

УДК (004.01/.08 /.6):55(1)

Н. Захарій, інж.,
Київський національний університет імені Тараса Шевченка, Київ
В. Поліновський, канд. техн. наук, директор
Інститут комп'ютерних технологій Університету "Україна", Київ

АНАЛІЗ ТА СИНТЕЗ РІЗНОРІДНИХ ГЕОПРОСТОРОВИХ ДАНИХ (ПОБУДОВА МЕТАМОДЕЛІ ДЛЯ ГЕОЛОГІЧНОЇ ІНФОРМАЦІЇ)

(Рекомендовано членом редакційної колегії д-ром геол. наук, доц. О.Є. Кошляковим)

Проаналізовано специфіку геологічної інформації. Запропоновано метод пошуку прихованих закономірностей та спосіб інтеграції різномірних геоданих в єдине, однорідне середовище. Різні методичні підходи до обробки геологічної інформації призводять до багатоваріантних, неоднозначних, суб'єктивних висновкам і, на думку авторів, вимагають розроблення уніфікованих методик обробки геологічної інформації та їх стандартизації (створення геометаданих).

Постановка проблеми. Сучасні наукові погляди та актуальність вирішення геологічних проблем динамічно змінюються, і закономірності, знайдені вчора, не будуть

працювати завтра. Сьогодні виграє той, хто, застосовуючи новітні технології, своєчасно виявляє нові закономірності й підлаштовує під них свої погляди на вирішення

поставлених задач. При цьому варто згадати про специфіку геологічної інформації – не структурованої, неформалізованої, надлишкової або не рівномірної, для аналізу та обробки якої недостатньо використовувати тільки програми технічного аналізу. Необхідна комплексна система, в якій аналітичні складові виявляють нові закономірності, а програма технічного аналізу реалізує їх.

Аналіз останніх досліджень та невирішені раніше проблеми. Інтелектуальні підходи до аналізу і обробки даних про стан об'єктів різної складності є досить ефективними і широко застосовуються. У [2] розглянута модель і інтегрована експертна система, яка призначена для дослідження методів стегаграфії і стегааналізу.

Щорічно в світі проходить багато конференцій, симпозіумів, форумів, які значною мірою висвітлюють питання використання мережевих моделей на основі нечіткої логіки для моделювання, аналізу і управління взаємодіючими процесами в технічних системах. Так в роботі [4] пропонуються загальні основи штучного інтелекту та стратегії і методи вирішення складних проблем.

Створення інфраструктур геопросторових даних будь-якого рівня ґрунтується на загальних основних складових, принципах і методах реалізації. До таких належать: базові набори геопросторових даних, бази метаданих та механізми обміну даними, стандарти на геопросторові дані, метадані та геоінформаційні сервіси, технологічні засоби інформаційно-комунікаційного середовища створення, оброблення та використання геопросторових даних [3]. Однак відсутність механізму геоаналізу просторової інформації не дозволяють їх використати в повній мірі для вирішення задач, що розглядаються в даній роботі.

Як вже було зазначено, геологічні дані являють собою сукупність різномірної інформації і часто є великим масивом накопичених знань якісного та кількісного характеру. Очевидно, що для виявлення прихованих знань потрібно застосовувати спеціальні методи аналізу даних, за допомогою яких добуваються знання із величезного об'єму інформації. Саме тому, стають актуальними використання нових методів пошуку прихованих закономірностей (Data Mining підхід) та методів і

способів інтеграції різномірних геоданих в єдине, однорідне середовище.

Такий підхід дає можливість більш швидко та ефективно вирішенні задач побудови моделей геопросторових об'єктів будь-якої складності. Так, наприклад, на сьогоднішній день, при побудові таких моделей дослідники стикаються з проблемами не структурованої, неформалізованої, надлишкової або не рівномірної геопросторової інформації яку необхідно перетворити в певну геопросторову модель, а вже з цією моделлю розпочати певні наукові дослідження. При цьому дослідники, фактично, методом проб та помилок доводиться вибирати, яку саму первинну інформацію D_0 (рис. 1) та за допомогою яких методів синтезу D_0^I необхідно брати, щоб створити D_0^I – структуровану (підготовлену) інформацію, яку буде використовувати геопросторова модель, як вхідну інформацію. Одним із можливих варіантів вирішення поставленої задачі є створення хмарин метаданих при формуванні дослідницької задачі.

Цілі статті. Пропонується розглядати підхід оцінки змін стану геологічного середовища та просторового моделювання екологічних наслідків на основі "метамоделі", що відображає застосування методів обробки та перетворення вхідної інформації для отримання якісно нових даних та практичних відповідей щодо змін геологічних параметрів та стану навколишнього середовища.

Викладення основного матеріалу. Створений ГІС-макету Західного Донбасу [1] забезпечує роботу з наборами карт для комплексного аналізу (кількісна і якісна інтерпретація різномірної інформації, побудова багатфакторних моделей) та дозволяє переходити до повнофункціональної геоінформаційної обробки геоданих та створення бази даних для підтримки рішення задач користування надрами (екологічної спрямованості).

Варто зазначити, що основним недоліком в цьому ГІС-макеті є неможливість автоматизованої обробки всього масиву різномірної інформації самими моделями. Саме тому, як інструмент підтримки прийняття рішення щодо вибору методів оцінки змін стану геологічного середовища була розроблена методологія за допомогою "алгоритму дерева пошуку рішень/метамоделі".

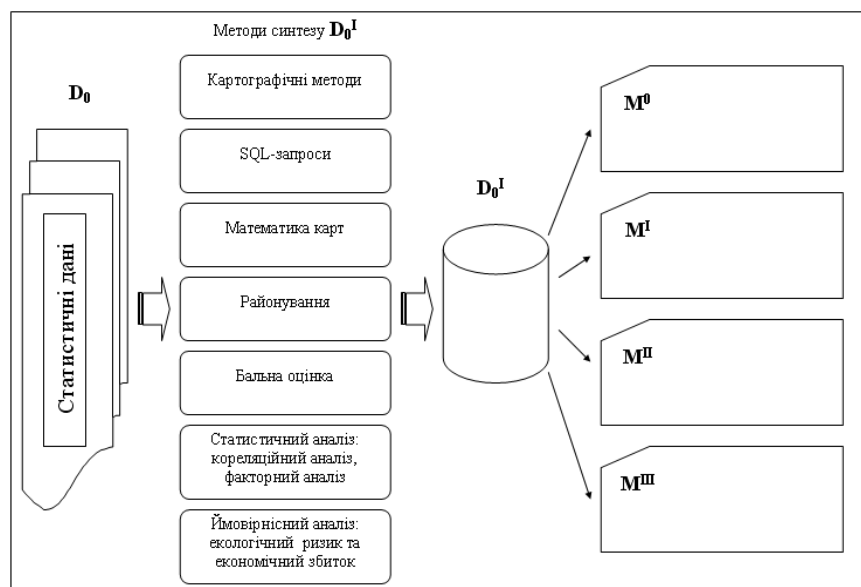


Рис. 1. Схема перетворення неструктурованої інформації D_0

Поняття *алгоритму*, яке введено у дослідженні, можна назвати поняттям алгоритму в інтуїтивному сенсі. Воно має поки що нечіткий, неформальний характер, посилається на деякі точно не визначені, але інтуїтивно

зрозумілі речі. Головна мета формалізації поняття алгоритму така: підійти до вирішення проблеми алгоритмічної можливості розв'язання різних геологічних задач, тобто відповісти на питання, чи може бути побудований

алгоритм, який приводить до вирішення завдання на основі згенерованих експертом-аналітиком сценаріїв.

Зрозуміло, що для нормальної роботи такого алгоритму необхідно на вхід його подавати однорідну D_0 (структуровану) інформацію. Для цих цілей запропоновано створювати геометадані до всього набору геологічної інформації.

Під терміном "геометадані", автори пропонують розглядати метадані, що не тільки вказують до якого саме типу відносяться геопросторові дані, а й "пояснюють" системі, всю сукупність методів, алгоритмів, тощо, які можна використовувати для цих геопросторових даних, з метою створення однорідних з точки зору системи баз геопросторових даних (*machine-readable*).

Поки що в ГІС-технологіях використовуються тільки так звані "Геопросторові метадані" (*Geospatial metadata*) для документування географічних цифрових ресурсів, що на думку авторів є звууженим і неповноцінним використанням всіх можливостей метаданих. На даному етапі геопросторові метадані включають в себе лише інформацію щодо виявлення, обмеження, ступеня, якості, просторових і тимчасових посилення, поширення, походження і технічного обслуговування цифрових географічних наборів даних [5].

Завдяки встановленню єдиного набору термінології *геометаданих*, на думку авторів, необхідно створити стандарт, який сприятиме не тільки належному використанню і ефективному пошуку географічних даних, а і дозволить проводити аналіз однорідної, з точки зору системи (*machine-readable*), інформації просторових геобаз даних та будувати сценарії розвитку змін та оцінки стану навколишнього природного середовища (альтернативних варіантів рішень геологічних задач на основі аналітико-експертної системи інтелектуального інформаційного аналізу геоданих).

Очевидно, що велика кількість параметрів геологічного простору, фільтрів/запитів для перетворень/перекласифікації даних, методів обробки інформації та аналізу набувають вигляду громіздких ієрархічних моделей і

ускладнюють аналітикам сприйняття задачі та побудову функціональних залежностей і визначення кореляції між параметрами. Саме тому, ці характеристики і параметри пропонується подавати у вигляді хмарин метаданих. Такий підхід дозволить автоматизувати дослідження за допомогою систем інтелектуального аналізу даних.

Кожен з цих параметрів або елементів належить до відповідного масиву (хмарин) метаданих. Отже, маємо чотири таких хмарин: хмара метаданих параметрів (ГІС) та опису змісту сцени: хмара метаданих фільтрів і перетворень, хмара метаданих алгоритмів та хмара метаданих геоаналізу (рис. 2).

Варто зазначити, що в більш загальному вигляді три останні хмаринки являють собою результуючу хмару синтезу задач. Зрозуміло, що аналітику незручно працювати одночасно з всіма трьома хмаринками метаданих "задач", тому в цій хмарі виділено абстрактну сутність – "вибір операцій/задач" (рис. 2) обраних метаданих, що в будь-який момент часу може бути повною реплікацією однієї з трьох хмарин (фільтри, алгоритми, геоаналіз) за вибором оператора. Тобто аналітик працює тільки з однією хмаринкою вибору метаданих.

Такий підхід є практичним, особливо для розподіленої роботи та паралельних обчислень. Більш детально механізм синтезу задач геоаналізу та обробки геологічної інформації розглянемо нижче, а зараз покажемо як формуються вибірки (рис. 2).

Аналітик звертається до хмари метаданих геологічної інформації, з якої він обирає необхідні параметри для проведення досліджень, далі запускається механізм відбору геологічних даних. Відповідний додаток звертається до банку даних різномірної геологічної (просторової) інформації та формує віртуальну вибірку геологічних даних за обраними параметрами. Такий механізм дозволяє сформувати декілька суттєво відмінних вибірок, які використовуватимуться для вирішення різних задач геоаналізу, обробки просторової інформації та статистичних досліджень, що є раціональним для розподіленої роботи та паралельних обчислень.



Рис. 2. Концепція [2] синтезу задачі геоаналізу

Побудуємо концептуальну модель процесу синтезу задач геоаналізу або обробки просторових даних та їх реалізації (рис. 3). Експерт-аналітик, звертається до хмари задач, а саме до хмари "0", з якої, за зазначеному вище способу, починає формувати послідовність задач (від 1 до n, рис. 3), які необхідно буде виконати для проведення геоаналізу або обробки просторових даних. Для кожної з цих задач, аналітик може задати певні параметри (рис. 3). Сукупність задач та їх пара-

метрів формують певний робочий простір, зазначеному на рис. 3 – як W_f , для обробки геологічної інформації, що містяться у віртуальній вибірці. Зрозуміло, що навіть не запускаючи ніяких обчислювальних процесів користувач може заздалегідь створити декілька W_f , і вже після цього запустити процес обробки просторових даних, що дозволить природно використовувати розподілені роботи та паралельні обчислення.

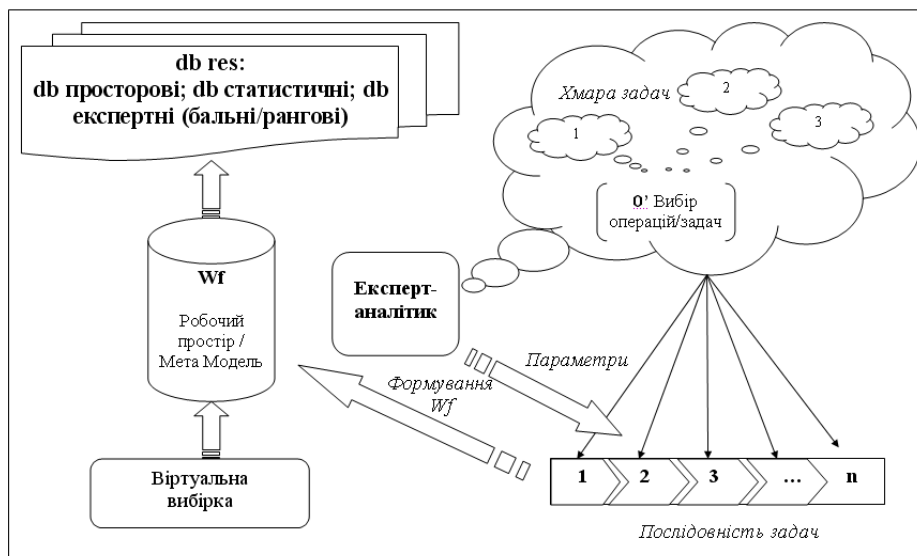


Рис. 3. Модель [2] синтезу задач для геоаналізу або обробки геологічної інформації

Результати руху геологічних даних через робочий простір, тобто послідовна обробка кожного шару даних задачами від 1 до n з певними параметрами, будуть записані в базу даних (**db_{res}**). При цьому, варто зазначити, що ця **db_{res}** база містить в собі три пов'язані між собою бази даних: **db_{просторові}** – база даних, що містить інформацію про кожному з робочих просторів, у яких виконувалися дослідження; **db_{статистичні}** – база даних, що містить статистичну інформацію (кількісну характеристику) та **db_{експертні}** – база даних, що містить інформацію про результати певних подій.

При формуванні робочого простору **Wf**, важливим є не тільки вид задач та з якими параметрами входять до нього, а й послідовність виконання цих задач, оскільки перетворення даних у геоаналізі не є комутативними операціями. Варто зазначити, що обробка кожного шару даних віртуальної вибірки певними алгоритмами геоаналізу або фільтрації, обробка інформації, та самі Data Mining алгоритми споживають багато обчислювальних ресурсів. Тому актуальною науково-прикладною задачею є оптимізація механізмів руху масивів даних через робочий простір при дослідженні просторової інформації методами геоаналізу або обробки просторових даних.

Згідно вищезазначеної концепції (див. рис. 2-3), отримані масиви статистик зберігаються в базі даних, структура якої дозволяє зберігати та отримувати статистику оброблену іншими модулями аналізу даних. Такий комплекс має досить широкі можливості вибірки та аналізу накопичених даних. За переліченими характеристиками комплекс геоаналітичних досліджень суттєво відрізняється від наявних рішень, більшість з яких є вузькоспеціалізованими і виконують аналіз геоданих лише одного типу та одним алгоритмом, а результати обробки представляються в своєму форматі, що часто унеможливує подальший аналіз.

Висновки. Отже, пропонується будувати *алгоритм дерева пошуку рішень (метамодель)*, як синтез методів збору, перетворення, обробки та аналізу інформації для розв'язку геологічних задач різної складності.

Data Mining підхід видобування даних надає аналітику засіб перевірки гіпотез при аналізі даних. І в підсумку для вирішення будь-яких завдань різної складності та сфер застосування результатів необхідний аналітик, який генерує гіпотези, а сучасні технології дозволяють їх обробляти. В той же час, якщо надати досліднику інструментарій по формуванню метаданих, потім ці дослідження можна буде повторювати в напівавтоматизованому режимі.

Автори вважають, що різні методичні підходи до обробки геологічної інформації призводять до багатоваріантних, неоднозначних, суб'єктивних висновкам і, на думку авторів, вимагають розроблення уніфікованих методик обробки геологічної інформації та їх стандартизації (закріплення в нормативно-правових документах).

Список використаних джерел

1. Захарій Н.В. Комплексний аналіз небезпечних змін геологічного середовища на прикладі вугледобувного регіону Західного Донбасу. / Захарій Н.В., Курило М.М. – Вісник КНУ імені Тараса Шевченка, Геологія. – К, 2012. – № 56. – С. 42-45.
2. Інформаційна технологія для дослідження методів стеганографії і стегоаналізу. / [Поліновський В.В., Корольов В.Ю., Герасименко В.А., Горинштейн М.Л.] – Міжвузівський збірник "Комп'ютерно-інтегровані технології: освіта, наука, виробництво" (м. Луцьк), 2011. – Випуск №5. – С. 236 – 242.
3. Карпінський Ю. Концептуальні створення національної інфраструктури геопросторових даних України / Карпінський Ю., Лященко А. // Сучасні досягнення геодезичної науки та виробництва, – 2005. – С. 295.
4. Люгер Д. Ф. Искусственный интеллект. Стратегия и методы решения сложных проблем = Artificial Intelligence. Structures and Strategies for Complex Problem Solving / Д.Ф. Люгер; пер. с англ. : – 4 изд. – М. : СПб. ; К. : Изд. Дом "Вильямс", 2005. – 864 с.
5. Geographic information/Geomatics <http://www.iso.org> ; Geospatial metadata <http://en.wikipedia.org>.

Надійшла до редколегії 04.04.13

Н. Захарій, інж.

Київський національний університет імені Тараса Шевченка, Київ
В. Полиновський, канд. техн. наук, директор
Інститута комп'ютерних технологій Університету "Україна", Київ

АНАЛИЗ И СИНТЕЗ РАЗНОРОДНЫХ ГЕОПРОСТРАНСТВЕННЫХ ДАННЫХ (ПОСТРОЕНИЕ МЕТАМОДЕЛИ ДЛЯ ГЕОЛОГИЧЕСКОЙ ИНФОРМАЦИИ)

Проанализирована специфика геологической информации. Предложен метод поиска скрытых закономерностей и способ интеграции разнородных геоданных в единую, однородную среду. Различные методические подходы к обработке геологической информации приводят к многовариантным, неоднозначным, субъективным выводам и, по мнению авторов, требуют разработки унифицированных методик обработки геологической информации и их стандартизации (создание геометаданных).

N. Zakcharyi, Engineer
Taras Shevchenko National University of Kyiv, Kyiv
V. Polinovskyi, Cand. Sci. (Tech.), director
Institute of Computer Technology of University "Ukraine", Kyiv

ANALYSIS AND SYNTHESIS OF DISSIMILAR GEOSPATIAL DATA (CONSTRUCTION OF METAMODELS FOR GEOLOGICAL INFORMATION)

Analyzed the specific geological information. Suggested the method of finding hidden regularities and way of integrating diverse geodata into a single, homogeneous environment. Various methodological approaches to processing of geological information leading to multivariate, controversial, subjective opinion and, according to the authors, requires the development of standardized methods processing of geological information and standardization (create geometadata).