

ДОСЛІДЖЕННЯ ІНСТРУМЕНТІВ ТРАНСФОРМАЦІЇ БАЗ ЗНАНЬ НАФТОГАЗОВОЇ ПРЕДМЕТНОЇ ОБЛАСТІ

Л.М.Гобир, Т.О.Ваврик, В.І.Шекета

ІФНТУНГ, 76019, Івано-Франківськ, вул. Карпатська, 15, тел. (03422) 42127

e-mail: sheketa@mail.ru

В предложенном исследовании представлено концепцию необходимой модификации, как эффективного инструмента трансформации баз знаний и их подготовки к применению формально-логического аппарата предикатных запросов, как multifunctional mechanism for supporting dialog with an user in the information systems on the basis of databases and knowledgebases of oil and gas subject domain.

In given paper the notion of necessary modification is introduced, as an tool for transformation of knowledge bases and preparing them for applying the formal-logical apparatus of predicate queries modifications, which are considered as a multifunctional mechanism for supporting dialog with an user in the information systems on the basis of databases and knowledgebases of oil and gas subject domain.

Одним із ключових питань, з точки зору математичного моделювання процесу побудови інформаційних інтелектуальних систем для нафтогазової предметної області [1], є спосіб представлення знань, на основі якого система повинна приймати рішення в певній ситуації. Представлення знань повинно бути задано способом, який дозволяє перехід до представлення фрагментів інформації про нафтогазовий об'єкт в термінах структур баз даних і знань.

Теорія модифікаційних предикатних запитів [2] є формально-логічним апаратом опису і вивчення процесів оновлення і модифікації баз даних і знань, побудови логічного висновку на основі баз даних і знань [2, 3, 5, 6]. В першому наближенні під базою даних будемо розуміти набір фактів. Основні ідеї такого підходу розглядаються в рамках конкретних реалізацій SQL, або реалізацій для WWW і WEB [7].

Розглядатимемо таку базу даних з точки зору фактів і процесів, що призводять до їх зміни, тобто з погляду семантики і синтаксису такого представлення. Під синтаксисом будемо розуміти набір правил для поєднання символів в логічно коректні вирази. Під семантикою будемо розуміти спосіб інтерпретації виразів, що одержуються в результаті конкретних реалізацій синтаксичних правил.

В роботі [7] автором введено поняття модифікаційних предикатних запитів як інструменту підтримки діалогу з користувачем інформаційних інтелектуальних систем в процесі побудови ним запитів, що розглядаються в ході динаміки процесу оновлення і модифікації баз даних і знань нафтогазової предметної області.

Проте, виконані дослідження показали, що рівень структуризації інформації, що наповнює реальні бази знань нафтогазової предметної області не відповідає рівню, на якому можна було би виконувати безпосереднє застосування концепції модифікаційних предикатних запитів.

Тому метою даної статті є побудова ефективного механізму трансформації вихідних баз знань предметної області для досягнення задовільного рівня логічної зв'язаності інформації,

що є необхідною передумовою застосування формально-логічного апарату модифікаційних предикатних запитів для інформаційних систем на основі баз даних і баз знань нафтогазової предметної області.

Ми будемо розвивати власний формально-логічний апарат на основі основних принципів абстрактного логічного програмування і теоретичних аспектів програмування в мові Prolog.

Розглядатимемо базу знань як набір інформаційних сутностей атомарних предикатів з деякого скінченного інформаційного простору O . Всі зміни, що відбуватимуться в базі знань, будемо розглядати як наслідок модифікаційних предикатних запитів, що генеруються ІС відповідно до вказівок користувача. Основою самих запитів є набір модифікаційних предикатних правил. Будемо розглядати два типи правил:

$$K_{B_+}(o) \ll K_{B_+}(o_1), \dots, K_{B_+}(o_l), \\ K_{B_-}(p_1), \dots, K_{B_-}(p_m); \quad (1)$$

$$K_{B_-}(o) \ll K_{B_+}(o_1), \dots, K_{B_+}(o_l), \\ K_{B_-}(p_1), \dots, K_{B_-}(p_m), \quad (2)$$

де $o, o_i, p_i \in O$, \ll – дескриптор модифікації. Основна ідея такого запису правил в тому, що $K_{B_+}(o)$ означає, що атомарний предикат o повинен бути включений в базу знань K_B , а K_{B_-} – що o – повинен бути виключений з бази знань.

Таким чином, модифікаційні предикатні правила задають логічні обмеження на базу знань. Зокрема, правило (1) накладає на базу знань таку умову: o належить базі знань, або принаймні одне з o_i , $1 \leq i \leq l$ не належить базі знань, або принаймні одне з p_j , $1 \leq j \leq m$ належить базі знань. Наприклад, правило (2) стверджує, що у випадку, коли o і всі o_i , $1 \leq i \leq l$ належать базі знань, а всі p_j , $1 \leq j \leq m$ не належать базі знань, ми можемо виключити o з бази знань, або вилучити одне з o_i , або додати одне з p_j .

Введемо поняття необхідної модифікації, яка буде описувати всі операції $K_{B_+}()$ і $K_{B_-}()$, що спричиняються модифікаційним запитом для вихідної бази знань K_B .

Означення 1. Нехай Q_m – модифікаційний предикатний запит. Позначимо через $\lambda_{nm}(Q_M)$ необхідні модифікації для K_B . В [7] показано, що $\lambda_{nm}(Q_M)$ фактично є найменшою моделлю для Q_m , якщо $\lambda_{nm}(Q_M)$ розглядати як Горн-програму Q_m , побудовану з незалежних атомарних предикатів виду $K_{B+}(o)$ і $K_{B-}(p)$.

Нехай, наприклад, модифікаційний предикатний запит Q_m має вигляд

$$\{K_{B+}(o) \ll K_{B-}(p) \ll K_{B+}(o), \\ K_{B-}(r) \ll K_{B-}(o)\},$$

тоді $\lambda_{nm}(Q_M) = \{K_{B+}(o), K_{B-}(p)\}$.

В результаті виконання модифікаційних предикатних запитів база знань змінюється, але статус деяких елементів залишається незмінним. Тобто згідно розглянутого принципу інерції [9] при створенні набору модифікаційних правил ми, власне, вказуємо тільки опис того, чи потрібно змінити певні елементи. Якщо для певного елемента немає правила модифікації, то, відповідно, згідно з принципом інерції він не буде змінюватися в ході виконання модифікації.

Таким чином, множину всіх модифікаційних літералів, що описують елементи бази знань, статус яких не змінюється в процесі переходу від бази знань K_{B_1} до бази знань K_{B_2} , Q_m і $K_{B_1} \gg K_{B_2}$, будемо називати множиною інерції для K_{B_1} , K_{B_2} і означати, як:

$$O_i(K_{B_1}, K_{B_2}) = \{K_{B+}(o) : o \in K_{B_1} \cap K_{B_2}\} \cup \\ \cup \{K_{B-}(o) : o \notin K_{B_1} \cup K_{B_2}\}.$$

Означення 2. Залишок для Q_m стосовно баз знань (K_{B_1}, K_{B_2}) $Q_M^{K_{B_1}, K_{B_2}}$ будемо називати новий модифікаційний запит, що одержується із Q_m шляхом видалення із тіла кожного модифікаційного правила в Q_m всіх модифікаційних літералів, які належать також і $O_i(K_{B_1}, K_{B_2})$.

Означення 3. Базу знань K_{B_2} будемо називати Q_m -модифікацією для K_{B_1} , якщо множина необхідних модифікацій для $Q_M^{K_{B_1}, K_{B_2}}$ є когерентною і якщо

$$K_{B_2} = K_{B_1} \circ \lambda_{nm}(Q_M^{K_{B_1}, K_{B_2}}).$$

Для ілюстрування наведених вище означень розглянемо такий приклад. Приклад 1. Припустимо: нам потрібно будувати процедуру логічного висновку для інтелектуальної системи на основі баз даних і знань нафтогазової предметної області Plast Researcher, виходячи із чотирьох характеристик нафтогазоносної породи: $\{пористість, насиченість, проникність, коефіцієнт_опору\}$.

Причому, виходячи із досвіду експертів по опрацюванню даних характеристик породи, ми матимемо справу із певним набором обмежень, а саме:

1. Характеристики “пористості” і “насиченості” є найбільш продуктивними для виконання процедури логічного висновку, тому одна із них повинна обов’язково бути присутньою у вихідній базі знань K_B^{noc} .

2. Використання характеристики “проникність” є неефективним без одночасного опрацювання характеристики “коефіцієнт_опору”. Це – так звані парні характеристики, і якщо в K_B^{noc} не буде включено характеристику “коефіцієнт_опору”, то не потрібно включати і характеристику “проникність”.

3. Спільне використання характеристик “коефіцієнт_опору” і “проникність” не є ефективним, оскільки вони мають різну фізичну природу і методики їх вимірювання базуються на різних принципах.

4. Неефективним є також спільне використання характеристик “насиченість”, “коефіцієнт_опору” відповідно до причин, викладених в пункті 3.

Згідно з переліченими обмеженнями як початковий склад бази знань приймемо $K_B^{noc} = \{пористість, проникність\}$.

Ми прагнемо сформувати базу знань K_B , яка б задовольняла всім чотирьом накладеним обмеженням. Таким чином, в термінах введених нами вище означень ми одержуємо $K_{B_1} = \{пористість, проникність\}$. Побудуємо модифікаційний предикатний запит Q_M :

$$\{K_{B+}(насиченість) \ll K_{B-}(пористість), \\ K_{B+}(пористість) \ll K_{B-}(насиченість),$$

$$K_{B+}(коефіцієнт_опору) \ll K_{B+}(проникність), \\ K_{B-}(проникність) \ll K_{B-}(коефіцієнт_опору), \\ K_{B-}(пористість) \ll K_{B+}(коефіцієнт_опору), \\ K_{B-}(коефіцієнт_опору) \ll K_{B+}(насиченість)\}.$$

Покажемо, що $K_{B_2} = \{пористість\}$ є Q_M -модифікацією для K_{B_1} . Виходимо з того, що

$$O = \{пористість, насиченість, проникність, коефіцієнт_опору\}.$$

Тоді

$$O_i(K_{B_1}, K_{B_2}) = \{K_{B+}(пористість),$$

$$K_{B-}(насиченість), K_{B-}(коефіцієнт_опору)\}$$

$$Q_M^{K_{B_1}, K_{B_2}} = \{K_{B+}(насиченість) \ll K_{B-}(пористість),$$

$$K_{B+}(пористість) \ll K_{B+}(коефіцієнт_опору) \ll \\ \ll K_{B+}(проникність),$$

$$K_{B-}(проникність) \ll K_{B-}(пористість) \ll$$

$$\ll K_{B+}(коефіцієнт_опору),$$

$$K_{B-}(коефіцієнт_опору) \ll K_{B+}(насиченість)\}.$$

Звідси

$$\lambda_{nm}(Q_M^{K_{B_1}, K_{B_2}}) =$$

$$= \{K_{B_+}(\text{пористість}), K_{B_-}(\text{проникність})\}.$$

Збережена властивість когерентності

$$K_{B_2} = K_{B_1} \circ \lambda_{nm}(Q_M^{K_{B_1}, K_{B_2}}).$$

Таким чином, ми показали, що $K_{B_2} \in Q_M$ -модифікацією для K_{B_1} .

Тепер покажемо, що K_B^{QM} є Q_M -модифікацією для K_B^{noch} . Припустимо, що правила з модифікаційного запиту Q_M є такими, що не всі з них є одночасно хибними, тоді E_M є найменшою моделлю для $(\Delta_R^{noch})^{E_M} \cup \langle (\Delta_R^{in})^{E_M} \rangle \cup Q_M$. Для того, щоб

обчислити залишок $Q_M \Big|_{K_B^{noch}, K_B^{QM}}$ для Q_M слід видалити з тіла кожного модифікаційного правила всі атомарні предикати $K_{B_+}(o)$, такі, що $o \in K_B^{noch} \cap K_B^{QM}$, і всі атомарні предикати $K_{B_-}(o)$, такі, що $o \in K_B^{'noch} \cap K_B^{'QM}$. Це і будуть ті атомарні предикати, що є істинними в E_M завдяки множині $\langle (\Delta_R^{in})^{E_M} \rangle$. Тому E_M залишається найменшою моделлю модифікованого запиту

$$(\Delta_R^{noch})^{E_M} \cup \langle (\Delta_R^{in})^{E_M} \rangle \cup (Q_M \Big|_{K_B^{noch}, K_B^{QM}}).$$

Тоді необхідна модифікація $\lambda_{nm}(Q_M \Big|_{K_B^{noch}, K_B^{QM}})$ є найменшою моделлю для $Q_M \Big|_{K_B^{noch}, K_B^{QM}}$. Розглянемо тепер дві множини

$$O_+ = \{o : K_{B_+}(o) \in \lambda_{nm}(Q_M \Big|_{K_B^{noch}, K_B^{QM}})\} \text{ і}$$

$$O_- = \{o : K_{B_-}(o) \in \lambda_{nm}(Q_M \Big|_{K_B^{noch}, K_B^{QM}})\}.$$

Тоді запит

$$(\Delta_R^{noch})^{E_M} \cup \langle (\Delta_R^{in})^{E_M} \rangle \cup (Q_M \Big|_{K_B^{noch}, K_B^{QM}})$$

складається із трьох незалежних частин, причому залишок $Q_M \Big|_{K_B^{noch}, K_B^{QM}}$ більше не містить атомарних предикатів, які би входили до двох інших частини. Тоді множина E_M буде складатися з:

$$\{K_{B_+}^{\in}(o) : o \in K_B^{noch}\} \cup \{K_{B_-}^{\in}(o) : o \in K_B^{'noch}\},$$

$$\{K_{B_+}(o) : o \in K_B^{QM} \cap K_B^{noch}\} \cup$$

$$\cup \{K_{B_-}(o) : o \in K_B^{'QM} \cap K_B^{'noch}\},$$

$$\{K_{B_+}(o) : o \in O_+\} \cup \{K_{B_-}(o) : o \in O_-\}$$

на основі чого робимо висновок, що

$$K_B^{QM} = \{o : K_{B_+}(o) \in E_M\} = (K_B^{QM} \cap K_B^{noch}) \cup O_+,$$

$$K_B^{'QM} = O \setminus K_B^{QM} = \{o : K_{B_-}(o) \in E_M\} =$$

$$= (K_B^{'QM} \cap K_B^{'noch}) \cup O_-.$$

Таким чином, множина необхідних модифікацій $\lambda_{nm}(Q_M \Big|_{K_B^{noch}, K_B^{QM}})$ є когерентною.

Також $K_B^{QM} = (K_B^{noch} \cup O_+) \setminus O_-$. Звідки випливає, що $K_B^{QM} \in Q_M$ -модифікацією для K_B^{noch} .

Розглядатимемо випадок, коли мітки задані у вигляді чисел з діапазону $[0;1]$.

Нехай тепер ми маємо дві характеристики свердловини o і p . У вихідній базі знань дані характеристики представлено з певним коефіцієнтом впевненості. Наприклад, якщо коефіцієнт впевненості в характеристиці $o \in C_F^1$, $(0 \leq C_F^1 \leq 1)$, тоді $K_B^{noch}(in^+) = \langle C_F^1, 1 - C_F^1 \rangle$, де

in^+ – атомарний предикат. Ми можемо інтерпретувати перший і другий елементи в кортежі $(C_F^1, 1 - C_F^1)$ як степені впевненості і невпевненості в надійності інформації. Задача, яка виникає в даній ситуації, полягає у виведенні дійсного значення характеристики, виходячи із суб'єктивної оцінки експерта. Таким чином, всі, Q_m -модифікації для вихідної бази знань

K_B^{noch}, K_B^{QM} повинні представляти реальні значення фізичних характеристик свердловини. Припустимо тепер, що перешкода H_1 не може занизити значення характеристики більше, ніж на 0.1. Тоді можна припустити, що характеристика o є присутньою, коли коефіцієнт впевненості в даному факті є 0.9 і більше. Припустимо тепер, що перешкода H_2 не може впливати на характеристику p більше ніж на 0.3. Тому, якщо коефіцієнт впевненості щодо відсутності характеристики є принаймні 0.7, дана характеристика повинна бути відсутньою. Ці передумови у сукупності з фактом, що для кожної характеристики беруться до уваги два значення $\{\text{присутня}, \text{відсутня}\}$, можуть бути представлені:

$$\{(K_{B_+}(o) : 1) \ll (K_{B_+}(o) : 0.9), (K_{B_-}(o) : 0.7),$$

$$(K_{B_-}(p) : 1) \ll (K_{B_+}(o) : 0.9), (K_{B_-}(p) : 0.7)\}, \quad (3)$$

$$(K_{B_+}(p) : 1) \ll (K_{B_+}(p) : 0.9), (K_{B_-}(o) : 0.7),$$

$$(K_{B_-}(o) : 1) \ll (K_{B_+}(p) : 0.9), (K_{B_-}(p) : 0.7)\}. \quad (4)$$

Перших два правила стверджують, що якщо коефіцієнт впевненості щодо наявності характеристики o є принаймні 0.9, а щодо відсутності p – принаймні 0.7, тоді характеристика o буде присутньою із коефіцієнтом впевненості 1, а характеристика p відсутньою з коефіцієнтом впевненості 1. Нехай тепер коефіцієнт

впевненості щодо присутності характеристик o і p складає відповідно 0.4 і 0.8. Тоді $K_B^{noch}(o) = \langle 0.4, 0.6 \rangle$, $K_B^{noch}(p) = \langle 0.8, 0.2 \rangle$. Після виконання Q_m - модифікації для K_B^{noch} ми одержимо реальні значення характеристики в базі знань, а саме:

$$K_B^{Q_m}(o) = \langle 0, 1 \rangle, K_B^{Q_m}(p) = \langle 1, 0 \rangle.$$

Для M введемо оціночну функцію Θ , яка буде описувати наявну інформацію про належність елементів O деякій множині T , $T \subseteq O$. Тому $\Theta(K_{B+}(o)) = C_F$ будемо інтерпретувати, як факт того, що $o \in T$ з впевненістю C_F . Оціночна функція Θ задовольняє модифікаційний атом $(K_{B+}(o) : C_F)$ якщо $\Theta(K_{B+}(o)) \geq C_F$. За аналогією Θ задовольняє $(K_{B-}(o) : C_F)$, якщо $\Theta(K_{B-}(o)) \geq C_F$. Загалом вважатимемо, що Θ задовольняє множину модифікаційних атомів, якщо вона задовольняє кожен з елементів з множини. Будемо вважати, що оціночна функція Θ – задовольняє модифікаційне правило з мітками, якщо вона задовольняє заголовок правила в усіх випадках, коли задовольняє тіло правила. Отже Θ задовольняє модифікаційному предикатному запиту з мітками, або є моделлю запиту, якщо вона задовольняє всім правилам в запиті.

Для даного модифікаційного предикатного запиту із мітками, введемо оператор Λ_{Q_m} на множині значень оціночної функції Θ . Нехай $\Lambda_{Q_m}(\Theta)$ – множина заголовків всіх правил в Q_m , тіла яких задовольняються оціночною функцією Θ . Таким чином, оператор Λ_{Q_m} введемо наступним чином

$$\Lambda_{Q_m}(\Theta) = \bigcup_n \{C_F \mid (r_n : C_F) \in \Lambda_{Q_m}(\Theta), n \geq 0\}, \quad (5)$$

де r_n – модифікаційне правило.

Очевидно, що під час побудови оціночних функцій $\Theta(M)$ для кожного елемента $o \in O$, можна поставити у відповідність пару елементів з M , що відповідають модифікаційним атомам $K_{B+}(o)$ і $K_{B-}(o)$. Таким чином, ми повинні розглянути структуру $M^2 = M \times M$ із введеною операцією \leq_l , яку означимо так:

$$(C_{F_1}^1, C_{F_2}^1) \leq_l (C_{F_1}^2, C_{F_2}^2), \text{ якщо} \\ C_{F_1}^1 \leq_l C_{F_1}^2 \text{ і } C_{F_2}^1 \leq_l C_{F_2}^2. \quad (6)$$

Якщо пару $(C_{F_1}^1, C_{F_2}^1)$ – розглядати як міру нашого знання про належність об'єкта o множині T , тоді якщо $C_{F_1}^1 \leq_l C_{F_1}^2$ і $C_{F_2}^1 \leq_l C_{F_2}^2$, тоді пара $(C_{F_1}^2, C_{F_2}^2)$ більш точно характеризує об'єкт o . Тому операцію \leq_l можна розглядати,

як відношення порядку, задане на множині коефіцієнтів впевненості для тверджень бази знань.

Введені означення є обгрунтованими, оскільки підчас їх побудови ми не виходили за рамки процедури обчислення обгрунтованих семантик, прийнятих в теорії абстрактного логічного програмування і стабільних семантик для абстрактних логічних програм.

Висновки

В даній роботі введено поняття необхідної модифікації як інструмента трансформації баз знань нафтогазової предметної області. Подальші дослідження даного напрямку будуть спрямовані на розширення формально-логічного апарату необхідних модифікацій в рамках побудови загальної теорії модифікаційних предикатних запитів для інформаційних інтелектуальних систем на основі баз даних і знань нафтогазової предметної області.

Література

- 1 Шекета В.І. Інформаційна система для прогнозування нафтогазоносних покладів: дисертація на здобуття наукового ступеня кандидата технічних наук: 17.09.99 / Шекета Василь Іванович. – Херсон, 1999. – 130 с.
- 2 Cooperative behavior through Request Modification/Proc. 10th Conference on the Entity-Relationship Approach. – 1991. – P. 607–621.
- 3 The potential and actual effectiveness of interactive query expansion /Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. – Philadelphia. – 1997. – P. 324-331.
- 4 Howe A.E. A meta-search engine that learns which search engines to query/Howe A.E. – AI Magazine. – №18(2). – 1997. – P. 20-28.
- 5 Architecture of a metasearch engine that supports user information needs / Proc. Eighth International Conference on Information Knowledge Management (CIKM'99). – Kansas City. – P. 210–216.
- 6 Glover E. Improving category specific web search by learning query modifications/ Glover E. , Flake G. , Lawrence S./ Proc. of Symposium on Applications and the Internet. – San Diego. – 2001. – P. 8–12.
- 7 Шекета В.І. Модифікаційні предикатні запити, як інструмент підтримки діалогу з користувачем в інформаційних системах на основі баз даних і знань / Шекета В.І. – Тернопіль: Вісник Тернопільського державного технічного університету. Серія Математичне моделювання. – 2003. – 231с.