

МЕТОД ЗМЕНШЕННЯ НЕВИЗНАЧЕНОСТІ ДАНИХ В СИСТЕМАХ МОНІТОРИНГУ НА ОСНОВІ ОНТОЛОГІЧНИХ ДОПОВНЕНЬ

Національний технічний університет України “Київський політехнічний інститут
імені Ігоря Сікорського”

art.kaplunov@gmail.com,
artem.volokita@kpi.ua

Вступ

Переосмислення світовим співтовариством інформації як стратегічного ресурсу, стимулювало розробку нових інформаційних технологій для отримання і обробки великих обсягів інформації, її зберігання, аналізу та надання користувачам. В рамках використання складних розподілених систем необхідно максимально підвищити їхню надійність і адаптивність. Основним механізмом контролю виступають системи моніторингу, тому їх постійне вдосконалення є однією з пріоритетних складових у розвитку розподілених систем.

В процесі моніторингу розподіленої системи частина інформації про стан об'єкту моніторингу може бути недоступною чи втраченою під час передачі, через що сукупність зібраної інформації для аналізу є неповною. Частково це можливо компенсувати за допомогою онтологічних доповнень, які шляхом логічного аналізу над об'єктами моніторингу дозволяють виконувати визначення їх атрибутів.

Таким чином тематика, якій присвячене дане дослідження є актуальною і має практичну цінність, а дослідження в рамках запропонованої тематики призначені для розширення можливостей систем моніторингу і поліпшення їх експлуатаційних характеристик.

Огляд літературних джерел

Онтології можуть містити тисячі тверджень і правил, відображаючи глибинні взаємозв'язки між інформаційними

об'єктами і їх атрибутами. При цьому накладні витрати, пов'язані з логічним висновком над онтологіями, залежать як від її розміру, так і від мови дескрипційної логіки, на якій вона заснована. Так у роботі [1] показано, що для онтологій на базі мови EL++ характерна поліноміальна залежність між розміром онтології і часом її обробки, що дозволяє ефективно застосувати методику онтологічного доповнення з урахуванням великого обсягу даних.

В роботі [2] був запропонований спосіб вирішення проблеми неоднозначності в процесах моніторингу неоднорідного проекту за допомогою онтологій, однак даний підхід не передбачає ні онтологічного доповнення, ні подальшого спрощення аналізу зібраної інформації.

Можливість створення онтологічних доповнень автоматизованими засобами на основі Data Mining аналізу розглянуто в роботі [3]. Авторами запропоновано визначити онтологію між класами в результаті попередньої класифікації даних. Класифікація даних полягає в розділенні набору даних на підмножини (класи), де усі дані в одному класі будуть подібними, а дані з різних – несхожими. Причому кількість кластерів, згенерованих алгоритмом класифікації, завжди буде меншою за кількість об'єктів, до яких було застосовано алгоритм класифікації, а усі об'єкти, що належать до одного кластеру, мають однакові властивості.

Таким чином, на основі аналізу літературних джерел, в яких розглядалася можливість застосування онтологічних доповнень для зменшення невизначеності інформації, зібраної системою моніторингу, було зроблено висновок, що ця тема не розкрита у повному обсязі та потребує подальшого дослідження. А створення онтологічних доповнень автоматизованими засобами на основі Data Mining аналізу є перспективним для реалізації підходом.

Виходячи з цього постає актуальна і доцільна задача розроблення методів використання онтологічних доповнень для зменшення невизначеності даних в системах моніторингу складних розподілених систем, якій присвячена дана стаття.

Мета та задачі дослідження

Метою є підвищення ефективності моніторингу розподілених систем, за рахунок зменшення невизначеності даних моніторингу, шляхом використання онтологічних доповнень.

Для реалізації мети дослідження запропоновано розробити:

- математичну модель представлення об'єктів системи моніторингу та їх атрибутів для визначення властивостей даних та реалізації їх онтологічного доповнення.
- метод зменшення невизначеності моніторингових даних в системах моніторингу складних розподілених систем, шляхом використання онтологічних доповнень.

Основна частина

Для опису запропонованої математичної моделі представимо об'єкти системи моніторингу та їх атрибути у вигляді наступних математичних співвідношень.

Визначимо поняття сукупності інформації, яка представляє собою дані моніторингу і об'єкту моніторингу. В основі цієї сукупності лежить множина об'єктів моніторингу (A), природа яких визначає всі їх ключові характеристики.

Класифікуємо об'єкти моніторингу за допомогою множини атрибутів (P).

Для кожного атрибута $\rho \in P$ визначено множину допустимих значень V_ρ ,

відоме як домен атрибута ρ . Атрибути можуть розділяти загальний простір значень і можуть відноситись до одного і того ж домену V_ρ .

Для того щоб асоціювати певну інформацію про об'єкт моніторингу, опишемо функцію визначення γ . Ця функція відображує множину $A \times P$ в множину $V = \cup_{\rho \in P} V_\rho$ таким чином, що $\gamma(\alpha, \rho) \in V_\rho$ для всіх $\alpha \in A$ і $\rho \in P$. Фактично за допомогою функції γ ми задаємо кожному об'єкту його опис у вигляді значень його атрибутів.

Якщо γ – частково визначена функція, то визначимо сукупність зібраної інформації як неповну, якщо ж γ – всюди визначена функція, то така сукупність зібраної інформації визначається як повна.

Таким чином, ми можемо визначити сукупність зібраної інформації S як

$$S = \langle A, P, V, \gamma \rangle,$$

де A – множина об'єктів моніторингу, P – множина атрибутів, $V = \cup_{\rho \in P} V_\rho$ де V_ρ – множина значень атрибута $\rho \in P$ за умови що $|V_\rho| > 1$, і γ – функція визначення $A \times P \rightarrow V$.

Тоді, доповнена сукупність зібраної інформації S^+ описується наступним чином:

$$S^+ = \langle A, P, V, \gamma, \omega \rangle,$$

де A – множина об'єктів моніторингу, P – множина атрибутів, $V = \cup_{\rho \in P} V_\rho$, де V_ρ – множина значень атрибута $\rho \in P$, γ – функція визначення $\gamma: A \times P \rightarrow V$ і ω – функція доповнення $\omega: A \times P \rightarrow V$.

Призначення функції ω полягає в додатковому визначенні атрибутів об'єкта x , виходячи з множини вже визначених значень, заданих за допомогою функції γ .

Для реалізації функції ω запропоновано використати підхід онтологічних доповнень. Зважаючи на те, що онтологія – це формалізація деякої області знань у вигляді множини понять і відносин між ними, стає можливим довизначення атрибутів об'єктів моніторингу за допомогою виконання логічного аналізу над самими об'єктами.

Представимо онтологічно доповнену сукупність зібраної інформації S_O^+ як розширення визначення S^+ :

$$S_O^+ = \langle A, P, V, \gamma, \omega, O \rangle,$$

де O – онтологія системи.

При такому підході онтології дозволяють здійснювати логічний аналіз над об'єктами $\alpha \in A$ і виконувати довизначення їх атрибутів, тим самим збільшуючи

кількість інформації в сукупності зібраної інформації.

Розглянутий метод базується на реалізації запропонованої математичної моделі для опису процесу моніторингу завдань в розподіленій системі.

У якості прикладу розглянемо онтологічно доповнену сукупність інформації S_O^+ , зібрану внаслідок моніторингу завдань в розподіленій обчислювальній системі:

$$A = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\},$$

$$P = \{\text{запущено задач, статус IP – адреса, задач в черзі}\},$$

$$V = \{V_{IP\text{-адреса}} \cup V_{\text{задач в черзі}} \cup V_{\text{запущено задач}} \cup V_{\text{статус}}\},$$

$$V_{IP\text{-адреса}} = \{0 \leq \alpha \leq 2^{32} - 1\},$$

$$V_{\text{задач в черзі}} = N, V_{\text{запущено задач}} = N,$$

$$V_{\text{статус}} = \{\text{простоює, зайнятий, недоступний}\}.$$

Функцію визначення γ представимо у вигляді таблиці (табл. 1), де розміщено дані моніторингу завдань в розподіленій обчислювальній системі. Як можна побачити, зібрані в процесі моніторингу дані є неповними.

Таблиця 1. Неповна сукупність моніторингових даних (γ)

A	IP-адреса	задач в черзі	запущено задач	статус
α_1	192.168.1.1	5	0	
α_2	192.168.1.2	10	3	
α_3	192.168.1.3	8	1	
α_4	192.168.1.4			простоює
α_5	192.168.1.5	7		зайнятий

На підставі аналізу характеристик атрибутів доповнимо сукупність зібраних даних моніторингу наступною онтологією:

$$O = \begin{cases} \text{Естатус. простоює} \equiv \exists \text{задач в черзі. } (=, 0) \\ \quad \sqcap \text{Езапущено задач. } (=, 0) \\ \text{Естатус зайняті} \equiv \exists \text{запущено задач. } (\geq, 1) \\ \text{Естатус. недоступний} \equiv \exists \text{задач в черзі. } (>, 0) \\ \quad \sqcap \text{Езапущено задач. } (=, 0) \end{cases}.$$

Доповнена сукупність моніторингових даних буде мати вигляд наведений в табл. 2.

Таблиця 2. Доповнена сукупність моніторингових даних ($\gamma + \omega$)

A	IP-адреса	задач в черзі	запущено задач	статус
α_1	192.168.1.1	5	0	недоступний
α_2	192.168.1.2	10	3	зайнятий
α_3	192.168.1.3	8	1	зайнятий
α_4	192.168.1.4	0	0	простоює
α_5	192.168.1.5	7	≥ 1	зайнятий

Отже за допомогою використання онтологічних доповнень вдалося зменшити неповноту сукупності зібраних даних моніторингу. При цьому для об'єкта $\alpha 5$ була зменшена невизначеність (ентропія) щодо параметра «запущено задач». Подібне зменшення невизначеності дозволяє спростити процес аналізу зібраних даних, а також вдосконалити процес прийняття рішень.

У прикладі використовуються створені вручну онтології. Проте, як було продемонстровано у дослідженні [3], створення онтологічних доповнень можливе автоматизованими засобами на основі Data Mining аналізу. Варто зазначити, що на даний момент це найменш вивчений із підходів і подібні інструменти не отримали належного розвитку, а створювані таким чином онтологічні доповнення все ще поступаються в якості онтологічним доповненням, створеним людиною вручну. Однак з розвитком подібних технологій цей варіант буде ставати більш привабливим для використання.

Висновки

Для вирішення завдання підвищення ефективності моніторингу розподілених систем в статті запропоновані наступні рішення.

- Розроблено математичну модель представлення об'єктів системи моніторингу та їх атрибутів для визначення властивостей даних та реалізації їх онтологічного доповнення.

- Запропоновано метод зменшення невизначеності моніторингових даних, який за рахунок використання онтологічних доповнень дозволяє зменшити

неповноту сукупності зібраної інформації, що в цілому забезпечує підвищення ефективності функціонування систем моніторингу складних розподілених систем.

- Теоретично обґрунтовано, що застосування онтологічних доповнень дозволяє збільшити обсяг інформації, отриманої з вхідних даних. Таким чином, користувач системи з доповненою сукупністю зібраних даних може отримувати більше релевантної інформації, ніж користувач традиційної системи на основі того ж набору вихідних даних. Однак застосування онтологічних доповнень відкриває можливість значно вдосконалити процес аналізу зібраної інформації, спростити його і зробити більш зручним для користувача.

Запропонований в статті підхід використання онтологічних доповнень має високу практичну цінність і може відігравати значну роль у розширенні можливостей систем моніторингу і поліпшення їх експлуатаційних характеристик.

Література

1. Постішний О.С. Эффективный логический анализ больших онтологий за полиномиальное время. *Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр.* 2012. В. 55. С. 194–198.

2. Dong H., Hussain F., Chang E. An ontology-based real-time project monitoring system in the cloud. *Journal of Universal Computer Science.* 2011. V. 17, Iss. 8. P. 1161–1182.

3. Touzi A., Massoud H., Ayadi A. Automatic ontology generation for data mining using FCA and clustering. 2013. 10 p.

Каплунов А.В., Волокита А.М.

МЕТОД ЗМЕНШЕННЯ НЕВИЗНАЧЕНОСТІ ДАНИХ В СИСТЕМАХ МОНІТОРИНГУ НА ОСНОВІ ОНТОЛОГІЧНИХ ДОПОВНЕНЬ

Пропозиція щодо використання онтологічних доповнень має перспективу у сфері моніторингу розподілених систем, коли у сукупності зібраних даних існує невизначеність, а отримати чіткі параметри традиційними методами вимірювання досить складно у зв'язку з рядом об'єктивних причин. Було запропоновано представити онтологічно доповнену сукупність зібраної інформації у вигляді:

$$S_0^+ = \langle A, P, V, \gamma, \omega, O \rangle$$

де O – онтологія системи. Таким чином онтології дозволяють проводити логічний аналіз над об'єктами моніторингу і виконувати довізначення їх атрибутів, тим самим збільшуючи кількість інформації в сукупності зібраної інформації. Додатково була розглянута перспектива створення онтологічних доповнень автоматизованими засобами на основі Data Mining аналізу. У статті представлені результати теоретичного дослідження можливостей використання доповненої сукупності зібраної інформації. Запропонований в дослідженні підхід може відігравати значну роль у розширенні можливостей систем моніторингу і поліпшення їх експлуатаційних характеристик.

Ключові слова: моніторинг, розподілені системи, онтологічні доповнення.

Kaplunov A.V., Volokyta A.M.

METHOD OF REDUCTION OF DATA UNCERTAINTY IN MONITORING SYSTEMS BASED ON ONTOLOGICAL ADDITIONS

The proposal for the use of ontological additions has a perspective in the field of distributed systems monitoring, when there is uncertainty in the totality of the collected data, and it is quite difficult to obtain clear parameters by traditional measurement methods due to a number of objective reasons. It was proposed to present an ontologically supplemented set of collected information in the form of:

$$S_0^+ = \langle A, P, V, \gamma, \omega, O \rangle$$

where O is the ontology of the system. In this way, ontologies make it possible to carry out a logical analysis of monitoring objects and perform redefinition of their attributes, thereby increasing the amount of information in the totality of collected information. In addition, the prospect of creating ontological additions by automated means based on Data Mining analysis was considered. The article presents the results of a theoretical study of the possibilities of using the expanded set of collected information. The approach proposed in the study can play a significant role in expanding the capabilities of monitoring systems and improving their operational characteristics.

Keywords: monitoring, distributed systems, ontological additions.