

# КОНТРОЛЬ, АВТОМАТИКА ТА ЕЛЕКТРОТЕХНІКА

УДК 681.3:622.276

## ПОБУДОВА ІНФОРМАЦІЙНОЇ ПРЕДИКАТНОЇ СХЕМИ ЯК СЕРЕДОВИЩА ВИКОНАННЯ ТРАНСФОРМАЦІЇ ЗАПИТІВ КОРИСТУВАЧА ЩОДО НАПІВСТРУКТУРОВАНОЇ ІНФОРМАЦІЇ НАФТОГАЗОВОЇ ПРЕДМЕТНОЇ ОБЛАСТІ

В.І. Шекета

ІФНТУНГ, кафедра прикладної математики, e-mail: [sheketa@rambler.ru](mailto:sheketa@rambler.ru), тел. 4-21-23

*Стаття посвячена проблемі представлення інформації, котора наповняє бази даних і бази знань нафтогазової предметної області в виде графов.*

*Выполнено построение информационной предикатной схемы на множестве унарных предикатов, что позволяет выполнять интерпретацию объектов нафтогазової предметної області путем выполнения сопоставления информационной предикатной схемы и всех подобъектов базы данных, которые ей соответствуют. Построенный таким образом формально-логический аппарат делает возможным выполнение трансформации условных SQL запросов на множестве минимальных совпадений информационной предикатной схемы и множества подобъектов базы данных формализованных на основе ориентированных графов с метками. Следующие исследования данного направления будут направлены на расширения абстрактного домена построеной информационной предикатной схемы к уровню домена информационно-поисковых задач нафтогазової предметної області на основе ограничений.*

Розглянемо проблему представлення інформації, що наповнює бази даних і бази знань нафтогазової предметної області у вигляді графів [1,2,3,4,8,9]. В ряді робіт представлення інформації у вигляді графової моделі розглядається як уніфікована ідея для представлення напівструктурованої інформації [5,6,7,13]. Результатом таких підходів є побудова різноманітних інформаційних систем і структур даних, які функціонують на основі процедур трансформції графів [10,11]. У той же час недослідженим залишається питання побудови формально-логічних структур для аналізу запитів користувача щодо баз даних і знань через введення формалізації предметної області на основі напрям-

*Given article is devoted to a problem of representation of the information which fills databases and knowledge bases of an oil-and-gas knowledge domain with means of graphs.*

*Construction of the information predicate circuit on the set of unary predicates that allows to carry out interpretation of objects of an oil-and-gas subject domain by means of comparison of the information predicate circuit and all subobjects of database which corresponds her is done.*

*The formal - logic apparatus constructed thus makes possible performance of transformation for conditional SQL queries over the set of the minimal concurrences of the information predicate circuit and set of subobjects of database formalized on the basis of oriented graphs with labels.*

*The forthcoming researches of the given problem will be directed on expansions of the abstract domain of the constructed information predicate circuit to a level of the domain of information retrieval tasks of an oil-and-gas subject domain on the basis of restrictions.*

лених графів з позначками.

Ціль даної статті полягає в побудові інформаційної предикатної схеми на множині унарних предикатів, що дасть змогу виконувати інтерпретацію об'єктів нафтогазової предметної області, представлених у вигляді напрямлених графів із позначками за допомогою скінченного набору предикатів, що в кінцевому підсумку дасть змогу розширити введене формальне означення до рівня формального апарату трансформції запитів, що працює з усіма підоб'єктами бази даних, які відповідають введеної інформаційній предикатній схемі.

Ми не будемо накладати ніяких особливих обмежень на графи, що використовуються. Зо-



крема. будемо використовувати написи (позначки) на вершинах і вітках графу, не зводитимемо граф тільки до випадку дерева і т. ін.

**Означення 1.** Упорядковану четвірку  $\Gamma=(B,R,f_1,f_2)$  вважатимемо напрямленим інформаційним графом, якщо  $B$  – множина вершин,  $R$  – множина дуг, а  $f_1$  і  $f_2$  – функції, які для кожної дуги визначають її початок і кінець. Як синонім для терміну вершина будемо використовувати термін “вузол”. Але будемо вважати обов’язковим використання терміну “дуга”, щоб підкреслити напрямленість інформаційних графів. В нашій моделі два вузли можуть бути з’єднані за допомогою більше, ніж з однієї дуги. Крім того, ми припускаємо використання петель. Введемо такі означення.

**Означення 2.** Нехай  $N$  – множина написів, тоді впорядкована п’ятірка  $\Gamma=(B,R,f_1,f_2,n)$  є  $(N-)$  описаним напрямленим графом, якщо  $(B,R,f_1,f_2,n)$  є напрямленим графом і  $n: B \cup R \rightarrow N$  є функцією, яка присвоює кожній вершині і дузі напис з  $N$ .

Таким чином, інформаційний об’єкт нафтогазової предметної області будемо представляти описаним напрямленим графом. В даному випадку під об’єктом ми розуміємо базу даних. Як приклад напівструктурованої бази даних побудуємо такий граф.

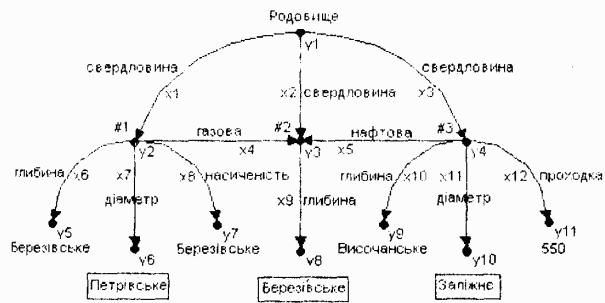


Рисунок 1 - Описаний напрямлений граф

**Означення 3.** Непорожню послідовність  $(b_{i_0}, r_{i_1}, b_{i_1}, \dots, r_{i_m}, b_{i_m})$  називатимемо обходом графу  $\Gamma=(B,R,f_1,f_2)$ , якщо  $f_1(r_{i_j})=b_{i_{j-1}}$  і  $f_2(r_{i_j})=b_{i_j}$  для всіх додатних  $j \leq m$ . Якщо всі дуги в обході попарно відмінні, то такий обхід називається маршрутом. Якщо до того ж всі вершини попарно відмінні – обхід називатимемо шляхом.

Як правило, ми будемо вказувати лише послідовність дуг коли задаватимемо обхід графу. Такий підхід дає змогу встановлювати відсутні вершини за допомогою відповідних функцій  $f_1$  і  $f_2$ . Винятком є лише обхід, що складається лише з однієї вершини. Кількість дуг в обході вважатимемо довжиною обходу. Обходи одиначної довжини називатимемо атомарними. В наших дослідженнях ми особливо зацікавлені в побудові маршрутів для графу. Основна перевага над простим обходом полягає в тому, що

множина всіх маршрутів завжди є скінченною. Основна перевага над операцією “шлях” полягає в тому, що ідея маршруту є більш загальною. Будемо позначати множину всіх непорожніх маршрутів через  $T'$  і включати цю множину в граф за потреби, тобто  $G=(B,R,T',f_1,f_2)$ . Введені нами функції  $s$  і  $t$  розширимо таким чином, щоб вони покривали операції обходу, маршруту і шляху в графі

$$f_1((b_{i_0}, r_{i_1}, b_{i_1}, \dots, r_{i_m}, b_{i_m})) := b_{i_0},$$

$$f_2((b_{i_0}, r_{i_1}, b_{i_1}, \dots, r_{i_m}, b_{i_m})) := b_{i_m}.$$

Ми вважатимемо два обходи  $o_1$  і  $o_2$  такими, що можуть бути об’єднаними, якщо  $f_2(o_1)=f_1(o_2)$ . Об’єднанням обходів  $o_1$  і  $o_2$ , де

$$o_1 = (b_{1_0}, r_{1_1}, b_{1_1}, \dots, r_{1_m}, b_{1_m}) \text{ і}$$

$$o_2 = (b_{2_0}, r_{2_1}, b_{2_1}, \dots, r_{2_n}, b_{2_n}) \text{ визначимо, як}$$

$$o_1 \circ o_2 = (b_{1_0}, r_{1_1}, b_{1_1}, \dots, r_{1_m}, b_{1_m} = b_{2_0}, r_{2_1}, b_{2_1}, \dots, r_{2_n}, b_{2_n})$$

Таке об’єднання є асоціативним, тобто для обходів  $f_1, f_2, f_3$  справедливо

$$(o_1 \circ o_2) \circ o_3 = o_1 \circ (o_2 \circ o_3).$$

Для специфікації запитів щодо бази даних введемо такі означення.

**Означення 4.** Об’єкт

$\Omega_2 = (B^{(\Omega_2)}, R^{(\Omega_2)}, f_1^{(\Omega_2)}, f_2^{(\Omega_2)}, n^{(\Omega_2)})$  є під-об’єктом об’єкта

$\Omega_1 = (B^{(\Omega_1)}, R^{(\Omega_1)}, f_1^{(\Omega_1)}, f_2^{(\Omega_1)}, n^{(\Omega_1)})$  якщо  $B^{(\Omega_2)} \subseteq B^{(\Omega_1)}, R^{(\Omega_2)} \subseteq R^{(\Omega_1)},$

$$f_1^{(\Omega_2)} \subseteq f_1^{(\Omega_1)} \Big|_{R^{(\Omega_2)}}, \quad f_2^{(\Omega_2)} \subseteq f_2^{(\Omega_1)} \Big|_{R^{(\Omega_2)}},$$

$$n^{(\Omega_2)} = n^{(\Omega_1)} \Big|_{B^{(\Omega_2)} \cup R^{(\Omega_2)}}$$

Позначатимемо даний факт через  $\Omega_2 \subseteq \Omega_1$ .

Для даного об’єкта  $\Omega$  позначатимемо всю множину його підоб’єктів через  $\theta(\Omega)$ .

**Лема 1.** Для заданого об’єкта  $\Omega$ , структура  $[\theta(\Omega), \subseteq]$  є частково впорядкованою множиною, тобто  $\subseteq$  є рефлексивним антисиметричним і транзитивним бінарним відношенням над  $\theta(\Omega)$ .

**Дов.**  $\subseteq$  є рефлексивним, тому що  $\Omega \subseteq \Omega$  є справедливим для будь-якого об’єкта  $\Omega$ . Нехай  $\Omega_1, \Omega_2$  – два об’єкти, причому  $\Omega_1 = (B^{(\Omega_1)}, R^{(\Omega_1)}, f_1^{(\Omega_1)}, f_2^{(\Omega_1)}, n^{(\Omega_1)})$  і  $\Omega_2 = (B^{(\Omega_2)}, R^{(\Omega_2)}, f_1^{(\Omega_2)}, f_2^{(\Omega_2)}, n^{(\Omega_2)})$ . Тоді  $\Omega_1 \subseteq \Omega_2$  і  $\Omega_2 \supseteq \Omega_1$  означають, що  $B^{(\Omega_1)} = B^{(\Omega_2)}$  і  $R^{(\Omega_1)} = R^{(\Omega_2)}$ , і таким чином  $\Omega_1 = \Omega_2$ . Звідки випливає, що  $\subseteq$  є антисиметричною. Нехай  $\Omega_3 = (B^{(\Omega_3)}, R^{(\Omega_3)}, f_1^{(\Omega_3)}, f_2^{(\Omega_3)}, n^{(\Omega_3)})$ .  $\Omega_1 \subseteq \Omega_2$  і  $\Omega_2 \supseteq \Omega_1$  означають  $B^{(\Omega_1)} \subseteq B^{(\Omega_2)}$  і  $R^{(\Omega_1)} \subseteq R^{(\Omega_2)}$ . Таким чином, ми довели твердження для  $f_1$ . Для функцій  $f_2$  і функції описів  $n$  доведення може бути здійснено подібним чином. Із  $\Omega_1 \subseteq \Omega_2$  і  $\Omega_2 \subseteq \Omega_3$  випливає  $f_1^{(\Omega_3)} = f_1^{(\Omega_2)} \Big|_{R^{(\Omega_3)}} = (f_1^{(\Omega_1)} \Big|_{R^{(\Omega_2)}}) \Big|_{R^{(\Omega_3)}}$ .



Оскільки  $R^{(\Omega_3)} \subseteq R^{(\Omega_2)}$ , то це є рівним  $f^{(\Omega_1)} \Big|_{R^{(\Omega_3)}}$ . Звідси випливає, що операція  $\subseteq$  є транзитивною.

**Лема 2.** Для заданого об'єкту  $\Omega$ ,  $[\theta(\Omega), \subseteq]$  є повною структурою, тобто кожна непорожня підмножина  $\theta(\Omega)$  має найменшу верхню і найбільшу нижню границі.

Дов. Нехай  $\Omega$  - заданий об'єкт і  $\Lambda = \{\Omega_1, \dots, \Omega_m\}$  підмножина  $\theta(\Omega)$ . Тоді

$$\Omega_{\text{mg}} = (\bigcap_i B^{(\Omega_i)}, \bigcap_i R^{(\Omega_i)}, f_1^{(\Omega)}, \dots)$$

$$\left( \bigcap_i R^{(\Omega_i)} \cdot f_2^{(\Omega)} \Big|_{\bigcap_i R^{(\Omega_i)}}, n^{(\Omega)} \Big|_{\bigcap_i B^{(\Omega_i)} \cup R^{(\Omega_i)}} \right)$$

є об'єктом перетину. Ми показали, що  $\Omega_{\text{mg}}$  є найбільшою нижньою границею для  $\Lambda$ . Побудова найменшої верхньої границі може бути здійснена подібним чином, використовуючи об'єкт об'єднання. В першу чергу слід зауважити, що  $\Omega_{\text{mg}}$  насправді є об'єктом. Якщо дуга належить  $R^{(\Omega_{\text{mg}})} = \bigcap_i R^{(\Omega_i)}$ , то вона належить і всім  $R^{(\Omega_i)}$ . Тоді її початкова і кінцева вершини належить  $B^{(\Omega_i)}$ , і таким чином також належить  $B^{(\Omega_{\text{mg}})}$ . Таким чином,  $\Omega_{\text{mg}} \subseteq \Omega_i$  справджується для всіх  $1 \leq i \leq m$ , тому що  $B^{(\Omega_{\text{mg}})} \subseteq B^{(\Omega_i)}$  і  $R^{(\Omega_{\text{mg}})} \subseteq R^{(\Omega_i)}$ , звідки  $\Omega_{\text{mg}}$  є нижньою границею для  $\Lambda$ . Припустимо, ми маємо іншу нижню границю  $\Omega'$  для  $\Lambda$ , і  $\Omega' \subseteq \Omega_{\text{mg}}$  не справджується, тоді існує вершина або дуга  $y \in B^{(\Omega')} \cup R^{(\Omega')}$ , причому  $y \notin B^{(\Omega_{\text{mg}})} \cup R^{(\Omega_{\text{mg}})}$ . Звідси випливає, що існує об'єкт  $\Omega_i$ , для якого  $y \notin B^{(\Omega_i)} \cup R^{(\Omega_i)}$ . Але це відразу ж приводить до  $\Omega' \not\subseteq \Omega_i$ , що суперечить факту, що  $\Omega'$  є нижньою границею для  $\Lambda$ .

Наведені леми доводять, що ми спроектували добре структуроване середовище, де означено поняття мінімального елемента.

**Відображення між графами**

Побудуємо означення відображення між графами з точки зору їх структури і використаємо ідею графового морфізму [5,12].

**Означення 5.** Під графовим морфізмом з одного напрямленого графу  $\Gamma_1 = (B^{(\Gamma_1)}, R^{(\Gamma_1)}, f_1^{(\Gamma_1)}, f_2^{(\Gamma_1)})$  в інший напрямлений граф  $\Gamma_2 = (B^{(\Gamma_2)}, R^{(\Gamma_2)}, f_1^{(\Gamma_2)}, f_2^{(\Gamma_2)})$  будемо розуміти функцію  $\sigma: B^{(\Gamma_1)} \cup R^{(\Gamma_1)} \rightarrow B^{(\Gamma_2)} \cup R^{(\Gamma_2)}$ , таку що:

1. Для всіх  $y \in B^{(\Gamma_1)} \cup R^{(\Gamma_1)}$  вірно, що  $\sigma(y) \in B^{(\Gamma_2)}$  тоді і тільки тоді, коли  $y \in B^{(\Gamma_1)}$  і  $\sigma(y) \in R^{(\Gamma_2)}$  тоді і тільки тоді, коли  $y \in R^{(\Gamma_1)}$ .

$$2. f_1^{(\Gamma_2)} \circ \sigma \Big|_{R^{(\Gamma_1)}} = \sigma \circ f_1^{(\Gamma_1)} \quad \text{і}$$

$$f_2^{(\Gamma_2)} \circ \sigma \Big|_{R^{(\Gamma_1)}} = \sigma \circ f_2^{(\Gamma_1)}.$$

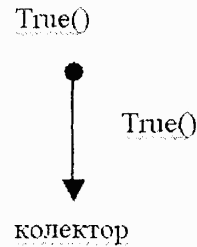
На відміну від підходу, при якому розглядається процедура трансформації графів[10,11] ми не вимагаємо збереження позначок при морфізмі графів. Ця вимога є надто жорсткою для нашого випадку, оскільки ми будемо розглядати різні види графів. Означений морфізм  $\sigma$  будемо розглядати, як частковий, якщо не всі елементи з  $B^{(\Gamma_1)} \cup R^{(\Gamma_1)}$  лежать в області визначення  $\sigma$ .

**Інформаційна предикатна схема**

**Означення 6.** Інформаційною предикатною схемою на множині унарних предикатів  $\Pi$  будемо називати об'єкт  $\eta = (B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)}, n^{(\eta)})$ , де кожному елементу приписаний відповідний предикат  $(n: B^{(\eta)} \cup R^{(\eta)} \rightarrow \Pi)$ .

При цьому ми використаємо два предикати: текстову константу і предикат True().

Щоб встановити зв'язок між інформаційною предикатною схемою та об'єктами, які вона описує, ми введемо поняття відповідності[8,9] між ними. Залежно від напрямку відображення будемо говорити, що ми співставляємо предикатну схему з об'єктом, або інтерпретуємо об'єкт за допомогою схеми.



**Рисунок 2 - Проста інформаційна предикатна схема**

**Означення 7.** Під співпаданням (або функцією співпадання) інформаційної предикатної схеми  $s$  з об'єктом  $\Omega$  будемо розуміти ізоморфне вбудовування предикатної схеми  $\eta$  в  $\Omega$ , тобто повний ін'єктивний графовий морфізм  $\sigma: \eta \rightarrow \Omega$ , такий, що для всіх  $y \in B^{(\eta)} \cup R^{(\eta)}$  предикат  $n^{(\eta)}(y)$  є істинним для  $n^{(\Omega)}(\sigma(y))$ .

Якщо існує співпадання інформаційної предикатної схеми  $\eta$  з об'єктом  $\Omega$ , то будемо говорити, що  $\Omega$  відповідає  $\eta$  (або що  $\Omega$  може бути інтерпретований з допомогою  $\eta$ ) і будемо називати  $\Omega$  екземпляром або співпаданням  $\eta$ .

Нехай  $\Omega$  - деяка база даних,  $\eta$  – інформаційна предикатна схема і  $\Omega_1 \subseteq \Omega$  є співпаданням  $\eta$ . Тоді кожен надоб'єкт  $\Omega_2$  для  $\Omega_1 \subseteq \Omega_2 \subseteq \Omega$  є також співпаданням  $\eta$ . Нехай  $\Phi^{(\eta)}(\Omega)$  позначає множину всіх співпадань  $\eta$  в  $\Omega$ . Оскільки  $\Phi^{(\eta)}(\Omega)$  є підмножиною  $\theta(\Omega)$ , то структура  $[\Phi^{(\eta)}(\Omega), \subseteq]$  є також частково впорядкованою множиною. Будемо називати мінімальний еле-



мент в цій частково впорядкованій множині мінімальним співпаданням або мінімальним екземпляром  $\eta$  в  $\Omega$ . Ми позначатимемо множини всіх мінімальних співпадань  $\eta$  в  $\Omega$  через  $\Phi_{\min}^{(\eta)}(\Omega)$ . На рисунку 3 ми побудуємо таку саму інформаційну предикатну схему, як і на рисунку 2, але цього разу відобразимо також мінімальні співпадання в базі даних.

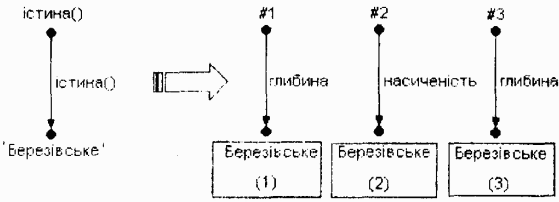


Рисунок 3 - Інформаційна предикатна схема та її мінімальні співпадання

Додавання змінних

Крок за кроком ми будемо вдосконалювати експресивність інформаційної предикатної схеми, щоби забезпечувати більшу гнучкість в представленні даних. Введемо означення змінних для інформаційної предикатної схеми, що дасть змогу нам підтримувати зв'язки між різними частинами бази даних, спираючись на введені поняття позначок для ребер інформаційного графу.

Нехай  $\eta$  - інформаційна предикатна схема,  $Z$ - множина змінних і  $z^{(n)}: B^{(n)} \cup R^{(n)} \rightarrow Z$  відображення множини вузлів і ребер інформаційної предикатної схеми в множини змінних. Тоді впорядкований кортеж  $(B^{(n)}, R^{(n)}, f_1^{(n)}, f_2^{(n)}, n^{(n)}, z^{(n)})$  будемо називати інформаційною предикатною схемою із змінними. Інтуїтивно зрозуміло, вузли і ребра, що відображаються на ті самі змінні є "з'єднаними", тобто їх написи повинні бути однаковими на екземплярі, що одержується. Тому ми додатково вимагаємо, що для відображення  $\sigma$ , що є співставленням  $\eta$  в об'єкт  $\Omega$ , щоби для всіх  $y_1, y_2 \in B^{(n)} \cup R^{(n)}$ , якщо  $z(y_1)$  і  $z(y_2)$  існує і  $z(y_1) = z(y_2)$ , то  $n^{(\Omega)}(\sigma(y_1)) = n^{(\Omega)}(\sigma(y_2))$ . Інформаційна предикатна схема зі змінними і мінімальними співпаданнями в базі даних представлені на рисунку 4. В цьому прикладі множина змінних  $Z$  складається лише з однієї змінної; і два вузли в ній. Цей факт відображається міткою  $Y: істина()$ .

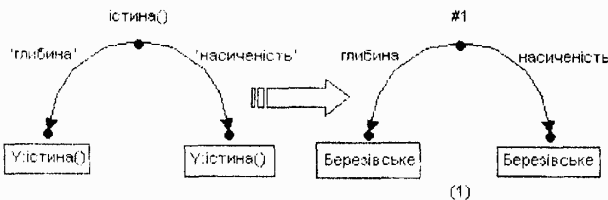


Рисунок 4 - Додавання змінних  
Додавання шляхів

Наведемо наше кінцеве означення інформаційної предикатної схеми. При цьому виходитимемо з того, що інформація бази даних є напівструктурованою. Фрагменти інформації можуть бути як комплексними, так і атомарними. Тому свого подальшого розвитку одержить і процедура обходу дерева.

Нехай  $\Gamma = (B, R, f_1, f_2)$  є напрямленим графом. Означене вище поняття маршруту є послідовністю дуг  $(r_{i_1}, \dots, r_{i_m})$ , де всі  $r_{i_j}$  є відмінними і існують вузли  $(b_{i_1}, \dots, b_{i_m})$ , такі, що для всіх  $r_{i_j}$   $f_1(r_{i_j}) = b_{i_{j-1}}$  і  $f_2(r_{i_j}) = b_{i_j}$ . Слід звернути увагу, що цей факт не вимагає, щоби  $b_{i_0}, \dots, b_{i_m}$  були відмінними. Кількість дуг в маршруті будемо називати довжиною маршруту. Множину всіх маршрутів в графі будемо позначати через  $T$ , а множину непорожніх - через  $T^+$ . Для непорожнього маршруту  $t_i = (r_{i_1}, \dots, r_{i_m}) \in T^+$  ми введемо функції  $f_{1T}, f_{2T}: T^+ \rightarrow B$ , які означимо класичним способом як  $f_{1T}(t_i) = f_1(r_{i_1})$  і  $f_{2T}(t_i) = f_2(r_{i_m})$  відповідно.

Тепер потрібно подати ідею відповідності між інформаційною предикатною схемою і об'єктами, що представляють шляхи. Для цього нам буде потрібно певне структурне відношення між графами, що дасть змогу пов'язати в певній залежності дуги і обходи графу. Як перший крок до досягнення цієї мети введемо поняття відповідного маршрутного графу.

**Означення 8.** Відповідний маршрутний граф для графу  $\Gamma = (B^{(\Gamma)}, R^{(\Gamma)}, f_1^{(\Gamma)}, f_2^{(\Gamma)})$  означимо як  $\Gamma_T = (B^{(\Gamma)}, T^{+(\Gamma)}, f_{1T}^{(\Gamma)}, f_{2T}^{(\Gamma)})$ .

Зрозуміло, що у відповідному маршрутному графі всі маршрути матеріалізуються як дуги. Введені означення є пов'язаним з ідеєю транзитивного закриття графу [11]. Але між цими ідеями є суттєва відмінність. Транзитивне закриття включає лише одну дугу для кожної пари досяжних вузлів, тоді як в нашому випадку ми включаємо кожен дугу, через яку вони можуть бути досягнуті. Тим не менше, маршрутний граф є завжди скінченим, тому що лише скінчена кількість маршрутів існує для кожного напрямленого графу.

**Лема 3.** Направлений граф завжди є підграфом відповідного маршрутного графу.

Ця лема справджується, тому що існує відображення  $r_i \rightarrow (r_i)$  дуг  $R$  в маршрути в  $T^+$ . Так само очевидно є і наступна лема.

**Лема 4.** Якщо напрямлений граф  $\Gamma_1$  є підграфом  $\Gamma_2$ , тоді відповідний маршрутний граф  $\Gamma_1$  є підграфом відповідного маршрутного графу  $\Gamma_2$ .

Тепер розширимо дані означення інформаційної предикатної схеми. Введемо дві додаткові функції  $\phi_{\min}$  і  $\phi_{\max}$ , що дозволять нам специфікувати обмеження довжини обходів в співпадаючих об'єктах. Крім того, використаємо множину змінних  $Z$ .

**Означення 9.** Нехай дано множину позначок  $N$  і множину змінних  $Z$ , тоді кортеж



$$(B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)}, n^{(\eta)}, z^{(\eta)}, \varphi_{\min}^{(\eta)}, \varphi_{\max}^{(\eta)})$$

будемо називати інформаційною предикатною схемою. Причому:

1.  $B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)}, n^{(\eta)}$  означається так само як в попередніх розділах.

2.  $z : B^{(\eta)} \cup R^{(\eta)} \rightarrow Z$  є частковим відображенням множини вузлів і дуг предикатної схеми в змінній.

3.  $\varphi_{\min}^{(\eta)} : R^{(\eta)} \rightarrow N^+$  і  $\varphi_{\max}^{(\eta)} : R^{(\eta)} \rightarrow N^+ \cup \{+\infty\}$  є обмеженнями на довжину.

Крім того, якщо для дуги  $r_i \in R^{(\eta)}$  існують зв'язані змінні  $z^{(\eta)}(r_i)$ , тоді  $\varphi_{\min}^{(\eta)}(r_i) = \varphi_{\max}^{(\eta)}(r_i) = 1$ .

Щоб розширити можливості інформаційної предикатної схеми переозначимо поняття співвідповідності між інформаційною предикатною схемою і об'єктами бази даних.

**Означення 10.** Нехай  $\eta = (B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)}, n^{(\eta)}, z^{(\eta)}, \varphi_{\min}^{(\eta)}, \varphi_{\max}^{(\eta)})$  - інформаційна предикатна схема і  $\Omega = (B^{(\Omega)}, R^{(\Omega)}, f_1^{(\Omega)}, f_2^{(\Omega)}, n^{(\Omega)})$  - об'єкт бази даних. Співставлення предикатної схеми  $\eta$  в  $\Omega$  є ізоморфним вбудовуванням  $\eta$  в  $\Omega$ , тобто ізоморфне вбудовування  $(B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)})$  в  $(B^{(\Omega)}, T^{(\Omega)}, f_1^{(\Omega)}, f_2^{(\Omega)})$  таке, що мають місце властивості:

1. Для всіх вузлів  $y \in B^{(\eta)}$  предикат  $n^{(\eta)}(y)$  є істинним для  $n^{(\Omega)}(\sigma(y))$ .

2. Для всіх  $y \in R^{(\eta)}$  предикат  $n^{(\eta)}(y)$  є істинним для міток  $n^{(\Omega)}(y_j)$  для всіх дуг  $y_j$  в маршруті  $\sigma(y)$ .

3. Для всіх елементів  $y_1, y_2 \in B^{(\eta)} \cup R^{(\eta)}$  для яких  $z^{(\eta)}(y_1)$  і  $z^{(\eta)}(y_2)$  існує і  $z^{(\eta)}(y_1) = z^{(\eta)}(y_2)$ , то і відповідні надписи є однаковими  $n^{(\Omega)}(\sigma(y_1)) = n^{(\Omega)}(\sigma(y_2))$ .

4. Для всіх дуг  $y \in R^{(\eta)}$  довжина маршруту  $\sigma(y)$  є принаймні  $\varphi_{\min}^{(\eta)}(y)$  і не більшою, ніж  $\varphi_{\max}^{(\eta)}(y)$ .

Якщо співпадання між інформаційною предикатною схемою  $\eta$  і об'єктом бази даних  $\Omega$  існує, ми будемо говорити, що  $\Omega$  відповідає  $\eta$ .

Наступна теорема підтверджує, що ми справді підсилили наше початкове означення інформаційної предикатної схеми і що наше нове означення не суперечить попередньому.

**Теорема 1.** Інформаційна схема  $\eta$  відповідає об'єкту  $\Omega$  в простій формі тоді і тільки тоді, коли вона відповідає  $\Omega$ , виходячи з передположення, що  $z^{(\eta)}$  - порожнє відображення, і  $\varphi_{\min}^{(\eta)}$  і  $\varphi_{\max}^{(\eta)}$  дорівнюють одиниці для всіх дуг в  $\eta$ .

**Дов.** Нехай  $\eta = (B^{(\eta)}, R^{(\eta)}, f_1^{(\eta)}, f_2^{(\eta)}, n^{(\eta)})$  - інформаційна предикатна схема. Нехай  $\Omega$  - об'єкт, що відповідає  $\eta$  в простій формі, і  $\sigma'$  є співпаданням між  $\eta$  і  $\Omega$ . Тоді ми можемо побудувати співпа-

дання  $\sigma$  між  $\eta$  і  $\Omega$  таким чином:

$$\sigma(y) := \begin{cases} \sigma'(y) & \text{якщо } y \in B^{(\eta)} \\ (\sigma'(y)) & \text{якщо } y \in R^{(\eta)} \end{cases}$$

Перш за все очевидно, що ми дійсно оголосили графовий морфізм  $m$ . Для кожної дуги у властивість  $f_{1T}^{(\Omega)}(\sigma(y)) = \sigma(f_1^{(\eta)}(y))$  справджується, тому що

$$f_{1T}^{(\Omega)}(\sigma(y)) = f_{1T}^{(\Omega)}((\sigma'(y))) = f_1^{(\Omega)}(\sigma'(y)) = \sigma'(f_1^{(\eta)}(y)) = \sigma(f_1^{(\eta)}(y)).$$

і  $\sigma'$  є графовим морфізмом. Аналогічно справджується  $f_{2T}^{(\Omega)}(\sigma(y)) = \sigma(f_2^{(\eta)}(y))$ .

Зараз ми перевіримо, чи  $\sigma$  задовольняє умовам співпадання. Умова на написи для вузлів дійсна для  $\sigma$ , оскільки вона дійсна для  $\sigma'$ . Кожна дуга  $y \in R^{(\eta)}$  відображається на  $\sigma'(y)$ . Цей маршрут завжди має довжину одиниця, що підтверджує істинність четвертої умови на довжину маршруту. Умова на написі для дуг також справджується, тому що вона справджується для  $\sigma'$ . І, нарешті, умова на зв'язаність змінних також є істинною, оскільки відображення змінних є порожнім.

І, навпаки, нехай  $\sigma$  є співпаданням  $\eta$  в  $\Omega$ ; і нехай  $z^{(\eta)}$  - порожнє відображення, і  $\varphi_{\min}^{(\eta)}$  і  $\varphi_{\max}^{(\eta)}$  дорівнює одиниці для всіх дуг в  $\eta$ . Тоді ми можемо спроекувати графовий морфізм  $\sigma'$  між  $\eta$  і  $\Omega$ , як показано нижче

$$\sigma'(y) := \begin{cases} \sigma(y) & \text{якщо } y \in B^{(\eta)} \\ \text{return\_1}(\sigma(y)) & \text{якщо } y \in R^{(\eta)} \end{cases}$$

В цьому означенні функція  $\text{return\_1}$  повертає першу дугу маршруту. Слід звернути увагу, що  $\sigma(y)$  завжди містить точно одну дугу, тому що  $\varphi_{\min}^{(\eta)}$  і  $\varphi_{\max}^{(\eta)}$  дорівнюють одиниці. Таким чином  $\sigma'$  зберігає порядок слідування «початок-кінець», тому що  $\sigma$  зберігає дану властивість

$$f_{1T}^{(\Omega)}(\sigma'(y)) = f_{1T}^{(\Omega)}(\text{return\_1}(\sigma(y))) = f_{1T}^{(\Omega)}(\sigma(y)) = \sigma(f_1^{(\eta)}(y)) = \sigma'(f_1^{(\eta)}(y)).$$

і те саме буде дійсним для функції  $f_2$ . Більше того, для всіх  $y \in B^{(\eta)} \cup R^{(\eta)}$  предикат  $n^{(\eta)}(y)$  дійсний для  $n^{(\Omega)}(\sigma'(y))$ , тому що він є дійсним для  $n^{(\Omega)}(\sigma(y))$ , якщо  $y$  є вершиною, і для всіх  $n^{(\Omega)}(\sigma(y))$ , якщо  $y$  є дугою.

Розглянемо приклад, зображений на рисунку 5. Знак «+» на першій дузі в схемі означає, що довжина обходів, які співпадають, лежить в межах  $(1; +\infty)$ . Будемо використовувати термін «обхід», коли будемо говорити про напівструктуровані дані, але ми будемо продовжувати використовувати термін «маршрут» коли буде потрібна математична коректність.

Розглянемо даний приклад більш детально. Рисунок 6 показує ту саму інформаційну предикатну схему у випадку другого співпадання. Слід зазначити, що співпадання ми розглядаємо



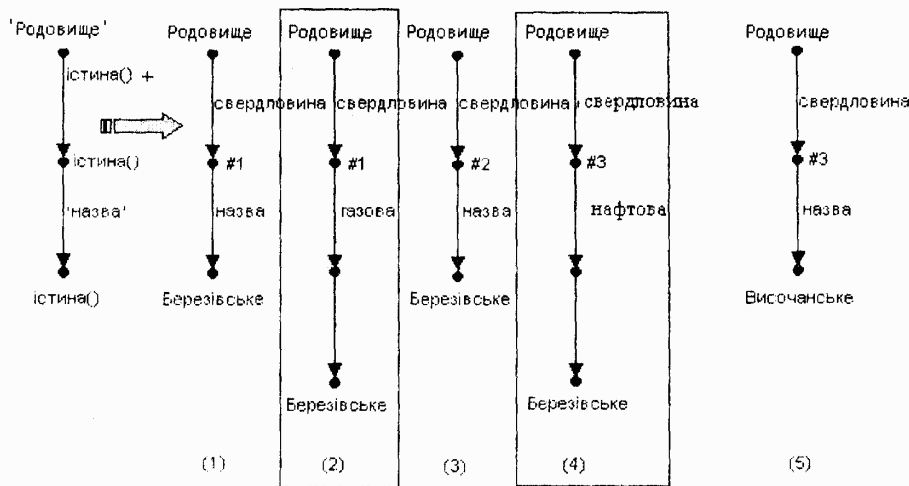


Рисунок 5 - Додавання шляхів

як підграф відповідного маршрутного графу бази даних. Співпадання інформаційної предикатної схеми позначені суцільними лініями, а пунктирна лінія позначає дуги решти відповідного маршрутного графу бази даних.

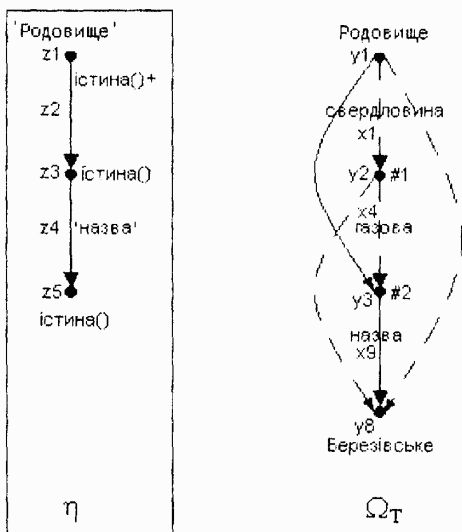


Рисунок 6 - Випадок відповідності

Для інформаційної предикатної схеми можна задати такий формальний опис:

$$B = \{y_1, y_3, y_5, \}, R = \{y_2, y_4\},$$

$$f_1 = \{(y_2, y_1), (y_4, y_3)\}, f_2 = \{(y_2, y_3), (y_4, y_5)\}$$

$$n = \{(y_1, 'родовище '), (y_2, 'істина ()),$$

$$(y_3, 'істина()), (y_4, 'назва'), (y_5, 'істина())\}.$$

$$z = \{ \}, \varphi_{\min} = \{(y_2, 1), (y_4, 1)\},$$

$$\varphi_{\max} = \{(y_2, +\infty), (y_4, 1)\}.$$

Співпадання  $\sigma$  між інформаційною предикатною схемою і частиною бази даних набуває такого вигляду:

$$\sigma = \{(y_1, z_1), (y_3, z_3), (y_5, z_5), (y_2, (r_1, r_4)), (y_4, (r_9))\}$$

Згідно з приведеним формальним описом співпадання  $\eta$  в  $\Omega$  будемо трактувати як факт того, що він є підоб'єктом  $\Omega$ . Хоча область  $\sigma(\eta)$  функції співпадання  $\Omega$  є підоб'єктом  $\Omega_T$ . Але при спробі адаптувати наведений опис до мінімального співпадання виникають деякі проблеми. Поняття мінімального співпадання є важливим для означення запиту по бази даних. Всі три співпадання є потенційно цікавими, але тільки перший є мінімальним, тому що він є підоб'єктом інших. З рисунка випливає, що всі три співпадання є результатом різних функцій співставлення. Діапазон їх відповідних областей є об'єктами, які не можна розглядати як підоб'єкти  $\Omega_T$ . Тому ми означимо мінімальне співпадання з огляду на функції співставлення. Для досягнення цього введемо функцію  $\Psi$ , яка одержує на вході підоб'єкт  $\Omega_T$  і дає на виході об'єкт  $\Omega$ . Фактично функція  $\Psi$  виконує декомпозицію маршрутів в дуги і додає вузли початку і кінця. Нехай  $\Omega'$  є підоб'єктом  $\Omega_T$ .

Тепер означимо множину мінімальних співпадань з  $\eta$  в  $\Omega$  як:  $\Phi_{\min}^{(\eta)}(\Omega) := \{\Psi(\sigma(\eta)), \sigma\}$

Таким чином, кожна функція  $\Psi(\sigma(\eta))$  є насправді співпаданням  $\eta$  в  $\Omega$ , тому що  $\eta$  може бути вбудований в  $\Psi(\sigma(\eta))_T$ , використовуючи  $\Omega$ , більше того наступна лема підтвердить, що ми дали дуже корисне означення мінімальних співпадань.

**Лема 5.** Для кожного співставлення  $\eta$  в  $\Omega$  існує співпадання  $\Omega'$  тобто для кожного елемента  $\Phi^{(\eta)}(\Omega)$  існує мінімальне співпадання  $\Phi_{\min}^{(\eta)}(\Omega)$ , яке є підоб'єктом  $\Omega'$ .

**Дов.** Нехай  $\Omega'$  є співпаданням  $\eta$  в  $\Omega$ . Звідси випливає, що існує функція співпадання  $\sigma$ , яка виконує вбудовування  $\eta$  в  $\Omega'_T$ . Оскільки  $\Omega' \subseteq \Omega$  і  $\Omega'_T \subseteq \Omega_T$ , відображення  $\sigma$  вбудовує  $\eta$  також і в  $\Omega_T$ . Таким чином,  $\Psi(\sigma(\eta))$  є елементом  $\Phi_{\min}^{(\eta)}(\Omega)$ . Якщо ми зможемо показати, що  $\Psi(\sigma(\eta)) \subseteq \Omega'$ , то наша лема буде доведена.



Нехай  $b$  – вузол в  $\Psi(\sigma(\eta))$ . Розглянемо два випадки: якщо  $b$  належить  $\sigma(\eta)$ , то тоді існує вузол  $b_\eta$  в  $\eta$ , таке що  $\sigma(b_\eta)=b$ . Звідки випливає  $b \in B^{(\Omega_\eta)}$  і, отже,  $b \in B^{(\Omega)}$ . Якщо все-таки  $b \notin \sigma(\eta)$ , то в цьому випадку існує дуга  $r$  в маршруті  $(r_1, \dots, r_j) \in \sigma(\eta)$ , таке що  $b$  є або початком або кінцем дуги  $r$ . Оскільки  $R^{(\Omega_\eta)} \subseteq R^{(\Omega)}$  і  $(r_1, \dots, r_j) \in R^{(\Omega_\eta)}$ , для всіх  $r_j$  в  $R^{(\Omega)}$ . Зокрема,  $r \in R^{(\Omega)}$ , і оскільки  $\Omega'$  є об'єктом,  $b \in B^{(\Omega')}$ .

Нехай тепер  $r$  є дугою в  $\Psi(\sigma(\eta))$ . Тоді існує дуга  $r_\eta$  в  $\eta$ , така що  $r \in \sigma(r_\eta)$ . Оскільки  $\sigma(r_\eta) \in R^{(\Omega_\eta)}$  і  $R^{(\Omega_\eta)} \subseteq R^{(\Omega)}$ , то звідси випливає, що всі дуги в  $\sigma(r_\eta)$  також належать і  $R^{(\Omega)}$ . Зокрема,  $r \in R^{(\Omega)}$ . Таким чином ми довели, що  $B^{(\Psi(\sigma(\eta)))} \subseteq B^{(\Omega)}$  і  $R^{(\Psi(\sigma(\eta)))} \subseteq R^{(\Omega)}$ ; оскільки  $\Psi(\sigma(\eta))$  і  $\Omega$  є об'єктами, то звідси  $\Psi(\sigma(\eta)) \subseteq \Omega'$ .

В попередніх розділах ми розглядали тільки непорожні маршрути і обходи. Причиною цього є легкість представлення функцій  $f_1, f_2$ . Але при більш детальному розгляді можна говорити також про порожні обходи. Ключовим моментом до розуміння цього підходу є те, що для кожного графу існує  $|B|$  порожніх обходів, тобто по одному для кожної вершини. Формально ці обходи складаються з однієї вершини, тому тепер означення функцій  $f_1, f_2$  стає простішим. Таким чином ідея порожнього обходу може бути інтегрована в означення інформаційної предикатної схеми.

### Прості запити по інформаційній предикатній системі

Вище ми бачили, що інформаційна предикатна схема сама по собі вже є простим видом запиту, який працює з всіма підоб'єктами бази даних, що відповідають предикатній схемі.

**Означення 11.** Під інформаційним предикатним запитом будемо розуміти  $\vartheta=(\eta)$ , де  $\eta$  є інформаційною предикатною схемою.

Розглядаючи таке означення, будемо оперувати не всією множиною співпадань предикатної схеми, але лише з мінімальними співпаданнями.

**Означення 12.** Під відповіддю на інформаційний предикатний запит  $\vartheta=(\eta)$  для бази даних  $\Omega$  будемо розуміти множину мінімальних співпадань при співставленні  $\eta$  з  $\Omega$ , тобто  $\Phi_{\min}^{(\eta)}(\Omega)$ .

Виходячи з вище даних двох означень ми можемо формулювати умовні запити. У грубому наближенні це відповідає операції вибірки в реляційній базі даних. Розглянемо класичний зразок SQL запиту:

*SELECT \* FROM свердловини WHERE назва="Петрівське"*

Але важко буде виразити запит

*SELECT родовище FROM свердловини WHERE назва="Петрівське"*

В цьому випадку ми в запиті хочемо знай-

ти назви родовищ в множині свердловин, хоча структура бази даних така, що множина свердловин складається з окремих підмножин, які в свою чергу є підмножинами множини родовищ.

**Означення 13.** Напрямленим запитом будемо називати кортеж  $\vartheta=(\eta_1, \eta_2)$ , де  $\eta_1$  – інформаційна предикатна схема,  $\eta_2$  – підоб'єкт  $\eta_1$ . Будемо називати  $\eta_2$  напрямом запиту.

**Означення 14.** Відповіддю на напрямлений запит  $\vartheta=(\eta_1, \eta_2)$  для бази даних  $\Omega$  будемо називати об'єднання мінімальних співпадань для  $\eta_2$  з усіма мінімальними співпаданнями для  $\eta_1$ , тобто  $\bigcup_{y \in \Phi_{\min}^{(\eta_1)}(\Omega)} \Phi_{\min}^{\eta_2}(y)$ .

Як приклад ми використовуємо другий з наведених вище SQL запитів.

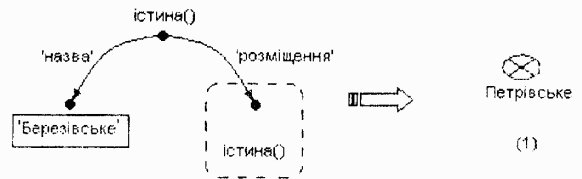


Рисунок 7 - Напрямленим запит

Запит на рисунку 7 виконує напрямлений пошук. Напряму запиту позначено пунктиром.

Таким чином, в даній статі виконано побудову інформаційної предикатної схеми на множині унарних предикатів, що дає змогу виконувати інтерпретацію об'єктів нафтогазової предметної області через виконання співставлення інформаційної предикатної схеми і всіх підоб'єктів бази даних, які їй відповідають. Одержаний результат є обґрунтованим, оскільки він є розширенням операції ін'єктивного графового морфізму.

В кінцевому підсумку побудований формально-логічний апарат дає змогу виконувати трансформацію умовних SQL запитів на множині мінімальних співпадань інформаційної предикатної схеми і множини підоб'єктів бази даних, формалізованих на основі напрямлених графів із позначками. Наступні дослідження даного напряму будуть спрямовані на розширення абстрактного домену побудованої інформаційної предикатної схеми до рівня домену інформаційно-пошукових задач на основі обмежень для нафтогазової предметної області.

### Літературні джерела

1. Глускин Л.М., Шор. Л.А., Шварц В.Я. Задачи и алгоритмы комбинаторики и теории графов. – Донецк: ДНИ, 1982. - 112с.
2. Кристофидес Н. Теория графов. Алгоритмический подход. - М.: Мир, 1978.-432с.
3. Оре О. Теория графов. - М.: Наука, 1980. - 336 с.
4. Харари Ф. Теория графов. - М.: Мир, 1973. - 302 с.
5. S. Abiteboul, D. Quass, J. McHugh, J. Widom, and J. Wiener. The Lorel query language for semistructured data. Journal of Digital



Libraries, 1(1):68-88, 1997.

6. P. Buneman. Semistructured data. In Proceedings of the Symposium on Principles of Database Systems

(PODS), pages 117-121, Tucson, AZ, USA, May 1997.

7. S. Chawathe, S. Abiteboul, and J. Widom. Representing and querying changes in semistructured data. In Proceedings of the International Conference on Data Engineering (ICDE), pages 4-13, Orlando, FL, USA, February 1998.

8. M. Fernandez and D. Suciu. Optimizing regular path expressions using graph schemas. In Proceedings of the International Conference on Data Engineering (ICDE), pages 14-23, Orlando, FL, USA, February 1998.

9. M. Henzinger, T. Henzinger, and P. Kopke. Computing simulations on finite and infinite graphs. In Proceedings of the Symposium on

перетворення достатньо мати 6-7 метрологічно атестованих точок, через які можна провести Foundations of Computer Science, pages 453-462, Milwaukee, WI, USA, October 1995.

10. M. Lowe. Algebraic approach to single-pushout graph transformation. Computer Science, 109:181-224, 1993.

11. Rudolf Maure. Utilizing restrictions techniques for Efficient graph pattern matching. In Proceedings of the International Workshop on Theory and Application of Graph Transformations (TAGT), Paderborn, Germany, November 1998.

12. A. Zuendorf. A heuristic for the subgraph isomorphism problem in executing PROGRES. Technical Report AIB 93-5, RWTH Aachen, Germany, 1993.

13. Солдатов В.Н., Чудинов И.Л., Ямпольский В.З. Банки данных в нефтяной промышленности. - Новосибирск: Наука, 1988.

УДК 681.121

## АНАЛІЗ ІНТЕРПОЛЬОВАНИХ ФУНКЦІЙ ПЕРЕТВОРЕННЯ РОБОЧИХ ЕТАЛОНІВ ОБ'ЄМУ ГАЗУ

**В.В.Вощинський**

*Спеціальне конструкторське бюро засобів автоматизації, 76000, м. Івано-Франківськ, вул. Акад. Сахарова, 32, тел. (03422) 2-34-13, e-mail: skbza@if.ukrtel.net*

*Приведены результаты исследования погрешностей интерполяции функций преобразования рабочего эталона объема газа.*

Економія природного газу як енергоносія можлива тільки при його достовірному обліку. Тобто, використання газу повинно бути забезпечене високоточними засобами вимірювання, до яких слід віднести лічильники газу турбінного і роторного типів. За останній час вітчизняна приладобудівна промисловість забезпечена такими засобами вимірювання з поліпшеними характеристиками (похибка лічильників газу – не більше 1,0%, діапазон вимірювання 1:30, 1:50, 1:100). А це, в свою чергу, ставить наукове завдання підвищення точності установок для калібрування і перевірки лічильників газу. Необхідно зазначити, що похибка установки з робочими еталонами об'єму газу [1], в основному, залежить від похибки робочих еталонів об'єму газу і є актуальною державною проблемою.

Використовуючи алгоритм метрологічної атестації робочих еталонів об'єму газу [2], можна отримати з високою точністю їх коефіцієнт перетворення  $K$  (імп/м<sup>3</sup>) на всьому діапазоні функції перетворення. Досвід роботи з метрологічної атестації робочих еталонів ЛГЕ-25; ЛГЕ-250 і ЛГЕ-2500 свідчить, що на функції їх

*The results of research of errors of interpolation of functions of transformation of the working standard of volume of gas are given.*

інтерполяційну криву. Але завдання полягає в тому, щоб інтерполяційна крива була максимально наближена до проміжних точок, які знаходяться на функції перетворення робочого еталона об'єму газу. Величина такого відхилення буде визначати похибку інтерполяції функції перетворення. Тому автором проведений аналіз відомих кривих [1, 4] для інтерполяції функції перетворення робочих еталонів об'єму газу, що і є науковою новизною даної роботи.

В нашому випадку коефіцієнт перетворення  $K$  залежно від витрати  $Q$  заданий таблично. Тоді згідно з [3] інтерполяційна функція може виглядати так:

$$K = \sum_{i=0}^N (A_i \cdot Q^i), \quad (1)$$

$$K = \sum_{i=0}^N (A_i \cdot Q^{i-1}), \quad (2)$$

$$K = \sum_{i=0}^N (A_i \cdot Q^{i-2}), \quad (3)$$

$$K = \sum_{i=0}^N (A_i \cdot Q^{i-3}), \quad (4)$$

