

**Т. А. ГОНЧАРЕНКО**, канд. техн. наук, доц., КНУБА, Київ,  
**В. М. МІХАЙЛЕНКО**, докт. техн. наук, проф., зав. каф., КНУБА,  
Київ

## **ЗАСТОСУВАННЯ МЕТОДІВ БАГАТОВИМІРНОГО АНАЛІЗУ ДАНИХ ДЛЯ МОДЕЛЮВАННЯ ТЕРИТОРІЇ ПІД ЗАБУДОВУ**

В статті розглянуто можливість застосування багатовимірної моделі даних для вирішення задачі вибору територій, які плануються під забудову. Запропоновано алгоритм вибору методу багатовимірний аналіз даних в залежності від кількості вимірів та якості вхідної інформації об'єкта, що досліджується. Показано доцільність застосування теорії нечітких множин для створення автоматизованих систем з неповною інформацією та високою складністю об'єкта управління, якою є територія під забудову. Лл.: 4. Бібліогр.: 14 назв.

**Ключові слова:** багатовимірний модель даних; автоматизована система; теорія нечітких множин; територія під забудову.

**Постановка проблеми.** Навколишній світ багатовимірний, в якому кожен об'єкт характеризується безліччю параметрів. При побудові моделі, як правило, доводиться знижувати розмірність реального світу, окреслюючи коло властивостей, що представляють інтерес для дослідження. Традиційні засоби аналізу даних, які представлені у вигляді реляційних баз даних, не можуть повною мірою задовольняти таким вимогам. Для аналізу інформації найбільш зручним способом її подання є багатовимірний модель даних. Така модель представляє собою гіперкуб, ребрами якого є послідовності значень одного з аналізованих параметрів – виміри, які стосуються досліджуваній предметній області [1]. Це дозволяє виконувати багатовимірний аналіз даних і вивчати інформацію про об'єкт відразу по декільком вимірам. Така множинність вимірів передбачає подання даних у вигляді багатовимірної моделі. На перетинах осей вимірювань розташовуються дані, що кількісно характеризують параметри, які досліджуються. Такі дані називають мірами (рис. 1).

**Мета статті.** При моделюванні території під забудову як просторової системи [2, 3] використовуються різні типи даних, які отримані з багатьох джерел інформації. Метою статті є дослідження можливості застосування багатовимірної моделі даних для вирішення задач, які виникають при плануванні будівельного простору. Такі задачі характеризуються складністю та інколи відсутністю повної інформації про об'єкт моделювання, що ускладнює розробку автоматизованої

системи управління території під забудову.

**Основна частина.** Простір, що представляє інтерес для дослідження, опишемо як множину вимірів  $N$ . Тоді будь-яка точка  $P$  в даному просторі може бути представлена  $(n_1, n_2, \dots, n_i, \dots, n_m)$ ,  $n_i \in N$  множиною вимірів  $(t_1, t_2, \dots, t_j, \dots, t_k)$ ,  $t_i \in T$ , де  $T$  – множина мір. Тоді багатовимірний простір визначається схемою  $M$ , атрибутами якої є множина вимірів і множина мір, і описується кортежем

$$M = \langle N, T \rangle. \quad (1)$$

Кортеж (1) характеризується множиною відносин  $m$  над цією схемою  $M$ . Для будь-якої точки  $P$  простору, існує проекція  $m[P]$  задана виразом:

$$\forall P \in N, \exists m[P], \forall p \in m[P]: p \in ALL, \quad (2)$$

де  $ALL$  – всі можливі значення виміру.

Багатовимірний простір  $S$  може бути визначено таким декартовим добутком:

$$S(m) = \{ \otimes_{P \in N} (m[P] \cup ALL) \cup \{0, \dots, 0\} \}. \quad (3)$$

Основним способом дослідження задач аналізу даних є їх формалізація моделі та її подальший аналіз. Зі збільшенням розмірів і складності системи істотно ускладнюється її моделювання за допомогою відомих математичних виразів, так як збільшується числом змінних і параметрів. В результаті, створення адекватної моделі стає практично неможливим. Автори [4 – 6] запропонували лінгвістичну модель, яка використовує не математичні вирази, а слова, що відображають якість (рис. 1).

Застосування словесної моделі не забезпечує кількісну точність математичного моделювання, але дозволяє створити достатньо якісну модель. У цьому предметом дослідження стає нечіткість мови опису системи.

Основна проблема аналізу таких багатовимірних матриць даних полягає в тому, що людський мозок не здатний одночасно оперувати більш ніж трьома вимірами простору. Для зведення багатовимірних даних до двох вимірів з мінімальними втратами інформації розроблена спеціальна група методів статистичного аналізу даних, яка називається багатовимірний аналіз даних.

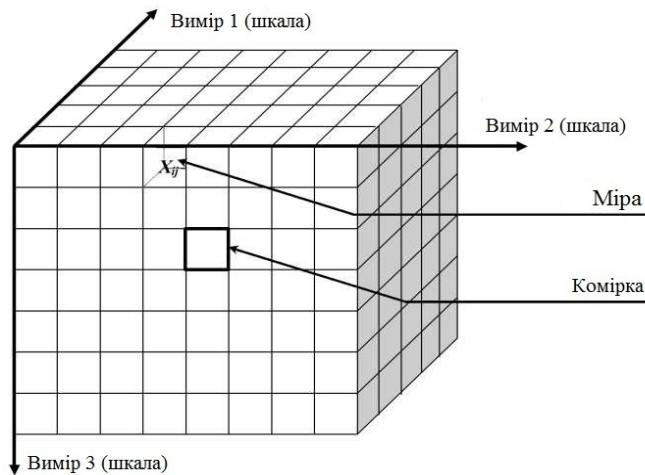


Рис. 1. Багатовимірна модель даних

Існує велика кількість підходів до аналізу багатовимірних даних. Залежно від того, яку саме відповідь необхідно отримати, проводиться і вибір методу. Відомі наступні методи багатовимірного аналізу:

- дерева класифікації;
- дискримінантний аналіз;
- багатовимірний дисперсійний аналіз;
- кластерний аналіз;
- багатовимірне шкалювання;
- аналіз головних компонент.

На рис. 2 надано процедуру вибору методу багатовимірного аналізу даних в залежності від приналежності об'єктів до груп.

Методи аналізу, що працюють тільки з однією змінною в певний момент часу, називаються одновимірними методами, які дозволяють досить просто отримати як первинну інформацію для аналізу, так і цілі ланцюжки взаємно незалежних результатів. Методи одновимірного аналізу використовують лінійні зв'язки між змінними, а методи багатовимірного будуються на припущенні існування при моделюванні як лінійних та і нелінійних зв'язків. Таким чином одновимірний аналіз є частковим випадком багатовимірного аналізу.

Багатовимірні дані при аналізі за одним виміром розглядаються як одновимірні. Послідовний аналіз по кожному виміру може розглядатися як одновимірний аналіз одновимірних даних. Отримані незалежні результати можуть бути представлені агрегуванням даних по кожному виміру з відповідними мірами. При встановленні відповідного ставлення в результаті може бути отримана нова багатовимірна модель, у

властивостях даних якої вже присутні залежності між вимірами. Такі дані вже не можуть аналізуватися тільки по одному вимірюванню як одномірні дані методами одновимірного аналізу (рис. 3). А результатом є новий багатовимірний інформаційний об'єкт.

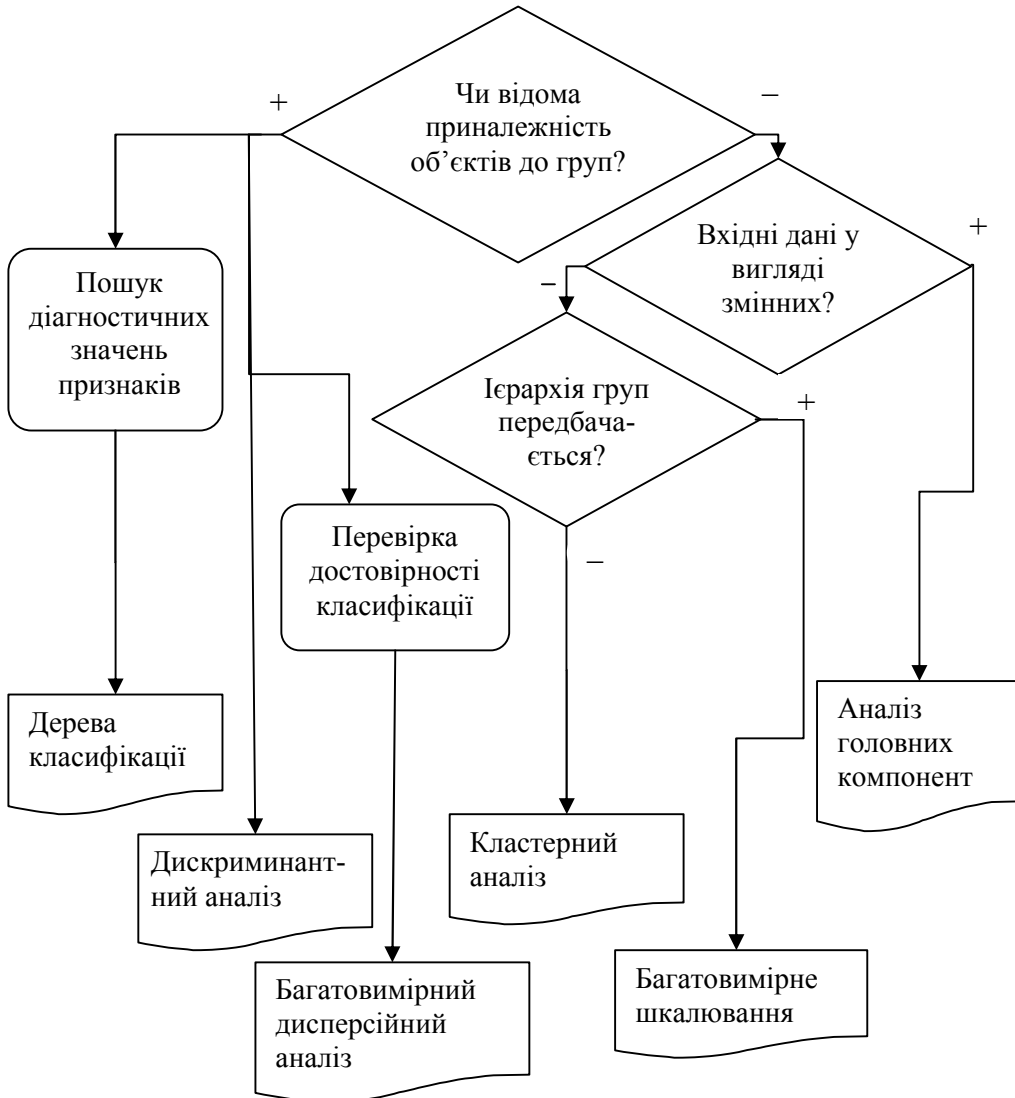


Рис. 2. Вибір методу багатовимірного аналізу даних

Як приклад багатовимірних даних розглянемо дані геоінформаційної системи для генерального планування території під забудову. Такі дані є поєднанням реляційних баз даних, в яких

зберігаються атрибутивні характеристики просторових об'єктів, з їх координатно організованою інформацією. Як правило, у цифровому вигляді координатні дані просторових об'єктів для моделювання території під забудову представлені у вигляді векторної моделі даних. В таких моделях координатні дані подаються у вигляді сукупності просторових координат  $\{X, Y, Z\}$ , де  $Z$  – залежна змінна від  $X, Y$ . У загальному випадку кожен точку просторового об'єкта з координатами  $(x, y)$  описує деякий вектор станів  $(z_1, \dots, z_n)$ .

Для всієї моделі території під забудову отримуємо набір векторів  $Z_1, \dots, Z_n$  – параметрів просторових об'єктів. Показники стану  $Z_1, \dots, Z_n$  можна поділити на:

- вхідні змінні  $C_i$  ( $i = 1, \dots, I$ ), які отримані будь яким способом;
- вихідні  $D_j$  ( $j = 1, \dots, J$ ), які виражаються через вхідні.

Взаємозв'язок між вхідними та вихідними даними можна описати такою функціональною залежністю:

$$D_j = F(C_1, \dots, C_I), j = 1, \dots, J, I + J = n. \quad (4)$$

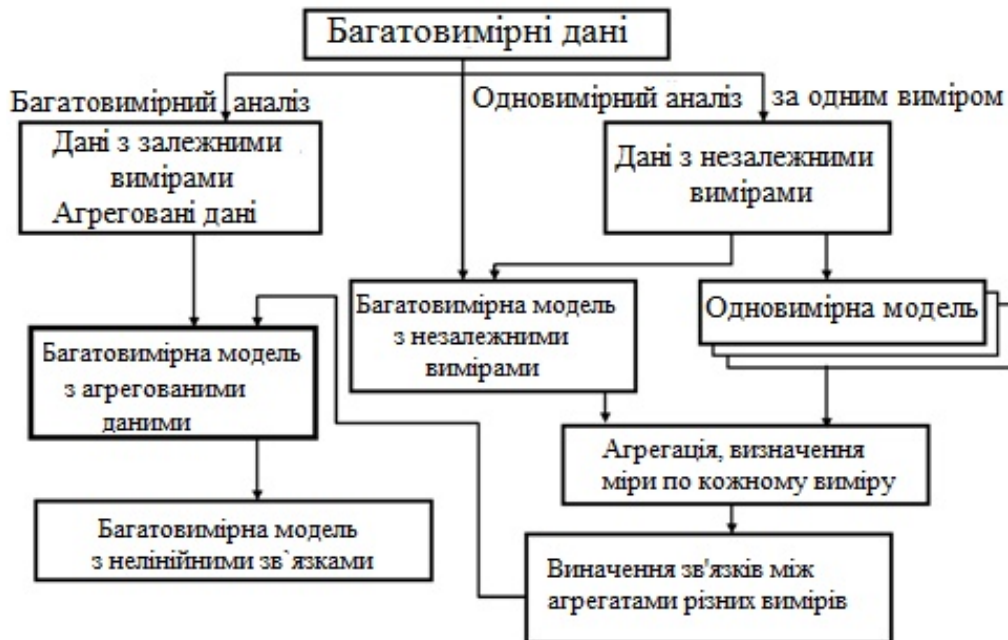


Рис. 3. Залежність мірності даних від типу аналізу

Створення якісної математичної моделі адекватної для вирішення практичного завдання завжди є непростим процесом, який інколи

призводить до втрати якісної інформації, яка складно піддається формалізації. Багато сучасних завдань управління просто не можуть бути вирішені класичними методами через дуже велику складність математичних моделей, які їх описують.

До появи теорії нечітких множин деякі характеристики з властивими їм невизначеностями ігнорувалися при моделюванні. На рис. 4 в залежності від рівня детермінованості об'єкта дослідження показані області найбільш ефективного застосування сучасних технологій управління. Для систем з неповною інформацією та високою складністю об'єкта управління оптимальними є нечіткі методи управління. Такими є нечіткі системи управління і нечіткі нейронні мережі.

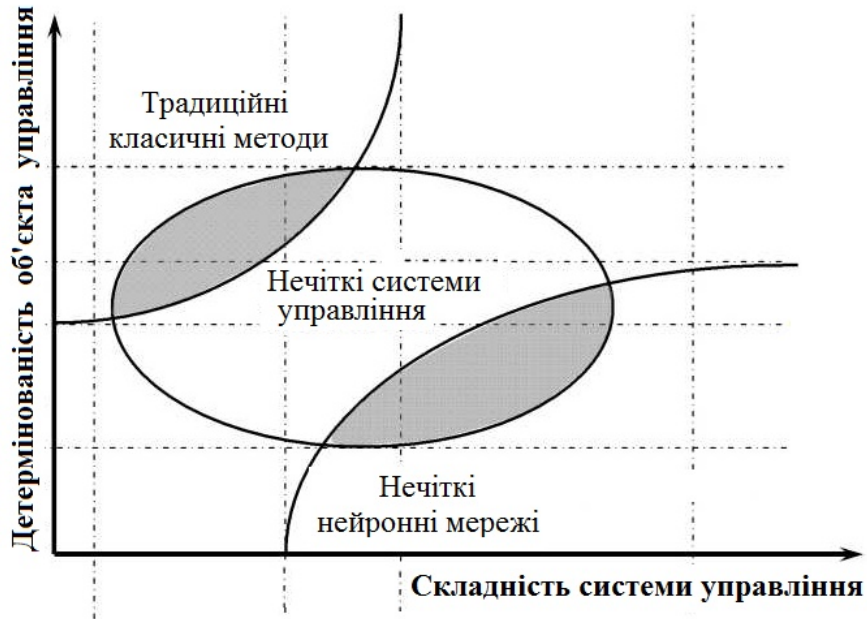


Рис. 4. Вибір сучасної технології управління в залежності від складності системи і рівня детермінованості об'єкта

В умовах застосування автоматизованих систем управління (АСУ) відбувається трансформація функцій людини, виникають нові зв'язки між людиною і системою. Деякі функції, які виконувала людина, повністю передаються системі. Комп'ютеризація сприяє розширенню можливостей суб'єкта, часом якісно змінюючи зміст його діяльності. І ця діяльність повинна бути чітко визначена і закріплена за суб'єктом так само, як і за автоматизованою системою (АС).

Найпростіша автоматизована система передбачає взаємозв'язок типу: людина - АС - людина, виділяючи тим самим два елементи: людина - АС і АС - людина. Для складних автоматизованих систем, розподілених у часі і просторі, кількість елементів збільшується до чотирьох:

- людина - АС;
- АС - людина;
- АС - АС;
- людина-людина.

Виконання таких взаємозв'язків багаторазово повторюється, що вимагає чіткої регламентації всіх виконуваних функцій, поділу операцій між людиною і машиною, введення інженерного терміна - технологічний процес.

Будь-яка система може розглядатися з різних сторін - з динамічної, структурної, логічної, фізичної. В результаті виникають різні архітектурні уявлення як окремі аспекти програмної архітектури, які розглядають специфічні властивості програмної системи. Розглядаючи цілі і обмеження у вигляді симетричних елементів логічної схеми можна досить просто сформулювати на їх основі рішення, яке є по суті вибором однієї або кількох з наявних альтернатив. При цьому нечітке рішення може розглядатися як деяка процедура, нечіткість якої є наслідком неточності формулювання поставлених цілей і обмежень, тобто вплив нечіткої цілі  $G$  і нечіткого обмеження  $C$  на вибір альтернатив характеризується їх перетином, яке утворює нечітка множина рішень  $D$ :

$$D = G \cap C. \quad (5)$$

Принципова відмінність між випадковістю і нечіткістю полягає в тому, що функція приналежності завжди є гіпотезою про характер цілей і наявних обмежень, що дозволяє будувати оцінки для альтернатив за допомогою формального апарату. Даний процес може бути описаний в формі алгоритму, відповідно до якого конкретні значення параметрів будь-якої складної властивості об'єднуються, а потім на цій основі робиться висновок про їх узагальнене значення.

Головний структурний елемент даного алгоритму - правило "якщо ..., то ...", що відображає властивість причинно-наслідкового зв'язку. Така система, спирається безпосередньо не на факти або об'єкти, а на відомості про них. Модель, що охоплює інформаційну систему, може бути представлена у вигляді метабази, в якій міститься інформація по кожному виду об'єкта обліку. З іншого боку інформаційна система може бути подана в вигляді функціональної системи у вигляді множини

функцій. Таким чином, цілі і обмеження задаються як нечіткі множини. Взаємозв'язок між ними може бути визначено відношенням (5).

**Висновки.** Розглянуто проблему моделювання будівельного простору з позицій методів, що враховують нечіткість описів моделі об'єкта управління. Для вирішення цієї проблеми запропоновано алгоритм вибору методу багатовимірного аналізу даних в залежності від груп об'єктів. Наукова новизна дослідження полягає у визначенні альтернатив нечіткої множини рішень в залежності від нечіткої цілі і нечітких обмежень, які існують при виборі територій, які плануються для забудови. Практична цінність результатів дослідження полягає в застосуванні запропонованого вибору сучасної технології управління в залежності від складності системи і рівня детермінованості об'єкта будівництва.

#### **Список літератури:**

1. Антонов В.В. Теоретические и прикладные аспекты построения моделей информационных систем / В.В. Антонов, Г.Г. Куликов, Д.В. Антонов. – LAP LAMBERT Academic Publishing GmbH & Co.KG, Germany, 2011. – 134 p.
2. Mihaylenko V. Modeling of Spatial Data on the Construction Site Based on Multidimensional Information Objects / V. Mihaylenko, T. Honcharenko, K. Chupryna, Yu. Andrashko, S. Budnik // International Journal of Engineering and Advanced Technology (IJEAT). – 2019. – Vol. 8. – P. 3934-3940 [Online]. Available: <https://www.ijeat.org/wp-content/uploads/papers/v8i6/F9057088619.pdf>.
3. Terentyev O. Multidimensional Space Structure for Adaptable Data Model / O. Terentyev, S. Tsiutsiura, T. Honcharenko, T. Lyashchenko // International Journal of Recent Technology and Engineering (IJRTE). – 2019. – Vol. 8. – Issue 3. – P. 7753-7758. [Online]. Available: <https://www.ijrte.org/wp-content/uploads/papers/v8i3/C6318098319.pdf>.
4. Артемьев С.А. ГИС конструктор со средствами анализа данных для создания информационно-аналитических систем / С.А. Артемьев, С.С. Замай, А.А. Путенко // Вестник КазНУ. – 2004. – Т. 9. – № 3 (42). – С. 188-192.
5. Alkathiri M. Multi-dimensional geospatial data mining in a distributed environment using MapReduce / M. Alkathiri, A. Jhummarwala, M.B .Potdar // J. Big Data. – 2019. – Vol. 6. – P. 82-90.
6. Evans M.R. Enabling spatial big data via CyberGIS: challenges and opportunities. CyberGIS for geospatial discovery and innovation / M.R. Evans, D. Oliver, K. Yang, X. Zhou, R.Y. Ali, S. Shekhar – Dordrecht: Springer, 2019.
7. Kravets R.B. Organization of the multidimensional implementation and analysis in data related bases / R.B. Kravets. – Available: <http://ena.lp.edu.ua/bitstream/ntb/9816/1/18.pdf>.
8. Hristodulo O.I. Integrated processing of spatial information based on multidimensional data models / O.I. Hristodulo. – DSc Thesis, Ufa State Aviation Technical University, 2012. – 395 p.
9. Morak M. Impact of Disjunction on Reasoning under Existential Rules / Morak M. – PhD in Computer Science. Thesis, University of Oxford, 2015.
10. Milani M. Ontology-Based Multidimensional Contexts with Applications to reality Data Specification and Extraction / M. Milani, L. Bertossi // In Proc. of the International



Symposium on Rules and Rule Markup Languages for the Semantic Web. – Springer LNCS 9202, 2015. – P. 277-293

**11.** *Body M.* Handling Evolutions in Multidimensional Structures / *M. Body* // IEEE 19th Int. Conf. on Data Engineering (ICDE), March 5-8, Bangalore, India, 2003.

**12.** *Nagpal S.* Complexity metric for multidimensional models for data warehouse / *S. Nagpal, A. Gosain, S. Sabharwal* // CUBE International Information Technology Conference, ACM, 2012. – P. 360-365.

**13.** *Chen W.-K.* Linear Networks and Systems / *W.-K. Chen*. – Belmont CA: Wadsworth, 1993. – P. 123-135.

**14.** *Goeken M.* Multidimensional reference models for data warehouse development / *M. Goeken, R. Knackstedt* // International Conference on Enterprise Information Systems, Citeseer, 2007. – P. 347-354.

### **References:**

**1.** Antonov, V.V., Kulikov, G.G., and Antonov, D.V. (2011), *Theoretical and applied aspects of building models of information systems*, LAP LAMBERT Academic Publishing GmbH & Co.KG, Germany, 134 p.

**2.** Mihaylenko, V., Honcharenko, T., Chupryna, K., Andrashko, Yu., and Budnik, S. (2019), "Modeling of Spatial Data on the Construction Site Based on Multidimensional Information Objects" in *International Journal of Engineering and Advanced Technology (IJEAT)* at Volume 8, August 2019, pp. 3934-3940 [Online]. Available: <https://www.ijeat.org/wp-content/uploads/papers/v8i6/F9057088619.pdf>.

**3.** Terentyev, O., Tsiutsiura, S., Honcharenko, T., and Lyashchenko, T. (2019), "Multidimensional Space Structure for Adaptable Data Model" in *International Journal of Recent Technology and Engineering (IJRTE)*, Volume 8, Issue 3, September 2019, pp. 7753-7758. [Online]. Available: <https://www.ijrte.org/wp-content/uploads/papers/v8i3/C6318098319.pdf>

**4.** Artemyev, S.A., Zamay, S.S., and Pitenko A.A. (2004), GIS constructor with data analysis tools for creating information-analytical systems // *Bulletin of KazNU*, Vol. 9, No. 3 (42), pp. 188-192.

**5.** Alkathiri, M., Jhummarwala, A., and Potdar, M.B. (2019), "Multi-dimensional geospatial data mining in a distributed environment using MapReduce". *J. Big Data*, Vol. 6, pp. 82-90.

**6.** Evans, M.R., Oliver, D, Yang, K, Zhou, X, Ali, R.Y., and Shekhar S. (2019), *Enabling spatial big data via CyberGIS: challenges and opportunities. CyberGIS for geospatial discovery and innovation*, Dordrecht: Springer.

**7.** Kravets, R.B. (2006), "Organization of the multidimensional implementation and analysis in data related bases", [Online]. Available: <http://ena.lp.edu.ua/bitstream/ntb/9816/1/18.pdf>.

**8.** Hristodulo, O.I. (2012), "Integrated processing of spatial information based on multidimensional data models", DSc Thesis, Ufa State Aviation Technical University, 395 p.

**9.** Morak, M. (2015), "Impact of Disjunction on Reasoning under Existential Rules". PhD in Computer Science. Thesis, University of Oxford.

**10.** Milani, M. and Bertossi, L. (2015), "Ontology-Based Multidimensional Contexts with Applications to reality Data Specification and Extraction". In *Proc. of the International Symposium on Rules and Rule Markup Languages for the Semantic Web*, Springer LNCS 9202, pp. 277-293

**11.** Body, M. (2003), "Handling Evolutions in Multidimensional Structures", *IEEE 19th Int. Conf. on Data Engineering (ICDE)*, March 5-8, Bangalore, India.

**12.** Nagpal, S., Gosain, A., and Sabharwal, S. (2012), "Complexity metric for multidimensional models for data warehouse", in: *CUBE International Information Technology Conference, ACM*, pp. 360-365.

13. Chen, W.-K. (1993), *Linear Networks and Systems* Belmont, CA: Wadsworth, pp. 123-135.

14. Goeken, M., and Knackstedt, R. (2007), "Multidimensional reference models for data warehouse development", in: *International Conference on Enterprise Information Systems, Citeseer*, pp. 347–354.

*Статтю представив д-р техн. наук, проф. КНУБА Терентьєв О.О.*

*Поступила (received) 10.10.2019*

Honcharenko Tetyana, Ph.D, Associate Professor  
Kyiv National University of Construction and Architecture  
Avenue Povitroflotsky, 31, Kyiv, Ukraine, 03037  
Tel: +38 093 022-82-01, e-mail: geocad@ukr.net  
ORCID ID: 0000-0003-2577-6916

Mihaylenko Victor, Dr.Sci.Tech, Professor  
Kyiv National University of Construction and Architecture  
Avenue Povitroflotsky, 31, Kyiv, Ukraine, 03037  
Tel: (044) 360-55-09, e-mail: kpm\_knuba@ukr.net  
ORCID ID: 0000-0002-9573-9873

УДК 004.942+045

**Застосування методів багатовимірного аналізу даних для моделювання території під забудову / Гончаренко Т.А., Михайленко В.М. // Вісник НТУ "ХПІ".** Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2019. – № 28 (1353). – С. 5 – 15.

В статті розглянуто можливість застосування багатовимірної моделі даних для вирішення задачі вибору територій, які плануються під забудову. Запропоновано алгоритм вибору методу багатовимірного аналізу даних в залежності від кількості вимірів та якості вхідної інформації об'єкта, що досліджується. Показано доцільність застосування теорії нечітких множин для створення автоматизованої системи з неповною інформацією та високою складністю об'єкта управління, якою є територія під забудову. Іл.: 4. Бібліогр.: 14 назв.

**Ключові слова:** багатовимірна модель даних; автоматизована система; теорія нечітких множин; територія під забудову.

УДК 004.942+045

**Применение методов многомерного анализа данных для моделирования территории под застройку / Гончаренко Т.А., Михайленко В.М. // Вестник НТУ "ХПИ".** Серія: Інформатика и моделирование. – Харьков: НТУ "ХПИ". – 2019. – № 28 (1353). – С. 5 – 15.

В статье рассмотрена возможность применения многомерной модели данных для решения задачи выбора территорий, планируемых под застройку. Предложен алгоритм выбора метода многомерного анализа данных в зависимости от количества измерений и качества входящей информации исследуемого объекта. Показана целесообразность применения теории нечетких множеств для создания автоматизированной системы с неполной информацией и высокой степенью сложности объекта управления, которым является территория под застройку. Ил.: 4. Библиогр.: 14 назв.

**Ключевые слова:** многомерная модель данных; автоматизированная система; теория нечетких множеств; территория под застройку.

UDK 004.942+004.045

**Application of multidimensional data analysis methods for modeling of the construction site // Honcharenko T.A., Mihaylenko V.M. // Herald of the National Technical University "KhPI".** Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2019. – № 28 (1353). – P. 5 – 15.

The article considers the possibility of using a multidimensional data model to solve the problem of choosing the territories planned for development. An algorithm is proposed for choosing a method of multivariate data analysis depending on the number of measurements and the quality of the incoming information of the investigated object. The expediency of using the theory of fuzzy sets to create an automated system with incomplete information and a high degree of complexity of the control object, which is the building area, is shown. Figs.: 4. Refs.: 14 titles.

**Keywords:** multidimensional data model; automated system; theory of fuzzy sets; construction site.