

## ВПОРЯДКУВАННЯ ІНФОРМАЦІЇ — ПОТОЧНА ПРОБЛЕМА ЕКОЛОГІЧНОЇ НАУКИ

*Описуються деякі особливості нової компонентної концепції екологічної інформації та можливості її застосування для класифікації екологічних повідомлень з використанням системи "оцінених" ключових слів.*

Нині вже зрозуміло, що реалізацію моделей сталого розвитку можна здійснити лише на основі ґрунтовних екологічних досліджень і доведенні їх результатів до широких верств населення. Лише тоді досягнення науки зможуть стати засобом підвищення екологічної культури та свідомості, ідентифікації свого місця в глобальній соціоекосистемі, знаходженні своєї екологічної ніші.

Але цей процес ускладнений тим, що сучасна екологія досі не перетворилась на чітку теоретико-методологічну систему. Це наукова дисципліна на етапі становлення, яке прискорено наявністю загрози незворотної деградації біосфери. В ній паралельно співіснують декілька методологій з різними підходами до визначення базових понять змісту екології, її зв'язків з іншими дисциплінами.

Різноманітність екологічних досліджень, різноманіття підходів до вивчення однакових проблем дезорієнтує дослідників, приводить до спотворення уявлення про об'єктивно існуючу реальність. Спроби поєднати ці підходи поки що не дали позитивного результату і призводили тільки до появи штучних синтетичних наукових напрямків.

Таке становище є незадовільним, воно гальмує розвиток екології як науки, вкрай утруднює інтеграцію зусиль науковців. Як казав видатний французький вчений Анрі Пуанкаре: "Наука — це насамперед класифікація". Не випадково питання класифікації екологічного знання сьогодні дедалі частіше аналізується в працях науковців.

Пошук спільного у різних підходах екологічної теорії і практики привів до висновку про необхідність формулювання нової концепції екологічної інформації та складу її джерел, головні риси якої описуються нижче.

Елементарним джерелом екологічної інформації є характер взаємодії біологічного об'єкта з екологічним фактором. Під біологічним об'єктом ми розуміємо всі живі структури організменного і надорганізменного рівня (деми, популяції, асоціації та інші — до біосферного рівня включно). Екологічні фактори, на нашу думку, — це будь-які об'єкти, процеси або явища природного чи антропогенного походження, взаємодія яких з біологічним об'єктом безпосередньо або опосередковано змінює параметри останнього.

Сукупність усіх екологічних факторів ми розподіляємо на 4 великі групи: абіотичні, біотичні, техногенні і соціогенні. Такий розподіл факторів схожий на традиційні класифікації виділенням абіотичних та біотичних факторів і відрізняється від них розмежуванням техногенних і соціальних факторів, в той час як у більшості класифікацій виділяють замість них одну групу антропогенних факторів.

Додатковий розподіл антропогенних факторів, на наш погляд, обумовлений об'єктивно існуючою різницею між характером техногенних і соціогенних факторів. Перші відрізняються притаманною їм штучною абіотичною природою і набагато сильнішим, порівняно з природними абіотичними факторами, ефектом впливу на елементи екосистем. Техногенні фактори — це головне джерело проблем, пов'язаних з погіршенням стану навколишнього середовища.

Соціогенні фактори відрізняються від трьох попередніх груп тим, що мають інформаційну природу, тобто, через них передаються порції інформації, повідомлення, вплив яких неадекватний потокам речовини й енергії, що переносяться, набагато перевищує його. Їх джерелом є соціальна структура і процеси, які протікають у людському суспільстві. Окрім того, виділення

соціогенних факторів в окрему групу постулює подвійну роль людини в природі: з одного боку — як біологічного об'єкта, а з іншого — як соціальної істоти.

Оскільки в сучасних умовах біологічний об'єкт зазнає впливу всіх груп факторів, то можна казати про окремі компоненти повідомлень екологічної інформації: абіотичну, біотичну, соціальну і техногенну, які містять відомості про характер взаємодії біологічного об'єкта з факторами чотирьох груп. З цих компонент, або складових, біотична відрізняється тим, що містить дані і про самий біологічний об'єкт і про біотичні фактори.

Екологічна інформація починає свій шлях до реципієнта з природної системи, і, зазнавши змін у процесі обробки й аналізу, перетворюється на інформаційні повідомлення, які частіш за все з'являються у вигляді наукових та інших публікацій. Якщо з такого повідомлення вилучити абіотичну, соціогенну й техногенну компоненту, або навіть всі їх разом, то воно збереже свою природу і певну адекватність реальній екологічній ситуації. Якщо ж з повідомлення вилучити біотичну компоненту, то інформаційне повідомлення втратить статус екологічного і перетвориться на технічне, геологічне, економічне або ще якесь інше.

Як бачимо, біотична компонента виконує особливу, базисну функцію в інформаційному комплексі екологічного повідомлення і є його основою. Вона і тільки вона є постійною незмінною частиною повідомлень екологічної інформації.

Отже, наведене дозволяє визначити різні підходи до класифікації різноманіття джерел екологічної інформації.

Перший підхід полягає в експертній оцінці екосистем різного рівня, їх поступовій декомпозиції, аналізі підсистем та їхніх елементів. Ймовірно, що такий підхід призведе до створення величезного реєстру екологічних об'єктів. Завдяки різноманіттю форм існування та адаптацій біологічних об'єктів обсяг такої роботи важко навіть уявити.

Спрощення процесу наукової інвентаризації джерел екологічної інформації можливе з використанням їх групування й узагальнення властивостей [2], або з обмеженням типів джерел екологічної інформації. В обох випадках це загрожувє втратою об'єктивності кінцевої класифікації.

Другий підхід базується на ідеї побудови класифікації з використанням всієї суми екологічних досліджень, виконаних науковцями, і результати яких були оприлюднені у вигляді наукових публікацій. В основі цієї ідеї лежить уявлення про наукову діяльність як про засіб об'єктивного відображення природи, світу.

Для розробки такої класифікації необхідно зробити репрезентативну вибірку наукових інформаційних джерел, причому єдиним критерієм відбору цих джерел є вимога обов'язкової наявності в них інформації про характер взаємодії біологічного об'єкта з будь-якою комбінацією факторів вищезначених груп. Взагалі, цей підхід вперше запроваджений нами в рамках екологічного дослідження.

Традиційно друкована інформація систематизується і класифікується в процесі реалізації різних бібліотечно-бібліографічних процесів. В цій сфері вже розроблено міцні підвалини теорії та методології створення як спеціальних, так і універсальних класифікацій. Останнім часом у зв'язку із швидкими темпами розвитку наук помітні тенденції до надання певної гнучкості цим системам, здатності розвиваючись вбирати нові розділи без порушення головного класифікаційного ряду. Рушійною силою цих процесів стала необхідність збереження адекватності існуючої класифікаційної системи реальній системі наукового знання, сумі теоретичних і практичних досягнень, зроблених з моменту останнього перегляду класифікації.

Причиною дивовижної нездатності універсальних класифікацій відображати поточний стан екологічного наукового знання полягає в застосуванні до комплексу екологічних дисциплін (які, як було зазначено вище, продовжують бурхливо розвиватись) так званого експертного методу, коли класифікаційний ряд визначається змішаними групами фахівців.

На наш погляд, при класифікації такого широкого кола напрямків досліджень, які входять до складного комплексу екологічних дисциплін, необхідно застосовувати об'єктивніші методи класифікації, які в своїй основі мали б досягнення теорії математичної статистики.

Методологія такого підходу така.

В межах будь-якої наукової дисципліни існує комплекс опорних базових лексичних одиниць, які зустрічаються в наукових текстах частіше і які мають більше смислове навантаження, ніж інші. Такі слова і деякі словосполучення назвемо ключовими. Їхня підвищена інформативність робить можливим їх використання для фіксації змісту інформаційних повідомлень, що корисно з точки зору зменшення обсягу даних, що зберігаються.

Комплекс ключових слів, які в теорії інформаційно-пошукових систем, завдяки їхній здатності описувати більші інформаційні елементи, звуть дескрипторами, утворює тезаурус наукової дисципліни — її лексико-семантичну основу, завдяки якій можливе існування наукової інформації і її передача.

Якщо тезаурус складається і використовується на рівні інформаційно-пошукових систем, то, як правило, в нього включають синоніми, омоніми, іншомовні еквіваленти й інші зв'язки. Створення екологічного українського тезаурусу — це значна аналітична і науково-дослідна робота, виконання якої може зайняти від 5 до 10 років [4].

Виділення набору ключових слів і словосполучень для комплексу екологічних дисциплін дозволить уніфікувати підходи до фіксації змісту екологічних повідомлень, легко формулювати предметні рубрики.

Систематизація екологічних лексичних одиниць вже розпочата авторами спеціальних словників і довідкових видань [6, 8, 10, 11 та ін.]. Набутий ними досвід є великою цінністю і є основою розробки тезаурусу.

Іншим джерелом для визначення дескрипторів є аналіз змісту наукових публікацій. Департаментом екології МФ НаУКМА було виконано аналіз відповідних джерел і визначено базову сукупність україномовних ключових слів для комплексу екологічних дисциплін. Вони використовуються як лексико-семантична база автоматизованої системи, зокрема, для опису змісту повідомлень екологічної інформації, дані про які зберігаються в ній.

З точки зору компонентної концепції екологічної інформації в семантиці кожного ключового слова у складі екологічного тезаурусу міститься певна інформація про відносини в діаді "біологічний об'єкт — екологічний фактор", але чітко передати значення таких смислових компонентів в більшості випадків неможливо без використання додаткових засобів.

Всю суму інформації про взаємодію елементів екологічної діади нами було типізовано і виявлено певне коло питань, відповіді на які відносно кожного з дескрипторів дозволяють чіткіше виявити його семантику. Ці питання стосуються всіх аспектів походження біологічних об'єктів, їх розвитку, існування і взаємодії з факторами довколишнього середовища. Типовими прикладами таких питань є: "Чи має відношення дескриптор до взаємодії біологічного об'єкта з техногенними факторами?" або "Чи має значення дескриптора відношення до вікової структури популяції?".

Всі питання сформульовані таким чином, щоб відповідь на них вимірювалась у бінарній шкалі булевих значень "так" або "ні", "1" або "0".

Завдяки цьому ми маємо можливість представити набір відповідей на означені питання у вигляді таблиці типу "об'єкт — ознака" розмірності  $N \times P$ , де  $N$  — кількість проаналізованих дескрипторів, що виступають у якості об'єктів,

а  $P$  — кількість питань, тобто ознак. Для кожного з об'єктів у такий таблиці набір значень ознак складає індивідуальний профіль дескриптора, який є семантичною мірою відношення значення дескриптора до біотичної складової екологічної інформації.

Якщо ми маємо набір повідомлень екологічної інформації, зміст яких описано за допомогою дескрипторів, то значення бінарних ознак цих дескрипторів успадковуються повідомленнями. Таким чином, ми отримуємо профілі відношення до біотичної складової вже для кожного з повідомлень, що відображаються в таблиці "об'єкт — ознака", аналогічній попередній, але в ній у якості об'єктів виступають не дескриптори, а повідомлення. Система ознак у цій таблиці тотожна системі ознак попередньої таблиці.

Така таблиця може бути оброблена з використанням алгоритмів кластерного аналізу на ЕОМ для отримання гіпотез про розбивку сукупності повідомлень на семантичні групи з використанням показників відношення повідомлень до біотичної складової в якості класифікаційних ознак. Оскільки вибірка повідомлень репрезентативна для їх генеральної сукупності, то на її базі можна розробити класифікацію всієї системи екологічного наукового знання.

З наведеної схеми процесу визначення значень ознак для повідомлень ми бачимо, що він відбувається через додатковий етап визначення змісту дескрипторів, а не безпосередньо. Це пояснюється тим, що вибірка повідомлень має бути дуже великою для адекватної репрезентації всього розмаїття повідомлень екологічної інформації. Використання системи "оцінених" дескрипторів значно зменшує витрати часу на опис змісту повідомлення і спрощує автоматизацію процесу класифікації.

Така класифікаційна система являє собою ієрархічну деревоподібну структуру, розгалужену відповідно до структури об'єктів екологічних досліджень. Вірогідно, що під час виконання розбивки багатовимірного ( $p$ ) простору ознак буде отримано не одну, а серію гіпотез про розподіл об'єктів і зв'язки між класами. Для визначення найбільш адекватної схеми розбивки можливе застосування коефіцієнтів кофенетичної кореляції, багатовимірного дисперсійного аналізу, методів Монте-Карло з використанням рандомізованих контрольних груп об'єктів.

Для визначення структури розділів, складення їх реєстру і присвоєння окремим групам класифікаційних ієрархічних індексів і назв, що відображають та узагальнюють зміст групованих об'єктів, слід додатково проаналізувати значення ознак об'єктів для окремих кластерів.

У такому вигляді класифікацію вже можна буде використати в місцях збереження екологічної інформації: бібліотеках, центрах НТІ замість відповідних розділів універсальних класифікацій, не порушуючи структуру головного класифікаційного ряду. З метою збереження функціонуючої системи каталогів і картотек можливе введення нової класифікації як системи індексації існуючих розділів бібліотечних класифікаційних систем, своєрідного класифікаційного покажчика.

Впорядкування системи збереження і використання екологічної інформації надасть велику користь науковцям, викладачам, студентам, які більше не будуть вимушені власноруч з'єднувати розрізнені гетерогенні екологічні дані.

Використання єдиної класифікаційної системи джерел екологічної інформації підвищить ефективність екологічної науки. Допоможе відшукати шляхи інтеграції зусиль і спільні засади досліджень.

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности.— М.: Финансы и статистика, 1989.— 607 с.

2. Беляев В. И. Моделирование морских систем.— К.: Наукова думка, 1987.— 201 с.

3. Бібліотечно-бібліографічні класифікації та інформаційно-пошукові системи.— К., 1995.— 108 с.

4. Вьюгин В. И. и др. Информационно-поисковый язык в биомедицине.

5. Колесников Л. А. Основы теории системного подхода.— К.: Наукова думка, 1988.— 176 с.

6. Кондратюк Є. М., Хархота Г. І. Словник-довідник з екології.— К.: Урожай, 1987.— 160 с.

7. Миркин Б. Г. Анализ качественных признаков и структур.— М.: Статистика, 1980.— 319 с.

8. Словарь-справочник по экологии / К. М. Сытник, А. В. Брайон, А. В. Городецкий, А. Г. Брайон.— К.: Наукова думка, 1994.

9. Факторный дискриминантный и кластерный анализ.— М.: Финансы и статистка. 1989.— 215 с.

10. Экологический энциклопедический словарь / Сост. Дедю И. И.— Кишинев, МСЭ.— 408 с.

11. Экология и охрана окружающей среды: Толковый терминологический словарь.— М.: Всемирный следопыт. 1998.— 479 с.

*Eliseyev V. V.*

## INFORMATION REGULARIZATION — THE CURRENT PROBLEM OF ECOLOGICAL SCIENCE

Some peculiarities of the new component concept of ecological information are described as well as possibilities of its use for ecological messages classification with the system of 'evaluated' keywords.