

# КОНФИГУРАЦИОННОЕ УПРАВЛЕНИЕ

Учебное пособие

П.Н. Советов

Москва – 2021

# Оглавление

<b>1. Введение</b>	<b>1</b>
1.1. Что такое конфигурационное управление . . . . .	1
1.2. Формальные определения . . . . .	1
1.3. Тематика книги . . . . .	3
<b>2. Работа в командной строке</b>	<b>4</b>
2.1. Командный интерпретатор . . . . .	6
2.2. Инструменты командной строки . . . . .	17
<b>3. Менеджеры пакетов</b>	<b>20</b>
3.1. Нумерация версий ПО . . . . .	20
3.2. Семантическая нумерация версий . . . . .	21
3.3. Управление пакетами . . . . .	21
3.4. Менеджер пакетов <code>apk</code> . . . . .	22
3.5. Менеджер пакетов <code>apt</code> . . . . .	23
3.6. Задача разрешения зависимостей пакетов . . . . .	25
<b>4. Конфигурационные языки</b>	<b>27</b>
4.1. Инфраструктура как код . . . . .	27
4.2. Формальные языки и грамматики . . . . .	28
4.3. Компьютерные языки . . . . .	31
4.4. Простые форматы описания конфигурации . . . . .	32
4.5. Языки общего назначения как конфигурационные . . . . .	38
4.6. Программируемые конфигурационные языки . . . . .	39
<b>5. Системы автоматизации сборки</b>	<b>41</b>
5.1. Виды систем сборки . . . . .	41
5.2. Топологическая сортировка . . . . .	43
5.3. Система сборки <code>Make</code> . . . . .	45

<b>6. Системы контроля версий</b>	<b>49</b>
<b>7. Документация как код</b>	<b>50</b>
7.1. Языки разметки . . . . .	50
7.2. Грамотное программирование . . . . .	53
7.3. Markdown и Pandoc . . . . .	55
7.4. Языки описания диаграмм . . . . .	56
7.5. Генераторы документации на основе исходных текстов . . . . .	59
<b>8. Вопросы виртуализации</b>	<b>61</b>
8.1. Что такое виртуализация . . . . .	61
8.2. Языковые виртуальные машины . . . . .	62
8.3. Виртуализация вычислительной системы . . . . .	65
8.4. Виртуализация приложения . . . . .	68
8.5. Виртуализация уровня ОС . . . . .	68
<b>9. Литература</b>	<b>70</b>



# 1. Введение

## 1.1. Что такое конфигурационное управление

Задача управления конфигурацией (configuration management) некоторой системы является типичной для инженерной деятельности. Под конфигурацией понимается состав элементов системы и взаимное их расположение. Конфигурацией можно управлять, отслеживая ее состояние и контролируя целостность изменений конфигурации, а также фиксируя эти изменения в документации.

Можно заметить, что конфигурационное управление в описанном виде представляется достаточно рутинной работой. К счастью, инструменты и подходы, разработанные для конфигурационного управления ПО (software configuration management), позволили автоматизировать многие задачи. Это, в частности, касается популярного инструмента git, который сегодня используется не только программистами, но даже некоторыми художниками и писателями для управления конфигурацией своих творений. Далее под конфигурационным управлением будет пониматься именно конфигурационное управление ПО.

Конфигурационное управление является частью программной инженерии, поэтому к нему применима следующая цитата:

«Программная инженерия – это то, что происходит с программированием при добавлении времени и других программистов»

(Russ Cox).

## 1.2. Формальные определения

Рассмотрим теперь несколько формальных определений конфигурационного управления.

В ГОСТ Р ИСО/МЭК 12207-2010 определены следующие термины:

**Базовая линия** (baseline) – спецификация или продукт, которые были официально рассмотрены и согласованы с тем, чтобы впоследствии служить основой для дальнейшего развития, и которые могут быть изменены только посредством официальных и контролируемых процедур изменения

**Составная часть конфигурации** (configuration item) – объект в пределах конфигурации, который удовлетворяет некоторой функции целевого применения и может быть однозначно идентифицирован в данный момент

С использованием этих терминов определена цель конфигурационного управления (менеджмента конфигурации):

Цель процесса менеджмента конфигурации состоит в установлении и поддержании целостности всех идентифицированных выходных результатов проекта или процесса обеспечения доступа к ним любой заинтересованной стороны.

В результате успешного осуществления процесса менеджмента конфигурации:

1. определяется стратегия менеджмента конфигурации;
2. определяются составные части, нуждающиеся в менеджменте конфигурации;
3. устанавливается базовая линия конфигурации;
4. осуществляется управление изменениями в составных частях, находящихся под менеджментом конфигурации;
5. осуществляется управление конфигурацией составных частей, входящих в выпуск;
6. статус составных частей, на которые распространяется менеджмент конфигурации, становится доступным на протяжении всего жизненного цикла.

Во введении к стандарту IEEE 828-2012 конфигурационное управление в системной и программной инженерии определено, как специальная дисциплина в рамках более крупной дисциплины конфигурационного управления. Целями конфигурационного управления является:

1. идентифицировать и задокументировать функциональные и физические характеристики любого продукта, компонента, результата или услуги;
2. управлять любыми изменениями этих характеристик;
3. вести записи и сообщать о каждом изменении и статусе его реализации;

4. поддерживать аудит продуктов, результатов, услуг или компонентов для проверки соответствия требованиям.

## 1.3. Тематика книги

В этой книге конфигурационное управление трактуется более широко, чем в приведенных выше формальных определениях. Тематика книги в некоторой степени пересекаются с заслуживающими внимания материалами из [1] и [2].

Рассматриваемые далее темы:

1. командная строка;
2. менеджеры пакетов;
3. конфигурационные языки;
4. системы автоматизации сборки;
5. системы контроля версий;
6. документация как код;
7. вопросы виртуализации.

Часто можно наблюдать таких «сапожников без сапог» – программистов, которые решают задачи конечных пользователей, но не занимаются автоматизацией собственных рутинных задач. Поэтому выбор тем книги обусловлен общей целью – стремлением к автоматизации процессов, связанных с разработкой ПО.

Акцент на сиюминутных технологиях и инструментах может привести к чрезвычайно быстрому устареванию материала. По этой причине основное внимание в книге уделено общим подходам, алгоритмам и использованию проверенных временем инструментов с открытым кодом.

**Благодарности.**

## 2. Работа в командной строке

Командная строка на экране монитора – имитация работы с телетайпом. Телетайп, в свою очередь, является электромеханической печатной машиной, которую можно подключить к компьютеру. Пользователь набирает текст, который печатается на рулоне бумаги. Компьютер печатает пользователю свои ответы.

Удивительно, но такой, казалось бы, устаревший способ общения с компьютером все еще активно используется, пусть и с более современными средствами ввода-вывода. Более того, многие задачи очень трудно решить без командной строки! Это касается, в частности, работы с системой контроля версий Git, с системой автоматизации сборки Make, с системой контейнеризации Docker и многими другими популярными сегодня программами. Командная строка в духе UNIX имеется в MacOS, Linux и Windows (WSL, Powershell). Стоит вспомнить и многочисленные фильмы о «хакерах» – если герой фильма решает за компьютером какие-то нетривиальные задачи, то, обычно, зрителю демонстрируется именно командная строка.

Командная строка, как ни странно, хорошо знакома и любителям старых текстовых игр. В этих играх для совершения какого-либо действия необходимо набрать с клавиатуры соответствующую команду в духе `go north`, `read book` или `take apple`.

Вот как выглядит пример диалога с пользователем в игре Zork (1979 г.):

```
West of House                                     Score: 0
↪ Moves: 4
ZORK
```

```
Welcome to ZORK.
Release 13 / Serial number 040826 / Inform v6.14 Library 6/7
```



West of House

This is an open field west of a white house, with a boarded front  
↪ door.

There is a small mailbox here.

A rubber mat saying 'Welcome to Zork!' lies by the door.

>open mailbox

You open the mailbox, revealing a small leaflet.

>take leaflet

Taken.

Операционная система UNIX была разработана в далеком 1969 году. UNIX изначально являлась операционной системой в первую очередь для разработчиков, которым удобнее всего автоматизировать свои действия с помощью командной строки. Сама по себе командная строка еще древнее UNIX.

Основная суть решений, принятых при использовании командной строки, сводится к положениям «философии UNIX» (Дуг Макилрой), которые можно выразить следующими пунктами:

- Пусть каждая программа решает хорошо одну задачу. Для новых задач создавайте новые программы, а не усложняйте старые новыми «возможностями».
- Предполагайте, что вывод каждой программы может стать входом другой, еще неизвестной программы. Не загромождайте вывод посторонней информацией. Избегайте строго выравненных столбчатых форматов и двоичных входных форматов. Не настаивайте на интерактивном вводе.
- Проектируйте и разрабатывайте ПО, даже операционные системы, таким образом, чтобы его можно было опробовать уже на ранних этапах, в идеале в течение недель. Не бойтесь выбрасывать неудачно реализованные части и пересоздавать их.
- Вместо ручного труда используйте инструменты для облегчения задач разработки, даже если придется отвлечься на создание этих инструментов, а сами инструменты могут впоследствии больше не понадобиться.

Далее будет рассматриваться современный вариант UNIX, популярная ОС Linux, разработанная в 1991 году Линусом Торвальдсом (в то время – студентом финского университета).

При работе с командной строкой необходимо учитывать структуру файловой системы. В Linux она имеет следующий вид:

```
localhost:~# tree -d -L 1 /
```

```
/
├── bin
├── dev
├── etc
├── home
├── lib
├── media
├── mnt
├── opt
├── proc
├── root
├── run
├── sbin
├── srv
├── sys
├── tmp
├── usr
└── var
```

Имя / является корнем (root) файловой системы. Внутри корня расположены следующие, наиболее значимые каталоги:

- bin. Исполняемые файлы системных утилит.
- dev. Устройства, к которым в Linux возможен доступ, как к файлам.
- etc. Системные конфигурационные файлы.
- home. Домашние каталоги пользователей.
- lib. Системные библиотеки.
- media. Подключаемые съемные диски: USB, CD-ROM и так далее.
- mnt. Подключаемые разделы.

## 2.1. Командный интерпретатор

За поддержку работы в командной строке отвечает специальная программа – интерпретатор оболочки ОС (shell). В случае Linux таким интерпретатором обычно является Bash, архитектура которого приведена на рис. 1. Под Bash далее будем понимать целое семейство работающих схожим образом интерпретаторов. Интерпретатор выполняет следующие основные действия

[3]:

1. Принимает строку от пользователя.
2. Разбирает эту строку и переводит во внутренний формат.
3. Осуществляет подстановку для различных специальных символов и имен.
4. Выполняет команду пользователя.
5. Выдает код выполнения.

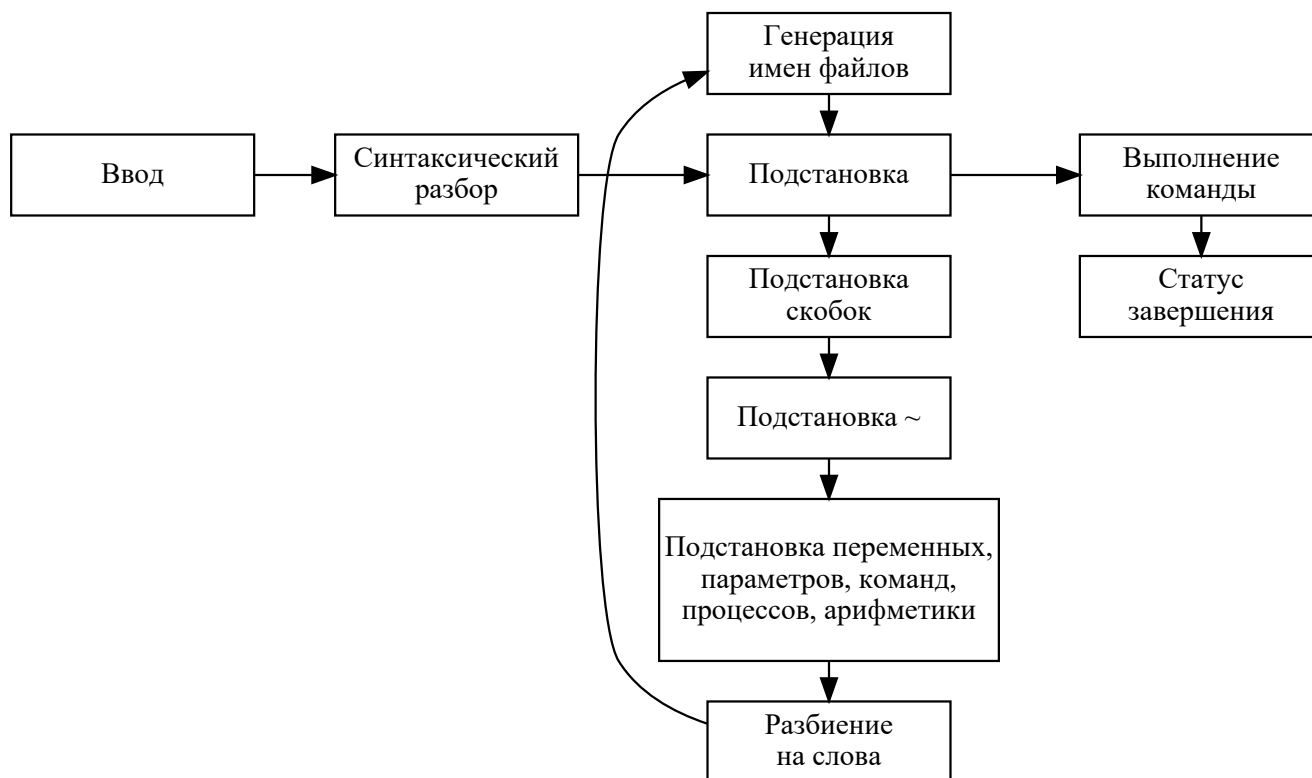


Рис. 1. Архитектура интерпретатора Bash

Ниже показан пример сеанса работы в Bash:

```
localhost:~# pwd
/root
localhost:~# ls -l
total 16
-rw-r--r--    1 root    root          114 Jul  5  2020 bench.py
-rw-r--r--    1 root    root           76 Jul  3  2020 hello.c
-rw-r--r--    1 root    root          22 Jun 26  2020 hello.js
-rw-r--r--    1 root    root         151 Jul  5  2020
↳  readme.txt
localhost:~# echo 'new file' > new_file.txt
```

```
localhost:~# cat new_file.txt
new file
localhost:~# mkdir new_dir
localhost:~# cp new_file.txt new_dir/
localhost:~# rm new_file.txt
localhost:~# ls -l
total 20
-rw-r--r--      1 root      root           114 Jul  5  2020 bench.py
-rw-r--r--      1 root      root            76 Jul  3  2020 hello.c
-rw-r--r--      1 root      root           22 Jun 26  2020 hello.js
drwxr-xr-x      2 root      root           66 Nov  4 17:16 new_dir
-rw-r--r--      1 root      root          151 Jul  5  2020
↳ readme.txt
localhost:~# ls -l new_dir/
total 4
-rw-r--r--      1 root      root            9 Nov  4 17:16
↳ new_file.txt
```

Обратите внимание на использование в приведенном сеансе команд, упрощенное описание которых дано ниже:

- `pwd`. Вывести имя текущего каталога.
- `ls`. Вывести содержимое каталога.
- `echo`. Вывести свой аргумент.
- `cat`. Вывести содержимое файла.
- `mkdir`. Создать каталог.
- `cp`. Скопировать файл.
- `rm`. Удалить файл.

Многие команды имеют ряд аргументов, это, в частности, касается `ls`, которая выше была вызвана с аргументом `-l`. Аргументы разделяются пробелами и имеют префикс `-`.

Узнать об аргументах, которые принимает команда, можно с помощью аргумента `--help`:

```
localhost:~# ls --help
BusyBox v1.31.1 () multi-call binary.
```

```
Usage: ls [-lAaCxdLHRFplinshrSXvctu] [-w WIDTH] [FILE]...
```

```
List directory contents
```

```
-l      One column output
-a      Include entries which start with .
-A      Like -a, but exclude . and ..
-x      List by lines
-d      List directory entries instead of contents
-L      Follow symlinks
-H      Follow symlinks on command line
-R      Recurse
-p      Append / to dir entries
-F      Append indicator (one of */=@|) to entries
-l      Long listing format
-i      List inode numbers
-n      List numeric UIDs and GIDs instead of names
-s      List allocated blocks
-lc     List ctime
-lu     List atime
--full-time    List full date and time
-h      Human readable sizes (1K 243M 2G)
--group-directories-first
-S      Sort by size
-X      Sort by extension
-v      Sort by version
-t      Sort by mtime
-tc     Sort by ctime
-tu     Sort by atime
-r      Reverse sort order
-w N    Format N columns wide
--color[={always,never,auto}]  Control coloring
```

Еще одним способом получить подробные сведения о конкретной команде является вызов вида `man <команда>`.

Без объяснений осталась строка `echo 'new file' > new_file.txt` в примере сеанса работы в командной строке выше. Здесь используется механизм перенаправления данных с помощью символов `<` (перенаправление ввода) и `>` (перенаправление вывода). В Linux имеется источник стандартного ввода `stdin` (код 0), а также два приемника стандартного вывода: `stdout` (код 1) и `stderr` (код 2, для ошибок). Организация ввода/вывода показана на рис. 2.

В примере ниже используется `stdout` и `stderr`:

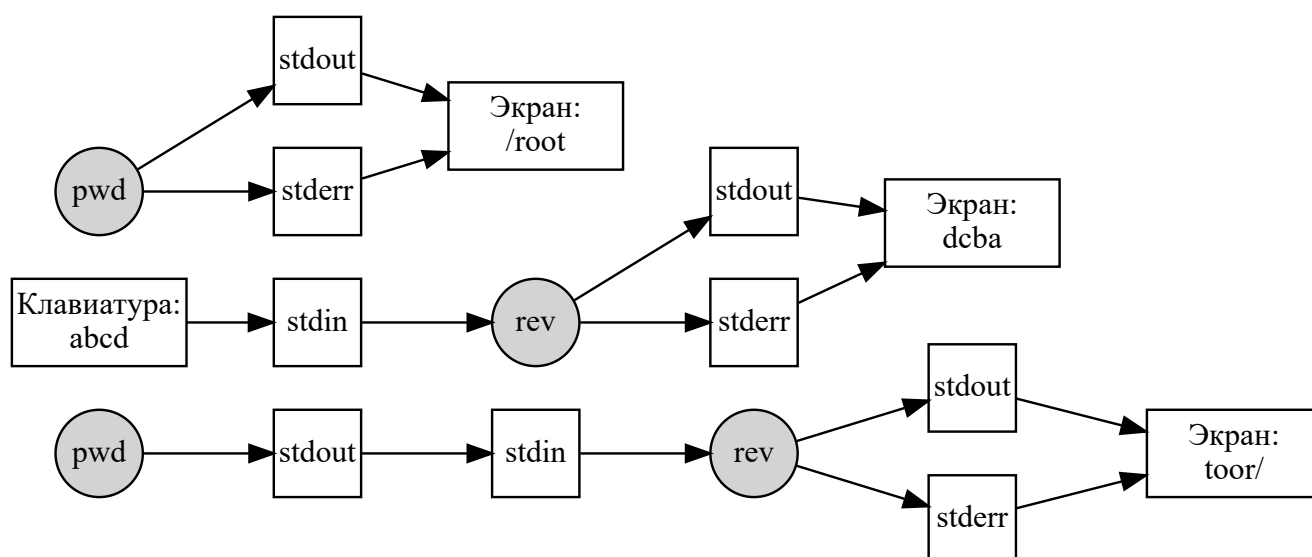


Рис. 2. Организация ввода/вывода

```
localhost:~# pwd
/root
localhost:~# pwd > pwd.txt
localhost:~# pwd --foo
sh: pwd: illegal option --
localhost:~# pwd --foo 2> err.txt
localhost:~# cat err.txt
sh: pwd: illegal option --
```

Обратите внимание на явное указание кода 2 при сохранении сообщения об ошибке в файл.

Перенаправление ввода/вывода превращается в очень мощную конструкцию при использовании такой организации команд, при которой вывод одной команды попадает на вход другой команды. Эта конструкция представляет собой конвейер и реализуется с помощью символа `|`, как показано в примере далее:

```
localhost:~# pwd > pwd.txt
localhost:~# rev --help
Usage: rev [options] [file ...]
```

Reverse lines characterwise.

Options:  
 -h, --help display this help

```
-V, --version  display version
```

For more details see `rev(1)`.

```
localhost:~# rev pwd.txt
toor/
localhost:~# pwd | rev
toor/
```

В Bash имеется удобный синтаксис для развертывания файловых путей (globbing). С помощью символов `*` (произвольная последовательность) и `?` (произвольный символ) реализуется подстановка имен файлов в духе регулярных выражений, как в примере ниже:

```
localhost:~# echo *
bench.py err.txt hello.c hello.js new_dir pwd.txt readme.txt rev
localhost:~# echo *.c
hello.c
localhost:~# echo p*
pwd.txt
localhost:~# echo *.*
bench.py hello.js
```

В Bash есть возможность задать переменные и, кроме того, имеется ряд уже определенных переменных. Обратите внимание на особенности создания переменных:

```
localhost:~# A = 42
sh: A: not found
localhost:~# A=42
localhost:~# A
sh: A: not found
localhost:~# echo $A
42
```

С помощью команды `set` можно, помимо прочего, узнать, какие переменные сейчас заданы для текущего пользователя:

```
localhost:~# set
A='42'
HISTFILE='/root/.ash_history'
HOME='/root'
HOSTNAME='localhost'
IFS=''
```

```
,  
LINENO=''  
OLDPWD='/'  
OPTIND='1'  
PAGER='less'  
PATH='/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin'  
PPID='1'  
PS1='\h:\w:$ '  
PS2='> '  
PS4='+ '  
PWD='/root'  
SHLVL='3'  
TERM='linux'  
TZ='UTC-03:00'  
_='/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin'  
script='/etc/profile.d/*.sh'
```

Особенно важна здесь переменная PATH, которая определяет те пути (разделенные с помощью :), где будет осуществляться поиск команд интерпретатором.

Linux является многопользовательской ОС и информация о зарегистрированных пользователях находится в конфигурационном файле /etc/passwd:

```
localhost:~# whoami  
root  
localhost:~# cat /etc/passwd  
root:x:0:0:root:/root:/bin/ash  
bin:x:1:1:bin:/bin:/sbin/nologin  
daemon:x:2:2:daemon:/sbin:/sbin/nologin  
adm:x:3:4:adm:/var/adm:/sbin/nologin  
lp:x:4:7:lp:/var/spool/lpd:/sbin/nologin  
sync:x:5:0:sync:/sbin:/bin/sync  
shutdown:x:6:0:shutdown:/sbin:/sbin/shutdown  
halt:x:7:0:halt:/sbin:/sbin/halt  
mail:x:8:12:mail:/var/mail:/sbin/nologin  
news:x:9:13:news:/usr/lib/news:/sbin/nologin  
uucp:x:10:14:uucp:/var/spool/uucppublic:/sbin/nologin  
operator:x:11:0:operator:/root:/sbin/nologin  
man:x:13:15:man:/usr/man:/sbin/nologin
```



```

postmaster:x:14:12:postmaster:/var/mail:/sbin/nologin
cron:x:16:16:cron:/var/spool/cron:/sbin/nologin
ftp:x:21:21::/var/lib/ftp:/sbin/nologin
sshd:x:22:22:sshd:/dev/null:/sbin/nologin
at:x:25:25:at:/var/spool/cron/atjobs:/sbin/nologin
squid:x:31:31:Squid:/var/cache/squid:/sbin/nologin
xfs:x:33:33:X Font Server:/etc/X11/fs:/sbin/nologin
games:x:35:35:games:/usr/games:/sbin/nologin
cyrus:x:85:12:./usr/cyrus:/sbin/nologin
vpopmail:x:89:89:./var/vpopmail:/sbin/nologin
ntp:x:123:123:NTP:/var/empty:/sbin/nologin
smmsp:x:209:209:smmsp:/var/spool/mqueue:/sbin/nologin
guest:x:405:100:guest:/dev/null:/sbin/nologin
nobody:x:65534:65534:nobody:./:/sbin/nologin
dhcp:x:100:101:dhcp:/var/lib/dhcp:/sbin/nologin
svn:x:101:102:svn:/var/svn:/sbin/nologin

```

Информация о каждом из пользователей занимает отдельную строку. Строка разделяется символом `:` на поля. Первое поле означает имя пользователя. В нашем случае это `root`. Последнее поле указывает путь к интерпретатору оболочки ОС. В нашем случае это компактный Bash-подобный интерпретатор `ash`.

Вспомним, как выглядит вывод команды `ls` в long-формате:

```

localhost:~# ls -l
total 20
-rw-r--r--    1 root    root          114 Jul  5  2020 bench.py
drwxr-xr-x    2 root    root           37 Nov  4 18:01 foo
-rw-r--r--    1 root    root           76 Jul  3  2020 hello.c
-rw-r--r--    1 root    root          22 Jun 26  2020 hello.js
-rw-r--r--    1 root    root          151 Jul  5  2020
↳  readme.txt

```

Первый столбец определяет права доступа и информацию о файле (-) или каталоге (d, как в случае с `foo`), закодированную в первом символе. Флаги доступа бывают следующих основных видов:

- `-`. Доступ запрещен.
- `r`. Имеется доступ на чтение.
- `w`. Имеется доступ на запись.
- `x`. Имеется доступ на исполнение (на вход в случае каталога).

Рассмотрим детали на примере с файлом `bench.py`, который имеет следующие права доступа:

```
-rw-r--r--  1 root    root          114 Jul  5  2020 bench.py
|[-][-][-]  [----]  [----]
| | | |      |      |
| | | |      |      +-----> 6. Группа
| | | |      +-----> 5. Владелец
| | | |
| | | +-----> 4. Права всех остальных
| | +-----> 3. Права группы
| +-----> 2. Права владельца
+-----> 1. Тип файла
```

При создании пользовательских команд необходимо указать права на исполнение, как показано в примере ниже:

```
localhost:~# echo "ls -l" > lsl
localhost:~# lsl
sh: lsl: not found
localhost:~# ./lsl
sh: ./lsl: Permission denied
localhost:~# chmod +x lsl
localhost:~# ./lsl
total 24
-rw-r--r--  1 root    root          114 Jul  5  2020 bench.py
drwxr-xr-x  2 root    root           37 Nov  4 18:01 foo
-rw-r--r--  1 root    root          76 Jul  3  2020 hello.c
-rw-r--r--  1 root    root          22 Jun 26  2020 hello.js
-rwxr-xr-x  1 root    root           6 Nov  4 18:44 lsl
-rw-r--r--  1 root    root          151 Jul  5  2020
↪  readme.txt
```

В Bash существует ряд специальных переменных, в частности:

- `$0`. Путь к запущенной программе.
- `$1`, `$2`, .... Аргументы программы.
- `$#`. Количество аргументов программы.
- `$@`. Список аргументов программы.
- `$?`. Значение результата выполнения программы (0 означает успешное выполнение).

Рассмотрим в качестве примера следующую программу `tests.sh`:

```
echo $0
echo $1 $2
echo $#
echo $@
```

Результат ее выполнения показан далее:

```
localhost:~# ./test.sh 1 2 3 4 5
./test.sh
1 2
5
1 2 3 4 5
localhost:~# echo $?
0
localhost:~# foo
sh: foo: not found
localhost:~# echo $?
127
```

Рассмотрим теперь более сложный пример пользовательской команды. Далее приведен код на языке Bash вычисления факториала:

```
#!/bin/sh
seq "$1" | xargs echo | tr " " "*" | bc
```

В первой строке указан интерпретатор, который будет использоваться для исполнения программы. По соглашению, такую строку необходимо всегда указывать первой в пользовательских скриптах. Далее используется ряд новых команд.

Команда `seq` генерирует последовательность чисел:

```
localhost:~# seq 5
1
2
3
4
5
```

Команда `xargs` форматирует список из стандартного ввода:

```
localhost:~# seq 5 | xargs
1 2 3 4 5
```

Команда `tr` осуществляет замену текстовых фрагментов:

```
localhost:~# seq 5 | xargs | tr " " "*"
1*2*3*4*5
```

Команда `bc` представляет собой калькулятор:

```
localhost:~# echo "2+2" | bc
4
```

Для вычислений в Bash можно также использовать скобки специального вида:

```
localhost:~# echo $((2 + 2))
4
```

Для получения результата команды в виде аргумента другой команды можно также использовать скобки специального вида:

```
localhost:~# echo "My folder is $(pwd)"
My folder is /root
```

В Bash имеются возможности полноценного языка программирования. Ниже приведен пример реализации факториала с использованием ветвлений и рекурсии:

```
#!/bin/sh
if [ "$1" -le 1 ] ; then
    echo 1
    return
fi
echo $(( $1 * $( ./fact.sh $(( $1 - 1 )) ) ))
```

Реализация факториала с использованием цикла:

```
#!/bin/sh
res=1
for i in $(seq 1 "$1"); do
    res=$((res * i))
done
echo $res
```

Существует веб-инструмент `ShellCheck` [4], которым можно пользоваться для проверки корректности Bash-скриптов.

## 2.2. Инструменты командной строки

Команда `grep` осуществляет поиск по образцу, определяемому регулярным выражением. Команда `sed` является строчным редактором, но главное ее использование состоит в замене по шаблону, как и в случае `grep`, заданному регулярным выражением.

В табл. 1 показаны примеры некоторых базовых элементов регулярного выражения.

Таблица 1. Некоторые базовые регулярные выражения

Символ	Действие
Буквы, числа, некоторые знаки	Обозначают сами себя
.	Любой символ
[множество символов]	Любой символ из множества
[^множество символов]	Любой символ не из множества
^	Начало строки
\$	Конец строки
~	Начало строки
выражение*	Повторение выражения 0 или более раз
выражение выражение	Последовательность из выражений

Веб-инструмент `regex101` [5] может помочь в поэлементном разборе сложных регулярных выражений.

С использованием `grep` и `sed` можно создать достаточно сложные схемы обработки данных. В частности, следующий код осуществляет проверку правописания для файла `README.md` на основе словаря из файла `unix-words`:

```
cat README.md | tr A-Z a-z | tr -cs A-Za-z '\n' | sort | uniq |
↪ grep -vx -f unix-words >out ; cat out | wc -l | sed 's/$/'
↪ misspelled words!/'
```

Еще более изощренным, чем `grep` и `sed`, является инструмент `awk`. `Awk` (по именам авторов – Aho, Weinberger, Kernighan) представляет собой язык программирования для обработки текстовых данных.

Ниже показан пример вывода колонки №5 из данных, предоставленных вызовом `ls -l`:

```
localhost:~# ls -l
total 36
-rw-r--r--      1 root      root      114 Jul  5  2020 bench.py
-rwxr-xr-x      1 root      root        51 Nov  4 18:56 fact.sh
-rwxr-xr-x      1 root      root        76 Nov  4 19:45 fact2.sh
drwxr-xr-x      2 root      root        37 Nov  4 18:01 foo
-rw-r--r--      1 root      root        76 Jul  3  2020 hello.c
-rw-r--r--      1 root      root       22 Jun 26  2020 hello.js
-rwxr-xr-x      1 root      root         6 Nov  4 18:44 lsl
-rw-r--r--      1 root      root      151 Jul  5  2020
↳  readme.txt
-rwxr-xr-x      1 root      root        36 Nov  4 18:50 test.sh
localhost:~# ls -l | awk '{ print $5 }'
```

```
114
51
76
37
76
22
6
151
36
```

Средствами `awk` легко подсчитать общий размер файлов:

```
localhost:~# ls -l | awk '{ s += $5 } END { print s }'
569
```

В заключение рассмотрим пример вывода на экран самой свежей новости с ресурса Hacker News:

```
#!/bin/sh
N=$(curl -s https://hacker-news.firebaseio.com/v0/topstories.json
↳ | jq '[0]')
curl -s "https://hacker-news.firebaseio.com/v0/item/$N.json" | jq
↳ '["title"]' | cowsay
```

Вот как выглядит вывод этой программы:

```
root@Server584432:~# ./hn.sh
-----
/ "Finishing my first game while working \"
```

```

\ full-time" /
-----
      ^__^
      (oo)\_______
          (__)\       )\/\
              ||----w |
              ||     ||

```

На тему анализа данных в командной строке существует целая книга [6].

# 3. Менеджеры пакетов

## 3.1. Нумерация версий ПО

В программе, состоящей из нескольких модулей (программных библиотек, пакетов), подключение этих модулей может происходить следующими способами:

- статическое связывание, которое осуществляется на этапе компиляции;
- динамическое связывание, происходящее на этапе выполнения программы.

Как программа, так и подключаемые к ней модули, обычно существуют в нескольких версиях.

Существуют различные схемы нумерации версий:

- последовательные значения;
- дата выпуска;
- хеш-сумма содержимого;
- различные экзотические схемы (например, последовательные цифры числа  $\pi$  или  $e$ ).

Проблема **ада зависимостей** (dependency hell) состоит в наличии конфликтных ситуаций между зависимостями различных модулей. Например, в системном каталоге библиотек может быть только одна версия библиотеки  $X$ , при этом одна программа требует  $X$  в версии  $A$ , а другая – в несовместимой с  $A$  версии  $B$ .

Для управления зависимостями между модулями необходима строгая и единая система нумерации версий, определяющая совместимые и несовместимые сочетания модулей.



## 3.2. Семантическая нумерация версий

В спецификации SemVer (semantic versioning) версия состоит из следующих компонентов:

1. мажорная часть, означающая несовместимые с предыдущими версиями изменения;
2. минорная часть, означающая, добавление обратно совместимой функциональности;
3. патч, к которому относятся обратно совместимые исправления ошибок.

Сравнение двух версий осуществляется слева направо, до определения первого расхождения. Таким образом устанавливается порядок между версиями, а также могут быть введены интервалы версий.

С помощью символа `^` указываются совместимый интервал версий. Например `~1.2.3` означает `>=1.2.3` и `<2.0.0`. Символ `~` определяет близкие друг к другу версии. Например `~1.2.3` означает `>=1.2.3` и `<1.3.0`.

## 3.3. Управление пакетами

Менеджер пакетов предназначен для автоматического управления установкой, настройкой и обновлением специального вида модулей, называемых пакетами.

Менеджеры пакетов бывают следующих видов:

1. уровня ОС (например, RPM в Linux);
2. уровня языка программирования (например, npm для JavaScript);
3. уровня приложения (например, Package Control для редактора Sublime Text, управление плагинами для среды Eclipse).

**Пакет** содержит в некотором файловом формате программный код и метаданные. К **метаданным** относятся:

- указание авторства, версии и описание пакета;
- список зависимостей пакета;
- хеш-значение содержимого пакета.

Для хранения пакетов используется **репозиторий**. Репозитории делятся на локальные, располагающиеся на компьютере пользователя и удаленные, размещаемые на сервере, предназначенном работы с конкретным менеджером пакетов.

## 3.4. Менеджер пакетов apk

Менеджер пакетов apk входит в дистрибутив Alpine ОС Linux.

Работа с apk осуществляется из командной строки. Для добавления нового пакета необходимо использовать опцию add:

```
/ # apk add python3
fetch https://dl-cdn.alpinelinux.org/alpine/v3.14/main/x86_64/
APKINDEX.tar.gz
fetch
  ↪ https://dl-cdn.alpinelinux.org/alpine/v3.14/community/x86_64/
APKINDEX.tar.gz
(1/13) Installing libbz2 (1.0.8-r1)
(2/13) Installing expat (2.4.1-r0)
(3/13) Installing libffi (3.3-r2)
(4/13) Installing gdbm (1.19-r0)
(5/13) Installing xz-libs (5.2.5-r0)
(6/13) Installing libgcc (10.3.1_git20210424-r2)
(7/13) Installing libstdc++ (10.3.1_git20210424-r2)
(8/13) Installing mpdecimal (2.5.1-r1)
(9/13) Installing ncurses-terminfo-base (6.2_p20210612-r0)
(10/13) Installing ncurses-libs (6.2_p20210612-r0)
(11/13) Installing readline (8.1.0-r0)
(12/13) Installing sqlite-libs (3.35.5-r0)
(13/13) Installing python3 (3.9.5-r1)
Executing busybox-1.33.1-r3.trigger
OK: 56 MiB in 27 packages
```

В первую очередь менеджером пакетов были загружены списки доступных пакетов из удаленного репозитория. В архиве APKINDEX.tar.gz можно найти информацию по устанавливаемому пакету и его зависимостям:

```
C:Q1A2S5Zfy6pdKftVzGVhkE5s9r1f4=
P:python3
V:3.9.5-r1
A:x86_64
S:13405337
I:47747072
T:A high-level scripting language
U:https://www.python.org/
L:PSF-2.0
```

```

o:python3
m:Natanael Copa <ncopa@alpinelinux.org>
t:1620852262
c:2d63700fe78744e22c497d2cf7f8610828f00544
D:so:libbz2.so.1 so:libc.musl-x86_64.so.1 so:libcrypto.so.1.1
  ↳ so:libexpat.so.1 so:libffi.so.7 so:libgdbm.so.6
  ↳ so:libgdbm_compat.so.4 so:liblzma.so.5 so:libmpdec.so.3
  ↳ so:libncursesw.so.6 so:libpanelw.so.6 so:libreadline.so.8
  ↳ so:libsqlite3.so.0 so:libssl.so.1.1 so:libz.so.1
p:so:libpython3.9.so.1.0=1.0 so:libpython3.so=0 cmd:2to3
  ↳ cmd:2to3-3.9 cmd:pydoc3 cmd:pydoc3.9 cmd:python3 cmd:python3.9
  ↳ py3.9:README.txt=3.9.5-r1

```

Здесь P: определяет имя пакета, а V: – его версию. С помощью D: в последних строках определяются зависимости, а с помощью p: – те модули, которые пакет предоставляет после своей установки.

## 3.5. Менеджер пакетов apt

В дистрибутиве Ubuntu ОС Linux используется менеджер пакетов apt.

Для установки пакета используется команда apt с опцией install:

```
apt install instead
```

В файле Packages, расположенном на удаленном репозитории, располагается информация о доступных для установке пакетах.

Пакет instead имеет следующее описание:

```

Package: instead
Architecture: amd64
Version: 3.2.1-1
Priority: optional
Section: universe/games
Origin: Ubuntu
Maintainer: Ubuntu Developers
  ↳ <ubuntu-devel-discuss@lists.ubuntu.com>
Original-Maintainer: Sam Protsenko <joe.skb7@gmail.com>
Bugs: https://bugs.launchpad.net/ubuntu/+filebug
Installed-Size: 527

```

```

Depends: libc6 (>= 2.14), libgdk-pixbuf2.0-0 (>= 2.22.0),
↳ libglib2.0-0 (>= 2.12.0), libgtk2.0-0 (>= 2.8.0), liblua5.1-0,
↳ libsdl2-2.0-0 (>= 2.0.8), libsdl2-image-2.0-0 (>= 2.0.2),
↳ libsdl2-mixer-2.0-0 (>= 2.0.2), libsdl2-ttf-2.0-0 (>= 2.0.14),
↳ zlib1g (>= 1:1.1.4), instead-data (= 3.2.1-1)
Filename: pool/universe/i/instead/instead_3.2.1-1_amd64.deb
Size: 224840
MD5sum: ae8edb2974cece3ff42ce664bcba8485
SHA1: 8e4f27d806d5822e2364c26b6a77502e6e1aada1
SHA256:
↳ f0d20dd6d7e3de3c2981bfe3a6fbab0f0be8ecc7bfd3f71736508bee0fe063df
Homepage: https://instead-hub.github.io/
Description: Simple text adventures/visual novels engine
Description-md5: ef0040d4434ac942fb089e9e171d022f

```

Как можно заметить, анализируя строку `Depends:`, помимо системных библиотек пакет `instead` зависит также от пакета `instead-data`, у которого также имеется описание:

```

Package: instead-data
Architecture: all
Version: 3.2.1-1
Priority: optional
Section: universe/games
Source: instead
Origin: Ubuntu
Maintainer: Ubuntu Developers
↳ <ubuntu-devel-discuss@lists.ubuntu.com>
Original-Maintainer: Sam Protsenko <joe.skb7@gmail.com>
Bugs: https://bugs.launchpad.net/ubuntu/+filebug
Installed-Size: 4092
Depends: fonts-liberation
Recommends: instead
Filename: pool/universe/i/instead/instead-data_3.2.1-1_all.deb
Size: 3681604
MD5sum: e244752081374392d024ab100648213a
SHA1: b560cbc4118b95b5633be48ed40b679652dc1322
SHA256:
↳ 06582dc02515a031297dc7ae2e280795f7052266dbb4aee1e2a8406d132ea2c8
Homepage: https://instead-hub.github.io/
Description: Data files for INSTEAD

```

Description-md5: abbaa9f2bdb5492dca18ea9558b57a9d

Зависимости пакета `instead` приведены в графе на рис. 3.

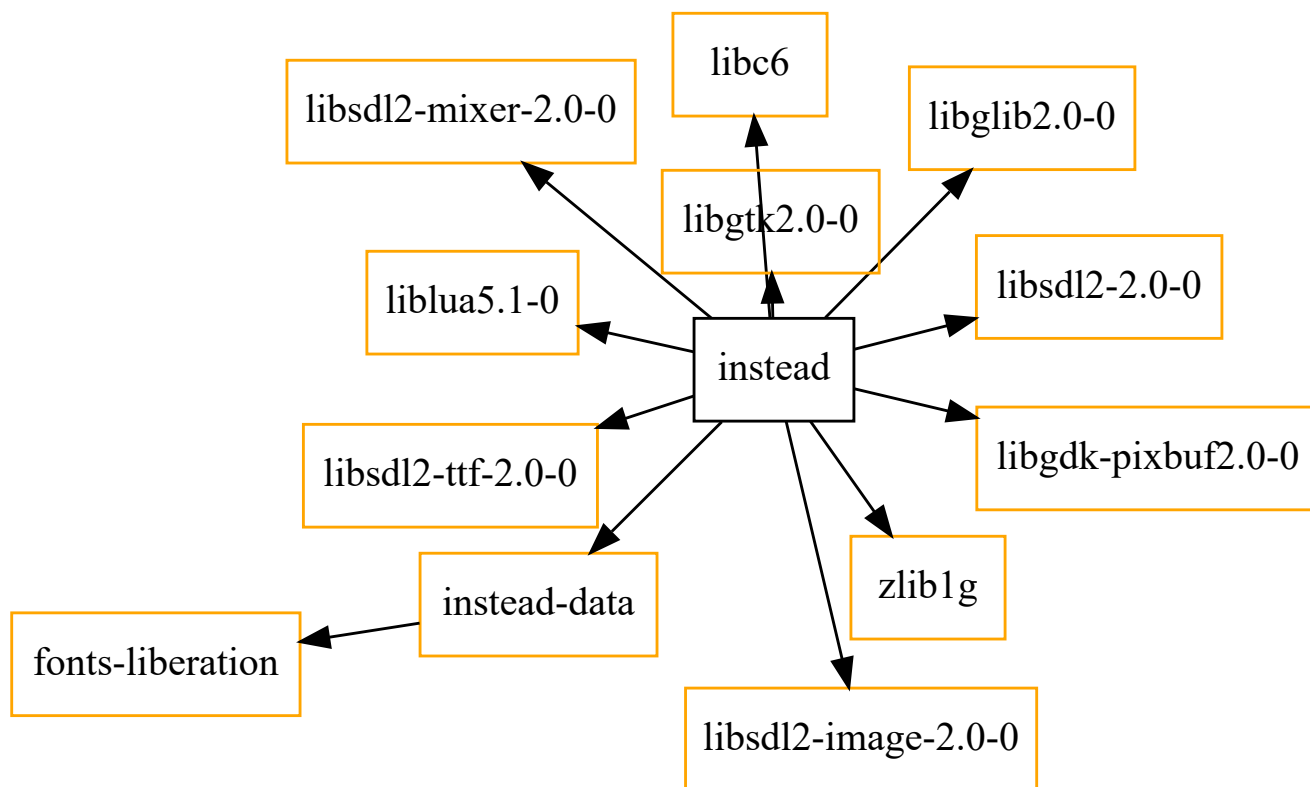


Рис. 3. Зависимости пакета instead

### 3.6. Задача разрешения зависимостей пакетов

Задача разрешения зависимостей, которую решает менеджер пакетов, в общем виде относится к труднорешаемым и может быть определена следующим образом.

Дано:

1. Состояние локального репозитория, содержащего уже установленные пакеты.
2. Состояние удаленного репозитория, в котором находятся все доступные к установке пакеты.
3. Имя устанавливаемого пакета.

Необходимо получить план установки нового пакета с разрешением всех его зависимостей. При этом желателен такой вариант решения, который требует установки минимального числа новых пакетов.

Пример задачи разрешения зависимостей показан на рис. 4. Обратите внимание, что из всех возможных версий пакета menu, менеджеру пакетов необходимо выбрать 1.0.0, иначе не получится разрешить зависимости пакета icons.

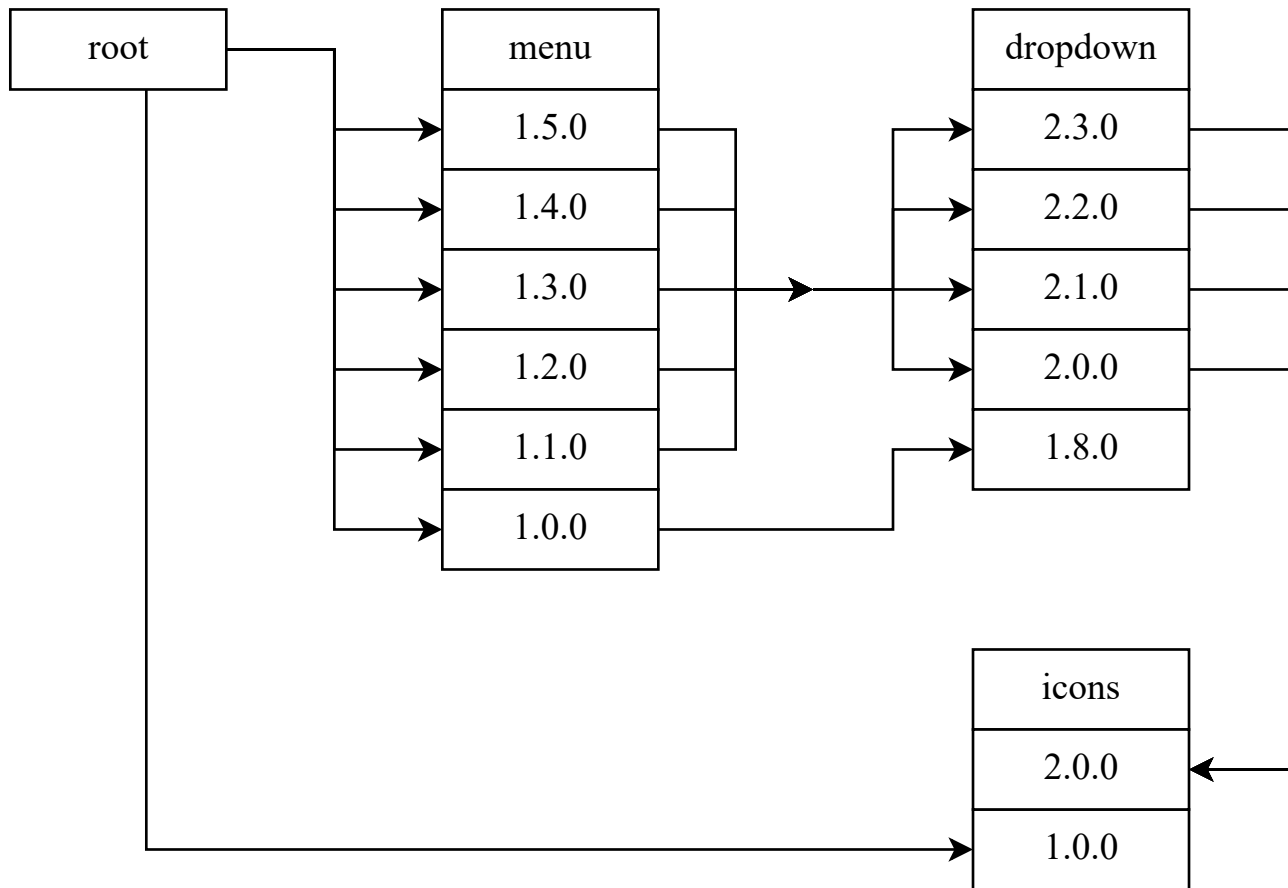


Рис. 4. Пример задачи разрешения зависимостей пакетов

Для задачи разрешения зависимостей в менеджерах пакетов сегодня все чаще используются универсальные решатели NP-полных задач: SAT-решатели [7], решатели для задачи программирования в ограничениях [8] и другие [9].

# 4. Конфигурационные языки

## 4.1. Инфраструктура как код

Параметры конфигурации ПО обычно хранятся в специальном виде, доступном для редактирования пользователями и другими программами. Для хранения может использоваться специальная база данных, но наиболее распространенным вариантом хранения настроек программы являются конфигурационные файлы, содержимое которых написано на одном из конфигурационных языков. Многие программные инструменты конфигурирования ПО используют конфигурационные языки.

**Конфигурационный язык** – язык описания параметров конфигурации ПО. **Конфигурационный файл** представлен на конфигурационном языке и предназначен для редактирования пользователями и другими программами.

К преимуществам конфигурационных языков относят:

- удобный человеко-читаемый синтаксис, который дает возможность внести изменения в конфигурацию быстрее, чем при использовании GUI-инструментов;
- удобный машинно-читаемый синтаксис для автоматической обработки параметров конфигурационного файла внешними программами;
- ограниченные средства программируемости, позволяющие автоматизировать рутинные действия.

В качестве конфигурационных языков могут использоваться:

- непрограммируемые форматы описания конфигурации,
- языки программирования общего назначения,

- программируемые предметно-ориентированные языки для задач конфигурационного управления.

Современный подход «инфраструктура как код» (infrastructure as code) предполагает широкое использование в конфигурационном управлении конфигурационных языков.

## 4.2. Формальные языки и грамматики

Компьютерные языки являются подмножеством формальных языков.

**Формальный язык** состоит из множества строк конечной длины, называемых словами. Слова состоят из символов, а символы содержатся в алфавите – конечном множестве символов.

В этом определении предполагается, что для установления принадлежности некоторой строки формальному языку необходимо сначала перечислить все возможные строки на этом языке, что, разумеется, является непрактичным подходом.

Рассмотрим иной способ определения языка. **Регулярный язык**, являющийся теоретической основой регулярных выражений, может быть описан следующим образом:

1. Пустая строка  $\epsilon$  или одиночный символ из алфавита являются регулярным языком.
2. Если  $A$  это регулярный язык, то  $A^*$  также является регулярным языком. Операция  $A^*$  обозначает повторение  $A$  0 или более раз.
3. Если  $A$  и  $B$  – регулярные языки, то  $A|B$  тоже является регулярным языком. Операция  $A|B$  обозначает выбор из  $A$  или  $B$ .
4. Если  $A$  и  $B$  – регулярные языки, то  $AB$  тоже является регулярным языком. Операция  $AB$  обозначает последовательность:  $A$ , за которым следует  $B$ .

Введенных базовых операций достаточно, чтобы определить дополнительные конструкции, использующиеся в регулярных выражениях. Например, конструкцию  $A^+$  можно определить, как  $AA^*$ , а конструкция  $A^?$  заменяется конструкцией  $A|\epsilon$ .

Мощности регулярных языков, тем не менее, не хватает для описания многих важных компьютерных языков. В частности, с помощью регулярного языка невозможно описать конструкции произвольной вложенности. Это, в том числе, касается разбора и HTML-кода, и даже простого выражения, в



котором присутствуют лишь правильно расставленные скобки с произвольной глубиной вложенности.

Другой способ определения формального языка – формальная грамматика, задающая общие правила построения языка. Этот способ используется для определения синтаксиса компьютерных языков. Формальная грамматика – метаязык, то есть язык, который определяет язык. В частности, для описания регулярных языков может быть задана регулярная грамматика.

Формальные грамматики на практике используются для следующих целей:

- лексический и синтаксический разбор компьютерного языка;
- порождение случайных фраз на компьютерном языке, например, для задач тестирования.

Более мощным, чем регулярные языки, формализмом описания компьютерных языков является **контекстно-свободная грамматика**.

Контекстно-свободную грамматику можно представить себе в виде программы на ограниченном языке программирования. В этом языке есть «имена функций» (нетерминалы, имена правил) и «определения функций» (тела правил). В определениях содержатся «вызовы функций» и «вывод символов» определяемого языка (терминалов). Работа программы начинается с «главной функции», называемой начальным нетерминалом.

Благодаря возможности использования рекурсии в грамматических правилах контекстно-свободные грамматики широко используются для определения синтаксиса компьютерных языков.

Контекстно-свободную грамматику можно представить в форме так называемой железнодорожной или синтаксической диаграммы.

На рис. 5 показан пример грамматики арифметического выражения.

В текстовой форме контекстно-свободная грамматика определяется с помощью формы Бэкуса-Наура (БНФ). На практике, в БНФ часто добавляются конструкции регулярных выражений для упрощения описания синтаксиса.

Ниже представлен пример грамматики арифметического выражения в БНФ:

```
<Значение> ::= <Число> | <Переменная>  
<Операция> ::= "+" | "-" | "*" | "/"
```

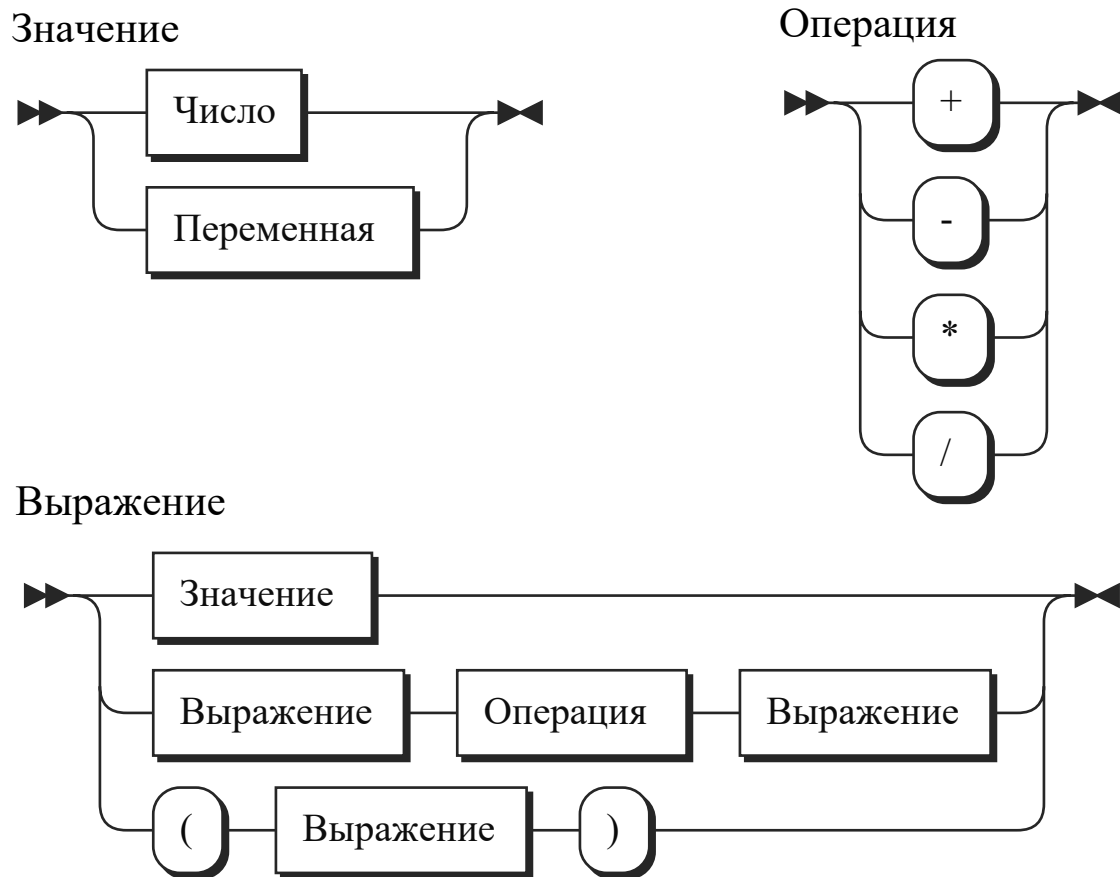


Рис. 5. Синтаксическая диаграмма грамматики арифметического выражения

$\langle \text{Выражение} \rangle ::= \langle \text{Значение} \rangle \mid \langle \text{Выражение} \rangle \langle \text{Операция} \rangle \langle \text{Выражение} \rangle \mid$   
 $\hookrightarrow " (" \langle \text{Выражение} \rangle ") "$

Для автоматизации задач лексического (уровень базовых лексических единиц) и синтаксического (уровень иерархических конструкций) разбора существуют специальные инструменты. Примерами таких инструментов являются ANTLR [10] для Java, SLY [11] для Python и Flex/Bison для C++.

### 4.3. Компьютерные языки

Рассмотрим представленную на рис. 6 упрощенную классификацию компьютерных языков.

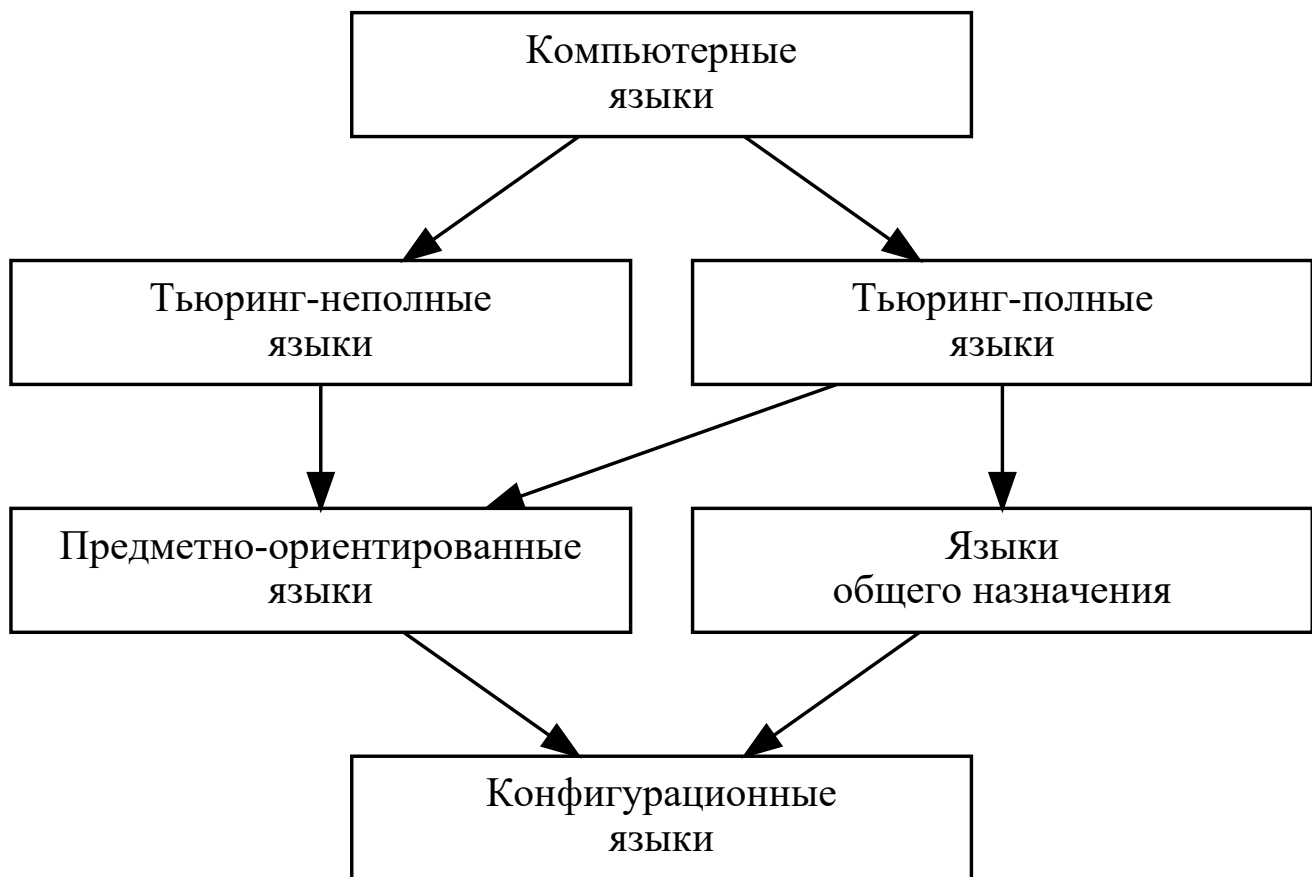


Рис. 6. Упрощенная классификация компьютерных языков

Тьюринг-полными называются языки, на которых можно реализовать модель машины Тьюринга, а значит – реализовать любой алгоритм. К таким языкам относится, в частности, большинство языков программирования.

Тьюринг-полнота необязательно является полезным свойством языка. Для Тьюринг-неполных, ограниченных языков может быть легче

осуществить статический анализ и преобразования на уровне компилятора.

Предметно-ориентированные языки (domain-specific language, DSL) ориентированы задачи некоторого узкого класса, поэтому, зачастую являются Тьюринг-неполными. Основное достоинство DSL – использование предметно-ориентированной нотации.

Известно множество специализированных нотаций. Например, нотация химических формул или нотная нотация. Математики древних цивилизаций описывали математические задачи на естественном языке с использованием сложных систем счисления. Поэтому установление даже некоторых тривиальных для современного школьника, оперирующего алгебраической нотацией, математических фактов могло быть затруднено для математика древности. Заслуживает внимания гипотеза лингвистической относительности, которая предполагает наличие влияния структуры языка на мышление.

## 4.4. Простые форматы описания конфигурации

Одним из старейших форматов описания как кода, так и данных являются **S-выражения** языка Lisp. Это нотация для древовидных структур, реализованных в виде вложенных списков.

Грамматика S-выражений в БНФ:

```
<s-exp> ::= <atom> | '(' <s-exp-list> ')'  
<s-exp-list> ::= <sexp> <s-exp-list> |  
<atom> ::= <symbol> | <integer> | #t | #f
```

Пример S-выражения:

```
(users  
  ((uid 1) (name root) (gid 1))  
  ((uid 108) (name peter) (gid 108))  
  ((uid 109) (name alex) (gid 109)))
```

К недостаткам S-выражений для описания конфигурации относятся:

- сложность чтения для человека синтаксиса с большим числом скобок;
- нет стандарта S-выражений для описания конфигурационных данных.

Классическим способом представления файлов конфигурации является формат **INI** из Windows, а также схожие с ним варианты **conf** из Unix. В

основе формата – последовательность пара ключ-значение. Пары схожего назначения объединяются в секции:

```
[секция_1]
параметр1=значение1
параметр2=значение2
```

```
[секция_2]
параметр1=значение1
параметр2=значение2
```

В (расширенной) БНФ:

```
ini ::= {section}
section ::= "[" name "]" "\n" {entry}
entry ::= key "=" value "\n"
```

Здесь фигурные скобки обозначают повторение 0 или более раз.

Пример INI-файла (system.ini):

```
; for 16-bit app support
[386Enh]
woafont=dosapp.fon
EGA80WOA.FON=EGA80WOA.FON
EGA40WOA.FON=EGA40WOA.FON
CGA80WOA.FON=CGA80WOA.FON
CGA40WOA.FON=CGA40WOA.FON

[drivers]
wave=mmdrv.dll
timer=timer.drv

[mci]
[network]
Bios=1438818414
SSID=29519693
```

К недостаткам INI для описания конфигурации относятся:

- отсутствие поддержки вложенных конструкций;
- отсутствие стандарта INI для описания конфигурационных данных.

Еще недавно наибольшую популярность среди форматов конфигураци-

онных данных имел **XML** (eXtensible Markup Language).

XML разрабатывался как язык с простым формальным синтаксисом, удобный для создания и обработки документов программами и одновременно удобный для чтения и создания документов человеком, с подчеркиванием нацеленности на использование в Интернете.

В самом базовом виде XML напоминает S-выражения, в которых вместо скобок используются открывающийся и закрывающийся именованные теги.

Пример XML-данных:

```
<?xml version="1.0" encoding="UTF-8"?>
<shiporder orderid="889923"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="shiporder.xsd">
  <orderperson>John Smith</orderperson>
  <shipto>
    <name>Ola Nordmann</name>
    <address>Langgt 23</address>
    <city>4000 Stavanger</city>
    <country>Norway</country>
  </shipto>
  <item>
    <title>Hide your heart</title>
    <quantity>1</quantity>
    <price>9.90</price>
  </item>
</shiporder>
```

Важной особенностью XML является поддержка специальных схем для определения корректности XML-данных: DTD (Document Type Definition), XML Schema и другие. DTD определяет:

- состав элементов, которые могут использоваться в XML документе;
- описание моделей содержания, т.е. правил вхождения одних элементов в другие;
- состав атрибутов, с какими элементами XML документа они могут использоваться;
- каким образом атрибуты могут применяться в элементах;
- описание сущностей, включаемых в XML документ.

В XML Schema дополнительно введены типы данных для элементов.

Фрагмент XML Schema описания элемента item из примера выше:

```
<xs:element name="item" maxOccurs="unbounded">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="title" type="xs:string"/>
      <xs:element name="note" type="xs:string" minOccurs="0"/>
      <xs:element name="quantity" type="xs:positiveInteger"/>
      <xs:element name="price" type="xs:decimal"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

К недостаткам XML для описания конфигурации относится высокая сложность формата, как для чтения человеком, так и для программного разбора.

Формат **JSON** (JavaScript Object Notation) основан на синтаксисе JavaScript. Изначально он использовался в качестве текстового формата обмена данными, но со временем все чаще стал применяться и в качестве формата описания конфигурации приложения.

В JSON используются следующие типы данных:

- число в формате double;
- строка;
- логическое значение true или false;
- массив значений любого поддерживаемого типа;
- объект – словарь пар ключ-значение, где ключ – строка, а значение – любой поддерживаемый тип;
- специальное значение null.

Пример JSON-данных:

```
{"menu": {
  "id": "file",
  "value": "File",
  "popup": {
    "menuitem": [
      {"value": "New", "onclick": "CreateNewDoc()"},
      {"value": "Open", "onclick": "OpenDoc()"},
      {"value": "Close", "onclick": "CloseDoc()"}
    ]
  }
}
```

```
}  
}}
```

Для проверки корректности JSON-данных в формат позже была добавлена схема – JSON Schema, которая во многом схожа с XML Schema.

К недостаткам JSON для описания конфигурации относится отсутствие поддержки комментариев.

Формат **YAML** (YAML Ain't Markup Language) предназначен для сериализации данных, но также часто используется в качестве конфигурационного языка. YAML с точки зрения синтаксиса имеет сходство с языком Python – для определения вложенности конструкций используются отступы. Особенностью YAML является возможность создания (с помощью `&`) и использования (с помощью `*`) ссылок на элементы документа. Это позволяет не приводить полностью повторно встречающиеся данные.

Пример YAML-файла:

```
-  
receipt:      Oz-Ware Purchase Invoice  
date:         2012-08-06  
customer:  
  first_name:  Dorothy  
  family_name: Gale  
  
items:  
  - part_no:   A4786  
    descrip:   Water Bucket (Filled)  
    price:     1.47  
    quantity:  4  
  
  - part_no:   E1628  
    descrip:   High Heeled "Ruby" Slippers  
    size:      8  
    price:     133.7  
    quantity:  1  
  
bill-to: &id001  
  street: |  
    123 Tornado Alley  
    Suite 16
```



```
    city:    East Centerville
    state:   KS

ship-to:   *id001
...
```

К недостаткам YAML для описания конфигурации относятся:

- сложность редактирования сильно вложенных элементов, проблемы с отступом;
- сложность стандарта YAML (размер файла спецификации больше, чем у XML).

Формат **TOML** (Tom's Obvious, Minimal Language) специально предназначен для использования в конфигурационных файлах. Его синтаксис основан на INI. Спецификация TOML не имеет БНФ-описания.

Пример TOML-данных:

```
# This is a TOML document

title = "TOML Example"

[owner]
name = "Tom Preston-Werner"
dob = 1979-05-27T07:32:00-08:00

[database]
enabled = true
ports = [ 8001, 8001, 8002 ]
data = [ ["delta", "phi"], [3.14] ]
temp_targets = { cpu = 79.5, case = 72.0 }

[servers]

[servers.alpha]
ip = "10.0.0.1"
role = "frontend"

[servers.beta]
ip = "10.0.0.2"
role = "backend"
```

К недостаткам TOML для описания конфигурации относится сложность спецификации языка.

## 4.5. Языки общего назначения как конфигурационные

Использование языка программирования общего назначения для описания конфигурации ПО обладает рядом преимуществ по сравнению с простыми непрограммируемыми конфигурационными форматами:

- поддержка функций;
- поддержка циклов;
- поддержка типов данных;
- поддержка импорта файлов.

Благодаря поддержке функций, циклов и разбиению файлов на модули достигается принцип DRY (Don't repeat yourself) – не допускать повторения одной и той же информации в программе. В результате конфигурационные файлы имеют компактную, удобную для редактирования форму.

Поддержка типов данных, в том числе пользовательских, позволяет осуществить на уровне типов проверку корректности программы – то, для чего в таких форматах, как XML, используются схемы.

Некоторые примеры использования языков общего назначения в качестве конфигурационных языков:

- В текстовом редакторе Emacs используется встроенный язык **Elisp** для описания конфигурации;
- В БД **Tarantool** используется Lua;
- Конфигурации на языке Python используются в **setuptools** и Jupyter;
- В **Gradle** конфигурация сборки может быть описана на Groovy и Kotlin.

К недостаткам языков общего назначения для задач конфигурирования относятся:

- небезопасное исполнение стороннего кода;
- сложность, чрезмерная выразительность языка (Тьюринг-полнота), мешающая задачам анализа и порождения конфигурационных данных.

## 4.6. Программируемые конфигурационные языки

К основным DRY-элементам программируемых конфигурационных языков, можно отнести:

- переменные;
- арифметические операции;
- функции в математическом смысле;
- ветвления;
- операции построения многоэлементных массивов и иных составных объектов;
- операции импорта сторонних модулей;

**Jsonnet** является расширенным вариантом JSON и представляет собой Тьюринг-полный конфигурационный язык с динамической типизацией.

Пример описания данных на Jsonnet:

```
local makeUser(user) = {
  local home = std.format("/home/%s", user),
  local privateKey = std.format("%s/.ssh/id_ed25519", home),
  local publicKey = std.format("%s.pub", privateKey),
  home: home,
  privateKey: privateKey,
  publicKey: publicKey
};

[
  makeUser("bill"),
  makeUser("jane")
]
```

Результат преобразования в JSON:

```
[
  {
    "home": "/home/bill",
    "privateKey": "/home/bill/.ssh/id_ed25519",
    "publicKey": "/home/bill/.ssh/id_ed25519.pub"
  },
  {
    "home": "/home/jane",
    "privateKey": "/home/jane/.ssh/id_ed25519",
    "publicKey": "/home/jane/.ssh/id_ed25519.pub"
  }
]
```

```

{
  "home": "/home/jane",
  "privateKey": "/home/jane/.ssh/id_ed25519",
  "publicKey": "/home/jane/.ssh/id_ed25519.pub"
}
]

```

Еще одним программируемым конфигурационным языком является **Dhall**. Этот язык со статической типизацией, основанный на принципах функционального программирования. Авторы определяют Dhall, как JSON + функции + типы + импорты. Важной особенностью Dhall является тот факт, что это Тьюринг-неполный язык. В Dhall имеется режим нормализации представления, при котором все языковые абстракции разворачиваются и результат представляет собой уровень представления JSON или YAML.

Пример описания данных на Dhall:

```

let makeUser = \(user : Text) ->
  let home      = "/home/${user}"
  let privateKey = "${home}/.ssh/id_ed25519"
  let publicKey  = "${privateKey}.pub"
  in { home = home
      , privateKey = privateKey
      , publicKey  = publicKey
    }

in [ makeUser "bill"
    , makeUser "jane"
  ]

```

# 5. Системы автоматизации сборки

## 5.1. Виды систем сборки

Сборка программного проекта может состоять из большого числа этапов, среди которых: компиляция модулей, подготовка файлов данных и преобразование их форматов, генерирование документации. Чтобы не повторять рутинные действия из раза в раз можно написать программу на языке интерпретатора командной строки. Тем не менее, работа простых сборочных скриптов на больших проектах часто занимает непозволительно больше время.

К системе сборки [12] могут быть предъявлены следующие требования:

- каждая сборочная задача выполняется не более одного раза, а сами задачи выполняются лишь в том случае, если они прямо или косвенно зависят от входных данных, которые изменились с предыдущей сборки проекта;
- задачи, которые прямо и косвенно не зависят друг от друга, имеется возможность выполнить параллельно.

К *основным элементам системы сборки* относятся:

- таблица, в которой содержатся ключи – задачи и значения – файлы или другие данные, определяющие результат выполнения задачи;
- план выполнения задач с определением действий по выполнению каждой задачи и указанием стартовой задачи;
- алгоритм планировщика задач и способ определения задач, которые не нуждаются в перестроении.

Система сборки принимает на вход таблицу и план выполнения задач и возвращает таблицу в актуальном состоянии – для стартовой задачи и всех

ее зависимостей выполнены необходимые действия.

Система сборки принимает описание задачи, целевой ключ и хранилище и возвращает измененное хранилище, в котором целевой ключ и всего его зависимости принимают актуальные значения.

*Таблица* может быть организована одним из следующих способов:

1. файловая система вместо отдельной таблицы и время модификации файлов. Если время модификации одного из файлов-зависимостей задачи более новое, чем время модификации файла-результата самой задачи – необходимо перестроить задачу;
2. таблица, использующая хеш-значения файлов в качестве ключей. Если хэш файла, связанного с задачей, изменился, то задачу необходимо перестроить.

*План выполнения задач* описывается на языке системы сборки. Такие языки можно отнести к классу конфигурационных языков.

Системы сборки различаются по типу используемого *алгоритма планировщика*:

- алгоритм топологической сортировки графа зависимостей задач;
- алгоритм на основе рестартов задач;
- алгоритм на основе приостановки задач;

Алгоритм на основе рестартов задач устроен следующим образом. Рассмотрим ситуацию, когда выполняется задача  $A$  и было установлено, что она имеет зависимую задачу  $B$ , которая должна быть выполнена в первую очередь. В этом случае выполнение задачи  $A$  отменяется, выполняется задача  $B$ , затем выполнение задачи  $A$  стартует повторно.

Алгоритм на основе приостановки задач отличается тем, что не отменяет выполнение задачи  $A$  полностью, а лишь приостанавливает это выполнение. После завершения задачи  $B$  система сборки возвращается к приостановленной ранее задаче  $A$  и продолжает ее выполнение.

Также системы сборки различаются по **типу зависимостей**:

- статические зависимости;
- динамические зависимости.

Статические зависимости устанавливаются на этапе составления плана выполнения задач. Динамические зависимости обнаруживаются в процессе сборки. Например, одна из задач формирует файл-список файлов-рисунков,

для каждого из которых необходимо выполнить отдельную задачу – преобразовать рисунок в некоторый графический формат.

В системе сборки может применяться **техника раннего среза** (early cutoff) – если задача выполнена, но ее результат не изменился с предыдущей сборки, то нет необходимости исполнять зависимые задачи, то есть процесс сборки можно завершить ранее. На практике такая ситуация определяется по хеш-значению файла, связанного с задачей. Например, если в файл main.c был добавлен новый комментарий, сборка может быть остановлена после определения отсутствия изменений в хеш-значении объектного файла – main.o.

При использовании облачной системы сборки скорость сборки может быть существенно увеличена, благодаря разделению результатов сборки между отдельными разработчиками. Облачная система сборки может поддерживать вариант сборки, при котором локально образуются только конечные результаты сборки, а все промежуточные файлы остаются в облаке.

На рис. 7 приведен пример сценария работы с облачной системой сборки. Пользователь совершает следующие действия:

1. Загружает исходные тексты, их хеш-значения 1, 2 и 3.
2. Затем пользователь запрашивает сборку main.exe. Система сборки определяет с помощью изучения истории предыдущих сборок, что кто-то уже скомпилировал ранее эти исходные тексты для main.exe. Результаты их сборки хранятся в облаке с хешами 4 (util.o) и 5 (main.o). Система сборки далее определяет, что для зависимостей с такими хешами есть готовый main.exe с хешем 6. По ключу 6 из облачного хранилища извлекается конечный результат;
3. Далее пользователь изменяет util.c, и его хеш становится равен 7. В облаке комбинации хешей (7, 2) не существует, то есть ранее никто еще не компилировал такой вариант исходного кода. Процесс продолжается до получения нового main.exe, после чего новые варианты файлов и их хеш-значения сохраняются в облаке.

## 5.2. Топологическая сортировка

Топологическая сортировка является популярным алгоритмом планировщика в системах сборки.

Рассмотрим пример, показанный на рис. 8. Здесь целевой задачей явля-

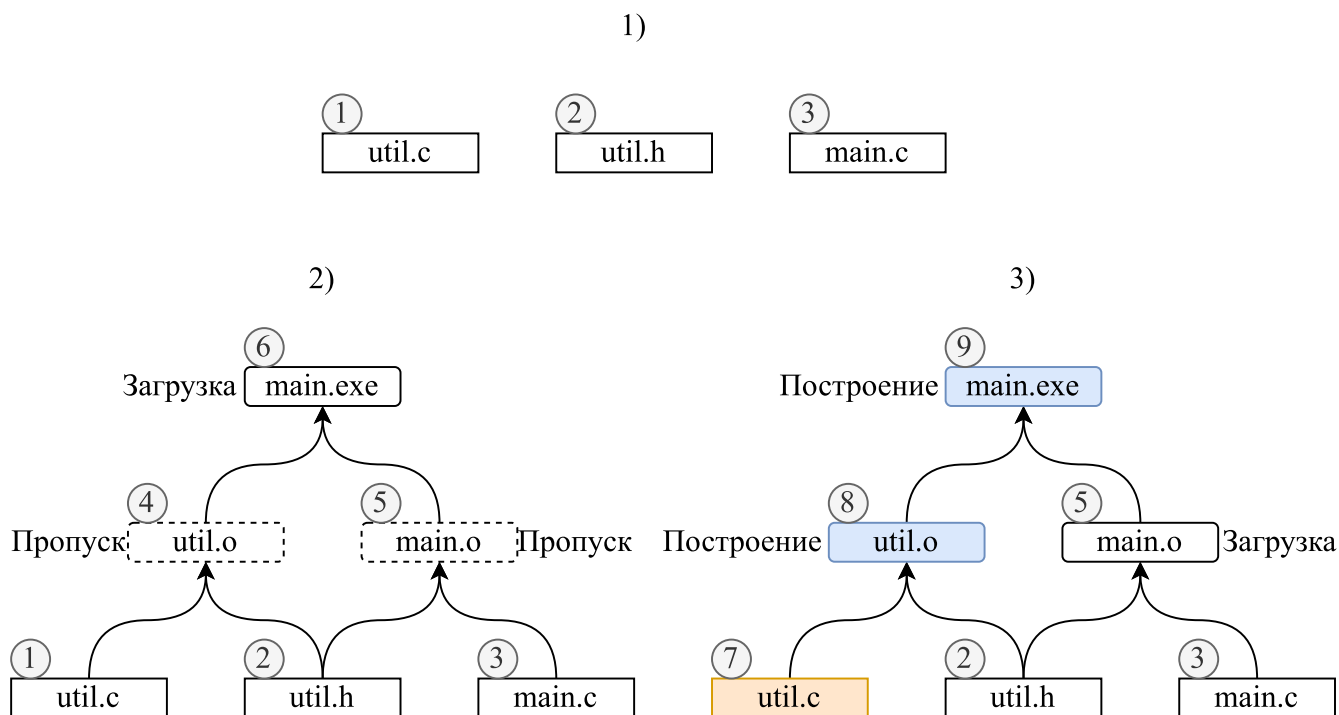


Рис. 7. Пример сценария работы с облачной системой сборки: 1) загрузка исходных текстов, 2) построение main.exe, 3) модификация util.c и пересборка

ется «Экипировка», для достижения которого необходимо выполнить подзадачи в корректном порядке. Таким порядком может быть следующий:

1 3 2 5 7 4 6 8

Легко заметить, что это не единственный корректный порядок. Следующий вариант тоже имеет право на существование:

1 3 4 2 6 5 7 8

Обратите внимание, что такие задачи, как, например, 4 и 5, можно было бы выполнить одновременно, поскольку они не зависят друг от друга. Одевающемуся человеку это выполнить проблематично, но в случае сборки ПО возможность параллельного выполнения подзадач является полезной.

Простой алгоритм топологической сортировки состоит из следующих шагов:

1. Найти узлы графа без входных зависимостей и добавить их к списку результатов.
2. Удалить ранее найденные узлы. Если узлов в графе не осталось, то вернуть результат. В противном случае перейти к п.1.



На практике чаще используется алгоритм топологической сортировки, основанный на обходе графа в глубину.

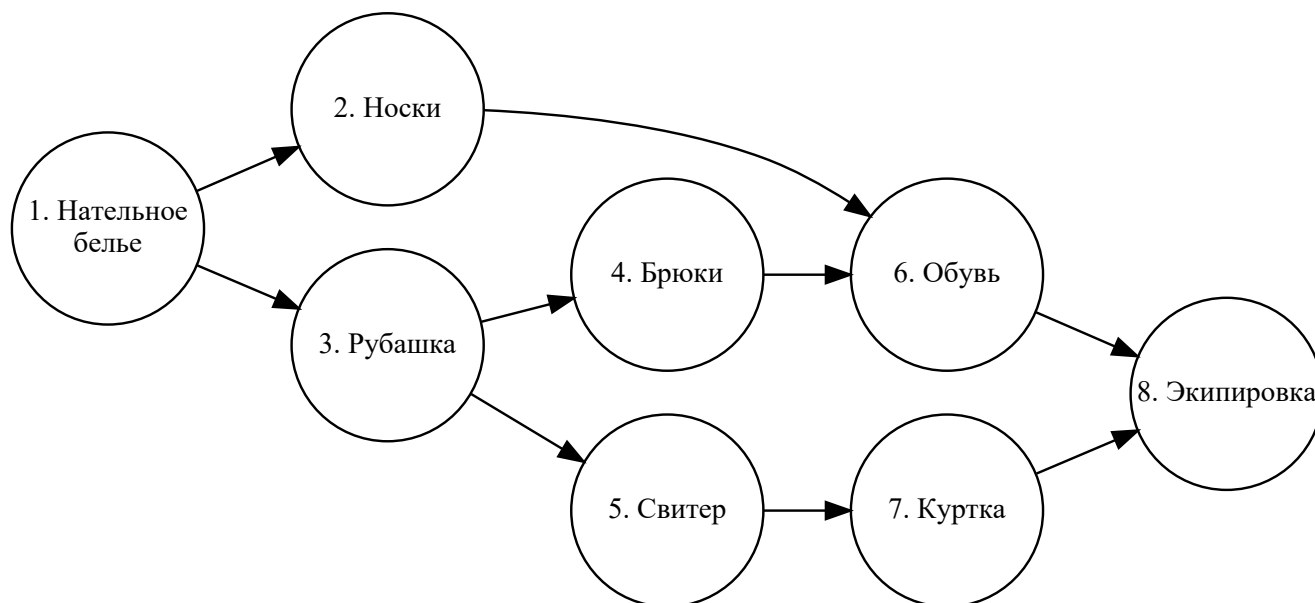


Рис. 8. Пример графа зависимостей задач

## 5.3. Система сборки Make

Система сборки Make является старейшей в своем классе (1976 г.) и при этом до сих пор активно используется разработчиками.

К основным характеристикам Make относятся:

- Использование файловой системы, вместо таблицы, а также времени модификации файлов для определения необходимости пересборки задач;
- Поддержка только статических зависимостей;
- Использование алгоритма топологической сортировки в качестве алгоритма планировщика.

В Make используется специальный конфигурационный файл с именем Makefile для определения плана сборки. Основными элементами декларативного языка Makefile являются определения переменных, а также правила сборки, состоящие из задачи-цели и ряда задач-зависимостей для этой цели.

Рассмотрим пример определения Makefile для графа «экипировки» на рис. 8.

```
equipment: shoes jacket
           @echo $$

underwear:
           @echo $$

socks: underwear
       @echo $$

shirt: underwear
      @echo $$

trousers: shirt
         @echo $$

sweater: shirt
        @echo $$

shoes: trousers
      @echo $$

jacket: sweater
       @echo $$
```

В этом примере имеется последовательность правил следующего вида:

```
цель: зависимости
      действие
```

Действия представляют собой последовательность строк, это команды для интерпретатора командой строки ОС. Командами могут быть, в частности, вызовы компилятора или вызовы инструментов преобразования форматов данных.

Обратите внимание, что строки действий выделяются символом табуляции. Именно табуляцией, а не пробелами. Если была получена ошибка `missing separator` при выполнении Makefile, то речь идет именно о путанице с пробелами и табуляциями.

Цели и зависимости представляют собой, с точки зрения make, имена файлов. У этих файлов и проверяется время последней модификации. Если файл цели задачи отсутствует, то сборка этой задачи всегда будет произве-

дена.

Вызов `make` в каталоге, содержащем приведенный выше `Makefile`, выдаст следующую информацию:

```
underwear
shirt
trousers
shoes
sweater
jacket
equipment
```

В правилах могут использоваться специальные переменные, среди которых:

- `$(C)`. Имя цели.
- `$(P)`. Имя первой зависимости.
- `$(S)`. Список имен всех зависимостей.

Переменные в `Makefile` определяются, как показано в примере ниже:

```
SHOW = @echo $(C)
%:
    $(SHOW)

equipment: shoes jacket
underwear:
socks: underwear
shirt: underwear
trousers: shirt
sweater: shirt
shoes: trousers
jacket: sweater
```

Здесь используется определение шаблона с помощью `%`, что обозначает произвольное имя. В рассматриваемом примере это приводит к выполнению указанного действия для всех целей.

Утилита `make` по умолчанию начинает выполнение с первой цели, указанной в `Makefile`. Можно также указать и конкретную цель для сборки:

```
# make trousers
underwear
```

shirt

trousers

В Makefile часто добавляются псевдоцели, такие как `all` (собрать все) и `clean` (очистить от временных файлов). Для того, чтобы утилита `make` могла отличить псевдоцели от файлов, используется специальная цель `.PHONY`.

## **6. Системы контроля версий**

# 7. Документация как код

Как известно, большинство программистов не любит писать документацию к своим проектам. На это есть причины. В частности, документацию трудно поддерживать в актуальном состоянии в процессе разработки программы. Кроме того, традиционный подход к ведению технической документации с использованием редакторов в духе Microsoft Word с точки зрения разработчика сильно отличается от процессов ведения программного проекта.

В связи с вышесказанным перспективным является подход «документация как код» (docs as code), основная идея которого в использовании для создания технической документации тех же процессов, что и для разработки программ. Подход «документация как код» отличается следующими особенностями:

- Использование текстовых языков разметки, удобных как для чтения человеком, так и с точки зрения машинной обработки.
- Использование текстовых языков описания графических материалов.
- Использование системы контроля версий для хранения проекта документации.
- Использование инструментов командной строки для автоматической проверки, сборки документации и непрерывной интеграции (continuous integration).
- Ориентация на выходной web-формат.

## 7.1. Языки разметки

Языки разметки, помимо очевидной возможности написания текстов, поддерживают специальные команды, отвечающие за внешний вид и структурные особенности документа. В отличие от обычных WYSIWYG-редакторов («что вижу на экране, то и получу в документе») язык разметки

позволяет документ «запрограммировать», при этом «программа» на языке разметки и ее результат в виде документа отличаются друг от друга.

Очевидным примером языка разметки является HTML, но для задач составления документации было создано множество специальных языков, в частности:

- LaTeX,
- Markdown,
- reStructuredText,
- AsciiDoc.

Важным достоинством языка разметки является удобство использование системы контроля версий – в истории репозитория легко отследить изменения, внесенные в документ. Этого не удалось бы добиться с двоичными форматами в духе docx.

Одной из важнейших проблем проектирования языка разметки является обеспечение необходимой гибкости в компьютерной верстке документа при использовании облегченного, почти «невидимого» для пользователя командного языка.

Один из древнейших и, пожалуй, самый мощный язык разметки – TeX, который используется в одноименной системе компьютерной верстки. TeX был разработан Д. Кнудом в 1978 году для задач написания литературы в области компьютерных наук. В 1984 году Л. Лэмпорт создал набор макросов для TeX под названием LaTeX. Сегодня LaTeX используется для написания статей в журналах по математике и физике, создания технических книг, дипломов и диссертаций.

LaTeX отличают средства автоматизации создания документов, это касается, в частности, построения списка литературы, нумерации элементов и ссылок на них, оптимизации размещения элементов на страницах и описания математических формул.

Ниже представлен пример простого документа в LaTeX:

```
\documentclass[14pt]{article} % Формат страниц
\usepackage{polyglossia} % Поддержка русского языка
\setmainlanguage{russian}
\setmainfont{Times New Roman} % Настройка шрифта

\title{Тестовый документ} % Заголовок
\author{П.Н. Советов} % Автор
```

```
\date{\today} % Дата создания
```

```
\begin{document} % Тело документа
```

```
\maketitle % Вставка заголовка
```

Это простой `\textbf{пример}` документа в `\LaTeX`.

```
\end{document}
```

Можно заметить, что команды в `LaTeX` предваряются символом `\` и могут иметь аргументы, заключенные в скобки различных форм.

Результат компиляции документа с помощью `xelatex` показан на рис. 9.

## Тестовый документ

П.Н. Советов

5 ноября 2021 г.

Это простой **пример** документа в `LaTeX`.

Рис. 9. Результат компиляции `LaTeX`-документа

Особый интерес представляет использующийся в `LaTeX` язык разметки математических формул. Формулы обрамляются символами `$` (встраивание в текст) или `$$` (в отдельной строке). Примеры простых формул представлены ниже:

```
$$
```

```
a^n + b^n = c^n
```

```
$$
```

```
$$
```

```
x_{1,2}=\frac{-b \pm \sqrt {b^2-4ac}}{2a}
```

```
$$
```



```

$$
f(x) = \frac{1}{\sigma \sqrt{2\pi}}
      e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}
$$

$$
\eta(T|a) = \sum_{v \in \text{vals}(a)} \left\{ \frac{|S_a(v)|}{|T|}
      \cdot \eta\left(S_a\left(\left(v\right)\right)\right) \right\}
$$

```

Результат компиляции формул представлен на рис. 10.

$$a^n + b^n = c^n$$

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\eta(T|a) = \sum_{v \in \text{vals}(a)} \frac{|S_a(v)|}{|T|} \cdot \eta(S_a(v))$$

Рис. 10. Полученные формулы

Знакомство с языком описания формул LaTeX очень полезно, поскольку этот язык или его подмножества используются во многих современных системах, например, в языке разметки Wikipedia.

## 7.2. Грамотное программирование

Как уже говорилось выше, особенную проблему представляет разделенность программного кода и документации на программный проект. Ранней попыткой решить указанную проблему является подход «грамотного программирования» (literate programming) Д. Кнута в виде системы WEB, созданной в 1984 году. WEB основана на системе TeX и позволяет вести документацию с внедренным в нее программным кодом.

С помощью инструмента **weave** из WEB-файла извлекается только часть, связанная с документацией. С помощью инструмента **tangle** из WEB-файла извлекается программный код. Схема работы WEB показана на рис. 11.

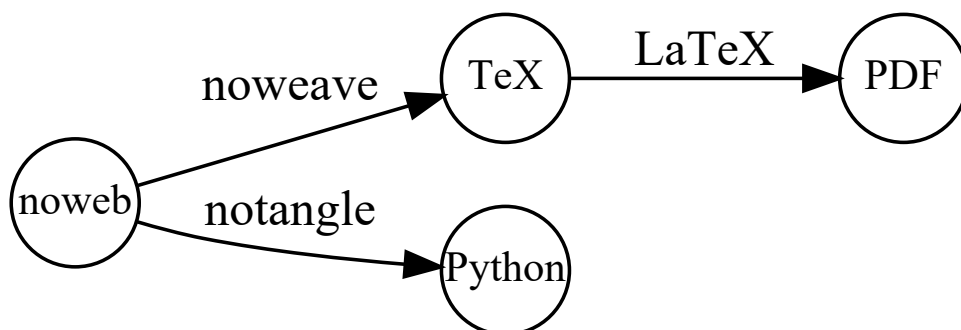


Рис. 11. Схема работы системы WEB

На рис. 12 показан пример LaTeX-документа, извлеченного из WEB-файла.

### Вставка в бинарное дерево

В этом разделе описывается реализация функции `insert` для вставки элемента в бинарное дерево.

```

<insert.py>≡
def insert(tree, key):
    <Случай, когда узел — None>
    <Случай, когда ключ-аргумент не равен ключу узла>

    Рассмотрим случай, когда узел представляет собой None.
    <Случай, когда узел — None>≡
    if not tree:
        tree = Node(key)

    Теперь рассмотрим случаи, когда ключ-аргумент не равен ключу узла.
    <Случай, когда ключ-аргумент не равен ключу узла>≡
    elif key < tree.key:
        tree = Node(tree.key, insert(tree.left, key), tree.right)
    elif key > tree.key:
        tree = Node(tree.key, tree.left, insert(tree.right, key))

    Наконец, рассмотрим оставшийся вариант.
    <insert.py>+≡
    return tree

    Все.
  
```

Рис. 12. Пример WEB-документа

Из приведенного WEB-документа может быть извлечен также следующий код:

```

def insert(tree, key):
    if not tree:
        tree = Node(key)
    elif key < tree.key:
        tree = Node(tree.key, insert(tree.left, key), tree.right)
    elif key > tree.key:
  
```

```

    tree = Node(tree.key, tree.left, insert(tree.right, key))
    return tree

```

В какой-то мере черты грамотного программирования унаследовала система Jupyter-блокнотов, в которой документы представлены в виде последовательности ячеек. Ячейка может либо содержать программный код, либо – документацию. Jupyter-блокноты используются, в основном, в области научно-технических расчетов и для анализа данных.

## 7.3. Markdown и Pandoc

Одним из простейших языков разметки является Markdown. Его авторы преследовали цель получения незаметного командного языка, чтобы файлы на Markdown было легко читать, как есть, даже без трансляции в какое-то выходное представление.

Ниже показаны примеры элементов синтаксиса Markdown:

```

# Заголовок 1
## Заголовок 2
## Заголовок 3

```

Параграф с текстом и *[ссылкой]* (<https://pandoc.org/>).

Параграф с текстом, выделенным **\*\*жирным\*\*** и *\*курсивом\**.

> *Важная цитата* (Автор)

Список элементов:

1. Первый.
1. Второй.
  - \* Вложенный первый.
  - \* Вложенный второй.
1. Третий.

Поле 1	Поле 2	
-----	-----	
Данные 1	Данные 3	
Данные 2	Данные 4	

```
```python
{
  "name": "Ivan",
  "last_name": "Drago",
  "age": 25
}
...`
```

Markdown, в силу простоты и читаемости своего синтаксиса, представляет собой возможную альтернативу LaTeX. Однако, при использовании более-менее сложного форматирования документов возможностей Markdown быстро перестанет хватать. В этой ситуации может помочь инструмент Pandoc, предназначенный для трансляции документов из одного представления в другое. Особенность Pandoc в том, что в нем поддерживается расширенный вариант Markdown, который, в свою очередь, можно дополнить рядом сторонних модулей.

Архитектура Pandoc показана на рис. 13. Работу Pandoc можно разбить на три этапа:

1. Одно из входных представлений с помощью поддерживаемых трансляторов для чтения (readers) преобразуется во внутреннее представление Pandoc – дерево абстрактного синтаксиса.
2. На уровне AST могут бы применены пользовательские фильтры (filters).
3. Полученное AST с помощью поддерживаемых трансляторов для записи (writers) преобразуется в одно из выходных представлений.

В варианте markdown от Pandoc поддерживаются некоторые элементы LaTeX, в частности, язык описания математических формул.

## 7.4. Языки описания диаграмм

Как и в случае текста, языки описания графических материалов позволяют легко вносить и отслеживать изменения в рисунок. Еще одним преимуществом таких языков является возможность автоматизации построения рисунков. Это касается, в частности, построения различных диаграмм на основе автоматического анализа классов и модулей из программного кода.

Одним из наиболее популярных инструментов в этой области является Graphviz, в котором для описания графов различного вида используется язык Dot.

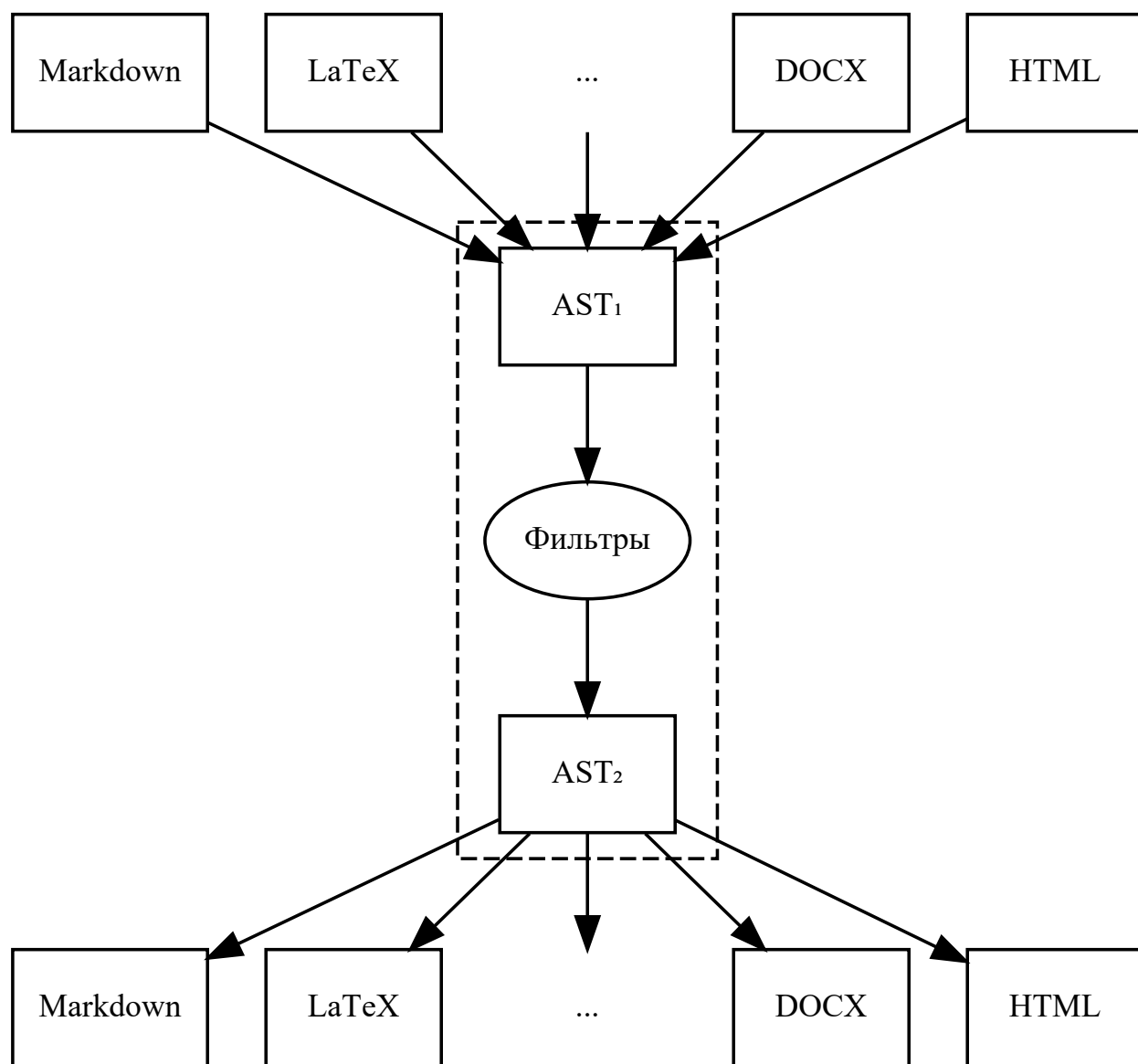


Рис. 13. Архитектура Pandoc

Пример кода на языке Dot показан далее:

```
digraph G {
  n1 [style=filled, color=brown1, label="1", shape=oval]
  n2 [style=filled, color=darkolivegreen1, label="2", shape=box]
  n3 [style=filled, color=aquamarine, label="3", shape=circle]

  n1 -> n2 -> n3
  n3 -> n1
}
```

Результат компиляции в графический файл представлен на рис. 14.

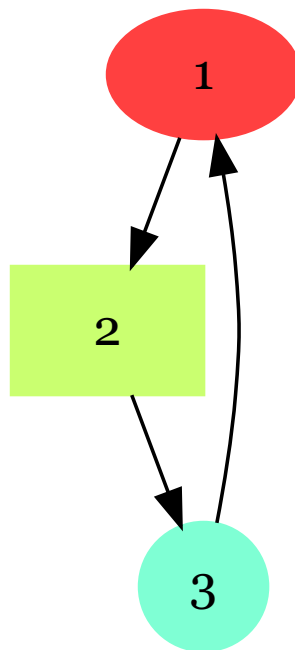


Рис. 14. Результат работы Graphviz

Еще одним популярным инструментом является PlantUML, предназначенный для создания как UML-диаграмм различного вида, так и для диаграмм иного вида (диаграммы Ганта, интеллект-карты и проч.).

Ниже представлен пример диаграммы, описанной на языке PlantUML:

```
@startuml
skinparam monochrome true
skinparam shadowing false
```

```
A -> B: шаг
```

```

activate B
B -> C: шаг

activate C
C --> C: действие
C -> B: шаг
deactivate C

B -> A: шаг
deactivate B
@enduml

```

Результат компиляции в графический файл представлен на рис. 15.

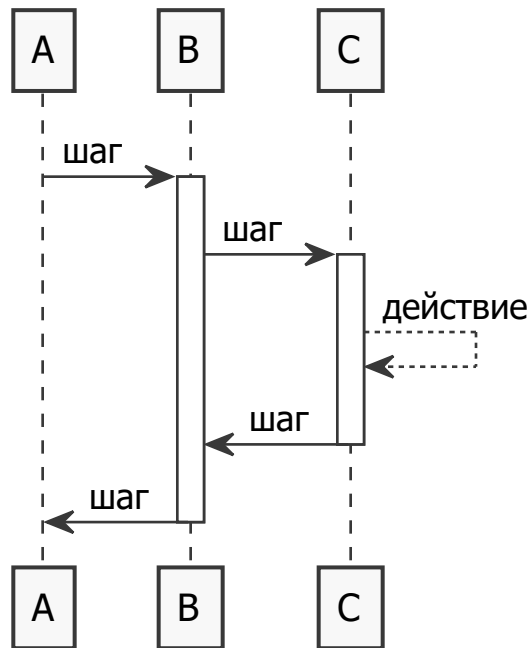


Рис. 15. Результат работы PlantUML

## 7.5. Генераторы документации на основе исходных текстов

Очевидным способом объединения документации и программного кода является подробное комментирование программного кода. При внесении изменений в программу имеет возможность сразу же обновить документирующие поведение программы комментарии в коде. Существуют инструменты, позволяющие автоматически извлечь из файла программы специальным об-

разом оформленные комментарии и оформить результат в виде справочной документации по программному модулю.

Примеры генераторов документации:

- Javadoc для языка Java.
- Doxygen для C++ и некоторых других языков.
- Расширения для системы Sphinx, поддерживающие целый ряд языков.

Ниже показан пример программы на языке C со специальными комментариями, поддерживаемыми системой Doxygen. Обратите внимание на специальные ключевые слова `\file`, `\brief` и `\param`:

```
/// \file main.cpp
/// Модуль main.

#include <stdio.h>

/// \brief Главная функция.
/// \param int argc Счетчик аргументов.
/// \param char **argv Указатель на аргументы.
int main(int argc, char **argv) {
    return 0;
}
```

Далее приведен пример программы на языке Питон, с комментариями, поддерживаемыми системой Sphinx. Обратите внимание на специальные ключевые слова `param`, `type` и `return`:

```
"""
Модуль main.
"""

def main(x, y):
    """
    Функция main.

    :param x: Параметр x.
    :type x: str
    :param y: Параметр y.
    :type y: int
    :return: Ничего не возвращает.
    """
```



# 8. Вопросы виртуализации

## 8.1. Что такое виртуализация

Определение слова «виртуальный» можно найти, к примеру, в толковом словаре Ожегова:

«ВИРТУАЛЬНЫЙ, -ая, -ое; -лен, -льна (спец.). Несуществующий, невозможный. Виртуальные миры. Виртуальная реальность (несуществующая, воображаемая). В. образ (в компьютерных играх)».

Некоторые сюжетные элементы на тему виртуальности из мира фантастики вполне встречаются и в компьютерном мире. Возьмем ситуацию, когда герой обнаруживает внутри виртуального мира еще один, внутренний виртуальный мир. Это похоже на запуск в браузере ОС Linux, в которой, в свою очередь, запускается программа DOSBox, имитирующая работу старого компьютера под управлением MS-DOS.

Можно вспомнить еще одну типичную для фантастических произведений ситуацию, когда герой пытается понять, в реальном ли он находится мире, или же – в виртуальном. Подобная проблема возникает, к примеру, для выполняемой компьютерной программы, которая пытается определить, не происходит ли нежелательный анализ ее поведения под управлением виртуальной машины.

В случае компьютерной системы «реальным миром» является хост-система, а «виртуальным миром» – виртуальная машина, гостевая система или, в более общем смысле, некоторый виртуальный ресурс.

**Виртуализацией** будем называть создание виртуальной (абстрактной, имитируемой) версии ресурса компьютерной системы. Для виртуализации типичными являются следующие свойства:

- изолированность виртуального ресурса от хост-системы и от других

- виртуальных ресурсов,
- разделение ресурсов хост-системы между виртуальными ресурсами.

На рис. 16 показаны основные виды виртуализации.

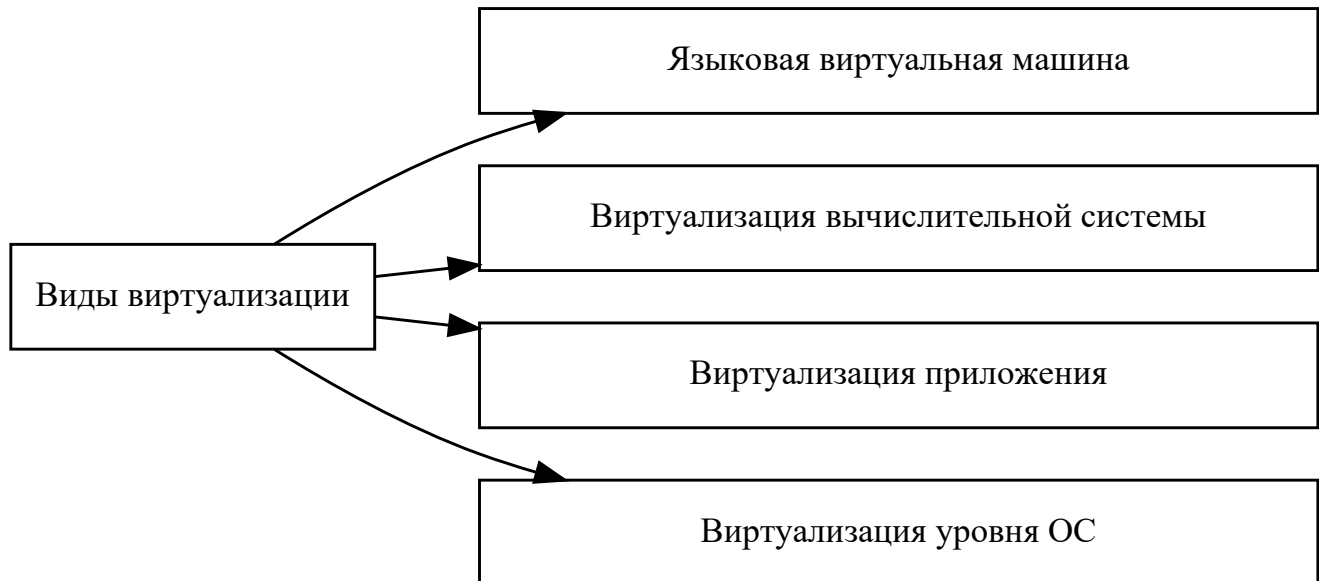


Рис. 16. Виды виртуализации

Понятие виртуальной машины связано с самим понятием алгоритма, определяющего процесс вычислений, который может быть выполнен на машине Тьюринга или на каком-то ином абстрактном вычислителе, полном по Тьюрингу. Доказательство полноты по Тьюрингу достигается демонстрацией реализации внутри интересующего нас вычислителя виртуальной машины, имитирующей работу какого-либо из вычислителей, Тьюринг-полнота которого известна.

На рис. 17 представлены основные техники обеспечения виртуализации.

## 8.2. Языковые виртуальные машины

Языковые виртуальные машины предназначены для исполнения программ на конкретном языке программирования (или семействе таких языков) в условиях различных программно-аппаратных платформах без перекомпиляции. Иными словами, упрощается портирование программ.

Рассмотрим классический пример П-кода (P-code) для виртуального выполнения программ на языке Паскаль, предложенный в середине 70-х годов. Специальный вариант компилятора Паскаля порождает платформонезави-

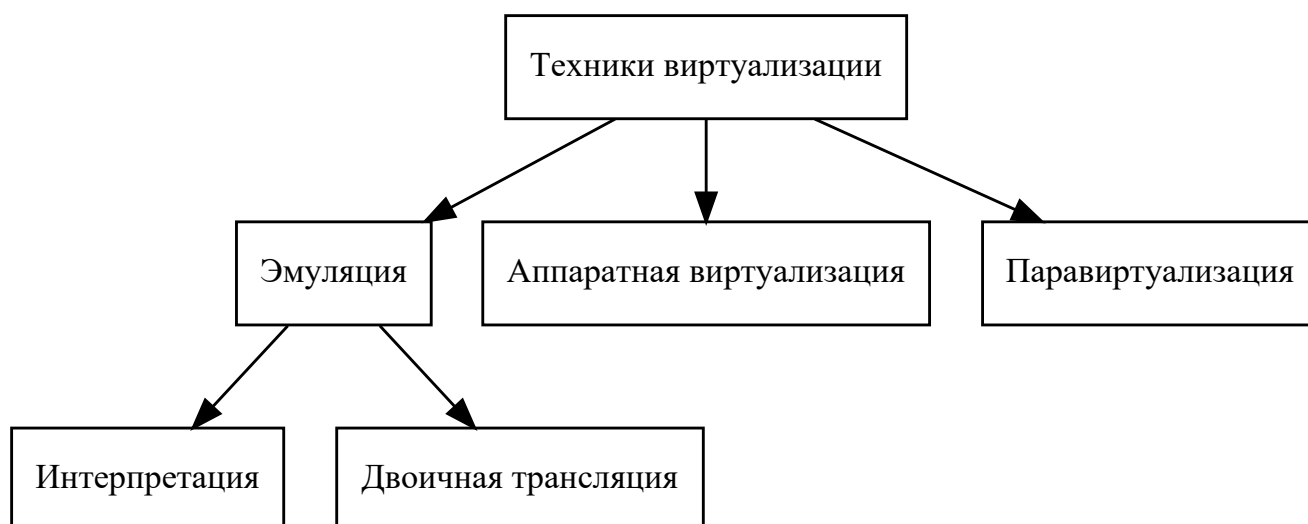


Рис. 17. Техники виртуализации

симый П-код. Для различных программно-аппаратных платформ реализованы интерпретаторы П-кода, также включающие библиотеку времени выполнения. В результате компилятор существует только в одном варианте и для каждой из новых целевых платформ достаточно реализовать небольшую программу – интерпретатор П-кода.

К ярким историческим примерам успешного портирования компьютерных игр на множество различных игровых платформ относятся текстовая игра Zork (1979), реализованная в коде виртуальной Z-машины, а также игра Another World (1991), имеющая изощренную виртуальную машину с поддержкой многопоточности, графики и звука.

К популярным современным языковым виртуальным машинам можно отнести:

- Java Virtual Machine (JVM). Виртуальная машина для Java.
- Common Language Runtime (CLR). Виртуальная машина среды .NET.
- CPython Virtual Machine. Виртуальная машина языка Python.
- WebAssembly (WASM). Виртуальная машина для веб-приложений.

**Языковая виртуальная машина** выполняет команды абстрактного процессора, ориентированного на конструкции конкретного языка или целого семейства языков. Программу в таком представлении принято называть байткодом. Такое название закрепилось исторически, оно связано с виртуальной машиной языка Smalltalk, в которой большинство команд кодировалось одним байтом.

Поскольку реальные процессоры, в большинстве случаев, не поддержива-

ют на аппаратном уровне исполнение **байткода**, то используется программная модель языковой машины.

Архитектуры виртуальных машин, как и архитектуры реальных процессоров, можно разделить на два класса:

- стековые машины,
- регистровые машины.

Рассмотрим байткод различных виртуальных машин для вычисления выражения

$$bb - 4ac.$$

В стековой модели вычислений это выражение представляется в постфиксной форме записи:

b b \* 4 a \* c \* -

В регистровой модели вычислений то же выражение может быть представлено трехадресным кодом:

```
t1 = b * b
t2 = 4 * a
t3 = t2 * c
t4 = t1 - t3
```

В JVM используется стековая модель вычислений:

```
0: iload_1
1: iload_1
2: imul
3: iconst_4
4: iload_0
5: imul
6: iload_2
7: imul
8: isub
9: ireturn
```

В CPython тоже используется стековая модель вычислений:

```
0 LOAD_FAST          1 (b)
2 LOAD_FAST          1 (b)
```

```

4 BINARY_MULTIPLY
6 LOAD_CONST          1 (4)
8 LOAD_FAST           0 (a)
10 BINARY_MULTIPLY
12 LOAD_FAST           2 (c)
14 BINARY_MULTIPLY
16 BINARY_SUBTRACT
18 RETURN_VALUE

```

Виртуальная машина языка Lua использует регистровую вычислительную модель (см. числа после имен операций):

```

1  MUL      3 1 1
2  MMBIN    1 1 8    ; __mul
3  MULK     4 0 0    ; 4
4  MMBINK   0 0 8 1  ; __mul 4 flip
5  MUL      4 4 2
6  MMBIN    4 2 8    ; __mul
7  SUB      3 3 4
8  MMBIN    3 4 7    ; __sub
9  RETURN1  3
10 RETURN0

```

Во многих случаях для выполнения байткода быстрого действия интерпретатора оказывается недостаточно. В этой связи может использоваться трансляция. Различают следующие варианты трансляции:

- АОТ (Ahead-of-Time). Полная трансляция байткода в машинный код целевой платформы перед ее запуском.
- JIT (Just-in-Time). Динамическая трансляция областей программы прямо во время интерпретации байткода.

## 8.3. Виртуализация вычислительной системы

Устаревание программного и аппаратного обеспечения является серьезной проблемой в области информационных технологий. Здесь помощь может оказать виртуализация. Устаревшее оборудование, к которому уже не найти современных драйверов, может быть заменено своей виртуальной программной версией. Программа, предназначенная для выполнения на редкой или устаревшей вычислительной системе, может быть выполнена в рамках виртуальной машины уже на современном компьютере. Виртуализация так-

же полезна для экономного управления вычислительными ресурсами. С использованием виртуализации пользователь на одном физическом компьютере может работать с несколькими виртуальными компьютерами, имеющим различные характеристики и использующими различные ОС. Этот подход используется для организации виртуальных серверов для веб-хостинга, а также применяется в сфере облачных вычислений.

Виртуализация уровня вычислительной системы имеет давнюю историю. Еще в конце 60-х годов компания IBM реализовала в своей ОС CP/CMS поддержку виртуализации на аппаратном уровне. Таким образом было организована работа с несколькими программными версиями компьютера System/360-370. С массовым развитием персональных компьютеров интерес к виртуализации сильно угас. Возобновление этого интереса произошло уже в 1999 году, когда компания VMware выпустила коммерчески успешное ПО VMware Workstation для виртуализации работы ОС Linux и Windows на компьютерах с архитектурой x86.

Для организации виртуальной машины требуется специальный программный модуль – монитор виртуальных машин или гипервизор. **Гипервизор** управляет доступом виртуальных машин к физическим ресурсам хост-машины. В этом отношении гипервизор ведет себя аналогично ядру операционной системы. Гипервизоры разделяют на следующие два типа:

1. Работает прямо на оборудовании хост-машины, без участия ОС.
2. Работает в рамках хост-ОС и, возможно, использует модуль ядра для управления физическими ресурсами хост-машины.

Теоретические основы виртуализуемости, определяющие, насколько эффективно работает гипервизор, были предложены в 1974 году Ж. Попеком (G. Popek) и Р. Голдбергом (R. Goldberg). Рассмотрим подробнее суть предложений этих авторов.

Выделены следующие типы команд компьютера:

- Привилегированные. Выполнение которых вызывает исключение, если это выполнение происходит вне режима ядра ОС.
- Чувствительные. Команды, которые позволяют управлять состоянием компьютера или отслеживать это состояние.
- Безвредные. Остальные команды.

Сформулированы следующие свойства, которыми должен обладать гипервизор:

- Безвредные команды выполняются аппаратурой напрямую, без при-

влечения гипервизора.

- Программы влияют на доступные им ресурсы системы только опосредованно, с помощью гипервизора.
- Программа под управлением гипервизора выполняется точно так же, как и без него, за исключением временных характеристик и объема доступных ресурсов.

Наконец, авторы сформулировали следующую **теорему виртуализуемости**: построение гипервизора, удовлетворяющего этим свойствам для заданного компьютера возможно, если множество чувствительных команд этого компьютера является подмножеством привилегированных команд.

Аппаратная виртуализация, основанная на обработке гипервизором исключений со стороны привилегированных команд использовалась еще в старых ОС от компании IBM. Тем не менее, для персональных компьютеров архитектуры x86 долгое время представленная выше теорема не выполнялась. В начале 2000-х реализация гипервизора на архитектуре x86-64 требовала серьезных ухищрений. Использовалась, в частности, динамическая трансляция программы с целью замены чувствительных команд на соответствующие обращения к гипервизору. Кроме того, применялась техника **паравиртуализации**, при использовании которой в гостевую ОС должны быть внесены изменения, включающие обращения к гипервизору. Наконец, в середине 2000-х компании Intel и AMD добавили в свои процессоры аппаратную поддержку виртуализации. Современные процессоры поддерживают аппаратную виртуализацию как процессора, так и ОЗУ, а также устройств ввода-вывода.

При использовании виртуальной машины с отличающейся от хост-машины процессорной архитектурой необходима **эмуляция**, которая может быть реализована в виде интерпретатора машинного кода, а также на основе статического (АОТ) или динамического (JIT) двоичного транслятора.

Одной из популярных программ для задач виртуализации вычислительных систем является QEMU. В этой программе реализована эмуляция ряда компьютеров различных архитектур с набором периферийных устройств. В QEMU поддерживается аппаратная виртуализация с использованием сторонних гипервизоров, в числе которых KVM и HAXM.

Рассмотрим пример использования QEMU. Предположим, мы хотим использовать дистрибутив Linux Alpine на виртуальной машине с архитектурой x86-64. В первую очередь создадим виртуальный жесткий диск разме-

ром 1 Гбайт, на который будет установлен дистрибутив Linux:

```
qemu-img create -f qcow2 alpine.qcow2 1G
```

Созданный жесткий диск доступен в виде файла `alpine.qcow2`. Теперь необходимо запустить виртуальную систему x86-64 с объемом ОЗУ в 1 Гбайт для установки Alpine на виртуальный жесткий диск. При этом используется виртуальный CD-ROM и ISO-файл `alpine-standard-3.14.2-x86_64.iso`:

```
qemu-system-x86_64 -m 1G -cdrom alpine-standard-3.14.2-x86_64.iso  
↪ -hda alpine.qcow2
```

После установки Alpine на жесткий диск можно перезагрузить виртуальную систему и далее работать с ней с помощью следующей команды:

```
qemu-system-x86_64 -m 1G -hda alpine.qcow2
```

## 8.4. Виртуализация приложения

Виртуализация приложения позволяет выполнять приложения на тех платформах, для которых они не были предназначены. При этом виртуализируются интерфейсы операционной системы. Примером виртуализации уровня приложения является система Wine, позволяющая запускать приложения для Windows внутри Linux.

В QEMU имеется эмуляция режима пользователя, которая позволяет выполнять Linux-приложения, скомпилированные для одной процессорной архитектуры, на другой процессорной архитектуре.

## 8.5. Виртуализация уровня ОС

При виртуализации уровня ОС виртуализация аппаратных ресурсов обычно отсутствует. Существует единственное хост-ядро ОС и набор изолированных друг от друга пользовательских пространств, подключаемых к нему. Такие подключаемые пользовательские пространства называются **контейнерами**. Каждый контейнер содержит отдельную копию файловой системы ОС с установленным набором программ. Использование контейнеров облегчает установку, использование и взаимодействие сложных программ. В отличие от виртуализации, контейнеризация не несет накладных расходов на выполнение гостевого ядра ОС.



Рассмотрим пример использования контейнеров в системе Docker. Установим контейнер дистрибутива Alpine:

```
docker pull alpine
```

Запустим (`run`) контейнер в интерактивном режиме (`-it`), вызвав командный интерпретатор `sh` в Alpine:

```
docker run -it alpine sh
```

Необходимо учитывать, что после окончания сеанса работы с контейнером все несохраненные данные исчезнут. Данные, которые необходимо сохранять, например, файлы базы данных, записываются в каталог хост-системы. Команда Docker `-v` указывает этот каталог, а также его представление в файловой системе контейнера (`/my_data`):

```
docker run -it -v "$(pwd)/my_data:/my_data" alpine sh
```

Для построения контейнера удобно использовать конфигурационный файл Dockerfile. Предположим, стоит задача построить контейнер на базе Alpine с добавленным интерпретатором языка Python. Для выполнения этой задачи Dockerfile имеет следующий вид:

```
FROM alpine
RUN apk add python3
```

Команда `FROM` указывает в этом примере базовый контейнер, а команда `RUN` – те команды, которые необходимо выполнить в процессе построения контейнера.

## 9. Литература

1. The missing semester of your CS education [Online]. — URL: <https://missing.csail.mit.edu/>.
2. *Irving D., Hertweck K., Johnston L., Ostblom J., Wickham C., Wilson G.* Research software engineering with python: Building software that makes research possible. — Chapman; Hall/CRC, 2021.
3. *Ramey C.* The bourne-again SHell // The Architecture of Open Source Applications. — 2011.
4. ShellCheck – shell script analysis tool [Online]. — URL: <https://www.shellcheck.net/>.
5. regex101: Build, test, and debug regex [Online]. — URL: <https://www.shellcheck.net/>.
6. *Janssens J.* Data science at the command line. — ” O’Reilly Media, Inc.”, 2021.
7. *Tucker C., Shuffelton D., Jhala R., Lerner S.* Opium: Optimal package install/uninstall manager / 29th international conference on software engineering (ICSE’07). — IEEE, 2007. — P. 178–188.
8. The package manager for dart [Online]. — URL: <https://github.com/dart-lang/pub>.
9. *Abate P., Di Cosmo R., Gousios G., Zacchiroli S.* Dependency solving is still hard, but we are getting better at it / 2020 IEEE 27th international conference on software analysis, evolution and reengineering (SANER). — IEEE, 2020. — P. 547–551.
10. ANTLR [Online]. — URL: <https://www.antlr.org/>.
11. Sly lex yacc [Online]. — URL: <https://github.com/dabeaz/sly>.
12. *Mokhov A., Mitchell N., Jones S. P.* Build systems à la carte: Theory and practice // Journal of Functional Programming. — Cambridge University Press, 2020. — Vol. 30.

