

Івасюк Василь Васильович
магістрант 1 курсу, групи ІПЗм-1
Карпатський національний університет імені Василя Стефаника
+380 (342) 59-60-58
vasyl.ivasiuk.21@pnu.edu.ua

ОГЛЯД СУЧАСНОГО СТАНУ ДОСЛІДЖЕННЯ ЗАЛЕЖНОСТІ ПРОГРАМНИХ ПАКЕТІВ

Постановка задачі. Одним з показником надійності коду є якість залежностей, які прямо або опосередковано успадковуються певними пакетами. Якщо їхня кількість є надмірною, або самі залежності видають невластиву поведінку, то це прямо впливає на функціональність продукту. Ще одним важливим показником є видимість, оскільки не завжди дослідники замислюються, як сильно їм допомагають ті чи інші пакети.

Мета дослідження. Метою дослідження є огляд публікацій, які розглядають питання оцінки залежності у різних екосистемах, які були опубліковані протягом останніх 5 років; визначення основних результатів робіт.

Результати дослідження. Автори роботи [1] досліджували питання ступені залежності. Вони замітили, що залежності пакетів **P** у мові програмуванні **R** можна представити в якості орієнтованого графу, а самі пакети розподілити за категоріями:

- сильні батьківські пакети - обов'язкові для установки (Depends, Imports, LinkingTo);
- слабкі батьківські пакети - опціональні для установки (Suggests, Enhances);
- сильнозалежні пакети - такі пакети, кількість яких отримана шляхом рекурсивного пошуку їх сильних батьківських пакетів;
- всезалежні пакети - такі пакети, кількість яких отримана шляхом рекурсивного пошуку сильних та слабких батьківських пакетів;
- дочірні пакети - такі пакети, чії батьки включають **P**.

Автори ввели метрику "важкості залежності", яка визначає, скільки додаткових залежностей вводить кожний батьківський пакет. Вони склали формули для визначення залежності по кожній з категорій. Потім на цій основі був написаний пакет *pkgndep*, який був використаний для створення індивідуальної теплової карти. Стаття [2] доповнює цю тему, додатково досліджуючи поширення важкості залежності від висхідних до низхідних пакетів. Першими можна вважати сильнозалежні пакети, а другими - такі пакети, які шукають дочірні пакети шляхом рекурсивного пошуку. Пакет *pkgndep* був оновлений з врахуванням цих змін, а в якості наборів даних були використані пакети з репозиторій *CRAN* та *Bioconductor*. Визначаючи патерни важкості залежностей, автори виявили, що лише невелика частина пакетів має важкі батьківські пакети в екосистемі. При цьому, сильна залежність зазвичай формується на самому початку. Так само, кількість пакетів, які одночасно успадковують залежності від двох батьків, також невисока. Якщо це й

відбувається, то вони здебільшого мають відношення батько-дитина. Також було визначено, що середня кількість унікальних залежностей спадає водночас зі збільшенням дочірніх пакетів. Зазвичай важкість залежності передається на великі відстані. Накопичення важкості відбувається за рахунок декількох важких батьків, або коли група пакетів успадковують один важкий пакет. Для зменшення важкості автори радять напряду імплементувати необхідні функції з важкого пакету; використовувати функціонально схожі, але легкі пакети; зробити батьківські пакети, які надають лише опціональні функції, необов'язковими для установки. Хоча ступінь залежності є сама по собі важливим фактором, не менш важливим є розумінням, якими вони є і який функціонал виконують. Дослідженню опосередкованих залежностей в біомедичних програмах приділена робота [3]. Для розгляду цього питання береться ті пакети, які мають мало згадок у наукових роботах, однак які володіють високим рівнем центральності Каца (Katz centrality). Такі пакети були прозвані *Nebraska*. Було підготовлено три набори даних з репозиторій *PyPI* (Python), *CRAN* та *Bioconductor* (R). Візуалізація мереж залежностей показала, що більшість пакетів відповідають типу *Nebraska*; автори також звернули увагу на відсутність циклів у біомедичних програмах.

Висновки та перспективи. Пошук рішень, які стосуються питання залежності пакетів є нагальним серед дослідників. Одне з рішень стосується важкості залежності, яка показує, наскільки ефективно задіються імпортовані пакети і як вони взаємодіють у мережі. Важливим для ефективності є необхідність залучати лише ті функції, які реально застосовуються в проєкті, а також забезпечують його надійність. Ще однією точкою дослідження є огляд пакетів з малою видимістю та високою центральністю, які несуть велику користь для проведення наукових досліджень. Аналіз досліджень підтверджує потребу приділення більшої уваги на питання залежності програмних пакетів [4].

Список використаних джерел

1. Z. Gu and D. Hübschmann, "Pkgndep: a tool for analyzing dependency heaviness of R packages", *Bioinformatics*, vol. 38, no. 17, pp. 4248–4251, Sep. 2022, doi: 10.1093/bioinformatics/btac449
2. Z. Gu, "On the dependency heaviness of CRAN/Bioconductor ecosystem", *J. Syst. Softw.*, p. 111610, Jan. 2023, doi: 10.1016/j.jss.2023.111610
3. E. M. Brown *et al.*, "Biomedical Open Source Software: Crucial Packages and Hidden Heroes", Nov. 2025. doi: 10.1119/1.4767505
4. M. Kuz *et al.*, "Methods for evaluating software accessibility," *Radio Electronics, Computer Science, Control*, no. 3, pp. 163-172, Sept. 22, 2025, doi: 10.15588/1607-3274-2025-3-15