

МАТЕМАТИЧНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ WORDPRESS ТЕМ НА PHP

Дорогань І.Ю.¹; Пухалевич А.В., к.т.н.²

^{1,2} Національний університет кораблебудування імені адмірала Макарова

^{1,2} Україна, Миколаїв

¹ 211252@nuos.edu.ua; ² andrii.pukhalevych@nuos.edu.ua

Анотація. Було зібрано ООП метрики 51 Wordpress теми розроблених з використанням мови програмування PHP. Побудовано математичну модель для оцінювання розміру Wordpress тем на PHP. Перевірено характеристики побудованої моделі.

Ключові слова: оцінювання розміру програмного забезпечення; PHP; Wordpress теми; математична модель; нелінійна регресія.

Вступна частина. Оцінка розміру програмного забезпечення – це важлива інформація на перших етапах розробки, тому що оцінка розміру раніше розроблених проєктів використовується для оцінювання трудомісткості розроблення програмного забезпечення, та дозволяє правильно розподілити витрати й зусилля розробників [1]. Отже, оцінювання розміру тем для системи керування вмістом WordPress, яка є найпопулярнішою такою системою в світі [2], є важливою задачею програмної інженерії, вирішення якої потребує побудови відповідних моделей.

Для оцінювання розміру коду використовуються лінійні та нелінійні регресійні моделі [3]. Існуючі моделі, такі як трифакторна нелінійна регресійна модель для оцінювання розміру PHP застосунків із відкритим кодом [4], не призначені для оцінювання розміру WordPress тем, що призводить до низької достовірності відповідного оцінювання. Тому, розробка математичної моделі для оцінювання розміру WordPress тем на PHP є актуальною задачею.

Метою дослідження є підвищення достовірності оцінювання Wordpress тем на PHP за рахунок побудови відповідної математичної моделі.

Основна частина. Wordpress теми, розміщені на вебсайті Wordpress (<https://wordpress.org/themes>), були розроблені з використанням мови програмування PHP. ООП метрики 51 Wordpress тем з вказаного сайту були отримані за допомогою утиліти PhpMetrics (<https://phpmetrics.org/>). Для побудови математичної моделі було відібрано такі метрики: загальна кількість тисяч рядків коду Y , кількість класів X_1 , середня кількість методів на клас X_2 , глибина дерева успадкування X_3 . Вибір цього набору метрик пов'язаний з тим, що між ними відсутня мультиколінеарність.

Відібрані дані було перевірено на нормальність закону розподілу. Виявилось, що розподіл даних не являється гаусівським, тому необхідно провести нормалізацію даних. Для нормалізації емпіричних даних було використано десятковий логарифм.

Після нормалізації з даних було видалено одну з тем, так як відстань Махаланобіса перевищувала критичне значення.

Для 50 Wordpress тем, що залишились, було побудовано лінійну регресійну модель для нормалізованих значень:

$$\lg(Y) = b_0 + b_1 * \lg(X_1) + b_2 * \lg(X_2) + b_3 * \lg(X_3) + \varepsilon,$$

де b_0, b_1, b_2, b_3 – це оцінки параметрів, ε – випадкова величина.

Методом найменших квадратів було отримано наступні значення оцінок параметрів лінійної регресії $b_0 = -1,54$, $b_1 = 0,97$, $b_2 = 1,05$, $b_3 = 0,19$.

Перевірка за критерієм Пірсона показала, що розподіл випадкової величини ε є гаусівським, так як значення $\chi_2 = 6,07$ є меншим, ніж критичне значення $\chi_{2кр} = 9,49$ для 4 ступенів вільності та рівня значущості 0,05.

Нелінійна регресійна модель для оцінювання розміру Wordpress тем у тисячах рядків коду в залежності від факторів X_1 , X_2 та X_3 :

$$Y = 10^{b_0 + b_1 * \lg(X_1) + b_2 * \lg(X_2) + b_3 * \lg(X_3) + \varepsilon}.$$

Після перетворення, нелінійна регресійна модель є такою:

$$Y = 10^{\varepsilon + b_0} X_1^{b_1} X_2^{b_2} X_3^{b_3}.$$

Побудована нелінійна регресійна модель була перевірена за множинним коефіцієнтом детермінації R^2 , середньою величиною відносної помилки MMRE і відсотком оцінок кількості рядків коду, для яких величина модулю відносної помилки не більша за 0,25 (PRED(0,25)). Допустимі значення MMRE і PRED(0,25) складають не більше 0,25 і не менше 0,75 відповідно. Значення R^2 , MMRE і PRED(0,25) для побудованої моделі дорівнюють відповідно 0,92, 0,11 і 0,90. Це свідчить про високу якість отриманої моделі.

Висновки. В результаті роботи було побудовано математичну модель, для оцінювання розміру Wordpress тем на PHP. Побудована модель має високі характеристики R^2 , MMRE, PRED(0,25), що дозволяє оцінювати розмір Wordpress тем з високою достовірністю. В майбутньому запланована розробка програмного забезпечення, яке реалізує побудовану модель для оцінювання розміру Wordpress тем.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Вікіпедія. (н.д.). Оцінка складності розробки ПЗ. https://uk.wikipedia.org/wiki/Оцінка_складності_розробки_ПЗ
- [2] W3Techs. (n.d.). Usage statistics and market share of WordPress. Retrieved October 19, 2022, from <https://w3techs.com/technologies/details/cm-wordpress>
- [3] Tan, H.B.K., Zhao, Y., & Zhang, H. (2009). Conceptual data model-based software size estimation for information systems. Transactions on Software Engineering and Methodology, vol. 19, issue 2, article No. 4.
- [4] Приходько, С.Б., & Ворона, М.В. (2021). Оцінювання розміру PHP-застосунків з відкритим кодом за нелінійними регресійними моделями з різними факторами. *Збірник наукових праць Національного університету кораблебудування імені адмірала Макарова*, 1(484)/2021, 92-98. DOI: [https://doi.org/10.15589/znp2021.1\(484\).13](https://doi.org/10.15589/znp2021.1(484).13)

Dorohan I.Y., Pukhalevych A.V.

A MATHEMATICAL MODEL FOR ESTIMATING THE SIZE OF WORDPRESS THEMES IN PHP

Abstract. The OOP metrics of 51 Wordpress themes developed using the PHP programming language were collected. A mathematical model was built for estimating the size of Wordpress themes in PHP. The characteristics of the built model were checked.

Keywords: software size estimation; PHP; Wordpress themes; mathematical model; non-linear regression.