

Валентина В. Іванова

МЕТОДОЛОГІЧНІ ПІДХОДИ ДО МОДЕЛЮВАННЯ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ ЕКОНОМІКИ

У статті обґрунтовано необхідність використання моделювання для організації інформаційного забезпечення економічних суб'єктів, а саме: теорії множин і математичної логіки. Розроблено модель інформаційної технології інформаційного забезпечення економічних суб'єктів. Запропоновано здійснювати структурування інформаційного потоку перед його надходженням на рівень обробки за допомогою кластеризації.

Ключові слова: інформаційне забезпечення; моделювання; інформаційна технологія; інформаційний потік; кластеризація.

Форм. 10. Табл. 2. Літ. 10.

Валентина В. Иванова

МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К МОДЕЛИРОВАНИЮ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ ЭКОНОМИКИ

В статье обоснована необходимость использования моделирования для организации информационного обеспечения экономических субъектов, а именно: теории множеств и математической логики. Разработана модель информационной технологии информационного обеспечения экономических субъектов. Предложено осуществлять структуризацию информационного потока перед его поступлением на уровень обработки с помощью кластеризации.

Ключевые слова: информационное обеспечение; моделирование; информационная технология; информационный поток; кластеризация.

Valentina V. Ivanova¹

METHODOLOGICAL APPROACHES TO MODELLING OF INFORMATION SUPPORT FOR ECONOMICS

The paper grounds the need to apply modelling, in particular, the fuzzy set theory and the mathematical logic for providing information support for economic entities. A model of information technology of ensuring information support for economic entities has been developed. It is recommended to structure the information flow prior to its entering the level of processing by means of clustering.

Keywords: information support; modelling; information technology; information flow; clustering.

Постановка проблеми. Інформація є життєво необхідним, ключовим ресурсом економіки, що загострює потребу не стільки в кількості, скільки в якості інформації та процесі забезпечення нею як економічних суб'єктів, так і суспільства у цілому.

Становлення економіки, заснованої на знаннях, глобалізаційні процеси та загострення конкуренції посилюють необхідність розширення зовнішніх інформаційних потоків і особливої уваги до своєчасності отримання необхідної інформації. Вирішення проблеми інформаційного забезпечення економіки в цих умовах потребує системного підходу, тобто її слід розглядати як сукупність декількох проблемних взаємопов'язаних напрямів дослідження, а саме: обґрунтування методологічних засад інформаційного забезпечення; моделювання його організаційного механізму; розробка інформаційної технології, яка повинна охоплювати весь технологічний процес створення й органі-

¹ Poltava University of Economics and Trade, Ukraine.

зації інформаційного забезпечення економічних суб'єктів. Використання при цьому відповідних методів і моделей з урахуванням конкретної економічної ситуації є необхідним і першочерговим завданням.

Аналіз останніх досліджень та публікацій. Рівень інформаційного забезпечення економіки залежить від наявності та якості інформаційних ресурсів і продуктів на основі інформації, сконцентрованих у суб'єктів, які здійснюють його стосовно економічних суб'єктів. Для впорядкування та формалізації вибору найкращих варіантів зазначених складових інформаційного забезпечення часто пропонується використовувати оптимізаційні моделі. Зокрема, оптимізаційну модель забезпечення споживачів статистичною інформацією на мезорівні, що мінімізує витрати на створення та діяльність таких об'єктів підчас вибору варіанта їх територіального розміщення [9]; оптимізаційну модель, що мінімізує витрати на обладнання базових центрів мережі інформаційних центрів при органах державного управління, а також передачу інформації каналами зв'язку [5]; оптимізаційну модель формування інформаційного фонду інформаційними одиницями (книгами, статтями, звітами, повідомленнями, аналітичними матеріалами тощо) для здійснення інформаційного забезпечення економіки, що максимізує обсяг інформації для прийняття управлінського рішення з урахуванням пріоритетів [6]. Для вирішення проблеми інформаційного забезпечення діяльності інвестиційно-посередницьких організацій пропонується використовувати інформаційно-адаптивну модель інвестиційних систем [1]. Для забезпечення інформацією управлінських процесів на мікрорівні пропонується інформаційна модель, що складається з 3 інформаційних блоків внутрішнього середовища [2].

Відмічаючи важливе значення інформації для економічного розвитку в цілому, більшість дослідників недостатньо використовують математичний апарат, зокрема, для визначення механізму реалізації інформаційного забезпечення, розробки інформаційної технології. Отже, економіко-математичне моделювання інформаційної сфери, особливо інформаційного забезпечення, потребує активізації.

Метою дослідження є розширення методологічних підходів до моделювання інформаційного забезпечення економіки шляхом залучення можливостей математичного апарату

Основні результати дослідження. Підвищення рівня інформаційного забезпечення економіки передбачає широке використання математичного апарату, відповідних підходів до моделювання організації інформаційного забезпечення економічних суб'єктів цим стратегічним ресурсом, зокрема, теорії множин і математичної логіки.

Для моделювання інформаційної технології інформаційного забезпечення економічних суб'єктів доцільно використати досвід розробки геоінформаційних систем і технологій, загальна схема якої будується на основі аналізу вхідних і вихідних інформаційних потоків. Відповідно до цього вхідні та вихідні інформаційні потоки можна представити у вигляді декартових добутків [8, 15]:

$$VP = (VP_y \times TZ_c \times TZ_m \times TZ_n); \quad (1)$$

$$IH = (VP_{IH} \times CM_1 \times CM_2), \quad (2)$$

де VP – вхідний потік; IH – вихідний потік; TZ_c – технічне завдання для збору інформації; TZ_m – технічне завдання для зберігання та моделювання; TZ_n – технічне завдання для представлення даних після кінцевої обробки; VP_y – множина первинних даних, отриманих завдяки різним технологіям; VP_{IH} – множина уніфікованих даних, отриманих після збору та первинної обробки; CM_1, CM_2 – формат представлення результату.

Узагальнена геоінформаційна система, враховуючи вищезазначене, може бути представлена як стратифікована трирівнева структура:

$$\begin{aligned} SUO:VP_y \times TZ_c \times NM &\rightarrow IH; \\ SUM:IH \times TZ_m \times ND &\rightarrow CM_1; \\ SUD:TZ_n \times CM_1 &\rightarrow CM_2, \end{aligned} \quad (3)$$

де SUO – системний рівень збору та первинної обробки інформації; SUM – системний рівень моделювання та зберігання; SUD – системний рівень представлення даних; NM, ND – нормативні вимоги до даних при моделюванні та представленні інформації відповідно.

Якщо розбити інформаційну технологію інформаційного забезпечення економічних суб'єктів на 4 рівні, через які проходять всі інформаційні одиниці під час інформаційного технологічного процесу (збір, обробка, збереження, використання для надання інформаційних послуг) для включення їх до інформаційного фонду, то слід відзначити, що вхідні та вихідні потоки даних рівнів неоднорідні, тобто кожному з них відповідає технологічно незалежна сукупність інформаційних одиниць.

Отже, інформаційна технологія інформаційного забезпечення економічних суб'єктів може бути представлена як чотирьохрівнева система, де для кожного із рівнів характерні певні процеси та визначені правила обробки інформаційних одиниць відповідно до встановлених вимог до їх кінцевого представлення на цьому рівні, тобто множини вхідних і вихідних інформаційних одиниць неоднорідні та можуть бути представлені у вигляді декартового добутку:

$$\begin{aligned} ZDO:IOD_0 \times PR_{ZDO} \times PRW_{ZDO} &\rightarrow IOD_K; \\ OID:IOD_K \times PR_{OID} \times PRW_{OID} &\rightarrow IOD_G; \\ ZBO:IOD_G \times PR_{ZBO} \times PRW_{ZBO} &\rightarrow IOD_Z; \\ PID:IOD_Z \times PR_{PID} \times PRW_{PID} &\rightarrow IOD_V, \end{aligned} \quad (4)$$

де ZDO – системний рівень збору інформації; OID – системний рівень обробки інформаційних одиниць; ZBO – системний рівень збереження інформації; PID – системний рівень використання інформаційних одиниць для надання інформаційних послуг; IOD_0 – множина інформаційних одиниць вхідного зовнішнього інформаційного потоку, що надходить при збиранні інформації; $IOD_K, IOD_G, IOD_Z, IOD_V$ – множини інформаційних одиниць, отриманих як результат їх опрацювання на певному рівні; $PR_{ZDO}, PR_{OID}, PR_{ZBO}, PR_{PID}$ – процеси, що здійснюються на відповідному рівні; $PRW_{ZDO}, PRW_{OID}, PRW_{ZBO}, PRW_{PID}$ – правила, що використовуються для роботи з інформаційними одиницями на відповідному рівні.

На першому рівні з використанням різних технологій здійснюється збір інформації, тобто інформаційних одиниць (книг, статей, звітів, повідомлень, аналітичних матеріалів тощо), але відповідно визначеним правилам наповнення інформаційного фонду, що використовується для інформаційного забезпечення. На цьому рівні інформаційні одиниці перевіряються на дуплетність для вирішення питання щодо їх включення у фонд.

На другому рівні для проведення обробки отримані інформаційні одиниці розподіляються за ступенем їх готовності на окремі підмножини (PM), що не перетинаються, а інформаційні одиниці кожної з них обробляються за єдиними для цієї групи правилами:

$$\begin{aligned} IOD &= \bigcup_{\delta=1}^{PM} IOD_{\delta}; \\ IOD_{\delta} \cap IOD_{\lambda} &= \emptyset; \\ \delta &\neq \lambda, \delta, \lambda = \overline{1, PM}. \end{aligned} \quad (5)$$

Процеси обробки інформаційних одиниць для включення їх в інформаційний фонд мають ланцюговий характер та здійснюються згідно з правилами перетворення множини однорідних інформаційних одиниць, що дозволяє представити їх у декількох варіантах у межах одного інформаційного простору: $PRV_S: IOD_S \Rightarrow [IOD_S]'$.

Інформаційні одиниці кожної групи проходять визначені для неї процеси обробки, які здійснюються за встановленими правилами. У результаті формується множина готових до використання інформаційних одиниць – IOD_G , що передається на рівень зберігання.

На третьому рівні інформаційні одиниці, підготовлені до використання, розміщуються на сервері для збереження згідно визначених правил фізичної організації даних у ЕОМ, а інформація про них заноситься до бази даних.

На четвертому рівні за допомогою комплексу процесів та правил інформаційні одиниці використовуються для надання інформаційних послуг, зокрема, здійснюються пошукові процеси, процеси отримання з архіву, відправка замовнику, організація доступу, візуалізація.

Теорію множин доцільно використати для формалізації й аналізу структури вхідного інформаційного потоку. Структуризація інформаційного потоку перед його надходженням на рівень обробки може здійснюватися за допомогою кластеризації.

При ієрархічній кластеризації створюється ієрархія кластерів, що починається з декількох найменших і закінчується одним великим, до якого включаються всі дані [7, 159].

Алгоритм неієрархічної кластеризації починається з прийняття рішення стосовно кількості стартових кластерів або мінімальної наближеності двох об'єктів, що мають бути у межах одного кластера. Прогін через базу даних може здійснюватися багато разів, починатися з довільної кластеризації, ітеративно поліпшуючи її.

Нехай \tilde{n} об'єктів належать множині $\tilde{Z} = \{\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_{\tilde{n}}\}$ та існує множина певних характеристик $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_{\tilde{n}}\}$, до властиві кожному елементу мно-

жини \check{Z} . Якщо x_{ij} є результатом виміру i -ї характеристики l_j об'єкта, вектор $X_j = [x_{ij}]$ відповідає кожному ряду вимірів, то для множини об'єктів I отримаємо множину векторів, а саме $X = \{X_1, X_2, \dots, X_{\check{n}}\}$, що описують множину I . Алгоритм кластеризації повинен забезпечити розподіл цієї множини на v кластерів на основі даних, що містяться у множині X , при цьому кожен об'єкт l_j повинен належати лише одному кластеру, об'єкти одного кластера мають бути схожими, об'єкти різних кластерів – несхожими.

Розподіл об'єктів здійснюється за певним критерієм оптимальності, (цільової функції), який дозволяє з'ясувати, коли буде досягнутий необхідний розподіл. Отже, виникає потреба у знаходженні ступеня внутрішньої однорідності та різнорідності. Якщо при кластеризації відстань між X_i та X_j виявиться достатньо малою, то i -й та j -й об'єкти попадатимуть в один кластер. Якщо ця відстань достатньо велика, то вони розподіляються у різні кластери.

Функція $d(X_i, X_j)$ є мірою відстані, якщо $d(X_i, X_j) \geq 0$ для всіх X_i і X_j ; $d(X_i, X_j) = 0$ тоді і тільки тоді, коли $X_i = X_j$; $d(X_i, X_j) = d(X_j, X_i)$; $d(X_j, X_i) \leq d(X_j, X_h) + d(X_h, X_i)$, де X_i, X_h, X_j – будь-які три вектори [7, 154].

Значення $d(X_i, X_j)$ є відстанню між X_i і X_j та еквівалентно відстані між l_i та l_j відповідно до характеристик $\{S_1, S_2, \dots, S_n\}$.

Функція $k(X_i, X_j)$ є функцією схожості, якщо $0 \leq k(X_i, X_j) \leq 1$ для $X_i \neq X_j$; $k(X_i, X_j) = 1$; $k(X_i, X_j) = k(X_j, X_i)$.

Отже, нехай $I = \{l_1, l_2, \dots, l_{\check{n}}\}$ та $J = \{J_1, J_2, \dots, J_{\check{n}}\}$ – це дві множини об'єктів. Нехай $S = \{S_1, S_2, \dots, S_n\}$ є множиною характеристик, що генерують дві множини виміру $X = \{X_1, X_2, \dots, X_{\check{n}_1}\}$ і $Y = \{Y_1, Y_2, \dots, Y_{\check{n}_2}\}$, які відповідають множинам I та J .

Величину $D_i(I, J)$ називають мінімальною локальною відстанню між кластерами I та J :

$$D_1(I, J) = \min k(X_i, X_j), \quad (6)$$

при цьому $i = 1, \dots, \check{n}_1$; $j = 1, \dots, \check{n}_2$.

Максимальною локальною відстанню між кластерами I та J є:

$$D_2(I, J) = \max k(X_i, X_j), \quad (7)$$

при цьому $i = 1, \dots, \check{n}_1$; $j = 1, \dots, \check{n}_2$.

Середньою відстанню між кластерами I та J , що відповідає даній функції відстані, є:

$$D_3 = \sum_{j=1}^{\check{n}_2} \sum_{i=1}^{\check{n}_1} \frac{d(X_i, X_j)}{n_1 n_2}. \quad (8)$$

Отже, вся вхідна інформація, що буде використовуватися для інформаційного забезпечення (всі інформаційні одиниці інформаційного фонду), перед початком обробки має бути зафіксована на магнітному носії у вигляді файлів і зберігатися на сервері. У подальшому інформаційні одиниці, що тимчасово зберігалися на сервері, передаються виконавцям. Координатор, відповідальний за отримання вхідного загального інформаційного потоку, крім співставлення його інформаційних одиниць із наявними одиницями в

інформаційному фонді, здійснює розподіл цього потоку за типом інформаційних одиниць потоку (типом документів) для спрямування його виконавцям, що здійснюють обробку документів. Для цього координатор доповнює імена файлів вхідного потоку, ідентифікуючи їх відповідно до типу документів, що в подальшому забезпечить можливість їх кластеризації за цією характеристикою (табл. 1).

Ідентифікація та кластеризація за виконавцями не здійснюється, оскільки важливим є саме типізація отриманих інформаційних одиниць, рішення стосовно якої приймається лише координатором для уникнення розбіжностей у процесі розподілу інформаційних одиниць за типами різними виконавцями, оскільки це здійснюється емпірично.

Таблиця 1. Вхідні дані для кластеризації вхідного інформаційного потоку за типом документів, авторська розробка

№	Ім'я файла	Тип документа
1	1zN3uH3V.doc	книга
2	1ymCd6wnr.pdf	книга
3	3abXoM3j2.doc	звіт
4	4GxvLfqPU.doc	аналітичні матеріали
5	1SFYR0Ce.doc	книга
6	1JYRT0T0.doc	книга
7	1nVT9H94U.pdf	книга
8	2mSWeHsiA.pdf	стаття
9	4jrsc6Y8Z.doc	аналітичні матеріали
10	2xPvtGo8t.pdf	стаття
11	5nYqLEqE.pdf	методичні матеріали
12	2xGzt0sMa.doc	стаття
13	5vYbAp2t.doc	методичні матеріали
14	3qOXISdM2.doc	звіт
15	3KW1HFXQT.pdf	звіт
16	4osw3GoLP.pdf	аналітичні матеріали
17	3JDFaQ7kL.pdf	звіт
18	3JtoEX0JA.pdf	звіт
19	29qb1WtaJ.doc	стаття
20	3Rt1Yment.pdf	звіт
21	3ucMrzZdk.doc	звіт
22	6OnKFj7Um.pdf	інші
23	3ph6qJYW2.doc	звіт
24	5nJ2fap.pdf	методичні матеріали

Нехай множина QW є множиною інформаційних одиниць вхідного загального інформаційного потоку в електронній формі, тобто $QW = \{QW_1, QW_2, \dots, QW_n\}$. Серед множини характеристик кожної з них розглянемо ту, що буде використовуватися для вирішення конкретного завдання кластеризації, тобто ім'я файлу.

Отже, $\forall QW_n$ властива така характеристика як ім'я файлу – AZ , що описується множиною $AZ = \{AZ_1, AZ_2, \dots, AZ_n\}$.

Оскільки кожне ім'я файла складається з окремих символів, то $\forall QW_n = AZ_n = \{az_{1n}, az_{2n}, \dots, az_{jn}\}$, при цьому az_{1n} – символ типу документа, що надається координатором: 1 – книга, 2 – стаття, 3 – звіт, 4 – аналітичні матеріали, 5 – методичні матеріали, 6 – інші.

Кількість кластерів інформаційних одиниць зафіксуємо відповідно до типів документів (6): множина KK – книга, CT – стаття, ZV – звіт, AM – аналітичні матеріали, MM – методичні матеріали, IM – інші, тобто $KK = \{KK_1, KK_2, \dots, KK_q\}$; $CT = \{CT_1, CT_2, \dots, CT_p\}$; $ZV = \{ZV_1, ZV_2, \dots, ZV_z\}$; $AM = \{AM_1, AM_2, \dots, AM_e\}$; $MM = \{MM_1, MM_2, \dots, MM_\zeta\}$; $IM = \{IM_1, IM_2, \dots, IM_\tau\}$, кожному з яких відповідає певна множина виміру – ім'я файла, тобто $KX = \{KX_1, KX_2, \dots, KX_q\}$; $CX = \{CX_1, CX_2, \dots, CX_p\}$; $ZX = \{ZX_1, ZX_2, \dots, ZX_z\}$; $AX = \{AX_1, AX_2, \dots, AX_e\}$; $MX = \{MX_1, MX_2, \dots, MX_\zeta\}$; $IX = \{IX_1, IX_2, \dots, IX_\tau\}$, відповідно $\forall KK_q = KX_q = \{kx_{1q}, kx_{2q}, \dots, kx_{aq}\}$; $\forall CT_p = CX_p = \{cx_{1p}, cx_{2p}, \dots, cx_{qp}\}$, за аналогією і для інших кластерів.

Отже, завдання кластеризації – на основі даних, що містяться у множині AZ , розбити множину об'єктів QW на кластери відповідно до ідентифікації типу документа в імені файла, використовуючи функцію схожості:

$$f_s(QW, KK) = f_s(AZ_{QW}, KX_{KK}) = (az_{1n} = kx_{1q}) \quad (9)$$

для документа типу «книга», що за аналогією можна представити для решти пар множин.

Якщо вхідний інформаційний потік є множиною QW , при цьому KS – множина всіх кластерів, що відповідають типам документів, DD_{QW} , PP_{KS} – множини характеристик, що відповідають QW і KS , то узагальнено функцію кластеризації можна представити:

$$f_s(QW, KS) = f_s(DD_{QW}, PP_{KS}) = (dd_{1qw} = pp_{1ks}). \quad (10)$$

У результаті кластеризації отримаємо 6 груп файлів відповідно до типу документа (табл. 2).

Таблиця 2. Результати кластеризації інформаційних одиниць за типом документа, авторська розробка

Номер кластера	Кількість документів у кластері	Тип документа
1	5	книга
2	4	стаття
3	8	звіт
4	3	аналітичні матеріали
5	3	методичні матеріали
6	1	інші

Розглянутий підхід до обробки вхідного інформаційного потоку сприяє визначенню чіткої системи фізичної організації інформаційних одиниць, зафіксованих на магнітному носії, на основі ієрархічних зв'язків і запропонованого способу їх ідентифікації, що дає змогу впорядкувати розміщення інформаційних одиниць, забезпечити централізовану організацію даних та спростує порядок доступу до них на етапах обробки, пошуку та надання інформації.

Економіко-математичне моделювання є науковим інструментарієм, що забезпечує обґрунтування прийняття рішень і виконання завдань. Інформація при цьому стає основою цих процесів, результатом яких є поява нових знань про об'єкт моделювання, що враховуються для вирішення поставлених завдань.

Висновки. Економіко-математичне моделювання необхідно використовувати у процесах організації та здійснення інформаційного забезпечення, що сприяє вирішенню проблеми надання економічним суб'єктам ефективної інформації. Оцінка методологічних підходів до моделювання інформаційного забезпечення дозволила розширити його загальний інструментарій шляхом використання теорії множин і кластерного аналізу. Подальші дослідження доцільно спрямувати на визначення та використання методів математичної логіки для вдосконалення процесів пошуку і надання інформації.

1. *Аксенова Е.А.* Совершенствование инвестиционной деятельности через систему информационных технологий // Предпринимательство: проблемы и решения. – Донецк: Кассиопея, 1998. – С. 103–108.
2. *Голячук Н.В.* Розробка моделі забезпечення інформацією керівництва підприємства // www.rusnauka.com.
3. *Донець Л.І., Баранцева С.М.* Інформаційне забезпечення стратегічного управління прибутком торговельного підприємства // Вісник Донецького державного університету економіки і торгівлі ім. М. Туган-Барановського. – Серія: Економічні науки. – 2007. – №3. – С. 87–92.
4. Інноваційний розвиток економіки: модель, система управління, державна політика / За ред. Л.І. Федулової. – К.: Основа, 2005. – 552 с.
5. *Калина Р.И.* О концепции информатизации Оренбургской области. – Оренбург: ИПКРО, 1999. – 24 с.
6. *Львович Э.М.* Математические методы и механизмы, оптимизации информационных ресурсов как фактора повышения эффективности управления социально-экономическим развитием региона // Вестник ВГТУ. – 2003. – Вып. 3. – С. 37–39.
7. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод – СПб.: БХВ-Петербург, 2004. – 336 с.
8. *Цветков В.Я.* Геоинформационные системы и технологии. – М.: Финансы и статистика, 1998. – 288 с.
9. *Чернышева Е.Н., Божко В.П.* Один из подходов к совершенствованию процесса обработки информации на районном уровне статистической информационной системы // Вопросы статистики. – 1997. – №7. – С. 19–23.
10. *Чорноус Г.* Інформаційне забезпечення проактивного управління // Вісник Київського національного торговельно-економічного університету. – 2012. – №5. – С. 102–114.

Стаття надійшла до редакції 22.10.2014.