

МЕТОД ДІЯЛЬНОСТІ ТА РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОГО АГЕНТА НА ОСНОВІ ОНТОЛОГІЧНОГО ПІДХОДУ ДЛЯ ПАРСИНГУ ПРИРОДОМОВНИХ СПЕЦИФІКАЦІЙ ВИМОГ ДО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Метою даного дослідження є автоматизація семантичного парсингу природомовної специфікації, зокрема, автоматизованого пошуку атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ шляхом розроблення інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до ПЗ на предмет пошуку атрибутів для визначення нефункційних характеристик-складових якості ПЗ.

Ключові слова: інтелектуальний агент на основі онтологічного підходу, специфікація вимог до програмного забезпечення (ПЗ), парсинг природомовних специфікацій вимог до ПЗ, нефункційні характеристики-складові якості ПЗ.

O.O. PAVLOVA, M.A. BODNAR, YE.H. HNATCHUK

Khmelnytskyi National University

METHOD OF FUNCTIONING AND REALIZATION OF ONTOLOGY-BASED INTELLIGENT AGENT FOR PARSING THE NATURAL-LANGUAGE SOFTWARE REQUIREMENTS SPECIFICATIONS

The paper is devoted to development of ontology-based intelligent agent for parsing the natural-language software requirements specifications. The paper develops a method of functioning, structure and realization of ontology-based intelligent agent for parsing the natural-language software requirements specifications. The simplicity of the method of functioning the intelligent agent provides high-speed parsing of natural-language specifications. The developed intelligent agent makes it possible to perform the analysis of natural-language specifications to determine the presence or absence of attributes needed to determine the non-functional characteristics-components of software quality. As a result of such parsing, the proposed intelligent agent determines the number and percentage of missing attributes, displays which attributes are missing for a particular subcharacteristics of non-functional characteristics, and also forms a real ontology for non-functional characteristics. The results of the operation of the intelligent agent are then used to assess the sufficiency of information (attributes) to determine the non-functional characteristics-components of software quality, and can also be used as recommendations for developers of specifications (requirements' engineers) with the purpose of providing the maximum quantity of attributes, which are necessary to determine the non-functional characteristics, in the specifications of software requirements.

Keywords: ontology-based intelligent agent, software requirements specification, parsing the natural-language software requirements specifications, non-functional characteristics-components of software quality.

Вступ

У сучасному світі розробка програмного забезпечення (ПЗ) – одна із найдорожчих індустрій. На сьогодні в світі витрачається більше 250 млрд USD щорічно на розроблення приблизно 175 тис. програмних проектів, при цьому значна кількість (в середньому до 70%) програмних проектів є неуспішними (в т. ч. 50% – проблемні проекти, 20% – провальні проекти) [1]. Тому, однією із основних вимог користувачів до сучасного ПЗ є його висока якість (якість ПЗ – це його здатність задовольнити заявлені і передбачувані потреби при використанні за певних умов [2, 3]). Суттєва кількість помилок (10–23% [4], до 56% [5]) вноситься у ПЗ на етапі формування вимог до ПЗ, причому вартість виправлення дефектів програмного забезпечення зростає експоненційно з кожним наступним етапом життєвого циклу програмного забезпечення [5]. Близько 50% дефектів вимог є результатом незрозумілих чи неоднозначних вимог; інші 50% – це результат незавершеності або недостатності специфікацій (неповні, недостатні або пропущені вимоги) [5]. Враховуючи цей факт, важливою задачею є саме визначення достатності інформації у специфікації вимог, зокрема, наскільки повно у специфікації відображена інформація щодо функцій та обмежень майбутнього ПЗ.

У [6] розроблено теоретичні засади інформаційної технології оцінювання достатності інформації щодо якості у специфікаціях вимог до ПЗ, але обмеженням такого рішення є його теоретична орієнтованість – і пошук інформації у специфікації, і оцінювання її достатності проводиться вручну. У [7] розроблено інтелектуальний агент на основі онтологічного підходу для оцінювання початкових етапів життєвого циклу ПЗ, призначений для аналізу специфікацій вимог до ПЗ на предмет достатності їх інформації щодо нефункційних характеристик-складових якості ПЗ (функційна придатність, ефективність, надійність, сумісність, зручність використання, можливість переносу, супроводжуваність, захищеність – відповідно до ISO 25010:2011). Але недоліком такого рішення є те, що на сьогодні вибір інформації щодо нефункційних характеристик зі специфікації вимог до реального ПЗ реалізується вручну, хоча необхідним є автоматизований семантичний розбір природомовної специфікації на предмет пошуку атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ.

На сьогодні відомі різні підходи до семантичного аналізу специфікацій. У [8] запропоновано методику, що забезпечує автоматичну допомогу розробникам шляхом трансформації природомовних вимог за допомогою UML-діаграм. Стаття [9] присвячена вирішенню проблеми перетворення бізнес-специфікацій, написаних природною мовою, у формальні моделі, придатні для використання при розробці інформаційних

систем. Методологія, запропонована у [10], є ефективною для мінімізації неоднозначності вимог. В роботі [11] запропоновано підхід до автоматичного видобування семантичної інформації зі специфікацій вимог до ПЗ шляхом поєднання методів маркування семантичної ролі та моделювання знань предметної галузі, який є ефективним і надійним для моделювання мінливості та спільності функційних вимог, але не підходить для роботи з нефункційними вимогами. У [12] розроблено метод для налаштування і створення комбінованого пар-сепа (аналізатора) для обробки та аналізу природомовних специфікацій, який дозволяє побудувати ефективний синтаксичний аналізатор для аналізу природомовних специфікацій вимог до промислового ПЗ. У [13] представлено онтологічний підхід до автоматизованої перевірки та вимірювання вимог до ПЗ, який використовується для виявлення невідповідностей, непослідовності та недоліків вимог до ПЗ. Автори [14] представляють прототип семантичної системи, що використовується для надання допомоги інженерам з вимог щодо видобування вимог з використанням напівформального представлення. У дослідженні [15] запропоновано структуру для інтеграції різнорідних вимог за допомогою онтологій, яка фокусується на мінімізації впливу неоднорідності вимог на якість ПЗ.

Отже, на сьогодні не вирішеною і актуальною є проблема аналізу специфікацій вимог до ПЗ на предмет відповідності нефункційних вимог потребам замовника (зокрема, достатності їх інформації) з метою підвищення успішності програмних проектів, розроблених за специфікаціями. Для вирішення такої проблеми слід розв'язати задачу автоматизації семантичного парсингу природомовної специфікації, зокрема, автоматизованого пошуку атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ. Як показав проведений аналіз, всі розглянуті методи семантичного аналізу специфікацій вимог до ПЗ не забезпечують перевірки відповідності нефункційних вимог специфікації потребам замовника (валідації), тому не спрямовані на підвищення успішності розробленого за специфікацією ПЗ. Невирішеність цієї задачі обумовлює необхідність розроблення інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до ПЗ на предмет пошуку атрибутів для визначення нефункційних характеристик-складових якості ПЗ, що і буде *метою даного дослідження*.

Метод діяльності та структура інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення

Інтелектуальний агент для парсингу природомовних специфікацій розробляється на основі онтологічного підходу. В якості відомих знань інтелектуальний агент використовує базу онтологію предметної галузі «Інженерія програмного забезпечення» (частина «Якість ПЗ»), розроблену у [16]. Інтелектуальний агент на основі онтологічного підходу для парсингу природомовних специфікацій приймає на вхід природомовну специфікацію вимог до ПЗ та проводить автоматичний аналіз вимог на предмет пошуку атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ.

Метод діяльності інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення складається з наступних етапів:

1) пошук кожного атрибута з базової онтології специфікації вимог (така онтологія розроблена у [16] та міститься у базі знань агента) у специфікації вимог до реального ПЗ (вважаємо, що специфікація формалізована та відповідає вимогам стандарту ISO 29148:2018);

2) якщо <атрибут_{*i*}> знайдено у специфікації вимог, то <атрибут_{*i*}> заноситься у множину наявних атрибутів, $i=1..138$ (оскільки, згідно зі стандартом ISO 25023:2016, є 138 різних атрибутів, від яких залежать нефункційні характеристики-складові якості ПЗ);

3) якщо <атрибут_{*i*}> не знайдено у специфікації вимог, то <атрибут_{*i*}> заноситься у множину відсутніх атрибутів, $i=1..138$;

4) з базової онтології для нефункційних характеристик-складових якості ПЗ, розробленої у [16], видаляються всі атрибути з множини відсутніх атрибутів;

5) відбувається перевірка, чи всі атрибути з множини наявних атрибутів залишилися у онтології після її модифікації на попередньому кроці;

6) відбувається збереження внесених змін – створення реальної онтології для нефункційних характеристик-складових якості ПЗ.

Результатом роботи розробленого інтелектуального агента є реальна онтологія для нефункційних характеристик-складових якості ПЗ, яка є вхідними даними для інтелектуального агента на основі онтологічного підходу для оцінювання початкових етапів життєвого циклу ПЗ, розробленого у [7]. Як додаткові результати функціонування агента для подальшої роботи можуть бути використані також множини наявних та відсутніх атрибутів у реальній специфікації вимог до ПЗ.

Структурна схема інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення представлена на рис. 1.

Реалізація інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення

Інтелектуальний агент для парсингу природомовних специфікацій реалізований мовою PHP в формі частини вільно поширюваного програмного забезпечення, доступного за посиланням – <https://olp-project.herokuapp.com>.

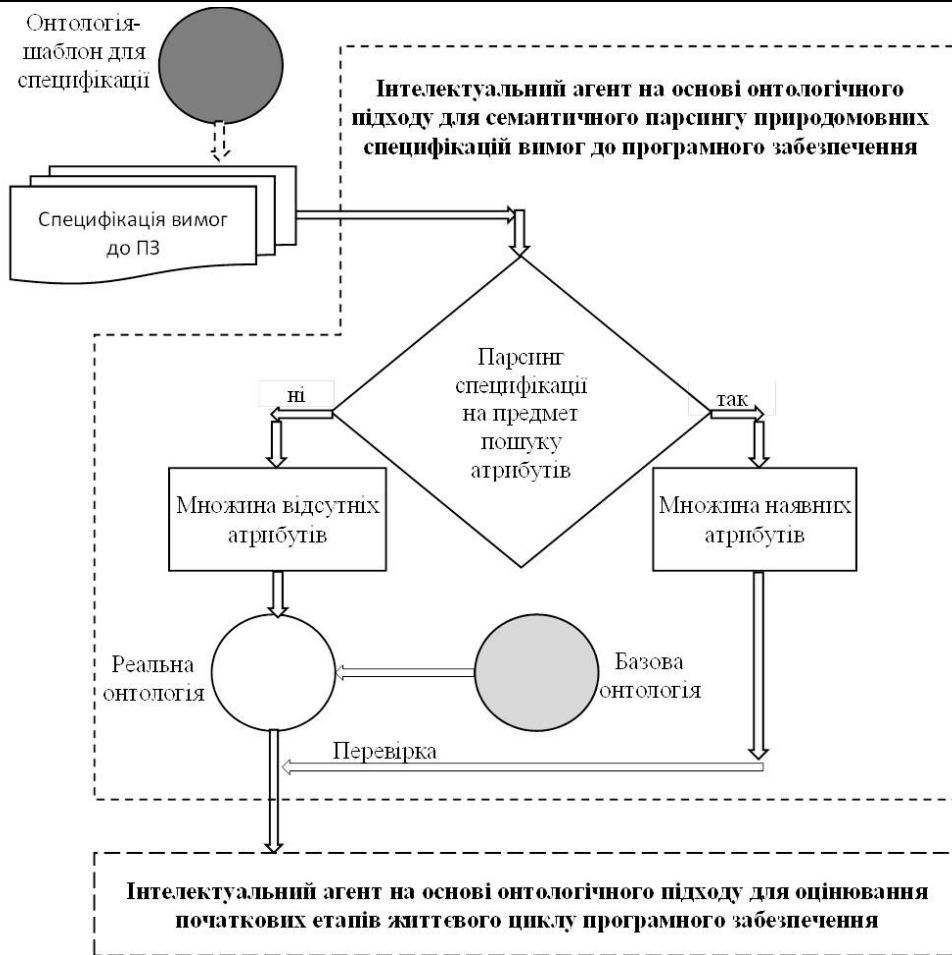


Рис. 1. Структура інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення

На вхід розробленого інтелектуального агента користувач повинен подати природомовну специфікацію вимог до ПЗ в pdf-форматі. Інтелектуальний агент проводить парсинг специфікації. Результатами роботи агента є: кількість відсутніх атрибутів (рис. 2), візуалізація відсутніх атрибутів з розподілом за підхарактеристиками нефункційних характеристик-складових якості, для визначення яких вони необхідні (рис. 3), а також реальна онтологія для нефункційних характеристик-складових якості в owl-форматі. На рис. 2 – рис. 3 представлено результати роботи інтелектуального агента на основі онтологічного підходу на прикладі парсингу природомовної специфікації вимог до програмної системи для підвищення безпеки програмного забезпечення комп'ютерних систем, розробленої колективом науковців Хмельницького національного університету в рамках ДБТ 1Б-2019.

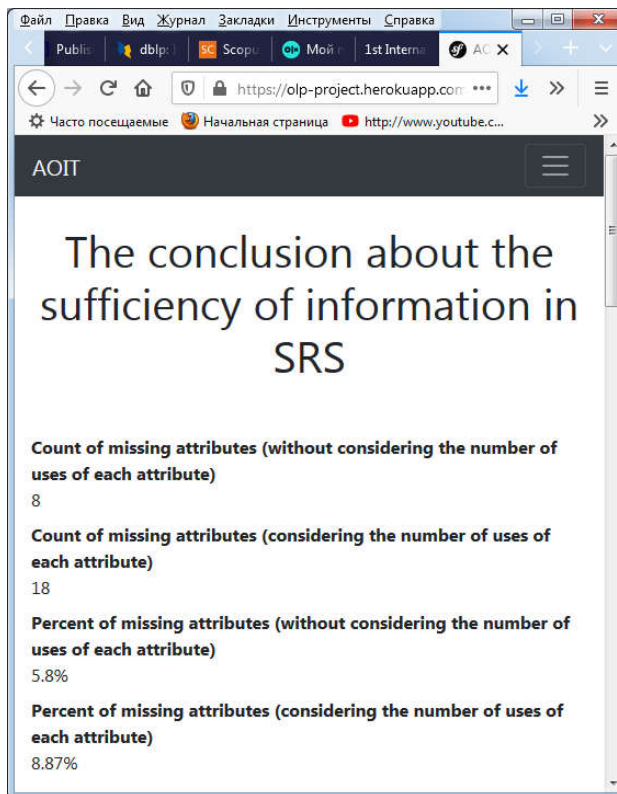


Рис. 2. Результати діяльності інтелектуального агента на основі онтологічного підходу на прикладі парсингу природомовної специфікації вимог до програмної системи для підвищення безпеки програмного забезпечення комп'ютерних систем: кількість та відсоток відсутніх атрибутів

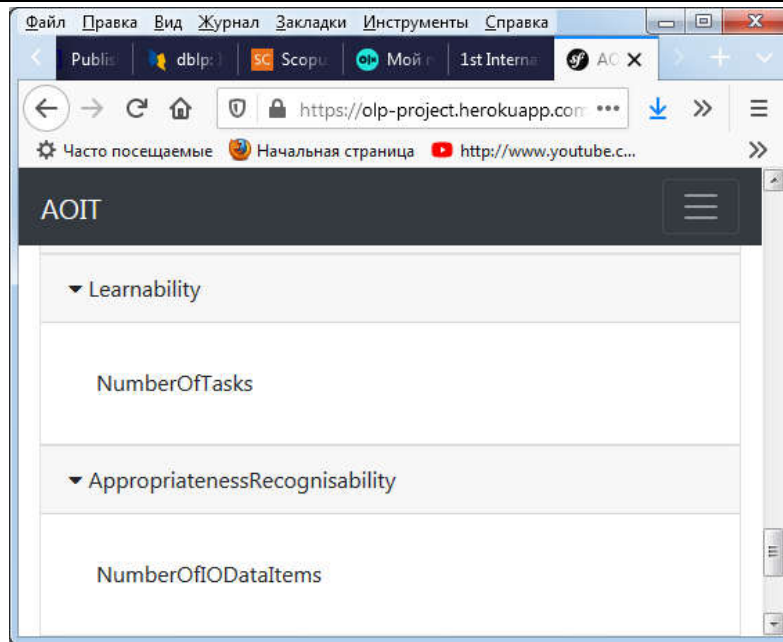


Рис. 3. Результати діяльності інтелектуального агента на основі онтологічного підходу на прикладі парсингу природомовної специфікації вимог до програмної системи для підвищення безпеки програмного забезпечення комп'ютерних систем: відсутні атрибути для підхарактеристик: розпізнавання доцільності та можливість вивчення характеристики, зручність використання

Висновки

Для автоматизації семантичного розбору природомовної специфікації з метою перевірки відповідності її нефункційних вимог потребам замовника необхідно виконати її формалізацію, наприклад, з використанням онтологій. Для такої формалізації було запропоновано використати раніше розроблені онтологію для специфікацій вимог до ПЗ та онтологію для нефункційних характеристик-складових якості ПЗ, які стали основою (відомими фактами) інтелектуального агента на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення на предмет пошуку атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ.

У статті розроблено метод діяльності, структуру та реалізовано інтелектуальний агент на основі онтологічного підходу для парсингу природомовних специфікацій вимог до програмного забезпечення. Простота методу діяльності інтелектуального агента забезпечує високу швидкість парсингу природомовних специфікацій. Розроблений інтелектуальний агент дає можливість виконувати розбір природомовних специфікацій на предмет встановлення наявності чи відсутності атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ. В результаті такого парсингу пропонується інтелектуальний агент визначає кількість та відсоток відсутніх атрибутів, відображає, яких атрибутів не вистачає для тієї чи іншої підхарактеристики нефункційної характеристики, а також формує реальну онтологію для нефункційних характеристик. Результати функціонування інтелектуального агента далі використовуються для оцінювання достатності інформації (атрибутів) для визначення нефункційних характеристик-складових якості, а також можуть бути використані як рекомендації для розробників специфікацій (інженерів з вимог) з метою забезпечення наявності у специфікаціях вимог до ПЗ якомога більшої кількості атрибутів, необхідних для визначення нефункційних характеристик-складових якості ПЗ.

References

1. Shane H. Standish Group 2015 Chaos Report – Q&A with Jennifer Lynch [Electronic resource] / H. Shane, S. Wojewoda. – Access mode: <http://www.infoq.com/articles/standish-chaos-2015>
2. ISO/IEC 25010:2011. Systems and software engineering. Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models. [Introduced 01.03.2011]. – Geneva (Switzerland), 2011. – 34 p. (International standard).
3. ISO/IEC TR 19759:2015. Software Engineering. Guide to the software engineering body of knowledge (SWEBOOK). [Introduced 01.10.2015]. – Geneva (Switzerland), 2015. – 336 p. (International standard).
4. McConnell S. Code complete / S. McConnell. – Redmond: Microsoft Press, 2013. – 896 p.
5. Leveraging Natural Language Processing in Requirements Analysis: How to eliminate over half of all design errors before they occur [Electronic resource] – Access mode: <http://qracorp.com/wp-content/uploads/2017/03/Leveraging-NLP-in-Requirements-Analysis.pdf>
6. Hovorushchenko T. Information technology of evaluating the sufficiency of information on quality in the software requirements specifications / T. Hovorushchenko, O. Pomorova. – CEUR-WS. – 2018. – Vol. 2104. – P. 555–570.
7. Hovorushchenko T. Method of Activity of Ontology-Based Intelligent Agent for Evaluating the Initial

Stages of the Software Lifecycle / T. Hovorushchenko, O. Pavlova. – Advances in Intelligent Systems and Computing. – 2019. – Vol. 836. – P. 169–178.

8. Gulia S. An Efficient Automated Design to Generate UML Diagram From Natural Language Specifications / S. Gulia, T. Choudhury. – The 6-th International Conference on Cloud System and Big Data Engineering: Proceedings (Noida, January 14-15, 2016). – Noida (India), 2016. – P. 641–648.

9. Formalising natural language specifications using a cognitive linguistic/configuration based approach. Information Systems / M. Selway, G. Grossman, W. Mayer, M. Stumptner M. – 2015. – Vol. 54. – P. 191–208.

10. Ali S. W. Process to Enhance the Quality of Software Requirement Specification Document / S.W. Alim Q. A. Ahmed, I. Shafi. – The International Conference on Engineering and Emerging Technologies: Proceedings (Lahore, February 22-23, 2018). – Lahore (Pakistan), 2018. – P. 113–118.

11. Wang Y. Semantic Information Extraction for Software Requirements using Semantic Role Labeling / Y. Wang. – The IEEE International Conference on Progress in Informatics and Computing: Proceedings (Nanjing, December 18-20, 2015). – Nanjing (China), 2015. – P. 332–337.

12. Iwama F. Constructing Parser for Industrial Software Specifications Containing Formal and Natural Language Description / F. Iwama, T. Nakamura, H. Takeuchi. – The 34-th International Conference on Software Engineering: Proceedings (Zurich, June 02-09, 2012). – Zurich (Switzerland), 2012. – P. 1012–1021.

13. Siegemund K. Contributions to Ontology-Driven Requirements Engineering: dissertation / K. Siegemund. – Dresden, 2014. – 249 p.

14. Ontology-Driven Guidance for Requirements Elicitation / S. Farfeleder, T. Moser, A. Krall, T. Stalhane, I. Omoroniya, H. Zojer. – Lecture Notes in Computer Science. – 2011. – Vol. 6644. – P. 212–226.

15. Integration of Heterogeneous Requirements using Ontologies / A. Mustafa, W. M. N. Wan-Kadir, N. Ibrahim, A. Shah, M. Younas. – International Journal of Advanced Computer Science and Applications. – 2018. – Vol. 9. – No. 5. – P. 213–218.

16. Hovorushchenko T. Models and methods of evaluation of information sufficiency for determining the software complexity and quality based on the metric analysis results / T. Hovorushchenko. – Central European Researchers Journal. – 2016. – Vol. 2. – P. 42–53.

Рецензія/Peer review : 15.05.2020 р.

Надрукована/Printed : 16.6.2020 р.

Рецензент: д.т.н., проф. О.С. Савенко