УДК 004.891.3 DOI: 10.30837/0135-1710.2021.177.069

С.Ф.ЧАЛИЙ, В.О. ЛЕЩИНСЬКИЙ, І.О. ЛЕЩИНСЬКА

ПРЕДСТАВЛЕННЯ УЗГОДЖЕНИХ ЗНАНЬ З УРАХУВАННЯМ ЇХ ЛОГІЧНОЇ НЕСУПЕРЕЧЛИВОСТІ ДЛЯ ЗАДАЧІ ПОБУДОВИ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ

Розроблено узагальнений підхід до узгодження знань для побудови пояснень в інтелектуальних інформаційних системах. Підхід забезпечує можливість формування множини узгоджених пояснень, що відображають різні аспекти процесу прийняття рішення та отриманого інтелектуальною системою результату, що спрощує застосування цих результатів у предметній області користувача. Запропоновано модель представлення узгоджених знань з урахуванням їх логічної несуперечливості. Модель дає можливість перевірити узгодженість знань безпосередньо при формуванні пояснень, що створює умови для формування пояснень за принципом чорного ящику та дає можливість доповнити функціонуючі системи можливостями пояснень, а також адаптувати пояснення при зміні їх функціональності.

1. Вступ

Потреба у ефективному використанні результатів роботи інформаційних систем при вирішенні погано формалізованих задач обумовлює важливість використання пояснень щодо процесу прийняття рішення в таких системах [1]. Формування пояснень є вельми актуальним для інтелектуальних інформаційних систем, які широко використовують методи машинного навчання, що не є «прозорими» для користувача. Непрозорість алгоритмів прийняття рішення призводить до виникнення проблеми практичного застосування отриманих у таких системах результатів. Дана проблема виникає внаслідок протиріччя між можливостями побудови ефективних рішень з використанням складних, «непрозорих» алгоритмів та обмеженими можливостями розуміння користувачем каузальних зв'язків, що визначають та обґрунтовують процес прийняття таких рішень. Вказане протиріччя знижує довіру користувача до отриманого рішення й ускладнює застосування отриманих результатів для вирішення практичних задач у предметній області користувача.

З метою вирішення вказаної проблеми і підвищення довіри користувача до результатів роботи інтелектуальних систем запропоновано концепцію інтеграції пояснень у процес роботи таких систем - XAI (Explainable Artificial Intelligence) [2]. Згідно з даною концепцією, пояснення представляє собою раціональне тлумачення результату роботи інтелектуальної системи та процесу його досягнення. Пояснення має обґрунтовувати коректність та відповідність рішення інформаційної системи завданням користувача. Для вирішення цієї задачі пояснення має відображати ключові причинно-наслідкові зв'язки, що привели до отриманого рішення. Сукупність таких каузальних залежностей дає можливість користувачеві не лише розуміти причини отриманого рішення з урахуванням властивостей вхідного набору даних, але й порівняти це рішення з його альтернативами з тим, щоб вибрати найкращий результат. Використання обмеженої кількості варіантів рішення з можливістю їх порівняння спрощує користувачеві вибір та обґрунтування отриманого результату [3].

Для побудови пояснень використовуються два альтернативних підходи: інтеграція пояснень у процес прийняття рішення на етапі проектування відповідної інформаційної системи або ж доповнення існуючої системи модулем пояснення [4-6]. В першому випадку використовуються ті ж множини вхідних даних та знань, що були використані для побудови рішення в інформаційній системі. Даний підхід дає можливість детально пояснити процес прийняття рішення, однак модуль пояснення потрібно передбачити при створенні інтелектуальної інформаційної системи. Необхідність удосконалення даного модулю може потребувати перебудови інформаційної системи в цілому. Другий підхід передбачає доповнення існуючих систем модулем пояснення. В даному випадку може бути використана підмножина вхідних даних, що використовується для прийняття рішення, а також додаткові дані, що відображають особливості предметної області та ключові події процесу прийняття рішення. Останній підхід забезпечує більшу гнучкість та можливість удосконалення пояснень у

порівнянні з першим підходом, однак при його використані виникає проблема узгодження знань. Концепція узгодження знань була запропонована в роботах [7-9]. Властивості узгоджених знань з урахуванням ймовірнісного аспекту розглянуто в роботах [10, 11]. Підхід до узгодження пояснення із знаннями, що використовуються для прийняття рішення в інтелектуальній системі, запропоновано в [12]. Сутність узгодження знань полягає у визначенні підмножини таких знань, які б не мали протиріч та аномалій при використанні альтернативних варіантів опису предметної області. Ключова перевага узгодження знань полягає у можливості використовувати неповні та неточні знання. Це дає можливість сформувати пояснення щодо результату роботи інформаційної системи, який відповідає знанням користувача щодо предметної області.

Однак існуючі підходи до узгодження знань орієнтовані на статичний, фіксований опис знань. В той же час при побудові пояснення необхідно враховувати можливості уточнення та поповнення знань, що використовуються при виконанні процесу прийняття рішення у інформаційній системі. Усунення даного недоліку потребує розробки та реалізації узагальненого підходу до узгодження знань при вирішенні задачі побудови пояснень в інтелектуальних інформаційних системах.

2. Мета і задачі дослідження

Метою роботи є розробка моделі узгоджених знань щодо процесу прийняття рішення в інтелектуальній системі, яка враховувала б їх логічну несуперечливість та забезпечувала можливість формування пояснення процесу та результату роботи такої системи.

Досягнення мети дослідження потребує вирішення таких задач:

- розробка узагальненого підходу до узгодження знань для побудови пояснень в інтелектуальних інформаційних системах;
- розробка моделі представлення узгоджених знань з урахуванням їх логічної несуперечливості.

3. Розробка узагальненого підходу до узгодження знань для побудови пояснень в інтелектуальних інформаційних системах

Визначимо ключові властивості узгоджених даних та знань, що використовуються для побудови пояснень, з урахуванням представленого у роботі [10] опису загальних властивостей структурованих знань. Згідно з даним підходом, система структурованих знань є узгодженою у тому випадку, коли кожне судження є наслідком решти знань, а також спричиняє інші знання у цій системі. Дане визначення охоплює не лише несуперечливість та повноту знань, але й можливість виводу із одних суджень інших в рамках єдиної системи знань, а також формування пояснень щодо отриманих в результатів виводу нових суджень. Базуючись на підході [10], узгодження знань для побудови пояснень розглянемо у аспектах логічної та ймовірнісної несуперечливості, в аспекті виводу на несуперечливих знаннях, а також в аспекті пояснень зв'язків між елементами єдиної системи узгоджених знань.

Перший аспект враховує неповноту знань про функціонування інтелектуальної системи, що потребує використання ймовірнісного підходу.

Другий аспект пов'язаний із формуванням нових узгоджених знань в результаті виводу. Процес виводу задає складні зв'язки між знаннями, тобто визначає цілісність системи узгоджених знань. Перші три характеристики узгоджених знань визначають їх формальну, логічну та ймовірнісну складову. Елементи системи знань можуть бути зв'язані, наприклад, каузальними або темпоральними відношеннями.

Такі залежності внаслідок неповноти знань можуть мати ймовірнісні характеристики [13]. Врахування цих характеристик дає можливість сформувати систему знань, на якій забезпечується логічний та ймовірнісний вивід. Тобто така система знань буде послідовно зв'язною, що дає можливість виконати формальне узгодження знань. Однак формальна узгодженість не враховує семантику цих знань, що може привести до неможливості їх практичного застосування.

Для того, щоб врахувати семантику, необхідно пояснити зв'язки між елементами знань, тобто задати для цих знань додаткове відношення пояснення. Дане відношення забезпечує семантичну зв'язність системи узгоджених знань. Кожен елемент такої системи може бути зв'язаний з іншим відношенням пояснення.

Таким чином, узгоджені знання для побудови пояснень в інтелектуальній інформаційній системі мають характеризуватись такими властивостями:

- логічною несуперечливістю фактів та гіпотез щодо каузальних зв'язків між цими фактами у складі пояснення; каузальність може розглядатись у широкому сенсі, в тому числі з урахуванням ймовірнісних підходів до визначення причинно-наслідкових зв'язків;

- сумісністю фактів та гіпотез у ймовірнісному аспекті, що передбачає логічну несуперечливість елементів знань із найбільшою ймовірністю їх застосування при побудові пояснення;

- можливістю ймовірнісного виводу пояснення на узгоджених в логічному та ймовірнісному аспектах знаннях; результати такого виводу мають ймовірнісну оцінку, фінальне пояснення відбирається за найбільшим значенням ймовірності;

- відсутністю аномальних залежностей для знань, що використовуються при побудові пояснень; аномальність в даному випадку розглядається як неможливість отримати пояснення щодо існуючої закономірності.

Слід зазначити, що аномалії в знаннях можуть бути виявлені шляхом співставлення результатів ймовірнісного виводу для різних властивостей предметної області при використанні однакових вхідних даних. Аномалія виникає у випадку, якщо результати такого виводу приводять до взаємовиключних висновків. Наприклад, якщо результати першого виводу свідчать про популярність рекомендованого товару серед користувачів рекомендаційної системи, а результати другого - про недостатню надійність цього товару.

У відповідності до першої, другої та третьої властивостей, пояснення фактів та гіпотез може бути забезпечено шляхом ймовірнісного виводу на узгоджених знаннях. Такий вивід дає можливість порівняти найбільш ймовірні гіпотези, що враховують не лише поточні факти, але й базові знання з предметної області. Отримані гіпотези мають бути перевірені з урахуванням відношення пояснення. Тобто кожна з гіпотез має бути пояснена через іншу гіпотезу.

Таким чином, можливість пояснити результати як традиційного логічного, так і ймовірнісного виводу є невід'ємною властивістю узгоджених знань. Це означає, що узгодження знань є необхідною умовою для побудови модулю пояснень, що доповнює інтелектуальну систему. Узгодження знань дає можливість використати принцип чорного ящику при побудові пояснень. Згідно з цим принципом, для формування тлумачень можуть бути використані не лише вхідні дані та знання, які використовує інтелектуальна система при прийнятті рішень, а й додаткові знання про предметну область, які спрощують розуміння принципів та алгоритмів роботи цієї системи для користувача [14].

Структуру представлення узгоджених знань для побудови пояснень в інтелектуальних системах представлено на рис. 1. Згідно з розглянутими властивостями таких знань, погодження проводиться на двох рівнях: безпосередньо на рівні знань та на рівні пояснень щодо цих знань.

На першому рівні окремо узгоджуються факти даних, подій та результатів. При використанні даних вибирається така їх підмножина, для якої можуть бути погоджені факти виникнення цих даних. Для подій враховуються не лише каузальні, але й темпоральні залежності. Узгодженість результатів досягається шляхом співставлення фактів отримання результатів, які відображають різні властивості отриманого рішення. Таким чином, на першому рівні узгоджуються однотипні елементи знань.

На другому рівні виконується пояснення як однотипних елементів знань (факти подій процесу прийняття рішення та зв'язки між цими фактами), так і елементів із різних типів знань. Пояснення подій формується на базі як окремих залежностей, так і отриманої в результаті виводу послідовності таких залежностей.

Пояснення, згідно з запропонованим підходом, має дуальні властивості:

- відношення пояснення використовується для перевірки узгодженості знань щодо процесу прийняття рішення в інтелектуальній інформаційній системі;

- пояснення формується для користувача інтелектуальної системи з використанням узгоджених знань на основі вхідних даних такої системи, а також інформації про події, що відображають послідовність прийняття рішення в такій системі; узгодження в даному випадку грає роль обмеження при представленні пояснень користувачеві.

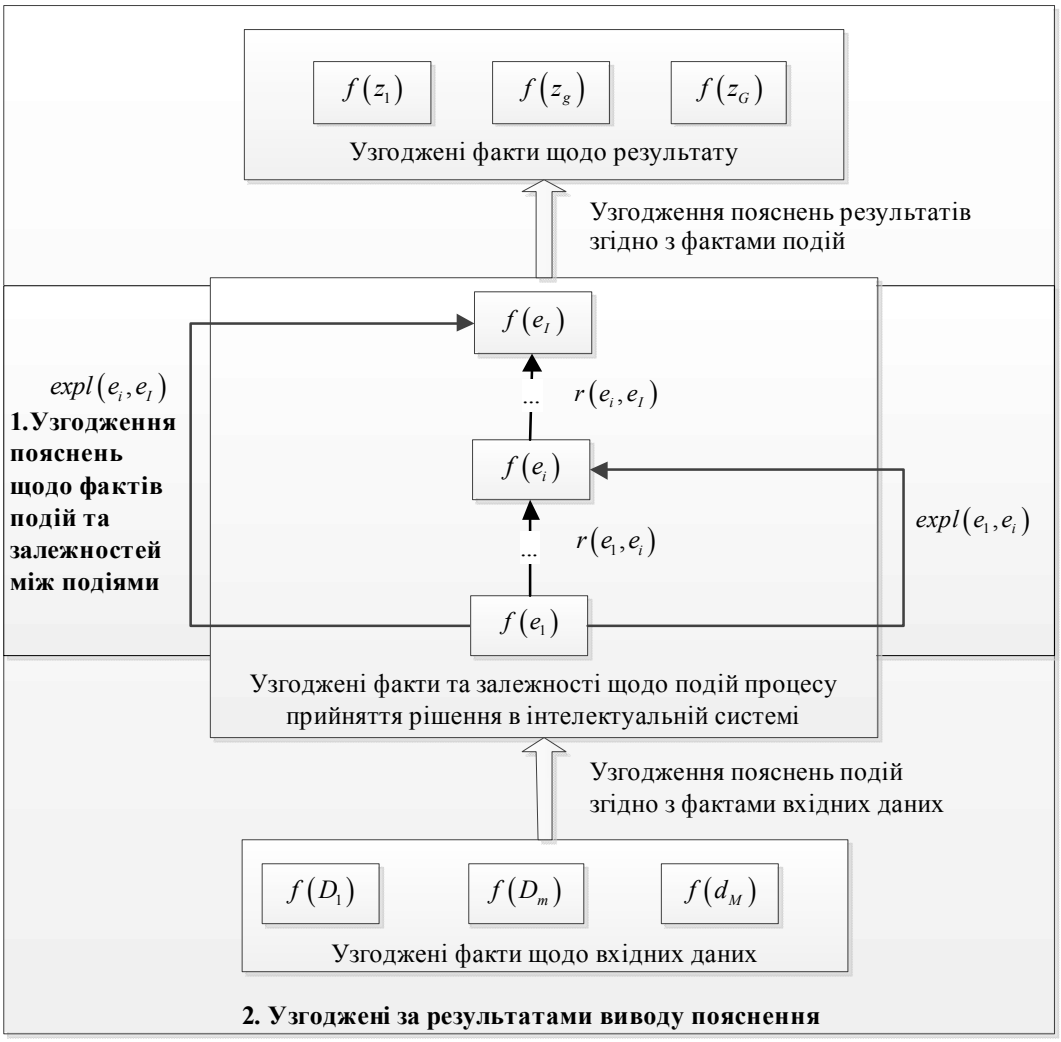

Рис. 1. Представлення узгоджених знань для побудови пояснень в інтелектуальних системах

Таким чином, узгодження знань використовуються як при формуванні цих знань, так і при побудові пояснення для користувача. Тому при постановці задачі побудови пояснення необхідно врахувати такі рівні узгодження знань:

- у логічному аспекті та в аспекті пояснення;
- у ймовірнісному аспекті та в аспекті пояснення.

У даному дослідженні головну увагу приділено логічному аспекту та аспекту пояснення у представленні узгоджених знань.

4. Розробка моделі представлення узгоджених знань з урахуванням їх логічної несуперечливості

Постановка задачі побудови тлумачення полягає у знаходженні такого упорядкованого набору пояснень, який би задовольняв умові узгодженості вхідних даних, знань та поточного або кінцевого результату роботи інтелектуальної інформаційної системи безпосередньо в аспекті пояснень:

$$\begin{array}{l} \textit{Знайти } Expl(D, E, Z) \\ \textit{при обмеженнях} \\ (\forall i) Expl_i(D) \hat{=} Expl_i(E) \hat{=} Expl_i(Z) \\ (\forall i \forall k) Expl_i(D_m) \hat{=} Expl_k(D_m), \\ Expl_i(e_j, e_l) \hat{=} Expl_k(e_j, e_l), \\ Expl_i(z_g) \hat{=} Expl_k(z_g) \end{array}, \quad (1)$$

де $Expl(D,E,Z)$ - множина пояснень щодо вхідних даних, подій та рішення інтелектуальної інформаційної системи; $Expl_i(D)$ - пояснення щодо вхідних даних; $Expl_i(E)$ - пояснення щодо подій процесу прийняття рішення і інтелектуальній системі; $Expl_i(Z)$ - пояснення щодо отриманого в результаті роботи інтелектуальної інформаційної системи рішення; $Expl_i(D) \hat{=} Expl_i(E) \hat{=} Expl_i(Z)$ - обмеження узгодженості пояснень щодо даних, подій та результату за умови, що вони описують єдиний *i*-й процес прийняття рішення в інтелектуальній системі; $Expl_i(D_m) \hat{=} Expl_k(D_m)$ - обмеження узгодженості різних пояснень щодо одного набору вхідних даних; $Expl_i(e_j,e_l) \hat{=} Expl_k(e_j,e_l)$ - обмеження узгодженості різних пояснень щодо елементів одного й того ж процесу прийняття рішення в інформаційній системі; $Expl_i(z_g) \hat{=} Expl_k(z_g)$ - обмеження узгодженості щодо одного із варіантів отриманого рішення.

Задача побудови пояснень у постановці (1) враховує логічну узгодженість знань. Такі знання відображають статичні та динамічні аспекти предметної області. Статичні властивості представляються у вигляді фактів, а динамічні - відношень між цими фактами.

Факти використання одного й того ж набору даних мають бути узгодженими між собою в сенсі логічної несуперечливості:

$$f_i(D_m) \hat{=} f_n(D_m) \,|\, \neg\left(f_i(D_m) \neg f_n(D_m)\right) \ , \tag{2}$$

де $f_i(D_m), f_n(D_m)$ - *i*-й та *n*-й факти використання підмножини вхідних даних D_m.

Факти щодо подій процесу прийняття рішення в інтелектуальній системі узгоджуються не лише в сенсі логічної несуперечливості, а й з урахуванням порядку їх виникнення в часі:

$$f(e_j) \hat{=} f(e_l) \,|\, \left(f(e_j) \neg f(e_l)\right), \left(e_j F^+ e_l\right) \vee \left(e_l F^+ e_j\right), \exists Expl(e_j,e_l) \ , \tag{3}$$

де $f\left(e_j\right)$, $f\left(e_l\right)$ - *i*-й та *n*-й факти щодо подій e_j та e_l; $e_j F^+ e_l$ - порядок подій у часі, що визначається темпоральним оператором F^+, тобто подія e_l обов'язково відбувається після події e_j, як безпосередньо за цією подією, так і після декількох проміжних подій; $Expl\left(e_j,e_l\right)$ - пояснення послідовності подій.

Додаткове темпоральне узгодження фактів виникнення подій є необхідним внаслідок того, що послідовність отримання результату є лінійною у часі для користувача. Важливість лінійного представлення подій для отримання пояснення пов'язана з тим, що на кожному етапі деревоподібного процесу прийняття рішення існує порядок у часі для кожної пари подій.

В залежності від вхідних даних та попередніх подій у кожній точці вибору відбувається перехід до однієї гілки процесу прийняття рішення. Тому для кожного сформованого рішення користувачеві необхідно пояснити лінійну послідовність дій, яка представлена упорядкованою послідовністю подій та привела до запропонованого йому результату.

Узгодженість результатів роботи інтелектуальної системи визначається в аспектах їх несуперечливості або можливості пояснення їх суперечливості:

$$(\forall g \forall q) f(z_g) \hat{=} f(z_q) \,|\, \neg\left(f(z_g) \neg f(z_q)\right) \vee \exists Expl\left(f(z_g), f(z_q)\right) , \tag{4}$$

де $f\left(z_g\right)$, $f\left(z_q\right)$ - факти отримання результатів $z_g, z_q \in Z$; $Expl\left(f\left(z_g\right), f\left(z_q\right)\right)$ - узгоджене пояснення щодо фактів отримання результатів роботи інтелектуальної системи.

Модель представлення узгоджених знань з урахуванням їх логічної несуперечливості або можливості пояснень виявленої суперечливості має вигляд:

$$M = \left\langle \begin{array}{l} D,E,Z : \left(\forall z_g \forall z_q\right) f(z_g) \hat{=} f(z_q), \\ \left(\forall e_j \forall e_l\right) f(e_j) \hat{=} f(e_l), \\ \left(\forall D_m \subseteq D\right) f_i(D_m) \hat{=} f_n(D_m) \end{array} \right\rangle . \tag{5}$$

Побудова узгоджених знань згідно з представленою моделлю створює умови для автоматизованого формування пояснень за принципом чорного ящику.

5. Висновки і перспективи подальших досліджень

Розроблено узагальнений підхід до узгодження знань для побудови пояснень в інтелектуальних інформаційних системах. Підхід передбачає узгодження знань з урахуванням їх несуперечливості у логічному та ймовірнісному аспектах, а також за відсутності аномальних елементів знань. У практичному плані підхід забезпечує можливість побудови набору узгоджених пояснень, що відображають різні аспекти процесу прийняття рішення та отриманого інтелектуальною системою результату, що спрощує застосування цих результатів у предметній області користувача.

Запропоновано модель представлення узгоджених знань з урахуванням їх логічної несуперечливості. Модель призначена для побудови пояснень та містить несуперечливі вхідні дані, події процесу прийняття рішення, для яких задано порядок у часі, а також несуперечливі результати роботи інтелектуальної системи. Модель дає можливість перевірити узгодженість знань у процесі формування пояснень щодо процесу та результатів роботи інтелектуальної системи, що створює умови для формування пояснень за принципом чорного ящику та доповнення функціонуючих інтелектуальних систем можливостями пояснень.

Подальші дослідження у напрямку узгодження знань при побудові пояснень пов'язані із погодженням знань із урахуванням ймовірнісної несуперечливості, що дасть можливість використати неповні та неточні знання при побудові пояснень. Використання ймовірнісного представлення знань також дає можливість запропонувати користувачеві декілька варіантів пояснення з урахуванням різних аспектів предметної області та упорядкованих за ймовірнісним показником.

Список літератури: 1. *Miller T.* Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence. 2019. №267. P. 1-38. 2. *Zhang Y., Chen X.* Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval. 2020. № 14(1). P. 1-101. 3. *Левикін В.М., Чала О.В.* Підтримка прийняття рішень в інформаціно-управляючих системах з використанням темпоральної бази знань. Науково-технічний журнал «Сучасні інформаційні системи», 2018 V. 2. №. 4. P. 101-107. 4. *Levykin V., Chala O.* Development of a method of probabilistic inference of sequences of business process activities to support business process management. Eastern-European Journal of Enterprise Technologies. 2018. № 5/3(95). P. 16-24. DOI: 10.15587/1729-4061.2018.142664. 5. *Phillips-Wren G.* Intelligent Systems to Support Human Decision Making. Artificial Intelligence: Concepts, Methodologies, Tools, and Applications. 2017. P. 3023-3036. http://doi:10.4018/978-1-5225-1759-7.ch125. 6. *Ribeiro M., Singh S., Guestrin C.* "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 2016. P. 97-101. DOI: 10.1145/2939672.2939778. 7. *Wang T., Lin Q.* Hybrid decision making: When interpretable models collaborate with black-box models. Journal of Machine Learning Research. 2019. № 1. P. 1-48. 8. *Thagard P., Verbeurgt K.* Coherence as constraint satisfaction. Cognitive Science. 1998. №. 22. P. 1-24. 9. *Thagard P.* Coherence, truth, and the development of scientific knowledge. Philosophy of Science. 2007. №. 74. P. 28-47. 10. *Thagard P.* Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. Applied Artificial Intelligence. 2004. №. 18. P. 231-249. 11. *Harwood W.* The Logic of Trust. PhD thesis, University of York. 2012. P. 245. 12. *Laurence B.J.* The Structure of Empirical Knowledge. Harvard University Press. 1985. P. 258. 13. *Chalyi S.F., Leshchynskyi V.O., Leshchynska I.O.* Explanation Model in an Intelligent Information System Based on the Concept of Knowledge Coherence. 2020. №1(3) . P. 19-23. https://doi.org/10.20998/2079-0023.2020.01.04. 14. *Chalyi S., Leshchynskyi V., Leshchynska I.* Designing explanations in the recommender systems based on the principle of a black box // Сучасні інформаційні системи. 2019. Т. 3, № 2. С. 47-51.

Надійшла до редколегії 29.06.2021

Чалий Сергій Федорович, доктор технічних наук, професор, професор кафедри інформаційних управляючих систем ХНУРЕ. Наукові інтереси: розробка моделей, методів і технологій автоматизованого управління бізнес-процесами (в тому числі із змінною структурою) в умовах неконтрольованих зовнішніх збурень. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. (057) 702 14 51.

Лещинський Володимир Олександрович, кандидат технічних наук, доцент кафедри програмної інженерії ХНУРЕ. Наукові інтереси: проектування, аналіз та рефакторінг коду програмного забезпечення. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. +38 (095) 203 17 50, +38 (098) 232 10 07.

Лещинська Ірина Олександрівна, кандидат технічних наук, доцент кафедри програмної інженерії ХНУРЕ. Наукові інтереси: проектування, аналіз та рефакторінг коду програмного забезпечення. Адреса: Україна, 61166, м. Харків, пр. Науки, 14, тел. +38 (095) 203 17 50, +38 (098) 232 10 07.