.5. Медленно растущие функции.....	388
Глава 11. Список публикаций.....	390
Предметный указатель	423

От авторов перевода

Это не первый наш опыт перевода книги, изданной много лет назад (Вирт, Гуткнехт). Но тем не менее мы взялись за перевод этой якобы старой книги, изданной 14 лет назад, потому что она не потеряла своей ценности и по сей день представляет интерес не только для профессионалов, но и для студентов, поскольку в ней подробно описывается, как путем незначительных усовершенствований давно и хорошо известных программистам структур данных можно добиться существенной эффективности работы с ними, что автор подкрепляет еще и доказательствами.

В сущности, книга представляет собой обзор различных способов модернизации давно известных структур данных с целью повышения их производительности. Поскольку в книге по большей части речь идет об оценке времени вычисления алгоритмов, то в таких случаях она подразумевается неявно. В тех же случаях, когда речь идет об оценке затрат памяти алгоритма, это явно оговаривается.

В целом книга представляет собой довольно полный обзор различных методов и приемов, позволяющих в той или иной степени ускорить работу с традиционными структурами данных, приводя необходимые обоснования (доказательства) их эффективности и во многих случаях – алгоритмы их реализации на языке программирования С. К сожалению, автор нередко приводит вместо программной реализации алгоритмов их словесное описание, что несколько затрудняет их понимание.

Книга может послужить хорошим справочником для всех тех, кто хочет понять, за счет каких приемов и методов можно ускорить работу с общеизвестными структурами данных.

Большинство так называемых автором «теорем» – это не столько теоремы, сколько выводы из предшествующих этим «теоремам» рассуждений. Книга насыщена такими «теоремами», но без доказательств. Обоснования этих «теорем» не всегда точны и убедительны. Те читатели, кто не интересуется ими, могут довериться автору и просто пропустить их. Ну а тем, кто хочет разобраться в них досконально, придется приложить определенные усилия или обратиться к многочисленным первоисточникам.

Конечно, эта книга адресована не столько программистам-любителям, сколько профессионалам, включая преподавателей и студентов, а также другим специалистам, интересующимся этой темой. Нужно сказать, что эта книга не для всех, а только для тех, кто хочет понять, с помощью каких методов и приемов можно достичь значительного ускорения работы алгоритмов с давно и многим известными структурами данных большого объема, включая базы данных.

Эта книга придется не всем по зубам, но мы приложили максимум усилий для того, чтобы ее перевод на русский язык хотя бы немного расширил круг ее читателей, и все найдут в ней немало познавательного и полезного.

О переводе

В главе 10 автор совершенно справедливо говорит о том, что термины и понятия в оригинальных работах не всегда удачны. Поэтому при переводе мы, конечно же, старались придерживаться авторской терминологии и тех оригинальных работ, на которые он ссылается, но изредка позволяли себе употреблять более понятные, с нашей точки зрения, русскоязычные термины. Что из этого получилось, судить читателю.

Несмотря на то что книга может заинтересовать довольно узкий круг читателей, мы надеемся, что наш перевод несколько расширит этот круг. Оригинал книги представляет собой скорее конспект лекций, чем детально продуманную монографию. Мы постарались превратить ее в полноценную монографию и, надеемся, интересную книгу для более широкого круга читателей – от студентов и преподавателей.

давателей до профессиональных программистов (специалистов), которые, как мы думаем, не разочаруются в ней.

В предисловии к книге автор пишет, что она родилась из его набросков лекций по данной теме. И к сожалению, этот набросочный стиль изложения материала отчасти сохранился и в самой книге. Поэтому при переводе книги нам пришлось немало потрудиться, чтобы превратить авторские наброски в полноценную монографию, качество которой оценит наш читатель.

Несколько слов о том, что красной нитью проходит по всей книге, – это оценка пространственно-временной сложности алгоритмов. Для них обычно используются обозначения: O – верхняя асимптотическая оценка сложности для худшего случая, Ω – нижняя асимптотическая оценка для худшего случая и Θ – жесткая асимптотическая оценка, когда сложность алгоритма для худшего случая не может выходить за пределы определяемых ею границ. Кроме того, нередко в книге идет речь об амортизированной (amortized) оценке вычислений, которую мы называли средней или усредненной. Чаще всего речь в книге идет именно об асимптотической оценке временной сложности алгоритмов для худшего случая и лишь изредка оценивается нижняя, жесткая, средняя и пространственная сложность алгоритмов.

При переводе книги мы позволили себе ради ее улучшения, помимо прочих незначительных вольностей, пронумеровать все пронумерованные авторские рисунки, а также обширный список публикаций.

Терминология

Сложность алгоритма – это оценка его стоимости, цены и прочих затрат времени или памяти.

Временная сложность – это количество выполняемых операций или просто время их выполнения.

Пространственная сложность – это затраты памяти на выполнение операций.

Чаще всего автор имеет в виду временную сложность.

hieght balanced tree – выровненное по высоте дерево или равновысокое дерево.

weight balanced tree – уравновешенное или равновесное дерево.

rebalancing – выравнивание.

Формулы

Некоторые несложные «многоэтажные» математические формулы в виде дробей мы иногда заменяли строчной записью, конечно же, без потери их смысла.

Так, например,

$$\frac{-1}{(\alpha \log \alpha + (1 - \alpha) \log(1 - \alpha))} \log n$$

мы заменили на

$$-1/(\alpha \log \alpha + (1 - \alpha) \log(1 - \alpha)) \log n.$$

Предисловие

Эта книга – учебник для слушателей курсов по структурам данных. Структура данных – это метод¹ представления данных для выполнения над ними некоторого набора операций. Классический пример структуры данных – это множество распознаваемых по значению ключа элементов в виде пар (*ключ, элемент*), которые можно добавлять, удалять или искать в этом множестве по заданному ключу. Структура, поддерживающая такие операции, называется словарем (*dictionary*). Словари могут быть реализованы разными способами, с разной степенью сложности и разными дополнительными операциями. Во многих первоисточниках было предложено и исследовано множество видов словарей, и некоторые из них² мы рассмотрим в этой книге.

Вообще говоря, структура данных – это своего рода набор высокоуровневых операций некоторой виртуальной машины: если алгоритм многократно выполняет некоторые операции, можно определить, какие из них выполняются чаще всего и как их лучше реализовать. Таким образом, основная задача структур данных – как реализовать набор операций над ними, предполагаемый результат которых нам известен.

Сегодня нет недостатка в книгах, в названии которых есть словосочетание «*структуры данных*», но они лишь поверхностно касаются этой темы, затрагивая только простейшие структуры – стек, очередь и некоторые виды выровненных деревьев поиска с изрядным количеством пафоса (*handwaving*). Серьезный интерес к структурам данных начали проявлять в 1970-х годах, а в первой половине 1980-х почти в каждом номере журнала *Communications ACM* была статья о них. Структуры данных стали центральной темой, отдельным пунктом в *классификаторе информатики*³ и стандартной частью учебных программ по информатике⁴. С выходом книги Вирта «*Структуры данных + Алгоритмы = Программы*» термин «*алгоритмы и структуры данных*» стал общепринятым в учебниках по этой теме. Правда, единственная монография Овермарса 1983 года [450], посвященная именно алгоритмическому аспекту структур данных, до сих пор находится в печати; это своего рода рекорд для серии книг *LNCS (Lecture Notes in Computer Science)*. Структурам данных уделяется внимание во многих приложениях и прежде всего в базах данных как индексным структурам. В том же контексте в монографиях Самета [482, 483] были изучены структуры для

¹ Эта книга не об объектно ориентированном программировании, поэтому термины «метод» и «объект» употребляются в ней в самом обычном смысле.

² Наиболее интересные и значимые. – *Прим. перев.*

³ Классификационный код Е.1. «Структуры данных». К сожалению, Классификатор информатики слишком груб, чтобы быть полезным.

⁴ АВЕТ (Accreditation Board for Engineering and Technology, Inc. – *Прим. перев.*) по-прежнему причисляет их к одной из пяти основных тем: алгоритмы, структуры данных, разработка ПО, языки программирования и архитектура компьютеров.

геометрических данных; подобные структуры данных были исследованы Лангетепе и Заманом [346] для компьютерной графики. В последнее время были подробно изучены строковые структуры данных, востребованные в первую очередь приложениями биоинформатики. Не иссякает поток публикаций по теории структур данных для вычислительной геометрии и комбинаторики. Однако во многих учебниках структуры данных рассматриваются только как примеры для объектно ориентированного программирования, исключая их важнейший алгоритмический аспект – как реализовать нетривиальные структуры данных, не выходя при этом за пределы худших оценок вычислительной сложности. Цель этой книги – сосредоточиться на структуре данных как основе любого алгоритма. Недавно изданный *Справочник по структурам данных* [411] – это еще один шаг в этом же направлении.

В книге приводится реальный код для многих обсуждаемых в ней структур данных и довольно подробная информация о большинстве других структур данных, реализация которых в книге не дается. Многие учебники избегают таких подробностей, что является одной из причин, по которой структуры данных не используются там, где они должны быть. Выбор рассматриваемых в этой книге структур данных почти всюду ограничивается только теми структурами, которые вписываются в модель так называемой *ссылочной машины* (pointer-machine)⁵, за исключением хеш-таблиц, которые рассматриваются из-за их практической важности. Все коды книги – сугубо иллюстративные и не предназначены для использования в качестве готовых модулей; их правильность не гарантируется. Большинство кодов доступны на моей домашней странице с минимальными возможностями их проверки.

Эта книга начиналась с набросков курса, который я читал в зимнем семестре 2000 года в Берлинском свободном университете (Free University of Berlin). Я благодарю Христиана Кнауэра, который был тогда моим ассистентом; мы с ним многому научились. Затем я представил этот курс в осенних семестрах 2004–2007 годов аспирантам Нью-Йоркского городского колледжа (City College of New York), а в июле 2006 года взял его за основу для летней школы по структурам данных в Корейском институте передовых технологий (Korean Advanced Institute of Science and Technology – KAIST). Работа над книгой была закончена в ноябре 2007 года.

Благодарю Эмили Войтек и Гюнтера Роте за выявленные в моих кодах ошибки, Отфрида Чонга – за организацию летней школы в KAIST, а также самих участников летней школы – за выявленные ими мои прочие ошибки. Благодарю Христиана Кнауэра и Хельмута Брасса за литературу из превосходных математических библиотек Берлинского свободного университета и Брауншвейгского технического университета, а также Яноша Паха за доступ к онлайн-журналам Института Куранта. Эта книга

⁵ Один из вариантов абстрактной машины с косвенной адресацией (Random Access Machine, сокращенно – RAM), предназначенной для теоретической оценки сложности алгоритмов. – *Прим. перев.*

была бы невозможна без доступа к хорошим библиотекам, и я старался цитировать только те источники, которые сам просматривал.

Данная книга не поддерживалась ни одним грантовым агентством.

Основные понятия

Структура данных моделирует некоторый абстрактный объект, обладающий набором операций, которые обычно можно разделить на:

- операции создания и удаления;
- операции обновления;
- операции обращения к данным (запросы).

В случае словаря мы создаем или удаляем объект, обновляем его, добавляя или удаляя его элементы, или запрашиваем его элемент путем обращения к нему. После создания объекта он изменяется посредством операций обновления. Операции обращения не изменяют сам абстрактный объект, но могут изменить вид объекта внутри структуры данных. Это называется *адаптивной* (adaptive) структурой данных, которая может перестраиваться после каждого обращения к ней с целью ускорения последующих обращений к ее данным.

Допускающие подобные обновления и обращения структуры данных называются *динамическими* (dynamic). Правда, есть более простые структуры, которые создаются только один раз и допускают обращения к своим элементам без их изменения; они называются *статическими* (static). Динамические структуры данных, как более общие, предпочтительнее, однако мы не должны пренебрегать и статическими структурами, так как они являются строительными блоками для динамических структур, хотя для некоторых довольно сложных объектов, с которыми мы столкнемся, динамическая структура вообще неизвестна.

Наша цель – определить и реализовать структуру данных так, чтобы она обеспечивала максимальную скорость работы с заданным абстрактным объектом. Размер структуры – еще один ее качественный показатель, но его влияние на скорость обычно незначительно. Для оценки скорости нужна некая мера. Обычно ею является размер самого объекта, а не его представление. Отметим, что в результате многократного применения операции обновления (добавления/удаления) размер объекта может даже уменьшиться. Наиболее общая мера сложности вычислений – это так называемая *худшая* (worst-case) сложность⁶. Таким образом, сложность операции с заданной структурой данных оценивается величиной $O(f(n))$, если в любом состоянии структуры данных после применения ряда операций обновления, приведших объект к размеру n , эта операция требует времени,

⁶ То есть *асимптотическая верхняя граница вычислительной сложности алгоритма*. Далее в книге автор называет эту оценку *временем выполнения алгоритма*, что не вполне правильно. – *Прим. перев.*

не превышающего $Cf(n)$ для некоторой константы C . Альтернативная, но более мягкая мера – это *средняя* (*amortized*) сложность⁷. Операция обновления имеет среднюю сложность $O(f(n))$, если существует такая функция $g(n)$, что любая последовательность из m таких операций, не увеличивающих размер n самого объекта, требует времени не более $g(n) + mCf(n)$, поэтому при многократном применении такой операции сложность вычислений в среднем не превысит $Cf(n)$.

Некоторые структуры данных можно отнести к *случайным* (*randomized*). Обращение к одному и тому же объекту выполняется неким случайным образом, и одни и те же действия над ним не всегда приводят к одинаковой последовательности шагов. В этом случае оценивается *ожидаемая* (*expected*) сложность операции; она тоже случайна и потому относится к худшей для всех объектов того же размера с теми же операциями над ними.

В некоторых случаях худшая оценка сложности вычислений $O(f(n))$ может ухудшаться из-за большого размера результата (выхода) операции, влияющего на ее сложность. В таких случаях имеет значение еще и чувствительность алгоритма к выходу, размер которого влияет на скорость алгоритма. Операция имеет *чувствительную к выходу* (*output-sensitive*) сложность $O(f(n) + k)$, если для объекта размера n и выхода размера k она требует времени не более $C(f(n) + k)$.

Для динамических структур данных время создания нового объекта обычно постоянно, поэтому нас будут интересовать главным образом сложность операций обновления объекта структуры и обращения к нему. Время удаления структуры размера n почти всегда равно $O(n)$. Для статических структур данных объект размера n создается сразу и не обновляется, поэтому в этом случае нас будет интересовать только время его создания (предварительной обработки) и время обращения к нему.

В этой книге $\log_a n$ – это логарифм по основанию a , но если основание логарифма не указано, то оно равно 2.

Для интервалов используется стиль записи Бурбаки: $[a, b]$ и $]a, b[$ – соответственно закрытый и открытый интервалы от a до b , а $]a, b]$ и $[a, b[$ – полуоткрытые интервалы с исключением, соответственно, начала и конца.

Подобно приведенному выше обозначению $O(\cdot)$ для асимптотической верхней границы вычислительной сложности, мы будем использовать также обозначения $\Omega(\cdot)$ – для нижней границы и $\Theta(\cdot)$ – для жесткой (верхней и нижней) границы. Неотрицательная функция f принадлежит $O(g(n))$ или $\Omega(g(n))$, если для некоторого положительного C и любых достаточно больших n выполняется условие $f(n) \leq Cg(n)$ или $f(n) \geq Cg(n)$ соответственно. Таким образом, функция f принадлежит $\Theta(g(n))$, если она принадлежит одно-

⁷ Обычно это некая усредненная оценка сложности выполнения цепочки операций над структурой данных, когда оценивается их *средняя* сложность в худшем случае. Она применяется, когда наиболее сложные операции выполняются довольно редко, но средняя сложность всей цепочки операций оказывается вполне приемлемой. Отметим, что средняя оценка не является вероятностной (ожидаемой). Это всего лишь средняя оценка $a = (\sum_{i=1}^n t_i)/n$, где t_1, t_2, \dots, t_n – сложность (или времена) выполнения в худшем случае цепочки из n операций над структурой данных. – *Прим. перев.*

временно $O(g(n))$ и $\Omega(g(n))$. Здесь «достаточно большие n » – это те значения, для которых $g(n)$ определена и положительна.

Примеры кода

Примеры кода в этой книге приведены в стандарте языка C, но они будут вполне понятны всем, кто знаком с любым другим императивным языком программирования.

В примерах кода символ «=» обозначает присваивание, а символ «==» – проверку на равенство. За пределами кода символ «=» имеет обычный математический смысл.

Логические операции отрицания, конъюнкции и дизъюнкции обозначаются в коде соответственно символами «!», «&&» и «||», а символ «%» обозначает операцию деления по модулю.

Разыменованье указателя (ссылки) выполняется операцией «*»: если `pt` – указатель на область памяти, то `*pt` – это сама область памяти. Указатели имеют тип, определяющий способ интерпретации содержимого этой области памяти. Таким образом, при объявлении указателя объявляется и тип области памяти, на которую он ссылается. Так, например, `<int *pt;>` объявляет указатель `pt` на область памяти типа `int`. Указатели – это переменные, которые допускают операции присваивания и добавления к ним целого числа (указательная арифметика). Если `pt` указывает на область памяти определенного типа, то `pt + 1` указывает на следующую область памяти такого же типа. Иными словами, каждый объявленный указатель трактует всю память как большой массив однотипных элементов. `NULL` – это указатель, который не указывает ни на какую область памяти, и его можно использовать в качестве особого значения в операциях сравнения.

Структура – это определяемый пользователем тип данных, представляющий собой последовательность именованных элементов (полей) любого типа. Элемент структуры тоже может быть структурой, но отличной от типа самой структуры, иначе структура становится рекурсивной, то есть бесконечной. Но зато ее элементом вполне может быть указатель на объект того же типа, что и сама структура. Именно такие структуры обычно бывают объектами в теории структур данных. Полям структур можно присваивать значения и использовать их подобно любым другим переменным. Если, например, переменная `z` определена как структура следующего типа C:

```
typedef struct { float x; float y; } C,
```

то ее элементы `z.x` и `z.y` имеют тип `float`. Если переменная `zpt` объявлена как указатель на объект типа C (`C *zpt;`), то она может ссылаться на `(*zpt).x` и `(*zpt).y`. Для такой часто используемой комбинации операций разыменовывания указателя и обращения к элементу структуры в языке C существует альтернативная операция `zpt->x` и `zpt->y`. Эти операции эквивалентны, но последняя предпочтительнее, поскольку снимает проблему приоритета

операций: разыменованное имеет более низкий, чем обращение к элементу структуры, приоритет, поэтому `(*zpt).x` – это не то же самое, что `*zpt.x`.

Во всех функциях мы избегаем рекурсивных вызовов, хотя в некоторых случаях они могут существенно упростить код. Однако значительные издержки рекурсивных вызовов функций вступают в противоречие с нашей целью достижения наивысшей скорости работы со структурами данных. При этом мы не противники рекурсивных функций, поскольку хороший компилятор сам может избавиться от них, используя встроенные функции или макросы.

Кроме того, в тексте книги объект данных нередко будет отождествляться со ссылкой на него.

Глава 1

Элементарные структуры

Элементарные структуры данных *стек* и *очередь* обычно изучаются в курсе «Программирование 2». Их объединяет иногда упоминаемая, но редко используемая на практике *двусторонняя очередь*. Стек и очередь – основные структуры данных, именно поэтому они будут подробно обсуждаться и использоваться для иллюстрации отдельных моментов реализации других структур данных.

1.1. Стек

Стек – простейшая из всех структур с очевидной интерпретацией: добавить (втолкнуть) объект в стек и затем удалить (вытолкнуть) его из стека с доступом только к верхнему объекту (вершине). По этой причине его иногда называют памятью типа LIFO (Last In, First Out – последним вошел, первым вышел)¹. В программировании стеки встречаются всюду, где есть вложенные блоки, локальные переменные, рекурсия или перебор с возвратами (backtracking). Типичные примеры программ с использованием стеков – это вычисление скобочных арифметических выражений с различными приоритетами операций или поиск пути в лабиринте методом перебора с возвратами.

Стек должен поддерживать, по крайней мере, следующие операции:

- `push(obj)` – добавить (втолкнуть) объект `obj` в вершину стека;
- `pop()` – удалить (вытолкнуть) объект из вершины стека;
- `stack_empty()` – проверить, пуст ли стек.

Реализация этих операций должна, конечно же, обеспечивать правильные результаты, поэтому необходимо как-то определить правильное поведение стека. Одним из способов может быть алгебраическое определение операций и возвращаемых ими значений. Для простых структур, подобных стеку, это возможно, но для понимания структуры такое определение не очень полезно. Вместо него можно определить каноническую реализацию на некой абстрактной машине с бесконечной памятью, которая выдает

¹ Типичный пример LIFO – пистолетная обойма с патронами. – *Прим. перев.*

правильные результаты для всех правильных цепочек операций (исключая операцию удаления из пустого стека). Если элементы стека имеют тип `item_t`, то такая реализация может выглядеть следующим образом:

```
int i=0;
item_t stack[∞];
int stack_empty(void)
{ return(i == 0);
}

void push( item_t x)
{ stack[i++] = x;
}

item_t pop(void)
{ return(stack[--i]);
}
```

Операции со стеком реализованы правильно, но в них есть одна проблема – это массив бесконечного размера, в котором всякая последовательность операций всегда будет правильной. Поэтому более реалистичной может быть следующая версия:

```
int i=0;
item_t stack[MAXSIZE];

int stack_empty(void)
{ return(i == 0);
}

int push(item_t x)
{ if( i < MAXSIZE )
  { stack[i++] = x;
    return(0);
  }
  else
    return(-1);
}

item_t pop(void)
{ return(stack[--i]);
}
```

Введенное ограничение максимального размера стека ограничивает правильное поведение стека, поэтому это не совсем то, что нам хотелось бы иметь. Но зато в этой реализации операция `push` возвращает сообще-

ние об ошибке (-1) при переполнении стека. Главный недостаток реализации структур данных в массиве заключается в том, что массивы имеют фиксированный размер, который должен быть задан заранее вне зависимости от истинного количества элементов структуры. Для устранения этого недостатка существует систематический способ, рассматриваемый в разделе 1.5, но предпочтение обычно отдается решению с динамически выделяемой памятью.

В этой реализации ошибка выдается только при *переполнении* (overflow) стека, но не при его *опустошении* (underflow), поскольку переполнение стека – это ошибка, порожденная самой структурой, что невозможно в идеальной реализации, тогда как опустошение стека – это ошибка неправильного использования структуры и, следовательно, результат работы программы, считающей стек «черным ящиком». Ее функция pop возвращает вершину стека, но если нужно «поймать» ошибку опустошения стека, то она должна возвращать еще и признак ошибки. Последняя претензия к первоначальной версии реализации стека состоит в том, что в одной и той же программе может понадобиться не один, а несколько стеков, которые придется создавать динамически. Поэтому нам нужны дополнительные операции по созданию и удалению стеков, а каждая операция со стеком должна знать, с каким именно стеком она работает. Одним из возможных вариантов реализации стека может быть следующий:

```
typedef struct {item_t *base; item_t *top; int size;} stack_t;

stack_t *create_stack(int size)
{ stack_t *st;
  st = (stack_t *) malloc(sizeof(stack_t));
  st->base = (item_t *) malloc(size * sizeof(item_t));
  st->size = size;
  st->top = st->base;
  return(st);
}

int stack_empty(stack_t *st)
{ return(st->base == st->top);
}

int push( item_t x, stack_t *st)
{ if (st->top < st->base + st->size)
  { *(st->top) = x;
    st->top += 1;
    return( 0 );
  }
  else
    return( -1 );
}
```

```

}

item_t pop(stack_t *st)
{ st->top -= 1;
  return(*(st->top));
}

item_t top_element(stack_t *st)
{ return(*(st->top - 1));
}

void remove_stack(stack_t *st)
{ free(st->base);
  free(st);
}

```

Сюда мы включаем только некоторые проверки безопасности и исключаем другие. В общем, наша политика безопасности состоит в том, чтобы включать только те проверки, которые, в отличие от идеального стека, выявляют ошибки, вызванные ограничениями этой реализации при правильном использовании стека и неограниченном объеме памяти базовой операционной системы. Кроме того, мы добавили еще одну полезную операцию, которая просто возвращает вершину стека без ее удаления.

Часто более предпочтительной реализацией стека является структура с динамически выделяемой памятью в виде связанного списка, когда добавление и удаление элемента выполняется в начале списка. Преимущество такого решения в том, что у такой структуры нет фиксированного размера, поэтому (считая память компьютера неограниченной) к ней всегда можно добавить новый элемент без проверки на ошибку переполнения стека. Она так же проста, как и структура в массиве, при наличии функций `get_node` и `return_node`, реализация которых будет приведена в разделе 1.4.

```

typedef struct st_t { item_t item; struct st_t *next; } stack_t;

stack_t *create_stack(void)
{ stack_t *st;
  st = get_node();
  st->next = NULL;
  return(st);
}

int stack_empty(stack_t *st)
{ return(st->next == NULL);
}

void push(item_t x, stack_t *st)

```

```

{ stack_t *tmp;
  tmp = get_node();
  tmp->item = x;
  tmp->next = st->next;
  st->next = tmp;
}

item_t pop(stack_t *st)
{ stack_t *tmp; item_t tmp_item;
  tmp = st->next;
  st->next = tmp->next;
  tmp_item = tmp->item;
  return_node(tmp);
  return(tmp_item);
}

item_t top_element(stack_t *st)
{ return(st->next->item);
}

void remove_stack(stack_t *st)
{ stack_t *tmp;
  do
  { tmp = st->next;
    return_node(st);
    st = tmp;
  }
  while( tmp != NULL );
}

```

Обратите внимание, что связный список начинается с пустого элемента (заголовка), поэтому пустой стек – это список из одного заголовка, а вершина стека – это уже следующий элемент списка. Это нужно для того, чтобы возвращаемый функцией `create_stack` и используемый во всех операциях указатель стека не изменялся ими. Поэтому указатель вершины стека не может быть его идентификатором. Поскольку извлекаемые из списка элементы становятся недоступными, нам нужны временные копии их значений в `pop` и `remove_stack`. Операция `remove_stack` должна возвращать все оставшиеся в стеке элементы. Однако нет оснований полагать, что удаляться будут только пустые стеки, поэтому при неудачной попытке удалить оставшиеся элементы не исключена утечка памяти.

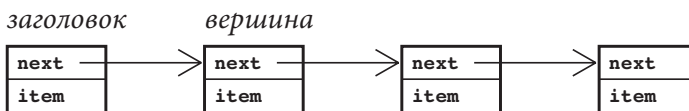


Рис. 1.1. Стек в виде списка из трех элементов

Реализация структуры с динамически выделяемой памятью более элегантна. Ей не грозит переполнение стека, и ей нужна только фактически используемая память, а не большие массивы, предельный размер которых определяется программистом. Один из недостатков такой реализации – возможное снижение производительности: разыменованное указание по времени не больше, чем увеличение индекса, однако адрес памяти, куда ссылается указатель, может находиться в любом месте памяти, тогда как следующий элемент массива находится вслед за предыдущим. Таким образом, структуры в массиве обычно очень хорошо работают с кешем, тогда как структуры с динамически выделяемой памятью могут создавать в кеше много пустот. Поэтому если мы совершенно уверены в максимально возможном размере стека (например, потому что его размер – всего лишь логарифм размера входных данных), то предпочтительна его реализация в массиве.

Если же нужно объединить достоинства обеих реализаций, можно использовать связный список блоков (массивов), по заполнении каждого из которых к списку добавляется новый блок. Такая реализация может выглядеть следующим образом:

```
typedef struct st_t { item_t *base;
                    item_t *top;
                    int size;
                    struct st_t *previous; } stack_t;

stack_t *create_stack(int size)
{ stack_t *st;
  st = (stack_t *) malloc(sizeof(stack_t));
  st->base = (item_t *) malloc(size * sizeof(item_t));
  st->size = size;
  st->top = st->base;
  st->previous = NULL;
  return(st);
}

int stack_empty(stack_t *st)
{ return( st->base == st->top && st->previous == NULL);
}

void push( item_t x, stack_t *st)
{ if( st->top < st->base + st->size )
  { *(st->top) = x;
    st->top += 1;
  }
  else
  { stack_t *new;
```

```
new = (stack_t *) malloc(sizeof(stack_t));
new->base = st->base;
new->top = st->top;
new->size = st->size;
new->previous = st->previous;
st->previous = new;
st->base = (item_t *) malloc(st->size * sizeof(item_t));
st->top = st->base + 1;
*(st->base) = x;
}
}

item_t pop(stack_t *st)
{ if( st->top == st->base )
  { stack_t *old;
    old = st->previous;
    st->previous = old->previous;
    free(st->base);
    st->base = old->base;
    st->top = old->top;
    st->size = old->size;
    free(old);
  }
  st->top -= 1;
  return(*(st->top));
}

item_t top_element(stack_t *st)
{ if( st->top == st->base )
  return(*(st->previous->top - 1));
else
  return(*(st->top - 1));
}

void remove_stack(stack_t *st)
{ stack_t *tmp;
  do
  { tmp = st->previous;
    free(st->base);
    free(st);
    st = tmp;
  }
  while( st != NULL );
}
```

Согласно нашей классификации операции `push` и `pop` обновляют структуру, тогда как `stack_empty` и `top_element` просто обращаются к ней. Очевидно, что реализация в массиве позволяет выполнять все действия за постоянное время, поскольку они включают только постоянное число операций. Реализация в виде связанного списка предполагает при выполнении операций `push` и `pop` разовые вызовы внешних функций `get_node` и `return_node`, поэтому они требуют постоянного времени, но лишь в предположении, что сами эти функции имеют фиксированное время выполнения. Реализацию с динамическим созданием элементов мы обсудим в разделе 1.4, но можем здесь (и во всех остальных структурах) допустить, что время их выполнения постоянно. Для связанного списка блоков вместо создания промежуточного слоя, как описано в разделе 1.4, используются стандартные операции управления памятью `malloc` и `free`, которые выделяют и освобождают большие области памяти. Обычно считается, что выделение и освобождение памяти – это операции с постоянным временем выполнения, но (в особенности для операции `free`) здесь возникают нетривиальные проблемы, поэтому следует избегать их частого употребления. В случае списка блоков это может произойти, например, при наличии большого количества операций `push` / `pop`, которые могут выводить за пределы выделенного блока. Так что незначительные преимущества связанного списка блоков, видимо, не стоят лишних проблем.

В операции `create_stack` выполняется только одно выделение памяти, и потому время ее выполнения при любой реализации должно быть постоянным. А вот операция `remove_stack`, очевидно, не имеет постоянного времени выполнения, поскольку должна удалять потенциально большую структуру: если в стеке n элементов, то время его удаления – $O(n)$.

1.2. Очередь

Очередь – почти такая же простая структура, как стек; элементы в ней хранятся так же, но в отличие от стека сначала извлекаются те, что были помещены в нее первыми, то есть это память типа FIFO (First In, First Out – первым вошел, первым вышел). Очереди полезны в задачах с циклической обработкой данных. Кроме того, они являются центральной структурой при поиске в ширину (Breadth-First Search, BFS). Поиск в ширину и поиск в глубину (Depth-First Search, DFS) различаются, по сути, только тем, что для хранения очередного объекта BFS использует очередь, а DFS – стек.

Очередь должна поддерживать, по крайней мере, следующие операции:

- `enqueue(obj)` – добавить объект `obj` в конец очереди;
- `dequeue()` – вернуть 1-й объект очереди с удалением его из очереди;
- `queue_empty()` – проверить, пуста ли очередь.

Разница между стеком и очередью, делающая ее чуть сложнее, – в том, что изменения происходят на обоих ее концах: на одном конце – добавле-

ния, на другом – удаления. Если очередь реализуется в массиве, то используемый ею отрезок массива как бы перемещается внутри него. При бесконечном массиве ее реализация не представляла бы никакой проблемы и выглядела бы так:

```
int lower=0; int upper=0;
item_t queue[∞];
int queue_empty(void)
{ return(lower == upper);
}

void enqueue(item_t x)
{ queue[upper++] = x;
}

item_t dequeue(void)
{ return(queue[lower++]);
}
```

На практике очередь в массиве ограниченной длины реализуется в виде кольца, когда индекс вычисляется по модулю длины массива. Тогда ее реализация могла бы выглядеть так:

```
typedef struct { item_t *base;
                int    front;
                int    rear;
                int    size; } queue_t;

queue_t *create_queue(int size)
{ queue_t *qu;
  qu = (queue_t *) malloc( sizeof(queue_t) );
  qu->base = (item_t *) malloc( size *sizeof(item_t) );
  qu->size = size;
  qu->front = qu->rear = 0;
  return(qu);
}

int queue_empty(queue_t *qu)
{ return(qu->front == qu->rear);
}

int enqueue(item_t x, queue_t *qu)
{ if( qu->front != ((qu->rear +2)% qu->size) )
  { qu->base[qu->rear] = x;
    qu->rear = ((qu->rear+1)%qu->size);
    return(0);
  }
```

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru