



Санкт-Петербургский государственный университет

Кафедра системного программирования

Проектирование и создание системы для анализа зависимостей пакетов в дистрибутивах GNU/Linux

Бурашников Артем Максимович, группа 22.Б07-мм

Научный руководитель: К.К. Смирнов, старший преподаватель кафедры ИАС

Санкт-Петербург
2023

- Вокруг каждого семейства дистрибутивов GNU/Linux своя экосистема
- Для установки программ используются пакеты, чьи форматы не унифицированы
- Хотим анализировать метаданные пакетов в контексте архитектуры RISC-V – может быть полезно как для бизнеса, так и для разработчиков
- Междистрибутивный и межархитектурный анализ — проблема

Постановка задачи

Целью работы является создание системы, которая помогает проводить междистрибутивный и межархитектурный анализ пакетов и их зависимостей в ОС Linux (приоритетная архитектура — RISC-V).

Задачи:

- Провести обзор инструментов, позволяющих анализировать метаданные пакетов, с целью выбора функциональности создаваемого инструмента
- Проведя анализ форматов метаданных, выбрать два дистрибутива, для которых обозначенная функциональность будет реализована
- Реализовать приложение на языке Python, учитывая сформированные требования

Цель обзора — выбрать требуемую для реализации функциональность

Критерии отбора инструментов:

- Поддерживается разработчиками
- Родной для дистрибутива или кросс-платформенный
- Может выводить информацию о метаданных для нескольких архитектур и/или дистрибутивов

Обзор существующих инструментов — сравнение

- **Apt** — пакетный менеджер для Debian и производных систем
- **Pactree** — пакетный менеджер для Arch Linux
- **Debtrees** — сторонняя утилита, распространяемая через apt
- **Repology** — веб-приложение агрегатор

	Apt	Pactree	Debtrees	Repology
Прямые зависимости	✓	✓	✓	✓
Обратные зависимости	✓	✓	✓	✓
Сравнение метаданных	✗	✗	✗	✗
Построение графа	✗	✓	✓	✗
Кросс-платформенность	✗	✗	✗	✓
API	CLI	CLI	CLI	REST API

Обзор существующих инструментов — вывод

- Сравнение по прямым зависимостям пакетов между различными архитектурами
- Вывод имен пакетов и архитектур, для которых можно произвести сравнение
- Агрегация пакетов из различных дистрибутивов
- Пользовательский интерфейс
- Кросс-платформенность

Существуют две основные категории пакетов:

- **rpm**: кодируют спецификации в бинарной форме
- **deb**: текстовое представление

Несмотря на различия форматов архивов, спецификации зависимостей в метаданных практически идентичны

Спецификация зависимостей

Отношения, используемые между пакетами, на примере спецификации Debian:

- 1 Depends
- 2 Recommends
- 3 Suggests
- 4 Enhances
- 5 Pre-Depends
- 6 Conflicts
- 7 Provides
- 8 Replaces

Выбор первичных дистрибутивов

- **Ubuntu**: метаданные в текстовом виде, простой синтаксис — позволило создать прототип и разработать интерфейс для пользователя
- **Fedora**: метаданные в репозитории представлены в файле `.sqlite`, использует пакеты `rpm` — созданное приложение работает с обоими форматами
- Обе ОС поддерживают RISC-V и входят в десятку самых популярных, согласно данным портала DistroWatch
- Помимо RISC-V, релизы и upstream репозитории выбраны произвольно
- Учтена возможность расширить приложение добавлением других релизов, репозиториях и дистрибутивов.

Реализовано на Python с использованием следующих библиотек и приложений:

- **Poetry**: менеджер зависимостей и система сборки
- **black**: форматтер для поддержания единого стиля кода
- **ruff**: линтер, использующий те же правила, что и flake8
- **mypy**: статическая проверка типов, для всех функций написаны аннотации типов

Для этих библиотек настроен CI

Приложение хранит метаданные в базах данных SQLite

- Упростило процесс внедрения метаданных Fedora
- Позволило переиспользовать метаданные, храня их на диске
- Позволило реализовать создание натривиальных запросов с помощью SQL

Реализация — базы данных

packages	
pkgKey	INTEGER PRIMARY KEY
pkgId	TEXT
name	TEXT
arch	TEXT
version	TEXT
release	TEXT
description	TEXT

depends	
name	TEXT
version	TEXT
release	TEXT
pkgKey	INTEGER
pre	BOOLEAN DEFAULT FALSE

recommends	
name	TEXT
version	TEXT
release	TEXT
pkgKey	INTEGER

.....

provides	
name	TEXT
version	TEXT
release	TEXT
pkgKey	INTEGER

Реализация — API

Для конечного пользователя с помощью библиотеки Click создано консольное приложение

- Реализация CLI требует значительно меньше времени, чем написание графического интерфейса
- Простота добавления новых команд
- Удовлетворяет требованиям

```
Usage: depinspect [OPTIONS] COMMAND [ARGS]...
```

```
Options:
```

```
--help          Show this message and exit.
```

```
Commands:
```

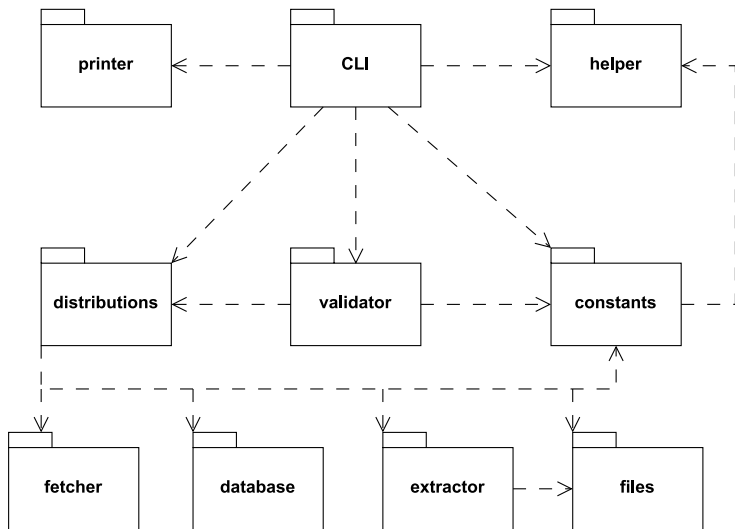
```
diff            Compare two packages.
```

```
find-divergent  List all packages that have divergent dependencies.
```

```
list-all       List stored architectures and packages for a given distro.
```

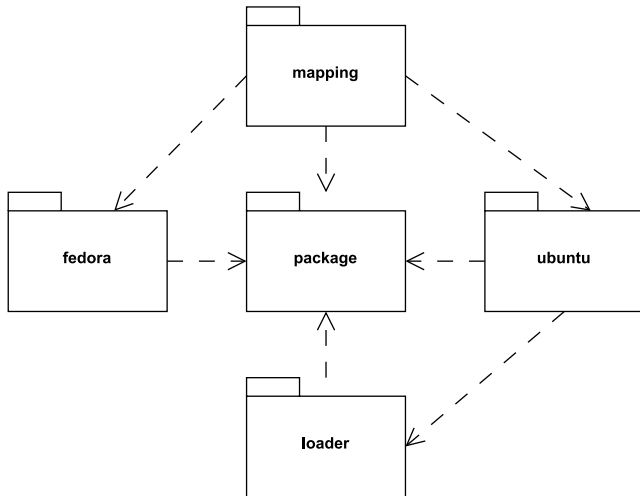
```
update         Update metadata stored in databases.
```

Реализация — архитектура



Реализация — архитектура

Детализация модуля distributions



Реализация — архитектура

Независимые компоненты:

- **CLI** — интерфейс для пользователя
- **Fetcher** — загрузка метаданных, указанных в конфигурационном файле проекта `pyproject.toml`
- **Extractor** — обработка загруженных разного рода архивов, содержащих метаданные
- **Database** — абстракция над запросами в базу данных с помощью языка запросов SQL
- **Printer** - вывод результата в консоль

Близко к архитектурному стилю «Каналы и Фильтры»

Проблема — во время поступления команды от пользователя заранее не известно, какой порядок действий нужно совершить:

- Для инициализации метаданных Ubuntu нужно выполнить целую цепочку (загрузить архивы, произвести парсинг текста, создать и наполнить базу данных)
- Для Fedora — загрузить и распаковать файлы `.sqlite`.

Решение — каждый класс знает, что с собой нужно делать

- Создан метакласс **Package**, имеющий набор полей, соответствующих именам полей метаданных, и декларирующий абстрактные методы, сопоставляемые консольным командам
- Дочерние классы **Ubuntu** и **Fedora** реализуют эти абстрактные методы по-своему

- Проведён обзор четырех инструментов, способных работать с метаданными пакетов, на основе анализа функциональности которых созданы четыре доступные для конечного пользователя команды приложения
- Внедрены метаданные пакетов Ubuntu и Fedora
- На языке Python реализовано приложение¹ с модульной архитектурой и консольным интерфейсом, позволяющее производить межархитектурный анализ зависимостей пакетов

¹Репозиторий приложения, <https://github.com/artem-burashnikov/depinspect>

Пример использования

```
depinspect diff -p ubuntu riscv64 apt -p ubuntu amd64 apt
```

These dependencies are present in both:

ubuntu - riscv64 - apt

ubuntu - amd64 - apt

=====

adduser

gpgv | gpgv2 | gpgv1

libapt-pkg6.0 (>= 2.4.5)

libc6 (>= 2.34)

libgnutls30 (>= 3.7.0)

libstdc++6 (>= 11)

libsystemd0

ubuntu-keyring

These dependencies are exclusive to:

ubuntu - riscv64 - apt

=====

libgcc-s1 (>= 3.4)

These dependencies are exclusive to:

ubuntu - amd64 - apt

=====

libgcc-s1 (>= 3.3.1)

libseccomp2 (>= 2.4.2)