

УДК 004.49:004.056.57:004.942

А. В. ШЕВЧЕНКО, аспірантка,
Державний університет телекомунікацій, Київ

ІДЕНТИФІКАЦІЯ ДИСФУНКЦІОНАЛЬНИХ СТАНІВ МЕДИЧНИХ СИСТЕМ ЗА ДОПОМОГОЮ ТАКСОНОМЕТРИЧНОГО МЕТОДУ

Виявлено потребу щодо прискорення ідентифікації самого факту присутності дисфункціонального стану. Поняття небезпеки «нульового дня» перенесено з галузі кібербезпеки в медицину. Проблему розв'язано за допомогою таксонометричного методу. Для виявлення дисфункціональних станів людини запропоновано знаходити нестандартні значення та відстежувати поведження показників стану людини. Таксонометричний метод та критерії визначення еталонних об'єктів було вдосконалено з урахуванням особливостей медицини.

Ключові слова: дисфункціональні стани людини; загрози «нульового дня»; таксонометричний метод; еталонний об'єкт.

Вступ

Сьогодні одним із головних факторів зростання ефективності досліджень у будь-якій галузі людської діяльності, зокрема в медицині, є впровадження інформаційних технологій. У такій суто гуманітарній сфері, як медицина, найбільш ефективно впровадження інформаційних технологій базується на вдалій формалізації фактів, що мають місце в медичній галузі. Велике значення в цьому сенсі мають засоби класифікації, зокрема міжнародний класифікатор хвороб (МКХ) [1], який містить найбільш сучасний перелік дисфункціональних станів, тобто хвороб та інших проблем, пов'язаних зі здоров'ям людини.

Розглянемо основні проблеми стосовно практичного використання МКХ.

1. МКХ включає в себе лише ті дисфункціональні стани, які досить добре вивчено медичною наукою.

2. Кількість дисфункціональних станів, що потребують класифікації (а далі й розробки засобів протидії), постійно зростає, насамперед із таких причин:

- зростання кількості вірусів та небезпечних мікроорганізмів унаслідок їх мутацій;
- зростання кількості дисфункціональних станів, які вдалось ідентифікувати;
- зростання кількості небезпечних зовнішніх умов, які можуть призводити до нових дисфункціональних станів;
- поєднання кількох дисфункціональних станів в одному організмі.

Головне завдання медицини — запобігання можливим дисфункціональним станам та усунення зазначених станів, що вже виникли. Аналіз історії розвитку міжнародного класифікатора хвороб, який тепер існує в 10-й версії [1], показує, що кількість описаних дисфункціональних станів зростає дуже високими темпами [2]. Нині тільки їх перелік займає понад 600 обліково-видавничих аркушів при кількості від 18 до 38 тис. знаків на аркуші [1] (усього близько 20 тис. знаків). Це означає, що поряд з новою версією МКХ завжди знайдеться певний дисфункціональний стан, який не

тільки ще не описали, а й досі не навчилися ідентифікувати. А як відомо, для того, щоб протидіяти дисфункції, потрібно передусім її ідентифікувати.

За аналогією з кібербезпекою невідомі раніше дисфункціональні стани людини можна називати загрозами «нульового дня», тобто загрозами, які тільки-но проявили себе у вигляді кібератаки або нової хвороби і щодо яких відлік часу на пошук засобів протидії лише розпочався (нульовий день пошуку засобів протидії). У нульовій точці відліку медичні, біологічні або інформаційні системи можуть виявитись беззахисними [3; 4]. Загрози «нульового дня» належать до найскладніших. Тому дослідження засобів ідентифікації дисфункціональних станів для вчасного початку пошуку протидії таким станам дедалі актуальніші.

Основна частина

Постає *актуальна суто практична наукова проблема в галузі медицини* — подальше прискорення процесів ідентифікації нових дисфункціональних станів. Адже медицина, як і всі інші галузі знань, давно працює на межі своїх можливостей. Тому вихід на якісно новий рівень результативності вимагає *міждисциплінарності* досліджень. Сформульовану наукову проблему пропонується розв'язати за допомогою методів інформаційних технологій. Для цього побудуємо таблицю, виділивши частини раніше невирішеної проблеми та сформулювавши мету досліджень щодо методів інформаційних технологій для медичної галузі.

Розподіл частин раніше невирішеної проблеми між галузями науки

Частина раніше невирішеної проблеми	Галузь науки, методи якої використовуються
1. Ідентифікація самого факту наявності дисфункціонального стану	Інформаційні технології
2. Деталізована ідентифікація дисфункціонального стану	Інформаційні технології та медицина
3. Усунення дисфункціонального стану, що має місце	Медицина
4. Розробка засобів запобігання появі дисфункціональних станів у майбутньому	Медицина

© А. В. Шевченко, 2018

Мета статті — розширення можливостей ідентифікації невідомих раніше дисфункціональних станів людини за допомогою методів інформаційних технологій, а саме таксонометричного методу.

Навіть якщо не вдається вчасно знайти способи захисту від загрози «нульового дня», але вдається ідентифікувати сам факт наявності загрози, це дозволяє вжити запобіжних заходів. Виявлення факту наявності загрози «нульового дня» є репрезентативною інформацією для прийняття управлінського рішення щодо застосування інструментів протидії дисфункціональному стану. Тому сконцентруємось на першому етапі (див. таблицю), тобто ідентифікації самого факту наявності дисфункціонального стану.

Традиційним методом ідентифікації небезпечних ситуацій, пов'язаних із дисфункціональними станами, є виявлення нестандартного перебігу інформаційних процесів та поведження об'єктів, що їх породжують, а отже, маркерами дисфункції є прояви нестандартності. Але життя змінюється, і те, що було нормою вчора може вже сьогодні втратити цей статус. До того ж біологічний організм може поводитись по-різному в різному віці, за різних умов, після різних перенесених раніше захворювань тощо. Тому **актуальним** є створення автоматичних адаптивних систем виявлення нестандартного поведження медичних і біологічних об'єктів у різних умовах.

Постає завдання оперативної кластеризації об'єктів по групах різного ступеня небезпеки щодо виникнення дисфункціонального стану. З огляду на великий обсяг інформації, стрімку мінливість біологічних структур та їхніх параметрів бажано, аби така ідентифікація виконувалась із максимальним залученням інформаційних технологій щодо виконання рутинної частки роботи.

Методи кластерного аналізу широко використовують для підтримки прийняття управлінських рішень у багатьох аспектах: соціальна поведінка [5], упорядкування знань [6], мобільні технології [7], інформаційні системи [8], розумні речі [9], програмне забезпечення [10–12] тощо. У нашому випадку можна було б просто поділити всі спостережувані біологічні об'єкти за рівнями дисфункціональності. Але якщо об'єкти постійно змінюють моделі свого поведження, стосовно якого визначатимемо рівні дисфункціональності, то обрання еталона «нормальної поведінки» значно ускладнюється.

У такій ситуації є сенс використати таксонометричний метод [13]. Як показує аналіз публікацій, цей метод особливо популярний серед дослідників країн колишнього СРСР. Адаптуємо таксонометричний метод під завдання щодо визначення еталонів поведження організму людини за певних умов.

Нехай маємо виборку з n станів людей, що є об'єктами спостереження. Це можуть бути різні люди, або ті самі особи в різні моменти часу. Наша інформаційна система здійснює постійний моніторинг цих об'єктів за формальними показниками, здобутими в результаті об'єктивних досліджень та на підставі суб'єктивних даних. Це можуть бути, наприклад:

- температура, пульс, тиск, показники крові, дані досліджень із використанням медичної діагностичної апаратури;
- суб'єктивні скарги пацієнтів;
- суб'єктивні дані експертів-лікарів (результати консилиуму) тощо.

Набір показників може бути довільний. Єдина вимога: вони мають бути вимірювані, тобто мати числове вираження в довільних шкалах. Перелік контрольованих показників (факторів) може змінюватись відповідно до таких обставин:

- захворювань, які вже були в людини;
- зовнішніх умов, в яких перебуває об'єкт спостереження;
- загального тону особи тощо.

Нехай загальна кількість контрольованих показників (факторів) дорівнює m . Якщо при кластеризації нових об'єктів інформація стосовно певних показників відсутня, то для збереження чинності наведених далі процедур значення таких показників до моменту їх реального визначення беруться такі, що дорівнюють показникам еталона або математичному сподіванню певного показника за доступними даними. Після отримання першого набору показників для всіх спостережуваних об'єктів будується матриця

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1n} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & & \vdots & & \vdots \\ x_{m1} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}.$$

Для визначення еталона поведження матрицю X нормуємо, знаходячи математичне сподівання та середнє квадратичне відхилення для кожного з показників:

$$\bar{X}_i = M[x_{ij}] = \frac{1}{n} \sum_{j=1}^n x_{ij},$$

$$D_i = M[x_{ij} - \bar{X}_i]^2 = \frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{X}_i)^2,$$

$$\sigma_i = \sqrt{D_i}.$$

За допомогою зазначених показників нормуємо кожний елемент початкової матриці показників

$$z_{ij} = \frac{x_{ij} - \bar{X}_i}{\sigma_i}.$$

Це перетворення звело всі показники до єдиної міри, що дозволяє вільно та з високим ступенем

адекватності порівнювати їх між собою та виконувати подальші перетворення. Далі потрібно побудувати еталонний об'єкт, який має найменшу дисфункцію, створивши набір еталонних показників, які й буде взято за характеристики еталонного об'єкта.

Згідно з класичним таксонометричним методом еталонними вважають або найменші, або найбільші показники серед усіх наявних:

$$Z_{i_etalon} = \min(z_{ij}) \text{ або } Z_{i_etalon} = \max(z_{ij}), \quad j = \overline{1, n}.$$

У нашому випадку діяти так можна, якщо показник має безпосередній (лінійний чи, можливо, нелінійний, але прямий) зв'язок із рівнем нормального стану людини. Таке співвідношення більш характерне для технічних систем. У разі біологічних систем частіше використовують поняття *діапазону показників, що відповідають нормальному стану*. Якщо така норма невідома, то припускаємо, що більшість людей, котрі підлягають спостереженню, не мають дисфункцій. Тому за відомим співвідношенням Парето 20:80 будемо підмножину початкової вибірки, куди включаємо 80% значень показника, найбільш наближених до математичного сподівання аналізованого показника. Можлива варіація розміру вибірки в діапазоні 51–80%. Якщо природу показників вивчено достатньо, то можлива ситуація, коли показники виходять за межі нормального стану лише в один бік. У такому разі відкидаються значення, що не входять у діапазон 51–80% лише з одного боку.

Після цього для нової вибірки знаходимо нове математичне сподівання, яке й беремо за еталон, тобто середину діапазону нормальних значень:

$$Z_{j_etalon} = M(z_{ij}) = \frac{1}{n} \sum_{j=1}^n z_{ij}, \quad j = \overline{1-n}.$$

Більш адекватним буде *зважене середнє з ваговими коефіцієнтами* β_j , які в оцінці збільшуватимуть вагу даних щодо людей, стан яких визначено більш точно за рахунок докладних попередніх досліджень:

$$Z_{i_etalon} = M(z_{ij}) = \frac{1}{n} \sum_{j=1}^n \beta_j z_{ij}, \quad j = \overline{1-n}.$$

Неодмінною умовою використання вагових коефіцієнтів є їх нормування такого виду: $\sum_{j=1}^n \beta_j = 1$.

У разі великої розбіжності значень певних показників можливе використання логарифмічної шкали.

Як іншу модифікацію класичного таксонометричного методу пропонуємо використання різних способів пошуку еталона для різних показників (мінімум, максимум, зважене середнє, зважене середнє за вибіркою, зменшеною за Парето). Тоді загальний вираз для еталона запишемо у вигляді

$$Z_{i_etalon} = etalon(z_{ij}), \quad j = \overline{1-n}.$$

Після визначення еталона потрібно знайти віддаленість усіх об'єктів від еталона. За міру віддаленості можна взяти квадрат евклідової відстані в m -вимірному просторі показників або евклідову відстань, або інші міри, традиційні для кластерного аналізу (міра Чебишова, Хеммінга, степенева відстань тощо). У класичному випадку використовуємо квадрат евклідової відстані у прямому вигляді

$$R_j = \sum_{i=1}^m (z_{ij} - Z_{i_etalon})^2,$$

або, з урахуванням коефіцієнтів ваги показників,

$$R_j = \sum_{i=1}^m \beta_i (z_{ij} - Z_{i_etalon})^2.$$

Найближчі об'єкти (люди) вважатимуться найбільш здоровими, а найвіддаленіші — потенційно найбільш дисфункціональними. Додаткову аналітичну інформацію можна буде отримати за результатами аналізу зміни в часі віддаленості об'єктів від еталона, із прогнозуванням збільшення рівня дисфункціональності та з визначенням основних тенденцій. Серед нових тенденцій можливе збільшення або поява нових дисфункціональних станів, а також перехід ознак дисфункціональності до категорії ознак нормального стану.

Наведений підхід простий за своєю суттю і його можна легко реалізувати у відповідному програмному забезпеченні. Але проблема полягає в тому, що кількість об'єктів, які потрібно досліджувати, може бути дуже велика. Тому наступним питанням є дослідження вимог і здатності звичайних бюджетних комп'ютерів щодо реалізації запропонованого алгоритму. Як середовище моделювання взято Матлаб, яке за своїм основним призначенням забезпечує високий ступінь оптимальності виконання матричних операцій.

Зміни параметрів для різних об'єктів спостереження ілюструє рис. 1, де по горизонталі наведено номери параметрів, а по вертикалі — їхні значення. Товста лінія відповідає еталону (в цьому разі — за критерієм максимуму).

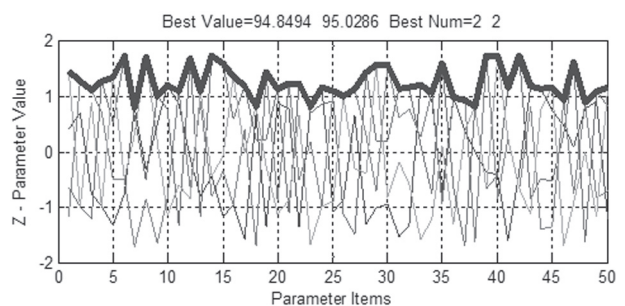


Рис. 1. Поточні значення параметрів для набору об'єктів та параметри еталонного об'єкта (мажоранта)

Дендрограму, в якій по горизонталі наведено номери об'єктів спостереження, а по вертикалі — їхні евклідові відстані від еталона в просторі показників, зображено на рис. 2.

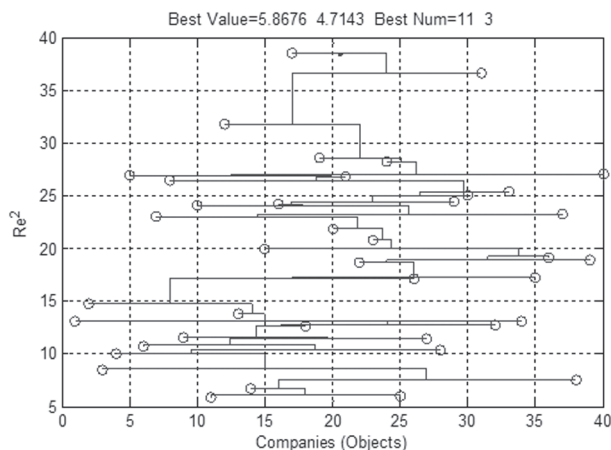


Рис. 2. Дендрограма об'єктів спостереження

Стан об'єктів із найменшими відстанями є найбільш нормальний. Існують певні групи об'єктів, віддалених від еталону, які при тривалому спостереженні можна об'єднати в кластер нетипових, але нормальних об'єктів. Хоча можливо, що це буде кластер зародження (виявлення) нового типу дисфункціональності за епідеміологічним типом. Конкретний тип кластера визначається по всій сукупності ознак при аналізі конкретної ситуації.

Вочевидь, зазначені об'єкти потребуватимуть особливої уваги.

Варто зазначити, що як підозрілі завжди сприймаються об'єкти в зоні великих відстаней. У нашому випадку йдеться про об'єкти за номерами 12, 17, 31.

Моделювання показало, що реалізація запропонованого алгоритму на звичайній бюджетній обчислювальній техніці можлива. Так, час розрахунків для 4000 об'єктів спостереження за 50 параметрами становив 2.136149 с, без урахування часу на реалізацію графіки.

Висновки

1. Найбільш небезпечними для біологічних систем є загрози «нульового дня».

2. Адаптований у статті таксонометричний метод дозволяє в автоматичному режимі виокремлювати потенційно дисфункціональні об'єкти для їх відстежування з метою організації вчасної протидії.

3. Додатковий вииграш у часі та точності ідентифікації дисфункцій можуть надати знання щодо основних закономірностей розвитку дисфункцій, зокрема епідеміологічного типу. У [3; 4] подано висновок щодо аналогії закономірностей біологічних та цифрових епідемій. Це дозволяє описати основні залежності розвитку дисфункціональних станів S-подібними залежностями [14; 15]. А далі слід лише уточнювати рішення з урахуванням нових статистичних даних.

4. Напрямок подальших пошуків є дослідження умов та динаміки трансформації потенційно дисфункціональних об'єктів у нормальні об'єкти,

що перебувають в особливому стані. Фактично йдеться про виявлення особливих станів, які насправді можуть вважатись нормальними.

Список використаної літератури

1. **ICD-10. International Statistical Classification of Diseases and Related Health Problems. Tenth Revision. Volume 1** // Geneva: World Health Organization, 1992. 698 с.

2. **WHO Historical Collection. Disease classifications and nomenclature documents** [Електронний ресурс]. 2018. Режим доступу:

<http://www.who.int/library/collections/historical/en/index1.html>

3. **Shevchenko V., Shcheblanin Ju., Shevchenko A. The Epidemiological Approach to Prognosis and Management of Information Incidents** // *Наука і техніка Повітряних Сил Збройних Сил України*. 2017. № 4(29). P. 145–150.

4. **Shevchenko V., Shevchenko A. The Epidemiological Approach to Information Security Incidents Forecasting for Decision Making Systems** // *13-th International Conference Perspective Technologies and Methods in MEMS Design (MEMSTECH). Proceeding. Polyana*. 2017. P. 174–177.

5. **Geiger D., Schulze T., Seedorf S. [a. o.]. Managing the crowd: towards a taxonomy of crowdsourcing processes** [Електронний ресурс] // *In Proceedings of the 17th Americas Conference on Information Systems (Sambamurthy V and Tanniru M, Eds), AIS, Detroit, MI*. Режим доступу:

<https://pdfs.semanticscholar.org/bb30/7fb8140d42e1a21c872eabfc9e400cf3347b.pdf>

6. **Hasan H. A taxonomy of modes of knowledge sharing between disparate group** [Електронний ресурс] // *In Proceedings of the 13th Pacific Asia Conference on Information Systems (Bapna R and Sambamurthy V, Eds), AIS, Hyderabad*. Режим доступу:

<https://pdfs.semanticscholar.org/04c9/7b30378b1ff0050cc003d212612847cd335e.pdf>

7. **Nickerson R., Varshney U., Muntermann J. and Isaac H. Taxonomy development in information systems: developing a taxonomy of mobile applications** [Електронний ресурс] // *In Proceedings of the European Conference on Information Systems*. Режим доступу:

https://halshs.archives-ouvertes.fr/file/index/docid/375103/filename/ECIS_2009_taxonomy_final_3.pdf

8. **Nickerson R., Muntermann J. and Varshney U. Taxonomy development in information systems: a literature survey and problem statement** [Електронний ресурс] // *In Proceedings of the 16th Americas Conference on Information Systems*. Режим доступу:

<https://pdfs.semanticscholar.org/f75d/d219dde068353678c69877fb5dfd7db12c95.pdf>

9. López T. S. *Taxonomy, technology and applications of smart objects* [Електронний ресурс] / López T. S., Ranasinghe D. C., Patkai B. and McFarlane D. // *Information Systems Frontiers* 11(4), 1–20. Режим доступу:

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.838.4246&rep=rep1&type=pdf>

10. Ducasse S. and Pollet D. *Software architecture reconstruction: a process-oriented taxonomy* [Електронний ресурс] // *IEEE Transactions on Software Engineering* 35(4), 573–591. Режим доступу:

<http://ieeexplore.ieee.org/document/4815276/?reload=true>

11. Krug S., Campidelli H. and Nickerson R. C. *A preliminary taxonomy for software failure impact: categorizing the impact on enterprises when software fails* [Електронний ресурс] // *In Proceedings of the 18th Americas Conference on Information Systems (Jessup L and Valacich J, Eds), Seattle, Washington*. Режим доступу:

<https://pdfs.semanticscholar.org/7e13/cf2f-1148b227fa465b22f984cf8f54e4c2e3.pdf>

12. Nickerson R., Varshney U. & Muntermann J. *A Method for Taxonomy Development and Its Application in Information Systems* [Електронний ресурс] // *Europ. Journ. of Information Systems*. May 2013, Vol.22, Issue 3, P. 336–359. Режим доступу:

<https://doi.org/10.1057/ejis.2012.26>

13. Шалаєва О. А., Колеснев В. И. *Таксонометрический метод рейтинговой оценки деятельности различных типов и форм сельскохозяйственных организаций* [Електронний ресурс] // *Проблемы экономики (Белорусская гос. с/хоз. акад.)* 2011. С. 237–244. Режим доступу:

<https://cyberleninka.ru/article/n/taksonometricheskij-metod-reytingovoy-otsenki-deyatelnosti-razlichnyh-tipov-i-form-selskohozyaystvennyh-organizatsiy>

14. Шевченко А. В., Шевченко В. Л. *Грубі моделі розвитку в медицині* // *Медична інформатика та інженерія*. 2009. №4. С. 52–55.

15. Шевченко В. Л. *Оптимізаційне моделювання в стратегічному плануванні*. Київ, 2011. 283 с.

Рецензент: канд. техн. наук. доцент С. В. Довбешко, Державний університет телекомунікацій, Київ.

А. В. Шевченко

ИДЕНТИФИКАЦИЯ ДИСФУНКЦИОНАЛЬНЫХ СОСТОЯНИЙ МЕДИЦИНСКИХ СИСТЕМ ПРИ ПОМОЩИ ТАКСОНОМЕТРИЧЕСКОГО МЕТОДА

Выявлена потребность ускорения идентификации самого факта наличия дисфункционального состояния. Понятие опасности «нулевого дня» перенесено из области кибербезопасности в медицину. Проблема решена при помощи таксонометрического метода. Для обнаружения дисфункциональных состояний человека предложено находить нестандартные значения и отслеживать поведение показателей состояния человека. Таксонометрический метод и критерии определения эталонных объектов были усовершенствованы с учетом особенностей медицины.

Ключевые слова: дисфункциональные состояния человека; угрозы «нулевого дня»; таксонометрический метод; эталонный объект.

A. V. Shevchenko

IDENTIFICATION OF DYSFUNCTIONAL STATES OF MEDICAL SYSTEMS BY TAXONOMETRIC METHOD

The need to accelerate the processes of identifying new dysfunctional states of a person is revealed. As a part of the previously unresolved problem, identification of the fact of the presence of a dysfunctional condition is determined. The notion of the danger of «zero day» is transferred from the field of cyber security to medicine. Threats of «zero day» are defined as the most complex. The problem is solved using the taxonomic method. It is suggested to reveal non-standard values and behavior of indicators of a person's state for revealing of dysfunctional states of a person. It is proposed to perform separate identification procedures for people of different ages who have suffered various diseases and are in different external conditions.

The use of cluster analysis in various fields of human activity was analyzed. Taxonomic method was chosen for tasks of identification of dysfunctional states of a person. The preliminary values of the indicators were normalized by means of mathematical expectation and mean-square deviation of indicators. The taxonomic method allowed the creation of benchmarks in the absence of prior information on new types of dysfunctional states. Taking into account the specifics of the medical industry, the criteria for determining the reference objects have been adjusted. To do this, it is proposed to create a subset of 51-80 percent of the primary sample, which does not include values with the highest mean square deviations. The mathematical expectation of a new subset is taken as a benchmark. Also, the possibility of using as a distance from the standard of other measures in addition to the Euclidean distance was considered, namely: the measure of Chebyshev, Heming, power distance. The possibility of implementing the proposed approach on ordinary budget personal computers has been experimentally tested. Based on the simulation results, a dendrogram was constructed in which the numbers of observation objects are located horizontally, and their Euclidean distance from the reference in the space of indicators is vertically.

Keywords: dysfunctional states of man; threats of «zero day»; taxonomic method; etalon object.