

РЕФЕРАТ

Расчетно–пояснительная записка 50 с., 2 рис., 1 табл., 55 ист., 1 прил.

УПРАВЛЕНИЕ ПАМЯТЬЮ, СБОРКА МУСОРА, JAVA, PYTHON

Цель работы: классификация алгоритмов распределения памяти в языках программирования с автоматической сборкой мусора.

В данной работе проводится обзор и сравнение алгоритмов автоматического управления памятью в языках программирования Python, Java, JavaScript, C# и Golang.

Среди рассмотренных автоматических менеджеров памяти были выделены менеджеры памяти языков программирования Python и Java как наиболее масштабируемые и оптимизированные для различных сценариев использования.

СОДЕРЖАНИЕ

РЕФЕРАТ	3
ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	7
ВВЕДЕНИЕ	9
1 Аналитический раздел	11
1.1. Управление памятью с точки зрения операционной системы	11
1.1.1. Иерархия памяти	11
1.1.2. Функции операционной системы по управлению памятью	12
1.2. Управление памятью с точки зрения приложений	12
1.2.1. Среда выполнения языка программирования	12
1.2.2. Управление памятью	13
1.2.2.1. Ручное управление памятью	15
1.2.2.2. Автоматическое управление памятью	17
1.2.3. Трассирующая сборка мусора	18
1.2.4. Подсчёт ссылок	20
1.3. Существующие решения в области автоматического управления памятью	21
1.3.1. Python	21
1.3.1.1. Структура объекта	21
1.3.1.2. Сборка мусора	22
1.3.2. Java	23
1.3.2.1. Управление памятью	23
1.3.2.2. Сборка мусора	24
1.3.3. Javascript	26
1.3.4. C#	27
1.3.4.1. Управление памятью	27
1.3.4.2. Сборка мусора	28
1.3.5. Golang	30
1.3.5.1. Структура кучи	30
1.3.5.2. Сборка мусора	31
1.3.6. Классификация существующих решений	33

1.4.	Классификация алгоритмов по требованиям к дополнительной памяти	36
1.5.	Постановка задачи	39
2	Конструкторский раздел	41
2.1.	Общая схема метода	41
2.2.	Выбор способа использования метода	41
2.3.	Описание необходимых структур данных	41
2.4.	Проектирование алгоритмов работы с памятью	41
2.4.1.	Алгоритм выделения памяти	41
2.4.2.	Алгоритм освобождения памяти	41
2.4.3.	Алгоритм сбора мусора	41
2.4.4.	Алгоритм сбора циклических ссылок	41
3	Технологический раздел	42
3.1.	Выбор языка программирования	42
3.2.	Компоненты программного обеспечения	42
3.2.1.	Аллокатор	42
3.2.2.	Поколение	42
3.2.3.	Сборщик мусора	42
3.2.4.	Анализатор объектов	42
3.2.5.	Мониторинг работы менеджера памяти	42
3.3.	Тестирование компонентов программного обеспечения	42
3.3.1.	Аллокатор	42
3.3.2.	Анализатор объектов	42
3.3.3.	Сборщик мусора	42
4	Исследовательский раздел	43
4.1.	Цель проводимых измерений	43
4.2.	Описание проводимых измерений	43
4.3.	Инструменты измерения времени работы программы	43
4.4.	Технические характеристики оборудования	43
4.5.	Результаты проведённых измерений	43
4.5.1.	Одиночный запуск программы	43
4.5.2.	Веб-сервер	43

4.5.2.1.	Открытая линейная нагрузка	43
4.5.2.2.	Открытая постоянная нагрузка	43
4.5.2.3.	Закрытая нагрузка	43
ЗАКЛЮЧЕНИЕ		44
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ		49
ПРИЛОЖЕНИЕ А		50

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

- 1) **Стек** — область адресного пространства процесса, предназначенная для хранения параметров функций, локальных переменных и адреса возврата после вызова функции.
- 2) **Куча** — область адресного пространства процесса, предназначенная для выделения памяти, динамически запрашиваемой программой.
- 3) **Внутренняя фрагментация** — явление, при котором аллокатор выделяет при каждом запросе больше памяти, чем фактически запрошено.
- 4) **Внешняя фрагментация** — явление, при котором свободная память разделена на множество мелких блоков, ни один из которых нельзя использовать для обслуживания запроса на выделение памяти.
- 5) **Управление памятью** — процесс координации и контроля использования памяти в вычислительной системе.
- 6) **Автоматическое управление памятью** — служба, которая автоматически перерабатывает память, которую программа не собирается использовать в дальнейшем.
- 7) **Мусор** — объект, о котором достоверно известно, что он не используется программой, так как является недостижимым¹ из других её объектов.
- 8) **Сбор мусора** — автоматическая переработка динамически выделяемой памяти.
- 9) **Сборщик мусора** — автоматический менеджер памяти, реализующий некоторый алгоритм сборки мусора.

¹ Термин «недостижимость» может иметь различные интерпретации в разных языках программирования

- 10) Барьер** — функция, которая принимает указатель на объект в памяти, анализирует его статус и в зависимости от него выполняет какие-либо действия с этим указателем или даже с объектом, на который он ссылается, после чего возвращает актуальное значение указателя, которое можно использовать для доступа к объекту.
- 11) Финализатор объекта** — функция, которая запускается, когда сборщик мусора определяет, что рассматриваемый объект недоступен в программе.
- 12) Финализация объекта** (finalization, termination) — процесс выполнения финализаторов объекта.

ВВЕДЕНИЕ

Память является одним из основных ресурсов любой вычислительной системы, требующих тщательного управления. Под памятью (memory) в работе будет подразумеваться оперативная память компьютера. Особая роль памяти объясняется тем, что процессор может выполнять инструкции программы только в том случае, если они находятся в памяти. Память распределяется между операционной системой компьютера и прикладными программами. [1]

Для выполнения вычислений в языках программирования используются объекты, которые могут быть представлены как простым типом данных (целые числа, символы, логические значения и т.д.) так и агрегированным (массивы, списки, деревья и т.д.). Значения объектов программ хранятся в памяти для быстрого доступа. Во многих языках программирования переменная в программном коде — это адрес объекта в памяти. [2] [3] [4] Когда переменная используется в программе, процесс считывает значение из памяти и обрабатывает его.

Большинство современных языков программирования активно использует динамическое распределение памяти, при котором выделение объектов осуществляется во время выполнения программы. Динамическое управление памятью вводит два основных примитива — функции выделения и освобождения памяти, за которые отвечает **аллокатор**.

Существует два способа управления динамической памятью — ручное и автоматическое. При ручном управлении памятью программист должен следить за освобождением выделенной памяти, что приводит к возможности возникновения ошибок. Более того, в некоторых ситуациях (например, при программировании на функциональных языках или в многопоточной среде) время жизни объекта не всегда очевидно для разработчика. [5] Автоматическое управление памятью избавляет программиста от необходимости вручную освобождать выделенную память, устраняя тем самым целый класс возможных ошибок и увеличивая безопасность разрабатываемых программ. Сборка мусора (garbage collection) за последние два десятилетия стала стандартом в области автоматического управления памятью, хотя её использование накладывает дополнительные расходы памяти и времени исполнения. На сегодняшний день среды времени выполнения (language runtime) многих популярных языков программирования, таких как Java, C#, Python и другие, активно используют сбор-

ку мусора.

Современные приложения, в особенности разработанные для длительной непрерывной работы, подвержены влиянию следующих факторов:

- 1) утечки памяти;
- 2) фрагментация памяти;
- 3) длительные паузы на сборку мусора.

Перечисленные факторы могут негативно сказываться на производительности и отказоустойчивости приложений. Снизить влияние данных факторов на работу программ и повысить эффективность работы с памятью можно на основе новых качеств методов автоматического управления памятью.

Целью данной работы является разработка метода автоматического управления памятью с гарантированным временем выполнения на основе подсчёта ссылок. Для достижения поставленной цели необходимо решить следующие задачи.

1. Провести анализ и классификацию существующих методов распределения памяти в языках программирования с автоматической сборкой мусора.
2. Провести классификацию алгоритмов для определения области применения разрабатываемого метода.
3. Спроектировать и реализовать метод автоматического управления памятью.
4. Определить границы применения реализованного метода.

1 Аналитический раздел

Управление памятью вычислительной системы выполняется на трёх уровнях.

1. Аппаратное обеспечение управления памятью (MMU, ОЗУ и т. д.).
2. Управление памятью операционной системы (виртуальная память, защита).
3. Управление памятью приложения (выделение и освобождение памяти, сборка мусора).

Аппаратное обеспечение управления памятью состоит из электронных устройств и связанных с ними схем, которые хранят состояние вычислительной системы. К этим устройствам относятся регистры процессора, кэш, ОЗУ, MMU (Memory Management Unit, блок управления памятью) и вторичная (дисковая) память. Конструкция запоминающих устройств имеет решающее значение для производительности современных вычислительных систем. Фактически пропускная способность памяти является основным фактором, влияющим на производительность системы. [6]

Далее будет рассмотрено подробно управление памятью с точки зрения операционной системы и приложений.

1.1. Управление памятью с точки зрения операционной системы

1.1.1. Иерархия памяти

В процессе развития аппаратного обеспечения была разработана концепция **иерархии памяти**, согласно которой компьютеры обладают несколькими мегабайтами очень быстродействующей, дорогой и энергозависимой кэш-памяти, несколькими гигабайтами памяти, средней как по скорости, так и по цене, а также несколькими терабайтами памяти на довольно медленных, сравнительно дешевых дисковых накопителях, не говоря уже о сменных накопителях, таких как DVD и флеш-устройства USB. Превратить эту иерархию в абстракцию, то есть в модель, а затем управлять этой абстракцией — и есть задача операционной системы. Та часть операционной системы, которая управляет иерархией памяти (или ее частью), называется **менеджером**, или **диспетчером**,

памяти [1]. Он должен следить за тем, какие части памяти используются, выделять память процессам, которые в ней нуждаются, и освобождать память, когда процессы завершат свою работу. Выбор, совершаемый менеджером памяти на этом этапе, может оказать существенное влияние на будущую эффективность программы, так как до 40% (в среднем 17%) времени её выполнения затрачивается на выделение и освобождение памяти. [7]

1.1.2. Функции операционной системы по управлению памятью

Помимо первоначального выделения памяти процессам при их создании операционная система должна также заниматься динамическим распределением памяти, то есть обслуживать запросы приложений на выделение им дополнительной памяти во время выполнения. После того, как приложение перестаёт нуждаться в дополнительной памяти, оно может вернуть её системе. Выделение случайного объёма памяти в случайные моменты времени из общего пула приводит к фрагментации памяти и, вследствие этого, к неэффективному её использованию. Дефрагментация памяти также является функцией операционной системы.

Таким образом, можно сформулировать следующие функции ОС по управлению памятью в мультипрограммных системах.

1. Отслеживание свободной и занятой памяти.
2. Управление виртуальными адресными пространствами процессов, в том числе отображение виртуального адресного пространства процесса на физическую память.
3. Динамическое распределение памяти.
4. Дефрагментация памяти.

1.2. Управление памятью с точки зрения приложений

1.2.1. Среда выполнения языка программирования

Менеджер памяти операционной системы с точки зрения выполняющихся на ней приложений обладает двумя серьёзными недостатками. Во-первых,

он является недетерминированным: процессы не могут напрямую влиять на решения, принимаемые операционной системой по управлению их адресными пространствами. Во-вторых, Менеджер памяти операционной системы не реализует те функции управления памятью, которые требуются конечному пользователю. Поэтому языки программирования реализуют собственную **среду выполнения** (language runtime), в которой реализуют необходимые пользователю функции для работы с памятью и создают собственные абстракции памяти.

Среда выполнения языка программирования может выполнять следующие функции: [8]

- 1) управление памятью;
- 2) обработка исключений;
- 3) сборка мусора;
- 4) контроль типов;
- 5) обеспечение безопасности;
- 6) управление потоками.

Управление потоками в среде выполнения языка подразумевает, что язык программирования организует многопоточное выполнение программы, используя возможности операционной системы, а также реализует некоторую **модель памяти**, которая для императивных языков программирования, поддерживающих многопоточное выполнение, определяет, какие записи в разделяемые переменные могут быть видны при чтениях, выполняемых другими потоками. [9]

1.2.2. Управление памятью

Менеджер памяти приложения должен учитывать следующие ограничения. [10]

1. **Нагрузка на процессор** — дополнительное время, затрачиваемое диспетчером памяти во время работы программы.
2. **Блокировки** — время, необходимое диспетчеру памяти для завершения операции и возврата управления программе. Это влияет на способность

программы оперативно реагировать на интерактивные события, а также на любое асинхронное событие, например связанное с сетевым взаимодействием.

3. **Накладные расходы памяти** — дополнительный объём памяти, затрачиваемый на администрирование, а также накладные расходы, связанные с внутренней и внешней фрагментацией.

Основная проблема управления памятью состоит в том, чтобы понять, когда следует сохранить содержащиеся в ней данные, а когда их следует уничтожить, чтобы память можно было использовать повторно. К сожалению, существует множество способов, которыми плохая практика управления памятью может повлиять на надежность и быстродействие программ, как при ручном, так и при автоматическом управлении памятью.

Ниже перечислены наиболее частые проблемы управления памятью. [10]

1. **Преждевременное освобождение памяти** (premature free). Когда программы освобождают память, но позже пытаются получить к ней доступ, их поведение становится неопределённым. Сохранившаяся ссылка на освобождённую память называется **висячим указателем** (dangling pointer).
2. **Утечки памяти** (memory leaks). Когда программы интенсивно выделяют память, не освобождая её, в конечном итоге память заканчивается.
3. **Внешняя фрагментация** (external fragmentation). Работа аллокатора по выделению и освобождению памяти может привести к тому, что он больше не сможет выделять достаточно большие области памяти, несмотря на наличие достаточного общего количества свободной памяти. Это связано с тем, что свободная память может быть разделена на множество небольших областей, разделённых используемыми данными. Данная проблема особенно критична в серверных приложениях, работающих в течение длительного времени.
4. **Неэффективное расположение ссылок** (poor locality of reference). Эта проблема связана с тем, как современные менеджеры памяти аппаратного обеспечения и операционной системы обращаются с памятью: после-

довательные обращения к памяти выполняются быстрее, если они производятся к соседним областям памяти. Если менеджер памяти разместит данные, которые программа будет использовать вместе, не в соседних областях памяти, то быстродействие программы уменьшится.

5. **Негибкий дизайн** (inflexible design). Менеджеры памяти могут вызвать серьезные проблемы с производительностью, если они были разработаны с одной целью, а используются для другой. Эти проблемы вызваны тем, что любое решение по управлению памятью опирается на предположения о том, как программа будет использовать выделенную память. Например, на стандартные размеры областей памяти, шаблоны ссылок или время жизни объектов. Если эти предположения неверны, то диспетчер памяти может работать с памятью менее эффективно.
6. **Межмодульное взаимодействие** (interface complexity). Если объекты передаются между модулями программы, то при проектировании интерфейсов модулей необходимо учитывать управление их памятью.

Управление памятью приложения объединяет две взаимосвязанные задачи: выделение памяти (allocation) и её переиспользование (recycling), когда она больше не требуется. За выделение памяти отвечает **аллокатор** [11]. Его необходимость обусловлена тем, что процессы, как правило, не могут заранее предсказать, сколько памяти им потребуется, поэтому они нуждаются в реализации дополнительной логики обслуживания изменяющихся запросов к памяти. Решение об освобождении и переиспользовании выделенной аллокатором памяти, которая больше не используется приложением, может быть принято либо программистом, либо средой выполнения языка. Соответственно, в зависимости от этого управление памятью в языке программирования может считаться либо ручным, либо автоматическим.

1.2.2.1. Ручное управление памятью

При ручном управлении памятью программист имеет прямой контроль над временем жизни объектов программы. Как правило, он осуществляется либо явными вызовами функций управления кучей (например, malloc и free в C), либо языковыми конструкциями, влияющими на стек управления (например,

объявлениями локальных переменных). Ключевой особенностью ручного менеджера памяти является то, что он даёт возможность явно указать, что заданная область памяти может быть освобождена и переиспользована. [10]

Преимущества ручного управления памятью:

- явное выделение и освобождение памяти делает программы более прозрачными для разработчика;
- ручные менеджеры памяти, как правило, используют память более экономно, так как программист может минимизировать время между моментом, когда выделенная память перестаёт использоваться, и её фактическим освобождением.

Недостатки ручного управления памятью:

- увеличение исходного кода программ за счёт того, что управление памятью, как правило, составляет значительную часть интерфейса любого модуля;
- повышение дублирования кода за счёт использования однотипных инструкций управления памятью;
- увеличение числа ошибок управления памятью из-за человеческого фактора.

К языкам с ручным управлением памятью относятся C, C++, Zig и другие. На таких языках программисты могут писать код, дублирующий поведение менеджера памяти либо путем выделения больших областей и их разделения для использования, либо путем внутреннего переиспользования этих блоков. Такой код называется **субаллокатором** (suballocator) [6], так как он работает поверх другого аллокатора. Субаллокаторы могут использовать как преимущество специальные знания о поведении программы, но в целом они менее эффективны, чем использование базового аллокатора. Также стоит отметить, что субаллокаторы могут быть неэффективными или ненадежными, тем самым создавая новый источник ошибок. [11]

1.2.2.2. Автоматическое управление памятью

Автоматическое управление памятью, как правило, работает либо как часть среды выполнения языка, либо как расширение. Автоматические менеджеры памяти обычно перерабатывают области памяти, которые недоступны из переменных программы (то есть области, к которым невозможно получить доступ с помощью указателей). Задачи автоматического управления памятью: [10]

- выделение памяти под новые объекты;
- идентификация используемых объектов;
- освобождение памяти, занятой неиспользуемыми объектами.

Преимущества автоматического управления памятью:

- освобождение разработчика от задачи управления памятью в своих программах;
- уменьшение объёма исходного программ за счёт отсутствия необходимости явно работать с памятью;
- уменьшение числа ошибок управления памятью;
- автоматическое управление памятью может быть более эффективным, чем ручное, за счёт применения соответствующих алгоритмов.

Недостатки автоматического управления памятью:

- время жизни объектов программы становится непрозрачным для разработчика, так как часть логики работы с памятью реализована на уровне языка;
- память, как правило, используется менее экономно, так как существует некоторая задержка между моментом, когда память перестаёт использоваться, и моментом её освобождения; такие задержки определяются алгоритмами автоматической переработки памяти.

Автоматическое управление памятью используется в большинстве современных языков программирования, среди которых C#, Golang, Haskell, Java, JavaScript, Python, Swift, частично Rust, C++ при использовании умных указателей и другие.

В языках с автоматическим управлением памятью часто необходимо выполнять действия над некоторыми объектами после того, как они перестали использоваться, и до того, как память, которую они занимают, может быть переиспользована. Для этого используется механизм финализации объектов, который, как правило, используется для освобождения ресурсов, когда работающий с ними объект перестаёт использоваться. Например, открытый файл может быть представлен объектом потока ввода-вывода. Когда менеджер памяти подтверждает, что этот объект больше не используется в программе, то можно быть гарантировать, что файл больше не используется программой и его нужно закрыть до повторного использования потока. Стоит отметить, что момент финализации объекта программы, как правило, не фиксирован. Этот факт может стать проблемой, если финализация используется для управления ограниченными ресурсами операционной системы, такими как файловые дескрипторы. [6]

Как правило, автоматическая сборка мусора реализуется либо методом подсчёта ссылок, либо с использованием трассирующего сборщика мусора. [12] Целью идеального сборщика мусора является освобождение пространства, используемого каждым объектом, как только он перестаёт использоваться программой. Стоит отметить, что для автоматического управления памятью также можно использовать гибридные методы. [13] [14]

1.2.3. Трассирующая сборка мусора

Трассирующий сборщик мусора [12] — автоматический менеджер памяти, который следует указателям, чтобы определить, какие области памяти доступны из переменных программы, называемых **корневым набором**). Далее будут описаны наиболее популярные алгоритмы сборки мусора. Стоит отметить, что алгоритмы могут сочетаться и заменяться во время выполнения программы в зависимости от параметров кучи, таких как заполненность и фрагментация. [15]

Алгоритм mark-sweep реализует рекурсивное определение достижимости объекта по указателям. Сборка мусора осуществляется в два этапа. Сначала

сборщик обходит граф объектов, начиная с «корней» (регистры, стеки потоков, глобальные переменные), через которые программа могла бы немедленно получить доступ к объектам, а затем, следуя указателям, сканирует каждый найденный объект. Такой обход называется **трассировкой**. На втором этапе, этапе очистки (sweep), сборщик проверяет каждый объект в куче: любой неразмеченный объект считается мусором и освобождается. [15]

Mark-sweep — это алгоритм косвенного сбора данных. В отличие от косвенных методов, прямые алгоритмы, такие как подсчёт ссылок, определяют достижимость объекта исключительно по самому объекту, без обращения к трассировке. [15]

Рассматриваемый алгоритм не обнаруживает мусор как таковой, а скорее идентифицирует все используемые объекты, а затем приходит к выводу, что всё остальное должно быть мусором. Стоит заметить, что сборщику необходимо размечать используемые объекты при каждом вызове. Для обеспечения согласованности при работе с памятью сборщик мусора, реализующий алгоритм mark-sweep, не должен работать параллельно или конкурентно с основной программой. Такой режим работы сборщика называют «**остановкой мира**» («stop the world»). [15]

Длительно работающие приложения, управляемые непемещающими сборщиками мусора (например, работающими по алгоритму mark-sweep), могут фрагментировать кучу, что отрицательно скажется на производительности приложений. Для устранения внешней фрагментации предлагается две стратегии. Первая, которой придерживается алгоритм mark-compact, заключается в уплотнении используемых объектов кучи, вторая — в перемещении объектов из одной области памяти в другую. Основное преимущество уплотнения кучи заключается в том, что она обеспечивает относительно быстрое последовательное распределение, просто проверяя ограничение кучи и находя свободный указатель, соответствующий запросу на выделение. [15]

Алгоритм **mark-compact** работает в несколько этапов. Первая фаза — фаза разметки. Затем дальнейшие этапы уплотнения сжимают используемые данные путем перемещения объектов и обновления значений указателей всех ссылок на объекты, которые были перемещены. Количество обходов кучи, порядок, в котором они выполняются, и способ перемещения объектов могут зависеть от реализации. Порядок уплотнения влияет на местоположение данных.

Сборщик может перемещать объекты в куче одним из трех способов. [15]

1.2.4. Подсчёт ссылок

Подсчёт ссылок предполагает отслеживание количества указателей на определённые области памяти из других областей. Он используется в качестве основы для некоторых методов автоматической переработки, которые не требуют отслеживания (трассировки). [12]

Системы подсчёта ссылок выполняют автоматическое управление памятью, сохраняя в каждом объекте, обычно в заголовке, число ссылок на данный объект. Объекты, на которые нет ссылок (счётчик ссылок равен нулю), не могут быть доступны вызывающей стороне. Следовательно, они не используются и могут быть переработаны. [6]

Преимущества подсчёта ссылок:

- накладные расходы по управлению памятью распределяются по всему времени работы программы; [15]
- устойчивость к высоким нагрузкам, так как потенциально подсчёт ссылок может переработать объект как только он перестаёт использоваться; [15]
- масштабируемость по объёму кучи, так как накладные расходы, как правило, зависят только от количества выполняемых операций с указателями на объекты, а не от объема хранимых данных; [15]
- может быть реализован без помощи среды выполнения языка и без её ведома. [15]

Недостатки подсчёта ссылок:

- требуются накладные расходы на операции чтения и записи для управления числом ссылок;
- операции с числом ссылок на объекты должны быть атомарными; [15]
- случай циклических ссылок требует отдельного рассмотрения; [15]
- подсчёт ссылок может вызывать паузы, например при освобождении ссылочных структур данных. [15]

По сравнению со сборкой мусора алгоритмы управления памятью, основанные на подсчёте ссылок, отличаются детерминированностью, предсказуемостью и меньшей вычислительной сложностью. [13]

Подсчёт ссылок наиболее полезен в ситуациях, когда можно гарантировать отсутствие циклических ссылок и сравнительно редкие модификации ссылочных структур данных. Такие условия могут иметь место в некоторых типах структур баз данных и некоторых файловых системах. Подсчёт ссылок также может быть полезен, если важно, чтобы объекты утилизировались своевременно, например, в системах с жёсткими ограничениями памяти. [12]

1.3. Существующие решения в области автоматического управления памятью

1.3.1. Python

Python [16] — интерпретируемый язык программирования с сильной динамической типизацией. Официально язык был представлен в 1991 году. Существует множество реализаций интерпретатора языка Python [17] [18] [19]. Среди них канонической считается реализация CPython, разработанная на языке C и официально представленная в 1994 году. [20]

1.3.1.1. Структура объекта

Все объекты программы на языке Python размещаются в **приватной куче** (private heap), управление которой обеспечивается встроенным в интерпретатор диспетчером памяти. [21]

Основной алгоритм сборки мусора, используемый CPython, — подсчёт ссылок. Для его реализации в структуре объекта Python предусмотрены поля, хранящие число ссылок и данные о типе объекта (**PyObject_HEAD**), а также указатели на элементы двусвязного списка объектов, отслеживаемых сборщиком мусора (**PyGC_Head**). Структура объекта Python представлена на рисунке 1.1. [22]

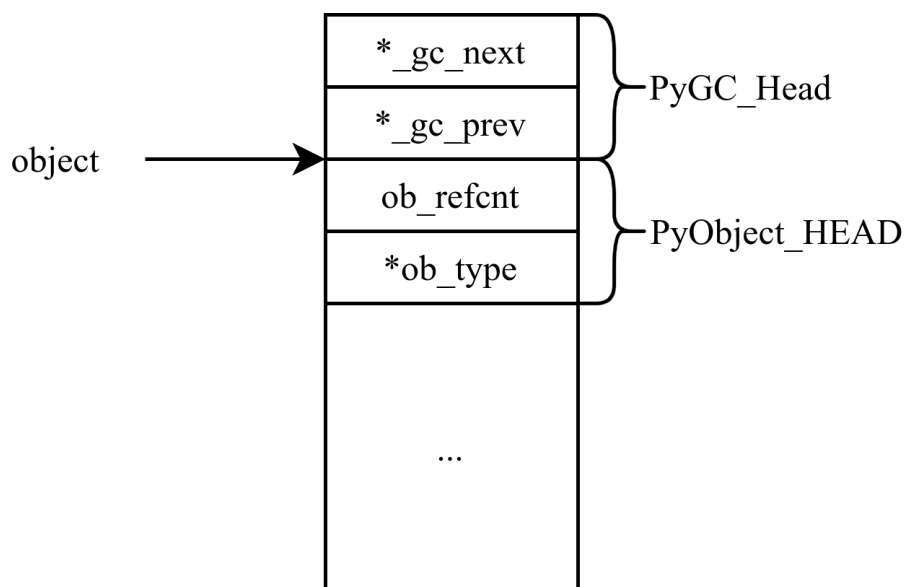


Рис. 1.1 – Структура объекта Python

Когда необходима дополнительная информация, связанная со сборщиком мусора, к полям `PyGC_Head` можно получить доступ с помощью адресной арифметики и приведения типа исходного объекта. [22]

Двусвязные списки используются по той причине, что они эффективно поддерживают операции, наиболее часто используемые при выполнении алгоритма сбора мусора, такие как перемещение объекта из одного раздела в другой, добавление нового объекта, полное удаление объекта, а также разбиение и объединение списков. [22]

1.3.1.2. Сборка мусора

Основная проблема подсчёта ссылок заключается в том, что он не обрабатывает циклические ссылки. Для её решения используется отдельный сборщик мусора, который занимается только очисткой объектов-контейнеров, то есть объектов, которые могут содержать ссылки на другие объекты. [22]

Алгоритм сборки мусора, который обрабатывает циклические ссылки, состоит в обходе двусвязного списка объектов и использовании метода пробного удаления [15]: объекты, счётчик ссылок на которые не превышает 1, помещаются в список условно недостижимых. Объекты, которые остались в этом списке после завершения обхода графа объектов, могут быть освобождены. [22]

Стоит отметить, что такая сборка мусора выполняется конкурентно с основной программой в одном выделенном потоке сборщика, не распределяя работу по сборке на несколько потоков приложения. [23]

1.3.2. Java

Java [24] — компилируемый язык программирования с сильной статической типизацией. Официально был представлен в 1995 году компанией Sun Microsystems.

Среда выполнения языка Java (Java Runtime Environment, JRE) состоит из **виртуальной машины Java** (Java Virtual Machine, JVM), основных классов и вспомогательных библиотек платформы Java. Рассматриваемый язык является кроссплатформенным: приложения, разработанные на Java, компилируются в **байт-код**, который сохраняется в файлах классов и запускается в JVM. Поэтому программы на языке Java можно запустить на всех аппаратных платформах и операционных системах, для которых реализована JVM. [24]

Существует множество реализаций JVM. [25] [26] [27] Среди них наиболее распространённой является HotSpot JVM, представленная в 1999 году компанией Sun Microsystems. [28]

1.3.2.1. Управление памятью

Частью JVM является **менеджер хранилища** (storage manager), который отвечает за управление жизненным циклом объектов Java: выделение новых объектов, сбор недоступных объектов и отправку уведомлений о недоступности по запросу. [29] Из соображений масштабируемости, каждый поток JVM имеет собственную область выделения памяти — **буфер выделения потока** (Thread-Local Allocation Buffer, TLAB). Каждый поток может выделять ресурсы из своей TLAB без координации с другими потоками, за исключением случаев, когда ему требуется новая TLAB. [29]

Было замечено, что большинство программ на языке Java следует **гипотезе поколений** [15], согласно которой в большинстве случаев «молодые» объекты (которые создаются позже), с гораздо большей вероятностью перестают использоваться раньше, чем «старые» объекты.. Чтобы учесть это свойство программ при распределении памяти, объекты Java разделяются на три поколения: молодое (young), старое (old) и постоянное (permanent). Управление поколениями может осуществляться по различным алгоритмам. Предполагается, что в молодом поколении будет больше объектов, чем в старом, а также будет выше плотность недоступных объектов. [29]

Молодое поколение должно поддерживать быстрое распределение, при этом ожидается, что большинство этих объектов относительно быстро станет недоступным. При сборе мусора в молодом поколении выявляются все достижимые объекты и копируются в старое поколение. Далее объекты, оставшиеся в молодом поколении, освобождаются. [29]

Объекты в **старом поколении** могут анализироваться сборщиком мусора реже, чем объекты в молодом поколении. Также сборщик мусора в зависимости от выбранной стратегии выполнять копирование и уплотнение объектов. [29] [30]

Помимо объектов, созданных программой на языке Java, существуют объекты, созданные и используемые JVM. Чтобы не путать их с объектами выполняемой программы, такие объекты выделяют в **постоянное поколение**. Данное название носит исторический характер и на самом деле объекты в нём не существуют на протяжении всего времени выполнения программы. [29] Начиная с Java 8 постоянное поколение было заменено **метапространством** (MetaSpace), предназначенным для хранения классов и метаданных JVM. [31]

1.3.2.2. Сборка мусора

HotSpot JVM предоставляет несколько сборщиков мусора, оптимизированных для различных типов приложений. Далее будут рассмотрены сборщики мусора, доступные в Java 21. [32]

Serial Collector работает по алгоритму mark-compact (см. п. 1.2.3.) и использует один поток для выполнения основной и второстепенной работы по сбору мусора, что может повысить его эффективность за счёт отсутствия накладных расходов на взаимодействие параллельных потоков. [32]

Parallel Collector отличается от Serial Collector использованием нескольких потоков для сборки мусора, количество которых может регулироваться с помощью аргументов командной строки. [32] Во время работы Parallel Collector все потоки основной программы приостанавливаются («stop the world», см. п. 1.2.3.). [33] Данный сборщик мусора также называют Throughput Collector, так как он может использовать несколько процессоров для ускорения сборки мусора и повышения пропускной способности при работе с памятью, что повышает производительность приложений. [24]

Garbage-First Collector (также G1) — это сборщик мусора, работающий

по алгоритму mark-sweep (см. п. 1.2.3.), предназначенный для многопроцессорных машин с большим объемом памяти. Он с высокой вероятностью соответствует целевому времени паузы для сборки мусора, обеспечивая при этом относительно высокую пропускную способность. Операции над всей кучей, такие как глобальная разметка объектов, выполняются конкурентно с потоками приложения. Это предотвращает паузы в работе приложений, пропорциональные размеру кучи и используемых данных. Начиная с версии Java 9, Garbage-First Collector выбирается по умолчанию в большинстве конфигураций оборудования и операционной системы или может быть явно включён с помощью параметров командной строки. Garbage-First Collector достигает высокой производительности при снижении длительности пауз за счет применения следующих методов. [34]

1. В отличие от Parallel Collector работа потоков приостанавливается только на некоторые этапы сборки мусора: в начале при разметке объектов из корневого набора (см. п.1.2.3.) и в конце при обнаружении изменений, произошедших за время сборки мусора. [33]
2. Куча разделяется на набор непрерывных областей одинакового размера, называемых регионами. После обнаружения используемых объектов в куче Garbage-First Collector определяет, какие регионы «почти пусты» (mostly empty). Именно в этих регионах в первую очередь будет выполняться сбор мусора и уплотнение неиспользуемых объектов, чтобы как можно раньше дать аллокатору непрерывные свободные области памяти наибольшего размера.
3. Используется **модель прогнозирования паузы** (pause prediction model) для достижения заданного пользователем целевого времени паузы на сборку мусора и на его основе выбирает количество регионов для обработки.
4. В отличие от Parallel Collector используемые объекты выбираются для копирования и уплотнения не из всей кучи, а только из некоторых её областей. Эта операция выполняется параллельно на нескольких процессорах, чтобы уменьшить время паузы на сборку мусора и увеличить пропускную способность. Таким образом, при каждой сборке мусора G1 непрерывно

работает над уменьшением фрагментации кучи, работая в течение заданного пользователем времени паузы.

Z Garbage Collector (также ZGC) — это сборщик мусора, работающий по алгоритму mark-compact (см. п. 1.2.3.) и нацеленный на минимизацию пауз. ZGC работает конкурентно с основной программой, не останавливая выполнение потоков приложения более чем на одну миллисекунду, при этом жертвуя пропускной способностью. Он предназначен для приложений, которым требуется низкая задержка на сборку мусора. Время пауз не зависит от размера используемых данных в куче. ZGC оптимизирован для работы с кучей размером от нескольких сотен мегабайт до 16 Тб. [35]

1.3.3. Javascript

JavaScript [36] — интерпретируемый язык программирования со слабой динамической типизацией, представленный в 1995 году. Стандартом языка является ECMAScript. [37] Наиболее широкое применение JavaScript получил в качестве языка сценариев веб-страниц, но также используется и в других программных продуктах, таких как node.js [38] и Apache CouchDB [39].

Модель времени выполнения (runtime model) в JavaScript основана на **цикле событий** (event loop), который отвечает за сбор и обработку событий, а также за выполнение подзадач из очереди (queued sub-tasks). [40]

Изначально интерпретаторы JavaScript использовали сборку мусора на основе подсчёта ссылок, не обрабатывая циклические ссылки, что могло приводить к систематическим утечкам памяти. По этой причине, начиная с 2012 года, все современные веб-браузеры оснащаются сборщиками мусора, работающими исключительно по алгоритму mark-sweep (ССЫЛКА НА КНИГУ). Все усовершенствования алгоритма сборки мусора в интерпретаторах JavaScript (инкрементальная, конкурентная и параллельная сборка мусора, а также применение алгоритма поколений) за последние несколько лет представляют собой усовершенствования данного алгоритма, не являясь новыми алгоритмами сборки мусора, поскольку дальнейшее сужение понятия «объект более не используется» не представляется возможным. [41]

1.3.4. C#

C# [42] — компилируемый язык программирования с сильной статической типизацией. Официально был представлен в 2001 году компанией Microsoft.

Программы на языке C# предназначены для запуска на виртуальной среде выполнения .NET, работающей с **общезыковой средой выполнения** (Common Language Runtime, CLR) и набором библиотек классов. Среда CLR — это реализация **общезыковой инфраструктуры языка** (Common Language Infrastructure, CLI), являющейся международным стандартом компании Microsoft. [42] Основные функции среды CLR: [43] [44]

- сборка мусора;
- обеспечение безопасности памяти и типов данных;
- поддержка языков программирования высокого уровня: C#, F# и Visual Basic;
- изоляция программ.

1.3.4.1. Управление памятью

Сборщик мусора среды Common Language Runtime управляет выделением и освобождением памяти для приложения. При инициализации нового процесса среда выполнения резервирует для него непрерывную область адресного пространства, называемого **управляемой кучей** (managed heap), которая является частью кучи адресного пространства процесса и предназначена для размещения всех ссылочных типов данных. Выделение памяти из управляемой кучи происходит быстрее, чем неуправляемое выделение памяти, так как среда выполнения выделяет память путём изменения указателя на следующий объект в куче, что аналогично выделению памяти на стеке. Кроме того, последовательное хранение объектов позволяет ускорить доступ к ним. Также стоит отметить, что все потоки процесса выделяют память в одной куче. [45]

Для повышения производительности среда выполнения выделяет память для больших объектов (размером не менее 85 000 байт) в отдельной куче, называемой **кучей больших объектов** (Large Object heap, LOH). Также во избе-

жание перемещения больших объектов в памяти куча больших объектов, как правило, не подлежит уплотнению. [46]

1.3.4.2. Сборка мусора

Сборщик мусора определяет оптимальное время для выполнения сбора на основе выделяемых ресурсов. Перед началом сборки мусора все управляемые потоки приостанавливаются, за исключением потока, который инициировал сборку. Период времени, в течение которого сборщик мусора активен, называется его **задержкой** (delay). Сборщик мусора работает по алгоритму mark-compact (см. п. 1.2.3.), разделяя сбор мусора на три фазы. [47]

1. **Фаза разметки:** определение неиспользуемых объектов путём исследования корневого набора программы (см. п. ??), каждый элемент которого либо ссылается на объект в управляемой куче, либо имеет значение null. Корневой набор включает статические поля, локальные переменные и параметры в стеке потока, а также регистры процессора. Сборщик мусора имеет доступ к списку активных «корней», который поддерживается JIT-компилятором и средой выполнения. С помощью этого списка исследуется корневой набор и строится граф доступных объектов.
2. **Фаза перемещения:** корректировка указателей, чтобы после сжатия используемых данных элементы корневого набора ссылались на новое местоположение объектов.
3. **Фаза уплотнения:** освобождение объектов, которые не попали в граф на этапе разметки, считаются недоступными и освобождаются. На данном этапе сборщик мусора проверяет управляемую кучу в поисках областей памяти, занятых недоступными объектами. При обнаружении каждого недостижимого объекта он использует функцию копирования для уплотнения доступных объектов, тем самым освобождая области памяти, выделенные для недоступных объектов. Он также устанавливает указатель управляемой кучи за последним доступным объектом.

Сбор мусора гарантированно происходит при выполнении одного из следующих условий. [47]

1. В системе обнаружена нехватка физической памяти. Данная ситуация определяется либо по уведомлению от операционной системы, либо по приближении к пределу объема физической памяти, доступной на вычислительной машине.
2. Объем памяти, используемый объектами в управляемой куче, превысил допустимое пороговое значение, которое может корректироваться во время выполнения программы.
3. Сборщик мусора вызван явно с помощью метода GC.Collect.

Работа сборщика мусора основана на следующих соображениях. [47]

- Уплотнение для части управляемой кучи выполняется быстрее, чем для всей управляемой кучи.
- Новые объекты, как правило, обладают меньшим временем жизни по сравнению со старыми объектами.
- Новые объекты, как правило, связаны друг с другом и используются приложением примерно в одно и то же время.

Для повышения производительности сборщика мусора и снижения среднего времени паузы на сборку, управляемая куча разделена на три поколения, обрабатываемые отдельно. Поскольку сжать часть управляемой кучи быстрее, чем всю кучу, эта схема позволяет сборщику мусора освобождать память только в определенном поколении, а не во всей управляемой куче при каждом вызове. [47] [45]

Поколение 0 содержит самые новые и, как правило, недолговечные объекты, такие как локальные переменные. Сбор мусора происходит чаще всего именно в этом поколении, при этом зачастую освобождается достаточно памяти, чтобы приложение могло продолжать выделять память. Объекты, оставшиеся после сборки мусора поколения 0, переводятся в поколение 1.

Поколение 1 служит буфером между новыми и старыми объектами. Если сбор мусора в поколении 0 не освобождает достаточно памяти для того, чтобы приложение могло создать новый объект, сборщик мусора может выполнить сборку в поколении 1, а затем в поколении 2. Объекты, оставшиеся после сборки мусора поколения 1, переводятся в поколение 2.

Поколение 2 содержит старые объекты, такие как статические данные, используемые на протяжении всего времени выполнения программы. Объекты, попавшие в поколение 2, остаются в нём до тех пор, пока они не перестанут использоваться в программе. Также при сборе мусора в поколении 2 выполняется сбор объектов в куче больших объектов (иногда называется поколением 3).

1.3.5. Golang

Golang (также Go) [4] — компилируемый язык программирования с сильной статической типизацией, разработанный компанией Google. Официально язык был представлен в ноябре 2009 года.

В любой программе на языке Golang присутствует **среда выполнения** (runtime) [48], то есть код, который выполняется без ведома программиста. Среда выполнения состоит из следующих компонентов.

1. **Планировщик** (Sheduler) [49], отвечающий за управление работой горутин. **Горутина** — это легковесный поток, управляемый средой выполнения языка Golang. [50]
2. **Аллокатор памяти.**
3. **Сборщик мусора.**

1.3.5.1. Структура кучи

Куча состоит из набора областей, называемых **аренами** (Arena), размер которых составляет 64 Мб в 64-разрядной версии и 4 Мб в 32-разрядной (heapArenaBytes). Начальный адрес каждой арены также выровнен по размеру арены. С каждой ареной связан объект heapArena [51], который хранит метаданные для этой арены: битовую карту кучи для всех слов (word) в арене и карту span для всех страниц в ней. Объекты heapArena выделяются вне кучи. Структура mheap содержит **карту арен**, которая охватывает всё доступное адресное пространство программы, поэтому его можно рассматривать как набор арен. Аллокатор старается сохранять арены смежными, чтобы большие span (и, следовательно, большие объекты) могли занимать одновременно несколько арен. [52]

1.3.5.2. Сборка мусора

В языке Golang для сборки мусора используется алгоритм **Concurrent Mark-Sweep**, являющийся модификацией алгоритма mark-sweep (см. п. 1.2.3.), предназначенного для сборки мусора конкурентно с основной программой. Сборщик мусора запускается параллельно с потоками основной программы и позволяет нескольким потокам сборки мусора выполняться параллельно. Алгоритм сборки мусора не является уплотняющим и не использует поколения (см. п. 1.3.2.1.). [53]

Сборщик мусора языка Golang использует трехцветную абстракцию, в рамках которой граф объектов разбивается на черные (предположительно используемые) и белые (возможно, неиспользуемые) объекты. Изначально каждый объект белый. Когда объект впервые обнаруживается во время трассировки, он окрашивается в серый цвет. Когда объект был отсканирован и были идентифицированы его дочерние элементы, он закрашивается чёрным. [15]

Для конкурентной разметки и очистки используются **барьеры записи** (write barrier), позволяющий избежать ситуации, когда чёрные объекты указывают на белые. Такое может произойти, например, при перемещении указателя в программе до того, как сборщик мусора успел его разметить. В такой ситуации основная программа скрывает объект от сборщика мусора. [53] Для реализации барьера записи используются операции «затенения» (shade), перемещающие указатели программы. Барьер записи «затеняет» как записываемый указатель, так и записываемое значение при любой записи по указателю [54].

Ниже представлено описание основных шагов алгоритма сборки мусора. [53]

1. Фаза завершения очистки (sweep termination).

- 1.1. «Остановка мира» («stop the world», см. п. 1.2.3.), приводящая к тому, что все потоки достигают **безопасной точки** (GC safe-point).
- 1.2. Очистка всех мусорных mspan. Неочищенные mspan будут только в том случае, если цикл сборки мусора был запущен раньше ожидаемого времени.

2. Фаза разметки (mark).

- 2.1. Включение барьера записи, постановка в очередь заданий по разметке объектов из корневого набора (см. п. 1.2.3.) и включение ассистентов (assistants), выполняющих свою работу во время выделения памяти. Никакие объекты не могут быть просканированы до тех пор, пока все потоки не включат барьер записи, что достигается с помощью «остановки мира».
- 2.2. «Запуск мира» («start the world»). С этого момента работа сборщика мусора выполняется обработчиками разметки (mark workers), запущенными планировщиком, и ассистентами. Выделяемые объекты отмечаются чёрным цветом.
- 2.3. Разметка объектов из корневого набора, включающая в себя сканирование стеков всех горутин, «затенение» всех глобальных переменных и любых указателей кучи в структурах данных среды выполнения вне кучи. При сканировании стека горутин останавливается, «затеняются» все указатели, найденные в её стеке, а затем горутин продолжает выполняться.
- 2.4. Очистка рабочей очереди от серых объектов. Каждый серый объект сканируется и отмечается чёрным, «затеняются» все указатели, найденные в объекте, что, в свою очередь, может привести к добавлению этих указателей в рабочую очередь.
- 2.5. Выполнение алгоритма распределенного завершения (distributed termination algorithm), чтобы определить, когда закончится выполнение заданий разметки объектов из корневого набора или серых объектов.
3. Фаза завершения разметки (mark termination).
 - 3.1. «Остановка мира».
 - 3.2. Отключение обработчиков и ассистентов.
 - 3.3. Очистка, включающая сброс кешей mcache.
4. Фаза очистки (sweep).
 - 4.1. Отключение барьера записи.

4.2. «Запуск мира». С этого момента выделяемые объекты становятся белыми. Также при необходимости аллокатор может очищать `mspan` перед использованием.

4.3. Выполнение параллельной очистки в фоновом режиме и во время выделения памяти.

1.3.6. Классификация существующих решений

Ниже приведены критерии классификации алгоритмов распределения памяти.

1. Разделение объектов на поколения.

Для оптимизации сборки мусора и уменьшения времени пауз в языках программирования может использоваться алгоритм поколений (см. п. 1.3.2.1.). В таблице сравнения будут использоваться обозначения «+» и «-» для определения использования алгоритма поколений менеджером памяти рассматриваемого языка программирования.

2. Отсутствие хранения вспомогательных данных в объектах.

Для реализации алгоритмов распределения памяти может понадобиться хранение дополнительных данных в выделяемых объектах, что создаёт зависимость объёма памяти, необходимого сборщику мусора, от количества отслеживаемых объектов. В таблице сравнения для обозначения факта хранения вспомогательных данных в объектах будет использоваться «-», иначе — «+».

3. Использование конкурентной сборки мусора.

Для снижения времени пауз на сборку мусора менеджеры памяти некоторых языков программирования могут выполнять её конкурентно с потоками основной программы. В таблице сравнения для обозначения факта использования конкурентной сборки мусора будет использоваться «+», иначе — «-».

4. Использование параллельной сборки мусора.

Менеджеры памяти некоторых языков программирования могут выполнять сборку мусора параллельно в нескольких потоках приложения, задействуя больше вычислительных ресурсов процессора для её ускорения.

В таблице сравнения для обозначения факта использования параллельной сборки мусора будет использоваться «+», иначе — «-».

5. Отсутствие остановки потоков основной программы на весь цикл сборки мусора.

Некоторые алгоритмы сборки мусора оптимизированы таким образом, чтобы останавливать работу потоков основной программы только на некоторых этапах сбора для минимизации времени пауз. В таблице сравнения для обозначения этого факта будет использоваться «+», иначе — «-».

6. Количество остановок потоков основной программы за один цикл сборки мусора.

Описанные выше характеристики алгоритмов распределения памяти, как правило, декларируются в документациях к соответствующим языкам программирования. Результаты сравнения алгоритмов распределения памяти приведены в таблице 1.1.

Таблица 1.1 – Сравнение алгоритмов распределения памяти

Язык программирования	Сборщик мусора	Разделение объектов на поколения	Отсутствие хранения вспомогательных данных в объектах	Использование конкурентной сборки мусора	Использование параллельной сборки мусора	Отсутствие остановки потоков основной программы на весь цикл сборки мусора	Количество остановок потоков основной программы за один цикл сборки мусора
Python	По умолчанию	+	-	+	-	+	1
Java	Serial	+	+	+	-	-	1
	Parallel	+	+	+	+	-	1
	Garbage-First	+	+	+	+	+	2
	ZGC	+	+	+	+	+	1
JavaScript	По умолчанию	-	+	-	-	-	1
C#	По умолчанию	+	+	+	+	-	1
Golang	По умолчанию	-	+	+	+	+	2

Среди рассмотренных интерпретируемых языков программирования наиболее оптимизированным для различных сценариев использования можно считать менеджер памяти языка Python за счёт применения алгоритма поколений и подсчёта ссылок.

Среди рассмотренных компилируемых языков программирования наиболее универсальным и масштабируемым можно считать менеджер памяти языка Java за счёт предоставления пользователю возможности выбора сборщика мусора для выполнения каждой программы, а также оптимизации времени пауз на сборку мусора.

1.4. Классификация алгоритмов по требованиям к дополнительной памяти

Теоретическое исследование ресурсных характеристик компьютерных алгоритмов предполагает введение соответствующих классификаций, отражающих различные ресурсные требования алгоритма и его программной реализации. [55]

Далее будет рассмотрена классификация компьютерных алгоритмов, в основу которой положен требуемый алгоритмом объём дополнительной памяти в принятой модели вычислений. В данном случае рассматривается только объём оперативной памяти, так как затраты памяти на внешних носителях требуют отдельного рассмотрения и выходят за рамки работы. Пусть D — конкретный вход алгоритма A , причем $|D| = n$, где n — длина или некоторая мера длины входа. Под объёмом памяти, требуемым алгоритмом A для входа, заданного множеством D , будем понимать максимальный объём памяти модели вычислений, задействованный в ходе выполнения алгоритма. Функцию объёма памяти алгоритма для входа D будем обозначать через $V_A(D)$. Значение функции $V_A(D)$ есть целое положительное число. [55]

Ресурсные требования алгоритма к объёму памяти определяются памятью входа, выхода и дополнительной памятью, задействованной алгоритмом в ходе его выполнения. В соответствии с этим будем различать следующие компоненты функции объёма памяти: память входа $V_I(D)$, память выхода $V_R(D)$ и дополнительная память $V_t(D)$. По объёму памяти входа (выхода) алгоритмы можно разделить на потоковые, обрабатывающие входные (выходные) данные поэлементно, и непотоковые, хранящие входные (выходные) данные в памяти целиком. [55]

Поскольку ресурсные компоненты $V_I(D)$ и $V_R(D)$ во многом определяются особенностями задачи, а для различных алгоритмов ее решения, обладающих статическими входом и выходом, фактически совпадают, то именно компонент $V_t(D)$ различает алгоритмы по ресурсным затратам памяти в рамках статического или потокового типа. Очевидно, что для современных компьютеров ограничение решаемых задач по ресурсу оперативной памяти не столь существенно, как ограничения по временной эффективности. Тем не менее известная дилемма «память-время» приводит к тому, что попытка улучшения вре-

менных характеристик приводит к ощутимым затратам памяти, даже при современных объемах. Актуальные размерности современных сложных задач также приводят к тому, что ресурс памяти становится значимым в комплексной оценке ресурсной эффективности алгоритма. Однако при сравнении алгоритмов, относящихся или к статическому, или к потоковому типу, основную роль играет именно дополнительная память, затраты которой обусловлены спецификой данного алгоритма. В связи с этим предлагается следующая классификация алгоритмов, основанная на оценке дополнительной памяти. [55]

Класс V0 — алгоритмы с нулевой дополнительной памятью.

$$\forall n > 0, \forall D \in D_n \quad V_t(D) = 0. \quad (1.1)$$

Алгоритмы этого класса либо вообще не требуют дополнительной памяти, либо используют ресурсы памяти входа и/или выхода в качестве необходимой дополнительной памяти по мере обработки элементов входа или по мере заполнения памяти результата. [55]

Класс VC — алгоритмы с фиксированной дополнительной памятью.

$$\forall n > 0, \forall D \in D_n \quad V_t(D) = \text{const} \neq 0. \quad (1.2)$$

Алгоритмы класса *VC* используют постоянный, не зависящий от длины и особенностей входа и длины выхода, объём дополнительной оперативной памяти. В реальных алгоритмах — это, как правило, память для хранения счетчиков циклов, промежуточных результатов вычислений и указателей на структуры. [55]

Класс VL — алгоритмы, дополнительная память которых линейно зависит от длины входа. Введём в рассмотрение функцию дополнительной памяти в худшем случае для всех допустимых входов с мерой n , обозначив её через $V_t^{\wedge}(n)$:

$$V_t^{\wedge}(n) = \max_{D \in D_n} V_t(D). \quad (1.3)$$

В этих обозначениях алгоритм принадлежит к типу с линейной дополнительной памятью, если

$$V_t^{\wedge}(n) = \Theta(n), \quad (1.4)$$

где обозначение Θ есть стандартное обозначение класса функций в асимптотическом анализе. [55]

Класс VQ — алгоритмы, дополнительная память которых квадратично зависит от длины входа. Для этого типа алгоритмов объём дополнительной памяти в худшем случае для входов размерности n пропорционален по порядку квадрату меры размерности:

$$\hat{V}_t(n) = \Theta(n^2). \quad (1.5)$$

Примером может служить стандартный алгоритм умножения длинных целых чисел, заданных побитно массивами длины n . [55]

Класс VP — алгоритмы, дополнительная память которых имеет полиномиальную надквадратичную зависимость. Это алгоритмы, требующие ресурса дополнительной памяти, по порядку большего, чем квадрат меры длины входа, но полиномиально зависящего от этой меры:

$$\exists k > 2 : \hat{V}_t(n) = \Theta(n^k). \quad (1.6)$$

К этому классу относятся, например, рекурсивные алгоритмы, порождающие дерево рекурсии с оценкой глубины $\Theta(n^k)$, $k > 1$, и требующие хранения в каждой вершине дерева дополнительного массива, имеющего длину порядка $\Theta(n)$. [55]

Класс VE — алгоритмы, дополнительная память которых экспоненциально зависит от длины входа. Формально это тип алгоритмов, требующих, в худшем случае, ресурса дополнительной памяти, по порядку экспоненциально зависящего от меры длины входа:

$$\exists \lambda > 0 : \hat{V}_t(n) = \Theta(e^{\lambda n}). \quad (1.7)$$

Реально это алгоритмы, использующие дополнительные массивы или более сложные структуры данных, размер которых определяется значениями элементов входа. Характерным примером могут служить алгоритм сортировки методом индексов или табличные алгоритмы, реализующие метод динамического программирования. [55]

1.5. Постановка задачи

Алгоритмы автоматического управления памятью в каждом из языков программирования обладают различными недостатками, среди которых однопоточная или неконкурентная сборка мусора, необходимость остановки потоков основной программы для её выполнения, а также отсутствие разделения объектов на поколения. Поэтому разрабатываемый метод должен быть ориентирован на устранение недостатков существующих решений, возникающих при реализации алгоритмов, относящихся к классам VL, VQ, VP, VE, а также VC при условии, что дополнительная память для работы такого алгоритма будет выделяться в куче, а не на стеке (например, во избежание переполнения стека).

На рисунке 1.2 представлена диаграмма IDEF0 нулевого уровня решаемой задачи.

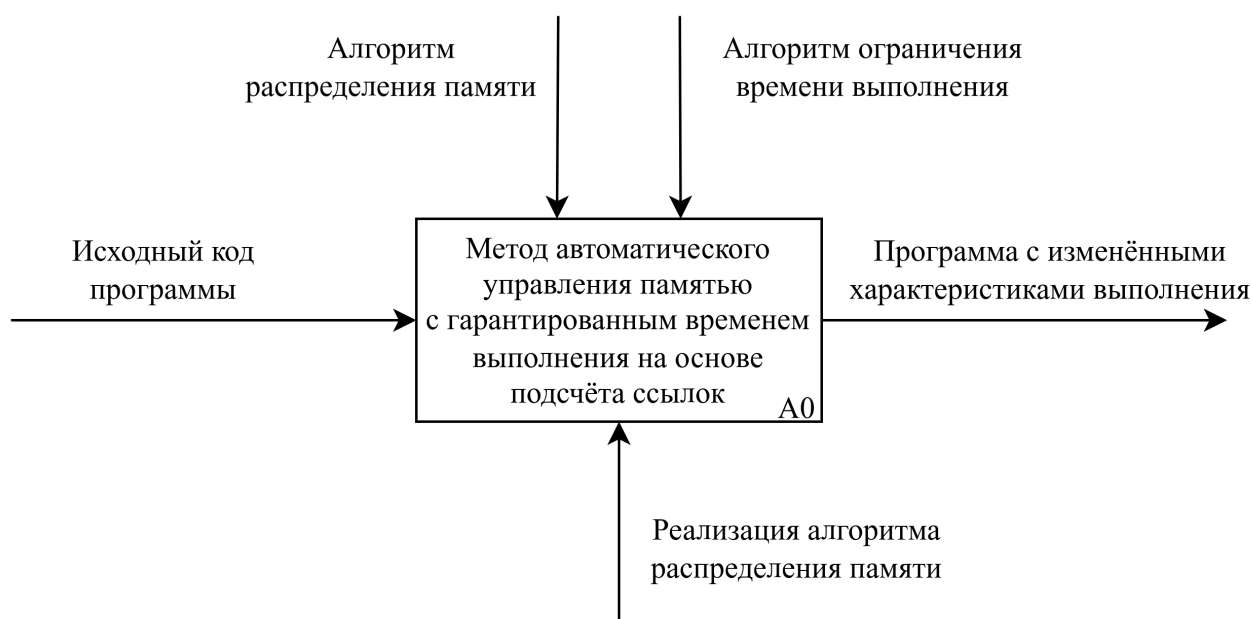


Рис. 1.2 – Диаграмма IDEF0 нулевого уровня

Под временем выполнения метода понимается время, которое основная программа затрачивает на вызовы реализации метода, то есть на выделение памяти и обращение к выделенным объектам. Гарантия времени выполнения метода подразумевает, что это время будет иметь некоторое ограничение, выражаемое асимптотической оценкой сложности алгоритмов выделения памяти и обращения к ней.

Подсчёт ссылок позволяет при освобождении памяти распределять накладные расходы по всему времени работы программы. При этом сбор циклических ссылок должен осуществляться конкурентно с основной программой, чтобы снизить накладные расходы на мутатор.

2 Конструкторский раздел

2.1. Общая схема метода

2.2. Выбор способа использования метода

2.3. Описание необходимых структур данных

2.4. Проектирование алгоритмов работы с памятью

2.4.1. Алгоритм выделения памяти

2.4.2. Алгоритм освобождения памяти

2.4.3. Алгоритм сбора мусора

2.4.4. Алгоритм сбора циклических ссылок

3 Технологический раздел

3.1. Выбор языка программирования

3.2. Компоненты программного обеспечения

3.2.1. Аллокатор

В том числе его инициализация

3.2.2. Поколение

3.2.3. Сборщик мусора

3.2.4. Анализатор объектов

3.2.5. Мониторинг работы менеджера памяти

3.3. Тестирование компонентов программного обеспечения

3.3.1. Аллокатор

3.3.2. Анализатор объектов

3.3.3. Сборщик мусора

4 Исследовательский раздел

4.1. Цель проводимых измерений

4.2. Описание проводимых измерений

4.3. Инструменты измерения времени работы программы

4.4. Технические характеристики оборудования

4.5. Результаты проведённых измерений

4.5.1. Одиночный запуск программы

4.5.2. Веб-сервер

4.5.2.1. Открытая линейная нагрузка

4.5.2.2. Открытая постоянная нагрузка

4.5.2.3. Закрытая нагрузка

Выводы из исследовательского раздела

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Таненбаум Э., Бос Х.* Современные операционные системы. — 4-е изд. — Питер, 2015.
2. ISO/IEC 9899:TC3 C standard [Электронный ресурс]. — Режим доступа, URL: <https://www.open-std.org/jtc1/sc22/wg14/www/docs/n1256.pdf> (дата обращения: 29.10.2023).
3. Standard C++ [Электронный ресурс]. — Режим доступа, URL: <https://isocpp.org/> (дата обращения: 29.10.2023).
4. The Go Programming Language documentation [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/> (дата обращения: 29.10.2023).
5. The Elixir programming language documentation [Электронный ресурс]. — Режим доступа, URL: <https://elixir-lang.org/> (дата обращения: 29.10.2023).
6. Memory Management Glossary [Электронный ресурс]. — Режим доступа, URL: <https://www.memorymanagement.org/glossary/> (дата обращения: 12.10.2023).
7. *Richardson O.* Reconsidering Custom Memory Allocation [Электронный ресурс]. — Режим доступа, URL: <https://www.cs.cornell.edu/courses/cs6120/2019fa/blog/custom-alloc/> (дата обращения: 12.10.2023).
8. Common Language Runtime (CLR) overview [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/standard/clr> (дата обращения: 12.10.2023).
9. A theory of memory models / V. Saraswat [и др.] // Proceedings of the 12th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming. — 2007. — DOI: 10.1145/1229428.1229469.
10. Memory Management Overview [Электронный ресурс]. — Режим доступа, URL: <https://www.memorymanagement.org/mmref/begin.html> (дата обращения: 15.10.2023).
11. Allocation techniques [Электронный ресурс]. — Режим доступа, URL: <https://www.memorymanagement.org/mmref/alloc.html#mmref-alloc> (дата обращения: 12.10.2023).

12. Recycling techniques [Электронный ресурс]. — Режим доступа, URL: <https://www.memorymanagement.org/mmref/recycle.html> (дата обращения: 15.10.2023).
13. *Sinha O., Maitland O.* A Unified Theory of Garbage Collection [Электронный ресурс]. — Режим доступа, URL: <https://www.cs.cornell.edu/courses/cs6120/2019fa/blog/custom-alloc/> (дата обращения: 12.10.2023).
14. *Blackburn S. M., McKinley K. S.* Ulterior Reference Counting: Fast Garbage Collection without a Long Wait // Proceedings of the 2003 ACM SIGPLAN Conference on Object-Oriented Programming Systems, Languages and Applications, OOPSLA 2003. — 2003. — DOI: 10.1145/949305.949336.
15. *Jones R., Hosking A., Moss E.* THE GARBAGE COLLECTION HANDBOOK. The Art of Automatic Memory Management. — CRC Press, 2012. — ISBN 978-1-4200-8279-1.
16. Python Programming Language documentation [Электронный ресурс]. — Режим доступа, URL: <https://www.python.org/> (дата обращения: 10.11.2023).
17. Jython documentation [Электронный ресурс]. — Режим доступа, URL: <https://www.jython.org/> (дата обращения: 10.11.2023).
18. IronPython: the Python programming language for .NET [Электронный ресурс]. — Режим доступа, URL: <https://ironpython.net/> (дата обращения: 10.11.2023).
19. PyPy documentation [Электронный ресурс]. — Режим доступа, URL: <https://www.python.org/> (дата обращения: 10.11.2023).
20. CPython definition [Электронный ресурс]. — Режим доступа, URL: <https://docs.python.org/3/glossary.html#term-CPython> (дата обращения: 10.11.2023).
21. Memory Management — Python 3.12.0 documentation [Электронный ресурс]. — Режим доступа, URL: <https://docs.python.org/3/c-api/memory.html> (дата обращения: 10.11.2023).
22. Garbage collector design [Электронный ресурс]. — Режим доступа, URL: <https://devguide.python.org/internals/garbage-collector/> (дата обращения: 10.11.2023).

23. PEP 556 — Threaded garbage collection [Электронный ресурс]. — Режим доступа, URL: <https://peps.python.org/pep-0556/> (дата обращения: 26.11.2023).
24. Java Garbage Collection Basics [Электронный ресурс]. — Режим доступа, URL: <https://www.oracle.com/webfolder/technetwork/tutorials/obe/java/gc01/index.html> (дата обращения: 16.11.2023).
25. Open J9 documentation [Электронный ресурс]. — Режим доступа, URL: <https://eclipse.dev/openj9/> (дата обращения: 16.11.2023).
26. Codename One: Cross-Platform App Development with Java [Электронный ресурс]. — Режим доступа, URL: <https://www.codenameone.com/> (дата обращения: 16.11.2023).
27. GraalVM documentation [Электронный ресурс]. — Режим доступа, URL: <https://www.graalvm.org/> (дата обращения: 16.11.2023).
28. The HotSpot Group [Электронный ресурс]. — Режим доступа, URL: <https://openjdk.org/groups/hotspot/> (дата обращения: 16.11.2023).
29. HotSpot VM Storage Management [Электронный ресурс]. — Режим доступа, URL: <https://openjdk.org/groups/hotspot/docs/StorageManagement.html> (дата обращения: 16.11.2023).
30. Memory Management in the Java HotSpot Virtual Machine [Электронный ресурс]. — Режим доступа, URL: <https://www.oracle.com/technetwork/java/javase/memorymanagement-whitepaper-150215.pdf> (дата обращения: 16.11.2023).
31. Troubleshooting Memory Issues in Java Applications [Электронный ресурс]. — Режим доступа, URL: https://www.oracle.com/webfolder/technetwork/tutorials/mooc/JVM_Troubleshooting/week1/lesson1.pdf (дата обращения: 16.11.2023).
32. Java SE 21. HotSpot Virtual Machine Garbage Collection Tuning Guide. Available Collectors [Электронный ресурс]. — Режим доступа, URL: <https://docs.oracle.com/en/java/javase/21/gctuning/available-collectors.html> (дата обращения: 16.11.2023).

33. Oracle JRockit Online Documentation Library Release 4.0. Understanding Memory Management [Электронный ресурс]. — Режим доступа, URL: https://docs.oracle.com/cd/E15289_01/JRSDK/garbage_collect.htm (дата обращения: 16.11.2023).
34. The Garbage First Garbage Collector [Электронный ресурс]. — Режим доступа, URL: <https://www.oracle.com/java/technologies/javase/hotspot-garbage-collection.html> (дата обращения: 16.11.2023).
35. The Z Garbage Collector (ZGC) [Электронный ресурс]. — Режим доступа, URL: <https://wiki.openjdk.org/display/zgc/Main> (дата обращения: 16.11.2023).
36. JavaScript documentation [Электронный ресурс]. — Режим доступа, URL: <https://developer.mozilla.org/en-US/docs/Web/JavaScript> (дата обращения: 12.11.2023).
37. ECMAScript 2023 language specification [Электронный ресурс]. — Режим доступа, URL: <https://ecma-international.org/publications-and-standards/standards/ecma-262/> (дата обращения: 12.11.2023).
38. Node.js documentation [Электронный ресурс]. — Режим доступа, URL: <https://nodejs.org> (дата обращения: 12.11.2023).
39. Apache CouchDB documentation [Электронный ресурс]. — Режим доступа, URL: <https://couchdb.apache.org/> (дата обращения: 12.11.2023).
40. The event loop [Электронный ресурс]. — Режим доступа, URL: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Event_loop (дата обращения: 12.11.2023).
41. JavaScript memory management [Электронный ресурс]. — Режим доступа, URL: <https://developer.mozilla.org/en-US/docs/Web/JavaScript> (дата обращения: 12.11.2023).
42. A tour of the C# language [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/csharp/tour-of-csharp/> (дата обращения: 17.11.2023).
43. What is .NET? Introduction and overview [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/core/introduction> (дата обращения: 17.11.2023).

44. Introduction to the Common Language Runtime (CLR) [Электронный ресурс]. — Режим доступа, URL: <https://github.com/dotnet/runtime/blob/main/docs/design/coreclr/botr/intro-to-clr.md> (дата обращения: 17.11.2023).
45. Automatic Memory Management [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/standard/automatic-memory-management> (дата обращения: 17.11.2023).
46. The large object heap on Windows systems [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/standard/garbage-collection/large-object-heap> (дата обращения: 17.11.2023).
47. Fundamentals of garbage collection [Электронный ресурс]. — Режим доступа, URL: <https://learn.microsoft.com/en-us/dotnet/standard/garbage-collection/fundamentals> (дата обращения: 17.11.2023).
48. runtime package (Golang) [Электронный ресурс]. — Режим доступа, URL: <https://pkg.go.dev/runtime> (дата обращения: 4.11.2023).
49. Golang source file src/runtime/proc.go [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/src/runtime/proc.go> (дата обращения: 4.11.2023).
50. A Tour of Go. Goroutines [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/tour/concurrency/1> (дата обращения: 4.11.2023).
51. Golang source file src/runtime/mheap.go [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/src/runtime/mheap.go> (дата обращения: 4.11.2023).
52. Golang source file src/runtime/malloc.go [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/src/runtime/malloc.go> (дата обращения: 4.11.2023).
53. Golang source file src/runtime/mgc.go [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/src/runtime/mgc.go> (дата обращения: 5.11.2023).
54. Golang source file src/runtime/mbarrier.go [Электронный ресурс]. — Режим доступа, URL: <https://go.dev/src/runtime/mbarrier.go> (дата обращения: 4.11.2023).
55. *Ульянов М. В.* Ресурсно-эффективные компьютерные алгоритмы. Разработка и анализ. — М.: ФИЗМАТЛИТ, 2008. — ISBN 978-5-9221-0950-5.

ПРИЛОЖЕНИЕ А

Презентация.