

ПОБУДОВА СЕМАНТИЧНИХ ШКАЛ ДЛЯ ІНТЕРПРЕТАЦІЇ ДАНИХ ЕКОЛОГІЧНОГО МОНІТОРИНГУ

Abstract. The article describes the semantic model of the domain (semantic space) for the tasks of analysis and meaningful interpretation of monitoring data of air. Proposed visualization of environmental indices in a two-dimensional. scales.

Вступ. Семантична модель предметної області (ПО) визначається як інформаційна модель, що відображує основні поняття ПО та певні відношення між ними. Взагалі, для формалізації ПО можна використовувати онтологічні, когнітивні та семантичні моделі ПО.

Найбільш поширеним є онтологічний підхід до моделювання ПО. Він включає об'єктний підхід, відомий в програмуванні, фреймово-мережевий формалізм представлення знань та механізм запитів з баз даних для виявлення знань про ПО. Для систематизації знань у галузі екологічного моніторингу побудовано семантичну модель ПО та розроблено семантичні шкали, які полегшують змістовну інтерпретацію результатів аналізу.

Постановка задачі. Для різних способів упорядкування кількісної інформації застосовуються різні типи шкал. Під *семантичним шкалюванням* ми розуміємо методи та засоби переходу від кількісних характеристик (зокрема, сукупності кількісних даних, одержаних в результаті екологічного моніторингу) до певних засобів змістовної інтерпретації цих даних.

При побудові баз знань інтелектуальних систем поняття семантичного простору експерта визначається як певна сукупність показників (просторових координат), за допомогою яких можна визначити та диференціювати об'єкти або поняття досліджуваної області [1]. Отже, на формальному рівні в екологічних дослідженнях семантичний простір будемо розглядати як систему координат (на площині або в багатовимірному просторі), де за значеннями окремих параметрів можна визначити екологічний стан певного об'єкту або траєкторію його руху (розвитку) в певному напрямку.

Сучасний математичний апарат багатовимірного аналізу даних (факторний аналіз, багатовимірне шкалювання, розпізнавання образів тощо) пропонує конструктивні засоби та технології для розробки багатовимірних моделей семантичної організації знань [2]. Розглянемо основні етапи та особливості розробки таких моделей для задач екологічної безпеки.

Вибір інформативних параметрів. Для екологічних досліджень це одна з найбільш актуальних задач, оскільки небезпека окремих ситуацій може посилюватися синергетичним ефектом від одночасної взаємодії кількох

факторів ризику. У певних випадках зону небезпеки (або нестабільності) можна виявити при дослідженні співвідношень між окремими параметрами. Вдалих вибір інформативних параметрів суттєво полегшує дослідження та візуальну інтерпретацію змін екологічного стану, а також визначення рівня стійкості певних територій відносно негативних чинників.

Критерії інформативності на даному етапі будемо обирати з практичної точки зору, тобто перевага надається тим показникам, які мають найбільший вплив на стан здоров'я населення прилеглих територій.

При розробці інтелектуальних технологій для організації бази знань запропоновано два альтернативних підходи до вибору інформативних параметрів, які можуть доповнювати один одного.

1. Експертний підхід. Вибір інформативних параметрів здійснюється експертами на основі результатів попереднього аналізу даних моніторингу територіальних систем. Зокрема, для аналізу даних про забруднення атмосферного повітря в містах України на експертному рівні було обрано індекс забруднення атмосфери ІЗА, розрахований в [3], та ряд окремих шкідливих речовин (діоксид азоту, чадний газ, формальдегід тощо).

2. Багатовимірні індекси. Процедуру вибору інформативних параметрів можна формалізувати за допомогою методів виявлення латентних факторів, які запропоновані для формалізації експертних знань в інтелектуальних системах [1]. Структурні методи аналізу багатовимірної інформації також забезпечують можливості для переходу до багатовимірних індексів, які можуть використовуватись як індикатори досліджуваних екологічних станів або тенденцій до їх змін.

Структурний аналіз даних. Для вирішення задач інтерпретації екологічної інформації та виявлення найбільш небезпечних екологічних ситуацій розроблено інтелектуальні технології структурування даних, що базуються на методах інтелектуального аналізу даних, або ДМ [2].

Структура багатовимірних даних може розглядатися як геометрична конфігурація точок, що представляють різні стани досліджуваних об'єктів у просторі вихідних параметрів. Необхідно одержати чітке уявлення про цю структуру, аналізуючи наявні кількісні показники та певні співвідношення між ними. В процесі аналізу експериментальні дані не просто замінюються іншими, більш загальними характеристиками, але отримують більш глибоку змістовну інтерпретацію, вказуючи на певні закономірності в вихідних даних. Загалом, методи ДМ спрямовані на виявлення й структурування тих знань, які можна одержати з експериментальних даних.

Велика кількість показників, в якій може розібратися тільки досвідчений фахівець з екології, в результаті статистичної обробки та використання ДМ отримує досить глибоку семантичну інтерпретацію, зрозумілу на рівні користувача системи. В практичних задачах еколого-енергетичного моніторингу та управління екологічною безпекою знання експертів та емпіричні дані доводиться інтегрувати в єдині багатовимірні структури,

представлені в доступному вигляді за допомогою сучасних засобів комп'ютерної графіки.

В DM факторний аналіз та деякі інші методи багатовимірної статистики розглядаються як структурні методи стиснення інформації, де в якості нових змінних (імен, типів) використовуються набори чисел, що відображають результати аналізу на семантичні шкали, тобто будуються багатовимірні індекси певних станів. Так, у випадку відображення на одну шкалу L екологічні стани будуть визначатись через певні інтервали, в яких відбуваються зміни чисельних значень уздовж цієї шкали. Двовимірні шкали забезпечують можливість перейти до графічної інтерпретації вихідних даних у вигляді точок на площині (L_1, L_2), а вибір трьох шкал спрямований на візуальну інтерпретацію даних у просторі трьох вибраних ознак.

З формальної точки зору структурний аналіз даних можна здійснити за допомогою відомих засобів матричної алгебри.

Множину різноманітних ситуацій $\{ S_1, S_2, \dots, S_n \}$ можна описати як багатовимірну структуру $x = \{x_1, x_2, \dots, x_n\}$, де $x_i = (x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)})$ - вектор значень аналізованих ознак (властивостей) x_1, x_2, \dots, x_m , що характеризують ситуацію S_i . Багатовимірну структуру можна розглядати також як матрицю попарних порівнянь:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

де a_{ij} виступає як міра схожості або відмінності ситуацій S_i і S_j , що в більшості випадків визначається як геометричне відстань між точками в багатовимірному просторі ознак.

Зокрема, для визначення відстані між парами ознак можна скористатись відомою формулою для евклідової відстані, де міра відмінності двох об'єктів i та j із значеннями k -ої ознаки x_{ik} і x_{jk} для об'єктів i та j відповідно матиме вигляд:

$$d_{ij} = \left[\sum_{k=1}^K (x_{ik} - x_{jk})^2 \right]^{1/2}$$

Отримана в результаті аналізу структурована множина ситуацій одержує інтерпретацію на змістовному рівні. Якщо використовувати сучасну термінологію щодо нормування викидів в галузі екологічної безпеки, можна виділити ситуації (або території) без перевищення гранично допустимих норм ГДК, ситуації з перевищенням цих норм по одному з параметрів, ситуації з перевищеннями норм для двох параметрів тощо.

Результати аналізу. Для інтерпретації екологічної інформації можна застосувати різні засоби візуалізації даних. Найбільш відомі з них – це візуалізація даних моніторингу у вигляді екологічних карт досліджуваних

територій і візуалізація графічних образів (patterns), які відображують розподіл досліджуваних структур у просторі інформативних ознак.

Для даних, що характеризують екологічний стан окремих територіальних систем, найбільш зручним вважається представлення інформації у вигляді карт техногенних навантажень на досліджувані території або карт ризиків. Для порівняння декількох ситуацій та прогнозування використовують серії тематичних карт.

Для аналізу та візуалізації інформації про значну кількість різних ситуацій більш ефективними будуть технології ДМ, орієнтовані на побудову шаблонів або зразків. Шаблони можна охарактеризувати як закономірності, властиві певним вибіркам даних та представлені у досить простій формі. Вибір шаблонів базується на виявлених раніше знаннях про екологічний стан досліджуваних територій, тобто створюється формальний опис набутого досвіту, який застосовується для прийняття рішень.

Для задач екологічного моніторингу шаблони можна представити як візуальні образи певних екологічних станів, відображені в координатний простір інформативних параметрів. В сучасних системах аналізу та візуалізації даних [3] реалізовано можливості візуального аналізу ситуацій як на площині (двовимірний візуальний аналіз), так і у просторі трьох вимірів (тривимірний візуальний аналіз).

На рис. 1 і рис. 2 наведено приклади двовимірних шкал, де стан атмосферного повітря відображено в системі координат, побудованій за даними ранжирування щодо перевищення ГДК для основних показників атмосферного забруднення – діоксиду азоту та формальдегіду, яке спостерігалось протягом 2010 р. Моніторинг проводився на 16 точкових об'єктах (постах спостереження), де відбувалося вимірювання концентрацій відповідних шкідливих речовин.

Найкращі еталони екологічного стану, побудовані за результатами вимірювання для найменш забруднених територій, відображено в початковій точці побудованої системи координат.

На рис. 3 зображено двовимірну шкалу, побудовану аналогічним чином для індексу атмосферного забруднення (ІЗА), обчисленого на основі значень окремих показників (за даними 2011 р.) за наступною формулою:

$$I = \sum_{i=1}^n \left(\frac{q_i}{\vec{A}\vec{A}\vec{E}_i} \right)^{C_i}$$

де: q_i – концентрація i -ої забруднюючої речовини (ЗР); $ГДК_i$ – середньодобова граничнодопустима концентрація i -ої ЗР; n – кількість ЗР, що досліджуються; C_i – безрозмірний коефіцієнт, що приводить рівень забруднення p -ою ЗР до рівня забруднення речовини третього класу небезпеки. Для речовин першого класу небезпеки $C_i=1,7$; для другого – $1,3$; для третього – $1,0$; для четвертого – $0,9$.

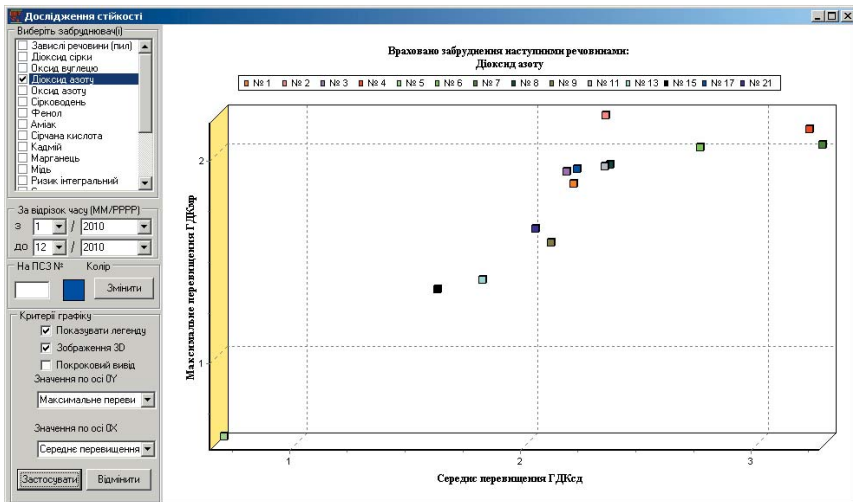


Рис. 1. Двовимірна шкала, що відображує перевищення норм забруднення для діоксиду азоту в м. Києві протягом 2010 р.

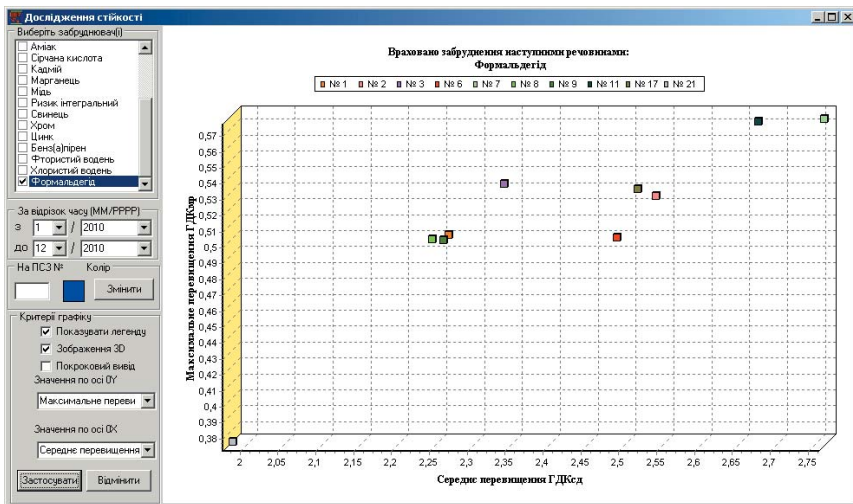


Рис. 2. Двовимірна шкала, що відображує перевищення норм забруднення для формальдегіду в м. Києві протягом 2010 р.

Ще один засіб інтерпретації даних моніторингу територіальних систем можна отримати з використанням сучасних уявлень про техногенний ризик [5 - 9]. Тоді кожен з досліджуваних територій можна віднести до одного із визначених класів ризику (території з помірним рівнем ризику, території з підвищеним рівнем ризику, території з високим ризиком тощо).

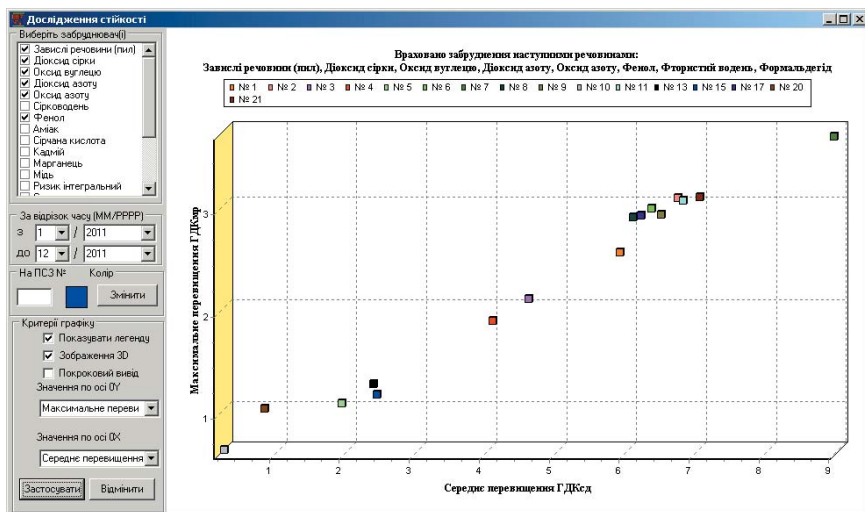


Рис. 3. Двовимірна шкала, що відображує перевищення норм забруднення для ІЗА в м. Києві протягом 2011 р.

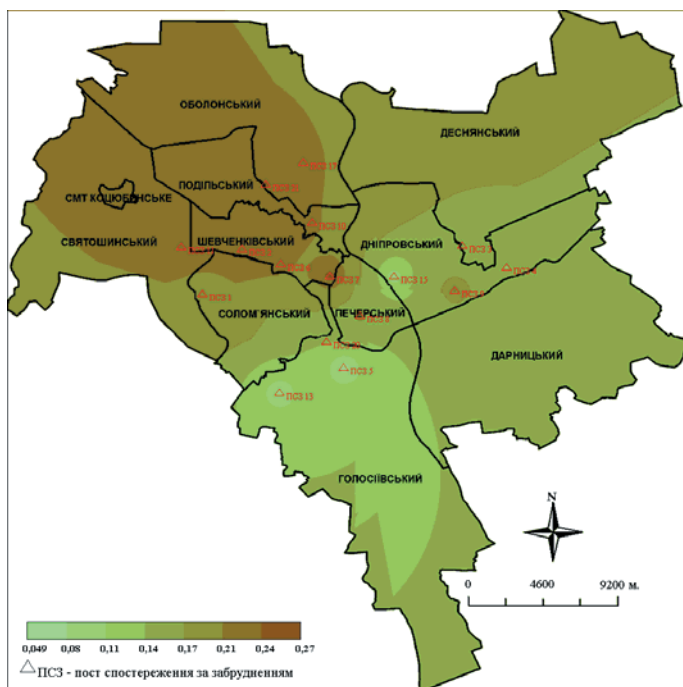


Рис. 4. Карта ризику хронічної інтоксикації внаслідок забруднення повітря Києва протягом 2011 року.

На рис. 4–5 наведено карти ризику, побудовані за даними моніторингу на постах спостереження м. Києва, одержаними протягом 2011 р. Для обчислення техногенних ризиків та побудови екологічних карт адаптовано та удосконалено методику, запропоновану в роботах [7, 10].

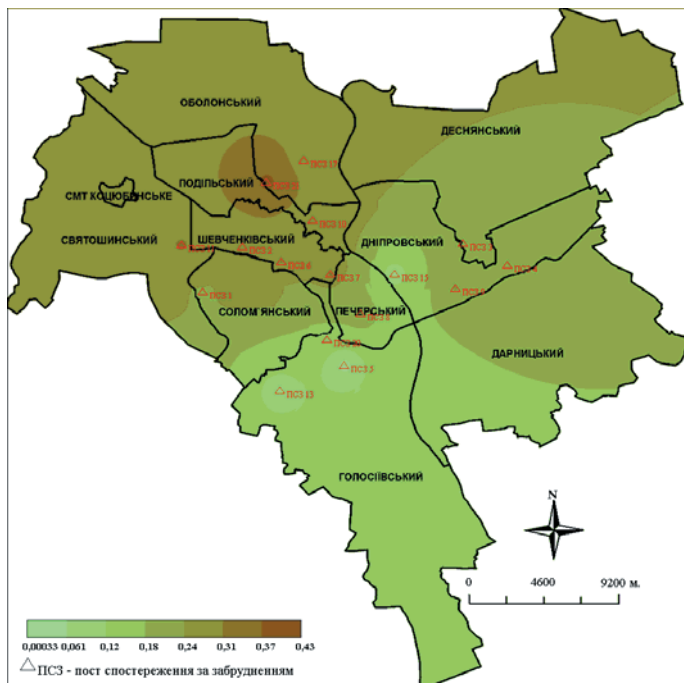


Рис. 5. Карта ризику миттєвих токсичних ефектів внаслідок забруднення повітря Києва за січень-грудень 2011 року

Значення ризиків для населення міста виявились досить великими. Особливо треба підкреслити незадовільний екологічний стан на північному заході, де ризик хронічних захворювань перевищує 0,2, тобто за таких умов більше 20 відсотків населення виділених районів можуть одержати (або вже мають) хронічні хвороби. Ризик миттєвих токсичних ефектів також виявився найбільшим для населення північно-західної частини міста.

Висновки. Розроблено семантичну модель предметної області для задач аналізу та змістовної інтерпретації даних еколого-енергетичного моніторингу атмосферного повітря. Для полегшення змістовної інтерпретації даних

моніторингу і результатів аналізу запропоновано засоби візуалізації екологічних індексів у вигляді двовимірних семантичних шкал.

Семантичні шкали, що відображують стан атмосферного повітря, можуть бути побудовані як на основі даних екологічного моніторингу, так і на основі екологічних індексів. Шкали, побудовані на основі екологічних індексів (зокрема, ІЗА) передають більш цілісне й інтегроване уявлення про екологічний стан досліджуваної території, оскільки одночасно враховується цілий ряд небезпечних показників.

Побудовано приклади семантичних шкал для інтерпретації результатів аналізу щодо перевищення норм забруднення атмосферного повітря та рівнів атмосферних ризиків в місті Києві протягом 2010 – 2011 рр. Виявлено досить великі значення ризиків для північно-західної частини міста, які можуть суттєво вплинути на показники захворюваності населення.

1. *Гаврилова Т.А., Хорошевский В.* Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
2. *Дюк В., Самойленко А.* Data Mining. Учебный курс. – СПб.: Питер, 2001. – 366 с.
3. *Боровиков В.* STATISTICA. Искусство анализа данных на компьютере. Для профессионалов. – СПб.: Питер, 2003. – 688 с.
4. Щомісячний бюлетень забруднення атмосферного повітря в Києві та містах Київської області. – К.: Центральна геофізична обсерваторія, 2005- 2011 рр.
5. *Алымов В.Т., Тарасова Н.П.* Техногенный риск: Анализ и оценка: Учебное пособие для вузов. – М.:ИКЦ «Академкнига», 2004. – 118 с.
6. *Швыряев А.А.* Оценка риска воздействия загрязнения атмосферы в исследуемом регионе : учебное пособие для вузов / А.А. Швыряев, В.В. Меньшиков. – М. : МГУ, 2004. – 124 с.
7. *Каменева І.П.* Методи визначення екологічного ризику за атмосферним фактором / І.П. Каменева, О.О.Попов, А.В. Яцишин, В.О. Артемчук. – Моделювання та інформаційні технології : зб. наук. пр. – К.: ІПМЕ ім. Г.С. Пухова НАН України, 2009. – № 53. – С. 23-32.
8. *Лисиченко Г.В.* Методологія оцінювання екологічних ризиків / Г.В. Лисиченко, Г.А. Хміль, С.В. Барбашев. — О. : Астропринт, 2011. — 368 с.
9. *Яцишин А.В.* Методы и технологии анализа рисков для здоровья на основе данных мониторинга /А.В. Яцишин, И.П. Каменева, В.А. Артемчук, А.А. Попов // IV Международная научная конференция «МОДЕЛИРОВАНИЕ-2012». Сборник трудов конференции., К., 2012. – С. 470-473.
10. Попов О.О. Математичне та комп'ютерне моделювання техногенних навантажень на атмосферу міста від стаціонарних точкових джерел забруднення. Автореф. канд. дис. К., 2010. – 20 с.

Поступила 21.02.2013р.