

# СИСТЕМНЕ УПРАВЛІННЯ

УДК 004.891.3

## ВИДІЛЕННЯ РЕЛЯЦІЙНИХ ЗАЛЕЖНОСТЕЙ БІЗНЕС-ПРОЦЕСУ НА ОСНОВІ АНАЛІЗУ ЙОГО ЛОГУ

**В. М. Левикін**, д-р техн. наук, проф.

**О. В. Чала**, канд. екон. наук, доц.

Харьковский национальный университет радиоэлектроники

E-mail: levykinvictor@gmail.com

*Виконано структурування об'єктів бізнес-процесу з урахуванням інформації про такі об'єкти в журналі реєстрації подій процесу. Запропоновано метод виділення реляційної складової знань бізнес-процесу шляхом аналізу його логу (журналу реєстрації подій). Метод передбачає виявлення залежностей на рівнях об'єктів процесу, атрибутів об'єктів та значень атрибутів. У практичному аспекті метод забезпечує можливість формалізації та подальшого включення в модель елементів персональних знань виконавців, відповідно до яких було змінено хід виконання процесу, після чого ці зміни зафіксовано в журналі.*

**Ключові слова:** знання-ємний бізнес-процес, інтелектуальний аналіз процесів, процесне управління.

*Structuring log information about business processes objects is made. The method of selecting the relational links between the objects of the business process by analyzing his journal suggested. The method involves identifying dependencies on process-level objects, object attributes and attribute values. In the practical aspect, the method makes it possible to find and include in the model the elements of personal knowledge of performers, which were used to change the process while it is running, and then recorded in the log.*

**Keywords:** knowledge-intensive business process, intelligent process analysis, process control.

### Постановка проблеми

Організація процесного підходу до управління підприємством передбачає побудову моделей бізнес-процесів (БП) та подальше управління БП з використанням їх моделей [1]. Модель БП містить алгоритм дій з вирішення відповідної функціональної задачі без урахування обмежень, що задає організаційна структура підприємства.

Цикл процесного управління передбачає побудову моделей БП «як повинно бути» на основі апріорних знань про процес, конфігурування моделей у процесній інформаційній системі, реалізацію БП, та удосконалення процесу за результатами його виконання. При удосконаленні бізнес-процесу виконується порівняння апріорної моделі БП «як повинно бути» та моделі процесу «як є, asis», що формується внаслідок аналізу роботи процесу. Під час порівняння визначають «вузькі місця» в моделі процесу. За результатами такого порівняння виконується коригування моделі «як повинно бути, tobe».

Формування моделі «як є» може бути ефективно виконано методами інтелектуального аналізу процесів (*process mining*) на основі дослідження логів (журналів реєстрації подій) таких процесів [2]. Результуюча модель «як є» форму-

ється у вигляді workflow — графа, що відображає можливі послідовності робіт, між якими встановлені причинно-наслідкові зв'язки.

Послідовність виконання дій процесу може змінюватись виконавцями з використанням персональних знань про контекст виконання БП [3]. Контекст складається з об'єктів, з якими взаємодіє БП [4]. Такі об'єкти характеризуються набором атрибутів та мають між собою статичні зв'язки, що впливають на хід виконання процесу та можуть привести до зміни послідовності дій реального процесу порівняно з моделлю «як повинно бути». Тому виявлення цих зв'язків на основі аналізу логів та формування з них складової знань бізнес процесу дозволяє обґрунтовано удосконалити модель БП «як повинно бути».

Таким чином, проблема виявлення залежностей між об'єктами контексту БП, є актуальною.

### Аналіз досліджень і публікацій

Традиційні методи інтелектуального аналізу логів процесів включають у модель всі варіанти їх поведінки без видалення незначущих дій [2]. У випадку багатоваріантних процесів, що змінюються виконавцями на основі своїх знань, це призводить до побудови workflow — моделей із значною кількістю вершин на дуг між ними [5],

аналіз та подальше удосконалення яких пов'язано із значними труднощами в силу когнітивних обмежень людини [6]. Існуючі підходи до фільтрації подій логу звичайно базуються на статистичних характеристиках [7] та не враховують причинно-наслідкові зв'язки, що привели до зміни ходу виконання бізнес-процесу.

Альтернативний підхід до моделювання процесів передбачає побудову опису об'єктів, з якими взаємодіє бізнес-процес, а також їх життєвого циклу [8]. При такому моделюванні враховуються зв'язки між об'єктами, однак не розглядається вплив цих зв'язків на послідовність дій бізнес-процесу в цілому [4], що не дозволяє врахувати вплив персональних знань виконавців та використати ці знання для удосконалення БП.

Тому питання виявлення реляційних залежностей між об'єктами бізнес-процесу на основі аналізу їх логів потребує подальшого розгляду.

### Постановка завдання

Мета роботи — розробка методу виявлення на основі аналізу логів БП реляційних залежностей між об'єктами контексту бізнес-процесу, що можуть впливати на послідовність його виконання.

Це дасть можливість формувати workflow — модель методами інтелектуального аналізу процесів, використовуючи лише таку підмножину подій логу, яка містить потрібні залежності. Такий підхід дозволяє отримати workflow — моделі підпроцесів БП з різною деталізацією.

Для досягнення поставленої мети необхідно вирішити такі задачі:

- виконати структурування об'єктів бізнес-процесу з урахуванням наявної інформації в журналі реєстрації подій;
- розробити метод виявлення реляційних залежностей з урахуванням ієрархічного опису об'єктів бізнес-процесу.

### Виклад основного матеріалу

Бізнес-процеси, що враховують вибір послідовності робіт виконавцями на основі своїх знань, відносяться до окремого класу БП — знання-ємних процесів (ЗБП) [9]. Запропонована авторами структурування таких процесів [10] передбачає поділ ЗБП на такі складові:

- контекст;
- послідовність робіт (workflow);
- знання, що визначають залежність між контекстом і діями процесу.

Контекст містить набір артефактів [8], під якими будемо розуміти всі об'єкти, які використовуються ЗБП (наприклад, продукт процесу), або впливають на ЗБП (наприклад, виконавці процесу). Між артефактами контексту існують взаємозв'язки, наприклад підприємство — під-

розділ — виконавець. Властивості та стан артефактів контексту, а також взаємозв'язків між артефактами впливають на хід виконання такого бізнес-процесу.

Послідовність робіт (workflow — складова) включає алгоритм виконання бізнес-процесу.

Рівень знань знання-ємного бізнес-процесу містить у собі два виду залежностей:

- залежності між елементами контексту;
- знання про вплив елементів контексту на виконання дій процесу.

Дана робота присвячена виявленню першого виду залежностей, що входять до складу ЗБП.

Рівень знань знання-ємного бізнес-процесу містить як явну, так і неявну складову. Явна складова задається апріорно і використовується при побудові моделі процесу. Неявна складова відображає вплив виконавців на хід виконання процесу.

Обидві складові можуть бути отримані на основі аналізу журналу. Це дає можливість видалити явну складову з множини залежностей, що отримані внаслідок аналізу логу, після порівняння моделі «як повинно бути» та результатів аналізу.

Підмножина, що залишиться, містить виявлену неявну складову. Після перевірки аналітиком впливу нових залежностей на результати процесу такі знання дають можливість удосконалити апріорну модель «як повинно бути».

Необхідною передумовою для виявлення вказаних залежностей є структурування вхідних даних — інформації про об'єкти (артефакти), що міститься в журналі подій процесу. Такий журнал охоплює множину трас процесу. Кожна траса містить послідовність подій, що фіксує одноразове виконання БП. Події логу, що входять до складу траси, відповідають діям бізнес-процесу. У більшості випадків одна подія містить запис про завершення відповідної дії процесу. Однак існують дії, яким відповідають декілька подій логу, наприклад: початок дії; затримка, що пов'язана з очікуванням ресурсів; завершення дії. У такому випадку до опису події в журналі включатимуть атрибут, що містить поточний стан дії процесу.

Кожна подія описується множиною атрибутів. Перелік атрибутів змінюється залежно від події та бізнес-процесу.

Обов'язковим атрибутом є часова мітка, що містить дату та час запису подій інформаційною системою.

Як правило, до складу атрибутів події входять: назва дії процесу; поточний стан дії процесу; ім'я виконавця дії; роль виконавця; об'єкт, що обробляється даною дією процесу.

Атрибути події можуть відображати структурні елементи одного й того ж об'єкту, наприклад:

- організація; підрозділ;
- активність (підпроцес); дія процесу; операція у складі дії.

Аналіз інформації в логах показав, що атрибути події процесу складають таку ієрархію: об'єкт контексту бізнес-процесу → атрибути об'єкту → значення атрибутів.

Інформація про складові вказаної ієрархії прив'язана до різних елементів логу. Так, атрибути та значення атрибутів входять до складу опису події. У більшості випадків така інформація розміщена у стандартизованому вигляді: «класифікатор: ім'я атрибута» = «значення атрибута». Класифікатор вказує артефакт або групу атрибутів, до яких належить даний атрибут. Наприклад, для логу у XES-форматі вираз `<stringkey="concept: name" value="10061280"/>` має таку структуру: ім'я (name) дії (concept) має значення "10061280".

Атрибути події можуть мати лише ім'я, без класифікатора, наприклад: `<stringkey="question" value="EMPTY"/>`. Додаткову інформацію про об'єкт, до якого має відношення атрибут без класифікатора, розміщено в опису лога. Однак у загальному випадку така інформація має бути надана додатково.

Аналіз атрибутів множини подій без урахування часової складової дозволяє встановити найбільш часто повторювані зв'язки між артефактами, атрибути яких представлені в записи подій журналу. Темпоральний аспект у даному випадку не розглядається, тому при виявленні таких реляційних залежностей доцільно використати не методи *processmining*, методи інтелектуального аналізу даних.

Запропонований авторами метод пошуку реляційних залежностей використовує для вирішення локальних задач на окремих етапах алгоритм пошуку предметних наборів FPG (Frequent Pattern-Growth Strategy). Даний алгоритм дозволяє задати послідовність атрибутів артефактів при побудові структури залежностей. Це дає можливість визначити залежності які між різними артефактами, так і між окремими атрибутами одного артефакту.

Головна ідея запропонованого метода полягає в тому, щоб побудувати реляційні залежності для всіх трьох рівнів ієрархії: артефакт; атрибут; значення атрибута. Тому на кожному етапі послідовно формуються набори, що були використані при описі подій.

При реалізації етапів запропонованого методу рівень підтримки визначено=2.

Такий рівень означає, що одна й та сама подія зустрічається щонайменше в двох трасах лога. Вибір такого рівня підтримки зумовлюється таким.

З результатами аналізу потрібно розділити типові рішення та унікальні, що, можливо, підходять тільки для окремих випадків виконання процесу. Такі окремі випадки недоцільно включати в загальну модель «як є» тому, що це створює труднощі при її аналізі внаслідок когнітивних обмежень людини. Для додаткового аналізу окремих випадків часткова модель БП може бути побудована з однієї унікальної траси процесу.

Якщо одна й та сама подія виникла в одних і тих самих умовах на різних трасах (тобто характеризується ідентичними наборами атрибутів, а також їх значень в журналі БП), то таку подію вважатимемо типовою, оскільки вона характерна як мінімум для двох екземплярів бізнес-процесу. Відповідно, будемо враховувати реляційні залежності, що використовуються при типовому виконанні бізнес-процесу. Метод виявлення реляційних залежностей містить такі етапи:

*Етап 1.* Визначення множини артефактів та їх атрибутів як вхідних даних.

Формально артефакт характеризується множиною атрибутів:

$$af = \{a_i^{af}\}, af \in Af,$$

де  $af$  — поточний артефакт;  $Af$  — множина артефактів контексту;  $a_i^{af}$  —  $i$ -атрибут артефакту.

Стан артефакту, зафіксований у подіях логу, визначається значеннями його атрибутів:

$$af_{tr}^e = \{a_i^{af}, v_{ij}\}, af \in Af, v_{ij} \in V_i, V_i \subset V,$$

де  $af_{tr}^e$  — стан артефакту у дискретний момент часу, коли виникла подія  $e$  та трасі  $t$ ;  $v_{ij}$  — поточне значення атрибута  $a_i$ ;  $V_i$  — множина можливих значень атрибута  $a_i^{af}$ .

Подія логу на трасі також визначається таким набором атрибутів об'єктів контексту, для яких задані значення в її описі:

$$e_{tr} = \bigcup_{af} \{a_i^{af} \mid \exists v_{ij}^{e,tr}\},$$

де  $e_{tr}$  — подія на трасі  $tr$  логу процесу;  $v_{ij}^{e,tr}$  — значення атрибута  $a_i^{af}$ , що зафіксовано в описі події  $e_{tr}$  на трасі  $tr$ .

У даному методі використовуються лише ті артефакти та їх атрибути, що зазначені в журналі подій процесу, тобто для вхідних даних повинна виконуватись умова:

$$A = \{A^{af}\} = \{\{a_i^{af}\}\} \cup \{a_i^{af}\} = \bigcup_{af, e, tr} \{a_i^{af} \mid \exists v_{i,j}^{e,tr}\}, (1)$$

де  $A$  — множина атрибутів — вхідних даних методу.

Умова (1) задає обмеження на вхідні дані в тому випадку, якщо є додаткова експертна інформація про склад артефактів процесу. При використанні лише інформації з подій логу обмеження на множину вхідних атрибутів має такий вигляд:

$$A = \{a_i\} = \bigcup_{e, tr} \{a_i^{e, tr}\} \quad (2)$$

Таким чином, на даному етапі формується множина артефактів відповідно до обмеження (1) та атрибутів відповідно до обмеження (2).

*Етап 2.* Виявлення залежностей між атрибутами артефактів.

На даному етапі виконуються такі кроки аналізу журналу подій.

$$A = \{a_i\} = \bigcup_{e, tr} \{a_i^{e, tr}\}.$$

2.1. Визначення підмножин атрибутів подій з підтримкою не менше 2.

2.2. Сортування атрибутів подій  $a_i^{e, tr}$  за рівнем зменшення їх підтримки.

2.3. Сортування подій за зменшенням підтримки першого атрибуту.

2.4. Побудова FP-дерева відповідно до алгоритму FPG.

2.5. Побудова умовних дерев та визначення атрибутів, між якими існують реляційні залежності.

*Етап 3.* Виявлення залежностей між значеннями атрибутів.

Залежності між значеннями атрибутів на даному етапі визначаються для кожного атрибуту  $a_i$ , що може бути зафіксований у різних подіях на різних трасах логу у вигляді  $a_i^{e, tr}$ .

Визначення залежностей відбувається аналогічно етапу 2.

Даний етап призначений для виявлення таких залежностей між значеннями атрибутів, які відсутні в моделі процесу, але були використані виконавцями для зміни послідовності його дій.

*Етап 4.* Виявлення залежностей між артефактами.

Даний етап виконується тільки у випадку істинності умови (1).

Особливість даного етапу полягає у тому, що безпосередньо артефакти не зафіксовані в журналі подій.

Тому рівень підтримки артефакту визначається як сума появ атрибутів цього артефакту в опису події. Оскільки одна й та ж подія може бути зафіксована на різних трасах процесу, то при визначенні рівня підтримки події з трас об'єднуються, тобто:

$$r(af, e) = \sum_{i, tr} |\{a_i^{af} | \exists v_{i, j}^{e, tr}\}|,$$

де  $r(af, e)$  — рівень підтримки артефакту  $af$  для події  $e$ .

### Висновки

Виконано структурування об'єктів бізнес-процесу з урахуванням інформації про такі об'єкти в журналі реєстрації подій.

Об'єкт характеризується множиною атрибутів. Стан об'єкту — поточними значеннями атрибутів. Інформація про атрибути об'єктів та їх значення завжди присутня в описі подій журналу. Інформація про зв'язок між об'єктами та атрибутами у повному обсязі у журналі не присутня. Тому для встановлення складу об'єктів потрібна додаткова експертна інформація.

Залежно від повноти вхідних даних, реляційні залежності можуть бути виявлені на двох або трьох рівнях: значення атрибутів — атрибути — об'єкти.

Запропоновано метод виділення, на основі аналізу логу, реляційних залежностей між об'єктами контексту бізнес-процесу.

Такі залежності можуть впливати на послідовність виконання бізнес-процесу. Метод передбачає виявлення залежностей на всіх трьох рівнях деталізації вхідних даних та подальшу побудову ієрархії залежностей.

У практичному аспекті метод створює можливість формалізації та подальшого включення в модель персональних знань виконавців, відповідно до яких було змінено хід виконання процесу.

### ЛІТЕРАТУРА

1. *Weske M.* Business Process Management: Concepts, Languages, Architectures / M. Weske — 2nd ed. — Presented at Springer-Verlag Berlin Heidelberg, 2012. — 403 p.
2. *Van der Aalst, W. M. P.* Process Mining: Discovery, Conformance and Enhancement of Business Processes / W. M. P. Van der Aalst. — Springer Berlin Heidelberg, 2011. — 352 p.
2. *Van der Aalst W. M. P.* Process Mining in the Large: A Tutorial / W. M. P. Van der Aalst // Business Intelligence. — Springer Science + Business Media, 2014. — P. 33–76.
3. *Smith E. A.* The role of tacit and explicit knowledge in the workplace / Elizabeth A. Smith // Journal of Knowledge Management. — 2001. — №5(4). — P. 311–321.
4. *Nigam A.* Business artifacts: An approach to operational specification/ A. Nigam, N. S. Caswell/ IBM Systems Journal. — 2003. — № 42(3). — P. 428–445.
5. *Van der Aalst W.M.P.* Process Mining: Discovering and Improving Spaghetti and Lasagna Processes / W.M.P. van der Aalst // In N. Chawla, I. King, and

A. Sperduti, editors, IEEE Symposium on Computational Intelligence and Data Mining, Paris, France, April 2011. — P. 13–20.

6. *Görg, C. Visual Representations / C. Görg, M. Pohl, E. Qeli, K. Xu // Human-Centered Visualization Environments. — Springer Science + Business Media. — 2007 — P. 163–230.*

7. *Van der Aalst, W. M. P. Process Mining: Discovery, Conformance and Enhancement of Business Processes / W. M. P. Van der Aalst. — Springer Berlin Heidelberg, 2011. — 352 p.*

8. *Hull R. Business Artifacts with Guard-Stage-Milestone Lifecycles: Managing Artifact Interactions with Conditions and Events// DEBS, — 2011. — P. 51–62.*

9. *Gronau N. Modeling and Analyzing knowledge intensive business processes with KMDL: Comprehensive insights into theory and practice (English) / N. Gronau. — Gito, 2012. — 522 p.*

10. *Левыкин В. М. Выделение элементов контекста знание-емких бизнес-процессов на основе анализа логов // В. М. Левыкин, О. В. Чалая // Технологический аудит и резервы производства. — 2016. — № 5/2(31). — С. 65–71.*

Стаття надійшла до редакції 24.11.2016