

# **ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ**

**Монографія**

*під ред. д.е.н., проф. Пономаренка В.С.*

**Харків**

**2019**

## УДК 004.78.2

Авторський колектив: к.т.н., доц. Н. О. Бринза, к.е.н., викл. О.В. Вільхівська – глава 1; д.е.н., проф. О.І. Пушкар, к.е.н., доц. Є.М. Грабовський – глава 2; к.ф.-м.н., доц. В. П. Бурдаєв – глава 3; к.т.н., доц. В. В. Браткевич, к.т.н., доц. В. Є. Климнюк – глава 4; ст.викл. Г.О. Плеханова, к.е.н., доц. О. Б. Плоха – глава 5; д.т.н., проф. О. Г. Руденко, д.т.н., проф. О. О. Безсонов – глава 6; д.т.н., проф. О. Г. Руденко, викл. Романюк О. С. – глава 7; д.т.н., проф. С.Г. Удовенко, к. т.н., доц. Л. Е. Чала – глава 8; к.т.н., доц. А. О. Поляков – глава 9, параграфи 9.1, 9.2; к.т.н., доц. В. М. Федорченко – глава 9, параграфи 9.3, 9.4; к.т.н., доц. Н. Г. Аксак, ст. викл. Д. М. Росінський – глава 10; к.т.н., доц. М. Ю. Лосев – глава 11; к.ф.-м.н., доц. В. В. Федько – глава 12; д.т.н., проф. М. М. Корабльов, к.т.н., ст. викл. О. О. Фомічов – глава 13.

### Рецензенти:

**Гороховатський В. О.** – доктор технічних наук, професор кафедри інформатики Харківського національного університету радіоелектроніки;

**Ляшенко С. О.** – доктор технічних наук, професор кафедри безпеки життєдіяльності Харківського національного технічного університету сільського господарства імені Петра Василенка;

**Бодянський Є. В.** – доктор технічних наук, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки.

*Рекомендовано до видання рішенням ученої ради Харківського національного економічного університету імені Семена Кузнеця (протокол № 9 від 27.05.2019 р.)*

I 76 **Інформаційні системи та технології:** монографія / за заг. ред. В. С. Пономаренка. - Х. : Видавництво «Стиль-іздат», 2019. - 462 с .

*В монографії розглянуті сучасний стан та перспективи розвитку інформаційних систем і технологій електронних мультимедійних видань. Монографія представляє інтерес як для фахівців, сфера діяльності яких безпосередньо пов'язана з розробкою інформаційних систем і технологій електронних мультимедійних видань, так і для більш широкого кола фахівців. Вона буде корисною викладачам, аспірантам і студентам, що спеціалізуються в області інформаційних технологій, і всім, хто серйозно цікавиться проблемами інформаційного суспільства.*

**За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор.**

## ЗМІСТ

ВСТУП	4
ГЛАВА 1. ВИКОРИСТАННЯ МОДЕЛЕЙ ЕКСПЕРТНОГО АНАЛІЗУ ДЛЯ ОЦІНЮВАННЯ СТАНУ ІТ-ГАЛУЗІ УКРАЇНИ	7
ГЛАВА 2. ТЕХНОЛОГІЯ СТВОРЕННЯ СПІЛЬНОГО КОНТЕНТУ В ВЕБ-БАЗОВАНОМУ НАВЧАЛЬНОМУ КАБІНЕТІ	22
ГЛАВА 3. INTEGRATION CHAT-BOTA @RIBS_karkas_bot WITH EXPERT SYSTEM	37
ГЛАВА 4. МОДЕЛІ ОЦІНЮВАННЯ ЕФЕКТУ ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ЕЛЕКТРОННОГО НАВЧАННЯ	52
ГЛАВА 5. КРАУДСОРСИНГОВА ПЛАТФОРМА ЯК ІТ-ІНСТРУМЕНТ ДЛЯ ЗАБЕЗПЕЧЕННЯ ГЕНДЕРНОЇ РІВНОСТІ В УНІВЕРСИТЕТІ	68
ГЛАВА 6. FORECAST ERROR MEASURES: CRITICAL REVIEW AND PRACTICAL RECOMMEND INTRODUCTION AND TASK STATEMENT	82
ГЛАВА 7. ANALYSIS OF THE INFLUENCE OF INTERFERENCE ON THE STATISTICAL PROPERTIES OF ESTIMATES OF NON-STATIONARY PARAMETERS	97
ГЛАВА 8. КЛАСИФІКАЦІЯ ЕЛЕКТРОННИХ НАУКОВО-ТЕХНІЧНИХ ТЕКСТІВ В ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМАХ	108
ГЛАВА 9. АНАЛІЗ БЕЗОПАСНОСТІ ПРОТОКОЛІВ ПЕРЕДАЧІ СООБЩЕНИЙ. В MESSENGER TELEGRAM	124
ГЛАВА 10. МЕТОДОЛОГІЧНА ОСНОВА ПРОЕКТУВАННЯ ПРОБЛЕМНО-ОРІЄНТОВАНИХ ВЕБ-СЕРВІСІВ	139
ГЛАВА 11. НЕЧІТКО-МНОЖИННА МЕТОДИКА ОЦІНКИ КОНКУРЕНТНОСПРОМОЖНОСТІ ПІДПРИЄМСТВА	155
ГЛАВА 12. DEVELOPMENT OF A MULTIMEDIA INTERACTIVE ELECTRONIC EDUCATIONAL EDITION USING THE MODULES OF VARIOUS FORMAT	170
ГЛАВА 13. КЛАСИФИКАЦИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ИММУННОЙ МОДЕЛИ КЛОНАЛЬНОГО ОТБОРА	185

## ВСТУП

Розвиток суспільства в сучасному світі характеризується дуже сильним впливом на нього комп'ютерних та інформаційних технологій (ІТ), які увірвалися в усі сфери людської діяльності, утворюючи глобальний інформаційний простір, що забезпечує поширення інформаційних потоків в суспільстві.

Складність інформаційних систем (ІС) невпинно зростає, відповідно набувають актуальності питання щодо ефективного управління процесами створення, тестування та впровадження таких систем. Комплексне вирішення цієї проблеми – складна й довготривала робота, що потребує високої кваліфікації задіяних у ній працівників. Адже, незважаючи на значний прогрес, багато питань у сфері автоматизації проектування ІС не піддається повній формалізації й виконується на інтуїтивному рівні з використанням досвіду та прогностичних здібностей фахівців, експертних оцінках, імітаційному моделюванню функціонування системи тощо.

Сьогодні інформаційні ресурси є не менш суттєвими, ніж трудові, матеріальні або енергетичні. Роль інформації стала настільки значною, що зараз можна говорити про інформаційну економіку (тобто таку, що ґрунтується на інформації), та інформаційну сферу управління (керівників, вчених, спеціалістів і службовців тощо).

Без достовірної, повної та своєчасно отриманої інформації неможливо керувати будь-яким виробництвом, своєчасно реагувати на виклики зовнішнього середовища. Саме тому на сьогодні проблема збирання та обробки інформації є одним з ключових завдань управління, вирішити яку можна виключно на базі використання сучасної обчислювальної техніки та новітніх інформаційних технологій. Слід відзначити, що сама обчислювальна техніка все ширше використовується як в управлінні виробничими процесами, так і на рівні економічної системи в цілому. Однак матеріальні витрати на зберігання, передавання та перероблення інформації все частіше перевищують аналогічні витрати, наприклад, на енергетику. І в цьому контексті проблема інформатизації набирає не тільки технічну, а й економічну складову.

Принциповим є той факт, що комп'ютеризація суспільства – це не тільки техніка, а й люди. Нова комп'ютерна ідеологія повинна обіймати не тільки технічні питання, вона має охоплювати у т. ч. підготовку відповідних спеціалістів і керівників. Керівник, що не володіє комп'ютером, не може сподіватися на прогрес у своїй роботі. Таким чином, комп'ютеризація у

загальному розумінні потребує комплексного підходу щодо вирішення широкого спектру проблем, – від суто технічних до освітніх та соціальних.

Необхідно також відзначити, що поява нових інформаційних систем і технологій, поряд з уже наявними, просунутими областями їх використання, здатна формувати такі напрямки, які будуть давати найбільший ефект і сприяти підвищенню технологічного рівня наукомістких виробництв.

У монографії представлені результати наукових досліджень в галузі інформаційних систем і технологій, що представлені на Міжнародній науково-практичній конференції «Інформаційні технології та системи», що відбулася 10-11 квітня 2019 року на базі Харківського національного економічного університету. У ній відображені найбільш цікаві результати за найсучаснішими науковими напрямками застосування, проектування та визначення проблем і перспектив ІТ в технічних системах, захисту інформації в інформаційних та комунікаційних системах тощо.

В главі 1 монографії сформовано тривимірний набір показників, що відображає стан ІТ-індустрії України та включає інвестиційну привабливість регіонів, внутрішній стан ІТ-ринку та зарплату. Побудовано математичну модель стану ІТ-індустрії України, яка представлена як комплексний показник. Проведено класифікацію регіонів з використанням кластерного аналізу.

Технологію створення спільного контенту в веб-базованому навчальному кабінеті запропоновано в главі 2 даної монографії. Розглянуті основні типи складності основних видів технологій створення спільного контенту. Аналізуються основні кроки реалізації педагогічної моделі формування спільного контенту.

В главі 3 представлені результати інтегрування чат-бота @ribs\_karkas\_bot з експертною системою для організації консультування в режимі онлайн. Розглянуто структуру взаємодії чат-бота і агентів експертної системи в онлайн режимі.

Глава 4 містить моделі оцінювання ефекту використання системи підтримки електронного навчання. Розглянута множина критеріїв для оцінювання ефекту застосування альтернативних варіантів систем електронного навчання.

Передумови розробки краудсорсингової платформи як ІТ-інструменту для забезпечення гендерної рівності в університеті розглянуті в главі 5.

В главі 6 монографії на основі результатів аналізу розроблено рекомендації щодо вибору критеріїв прогнозування при вирішенні практичних задач.

Задача ідентифікації нестационарного об'єкту з допомогою найбільш простого в обчислювальному відношенні однокрокового алгоритму модифікованого Качмажа розглянута в главі 7. Визначено умови збіжності регуляризованого алгоритму Качмажа при оцінюванні нестационарних параметрів, які описано Марківською моделлю першого порядку, при наявності завад вимірів. Отримані неасимптотичні оцінки є досить загальними і залежать як від ступеня нестационарності об'єкту, так і від статистичних характеристик завад.

В главі 8 пропонується класифікація електронних науково-технічних текстів в інформаційно-пошукових системах. Кінцевою метою такої класифікації є пошук у колекції класів документів, які є близькими до заданих еталонів за запропонованими критеріями.

Аналіз безпеки протоколів передачі повідомлень в messenger Telegram розглянуто в главі 9. Проаналізовано первинну реєстрацію клієнта на сервері, процедуру встановлення безпечного прозорого зашифроване з'єднання з іншим вузлом, а також процедуру генерації ключових даних.

В главі 10 сформовано методологічну основу проектування проблемно-орієнтованих веб-сервісів для розв'язання наукомістких проблем, яка ґрунтується на комплексному використанні різних технологій, методів та моделей для синтезу, інтелектуалізації, підвищення ефективності та застосування розподілених спеціалізованих комп'ютерних систем у рамках єдиної технології створення проблемно-орієнтованих систем. Розроблено модель інформаційної безпеки для захисту медичної інформації та компонентів інформаційної системи.

В главі 11 пропонується нечітко-множинна методика оцінки конкурентноспроможності підприємства.

Глава 12 монографії присвячена питанням побудови та використання мультимедійного інтерактивного електронного навчального посібника в галузі знань «Інформаційні технології». Розглянуто переваги використання pdf-формату як первинного для реалізації електронного навчального посібника.

В главі 13 розглянуто вирішення задачі класифікації даних на основі імунної моделі клонального відбору, в якій в результаті взаємодії антитіл з антигенами залишаються антитіла, які характеризуються високими афінністю до розпізнаваним антигенів. Проведено експериментальні дослідження, що підтверджують ефективність запропонованих алгоритмів.

# ГЛАВА 1

## ВИКОРИСТАННЯ МОДЕЛЕЙ ЕКСПЕРТНОГО АНАЛІЗУ ДЛЯ ОЦІНЮВАННЯ СТАНУ ІТ-ГАЛУЗІ УКРАЇНИ

### Вступ та постановка задачі

Український ІТ-ринок має величезний потенціал до зростання, проте нестабільність податкового регулювання, нестача спеціалістів і низка регуляторних бар'єрів є стримуючими факторами для розвитку галузі. У зв'язку з обранням Україною курсу на євроінтеграцію, сфера розробки програмного забезпечення є чи не єдиною експортною галуззю України, що стабільно зростає. Лише за 2018 р. приріст збільшився на 26 %. Також галузь виступає потужним економічним мультиплікатором: 1 робоче місце в ІТ створює ще 3 у суміжних професіях. Заробітна плата ІТ-спеціаліста у 5 разів вища за середню в країні. При цьому високий рівень оплати зберігається не лише в столиці, але й в регіонах.

Але й у цьому бізнесі є свої проблеми. Зокрема, спостерігається серйозна нестача молодих спеціалістів, які спроможні вирішувати нові завдання, що постійно з'являються та змінюються. Ця нестача пов'язана з тим, що одночасно існують проблеми як з підготовки конкурентоспроможних спеціалістів профільними навчальними закладами, так й «вимивання» кваліфікованих кадрів не тільки з менш благополучних регіонів в крупні вітчизняні ІТ-центри, а й за межі країни. Отже, проведення аналізу такої ситуації за допомогою сучасних математичних підходів та оцінка впливу виявлених факторів на результат, є актуальними питаннями.

Процеси, що відбуваються на ІТ-ринку праці, важливі і привертають увагу не тільки вчених, але й представників бізнесу. Багато питань щодо прогнозів розвитку, перегріву ІТ-ринку, створення ІТ-кластерів, визначення переваг і недоліків української ІТ-галузі, і багато іншого піднімаються в роботах Р. А. Винничук і Т. В. Склярчук [1], Е. Рубіна [2], в інтерв'ю з представниками українського бізнесу, такими як Дарина Стремetsька [3] та Анна Грабовська [4], ІТ-співтовариством доу і іншими. В своїй роботі Н. О. Бринза та А. А. Гаврилова [5] провели дослідження стану ІТ-галузі на основі виділення показників, але не за всіма регіонами України. Це ставить під сумнів зроблені висновки стосовно ступеня впливу кожного показника на результат та адекватної оцінки запропонованої цими авторами моделі.

Метою дослідження в даній роботі є проведення аналізу стану ІТ-галузі за більшою кількістю регіонів, ніж це було зроблено в роботі Н. О. Бринзи та А.

А. Гаврилової. В основу дослідження покласти сформовану модель експертної оцінки показників [5], які характеризують дану галузь, виявити значність впливу факторів на результат, перевірити адекватність наведеної моделі, а також провести упорядкування регіонів в порівняно однорідні групи із застосуванням математичних методів.

### **Основна частина**

Позитивні зміни в ІТ-галузі України пов'язані з появою якісних навчальних ІТ-програм у ЗВО, привабливі умови праці ІТ-компаній в Україні стрімко збільшує інтерес до галузі серед фахівців різних сфер [5].

Аналітиками спільноти програмістів на сайті dou.ua були розміщені результати опитування експертів (топ-менеджери відомих ІТ-компаній, які працюють в Україні), що працюють в ІТ-бізнесі, який був спрямований на виявлення позитивних та негативних факторів ведення ІТ-бізнесу в регіонах України. На підставі результатів даних опитувань були виділені показники, що характеризують стан ІТ-галузі (2015-2018 рр.): ІТ-кластер, ЗВО та ІТ, перегрів ІТ-ринку, вимивання кадрів, авіасполучення, співпраця ІТ-індустрії й місцевої влади, коворкінг, рівень кваліфікації ІТ-кадрів, комфортність проживання та ведення бізнесу, превалювання в напрямку діяльності, близькість до Єврозони, рівень володіння англійською мовою, актуальність знань, придбаних в ЗВО по ІТ-профілю, конкуренція на ІТ-ринку, близькість до вітчизняних ІТ-центрам, наявність інвесторів, попит на ІТ-кадри, кількість профільних ЗВО, кількість фахівців ІТ-сфери, рівень зарплати [5].

В роботі [5] для кожного показника стану ІТ-галузі були визначені 23 критерії оцінювання, які в свою чергу були розподілені на три категорії: «Інвестиційна привабливість регіонів України», «Внутрішній стан ІТ-ринку України», «Заробітна плата». Вхідні дані за всіма критеріями, відповідно до категорії представлено у табл. 1 – 2.

Багатофакторна оцінка показників, за якими можна оцінити стан розвитку ІТ-галузі, оцінюється кортежем, що враховує інвестиційну привабливість регіонів України ( $F_j^{IN}$ ), внутрішній стан ІТ-ринку України ( $F_j^{VN}$ ) та заробітну плату ( $F_j^{ZP}$ ) [5]:

$$F_j = \langle F_j^{IN}, F_j^{VN}, F_j^{ZP} \rangle, \quad (1)$$



Таблиця 1

## Дані групи «Внутрішній стан ІТ-ринку України» за Рівненським, Полтавським та Херсонським регіонами

Критерій \ Регіон	ІТ-кластер	ЗВО та ІТ	перегрів ІТ-ринку	вимивання кадрів	співпраця ІТ-індустрії та влади	коворкінг	рівень кваліфікації ІТ-кадрів	превалювання в напрямку діяльності	актуальність знань	кількість профільних ЗВО/факультетів	конкуренція на ІТ-ринку	близькість до ІТ-центрів	кількість фахівців ІТ-сфери, люд.	наявність інвесторів	попит на ІТ-кадри
Рівне	+	+	+	+	-	-	середній	А і П	не відповідають вимогам	4	-	+	1000	+	+
Полтава	-	+	-	+	-	-	низький	А	не відповідають вимогам	5	-	+	1000	-	+
Херсон	+	+	-	+	-	-	середній	А і П	не відповідають вимогам	3	-	-	1000	+	+

Таблиця 2

Дані груп «Інвестиційна привабливість регіонів України» та «Заробітна плата, USD» за Рівненським, Полтавським та Херсонським регіонами

Регіон Критерій	Інвестиційна привабливість регіонів України								Заробітна плата, USD		
	авіаспо- лучення	еколо- гія	транс- портна розв'язка	рівень дозвілля (відпо- чинок)	рівень безпе- ки прожи- вання	рівень життя	близь- кість до Євро- зони	рівень воло- діння англій- ською	зарпла- та Junior	зарпла- та Middle	зарпла- та Senior
Рівне	-	сприят- лива	гарна	середній	висо- кий	доступ- ний	-	низький	450	1150	2250
Полтава	-	сприят- лива	гарна	середній	висо- кий	доступ- ний	-	низький	300	1200	-
Херсон	+	сприят- лива	погана	середній	висо- кий	доступ- ний	-	середній	400	700	1900

З урахуванням цього завдання оцінка показників перетворюється в задачу багатокритеріальної оптимізації виду [5]:

$$O_j^o = \arg G[\langle F_j(O) \rangle], \quad (2)$$

де  $G$  – оператор, який визначає структуру моделі;  $\langle F_j(O_j) \rangle$  – тривимірний кортеж, компонентами якого є різнорідні по розмірності, за інтервалами зміни і вимірювальним шкалами величини.

У загальному випадку завдання багатокритеріальної оптимізації є некоректними за Адамаром, тому що не мають однозначної відповіді [6] і вимагають регуляризації [7].

Загальний підхід до регуляризації задач багатокритеріальної оптимізації полягає в їх скаляризації. В даний час існує безліч різних способів регуляризації: принцип головного критерію, послідовна оптимізація, функціонально-вартісний аналіз і таке інше, але найбільш аргументованим є підхід, заснований на теорії корисності [8], який полягає у формуванні узагальненої скалярної оцінки (функції корисності) допустимих рішень. В даному випадку функція корисності має вигляд [5]:

$$P(O_j^o) = Q[\langle \lambda_j \rangle, \langle F_j(O_j) \rangle], \quad j = \overline{1, n}, \quad (3)$$

де  $\langle \lambda_j \rangle$  – кортеж коефіцієнтів ізоморфізму;  $\langle F_j(O_j) \rangle$  – кортеж різнорідних оцінок показників розвитку ІТ-галузі.

Для конструктивного використання моделі (3) необхідно вирішити задачу її структурно-параметричної ідентифікації, тобто визначити структуру (вид оператора  $Q$ ) і параметри (значення коефіцієнтів ізоморфізму  $\lambda_j$ ). Рішення зазначених завдань є евристичними процедурами, заснованими на реалізації методології експертного оцінювання.

В якості узагальненої оцінки ефективності локальної категорії приймають адитивну модель виду [5]:

$$P(O_j) = \sum_{j=1}^3 \lambda_j F_{ji}(O_i), \quad (4)$$

де  $F_{ji}(O_i)$ ,  $j = \overline{1, 3}$  – інвестиційна привабливість регіонів України ( $j=1$ ), внутрішній стан ІТ-ринку України ( $j=2$ ) і заробітна плата ( $j=3$ ) оцінки, виміряні в натуральних показниках;  $\lambda_j$  – коефіцієнти ізоморфізму.

Коефіцієнти ізоморфізму  $\lambda_j$  в моделі (4) виконують дві функції: призводять різнорідні величини до однієї розмірності; враховують різну відносну важливість.

Така подвійність ускладнює експертну параметричну ідентифікацію моделі (4).

Для подолання цих труднощів модель перетворено шляхом приведення її до безрозмірною форми і представлено у вигляді [5]:

$$P(O_j) = \sum_{j=1}^n a_j F_{ji}(O_i), \quad (5)$$

де  $F_{ji}(O_i)$  – значення  $j$ -ої категорії з врахуванням і показників;  $a_j$  – безрозмірні коефіцієнти відносної важливості, що відповідають наступним обмеженням:

$$0 \leq a_j \leq 1, \quad \forall j = \overline{1,3}; \quad \sum_{j=1}^3 a_j = 1. \quad (6)$$

Модель оцінювання  $F_{ji}(O_i)$  має вигляд [9]:

$$F_{ji}(O_i) = \sum_{i=1}^n b_{ji} O_{ji}^H, \quad (7)$$

де  $O_{ji}^H$  – нормалізовані критерії  $j$ -ої категорії;  $b_{ji}$  – безрозмірні коефіцієнти відносної важливості нормалізованих приватних критеріїв. Їх нормалізація проводиться за формулою [5]:

$$O_{ji}^H(x) = \left[ \frac{O_{ji} - O_{ji}^G}{O_{ji}^K - O_{ji}^G} \right], \quad (8)$$

де  $O_{ji}$  – значення критерію;  $O_{ji}^K$ ,  $O_{ji}^G$  – відповідно найкраще і найгірше значення критерію, яке він приймає на області допустимих рішень.

З урахуванням проведеного раніше перетворення функцій, модель оцінки категорій, цільова функція має вигляд [5]:

$$O_j^o = \arg \sum_{j=1}^3 a_j \left\{ \sum_{i=1}^n b_{ji} \left[ \frac{O_{ji} - O_{ji}^G}{O_{ji}^K - O_{ji}^G} \right] \right\}. \quad (9)$$

Таким чином, за допомогою використання принципів побудови моделі прийняття багатокритеріальних рішень в умовах детермінованих вихідних даних, реалізовано етап синтезу методології прийняття рішень в умовах багатокритеріальності і інтервальної невизначеності.

На основі моделі (7) були виконані розрахунки по визначенню оцінки показників стану ІТ-галузі на основі узагальнених даних за опитуванням експертів.

Кожний з критеріїв має свою шкалу оцінювання [5]. Взагалі використовувалися двоохрівневі та трьохрівневі шкали.

Результати розрахунку за Рівненським, Полтавським та Херсонським регіонами подано у табл. 3.

**Таблиця 3**

**Значення показників стану ІТ-галузі за кожною категорією за Рівненським, Полтавським та Херсонським регіонами**

Регіон	Внутрішній стан ІТ-ринку України ( $F_j^{VN}$ )	Інвестиційна привабливість регіонів України ( $F_j^{IN}$ )	Заробітна плата ( $F_j^{ZP}$ )
Вінниця	0,371889401	0,675371224	0,465240642
Миколаїв	0,180832907	0,455709165	0,490813108
Запоріжжя	0,299982932	0,339477727	0,406965583
Дніпро	0,519952210	0,206093190	0,804538599
Одеса	0,413807817	0,473630312	0,941176471
Луцьк	0,274074074	0,444444444	0,431372549
Івано-Франківськ	0,550605905	0,615207373	0,587686823
Житомир	0,344427377	0,283922171	0,506101741
Хмельницький	0,268817204	0,559139785	0,116550117
Харків	0,866666667	0,333333333	0,703962704
Тернопіль	0,474534904	0,611879160	0,304675716
Ужгород	0,368509985	0,780849974	0,094405594
Львів	0,764225977	0,878648233	0,776402029
Черкаси	0,378067930	0,228878648	0,279720280
Чернівці	0,285936167	0,729646697	0,536404772
Суми	0,443659328	0,282642089	0,290004114
Чернігів	0,584400068	0,338197645	0,505635541
Рівне	0,591039427	0,448028674	0,385986562
Полтава	0,298446834	0,448028674	0,177156177
Херсон	0,450298686	0,503584229	0,155354449

За кожною групою показників максимальні та мінімальні значення, що впливає на комплексний показник, не змінилися [5]. Так, за групою «Внутрішній стан ІТ-ринку» мінімальне значення спостерігається у Миколаївському регіоні (0,180833), а максимальне – у Харківському (0,866667). За групою «Інвестиційна привабливість регіонів» максимум належить Львівському регіону (0,878648), а мінімум – Дніпропетровському

регіону (0,206093). За групою «Заробітна плата» максимальне значення за Одеським регіоном (0,9411765), а мінімальне – за Ужгородським (0,0944056).

Визначення комплексного показника стану ІТ-галузі регіону за моделлю (5) має вигляд [5]:

$$P(O_i) = \frac{F_i^{VN} + F_i^{IN} + F_i^{ZP}}{3}, \quad i = \overline{1,20}. \quad (10)$$

Результати розрахунку комплексного показника стану ІТ-галузі за кожним регіоном України представлені в табл. 4.

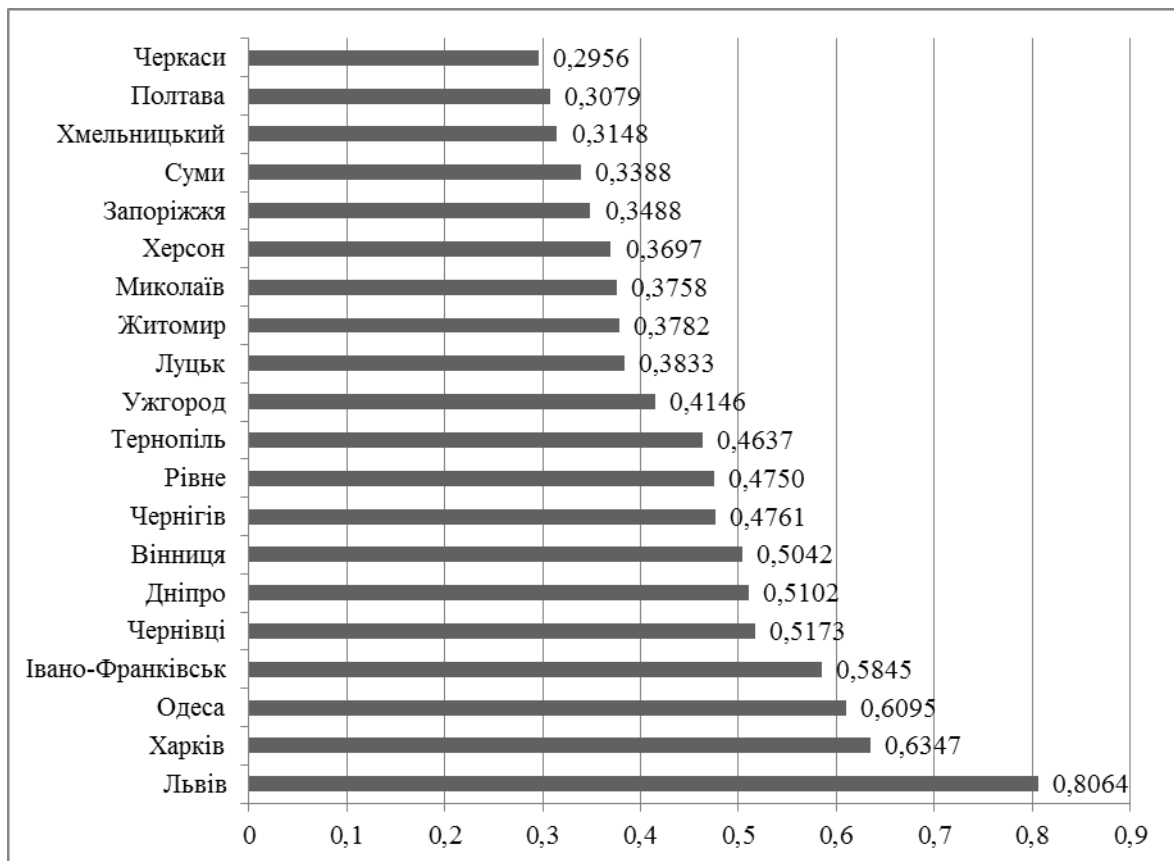
**Таблиця 4**

**Значення комплексного показника стану ІТ-галузі з врахуванням Рівненського, Полтавського та Херсонського регіонів**

Значення показника	Регіон
0,8064254	Львів
0,634654	Харків
0,609538	Одеса
0,584500	Івано-Франківськ
0,517000	Чернівці
0,510190	Дніпро
0,504167	Вінниця
0,476078	Чернігів
0,475018	Рівне
0,463697	Тернопіль
0,414589	Ужгород
0,383297	Луцьк
0,378150	Житомир
0,375785	Миколаїв
0,369746	Херсон
0,348809	Запоріжжя
0,338769	Суми
0,314836	Хмельницький
0,307877	Полтава
0,295556	Черкаси

Із введенням в дослідження додатково трьох регіонів тенденції стосовно найменшого та найбільшого значення комплексного показника стану ІТ-галузі, не змінилися: попереду Львів (0,806425), Харків (0,634654) та Одеса (0,609538), слабкіший розвиток – Хмельницький (0,314836), Полтава (0,307877) та Черкаси (0,295556) (рис. 1).

За результатами сформованої моделі та проведеними за нею розрахунками, необхідно з'ясувати силу зв'язку між обраними факторами та комплексним показником, а також значність впливу факторів на комплексний показник. Таке дослідження було проведено за допомогою кореляційно-регресійного аналізу.



**Рис. 1. Ранжування регіонів України за комплексним показником стану ІТ-галузі**

Розрахуємо показник тісноти зв'язку. Таким показником є вибірковий лінійний коефіцієнт кореляції, який розраховується за формулою [10]:

$$r_{xy} = \frac{\overline{x * y} - \bar{x} * \bar{y}}{s(x) * s(y)} \quad (11)$$

де  $\overline{x * y}$ ,  $\bar{x}$ ,  $\bar{y}$  – середні значення ( $\bar{x} = \frac{\sum x_i}{n}$ ,  $\bar{y} = \frac{\sum y_i}{n}$ ,  $\overline{x * y} = \frac{\sum (x_i * y_i)}{n}$ );

$s(x)$ ,  $s(y)$  – середньоквадратичні відхилення

$$(s(x) = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}, s(y) = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}).$$

Коефіцієнти парної кореляції між факторами «Внутрішній стан ІТ-ринку України» ( $F_j^{VN}$ ), «Інвестиційна привабливість регіонів України» ( $F_j^{IN}$ ), «Заробітна плата» ( $F_j^{ZP}$ ) та комплексним показником стану ІТ-галузі регіону та їх інтерпретація наведені в табл. 6.

**Результати розрахунку коефіцієнтів парної кореляції за двадцятьма регіонами України**

Назва показника	Значення коефіцієнта кореляції	Оцінка значення коефіцієнта кореляції за шкалою Чеддока [10]
Внутрішній стан ІТ-ринку України	0,726	сильна та пряма
Інвестиційна привабливість регіонів України	0,459	помірна та пряма
Заробітна плата	0,769	сильна та пряма

Тісноту спільного впливу чинників на результат оцінює індекс множинної кореляції. На відміну від парного коефіцієнта кореляції, який може приймати негативні значення, він приймає значення від 0 до 1 [10]. Чим щільніше фактичні значення  $y_i$  розташовуються щодо лінії регресії, тим менше залишкова дисперсія  $i$ , отже, більше величина  $R_y(x_1, \dots, x_m)$ . Таким чином, при значенні  $R$  близькому до 1, рівняння регресії краще описує фактичні дані і фактори сильніше впливають на результат. При значенні  $R$  близькому до 0 рівняння регресії погано описує фактичні дані і фактори чинять слабкий вплив на результат [10]:

$$R = \sqrt{1 - \frac{S_e^2}{\sum (y_i - \bar{y})^2}}. \quad (12)$$

Коефіцієнт множинної кореляції  $R = \sqrt{1 - \frac{0}{0,792}} = 1$ , отже зв'язок між ознакою  $Y$  та факторами  $X_i$  сильний.

Рівняння множинної регресії наступне:  
 $Y = 0,000021 + 0,3335X_1 + 0,3332X_2 + 0,3332X_3$ . Можлива економічна інтерпретація параметрів моделі така: збільшення  $X_1$  на 1 од. вим. призводить до збільшення  $Y$  в середньому на 0,3335 од. вим.; збільшення за  $X_2$  та  $X_3$  на 1 од. вим. призводить до збільшення  $Y$  в середньому на 0,3332 од. вим. за кожним фактором.



Перейдемо до статистичного аналізу отриманого рівняння регресії: перевіріть значимості рівняння і його коефіцієнтів. Об'єктивною оцінкою є скоригований коефіцієнт детермінації [10]:

$$\bar{R}^2 = 1 - (1 - R^2) * \frac{n - 1}{n - m - 1} \quad (13)$$

Чим ближче цей коефіцієнт до одиниці, тим більше рівняння регресії пояснює поведінку  $Y$ . В нашому випадку він дорівнює 1. За максимальним коефіцієнтом  $\beta_3=0,601$  можна зробити висновок, що найбільший вплив на результат  $Y$  оказує фактор  $X_3$ .

Також необхідно перевірити побудовану модель на адекватність реальному процесу шляхом порівняння даних, розрахованих за моделлю і визначених експериментально.

Для перевірки адекватності моделей обраний критерій Фішера [10]:

$$F_e = \frac{S_{ad}^2}{S_e^2}, \quad (14)$$

що є відношенням двох дисперсій – відтворюваності та адекватності.

Дисперсія відтворюваності [10]:

$$S_e^2 = \frac{\sum_{i=1}^n S_{ni}^2}{n}, \quad (15)$$

де  $S_{ni}^2$  – середньоквадратичне відхилення при вимірюванні  $i$ -го результату.

Розсіювання експериментальних точок щодо розрахункових характеризується залишковою дисперсією або дисперсією адекватності [11]

$$S_{ad}^2 = \frac{\sum_{i=1}^m (y_i - y_{pi})^2}{m - l - 1}, \quad (16)$$

де  $y_i, y_{pi}$  – розрахункові та експериментальні значення;  $m$  – число порівнюваних значень;  $l$  – число коефіцієнтів, визначених за дослідними даними.

Теоретична залежність покладається адекватною дослідній, якщо отримане експериментальне значення критерію Фішера  $F_e$  менше табличного  $F_{табл}$ , обраного за кількістю ступенів свободи чисельника і знаменника, а також довірчої ймовірності  $\alpha$ :

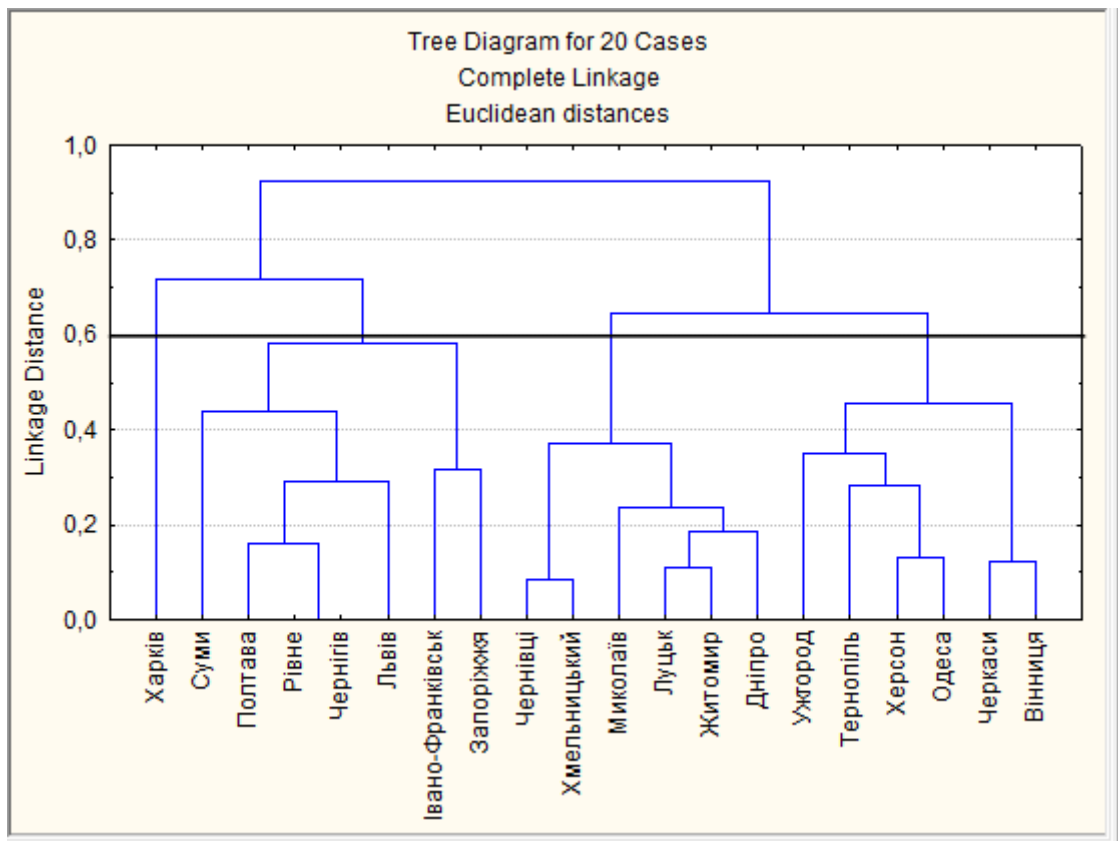
$$F_e < F_{табл}. \quad (17)$$

В даному випадку табличне значення при ступенях свободи  $k_1 = 3$  та  $k_2 = n - m - 1 = 20 - 3 - 1 = 16$ ,  $F_{\text{табл}}(3;16) = 3,24$ , а експериментальне –  $F_e = 1,25$ .

Експериментальне значення критерію Фішера не перевищує табличного значення. Тому розроблену модель можна вважати адекватною і покласти в основу подальших досліджень.

Для того, щоб з'ясувати які з регіонів за своїми характеристиками (факторами) близькі один до одного, а які ні, використано кластерний аналіз, в якому використовується деяка міра подібності або відстань між об'єктами. Кластерний аналіз проводили в пакеті STATISTICA [12].

Типовим результатом такої кластеризації є ієрархічне дерево. Процес вибору кількості кластерів найбільш ефективно відображає процедура ієрархічної класифікації об'єктів (рис. 2).



**Рис. 2. Дендродіаграма ієрархічної агломеративної кластеризації регіонів за комплексним показником стану ІТ-галузі**

На даній діаграмі горизонтальна ось являє собою спостереження, вертикальна ось представляє відстань об'єднання. Для кожного вузла в графі (там, де формується новий кластер) представлена величина відстані, для якого відповідні фактори зв'язуються в новий єдиний кластер [12]. Коли дані

мають ясну "структуру" в термінах кластерів об'єктів, схожих між собою, тоді ця структура, швидше за все, повинна бути відображена в ієрархічному дереві різними гілками. В результаті успішного аналізу методом об'єднання з'являється можливість виявити кластери (гілки) і інтерпретувати їх.

Метод деревовидної кластеризації було використано при формуванні кластерів несходства або відстані між об'єктами. Найбільш прямий шлях обчислення відстаней між об'єктами в багатовимірному просторі полягає в обчисленні евклідових відстаней [12]:

$$d_{Eij} = \left( \sum_{l=1}^m (x_i^l - x_j^l)^2 \right)^{\frac{1}{2}}. \quad (18)$$

Правило, за яким об'єднувались кластери – одиночний зв'язок (відстань «Найближчого сусіда») [12]:

$$\rho_{\min}(K_i, K_j) = \min_{x_i \in K_i, x_j \in K_j} \rho(x_i, x_j) \quad (19)$$

З рис. 2 видно, що кількість кластерів при 60% розрізі дорівнює трьом: перший кластер включає такі об'єкти, як Харків, Суми, Полтава, Рівне, Чернігів, Львів, Івано-Франківськ, Запоріжжя; до другого кластеру входять Ужгород, Тернопіль, Херсон, Одеса, Черкаси, Вінниця; до третього кластеру – Чернівці, Хмельницький, Миколаїв, Луцьк, Житомир, Дніпро.

Для уточнення щодо отриманого розподілу регіонів за кластерами при проведенні дослідження також було використано метод k-середніх (k-means clustering), який відрізняється від ієрархічних агломеративних методів тим, що вже до проведення розрахунків відомо кількість кластерів (3 в нашому випадку) та ці кластери розташовані на великій відстані один від одного (рис. 3).

Members of Cluster Number 1 and Distances from Respective Cluster contains 7 variables		Members of Cluster Number 2 and Distances from Respective Cluster contains 5 variables		Members of Cluster Number 3 and Distances from Respective Cluster contains 8 variables	
Variable	Distance	Variable	Distance	Variable	Distance
Запоріжжя	0,166924	Вінниця	0,094008	Дніпро	0,130886
Івано-Франківськ	0,185370	Одеса	0,135125	Житомир	0,039378
Львів	0,115235	Тернопіль	0,077643	Луцьк	0,072346
Суми	0,163105	Ужгород	0,136502	Миколаїв	0,091464
Харків	0,264853	Черкаси	0,144089	Хмельницький	0,096540
Чернігів	0,126594			Чернівці	0,085837
Рівне	0,126594			Полтава	0,146465
				Херсон	0,125221

**Рис. 3. Розподілення регіонів за групами методом k-середніх за комплексним показником стану ІТ-галузі**

За результатами проведеного аналізу даним методом розподіл регіонів за кластерами наступний: перший кластер – Харків, Суми, Рівне, Чернігів, Львів, Івано-Франківськ, Запоріжжя; другий кластер – Ужгород, Тернопіль, Одеса, Черкаси, Вінниця; третій кластер – Чернівці, Хмельницький, Миколаїв, Луцьк, Житомир, Дніпро, Полтава, Херсон.

З представленою аналізу видно, що до складу сукупності, що аналізували, входили чотири великі ІТ-центри України (Харків, Львів, Дніпро, Одеса) з п'яти (за Києвом інформація відсутня). Не всі кластери за двома методами кластерного аналізу включають однаковий склад регіонів, але розподіл флагманів за кластерами збігається у обох випадках, що дає привід стверджувати про наявність трьох шляхів розвитку ІТ-індустрії в Україні та обґрунтовує подальше вивчення цього питання.

### **Висновок**

В ході дослідження було проведено опитування експертів за додатковими трьома регіонами, оцінено за цими регіонами кожний з показників стану ІТ-галузі та проведено розрахунки за моделлю оцінки регіонів згідно з критеріями оцінювання. За допомогою кореляційно-регресійного аналізу з'ясовано силу зв'язку між обраними факторами та комплексним показником і виявлено значність впливу факторів (показників) на комплексний показник з урахуванням внесених змін в набір даних. Отримані результати дозволили стверджувати, що фактори, які входять до моделі, мають значущий вплив на комплексний показник та однозначно визначають його стан навіть з врахуванням внесених змін. Також було визначено, що й при доданих показниках дана модель за критерієм Фішера є адекватною. Для розподілу регіонів за розрахованими комплексними показниками стану ІТ-галузі було використано кластерного аналізу у межах математичного пакету STATISTICA. Отримані результати підтверджують гіпотезу про наявність трьох шляхів розвитку ІТ-індустрії в Україні та надають підґрунтя для проведення додаткових досліджень.

### **Список літератури**

1. Vynnychuk, R. O. & Skljарuk, T. V. (2015) Особливості розвитку ІТ-ринку в Україні: стан та тенденції [*Osoblyvosti rozvytku IT-rynku v Ukraini: stan ta tendenciji*]. Visnyk Nacionaljnogho universytetu "Ljvivsjka politekhnika". Loghistryk, .833, 3-8.
2. Rubin, Je. (2018) Перемога под елочку: ІТ-кластер координує програму облсовета «ІТ-Харьковщина» [*Peremoga pod elochku: IT-klaster*

*koordiniruet programmu oblsoveta «IT- Kharjkovshhyna»*]. Retrieved from <https://dou.ua/lenta/columns/it-kharkiv/>.

3. Stremetsjka, Daryna. (2017) Регулювання і перспективи ІТ-ринку України [*Reghuljuvannja i perspektyvu IT-rynku Ukrainy*]. Retrieved from <https://dou.ua/forums/topic/22075/>.

4. Grabovskaja, Anna. (2019) Токсичный HR: действия, которые отравляют команду. [*Toksichnyj HR: dejstvija, kotorye otravljajut komandu*]. Retrieved from <https://dou.ua/lenta/articles/toxic-hr/>.

5. Brynza, N. A. & Gavrilova, A. A. (2018) *Многофакторная оценка показателей развития IT-отрасли в регионах Украины* [*Mnogofaktornaja ocenka pokazatelej razvitija IT-otrasli v regionah Ukrainy*]. Zbirnyk naukovykh pracj Kharkivskogo nacionaljnogo universytetu Povitrjanykh Syl, 2(56), 159-169.

6. Podinovskiy, V.V.& Nogin, V.D. (Ed.) (1982). *Парето-оптимальные решения многокритериальных задач* [*Pareto-optimalnye reshenija mnogokriterial'nyh zadach*]. (254 p.). Moscow: Nauka.

7. Tikhonov, A.N. & Arsenin, V.Ya. (Ed.) (1986). *Методы решения некорректных задач* [*Metody reshenija nekorrektnyh zadach*] (286 p.). Moscow: Nauka.

8. Fishburn, P. (Ed.) (1978). *Теория полезности для принятия решений* [*Teorija poleznosti dlja prinjatija reshenij*] (352 p.). Moscow: Nauka.

9. Kuzmin, I.V. & Petrov, E.G. & Alferov, I.A. & Evseev, V.V. & Migunova, L.V. (Ed.) (1978). *Автоматизированные системы управления городским хозяйством* [*Avtomatizirovannye sistemy upravlenija gorodskim hozjajstvom*] (144 p.). Kyiv, Budiveljnyk.

10. Aliev, T. A. (1991) *Экспериментальный анализ* [*Jeksperimental'nyj analiz*] (272 p.). Moscow: Mashinostroenie.

11. Tavanjuk, T.Ja. & Nikolaenko, A.P. & Romanchenko, A.V. & Shumakova T.A. (2017) Оценка адекватности математических моделей характеристик электрогидравлических следящих приводов [*Ocenka adekvatnosti matematicheskikh modelej harakteristik jelektrogidravlicheskih sledjashhih privodov*]/. Visnyk skhidnoukrajinskogo nacionaljnogo universytetu imeni Volodymyra Dalja, 7 (237), 89-94.

12. Буреева Н. Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007, 112 с.

## ГЛАВА 2

### ТЕХНОЛОГІЯ СТВОРЕННЯ СПІЛЬНОГО КОНТЕНТУ В ВЕБ- БАЗОВАНОМУ НАВЧАЛЬНОМУ КАБІНЕТІ

#### **Вступ і постановка задачі**

З ростом технологічних можливостей Інтернету як середовища для комунікації та здійснення освітнього процесу з'явилися умови для передачі різного роду інформації на будь-які відстані і за короткі проміжки часу. Через глобальну комп'ютерну мережу Інтернет можливий миттєвий доступ до світових інформаційних ресурсів (електронних бібліотек, баз даних, сховищ файлів, тощо). Стверджується розуміння, що Інтернет – це середовище для ведення комунікації, що має свою специфіку, пов'язану з отриманням вигоди [1].

На сьогодні багато закладів вищої освіти усвідомлюють важливість і необхідність ресурсів для створення спільного контенту у зв'язку з інформатизацією освіти. Освітняни починають розуміти, що Інтернет являє собою нову сферу освітнього процесу, що доносить інформацію користувачам 24 години на добу і до неї залучаються все більше людей. Максимальне наближення освітнього процесу до студентів та викладачів за допомогою реурсів мережі Інтернет дозволяє здійснити веб-базований навчальний кабінет для формування спільного контенту.

Сьогодні часто можна побачити веб-базовані навчальні кабінети, які не виконують своїх функцій, незручні у використанні або мають невідповідний дизайн. Брак знань і умінь в області створення навчальних веб-базованих кабінетів призводить до великої кількості помилок, починаючи з характеру інформації, колірному рішенню та фірмового стилю сайту і закінчуючи відсутністю зворотного зв'язку з цільовою аудиторією сайту [2-5]. Це призводить до втрати одного з унікальних каналів комунікацій, що сприяє втраті інтересу до навчальних дисциплін. Проблема недооцінки навчального кабінету як каналу комунікації може виявитися згубною для більшості навчальних закладів. Таким чином, у науковій та практичній площині виявляється актуальною проблема створення ефективного навчального веб-базованого кабінету для формування спільного контенту.

Метою даного розділу є розробка і реалізація технології спільного формування контенту в веб-базованому навчальному кабінеті.

## Основна частина

Визначення основних елементів веб-базованого навчального кабінету здійснимо на основі використання методу експертних оцінок.

Першим етапом експертного опитування є розробка основних параметрів оцінки веб-базованого навчального кабінету. Спочатку доцільно проводити групову дискусію, в ході якої експерти виділяють найбільш інформативні характеристики продукції.

Перший етап. Використання методу експертних оцінок для визначення елементів які необхідно використовувати при створенні веб-базованого кабінету для створення спільного контенту. У якості експертів були обрані 19 студентів кафедри комп'ютерних систем і технологій ХНЕУ ім. С. Кузнеця та 2 незалежних респондента зі сфери бізнес-діяльності.

В ході групової дискусії було виділено 20 найбільш інформативних для них елементів веб-базованих навчальних кабінетів, зрозумілих для користувача:

- 1) кабінет користувача;
- 2) календар;
- 3) сховище файлів;
- 4) постановка цілей;
- 5) налаштування параметрів;
- 6) інформація про користувачів;
- 7) ресурси для контенту;
- 8) коментування проектів (звітів);
- 9) блог;
- 10) корисні посилання;
- 11) новини;
- 12) поштові повідомлення;
- 13) структура ресурсу (меню);
- 14) зворотній зв'язок;
- 15) перегляд завдань;
- 16) реєстрація;
- 17) нотатки;
- 18) комунікативні ресурси
- 19) звіти;
- 20) форум запитань.

Наступним кроком експерти надали кожному із 20 елементів відповідний ранг, в залежності від ступеню його важливості. Ранжування

проводилося за наступною схемою: найбільш важливому елементу присвоюється значення рангу 1, наступний за значимістю – ранг 2, і так далі до 20. Якщо декілька характеристик за поглядом респондента рівноцінні, то їм надаються однакові, так звані пов’язані, ранги.

За даними обчислень отримано коефіцієнт конкордації:

$$W = \frac{S}{\left( \frac{1}{12} m p^2 (m^2 - 1) - p \sum_{j=1}^p T_j \right)} =$$

$$= \frac{165239,46}{\left( \frac{1}{12} * 20 * 21^2 * (20^2 - 1) - 21 * 13,5 \right)} = 0,864$$

Показник  $W = 0,864$  наближений до 1, що вказує на узгодженість оцінок експертів. Отже, з врахуванням високого ступеня узгодженості думок експертів на основі сумарних оцінок рангів для кожного елементу є можливість відібрати за найменшими значеннями 10 основних елементів, які потрібно використовувати при створенні веб-базованого навчального кабінету. Результати ранжування наведено в табл. 1.

**Таблиця 1**

**Результати ранжування елементів веб-ресурсу (експертна оцінка)**

Елемент (m) № респондента	1	2	3	4	5	6	7	8	9	10
1	1	7	5	9	13	18	3	15	11	1
2	2	7	3	10	14	15	3	16	6	2
3	1	6	3	8	11	14	3	13	8	2
4	1	9	2	10	11	17	2	16	6	3
5	1	10	4	9	13	16	3	16	8	2
6	2	7	5	9	12	13	4	18	6	1
7	1	3	1	7	15	17	3	16	8	2
8	1	7	4	8	14	16	2	15	10	1
9	1	8	3	7	13	15	4	14	11	2
10	1	8	5	9	13	16	3	14	9	2
11	1	8	2	7	15	14	3	17	5	2
12	2	7	3	6	14	15	5	13	8	1
13	1	3	1	7	15	17	3	16	8	2
14	1	7	5	9	13	18	3	15	11	1
15	2	7	3	10	14	15	3	16	6	2
16	1	8	3	7	13	15	4	14	11	2
17	1	8	5	9	13	16	3	14	9	2
18	1	8	2	7	15	14	3	17	5	2
19	2	7	3	6	14	15	5	13	8	1
20	1	3	1	7	15	17	3	16	8	2
21	1	7	4	8	14	16	2	15	10	1
Сума	26	138	62	160	271	311	64	304	161	35



Закінчення табл. 1

Елемент (m) № респондента	11	12	13	14	15	16	17	18	19	20
1	4	8	2	6	12	4	14	16	17	10
2	4	5	1	8	9	6	11	12	13	17
3	4	7	4	6	10	5	9	12	15	16
4	4	8	5	7	10	6	12	13	14	15
5	5	9	7	11	12	6	15	14	18	17
6	3	8	2	10	11	4	14	17	16	15
7	5	6	4	11	9	10	14	13	12	18
8	3	6	5	12	7	9	13	11	17	18
9	5	6	3	9	12	6	17	10	16	18
10	4	7	6	10	11	7	12	15	17	18
11	4	9	3	11	10	6	13	12	16	18
12	4	9	4	10	11	6	16	12	17	18
13	5	6	4	11	9	10	14	13	12	18
14	4	8	2	6	12	4	14	16	17	10
15	4	5	1	8	9	6	11	12	13	17
16	5	6	3	9	12	6	17	10	16	18
17	4	7	6	10	11	7	12	15	17	18
18	4	9	3	11	10	6	13	12	16	18
19	4	9	4	10	11	6	16	12	17	18
20	5	6	4	11	9	10	14	13	12	18
21	3	6	5	12	7	9	13	11	17	18
Сума	83	142	76	193	202	135	270	255	308	341

Як свідчать результати дослідження методом експертних оцінок до веб-ресурсу доцільно включити елементи 1, 2, 3, 7, 9, 10, 11, 12, 13, 16, а саме: кабінет користувача, календар, сховище файлів, блог, корисні посилання, новини за дисципліною (оголошення), поштові повідомлення, структура ресурсу (меню), ресурси для створення контенту та реєстрацію до навчального ресурсу. Для встановлення міри узгодженості використано коефіцієнт конкордації, розраховано середнє значення сумарних рангів, відхилення від середнього значення сумарних рангів, квадратичне відхилення від середнього значення сумарних рангів і сумарне квадратичне відхилення (табл. 2).

Таблиця 2

## Матриця рангів важливості елементів веб-ресурсу

Елемент (m) № респондента	1	2	3	4	5	6	7
1	1,5	9	7	11	15	20	4
2	2,5	10	4,5	13	17	18	4,5
3	1	8,5	3,5	11,5	15	18	3,5
4	1	11	2,5	12,5	14	20	2,5
5	1	11	4	9,5	14	17,5	3
6	2,5	9	7	11	14	15	5,5
7	1,5	4,5	1,5	9	17	19	4,5
8	1,5	8,5	5	10	16	18	3
9	1	10	3,5	9	15	17	5
10	1	9	5	10,5	15	18	3
11	1	10	2,5	9	17	16	4,5
12	2	9	3	7,5	16	17	6
13	1	11	4	9,5	14	17,5	3
14	2,5	9	7	11	14	15	5,5
15	1,5	4,5	1,5	9	17	19	4,5
16	1,5	9	7	11	15	20	4
17	2,5	10	4,5	13	17	18	4,5
18	1	9	5	10,5	15	18	3
19	1	10	2,5	9	17	16	4,5
20	2	9	3	7,5	16	17	6
21	1	9	5	10,5	15	18	3
Сума ранга (а)	17,5	109,5	49	133,5	185	213,5	49
Середнє значення	126,4						
Відхилення	-108,9	-16,9	-77,4	7,1	58,6	87,1	-77,4
Квадратичне відхилення	11859,2	285,61	5990	50,41	3433,96	7586,41	5990,76
Сумарне квадратичне відхилення	165239,46						

Елемент (m) № респондента	8	9	10	11	12	13	14
1	17	13	1,5	5,5	10	3	8
2	19	8,5	2,5	6	7	1	11
3	17	11,5	2	5,5	10	5,5	8,5
4	19	7,5	4	5	10	6	9
5	17,5	8	2	5	9,5	7	12
6	20	8	1	4	10	2,5	12
7	18	10	3	7	8	6	13
8	17	12	1,5	4	7	6	14
9	16	13	2	6	7,5	3,5	11
10	16	10,5	2	4	7,5	6	12
11	19	7	2,5	6	11	4,5	13
12	15	10	1	4,5	11	4,5	12
13	17,5	8	2	2,5	9	7	11
14	20	8	1	1,5	4,5	1,5	9
15	18	10	3	1,5	9	7	11
16	17	13	1,5	2,5	10	4,5	13
17	19	8,5	2,5	4	7	6	14
18	16	10,5	2	6	7,5	3,5	11
19	19	7	2,5	4	7,5	6	12
20	15	10	1	5	9,5	7	12
21	16	10,5	2	4	10	2,5	12
Сума ранга (a)	210,5	119	25	62,5	108,5	55	135,5
Середнє значення	126,4						
Відхилення	-84,1	-7,4	-101,4	-63,9	-17,9	-77,4	9,1
Квадратичне відхилення	7072,81	54,76	10281	8083,21	720,4	4990,76	92,81
Сумарне квадратичне відхилення	165239,46						

Елемент (m) № респондента	15	16	17	18	19	20	T
1	14	5,5	16	18	19	12	1
2	12	8,5	14	15	16	20	1,5
3	14	7	13	16	19	20	1,5
4	12,5	7,5	15	16	17	18	1,5
5	13	6	16	15	20	19	1
6	13	5,5	16	19	18	17	1
7	11	12	16	15	14	20	1
8	8,5	11	15	13	19	20	1
9	14	7,5	19	12	18	20	1
10	13	7,5	14	17	18	20	1
11	12	8	15	14	18	20	1
12	13	7,5	18	14	19	20	1
13	14	15	5,5	20	8	1	2,5
14	17	19	4,5	18	10	3	1,5
15	15	20	4	17	13	1,5	1,5
16	17	18	4,5	19	8,5	2,5	2,5
17	8,5	11	15	13	19	20	1
18	14	7,5	19	12	18	20	1
19	13	7,5	14	17	18	20	1
20	13	6	16	15	20	19	1
21	13	5,5	16	19	18	17	1
Сума ранга (a)	150	93,5	187	184	216	226	13
Середнє значення	126,4						
Відхилення	9,1	23,6	32,9	60,6	57,6	89,6	99,6
Квадратичне відхилення	92,81	116,96	1082,41	672,36	3317,76	8028,16	9920,16
Сумарне квадратичне відхилення	165239,46						

Для встановлення міри достовірності експертних оцінок доцільно використовувати коефіцієнт конкордації. Для цього розраховано середнє значення сумарних рангів, відхилення від середнього значення, квадратичне відхилення від середнього значення і сумарне відхилення.

Згідно отриманих оцінок було визначено елементи, що доцільно включати до веб-базованого навчального кабінету. До веб-ресурсу доцільно включити наступні елементи:

- 1) кабінет користувача;
- 2) календар;
- 3) сховище файлів;
- 4) блог;
- 5) корисні посилання;
- 6) новини за дисципліною (оголошення);
- 7) поштові повідомлення;
- 8) структура ресурсу (меню);
- 9) ресурси для створення контенту;
- 10) реєстрація.

Таким чином, запропонована технологія розробки веб-базованого навчального кабінету дозволяє обрати найбільш значимі елементи до структури веб-ресурсу.

Наступним етапом розробки продукту є технологія формування спільного контенту.

Технологія формування контенту в веб-базованому навчальному кабінеті формується шляхом створення контенту в кабінеті з раніше обраними характеристиками.

Так як умови групового формування контенту не задовольняють попередні педагогічні моделі, необхідно запропонувати модель, яка б передбачала обмін знаннями між викладачем та групами студентів.

Необхідно передбачити те, що викладач контролює не весь навчальний процес студентів за дисципліною, а лише підключається на окремих ітераціях при проведенні лекційних занять, підготовці самостійного матеріалу та його презентації перед аудиторією, також при закріпленні матеріалу.

Групова навчальна діяльність, на відміну від фронтальної та індивідуальної, не ізолює студентів один від одного, а навпаки, дає змогу реалізувати природне прагнення до спілкування, взаємодопомоги та співпраці. У цьому випадку, студентам набагато простіше звертатись за поясненням один до одного, ніж до викладача.

Така навчальна діяльність сприяє активізації і результативності навчання, самостійності, умінню доводити і відстоювати свою думку, а також прислухатися до думки товаришів, культурі ведення діалогу, відповідальності за результати своєї праці.

Таким чином під технологією формування спільного контенту в веб-базованому навчальному кабінеті буде розумітися алгоритм створення контенту (ланцюг, за яким будується контент в розроблюваному ресурсі).

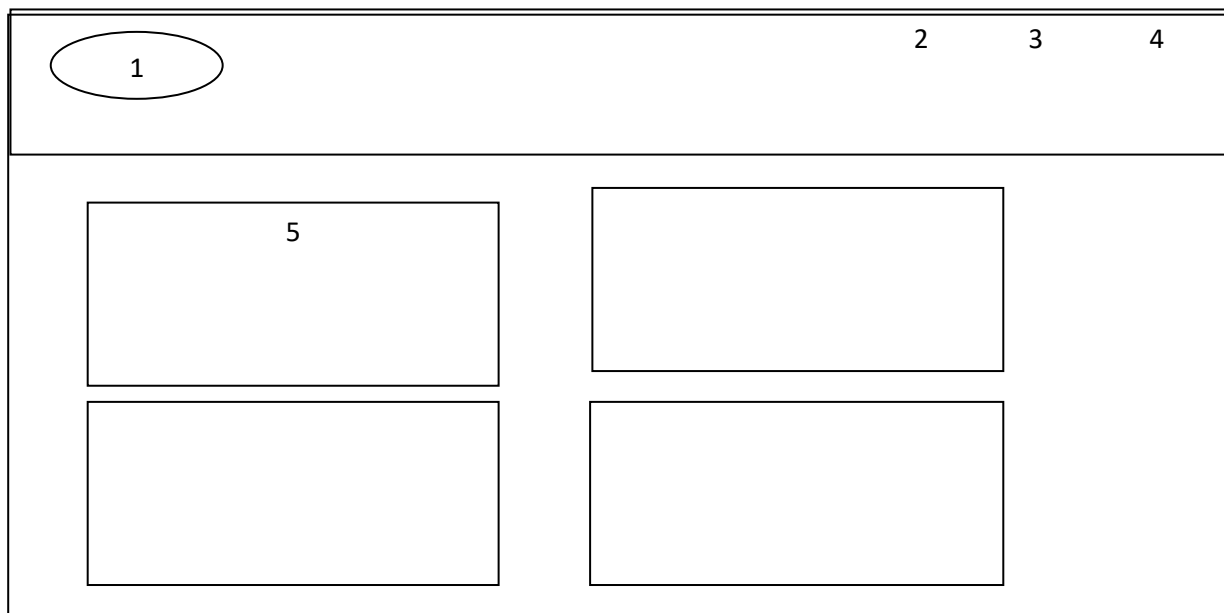
Розробка зовнішньої структури веб-базованого кабінету тісно пов'язана з технічним створенням сайту, з технічними можливостями системи, за допомогою якої створюється розроблюваний ресурс.

Зовнішня структура сайту буде складатися з таких елементів (рис. 1):

- 1) назва ресурсу;
- 2) меню головної сторінки;
- 3) меню допоміжної інформації дисципліни;
- 4) перехід до практичної частини дисципліни (кабінет);
- 5) контент за дисципліною (анотація, план дисципліни та інші матеріали).

Меню допоміжної інформації дисципліни містить теоретичні аспекти за дисципліною, завдання до лабораторних робіт та перелік програмних рішень, які входять до навчального кабінету.

Під переходом до практичної частини дисципліни розуміється перехід до навчального класу (матеріалів лабораторних та практичних робіт) та ресурсів, за якими можна створювати контент.



**Рис. 1. Зовнішня структура веб-ресурсу**

Структура сайту повинна визначатися на перших етапах створення проекту до початку розробки дизайну. Грамотна структура сайту дозволяє не тільки підвищити популярність ресурсу серед користувачів, але і сприяє просуванню сайту в пошукових системах.

Технології формування веб-базованого навчального кабінету можна поділити на три типи складності.

Технологія створення веб-базованого навчального кабінету за допомогою технології типу 1 складається з наступних етапів:

1) пошук та підготовка матеріалів, які будуть опубліковані в веб-ресурсі (робота з викладачем):

а) підготовка тексту;

б) пошук та підготовка зображень.

2) вибір ресурсу. На цьому етапі обирається ресурс, на якому буде розміщено веб-базований навчальний кабінет (або курс дисципліни) для подальшої роботи з ним;

3) вибір шаблонів та параметрів оформлення. На цьому етапі вибирається варіант оформлення, розмір сторінок та полів, штифти. Можна задати критерії захисту від копіювання;

4) імпорт текстової та ілюстративної інформації. На цьому етапі вся підготовлена інформація об'єднується в одне ціле;

5) компілювання або збереження ресурсу. На цьому етапі весь контент поєднується з обраними налаштуваннями для створення єдиного оформлення для веб-кабінету.

Як висновок, можна сказати, що дана технологія створення навчального веб-кабінету дуже проста у використанні та дозволяє швидко виконувати поставлені завдання. Отже, для створення однотипних початкових курсів, які не потребують ексклюзивного оформлення та створення спільного контенту можна використовувати дану технологію розробки базованого навчального кабінету, побудувавши його на вже готових ресурсах навчального типу.

Технологія створення веб-базованих навчальних кабінетів на основі технології типу 2 більш детальна та складніша за попередню, через це можливо виконати набагато більше різноманітних операцій та отримати більш якісний результат. При використанні цієї технології можна застосовувати вбудовані скрипти для автоматизації процесу верстання, або майстри сторінок, якщо їх дуже багато, як в моєму випадку, тим самим полегшити роботу над веб-ресурсом.

Для створення веб-базованого навчального кабінету потрібно пройти наступні етапи:

1) підготовка контенту:

а) текстової інформації;

- б) підготовка зображень;
- в) створення або редагування модулів для обробки відео;
- г) створення або редагування модулів для обробки аудіо;
- 2) вибір платформи. На цьому етапі визначається яким чином (за допомогою якої платформи) буде будуватися навчальний кабінет;
- 3) створення макету ресурсу:
  - а) вибір теми та основного оформлення кабінету;
  - б) створення стилів для тексту;
  - в) створення та розміщення елементів управління;
- 4) розміщення контенту:
  - а) розташування тексту;
  - б) розташування зображень;
  - в) розташування відео;
- 5) верстання. На цьому етапі вирівнюється та компонуються текст;
  - б) наповнення:
    - а) створення гіперпосилань;
    - б) створення випадаючого меню;
    - в) створення списку адміністраторів веб-базованого навчального кабінету;
    - г) створення аккаунтів закріплених за ресурсом (почтові повідомлення, редактори відео, аудіо);
- 7) публікація ресурсу. В основному ресурси будуються на обмежених серверах, а потім публікуються до загальної мережі;
- 8) тестування та виправлення помилок.

За допомогою такої технології можна розробити веб-базований навчальний кабінет як для користування на комп'ютері, так і для користування на мобільних пристроях завдяки адаптивному дизайну. Один із суттєвих недоліків цієї технології – складність та масштабність, бо для її використання потрібно набагато більше знань та навичок, ніж для попередньої, адже є потреба в різних видах програмного забезпечення тощо.

Технологія створення веб-базованих навчальних кабінетів на основі технології типу 3 складається з наступних етапів:

- 1) підготовка контенту:
  - а) текстової інформації;
  - б) підготовка зображень;
  - в) створення або редагування модулів для обробки відео;
  - г) створення або редагування модулів для обробки аудіо;



2) вибір мови. На цьому етапі необхідно обрати за допомогою якої мови програмування буде створюватися веб-ресурс;

3) створення кабінету. За допомогою обраних мов програмування створюється макет веб-базованого навчального кабінету;

4) розміщення контенту. На цьому етапі підготовлений раніше кабінет розміщується на розроблюваному кабінеті;

5) тестування та виправлення помилок.

Кожна технологія використовується з різною метою, тому має певні переваги та недоліки, таким чином:

1) технологія першого типу для полегшеного користування. За її допомогою можна об'єднати текстову та графічну інформацію для розміщення. На жаль, за допомогою цієї технології створити ресурс з авторськими дизайном елементів управління – дуже складно;

2) технологія другого типу потребує більше знань та умінь, але в результаті можна отримати унікальний ресурс, яким можна користуватися на різних девайсах. Ця технологія дозволяє використовувати різні способи автоматизації для спрощення процесу проектування веб-базованого навчального кабінету;

3) технологія третього типу розрахована на людей, які добре технічно підготовані. При виконанні веб-ресурсу за допомогою даної технології необхідно володіти мовами програмування. Також ця технологія не є досконалою, бо не містить детальних етапів підготовки контенту, що збільшує час розробки видання.

Педагогічна модель формування спільного контенту має складатися з наступних аспектів.

По-перше, зареєструвавшись студент має змогу обрати тему за дисципліною з переліку тем згідно календарного плану та доступ до форуму обговорення для того, щоб теми не повторювались.

Наступним етапом є збір необхідних матеріалів в мережі Інтернет (графічні матеріали, текст, аудіо- та відеоматеріали) для підготовки презентації та доповіді.

Третім етапом є компоновка та обробка матеріалу за допомогою ресурсів кабінету. Під цим розуміється редагування матеріалів, які були знайдено шляхом пошуку в мережі Інтернет (етап 2). Веб-базований кабінет надає змогу імпортувати та редагувати зображення для презентації завдяки імпортованому зовнішньому ресурсу. Також користувачу надається

можливість редагування відеоматеріалів з простору YouTube, а саме обрізка відео.

Четвертим етапом формування презентації засобами веб-базованого навчального кабінету є обговорення матеріалу з відповідальною групою для подальшої підготовки матеріалу. Обговорення також може бути включено до першого етапу, вибору теми, задля розподілу обов'язків.

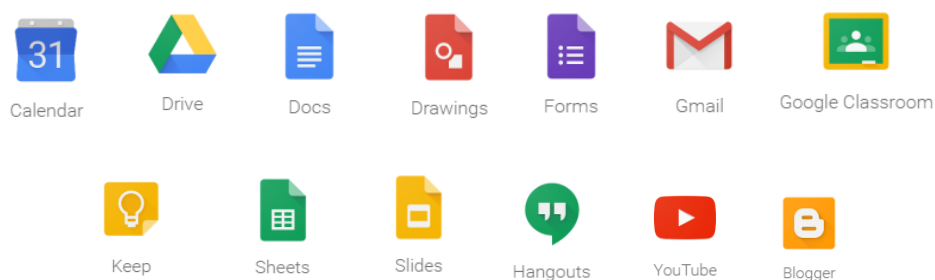
Після обговорення матеріалу необхідно коректувати матеріал та підготувати матеріал для презентації.

Отже, користувач, зареєструвавшись у початковому ресурсі, має змогу працювати з персональним кабінетом, до якого входить календар, поштові повідомлення. Але в самому веб-базованому кабінеті користувач має змогу в групі створювати контент, за допомогою спеціальних додатків та віджетів. Також користувачу надається можливість обмінюватися контентом за допомогою хмарних технологій.

Кожен користувач має своє уявлення щодо створення контенту в тій чи іншій ситуації. Таким чином, розробка продукту націлена на те, щоб користувачі створювали контент групою, вирішуючи спливаючі в міру надходження питання, які будуть виникати при виконанні завдань, за допомогою елементів, які будуть включені до розроблюваного продукту.

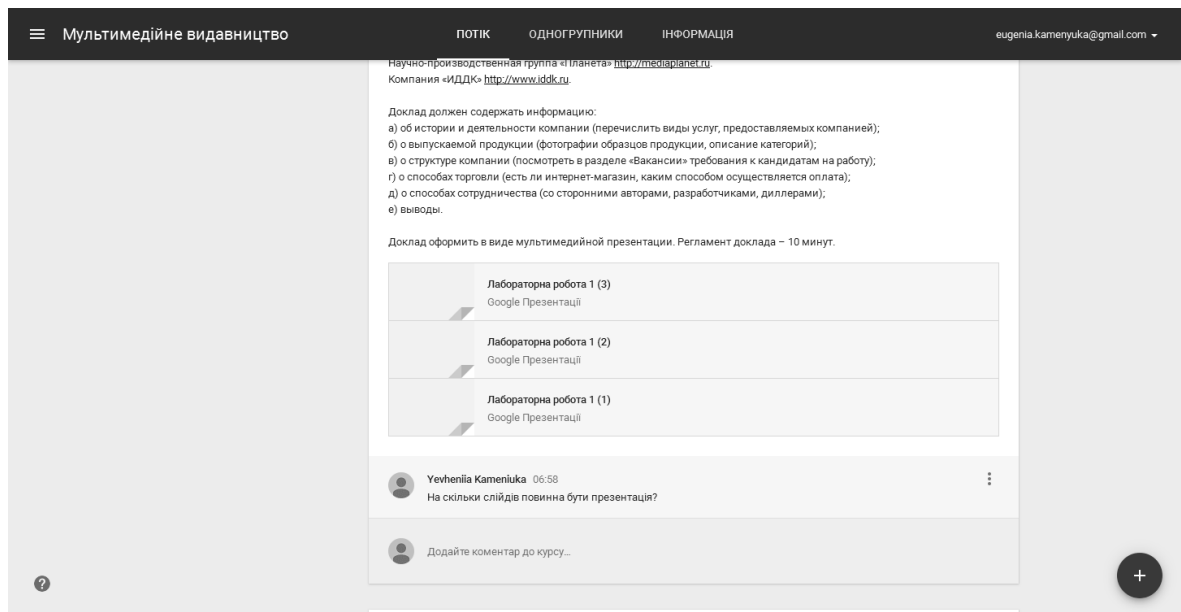
При розробці веб-базованого навчального кабінету було прийнято рішення використовувати систему управління контентом Google Sites.

Система Google Sites пропонує у своїй системі власні «засоби для освіти», які надалі використовувалися при розробці веб-ресурсу (рис. 2).



**Рис. 2. Засоби Google для створення спільного контенту**

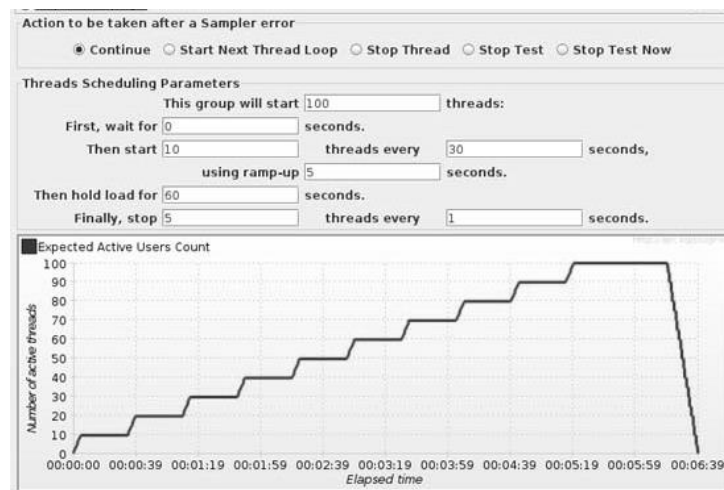
На рис. 3. наведено фрагмент сторінки, яка містить інтегровані модулі для створення спільного контенту за дисципліною.



**Рис. 3. Інтерфейс веб-ресурсу з модулем створення презентації**

Змінюючи доступ, наприклад, до презентації можна підключити колег по групі для подальшого редагування презентації та обговорення ідей.

На рис. 4 наведено аналіз тестування продуктивності ресурсу, для якого було створено 100 віртуальних користувачів, за допомогою ресурсу Apache JMeter.



**Рис. 4. Результат тестування у середовищі Apache JMeter**

З графіку видно, що незважаючи на кількість користувачів, запити до серверу йдуть стабільно згідно встановленим проміжком часу. Можна зробити висновок, що розроблюваний ресурс буде працювати стабільно та без збоїв в системі.

## **Висновки**

Таким чином, технологія спільного формування контенту в веб-базованому навчальному кабінеті дозволяє вирішити наступні проблеми:

1) доступність. Ресурс забезпечує більшу доступність і більший потенціал для залучення студентів, а також розвиток концепції навчання дистанційно. Подібні умови можуть бути особливо привабливими для студентів з обмеженими можливостями та студентів, які не можуть відвідувати заняття;

2) широке залучення студентів. Основною метою даного ресурсу є залучення студентів, поліпшення успішності студента, розширення його освітньої практики;

3) реалізація принципів безперервної освіти. Ресурс сприяє системі «освіти протягом усього життя», змушуючи студента задуматися про цілі і значення свого власного навчання і отримання освіти для підтримки конкурентоспроможності під час пошуку роботи.

Веб-базований кабінет для формування спільного контенту має значно підвищити ефективність навчального процесу за рахунок використання передових веб-технологій для формування спільного контенту.

## **Література**

1. J. Ozga, "Governing knowledge: data, inspection and education policy in Europe", *Globalisation, Societies and Education*, № 10(4), p. 439 – 455, 2012.

2. Касаткін Л. В. "Специфіка відображення інформаційних ресурсів (інформація про фонди та користувачів)", *Наукові праці Національної бібліотеки України ім. В. І. Вернадського*, № 35, с. 365 – 374, 2013.

3. S. Robertson, "Placing teachers in global governance agendas", *Comparative Education Review*, № 56, p. 584–607, 2012.

4. P. Zgaga, "The future of European teacher education in the heavy seas of higher education", *An International Journal of Teachers' Professional Development*, № 17(3), p. 347–361, 2013.

5. В. Лапінський, "Електронні засоби навчального призначення – світовий досвід й українська освіта", *Педагогіка вищої школи: методологія, теорія, технології*, № 3, с.487-495, 2013

### INTEGRATION CHAT-BOTA @RIBS\_karkas\_bot WITH EXPERT SYSTEM

**Introduction and formulation of the problem.** In the business environment, the TELEGRAM free manager has become the corporate communications standard. This is due to the following reasons: a high degree of data encryption in it, work stability, the ability to transfer large amounts of information, open protocol, cross-platform.

On the other hand, what is very important for integrating the TELEGRAM manager with other applications is that developers provide a library based on an API for working with chat bots.

A bot (chat bot, interlocutor) is an application that simulates human communication based on elements of artificial intelligence. Today, bots can communicate with each other to achieve their goals in other words, they can be used to as agents in multi-agent systems [1].

One of the first programs implementing the chat bot concept was the ELIZA program, which imitated the behavior of a psychotherapist during the initial interview of a patient [1]. The idea of implementing this program was to find key words or combinations in the text of communication, in order to ask a question to maintain a dialogue with the interlocutor. If the keyword is found in the database, then the question to the interlocutor was asked in accordance with a pre-prepared question template or the approval of the interlocutor was transformed into a question. If the word combination was not found, the program asked the interlocutor with general questions, for example, "Why do you think so?"

There are several strategies for implementing this dialogue:

1. The question is select from the list of questions, according to the higher frequency of use of the question in the subject area.

2. The bot collects questions and phrases used by the interlocutors, thus learning and increasing their content of the subject area.

3. A syntactic approach based on the grammatical analysis of an interlocutor's phrase and provided with rules of the "if-that" kind.

Naturally, the lack of such a logical dialogue between the bot and the interlocutor became the disadvantage of such communication. The interlocutor appeared illusion that the bot understands it, although in fact it is not. In other words, the chat bot lacks the implementation of interaction with the output machine, as is widely used in expert and expert training systems [1].

In recent decades, with the advent of the smartphone, the popularity of the concept of artificial intelligence in relation to messaging applications has increased. The global chat bots market is expected to grow in the coming years.

One of the main advantages of chat bots in customer service is that interlocutors can freely ask questions that they would not ask a customer service representative or company manager. In addition, the bot is able to instantly answer questions.

Chat bots are usually integrated into interactive systems, for example, virtual assistants, which enables them to naturally communicate or participate in casual conversations that are not related to the areas of their main expert systems.

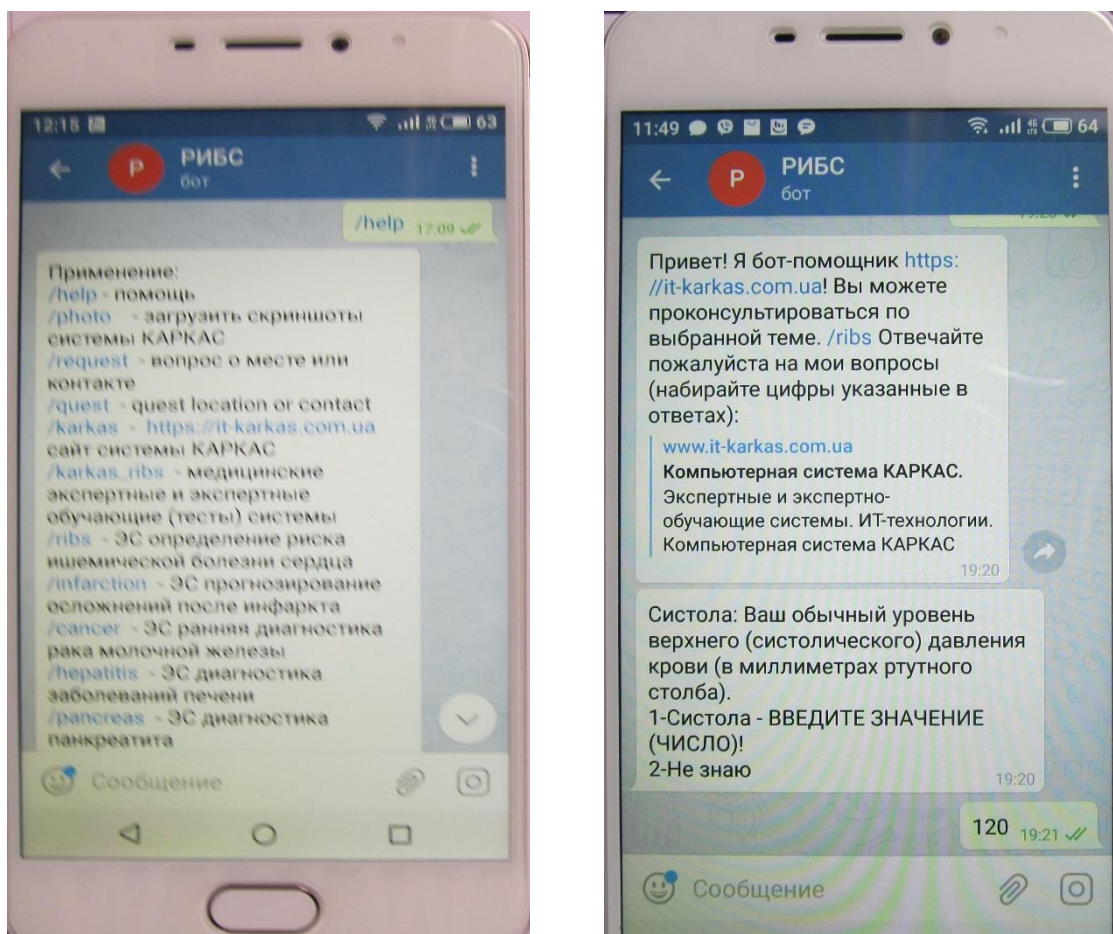
In most cases, chat bots use messaging applications to communicate with customers. A person can type or ask a question, and the chat bot will respond with the correct information. Depending on the situation, many chat bots can learn from what the client says to personalize the interaction and build up the previous interaction.

A chat bot can be viewed as a question-answer system (QA-system) with machine learning elements, namely, natural language parsing functions, an inference machine, and a communication module with external applications. The actual problem for chat bots of QA-systems is the creation of an inference machine that determines the relevance of knowledge to a given question.

The paradigm of integrating chat bots to work with expert systems is now becoming increasingly relevant [2].

Using the Telegram API library, the @Ribs\_karkas\_bot bot was created for online consultation of the user with a tool for creating knowledge bases with the "KARKAS" system [3 – 5].

Chat bots of the TELEGRAM messenger, as interlocutors, when working with the "KARKAS" system, provide more opportunities for mobile consultation with the expert system via a smartphone, which, for example, is important for making effective decisions in different subject areas such as medicine, ecology, business. In other words, you can now send a text message to the @Ribs\_karkas\_bot bot (a bot to determine the risk of coronary heart disease) and get instantly necessary information, that is, carry out a real-time consultation. The contents of the chatbot / help command are shown in fig. 1.



**Fig. 1. Type of command / help chatbot @Ribs\_karkas\_bot**

The @Ribs\_karkas\_bot bot allows you to conduct an online consultation with the following prototypes of expert systems:

- the "RIBS" expert system [3 – 5] – designed to determine the risk of coronary heart disease (CHD) in a practically healthy person. The urgency of developing a system is that currently in medicine there is a clearly expressed process of transition to the concept of prevention of coronary artery disease, that is, to the concept of risk factors associated with the lifestyle of a particular patient. The purpose of the system is to recognize the presence of risk factors for CHD with an emphasis on the individual lifestyle of the patient, using the knowledge of experts. The features of the system include the fact that with its help the patient is diagnosed: type of coronary behavior, degree of socio-psychological support, level of physical activity, degree of adequacy of rest;

- the "INFARKT" expert system [3 – 5] – helps doctors diagnose infarction patients, evaluate their condition and forecast the development of the following complications of myocardial infarction: fibrillation, acute left ventricular failure, chronic heart failure, arrhythmia, thromboembolism, myocardial rupture, recurrent infarction. The inference engine implements the Bayesian decision method.

Consultation with the "INFARKT" expert system in case of 28 symptoms takes 10 minutes, and the option of express consultation (7 symptoms) takes 3 minutes;

– the "GEPATIT" expert system [3 – 5] – is designed to diagnose acute and chronic liver diseases. The system makes it possible: to recognize the cause of liver disease and, if possible, by eliminating it, obtain a therapeutic effect, specifically include drugs for the treatment of liver diseases, conduct a statistical evaluation of therapeutic interventions in patients;

– the "ADBC" expert system [3 – 5] (automatic diagnosis of breast cancer) is intended for the early diagnosis of mammary tumors. Diagnostics relies on the knowledge of an oncologist expert, who are grouped in the following sections: thermography, anamnesis, physical examinations, echotomography. The system allows to classify the following tumors: lipoma, fibroadenoma, fibrocystic mastopathy diffuse; fibrocystic mastopathy localized, DFA (diffuse), DFA (localized), mastitis.

The "KARKAS" system is a toolkit for developing prototypes of knowledge bases for expert and expert training systems. The knowledge representation is based on a hierarchical functional system that is generated by the "KARKAS" system based on production rules and frames. The output machine uses a hierarchical functional system during user consultation. The user can select various modes of operation of the output machine: use direct output, reverse output, indirect output, Bayes formula, criteria tables, when the product's consequent is a list of parameters.

The online consultation module (interlocutor) allows you to communicate with the TELEGRAM messenger with the "KARKAS" system knowledge bases via the Internet, in other words, to provide real-time consultation.

**Main part.** Messenger TELEGRAM has millions of active users and is the fastest messaging application. It works on all devices on mobile and desktop platforms.

The "KARKAS" system is a tool for building knowledge base models. The structure of the subject area can be diverse, for example, the choice of solutions among a specific set of options, the use of unreliable knowledge. The "KARKAS" system allows both to develop models of knowledge bases and can be used for testing and teaching students on a local network. The "KARKAS" system using chat bots: @Ribs\_karkas\_bot, @test\_karkas\_bot, @it\_karkas\_bot allows online consultation with users and testing students' knowledge in various subject areas: computer graphics, database technology, web analytics, business intelligence systems.

The "KARKAS" system is built according to the modular principle and for this reason it has the ability to connect other additional modules. The following



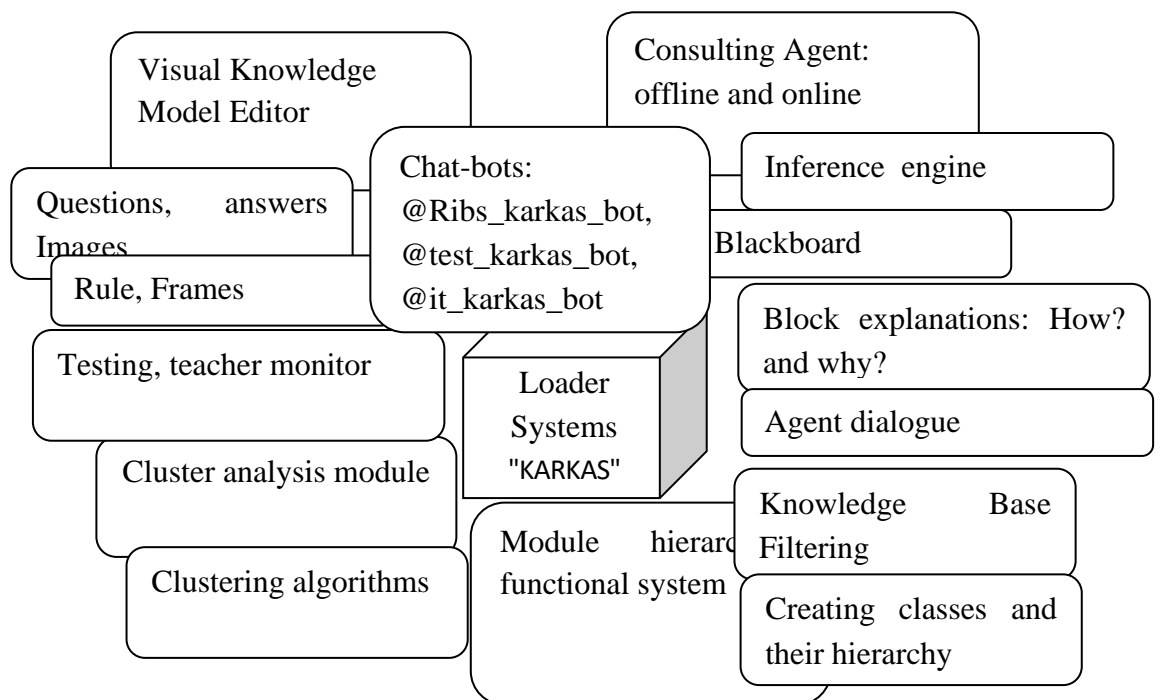
main modules can be distinguished in the system architecture: loader; visual editor for developing knowledge base models; offline and online consultation agent; dialogue agent; data clustering module.

The loader performs system startup and coordinates the interaction of all modules. The visual editor of knowledge base models allows you to create for expert and expert-training systems: questions with answers, products, frames, filtering the knowledge base, the hierarchy of knowledge base classes, a hierarchical functional system. There is an opportunity to place the knowledge base on the site <https://it-karkas.com.ua>.

The consultation agent contains the following components: inference engine, an explanation block, a notice board, a fact base monitor.

The data clustering module allows for interactive and intelligent data classification.

To test the user's knowledge, a test generator is used, and a teacher's monitor is used to monitor the tests. The modules of the system are presented in the form of software passive agents (their interaction is carried out using parameter arrays and through the general computer memory). The environment of their interaction is the system "KARKAS" fig. 2.



**Fig. 2. Tag cloud system "KARKAS"**

The system has two modes of use:

- the cognitive mode – a specialist who models the domain ontology and constructs a knowledge base model;
- the problem specialist's mode (user qualification is not high enough, and therefore he needs help and support from the system).

The system combines various approaches in building knowledge-based systems (table 1).

**Table 1**

**Components of the system "KARKAS"**

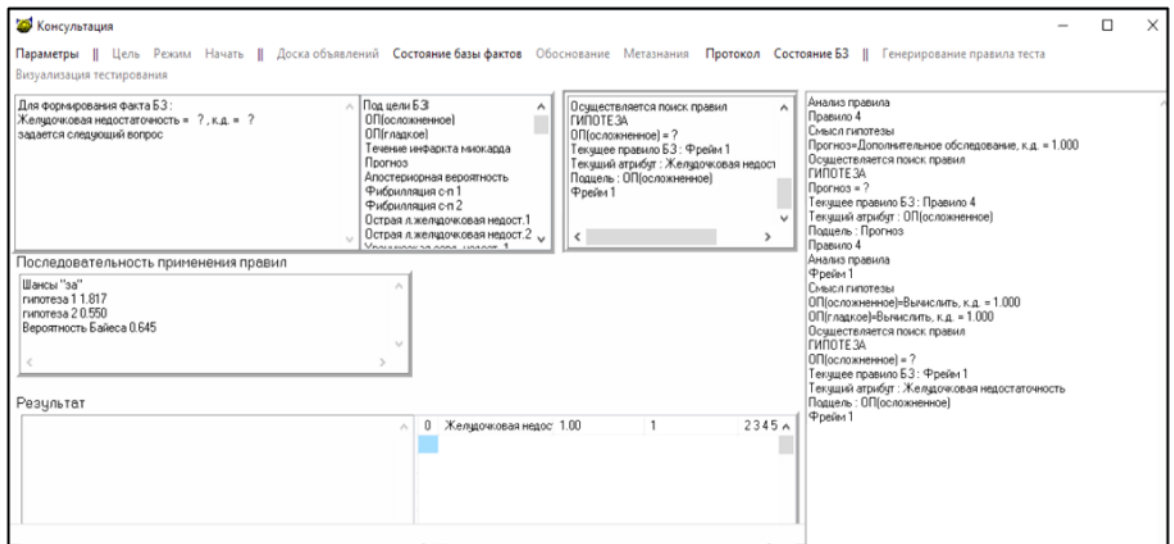
Expert Systems	Expert training systems	Multi-agent systems
Visual Knowledge Base Editor	Training	Help agents
Hierarchical functional system	Testing, intellectual tests	Chat bots: @Ribs_karkas_bot
Inference engine, Unit explanation	Teacher's monitor	@test_karkas_bot @it_karkas_bot

The components of the system are interconnected, exchange information with each other and allow to adequately display the ontology of the subject area during its operation.

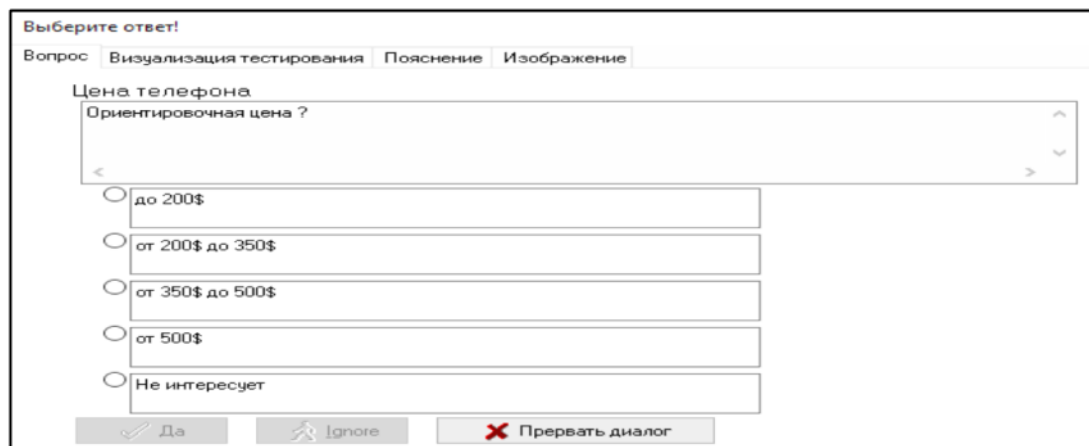
The system is implemented using the Embarcadero® RAD Studio 10.2 Tokyo programming environment and contains the following files: karkasW.exe (system loader), karkased.exe (visual editor of knowledge bases), karkascon.exe (consultation agent), karkas.hlp (help), pattern.knb (knowledge base pattern).

For the MS Windows XP operating system, the following agents are used: msagent.exe (MS agent server), merlin, peddy (agent characters), and the speech recognition engine (Microsoft Speech API).

To integrate the KARKAS system with the @Ribs\_karkas\_bot chat bot, a consultation agent is used (Fig. 3) and a dialogue agent (Fig. 4).

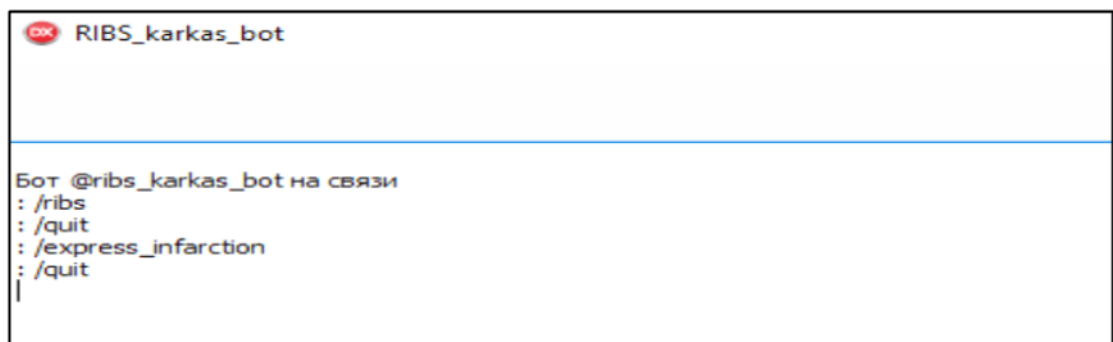


**Fig. 4. Type of form agent consultation system "KARKAS"**



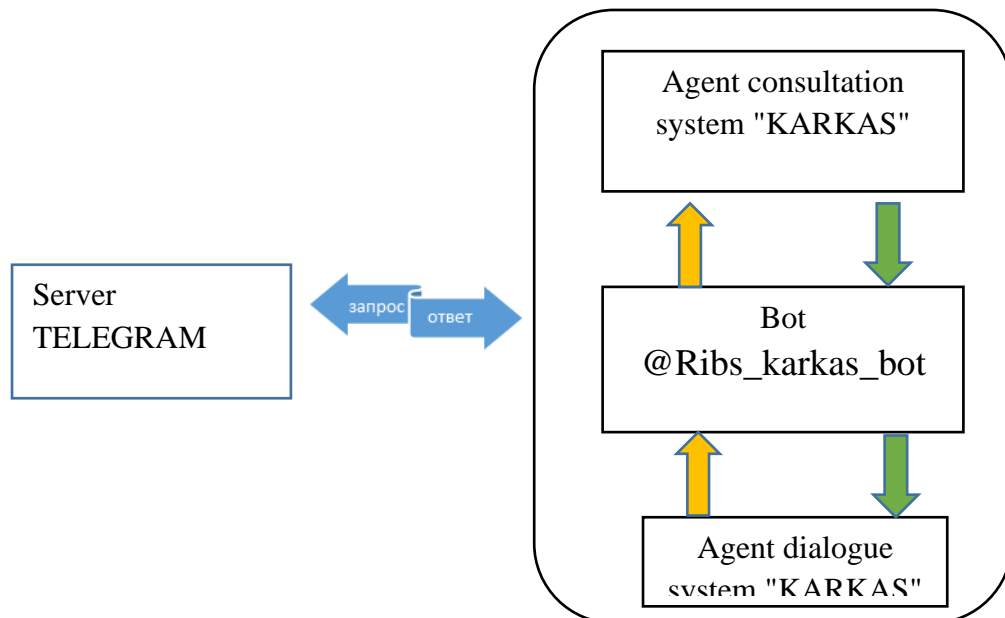
**Fig. 5. View of the form of the agent of the dialogue system "KARKAS"**

The main form of the chat bot is as follows fig. 6.



**Fig. 6. The main form of the chatbot loader @Ribs\_karkas\_bot**

Integration of the chatbot with the consultation module of the KARKAS system consists in exchanging messages between them, that is, sending and receiving requests for working with TELEGRAM servers fig. 7



**Fig. 7. Architecture of chatbot integration with KARKAS system**

The following components are used to work with queries with TELEGRAM servers:

1. for parsing JSON objects the superobject library is used;
2. for implementation through the https protocol, the OpenSSL libraries were used: libeay32.dll and ssleay32.dll;
3. Indy 10 library and the following procedure are used to send http requests and download knowledge bases via ftp protocol from <https://it-karkas.com.ua>:

```

procedure TFrmKarkasBot.BitBtnFTPloadES_KNBClick(Sender: TObject);
begin
  FtpClient1.HostName:='it-karkas.com.ua';
  FtpClient1.UserName:='it-karkas';
  FtpClient1.Password:='*****';
  FtpClient1.Port:='21';
  FtpClient1.LocalFileName:=DirKarkas+'\'+SNameFileSiteKnb;
  FtpClient1.HostDirName:= '/httpdocs/uploads/files/public-
folder/manual/knb/es_knb';
  FtpClient1.HostFileName := extractfilename(FtpClient1.LocalFileName);
  FtpClient1.Passive := True;
  try
    FtpClient1.Connect;
    FtpClient1.Cwd;
    FtpClient1.Binary := True;
    FtpClient1.Get;
    FtpClient1.Quit;
  except
    ShowMessage ('Ошибка загрузки файлов');
  end;
end;
  
```

Consider a number of functions and procedures of the agent consultation and dialogue system "KARKAS", which allows the bot to exchange messages with the channels available in TELEGRAM, using the TELEGRAM API and JSON via the secure HTTPS protocol.

Since the user communicates only with the chat bot, the problem arises in transmitting and receiving messages between the bot, consultation and dialogue modules in an automatic mode, i.e. without his participation. To realize this possibility, the WM\_SENDBOT constant was introduced, with the help of which messages are transmitted and received for the modules to function without user intervention.

Agents communicate with each other to perform the following operations:

1. push buttons, check boxes, radio buttons;
2. sending and receiving messages between visual objects on the form.

Thus, the above modules perform the functions of agents and in this sense, the implemented chat bot @Ribs\_karkas\_bot in the "KARKAS" system can be considered as a multiagent system.

For the interaction of the main form of the chatbot module with the consultation and dialogue forms of the "KARKAS" system, the following procedures are required:

1. To transmit a message from the chatbot to the agents of consultation and dialogue, a variable was introduced FSendMessage:

```
private
    FSendMessage: Cardinal;
```

A virtual constructor has been declared and the WndProc procedure has been redefined, which receives a message from the agents and sends them a message:

```
public
    constructor Create(AOwner:TComponent);override;
    constructor TFrmKarkasBot.Create(AOwner:TComponent);
    begin
        SendMessage:=RegisterWindowMessage('WM_SENDBOT');
    inherited;
    end;
    procedure TFrmKarkasBot.WndProc(var Msg: TMessage);
    begin
        if Msg.Msg=FSendNumberMessage then begin
            BtnSendMsgClick(Self);
        end else inherited;end;
```

2. The user can interrupt the dialogue with the bot with the command: /quit, which is implemented in the following code:

```
procedure TCaseAnswer.BitBtnQuitClick(Sender: TObject);
var
  TempWnd,Wnd: THandle;
begin
  if F_Consult.Visible then
  begin
    Wnd := FindWindow(NIL, 'Выберите ответ!');
    if Wnd <> 0 then begin
      Wnd:=FindWindowEx(Wnd, 0, 'TTabbedNotebook', nil);
      Wnd:=FindWindowEx(Wnd, 0, 'TabPage', nil);
      Wnd:=FindWindowEx(Wnd, 0, 'TPanel', nil);
      TempWnd:=Wnd;
      Wnd:=FindWindowEx(TempWnd, 0, 'TBitBtn', 'Прервать диалог');
      if Wnd <> 0 then begin
        SendMessage(Wnd, BM_CLICK, 0, 0);
      end;
    end; end else Bot:=0;
  end;
```

3. Bot launches consultation module procedure:

```
procedure TFrmKarkasBot.BitBtnRunESClick(Sender: TObject);
begin
  BotBusy:=1;
  F_Consult.Show;
  if Bot=1 then begin F_Consult.Close;
    BotBusy:=0;//снятие блокировки для запуска бота
  end; //если консультация загружена очистка указателей
  TrimWorkingSet;
  Bot:=0;
end;
```

4. Due to the fact that the answers to the questions of the expert system can be very long, therefore, to display them, the messenger buttons are not used, but the text is used. Therefore, chat bot offers numbered answers to the question of the expert system. If the question involves the choice of multiple answers, the answers are entered in numbers separated by commas. If, as an answer, you need to dial a number, for example, an age value, then the user in the chat dials the number that the bot sends to the inference engine to the expert system in fig. 1.

### **Bot interaction script with agent consultation.**

1. You can activate the application (ribs\_karkas\_bot.exe) that starts the bot on a resource that has access to the Internet (hosting, home computer). Then in the messenger TELEGRAM start it: @Ribs\_karkas\_bot. Type the commands /help or /start the bot and the bot will offer to choose the commands to run expert systems, tests (fig. 1).

2. For example, when you select the /ribs command, the following operations are performed:

- knowledge base ribs.knb is loaded from <https://it-karkas.com.ua>;
- the consultation agent is started and the expert system inference engine is activated;
- dialogue agent is activated.

3. The result of the expert system consultation is transmitted to the bot by the broadcast protocol by the following procedure:

```
procedure ResultMainBot;
var i2 : Integer;
begin
  if (Variant_ES=1)OR (Variant_ES=2) then
    begin
      StrTemp:="";
      for i2:=0 to MemoResult.Lines.Count-1 do
        begin
          StrTemp:=StrTemp+MemoResult.Lines.Strings[i2] + ' ';
        end;
      EdtMessage.Text:=StrTemp;
      case Bot of
        0 : begin end;//no bot
        1,2: begin // bot main
          NumberAns:=999;
          SendQA:=EdtMessage.Text;
          Recipients:=BSM_APPLICATIONS;
          BroadcastSystemMessage(BSF_POSTMESSAGE,@Recipients,FSendNumberMessage,0,0);
        end;
      end;//case
    end;
end;
```

Message handling by the dialog agent includes the following procedures:

1. The dialog agent sends the bot a message with the following procedure:

```
procedure TCaseAnswer.BitBtnMainClick(Sender: TObject);
var
  Recipients:DWORD;
  StrTemp, usage: String;
begin
  StrTemp:="";
  for i := 0 to MemoQus.Lines.Count-1 do
  begin
    StrTemp:=StrTemp+MemoQus.Lines.Strings[i] + ' ';
  end;
  //EdtMessage.Text:=AnsCurrAtribut^.Name+' '+StrTemp+'
'+ESquestion;
  EdtMessage.Text:=AnsCurrAtribut^.Name+' '+StrTemp+ #13#10;
  usage:="";
  for i := 1 to NumberAns do usage:=usage+IntToStr(i)+'-'+
ESAnswer[i]+#13#10;
  usage:=EdtMessage.Text+usage;
  SendQA:=usage;
  Recipients:=BSM_APPLICATIONS;
  BroadcastSystemMessage(BSF_POSTMESSAGE, @Recipients, FSendNum
berMessage,0,0); //послать сообщение объектам управление
end;
```

2. In order to receive messages to the dialogue agent from the bot, it is required to constantly send a request to the TELEGRAM server for GetUpdate updates. This is implemented in the following procedure, in which the parameter Timeout = 30 seconds sets how long to wait for a response from the server and the parameter Offset sets the last processed message plus the following message:



```

procedure TCaseAnswer.BitBtnParsingClick(Sender: TObject);
var
  JsObject1, JsObject2: ISuperObject;
  JSONArray: TSuperArray; s,h: string; i: integer;
  RHttp: TIdHTTP;
  RSSL: TIdSSLIOHandlerSocketOpenSSL;
  Index: integer;
begin
  try RHttp:=TIdHTTP.Create; except end;
  try RSSL:=TIdSSLIOHandlerSocketOpenSSL.Create; except end;
  RHttp.IOHandler:=RSSL;
  while True do
  begin
    JsObject1:=SO(RHTTP.Get(BaseUrl+S_Token_bot+'/getUpdates?offset='+
IntToStr(Offset)+'&timeout=30'));
    JsObject2:= JsObject1.A['result'].N[0];
    Offset:=JsObject2.i['update_id'] + 1; // In case there were no messages,
Offset will become equal to 1.
    if Offset<>1 then
    begin
      if JsObject2.s['message.text']<>" then begin
        MemoBot.lines.add(JsObject2.s['message.text']); // Message text
        Index := MemoBot.Lines.Count - 1;
        if Index < 0 then
          raise Exception.Create('Memo пуст')
        else
          begin
            EdtFromBot.Text := MemoBot.Lines[Index];
            Try RHttp.Disconnect; except end;
            RHttp.Free; RSSL.Free; Exit; end; end; end; end; RHttp.Free; RSSL.Free;
          end;

```

Thus, the @Ribs\_karkas\_bot chat bot algorithm consists of the following steps:

- Step 1. Run the application on the computer ribs\_karkas\_bot.exe (chat bot).
- Step 2. Select the commands: / help or / start, then, for example, / ribs.
- Step 3. The bot launches the agent of the "KARKAS" system.
- Step 4. The "KARKAS" inference engine is activated.

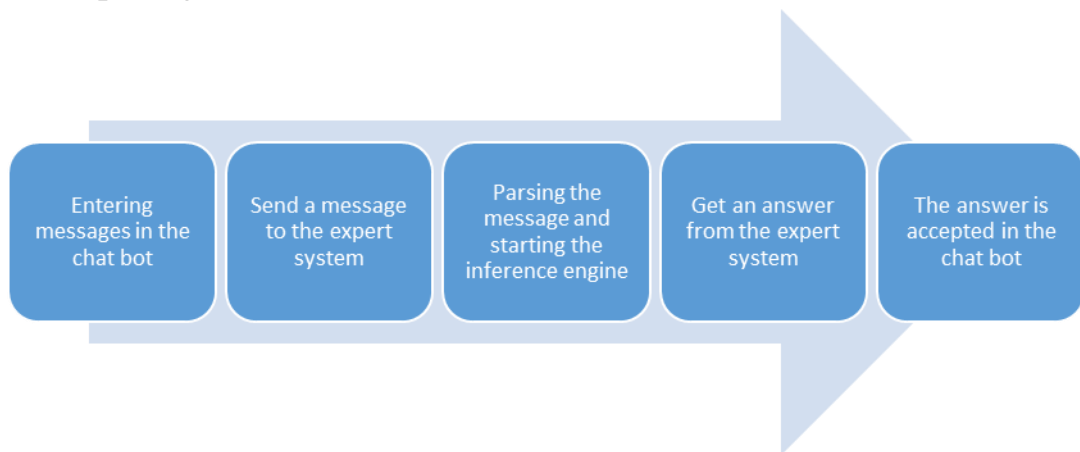
Step 5. Forms a hierarchical functional system for dialogue with the user "KARKAS".

Step 6. The agent of dialogue is activated and the message with the text of a question and answers a bot is sent. It accepts the message as a JSON object, parses it, displays the message in the chat and waits for the user's response.

Step 7. The user in the chatbot selects or enters the answer and the bot sends it to the output machine to the expert system.

Step8. An expert advice expert system accepts the message and sends the inference engine. The dialogue agent performs the next step according to the constructed hierarchical functional system for the dynamic purpose of the consultation. The purpose of the consultation is formed during the dialogue with the user. In other words, it dynamically depends on the user's answers to the questions of the expert system.

Step 9. The iterative consultation process continues until the inference engine receives the result. The user can stop the consultation at any time with the command / quit fig. 9.



**Fig. 9. Chat bot interaction @Ribs\_karkas\_bot step with expert system**

**Conclusions.** As a result, a fully functioning chat bot @RIBS\_karkas\_bot was created, which is integrated into the "KARKAS" system, which allows for online consultation with the RIBS expert system to determine the risk factor for coronary heart disease. After the application is deployed, it is planned to significantly expand the functionality of the bot.

The "KARKAS" system is a toolkit for developing prototypes of knowledge bases for expert and expert-training systems both offline and online on smartphones. The knowledge representation is based on a hierarchical functional system that is generated by the "KARKAS" system based on production rules and frames. The inference engine uses a hierarchical functional system during user consultation. The user can select various output machine modes: use direct output, reverse output, Bayes formula, criteria tables.

Using the "KARKAS" system, a number of expert system prototypes have been developed in the following subject areas: medicine, economics, mobile communications, and cluster analysis of multidimensional data.

The KARKAS system using chat bots: @Ribs\_karkas\_bot, @test\_karkas\_bot, @it\_karkas\_bot allows online consultation with users and testing students' knowledge in various subject areas: computer graphics, database technology, web analytics, business intelligence systems.

### **Literature**

1. Рассел С. Искусственный интеллект: современный поход / С. Рассел, П. Норвиг. – 2-е изд.; [Пер. с англ. – М.: Изд. дом "Вильямс", 2006. – 1408 с.

2. Бурдаев В.П. Інтегрування месенджерів з системою "КАРКАС". // Тези доповідей: міжнарод. наук.-практ. конф. Проблеми і перспективи розвитку ІТ-індустрії, Харків, 2018, с.7.

3. Бурдаев В. П., Системы навчання з елементами штучного інтелекту. Харків, Україна: ХНЕУ, 2009.

4. Бурдаев В. П., Сложность динамических систем: монография. Saarbrucken, LAP LAMBERT Academic Publishing, 2015.

5. Burdaev V. P. "About one concept of constructing a temporal knowledge base", in Proc. of the 1<sup>st</sup> International Congress Fundamental and Applied Studies in the Pacific and Atlantic Oceans Countries, Tokyo University Press, 2014, pp. 272–276.

## ГЛАВА 4

### МОДЕЛІ ОЦІНЮВАННЯ ЕФЕКТУ ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ЕЛЕКТРОННОГО НАВЧАННЯ

#### Вступ і постановка задачі

Одним з основних засобів реалізації технологій навчання та контролю в професійній освіті стали системи підтримки електронного навчання (СПЕН), які забезпечують інформаційну, методичну та технологічну підтримку освіти.

Система оцінки результатів дистанційного навчання (СОЕН) є обов'язковим атрибутом сучасних СПЕН. Від того які критерії закладені при її розробці багато в чому залежить ефективність подальшого застосування всієї системи електронного навчання.

Під ефективністю використання СПЕН та СОЕН розуміється відношення результату щодо освітніх досягнень студентів до відповідних витрат необхідних для досягнення цього результату. Тобто відношення ефекту до витрат, де під витратами розуміється вартість систем СПЕН та СОЕН, на базі якої здійснюється навчання.

В якості ефекту надалі розглядаються компетентності (знання, уміння, навички), які здобувають студенти в проміжок часу між початком і закінченням процесу навчання.

В рамках цього дослідження основна увага надається створенню моделі взаємозв'язку показників оцінки ефекту використання системи СОЕН та розробленню відповідної методики розрахунку комплексного показника оцінки ефекту застосування системи СПЕН в учбовому процесі.

Оцінка учбового ефекту не є тривіальним завданням оскільки залежить від різноманітними факторами, кожен з яких визначається безліччю взаємопов'язаних критеріїв. Успішність його рішення багато в чому визначається інтуїцією і досвідом розробника системи електронного навчання. Як наслідок, в процедурі формування інтегральної оцінки розробники системи повинні значну увагу приділяти експертним методам, що дозволяють кількісно описувати основні характеристики учбового ефекту та значно зменшити інтуїтивну складову прийняття технологічних рішень.

**Метою дослідження** є створення моделі взаємозв'язку показників оцінки ефекту використання системи СОЕН

## **Основна частина**

Дане дослідження базується на використанні методу аналізу ієрархій (MAI) та його подальшого розвитку – методу аналізу систем [6].

Проведення дослідження побудоване на таких припущеннях.

Припущення 1. Критерії, які були покладені в основу проектування (створення) систем оцінки ефективності використання СПЕН також повинні бути враховані і при оцінюванні результатів її роботи, тобто оцінки ефективності системи СОЕН.

Припущення 2. Сучасні системи підтримки електронного навчання в вищій школі обов'язково мають вбудовану систему оцінки результатів навчання.

З припущення 1 слід висновок, що засади створення СПЕН і засади стосовно оцінки ефективності їх використання повинні мати єдину базу критеріїв.

Припущення 2 дозволяє стверджувати, що можливості системи СОЕН будуть в значній мірі визначатися якісними характеристиками усієї системи підтримки електронного навчання - СПЕН.

Засади створення СПЕН і засади стосовно оцінки ефективності їх використання повинні мати єдину базу критеріїв, а можливості системи СОЕН будуть в значній мірі визначатися якісними характеристиками усієї системи підтримки електронного навчання – СПЕН.

Взаємозв'язок засад, які розглядаються, наведено на рис. 1.

Як слід з рисунка, у якості підгрунтя засад створення системи оцінки СОЕН і засад створення системи СПЕН обрано процес формування низки відповідних критеріїв та розроблення на їх основі багаторівневого графа, який в подальшому будемо називати ранговою моделлю. Модель дозволяє визначити значущість (важливість, ранг) кожного з взаємопов'язаних критеріїв. Доцільність застосування рангової моделі виникає коли кількість критеріїв перевищує п'яти – семі. Потрібність в цьому, як правило, суттєва на початковому етапі проектування систем СПЕН і СОЕН, коли мають місце ресурсні (тимчасові або фінансові) обмеження і необхідно виділити критерії, які слід враховувати в першу чергу.



**Рис. 1. Взаємозв'язок засад оцінки ефективності використання систем електронного навчання (СПЕН) та засад створення відповідної системи оцінювання (СОЕН).**

У якості прикладу надалі розглянута дев'яти елементна множина якісних характеристики системи підтримки електронно навчання.

*Критерій 1.* Технології розробки курсу.

Технології розробки курсу можуть бути різноманітними: Web-технології, телекомунікаційні технології, хмарні технології, технології мобільного навчання, технології дистанційного навчання тощо

*Критерій 2.* Моделі курсів і їх педагогічний дизайн.

Модель курсу визначає специфіку організації та подачі матеріалу.

*Критерій 3.* Наявність навчального плану курсу.

Наявність навчального плану відображає авторство, структуру, тематику розділів або модулів, його технічну складову і ресурсну базу.

*Критерій 4.* Технології доставки курсу.

Технології доставки і доступу до курсів і сервісів припускають асинхронний або синхронний формат навчання на основі використання електронних навчальних оболонок, навчальних можливостей Google у вигляді сайтів і блогів.

*Критерій 5.* Простота доступу до курсів і сервісів.

*Критерій 6.* Зручність навігації курсу.

Педагогічний дизайн і зручність навігації повинні підтримувати навчальний процес і бути зручними для користувачів.

*Критерій 7.* Наявність методичних рекомендацій по роботі з матеріалами курсу.

Наявність методичних рекомендацій по роботі з матеріалами курсу допомагає як студентам, так і викладачам

*Критерій 8.* Наявність технологій взаємодії.

Технології взаємодії зі слухачами припускають використання блогу, чату, форуму, скайпу, електронної пошти, телефону, очних зустрічей або консультацій.

*Критерій 9.* Адаптивність і персоналізація курсу.

Адаптивність і персоналізація курсу дозволяють врахувати контактність аудиторії, її готовність до сприйняття матеріалу, коригувати наповнення контенту, балансувати інформаційне навантаження, вибудовувати індивідуальні траєкторії навчання.

Таким чином є дев'яти елементний безліч частково або повністю пов'язаних критеріїв, які розглядаються в контекстному відношенні як система, опис якої може прийняти одну з двох різних, проте пов'язаних форм: бінарної матриці або спрямованого графа (мережі) для геометричного уявлення відносин.

Виникає задача перетворити вихідний граф в систему з рівнями (рангову модель) [6].

*Крок 1.* Побудова геометричного уявлення відносин між критеріями.

Поточна безліч вершин  $N$  визначена наступним чином:

$$N = \{ h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9 \},$$

де  $h_1--h_9$  – перелік критеріїв, які були розглянути в попередньому розділі.

Безліч  $N$  може бути уявлена як сукупність вершин-критеріїв (рис. 2),



**Рис. 2. Граф взаємозв'язків критеріїв.**

Зв'язки між вершинами (критеріями) визначаються як результат суджень експерта на питання:

Питання 1. Чи є взаємозв'язок між поточною парою критеріїв? Якщо відповідь "так", то відповідна пара з'єднується лінією.

Питання 2. Який з критеріїв є більш важливим (домінуючим)? Зазначений критерій позначається стрілкою, яка на нього вказує.

*Крок 2. Побудова матриці суміжності.*

Матриця суміжності (табл.1)  $B = \{ b_{ij} \}$  заповнюється на підставі бінарного відношення «залежить від» і визначається наступним чином:

$b = 1$ , якщо критерій  $i$  залежить від критерія  $j$

$b = 0$ , – в протилежному випадку.



Матриця суміжності  $V = \{b_{ij}\}$ 

$j \rightarrow$ $i \downarrow$	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$	$h_6$	$h_7$	$h_8$	$h_9$
$h_1$		1	1		1	1		1	1
$h_2$			1						
$h_3$							1	1	1
$h_4$	1					1		1	
$h_5$									
$h_6$					1				
$h_7$									1
$h_8$									1
$h_9$									

*Крок 3. Побудова матриці досяжності.*

Матриця досяжності визначається як бінарна матриця, в якій елементами є одиниці, якщо вершина графа будь-яким шляхом досяжна з іншої вершини, в іншому випадку елементи її – нулі. В роботі [6] надана формальна процедура її визначення:

$$(I + V)^{k-1} \leq (I + V)^k = (I + V)^{k+1}, \quad (1)$$

згідно з якої на базі матриці  $V$  формується бінарна матриця  $(I + V)$ , де  $I$  – одинична матриця. Показано, що існує найменше ціле  $k$ , при якому кожен елемент матриці  $(I + V)^{k-1}$  менше відповідного елемента матриці  $(I + V)^k$  або дорівнює йому, а відповідні елементи матриць  $(I + V)^k$  і  $(I + V)^{k+1}$  рівні. Матриця в правій частині виразу (1) називається матрицею досяжності.

При малій кількості критеріїв (до 7-10) побудова матриці досяжності може бути виконана шляхом прямого відстеження зв'язків у вихідному графі або – як результат аналізу відповідних рядків матриці суміжності.

Так, з розглядання першого рядку матриці суміжності можна бачити, що відносно вершини  $h_1$  досяжними є вершини  $h_2, h_3, h_5, h_6, h_8, h_9$ . А з урахуванням самої вершини  $h_1$  досяжними будуть вершини:  $h_1, h_2, h_3, h_5, h_6, h_7, h_8, h_9$ . Тому в відповідних клітинах першого рядку матриці досяжності слід записати одиниці.

Більш детально ця мнемонічна процедура розглянута в роботі [7].  
Кінцевий результат виконання другого кроку наведено в табл. 2.

**Таблиця 2.**

**Матриця досяжності**

$j \rightarrow$ $i \downarrow$	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$	$h_6$	$h_7$	$h_8$	$h_9$
$h_1$	1	1	1		1	1	1	1	1
$h_2$		1	1				1	1	1
$h_3$			1				1	1	1
$h_4$	1	1	1	1	1	1	1	1	1
$h_5$					1				
$h_6$					1	1			
$h_7$							1		1
$h_8$								1	1
$h_9$									1

*Крок 4. Аналіз матриці досяжності і побудова на його основі системи з рівнями (рангової моделі).*

Використання матриці досяжності дозволяє розділити поточну безліч вершин  $H = \{ h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9 \}$  на безліч рівнів, а також розділити кожен рівень на відповідні підмножини.

Таким чином, з безліч вершин  $H$  можна виділити два види множин: досяжні вершини ДВ ( $h_i$ ) – безліч вершин, яких можна досягти з вершини  $h_i$ ;

вершини попередниці: ВП ( $h_i$ ) безліч вершини, з яких можна досягти вершину  $h_i$ .

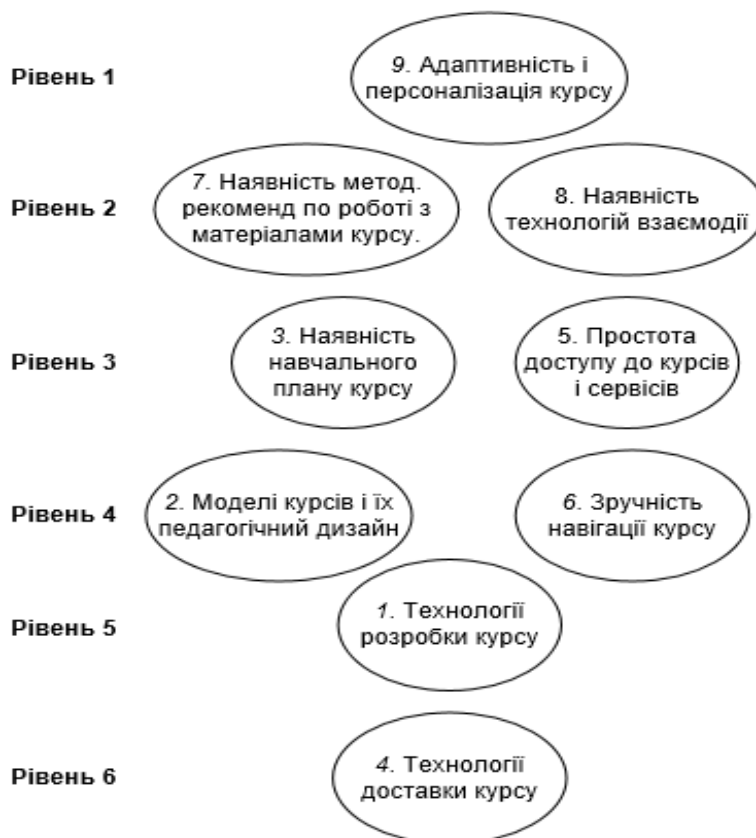
Безліч тих вершин  $h_i$ , для яких виконується  $ВП(h_i) = ДВ(h_i) \cap ВП(h_i)$ , що не можуть бути досяжним з будь-якої з решти вершин  $H$ , згідно [6] визначаються як рівень ієрархії.

На рис. 3. надано кінцевий результат даної процедури побудови рангової моделі. З рисунка слідує, що модель має шість рівнів і найбільш важливішим критерієм (перший рівень) серед якісних показників оцінки ефекту використання системи СПЕН є критерій №9 «Адаптивність і персоналізація курсу».

Розглянута рангова модель дозволяє оцінити ступінь впливу поточної множини критеріїв на остаточний результат оцінювання.

Не менш важливим для розробника системи СПЕН є пошук найбільш суттєвих зв'язків між критеріями оцінки її якості та кількісне оцінювання значущості найбільш значущих з них.

Для вирішення цієї проблеми пропонується процедура оптимізації, яка заснована на переборі і подальшому аналізу відгуку на розрив усіх (або вибіркових) зв'язків у вихідному графі. В якості індикатора на суттєвість зв'язку пропонується використовувати зміну рівнів критеріїв у відповідній рангової моделі.



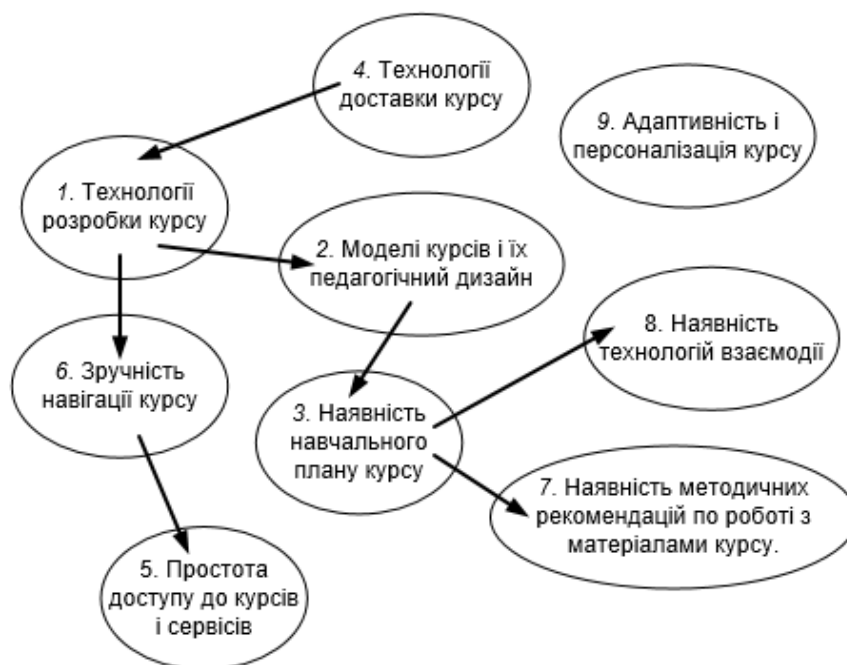
**Рис. 3. Рангова модель якісних показників оцінки ефекту використання системи СПЕН.**

На рис. 4 наведено остаточний результат виконання цієї процедури. *Кількісне оцінювання значущості зв'язків оптимізованого графу* пропонується виконувати на основі поняття «вектору конфігурації рангової моделі», яке визначається у вигляді N-елементного вектору, де N – кількість рівнів моделі. Так, рангової моделі, яка наведена на рис. 3 відповідає наступний вектор конфігурації:

$$\{9; (7, 8); (3, 5), (2, 6); 1; 4\},$$

де кожен елемент містить підмножину критеріїв відповідного рівня, починаючи від найстаршого – першого.

Кожному поточному вектору конфігурації співвідноситься його вага, яка обчислюється, як адитивна згортка добутків рангу вектору на кількість критеріїв цього рангу. Так, вага вектору конфігурації {9; (7, 8); (3, 5), (2, 6); 1; 4} дорівнює 30.

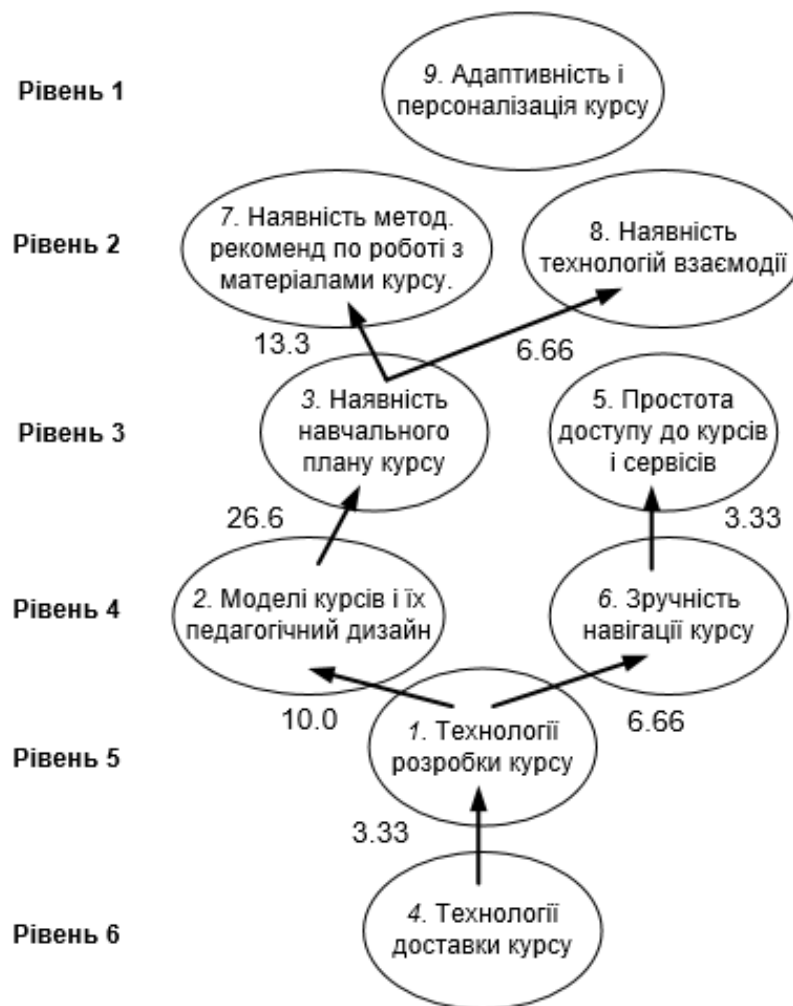


**Рис. 4. Результат оптимізації зв'язків поточного графа.**

Кількісне оцінювання значущості зв'язків оптимізованого графу (рангової моделі) пропонується обчислювати як абсолютну різницю вагових коефіцієнтів векторів конфігурації початкової і поточної рангових моделей, де під поточною конфігурацією розуміється конфігурація, що отримана при розриву зв'язку, який досліджується.

На рис. 5 наведено остаточний результат кількісного оцінювання домінування суміжних критеріїв рангової моделі. Кількісні значення на дугах графа нормовані відносно ваги початкового вектору конфігурації і збільшені на коефіцієнт, який дорівнює 100.

Отримані результати можуть бути застосовані в процесі розроблення ієрархічних або холархічних моделях обґрунтування альтернативних варіантів систем електронного навчання.



**Рис. 5. Кількісна оцінка домінування суміжних критеріїв рангової моделі оцінювання учбового ефекту системи СПЕН.**

### **Розроблення моделі взаємозв'язку критеріїв оцінювання альтернативних варіантів систем електронного навчання.**

В роботах [4; 5] розглядаються варіанти показників оцінки результатів навчання. Різноманітність варіантів пояснюється поточними цілями навчання і тому вибір конкретного варіанту цілком залежить від набору компетентностей, якими повинні оволодіти студенти в рамках дисципліни, що вивчається.

Розрізняють абсолютні і відносні показники оцінки ефекту навчання. Абсолютні показники незалежні від конкретної предметної області, в той час як відносні тісно прив'язані до особливостей поточної дисципліни.

У даній роботі в якості критеріальної бази оцінювання обрано набір абсолютних показників, які згруповані у відповідні кластери.

Кожен кластер містить набір взаємозв'язаних критеріїв, що утворюють в загальному випадку холархіческую мережу. Нижче подано стислу характеристику відповідних груп критеріїв.

*Кластер 1. Якісні характеристики СЕН.*

*Критерій 1.1. Технології розробки курсу.*

Технології розробки курсу можуть бути різноманітними: Web-технології, телекомунікаційні технології, хмарні технології, технології мобільного навчання, технології дистанційного навчання тощо

*Критерій 1.2. Моделі курсів і їх педагогічний дизайн.*

Модель курсу визначає специфіку організації та подачі матеріалу.

*Критерій 1.3. Наявність навчального плану курсу.*

Наявність навчального плану відображає авторство, структуру, тематику розділів або модулів, його технічну складову і ресурсну базу.

*Критерій 1.4. Технології доставки курсу.*

Технології доставки і доступу до курсів і сервісів припускають асинхронний або синхронний формат навчання на основі використання електронних навчальних оболонок, навчальних можливостей Google у вигляді сайтів і блогів.

*Критерій 1.5. Простота доступу до курсів і сервісів.*

*Критерій 1.6. Зручність навігації курсу.*

Педагогічний дизайн і зручність навігації повинні підтримувати навчальний процес і бути зручними для користувачів.

*Критерій 1.7. Наявність методичних рекомендацій по роботі з матеріалами курсу.*

Наявність методичних рекомендацій по роботі з матеріалами курсу допомагає як студентам, так і викладачам

*Критерій 1.8. Наявність технологій взаємодії.*

Технології взаємодії зі слухачами припускають використання блогу, чату, форуму, скайпу, електронної пошти, телефону, очних зустрічей або консультацій.

*Критерій 1.9. Адаптивність і персоналізація курсу.*

Адаптивність і персоналізація курсу дозволяють врахувати контактність аудиторії, її готовність до сприйняття матеріалу, коригувати наповнення контенту, балансувати інформаційне навантаження, вибудовувати індивідуальні траєкторії навчання.

## Кластер 2. Підходи до оцінки освітніх досягнень студентів.

*Критерій 2.1.* Наявність критеріально-орієнтованого підходу, що дозволяє оцінити, наскільки студенти досягли заданого рівня знань і умінь відповідно освітнього стандарту.

*Критерій 2.2.* Наявність нормативно-орієнтованого підходу, в основі якого є статистичні норми, які визначаються для даної групи студентів.

*Критерій 2.3.* Наявність підходу, який орієнтований на індивідуальні норми для кожного студента, тобто реальний рівень його розвитку в конкретний момент часу.

## Кластер 3. Результати навчання (компетентності).

Як результатів навчання тут вказується набір компетентностей, якими повинен опанувати студент після вивчення конкретної дисципліни. На даному етапі позначимо їх в загальному вигляді і обмежимося двома наборами компетентностей.

*Критерій 3.1.* Досягнення результату 1.

*Критерій 3.2.* Досягнення результату 2.

## Кластер 4. Тестові завдання.

*Критерій 4.1.* Наявність завдання для вибору найбільш правильної відповіді, з ряду відповідей (правильних в різної ступені), або для вибору кількох правильних відповідей.

Завдання з вибором однієї найбільш правильної відповіді в тестовій практиці поширені досить широко, що пояснюється технологічністю цієї форми, в першу чергу, для автоматизації завдань.

*Критерій 4. 2.* Наявність завдання у відкритій формі, при відсутності будь-яких готових відповідей.

У завданнях відкритої форми готові відповіді не пропонуються, і випробовуваний сам повинен дописати відповідь, який свідчить про наявність чи відсутність необхідних знань. Завдання у відкритій формі вимагають максимального творчості від студентів і значних витрат викладацької праці при перевірці.

*Критерій 4.3.* Наявність завдання на встановлення відповідності.

В завданнях на встановлення відповідності треба встановити відповідність елементів однієї групи елементів іншої групи.

*Критерій 4.4.* Наявність завдання на встановлення правильної послідовності.

Завдання на встановлення правильної послідовності дозволяють не тільки перевірити знання, вміння і навички по визначенню послідовності різних дій, операцій, розрахунків, а й дають можливість ефективно формувати їх в навчальному процесі.

#### Кластер 5. Види оцінювання.

##### *Критерій 5.1.* Наявність діагностичного оцінювання.

Визначає здібності і підготовленість до вивчення програми, а також можливі навчальні проблеми. Воно часто проводиться протягом навчального процесу для контролю розуміння і засвоєння навчального матеріалу окремими студентами. Основна функція діагностичного оцінювання - коригувальна.

##### *Критерій 5.2.* Наявність поточного оцінювання.

Забезпечує своєчасний зворотний зв'язок, сприяє поліпшенню навчального процесу, допомагає виявити відхилення від запланованої програми і прийняти оперативні рішення щодо корекції процесу засвоєння навчального матеріалу. Під час проведення поточного оцінювання важливу роль відіграє використання елементів самооцінки з боку студентів, що дозволяє досягати глибини розуміння і засвоєння навчального матеріалу, формувати стійкі навички, розвивати вміння і в кінцевому підсумку досягати високої якості освоєння освітньої програми.

##### *Критерій 5.3.* Наявність підсумкового оцінювання.

Забезпечує вимір і оцінку досягнення студентами запланованих результатів навчання, виявляє сильні і слабкі сторони навчального процесу.

#### Кластер 6. Альтернативні системи оцінки досягнення ефекту електронного навчання.

В якості альтернативних систем були обрані три найбільш поширені системи електронного навчання.

*6.1. Система тестування INDIGO [8].*

*6.2. Система електронного навчання Moodle [9].*

*6.3. Система електронного навчання Web Tutor [10].*

Взаємозв'язки кластерів і критеріїв всередині кожного з кластерів, як правило, носять холархічний характер.

На рис. 6 надано варіант холархічної моделі взаємозв'язку кластерів. Модель побудована як результат експертного опитування, що було



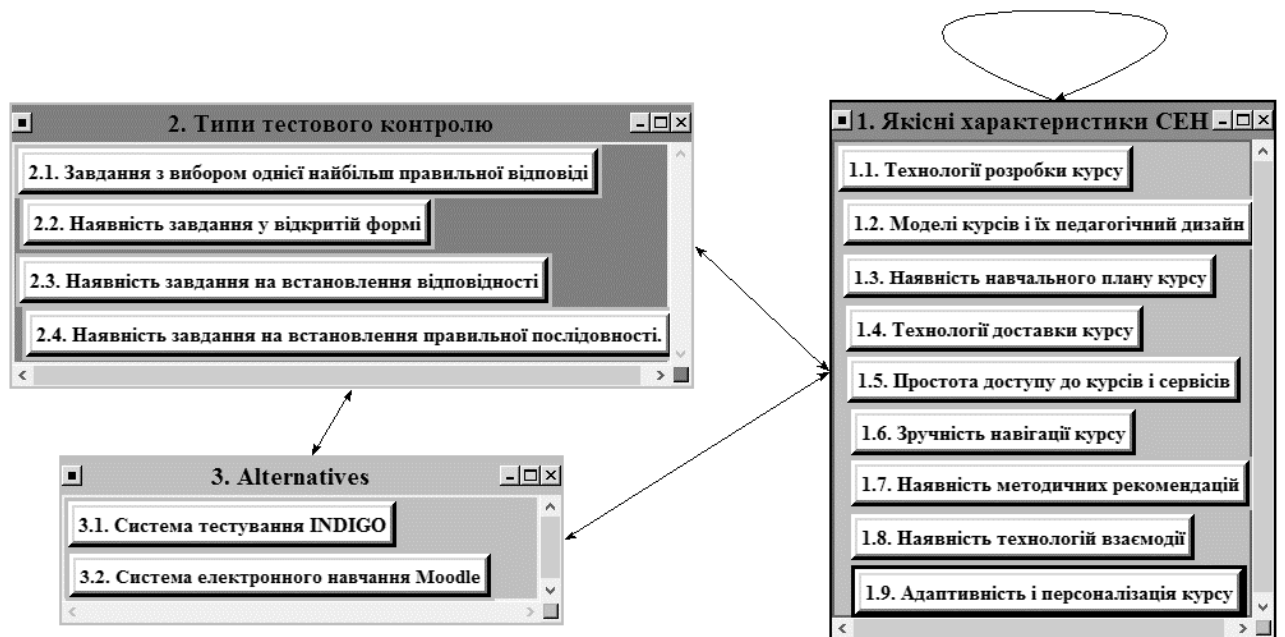
проведено серед викладачів кафедри комп'ютерних систем і технологій ХНЕУ ім. С. Кузнеця.



**Рис. 6. Модель взаємозв'язку кластерів оцінювання альтернативних варіантів систем СОЕН.**

Запропонована модель може бути рекомендована для обчислення граничних глобальних пріоритетів ефекту використання альтернативних варіантів систем СЕН, а також локальних пріоритетів відповідних кластерів і критеріїв.

Опис процедури розрахунку пріоритетів в повному обсязі є досить трудомістким, оскільки вимагає формування та подальшого заповнення експертами значної кількості матриць парних порівнянь. Тому для спрощеного подальшого розгляду доцільно залишені тільки три кластери (рис. 7), які відповідали першим двом найбільш важливим рівням (див. рис. 6).



**Рис. 7. Спрощений варіант моделі взаємозв'язку кластерів оцінювання альтернативних варіантів систем СОЕН.**

## **Висновки**

Розглянута множина критеріїв для оцінювання ефекту застосування альтернативних варіантів систем електронного навчання.

Розроблена рангова модель критеріїв оцінювання учбового ефекту системи підтримки електронного навчання.

Розроблена холархічна модель взаємозв'язку критеріїв оцінювання альтернативних варіантів систем електронного навчання.

Удосконалення моделей може здійснюватися шляхом застосування теорії нечітких множин у процесі формування матриць парних порівнянь.

## **Література**

1. Войтович И. К.. КРИТЕРИИ ЭФФЕКТИВНОСТИ ЭЛЕКТРОННОГО ОБУЧЕНИЯ И КАЧЕСТВА ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ В ВУЗЕ. [Електронний ресурс]. Доступно:

[https://vestnik.tspu.edu.ru/files/vestnik/PDF/articles/voytovich\\_i.\\_k.\\_152\\_156\\_4\\_14\\_5\\_2014.pdf](https://vestnik.tspu.edu.ru/files/vestnik/PDF/articles/voytovich_i._k._152_156_4_14_5_2014.pdf) /Вестник ТГПУ (TSPU Bulletin). 2014. 4 (145)/

2. ЯСКЕВИЧ С. В., МАКОВСКАЯ Е. В. Обзор методик оценки эффективности электронного обучения. г. Минск, Государственное

учреждение образования «Институт бизнеса и менеджмента технологий» Белорусского государственного университета,

3. Гущина Н.І., "Мультимедійний комплекс: на допомогу вчителю" [Електронний ресурс]. Доступно: <http://www.air-edu.com.ua/wp-content/uploads/2015/04/Methodichniy-posibnik-dlya-vchitelya-Multimediyniy-kompleks-Epson.pdf>. Дата звернення: Вер. 15, 2017.

4.. Оценка эффективности электронного дистанционного обучения. [Електронний ресурс]. Доступно: <http://eng.websoft.ru/db/wb/26D3C1E6AD7BC171442579820031F250/doc.html>

5. Оценка эффективности процесса обучения при наличии сложных открытых задач с помощью экспертных методов В.А. Латыпова Инженерный вестник Дона, №1 (2016) [Електронний ресурс]. Доступно: [ivdon.ru/ru/magazine/archive/n1y2016/3540](http://ivdon.ru/ru/magazine/archive/n1y2016/3540)

6. Саати Томас Л. Принятие решений при зависимостях и обратных связях: Аналитические сети. Пер. с англ. / Науч. ред. А. В. Андрейчиков, О. Н. Андрейчикова. — М.: Издательство ЛКИ, 2008. — 360 с.3.

7. Браткевич В. В. Оценка качества систем поддержки E-learning – Системи обробки інформації. Проблеми і перспективи розвитку ІТ-індустрії Випуск 4(141). Харків-2016, стр. 219-222.

8. NDIGO Програма для створення тестів та онлайн тестування студентів. [Текст]. – Режим доступу: <https://medytsyna.com/indigo-programa-dlya-stvorennya-testiv-ta-onlajn-testuvannya-studentiv/>

9. Основные возможности Moodle [Текст]. – Режим доступу: <http://moodlelearn.ru/mod/page/view.php?id=174>

## ГЛАВА 5

### КРАУДСОРСИНГОВА ПЛАТФОРМА ЯК ІТ-ІНСТРУМЕНТ ДЛЯ ЗАБЕЗПЕЧЕННЯ ГЕНДЕРНОЇ РІВНОСТІ В УНІВЕРСИТЕТІ

#### Вступ і постановка задачі

В системі вищої освіти України в останні роки спостерігаються значні зміни у законодавчому забезпеченні. Так, дедалі більше уваги в законодавчих актах приділяється питанням забезпечення рівних можливостей та недискримінації в освіті та науковій сфері, в тому числі питанням забезпечення гендерної рівності. Така ситуація вимагає розробки та впровадження сучасних методик та інструментів для збору та аналізу даних в гендерному розрізі, для обговорення проблем гендерної нерівності та обміну передовим досвідом. Сучасні ІТ-інструменти значно спрощують як процеси збору та аналізу кількісних даних в ході гендерного аудиту, так і процеси початкового якісного аналізу проблем гендерного паритету, і процеси обговорення, розробки та впровадження змін, і обмін досвідом між освітніми та науковими установами.

Сьогодні в Україні реалізується достатня кількість ініціатив, спрямованих на забезпечення принципів гендерної рівності в різних сферах соціальних відносин. Однією з таких ініціатив є проект EQUAL-IST «Планування гендерної рівності в наукових дослідженнях у галузі комп'ютерних наук та ІТ» [2], який фінансується ЄС у межах програми HORIZON2020 та має на меті впровадження структурних змін для підвищення гендерної рівності в університетах, що займаються освітою та науковими дослідженнями в галузі комп'ютерних наук та інформаційних технологій. Метою роботи української частини проектної команди є реалізація принципів рівних можливостей, окреслених Гендерною стратегією ЄС [1], в академічному середовищі українських ВИШів загалом та в Харківському національному економічному університеті імені Семена Кузнеця (ХНЕУ) зокрема. Починаючи з червня 2016 року (дата початку проекту) консорціумом проекту, який налічує 9 освітніх та науково-дослідницьких установ з різних країн Європи, було виконано такі ключові етапи: проаналізовано накопичені в Європі кращі практики щодо розробки та впровадження планів гендерної рівності в університетах; розроблено та впроваджено краудсорсингову онлайн-платформу [3] для збирання ідей щодо подолання гендерної нерівності та методик спільної роботи; адаптовано методик гендерного аудиту для ІТ-факультетів та кафедр; проведено

гендерний аудит в семи університетах Європи, в тому числі в ХНЕУ ім. С. Кузнеця. Досвід університетів, набутий протягом трирічного проекту EQUAL-IST довів, що краудсорсингова платформа є ефективним ІТ-інструментом для проведення гендерного аудиту та забезпечення структурних змін в університетах. Саме тому метою даного дослідження є узагальнення досвіду розробки та використання краудсорсингової онлайн-платформи в якості ІТ-інструменту для підтримки процесів гендерного аудиту університету та процесів розробки «Плану дій щодо підтримки гендерної рівності» (ПГР) в ХНЕУ ім. С. Кузнеця загалом та на кафедрі інформаційних систем зокрема.

Обов'язковою передумовою розроблення якісного ПГР є гендерний аудит. Гендерний аудит (ГА) – це сучасний моніторинговий інструмент просування політики гендерної рівності та формування гендерно-чутливого середовища в організації, де значну роль відіграють не лише кількісні, але й якісні показники. Особливість ГА полягає в тому, що застосовується методика самооцінки, демократичного залучення, коли всі зацікавлені працівники мають змогу висловитися з питань, які стосуються гендерної проблематики. Методика гендерного аудиту, яка використовувалася в рамках проекту EQUAL-IST була заснована на методиці Міжнародної Організації Праці (МОП) [4], однак була істотно адаптована під потреби ІТ-факультетів університетів. Слід зазначити, що досвід адаптації методики МОП до потреб університетів не є новим для України, така робота була проведена Мережею гендерних центрів ВНЗ України [5], а досвід адаптації методики під потреби ІТ-факультетів є інноваційним не лише для України, а й для Європи.

Отже, в основу ПГР ХНЕУ ім. С. Кузнеця було покладено дані внутрішнього ГА, результати дискусій з членами робочої групи, результати аналізу кращих практик освітніх та наукових установ із досягнення гендерної рівності на ІТ-факультетах, а також ідеї, зібрані на краудсорсинговій платформі CrowdEquality [3] - першій Європейській онлайн-платформі, яка збирає проблеми щодо гендерної рівності в університетах, ідеї їх подолання та створює умови для їх обговорення.

### **Основна частина**

Ідея краудсорсингу (англ. crowdsourcing, від crowd - натовп і sourcing - використання ресурсів) не є новою і має на меті залучення на добровільних засадах широкого кола осіб до вирішення проблем інноваційного типу із застосуванням інформаційно-комунікаційних технологій. Серед різновидів

краудсорсингових проектів виділяються платформи для голосування та збору ідей, для обробки наукових матеріалів, для проведення онлайн-конкурсів, для організації спільної роботи з контентом і підтримки колективних інновацій тощо. Особливість платформи CrowdEquality полягає в тому, що вона є першим в Європі краудсорсинговим інструментом для збирання ідей та підтримки дискусії про ідеї та ініціативи з просування гендерної рівності та різноманіття. Ці ідеї апелюють до специфічних завдань, які було визначено у 7 Європейських університетах, що є учасниками проекту EQUAL-IST, і які мали бути вирішені практично. Основною метою платформи є підтримка спільної розробки профільних Планів гендерної рівності. Профільний План гендерної рівності, розроблений для кожного університету-партнера, є документом, що містить плани з реалізації політики і конкретні заходи (дії), що мають бути виконані на практиці з метою досягнення поставлених завдань, що стосуються рівності та різноманіття в університеті. Контент, що розміщено на платформі, доповнює План гендерної рівності. А ПГР містить, в свою чергу, ідеї, генеровані користувачами платформи. Кращі ідеї, визначені за кількістю вподобань та результатами голосування, були включені до профільного Плану гендерної рівності відповідного університету-партнера.

Основний функціонал платформи передбачає можливість завантаження ідей внутрішніми користувачами платформи. При чому, ідеї мають бути віднесені до конкретних тематичних проблем, виявлених у кожному університеті-партнері. Для кожного університету-партнера платформа надає окремий простір, де представлені конкретні проблеми відповідного університету-партнера, а також ідеї та коментарі до цих ідей.

Виклики (проблеми), а також ідеї щодо їх вирішення, визначалися робочими групами з розробки Плану гендерної рівності заздалегідь під час аналізу поточних процесів «як є» у кожному університеті-партнері. Тому кожен виклик (проблема) співвідносився з конкретним університетом-партнером, а кожна ідея – з однією із проблем. Крім того, кожна проблема (та відповідні ідеї щодо її вирішення) була віднесена до однієї (або декількох) областей інтервенції: процеси менеджменту та управління персоналом; навчальний процес; комунікації; проведення наукових досліджень.

В залежності від ролі користувача, для кожного користувача інформація, що зберігається на платформі, є або повністю прихованою, або прихованою частково, або повністю відкритою. Кожна описана на платформі проблема та кожна ідея за замовчуванням позначені як «доступні для

публічного перегляду після знеособлення», окрім випадків, коли користувач встановлює додаткові обмеження. Знеособлення означає, що інформація про автора, що його ідентифікує, не буде видимою будь-яким іншим користувачам.

Перші ідеї, представлені на платформі, були запропоновані робочими групами з розробки ПГР на основі дискусій у командах. Під час завантаження ідеї на платформу, в її описі можна вказати, чи така ж або схожа ідея вже була раніше реалізована іншими університетами, і, якщо це так, надати відповідні посилання або файли (всі файли та посилання узагальнені в базі даних, але також доступні у відкритому доступі). Це дозволяє не лише отримати більше інформації про сутність ідеї, але й надати додаткове обґрунтування, чому була обрана саме така ідея.

Платформа забезпечує зворотний зв'язок (можливість залишати коментарі) щодо: кожної проблеми; кожної ідеї, щодо вирішення проблеми; кожного коментаря.

Також платформа надає можливість «вподобати» конкретну ідею або коментар. Після подання та обговорення достатньої кількості ідей для вирішення кожного завдання, зареєстровані користувачі мали змогу проголосувати за найкращі ідеї.

Під час реєстрації кожному користувачеві система пропонує надати інформацію про його стать, вік, професію (професор, науковий співробітник, член адміністративного персоналу, студент або зовнішній учасник, наприклад, гендерний експерт / експертка) тощо та зображення. Кожній особі, що проходить реєстрацію, пропонується використовувати його / її університетську адресу електронної пошти. Оскільки система зберігає список усіх дозволених доменів електронної пошти для 7 університетів-партнерів, то на основі цієї інформації, встановлюється належність особи до певного університету-партнера та визначається роль користувача і відповідно його права.

Було відокремлено такі ролі користувачів:

1. Члени консорціуму EQUAL-IST та члени робочих груп з розробки Планів гендерної рівності. Користувачі цієї категорії можуть переглядати і робити внесок у зміст усіх розділів платформи.

2. Співробітники / співробітниці та студенти / студентки університету-партнера, які не є членами консорціуму проекту EQUAL-IST. Цю категорію користувачів утворюють особи, знайомі з проблемами, які існують в університеті-партнері, а також із поточною ситуацією в

університеті (які ідеї можливо реалізувати в університеті-партнері, а які - ні). Вони можуть переглянути та зробити свій вклад (ідеї, проблеми, коментарі) у розділі свого університета-партнера.

3. Зовнішні учасники є особами, які мають достатні знання про контекст / конкретні проблеми в університеті-партнері і чия думка є цінною для університета (наприклад, експерти / експертки з питань гендерної рівності).

4. Не зареєстровані користувачі. Ця група охоплює всіх інших людей, які не належать ні до спільноти жодного з університетів-партнерів, ні до зовнішніх учасників. Зовнішні користувачі можуть бачити знеособлений зміст платформи у її публічній зоні, без необхідності реєстрації для його перегляду.

5. Зовнішні користувачі. Зовнішні користувачі мають можливість пройти реєстрацію і потім зробити свій внесок. Після цього вони можуть вподобати ідею або коментар чи проголосувати за ідею. Проте вони не можуть опублікувати новий коментар.

6. Модератор - спеціаліст, відповідальний за ефективність роботи платформи; це роль з найвищим рівнем прав користувача; модератор може переглядати зміст та робити свій внесок у роботу всіх розділів платформи, управляти правами користувачів, відстежувати їх дії та відгуки користувачів, вилучати / експортувати та аналізувати дані, отримані через платформу.

Крім того, модератор має доступ до функцій аналізу даних, а саме:

можливість аналізувати дані, пов'язані з профілями користувачів;

можливість відокремленого аналізу контенту / вподобань / голосування, представлених співробітниками і студентами певного університета-партнера та / або зовнішніми користувачами;

можливість розрахувати показники "популярності платформи" на основі інформації про загальну кількість відвідувань платформи, кількість підписок на оновлення, кількість зареєстрованих користувачів;

можливість аналізувати дані, пов'язані з представленими ідеями та коментарями (з можливістю візуалізувати результати аналізу у таблицях та діаграмах), такі, наприклад, як: кількість запропонованих ідей (в розрізі університетів, користувачів, проблем, з якими співвіднесені ідеї), кількість коментарів (в розрізі університетів, користувачів, проблем, ідей), кількість вподобань (в розрізі коментарів, користувачів, ідей), кількість відвідувань.



Для заохочення користувачів до участі в роботі платформи використовувалися такі інструменти: інструменти гейміфікації, сповіщення, нагорода «найбільш активний користувач місяця».

Приклади екранних форм краудсорсингової платформи наведено на рис. 1 - 4.

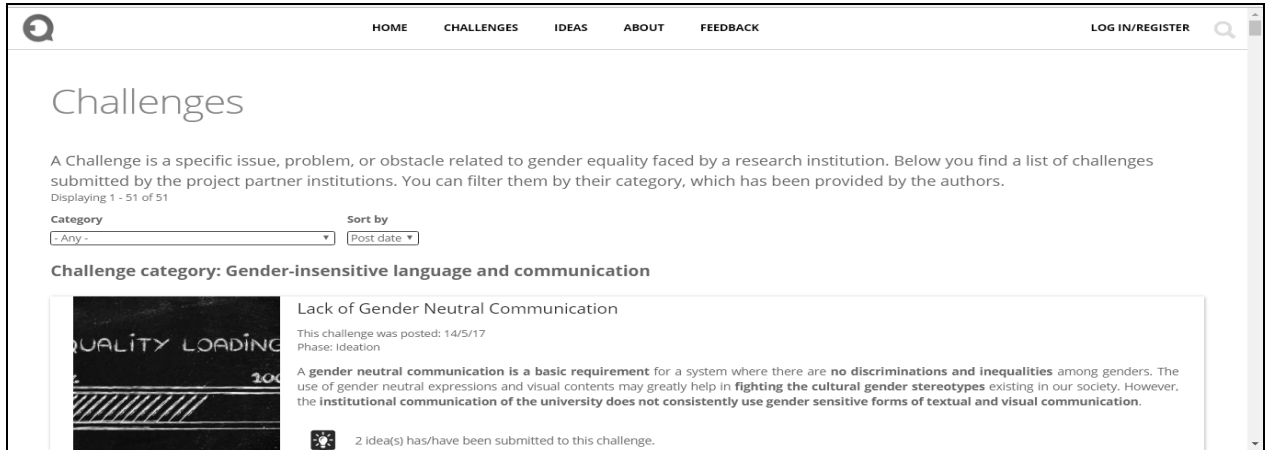


Рис. 1. Перелік проблем гендерної нерівності

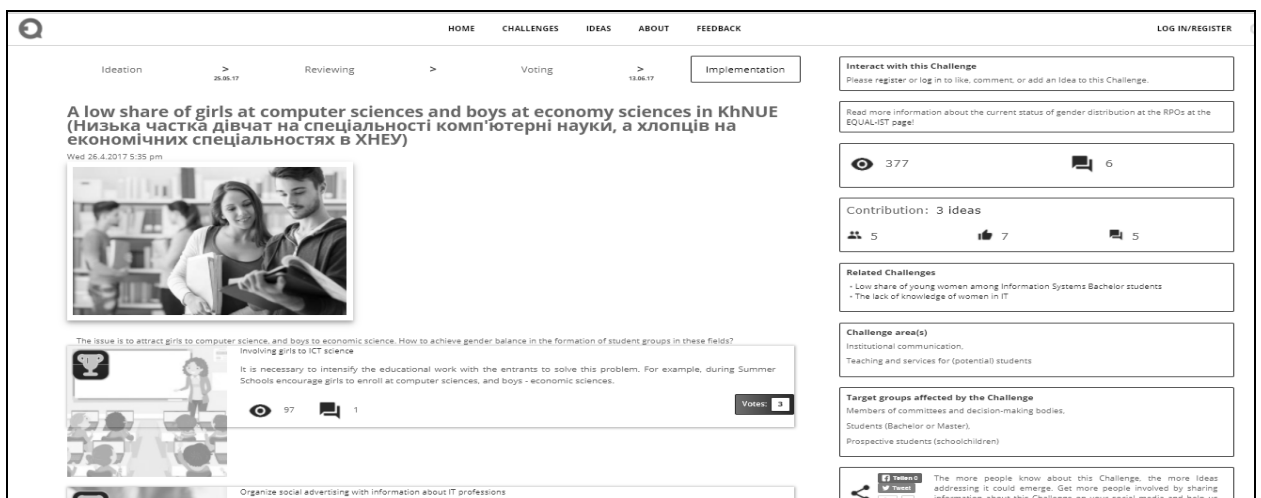


Рис. 2. Опис проблеми

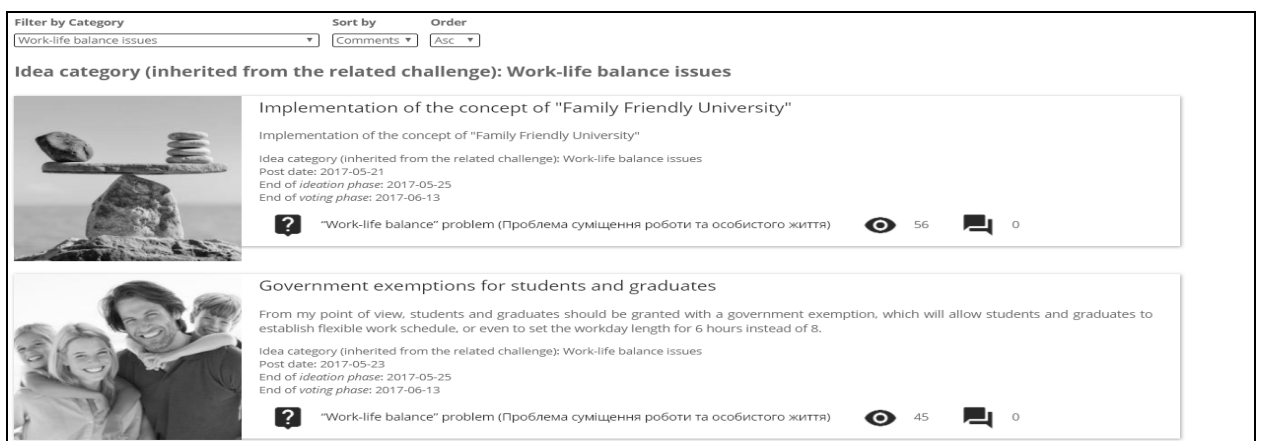


Рис. 3. Перелік ідей для подолання проблем гендерної нерівності



**Рис. 4. Опис ідеї щодо подолання проблем гендерної нерівності**

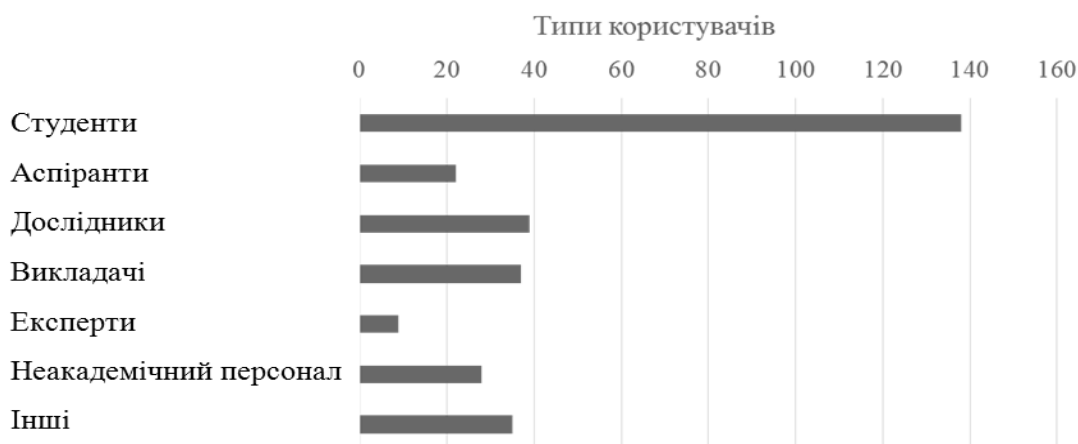
Оскільки така платформа фактично є інструментом для забезпечення процесів розробки ПГР, то на різних етапах роботи платформи, її функціонал змінювався (рис. 5).



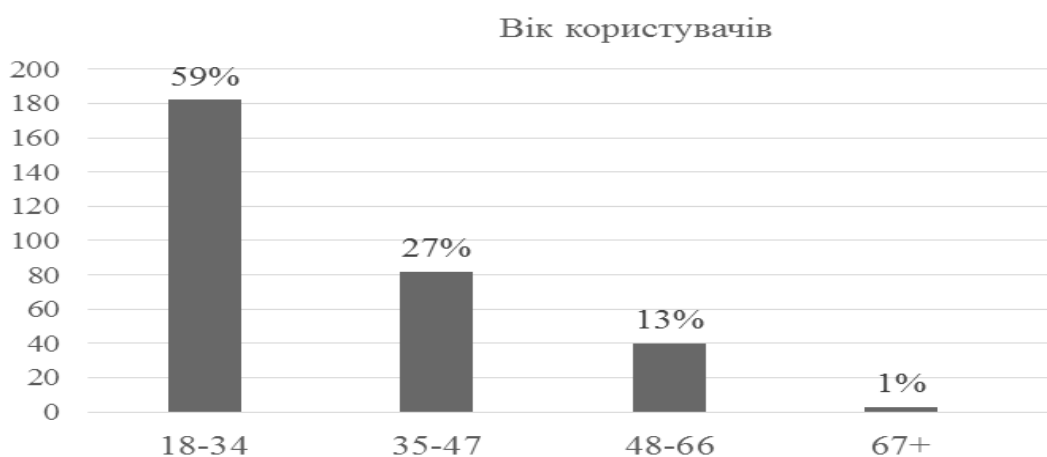
**Рис. 5. Етапи роботи краудсорсингової онлайн-платформи**

Так, якщо на початкових етапах роботи платформи можна було розміщувати проблеми та ідеї для їх подолання, то на етапі «голосування» ці функції були заблоковані, натомість, користувачі мали можливість голосувати за найкращі ідеї. На етапі розробки політики гендерної рівності можна було використовувати накопичені дані: проблеми, ідеї, кількість голосів за кожен ідею тощо.

Загальна кількість користувачів платформи склала 308 осіб. Детальні статистичні дані щодо використання краудсорсингової онлайн-платформи в університетах-партнерах наведені на рис. 6 - 8.



**Рис. 6. Розподіл користувачів платформи за типами**



**Рис. 7. Розподіл користувачів платформи за віком**



**Рис. 8. Розподіл користувачів за організаціями (університетами), які вони представляють:**

UNIMORE - Universita Degli Studi di Modena e Reggio Emilia (Італія)

KhNUE - Харківський національний економічний університет ім. С. Кузнеця (Україна)

WWU - University of Muenster (Німеччина)

KUT - Kaunas University of Technology (Литва)

UNILI - University of Liechtenstein (Ліхтенштейн)

MINHO - University of Minho (Португалія)

UTU - University of Turku (Фінляндія)

ЕХТ – зовнішні користувачі

Статистика в розрізі університетів щодо розміщення користувачами проблем та ідей на платформі наведена в табл. 1 - 2.

**Таблиця 1**

**Статистика щодо розміщення проблем на платформі в розрізі університетів**

Місяць	Квітень				Травень				Червень		Загалом
	1	2	3	4	5	6	7	8	9	10	
WWU	5	2	0	4	0	0	0	0	0	0	11
UNILI	2	0	0	5	0	2	0	0	0	0	9
UTU	2	0	0	0	4	0	0	0	0	0	6
KUT	2	0	0	0	2	2	0	2	0	0	8
UNIMORE	6	0	0	2	0	1	0	1	0	0	10
MINHO	0	0	0	0	0	0	0	0	0	0	0
KhNUE	3	0	1	2	2	2	0	0	0	0	10
Загалом	20	2	1	13	8	7	0	3	0	0	54

**Таблиця 2**

**Статистика щодо розміщення ідей на платформі в розрізі університетів**

Місяць	Квітень				Травень				Червень		Загалом
	1	2	3	4	5	6	7	8	9	10	
WWU	3	3	0	2	0	0	0	0	8	0	16
UNILI	0	0	0	0	0	1	4	7	2	0	14
UTU	0	0	0	0	1	6	0	0	0	0	7
KUT	0	0	0	0	1	5	0	2	0	0	8
UNIMORE	4	4	2	0	2	2	0	0	0	0	14
MINHO	0	0	0	0	0	0	0	0	0	0	0
KhNUE	2	13	2	3	0	1	3	7	1	0	32
Загалом	9	20	4	5	4	15	7	16	11	0	91

Крім загальної статистики використання платформи зовнішніми та внутрішніми користувачами, система накопичувала інформацію щодо результатів голосування за кращі ідеї. Для кожної ідеї збиралася така інформація:

Скільки загалом голосів отримала ідея?

Скільки загалом користувачів брали участь в голосуванні?

Скільки в голосів ідея отримала від внутрішніх користувачів?

Скільки в голосів ідея отримала від зовнішніх користувачів?

В результаті гендерного аудиту та використання краудсорсингової онлайн-платформи було виявлено та проведено ранжування основних проблем в області гендерного паритету як в ХНЕУ ім. С. Кузнеця, так і в інших університетах консорціуму проекту EQUAL-IST. За результатами першого етапу робіт, було виявлено такі основні проблеми для подальшого опрацювання:

- 1) дисбаланс між роботою / навчанням та особистим життям;
- 2) вертикальна сегрегація;
- 3) горизонтальна сегрегація;
- 4) відсутність адміністративно-правових механізмів підтримки гендерної рівності в університеті;
- 5) проблеми гендерної нерівності не беруться до уваги в процесі прийняття рішень;
- 6) відсутність гендерної культури та обізнаності щодо питань гендерної рівності;
- 7) гендерний дисбаланс серед студентів / студенток;
- 8) гендерно-нечутливі комунікації.

Для вирішення виявлених проблем в ПГР було запропоновано здійснити ряд дій, спрямованих на два рівні імплементації – загальноуніверситетський та кафедральний (кафедра інформаційних систем).

Крім того, на платформі CrowdEquality було зібрано більше 30 ідей щодо подолання виявлених проблем в ХНЕУ ім. С. Кузнеця. Кращі ідеї, запропоновані на платформі, були використані в університеті на етапі розробки політики гендерної рівності (а саме, при розробці планів гендерної рівності), причому всі ідеї були розподілені на чотири категорії відповідно до розділів плану: процеси менеджменту та управління персоналом; навчальний процес; комунікації; проведення наукових досліджень.

Причому, в якості основного джерела ідей для планування дій виступила саме краудсорсингова платформа (табл. 3).

**Таблиця 3**

**План дій для досягнення гендерного паритету в ХНЕУ ім. С. Кузнеця**

Заплановані дії	Джерело інформації
<i>Сфера інтервенції: Процеси менеджменту та управління персоналом</i>	
<b>Проблема 1:</b> Дисбаланс між роботою / навчанням та особистим життям і відсутність можливостей для підтримки балансу	
<b>Ціль 1:</b> Досягти балансу між роботою / навчанням та особистим життям серед персоналу та студентства	
<b>Дія 1.</b> Запровадити положення щодо забезпечення пріоритетних умов при складанні розкладу занять для викладачів / викладачок, що мають маленьких дітей, дітей з особливими потребами, багатодітних родин, вагітних жінок	Краудсорсингова платформа
<b>Дія 2.</b> Провести дослідження щодо виявлення потреб у ресурсах для відкриття дитячої кімнати в університеті	Краудсорсингова платформа:
<b>Дія 3.</b> Впровадити положення щодо забезпечення дистанційної роботи для викладачів/викладачок, що мають маленьких дітей, дітей з особливими потребами, багатодітних родин, вагітних жінок	Зустріч Робочої групи
<b>Дія 4.</b> Розвивати культуру рівних сімейних обов'язків серед студентства і викладацького складу	Зустріч Робочої групи
<b>Дія 5.</b> Створити і підтримувати розділ на веб-сайті університету з інформацією щодо проблем досягнення гендерної рівності, балансу роботи та особистого життя, регулювання сфери гендерної рівності	Зустріч Робочої групи
<b>Проблема 2:</b> Проблема «скляної стелі» (вертикальної сегрегації)	
<b>Ціль 2:</b> Заохочувати досягнення гендерної рівності (вертикальний принцип) за допомогою інформаційно-просвітницької діяльності	
<b>Дія 6.</b> Проводити семінари та круглі столи з питань гендерної рівності для управлінського персоналу	Краудсорсингова платформа
<b>Дія 7.</b> Зробити процедури та критерії відбору персоналу прозорими	Краудсорсингова платформа
<b>Проблема 3:</b> Концентрація жінок в певних «жіночих» секторах зайнятості (горизонтальна сегрегація)	
<b>Ціль 3:</b> Заохочувати досягнення гендерної рівності (горизонтальний принцип) за допомогою інформаційно-просвітницької діяльності	
<b>Дія 8.</b> Проводити інформаційно-просвітницьку роботу з питань гендерної рівності для персоналу	Краудсорсингова платформа
<b>Проблема 4:</b> Відсутність механізмів підтримки гендерної рівності	
<b>Ціль 4:</b> Забезпечити стабільність дій в межах ППР шляхом створення механізмів досягнення гендерної рівності	
<b>Дія 9.</b> Створити Комісію з вирішення проблем гендерної нерівності та розробити механізми її функціонування	Краудсорсингова платформа
<b>Проблема 4:</b> Відсутність механізмів підтримки гендерної рівності	
<b>Ціль 5:</b> Забезпечити стабільність дій в межах ППР шляхом включення цілей та заходів з досягнення гендерної рівності в стратегічні документи університету	
<b>Дія 10.</b> Забезпечити підтримку співробітницями і	Зустріч Робочої групи та

співробітниками університету окремих положень Європейської хартії для дослідників	адміністрації університету
<b>Дія 11.</b> Запропонувати та впровадити зміни до чинної Колективної угоди між адміністрацією університету та профспілковою організацією з метою забезпечення реалізації ПГР	Зустріч Робочої групи і Голови Первинної профспілкової організації університету
<b>Проблема 5:</b> Проблеми гендерної рівності не беруться до уваги в процесі прийняття рішень <b>Ціль 6:</b> Заохотити осіб, що приймають рішення, до підтримки ідей та принципів гендерної рівності	
<b>Дія 12.</b> Зібрати, проаналізувати та оприлюднити статистичні дані з урахуванням гендерної компоненти в університеті	Зустріч Робочої групи
<i>Сфера інтервенції: Навчальний процес</i>	
<b>Проблема 6:</b> Відсутність гендерної культури та обізнаності щодо питань гендерної рівності <b>Ціль 7:</b> Підвищити обізнаність щодо питань гендерної рівності, сформувати гендерну культуру наукових співробітників/ співробітниць та студенток/ студентів	
<b>Дія 13.</b> Провести інформаційно-просвітницькі заходи з питань гендерної рівності для студентства та викладацького складу	Краудсорсингова платформа
<b>Дія 14.</b> Розробити та розповсюдити друковані та електронні інформаційно-просвітницькі матеріали про гендерну рівність	Зустріч Робочої групи
<b>Проблема 6:</b> Відсутність гендерної культури та обізнаності щодо питань гендерної рівності <b>Ціль 8:</b> Використовувати гендерний підхід при розробці навчальних матеріалів	
<b>Дія 15.</b> Провести пілотну антидискримінаційну експертизу навчальних матеріалів	Зустріч Робочої групи
<b>Проблема 7:</b> Гендерний дисбаланс серед студентства економічних спеціальностей та факультету економічної інформатики <b>Ціль 9:</b> Покращити показники гендерного балансу серед студентства економічних спеціальностей та спеціальності «Комп'ютерні науки»	
<b>Дія 16.</b> Провести заходи з метою підвищення обізнаності та інформованості про жінок в ІТ-індустрії та можливості їх кар'єрного зростання	Краудсорсингова платформа
<b>Дія 17.</b> Заснувати спеціальні номінації для жінок/ жіночих команд в існуючих чемпіонатах та змаганнях в ІТ-сфері	Краудсорсингова платформа
<b>Дія 18.</b> Провести пілотну антидискримінаційну експертизу маркетингових та рекламних матеріалів для бакалаврських та магістерських програм в ІТ-сфері	Краудсорсингова платформа
<i>Сфера інтервенції: Комунікації</i>	
<b>Дія 19.</b> Забезпечити комунікаційну та візуальну підтримку гендерної рівності в медіа-контенті університету	Краудсорсингова платформа

Протягом 2017-2019 навчального року в ХНЕУ ім. С. Кузнеця відбувалася реалізація ПГР, в рамках якої регулярно проводилися семінари, відкриті лекції, дебати, ділові ігри та інші заходи, спрямовані на виявлення та подолання стереотипів, гендерної нерівності та дискримінації в сфері освіти та науки, на популяризацію ІТ-кар'єри серед дівчат, на пропаганду культури рівних обов'язків в родині. Розпочата пілотна антидискримінаційна

експертиза навчальних матеріалів і реклами ІТ-спеціальностей; університеті створено Комісію з питань гендерної рівності та протидії дискримінації. Окрема увага приділялася забезпеченню балансу між роботою та особистим життям співробітників і співробітниць.

### **Висновки**

Таким чином, досвід університетів консорціуму проекту EQUAL-IST загалом, та ХНЕУ ім. С. Кузнеця зокрема, довів, що краудсорсингова онлайн-платформа [crowdequality.eu](http://crowdequality.eu) виявилася ефективним ІТ-інструментом для залучення в процес розробки Плану гендерної рівності університетської спільноти в широкому сенсі – студентів, викладацького складу, адміністративного персоналу, а також засобом для виконання низки інших важливих завдань: оцінки, моніторингу, налагодження комунікацій, впровадження гендерно-чутливих методів роботи, просвітництва. Досвід використання краудсорсингової платформи та збирання й аналізу ідей щодо подолання проблем гендерної нерівності є інноваційним не лише для України, а й для Європи. Крім того, ХНЕУ ім. С. Кузнеця став піонером серед українських університетів не лише в питаннях використання подібного ІТ-інструмента, а також і в питаннях розробки та реалізації Плану гендерної рівності. Платформу CrowdEquality рекомендовано для подальшого використання європейськими університетами для підтримки як процесів початкового аналізу проблем гендерного паритету, так і для підтримки процесів розробки та впровадження ПГР. Подальша робота проектною командою EQUAL-IST буде пов'язана саме з поширенням серед вишів набутого досвіду щодо використання даної платформи та щодо процесів розробки та реалізації ПГР.

### **Література:**

1. Strategic engagement for gender equality 2016-2019 [Електронний ресурс] / Сайт ЄС. – Режим доступу: [http://ec.europa.eu/justice/gender-equality/document/files/strategic\\_engagement\\_en.pdf](http://ec.europa.eu/justice/gender-equality/document/files/strategic_engagement_en.pdf)
2. Про проект EQUAL-IST [Електронний ресурс] / Офіційний сайт проекту EQUAL-IST. – Режим доступу: <https://equal-ist.eu/overview/>



3. Краудсорсингова онлайн-платформа [Електронний ресурс] / CrowdEquality - the Idea Crowdsourcing Platform for Promoting Gender Equality and Diversity. – Режим доступу: [www.crowdequality.eu](http://www.crowdequality.eu)

4. A manual for gender audit facilitators: The ILO participatory gender audit methodology, 2nd Edition. [Електронний ресурс] / International Labour Office. – Режим доступу: [http://www.ilo.org/wcmsp5/groups/public/---dgreports/---gender/documents/publication/wcms\\_187411.pdf](http://www.ilo.org/wcmsp5/groups/public/---dgreports/---gender/documents/publication/wcms_187411.pdf)

5. Гендерний аудит діяльності вищих навчальних закладів. Практичний посібник. Світайло Н., Савельєва Ю., Давліканова О. [Електронний ресурс] / Фонд імені Фрідріха Еберта. – Режим доступу: <http://fes.kiev.ua/n/cms/fileadmin/upload2/Gender.pdf>.

## ГЛИАБА 6

### FORECAST ERROR MEASURES: CRITICAL REVIEW AND PRACTICAL RECOMMEND INTRODUCTION AND TASK STATEMENT

The urgency of the task of forecasting time series (CR) is particularly increasing recently, when powerful computing tools for collecting and processing information [1-4] became available.

To estimate the prediction quality, different criteria are used: estimates of forecast errors, speed of calculation, interpretation, etc. In the study of CR, the reliability of the selected forecasting model is estimated by a periodic comparison of the actual and predicted values of a series. When solving practical problems, the analysis of a prediction error or forecast accuracy is considered more significant.

Absolute accuracy of the forecast is not accepted, because for each particular task, it can radically different, so estimate the forecast error, which is characterized by some metric.

Given the difficulties of forecasting and errors in forecasts, there is a need to develop recommendations for choosing a quality prediction estimation made on the basis of the analyzed literature.

In this paper, the results of the analysis of existing and fairly common estimates of forecasting errors (metrics) used in research on prediction are given. When presenting the material, the classification of criteria (metrics) used in [5] is used.

#### **Criteria based on absolute mistakes**

Mean Forecast Error (MFE):

$$MFE = \frac{1}{N} \sum_{i=1}^N e_i. \quad (1)$$

Mean Absolute Error(MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |e_i| = \underset{i=1,N}{mean} |e_i| \quad (2)$$

determines the magnitude of the forecast error as a whole.

Like the MFE, this metric also depends on the scale of measurement and the use of conversion data.

There are two drawbacks in the MAE metrics, which are manifested only in the case of analyzing the predictive accuracy of the models:

- 1) its meaning is difficult to interpret;
- 2) using this metric, you can compare only between models from a single row of data. In practice, the task often is to understand how a model of several lines of data is in itself compared to other models. However, the values obtained from different rows MAE can not be combined with each other.

Mean Squared Error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N e_i^2. \quad (3)$$

This metric takes into account the presence of rare extreme mistakes.

Sum of Squared Error (SSE):

$$SSE = \sum_{i=1}^N e_i^2. \quad (4)$$

This metric has similar characteristics to MSE.

Signed Mean Squared Error (SMSE):

$$SMSE = \frac{1}{N} \sum_{i=1}^N \left( \frac{e_i}{|e_i|} \right) e_i^2. \quad (5)$$

This metric depends on the scale of the measurement and transformation of input data and normalizes the received value of MSE.

Square Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}. \quad (6)$$

This metric has similar characteristics to MSE.

Fourth Root Mean Quadrupled Error (FRMQE, R4MS4E):

$$FRMQE = R4MS4E = \sqrt[4]{\sum_{i=1}^N e_i^4}. \quad (7)$$

Normalized Average Quadratic Error (NMSE):

$$NMSE = \frac{MSE}{\sigma^2} = \frac{1}{N\sigma^2} \sum_{i=1}^N e_i^2. \quad (8)$$

This metric normalizes the received MSE value.

Median Absolute Error (MdAE):

$$MdAE = \underset{i=1,N}{median} |e_i| \quad (9)$$

Standard Deviation (SD):

$$SD = \sum_{i=1}^N \sqrt{\frac{(y_i - ME)^2}{N}}. \quad (10)$$

Absolute error measurement has the following disadvantages:

- 1) Dependence on scale [5] (therefore, if the forecast task includes objects with different scales or values, an absolute error as a metric can not be applied);
- 2) high impact of emissions in the estimation of the forecast efficiency (so if the data contain emissions, then the absolute values of the errors are ineffective).

RMSE, MSE have low reliability (robustness): the results may vary depending on the different parts of the data.

### **Criteria based on the calculation of interest rates**

Median Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{e_i}{y_i} \right| \times 100. \quad (11)$$

As you can see, MAPE has no dimension, with a very simple interpretation. It can be measured in fractions or percentages, it is easy to compose from different rows. You can even calculate MAPE and examine its distribution using statistical analysis tools.

As a result of a detailed comparison of MAPE with other metrics when choosing a prediction model in a series of papers, it has been shown that MAPE is an unreliable criterion when using this metric to choose between competing predictive models. In particular, there is a high probability that MAPE will choose a model that gives a rating below the average (that is, "underestimates") of the predicted average model, MAPE systematically and consistently values models that give too optimistic prognosis. By dividing the actual values, this metric was sensitive to scale, and with a small value of the actual value of the MAPE can be very large.

Average Mean Arctangent Absolute Percentage Error (MAAPE):

$$MAAPE = \frac{1}{N} \sum_{i=1}^N AAPE_i, \quad (12)$$

where

$$AAPE_i = \arctg \left( \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right).$$

This metric allows you to slightly improve the properties of MAPE.

Mean Percentage Error (MPE):

$$MPE = \frac{1}{N} \sum_{i=1}^N \left( \frac{e_i}{y_i} \right) \times 100 \quad (13)$$

is designed to evaluate not the accuracy of the forecast, but its displacement.

One can notice that it is similar to MAPE, has the same advantages and disadvantages as MAPE, but differs only in the absence of modules in the numerator. MPE shows the percentage offset of the forecast.

This metric is a fairly good indicator of a complete forecast error, but it is not so intuitive as unlike other metrics.

Median Absolute Percentage Error, MdAPE:

$$MdAPE = \underset{i=1,N}{\text{median}} |\bar{e}_i|, \quad (14)$$

where

$$\bar{e}_i = e_i / y_i.$$

Square Root Medium Square Percentage Error (RMSPE):

$$RMSPE = \sqrt{\underset{i=1,N}{\text{median}} (\bar{e}_i)^2}. \quad (15)$$

For this group of criteria, the following disadvantages are typical:

1) there is a division to zero, when the actual value is zero;  
 2) asymmetry - the value of the errors is different, if the predicted value is greater or less than actual [2,5];

3) emissions significantly affect the result, especially if the deviation is much larger than the maximum value of "normal" cases [4];

4) the calculation of the error is biased, which may lead to an incorrect estimation of the forecasting model.

### **Symmetric criteria**

Cymetric MAPE (Symmetric MAPE, SMAPE):

$$SMAPE = \frac{1}{N} \sum_{i=1}^N \frac{2|e_i|}{|y_i| + |\hat{y}_i|} \quad (16)$$

uses the arithmetic mean of actual and projected values, applies to time series whose actual values are significantly greater than 1.

Despite the fact that the metric should be symmetric, but it does not end there.

Metering Median APE (Symmetric Median APE, SMdAPE):

$$SMdAPE = \underset{i=1,N}{median} \left| \frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)} \right|. \quad (17)$$

Modified MAPE (Modified sMAPE, msMAPE):

$$msMAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2} + s_i} \quad (18)$$

$$s_i = \frac{1}{i-1} \sum_{k=1}^{i-1} |y_i - \bar{y}_i|, \bar{y}_i = \frac{1}{i-1} \sum_{k=1}^{i-1} y_k. \quad (19)$$

This group of criteria has the following drawbacks:

- 1) regardless of the name, they are asymmetrical [5];
- 2) if the actual value is equal to the forecast but with the opposite sign, or both of these values are zero, then there is a division into zero;
- 3) the values of the criteria significantly affect the emissions in the same way as the percentage errors;
- 4) There is a problem interpreting the results.

### **Relative criteria**

Average Relative Absolute Error (MRAE):

$$MRAE = \text{mean} \frac{|e_i|}{(y_i - \hat{y}_i^*)}, \quad (20)$$

where  $\hat{y}_i^*$  - predictive value of the base (reference) model (in practice, usually choose  $\hat{y}_i^* = y_{i-1}$ ).

Median Relative Absolute Error (MdRAE):

$$MdRAE = \text{median}_{i=1,N} \left| \frac{|e_i|}{(y_i - \hat{y}_i^*)} \right|. \quad (21)$$

Geometric Mean Relative Absolute Error (GMRAE):

$$GMRAE = \sqrt[m]{\prod_{i=1}^m \frac{|y_i - \hat{y}_i|}{|y_i - \tilde{y}_i|}} \quad (22)$$

where  $y_i - \tilde{y}_i$  is the mistake of some other model.

GMRAE shows how many times the constructed model was worse (or better) than the one chosen for comparison (model-benchmark). If  $GMRAE > 1$ , then the chosen model was less accurate, in the opposite situation - more accurate. However, the choice of the second model remains. GMRAE can not be used in situations where at least one of the mistakes of one of the models has turned out to be zero - in this case, GMRAE becomes equal or zero or infinity.

Based on the above formulas, the division into a zero error occurs if the predictive value is obtained using the model, is equal to the actual value.

**Scaled bugs**

Average Absolute Scaled Error (MASE):

$$MASE = \frac{T-1}{N} \frac{\sum_{i=1}^N |e_{T+i}|}{\sum_{j=2}^T |y_j - y_{j-1}|}. \quad (23)$$

The sum in the numerator corresponds to the test sample, and in the denominator - the training. The second is, in fact, the average absolute error forecast for the Naive method. It equals the average absolute deviation of the series in the first differences. This value, in essence, shows how the training sample is predictable and can be equal to zero only when all values in the training sample are

equal to each other, which corresponds to the absence of any changes in a series of data, that is, the situation is practically impossible in practice.

It should be noted that the module in the denominator (23) is needed for situations in which both positive and negative values of variables are present in the prediction, the average value of which can be close to zero.

Average absolute scalable error (sMAE - scaled MAE, sometimes referred to as MAE / mean):

$$sMAE = \frac{1}{N} \sum_{i=1}^N \frac{|e_i|}{y_i}. \quad (24)$$

Note that the sMAE denominator can be zero in only one case - if all the actual values in the training sample are zero. The sMAE criterion is practically the ideal indicator. He has only one small disadvantage - he is not limited to the top.

Root Square Mean Square Scaled Error (RMSSE):

$$RMSSE = \sqrt{\text{mean}\left(\frac{|e_i|}{\frac{1}{T-1} \sum_{j=2}^T |y_j - y_{j-1}|}\right)^2}. \quad (25)$$

Relative criteria are symmetric and resistant to emissions, but have the following disadvantages:

- if the real values in the forecast horizon are equal, then there is a division into zero.
- a slight bias of estimates is possible.

### **Criteria based on the calculation of relative variables**

Relative Mean Absolute Error, RelMAE:

$$RelMAE = \frac{MAE}{MAE^*}, \quad (26)$$

where MAE and MAE \* mean absolute errors for the analyzed forecasting model and the base (reference) model, respectively, calculated by the formula (2).

Relative Square Root Mean Squared Error (RRMSE):



$$RRMSE = \frac{RMSE}{RMSE^*}. \quad (27)$$

Log mean square error error (LMR):

$$LMR = \log\left(\frac{RMSE}{RMSE^*}\right). \quad (28)$$

These criteria determine the number of cases where the prediction model analyzed exceeds the base, but does not estimate the difference value.

The disadvantages of these criteria are the possibility of division into a zero error if the base predictive error is zero.

### **Other criteria**

Criterion of direction accuracy (direction accuracy, DA):

$$DA = \frac{1}{N} \sum_{i=1}^N a_i, \quad (29)$$

$$\text{where } a_i = \begin{cases} 1, & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - \hat{y}_i) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Modified direction accuracy (MDA):

$$MDA = \begin{cases} \frac{\sum_{i=0}^{N-1} D_i}{N-1}, \\ \end{cases} \quad (30)$$

$$D_i = (A_i - F_i)^2$$

$$A_i = 1, \text{ if } y_{i+1} - y_i \leq 0,$$

$$A_i = 0, \text{ if } y_{i+1} - y_i > 0$$

$$F_i = 0, \text{ if } \hat{y}_{i+1} - \hat{y}_i \leq 0,$$

$$F_i = 0, \text{ if } \hat{y}_{i+1} - \hat{y}_i > 0.$$

Directional symmetry (DS):

$$DS = \frac{100}{N} \sum_{i=1}^N d_i, \quad (31)$$

$$\text{where } d_i = \begin{cases} 1 & \text{if } (y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

DS estimates the accuracy of the predicted direction.

Weighted directional symmetry (WDS):

$$WDS = \frac{\sum_{i=1}^N d_i |y_i - \hat{y}_i|}{\sum_{i=1}^N d'_i |y_i - \hat{y}_i|}, \quad (32)$$

$$\text{where } d'_i = \begin{cases} 1 & \text{if } (y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

WDS measures the magnitude and direction of forecasting error. In this case, the errors associated with the wrongly predicted direction are penalized and the correct predicted direction is encouraged. The smaller the value of WDS, the better the prediction in terms of its accuracy and direction.

Sign criterion

$$Sign = \frac{1}{N} \sum_{i=1}^N z_i, \quad (33)$$

$$\text{Where } z_i = \begin{cases} 1, & \text{if } (y_{i+1})(\hat{y}_{i+1}) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Akaike Information Criteria (AIC):

$$AIC = \log MSE + \frac{2q}{N}, \quad (34)$$

where  $q$  is the total number of model parameters;  $N$ - number of observations.

Features of AIC:

- 1) it depends on the assumption of normal data;
- 2) this criterion is not feasible; that is, with increasing data it is prone to the so-called overfitting;
- 3) the criterion AIC in the general case is asymptotically effective.

### Bayesian Information Criteria (BIC):

$$BIC = \log MSE + q \frac{\log N}{N}. \quad (35)$$

The features of this criterion are as follows:

1) this criterion introduces more stringent penalty characteristics than AIC; that is, BIC "fines" models with a large number of parameters to a greater extent than does AIC;

2) the BIC criterion is a strictly valid criterion: if the data is actually generated by a model that is adequate to the real one, then the BIC will select a true model with unit probability, taking into account the fact that the size of the input data will tend to be infinite;

3) BIC is less effective than AIC.

### Kullback Information Criterion (KIC):

$$KIC = N \ln \left[ \frac{1}{N} \sum_{i=1}^N e_i^2 \right] + 3S \ln N. \quad (36)$$

This criterion is less used than others.

### Hennan-Quinn criterion (HQ)

$$HQ = N \ln \left[ \frac{1}{N} \sum_{i=1}^N e_i^2 \right] + 2S \ln(\ln N). \quad (37)$$

It should be noted that in the particular case of a normal distribution, the criterion AIC is applied in the form

$$AIC(S) = RSS(S) + 2S, \quad (38)$$

the so-called Akaike-Mellous criterion.

### Weighted Information Criteria (WIC)

Several papers described attempts to combine statistical and informational criteria in one weighted criterion

$$WIC = 0,1(AIC + BIC) + 0,2(RMSE + MAPE) + 0,2((1 - DA) + MDA). \quad (39)$$

It should be noted that the coefficients used in this criterion were asked heuristically not clear for any reason.

### Adaptive Weighted Information Criteria (AWIC)

In the development of this idea, it was suggested instead of fixed use coefficients, which are calculated by solving some optimization problem (with the criteria and taken with a weight of 0.1). The corresponding criterion was called adaptive (Adaptive WIC, AWIC)

$$AWIC = 0,1(AIC + BIC) + w_1RMSE + w_2MAPE + w_3(1 - DA) + w_4MD. \quad (40)$$

It should, however, be noted that at the same time

- 1) the criterion for weighting coefficients AIC and BIC is not defined;
- 2) there are no guidelines or recommendations for the use of specific statistical criteria that take into account the time series behavior;
- 3) heuristic methods have not been successful in the assessment of weights for combined methods;
- 4) this criterion does not take into account the characteristics of time series.

#### Tale's U-statistics

This important measure is calculated by the formula:

$$U = \frac{\sqrt{\sum_{i=1}^N (FPE_i - APE_i)^2}}{\sqrt{APE_i^2}}, \quad (41)$$

where  $FPE_i = \frac{\hat{y}_i - y_i}{y_i}$  is the relative change in forecast data;

$APE_i = \frac{y_{i+1} - y_i}{y_i}$  is relative change of factual data

or

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \hat{y}_i^2} \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}}. \quad (42)$$

This indicator allows comparisons of formal methods of forecasting with the simplest ("naive") methods.

As can be seen from expression (42), by raising relative errors to the square, the effect of mutual compensation of errors is eliminated and the effect is provided when significant errors are more important than small ones.

As a basis for comparison, two simple methods are used:

a) the method of "random walk" (random walk) - is a mathematical model of the process of random changes - steps in discrete moments of time. It is assumed that the change at each step does not depend on previous changes and on time;

b) the method of seasonal consideration - this method considers the possibility of seasonality in the time series. Since the high percentage of data fluctuations within the time series is due to seasonality, this method may have better results than the first, while remaining fairly simple.

Taylor's U-statistics fluctuates within the range  $[0; \infty]$  and is easy to interpret:

- if  $U = 1$ , the basic basic methods are as precise as the formal method of forecasting used;

- if  $U < 1$ , the formal prediction method used is better than the basic methods.

The approximation of the value of U-statistics to 0 indicates an increase in prediction quality compared to the basic methods;

- if  $U > 1$ , it makes no sense to use the appropriate formal forecasting method, since the basic methods give better results.

Taylor's U-statistics can be used to estimate the quality of forecasts obtained on the basis of different methods and models.

Can be used in assessing the quality of forecasts obtained on the basis of different methods and models.

Disadvantages:

1) is dependent on scale and data conversion;

2) its value significantly depends on the variance of the predicted series (when predicting processes with low dispersion under a clearly unsuccessful prediction, U-statistics may have a small value due to a low variation of the attribute).

Modified Tail U-statistics:

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \hat{y}_i^2} \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}} \frac{\sigma_y^2}{\sigma_{\hat{y}}^2}, \text{ if } \sigma_y^2 > \sigma_{\hat{y}}^2, \quad (43)$$

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \hat{y}_i^2} \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}} \frac{\sigma_{\hat{y}}^2}{\sigma_y^2}, \text{ if } \sigma_y^2 < \sigma_{\hat{y}}^2 \quad (44)$$

aims to eliminate the lack of traditional statistics.

We reviewed some important metrics (metrics) for estimating the accuracy of the forecast. Each of these measures has some unique properties that are different from others. In experiments it is better to use several performance criteria, which will help gain knowledge about the amount, size and direction of the total predicted error. In addition, there are still plenty of metrics that are some of the modifications discussed above, such as the Mean Relative Absolute Error

(MRAE), Median Absolute Percentage Error, MdAPE, Median Relative Absolute Error (Median Relative Absolute Error, MdRAE), Root Mean Square Percentage Error (RMSPE), Relative Mean Absolute Error, RelMAE, Relative Root Mean Square Error, RelRMSE, Symmetric Median APE (Symmetric MeDiA Mean Absolute Percentage Error, sMdAPE) and other]

### **Analysis of recommendations for choosing prediction estimates**

One of the most difficult problems in forecasting is the question of choosing the most appropriate measure (criterion) for estimating the quality of the forecast.

At this time, there are quite a few studies devoted to the analysis of the criteria for forecasting quality and the development of recommendations for their effective choice. Let's consider some of them.

If the forecast estimation is carried out for one-dimensional time series and preliminary data processing (data cleansing, detection of anomalies) has been performed, it is wise to choose MAE, MdAE, RMSE. When using different scales, these criteria do not apply.

As for the comparisons of MSE and MAE criteria, it should be noted that there is only one situation in which MSE-based errors can be preferred to MAE-based errors-this is an estimate of integer demand. In this case, the number of zeros in the sample can be so large (it can easily be transported over 50%), that all errors based on the modules will prefer a zero forecast (for example, when selling it means that we do not sell anything, therefore, nothing to store - depressive prognosis). This is due to the fact that the minimum MAE corresponds to the median value, while at least the MSE corresponds to the average.

Both of these MAE and MSE estimates are well suited, for example, to determine the accuracy of the prediction of the same sequence for different model parameters or when using different models, but they show lower forecast errors than with the naive method. That is, the accuracy of the results of the forecast exceeds the accuracy of the naive method. Conversely, if the values of relative estimates are greater than one, then the accuracy of the results of the forecast is, on average, worse than with the naive method of forecasting.

These estimates are equally suited to determine the accuracy of the forecast for two or more steps forward. To do this, you only need to use the value of forecast errors for the appropriate number of steps forward instead of a one-step forecast error.

If the set data does not contain emissions and the time period for which the calculations were made are not so large that they are not close to the data value, it is expedient to use scalable estimates.

If the data has seasonal or cyclic patterns, it is recommended to use the normalized errors when the normalization rates can be calculated within an interval equal to the cycle or season.

When estimating the accuracy of several forecasts, a method of aggregation is required. If you accept the arithmetic mean, then you should use the average absolute percentage error (MAPE).

For a variety of cases, you can provide the following recommendations.

Despite the fact that percentage errors are commonly used for prediction in real problems, asymmetry does not recommend using them. If the range of values lies in the positive half-plane, and in the data there are no emissions, it is expedient to use the symmetrical measures of error.

If the data is "polluted", that is, it contains emissions, it is expedient to apply large-scale measures such as MASE, inRSE. In this case, it is necessary that

- 1) the prediction horizon was sufficient;
- 2) there were no identical values predicting my rank;
- 3) the value of the normalized coefficient should not be equal to zero.

After calculation, the results are analyzed for the division into zero errors and the presence of contradictions.

Due to the fact that various estimates of data yield different estimates of prediction errors, in the absence of the results of the preliminary analysis and the need to generalize the methods instead of one metric of forecasting, it is wise to use a certain set of error metrics.

It should be noted that sometimes in the literature there is a comparison of models based on RMSE. Such a comparison in the general case is incorrect and usually leads to the selection of the wrong, that is, not the most accurate model. This is due to the fact that the RMSE is highly exposed to emissions (through the square in the formula). Therefore, as a result of using this criterion, preference is given to less accurate models that indirectly predict the data, but do not allow any major mistake. Obviously, any coefficients based on RMSE will also result in distorted results.

Due to the fact that different estimates of prediction errors are obtained on different sets of data, and each criterion has deficiencies that may lead to inaccurate prediction quality prediction, one universal criterion can not be identified. Therefore, it is wise to use a certain set of error metrics.

## LITERATURE

1. Bowerman B. Forecasting: methods and applications. 4<sup>th</sup> ed. / B. Bowerman, R. O'Connell, A. Koehler. – Belmont, CA.: Thomson Brooks, 2005. – 384 p.
2. Hyndman R.J. Forecasting: principles and practice. / R. J. Hyndman, G. Athanasopoulos. // 2<sup>nd</sup> ed, 2017. OTexts. <http://otexts.com/fpp>
3. Gooijer J.G.D. 25 Years of Time Series Forecasting. / J.G.D. Gooijer, R.J.Hyndman // Int. Journal of Forecasting.- 2006.-22(4).- Pp. 679-688.
4. Allende H. Artificial neural networks in time series forecasting: a comparative analysis. / H. Allende, C. Moraga. R. Salas // Kybernetika. – 2002. – 6. – Pp.685–707.
5. Shcherbakov M.V. A Survey of Forecast Error Measures / M.V. Shcherbakov, A.Brebels, N.L. Shcherbakova, A.P. Tyukov, T.A. Janovsky, V. Kamaev. // World Applied Sciences Journal 24 (Information Technologies in Modern Industry, Education & Society), –2013. – Pp.171–176.



## ANALYSIS OF THE INFLUENCE OF INTERFERENCE ON THE STATISTICAL PROPERTIES OF ESTIMATES OF NON-STATIONARY PARAMETERS

**Introduction** The extensive use of adaptive control methods is caused by the difficulties of obtaining an adequate mathematical description of real systems characterized by nonlinear dependencies, dynamic properties that are difficult to model, unsteady characteristics, the presence of uncontrollable noise and interference and their correlation, the absence of a priori information about the statistical properties of signals and interference, etc.

Adequate mathematical apparatus for solving the problem of synthesizing optimal control laws under uncertainty is the theory of discrete adaptive control systems, among which adaptive systems with an identifier in a loop (ASI) are widely used. Such an approach ensures timely and correct development of a decision on the change in control actions. The effective solution of the problem is associated with the introduction of an identifier into the control loop, which operates according to some adaptive algorithm. Built in this way, the system works as follows. The identifier estimates model parameters at discrete points in time. This estimate is fed to the regulator to form a control that minimizes the deviation of the current value of the object output from the required one. Thus, the model obtained in the identification process plays the role of a predictor, allowing to determine the value of the output variable of the object, which will be established at its output when certain signals are input to the input.

As already noted, the application of the adaptive approach is based on the replacement of a complex non-linear model with a linear model with variable parameters, which are evaluated in real time. As such a model, the pseudo-linear regression equation is usually used. Note that various equations describing linear and nonlinear dynamic objects can be reduced to this equation.

In this regard, certain requirements are imposed on the identifier operating in ASI, the main of which are the following:

- identification algorithm should ensure the convergence of parameter estimates to the true values of the parameters in the whole area of discrete changes of input variables;
- this convergence should not disappear when the regulator is turned on;
- computational procedure that implements the algorithm should be fairly simple;

– identification algorithm should provide tracking of variable parameters of the object.

Practically all currently used recurrent algorithms are obtained by minimizing the quadratic functional and used in the construction of the assessment direct measurements of input and output signals.

Thus, we can draw the following conclusion: for practical application of control algorithms it is necessary that they be adaptive, stable, non-linear, and also simple to implement and understand.

It should be noted that the quality of control systems, which largely depends on the calculation of control algorithms, in turn, requires simplifying the models. In some cases, it may be justified to manage a dynamic object using simple static models.

As noted above, when implementing adaptive systems, pseudo-linear regression models are widely used.

$$y_{n+1} = \theta^T x_{n+1} + \xi_{n+1}, \quad (1)$$

$$\tilde{x}_{n+1} = x_{n+1} + \zeta_{n+1}$$

where  $y_{n+1}$  is the observed output signal;  $x_{n+1} = (x_{1,n+1}, x_{2,n+1}, \dots, x_{N,n+1})^T$  - vector of input signals;  $\theta = (\theta_1, \theta_2, \dots, \theta_N)^T$  - vector of the required parameters  $N \times 1$ ;  $\xi_{n+1}$  - interference at the output of the object;  $\zeta_{n+1} = (\zeta_{1,n+1}, \zeta_{2,n+1}, \dots, \zeta_{N,n+1})^T$  - interference vector at the input of the object  $N \times 1$ , and reduced to minimizing some pre-selected quality functional (identification criterion). The quadratic functional that is most widely used in practice leads to various identification algorithms that allow one to obtain estimates of the desired vector  $\theta^*$  under normal distributions of interference, i.e.  $\xi_{n+1} \sim N(0, \sigma_\xi^2)$ .

However, the task of identification is significantly complicated if the parameters change (drift) in time, i.e.  $\theta_n = \text{var}$ .

The results of the majority of works devoted to this problem are mainly of a recommendatory nature and are weakly related to the nature of the nonstationarity of the object under study. So, for estimating non-stationary parameters, usually modified LS algorithms (using a sliding window or exponential smoothing, etc.), the Kaczmarz algorithm (and its dynamic variant), dynamic adaptation algorithms,

etc. are used. The lack of sufficiently general recommendations for choosing non-stationary parameters estimation algorithms apparently, the difficulties of theoretical studies of the properties of these algorithms in non-stationary conditions.

The effectiveness of applying one or another algorithm for estimating drifting parameters essentially depends on the amount of a priori information about the character of the drift. In accordance with this, adaptive identification algorithms for non-stationary parameters can be divided into two classes:

- 1) algorithms for estimating parameters with a known law of their drift
- 2) parameters estimation algorithms with an unknown drift law.

The first direction originates from the work of Dupac. In his paper, a generalization of the stochastic approximation method to the case when the root drift of the regression equation is close to linear is proposed. In most of the subsequent works, the case was considered when the drift is either on average close to linear or attenuates. In [1], the case was considered when the drift was parametrized, and the identification problem was reduced to estimating the unknown parameter  $\theta_n$  by the multidimensional method of stochastic approximation.

The lack of information about the nature of the drift  $\theta_n$  requires the development of identification algorithms that use the minimum amount of information and preserve operability in a wide range of variation  $\theta_n$ .

The algorithms of the second direction in the first place should include the recurrent least squares method with exponential weighting of information. Despite the large number of works devoted to the study of this method, in particular, the choice of the optimal weighing parameter, the question of choosing this parameter has not yet been resolved. The natural price for the lack of a priori information about the nature of the parameters drift is the delay in estimating the change in parameters, and, consequently, less tracking accuracy, which is due to the inertia of the algorithms.

Among the most computationally simple one-step identification algorithms, the Kaczmarz algorithm proposed in [2] for solving systems of linear algebraic equations, and subsequently used to solve the identification problem in constructing a model of type (1), is the most effective.

To improve the computational stability of the algorithm Kaczmarz V.M. Chadeev [3] proposed its modification - a regularized algorithm.

In [4], non-asymptotic and asymptotic estimates of the rate of convergence were obtained, in particular, the Kaczmarz regularized algorithm, which showed

that the introduction of a regularizing additive, improving the stability of the algorithms, slows down the rate of their convergence. However, the speed of these algorithms is often insufficient when estimating non-stationary parameters. Knowledge (or approximation) of the law of drift makes it possible to obtain effective algorithms for tracking non-stationary parameters. In a number of works, an algorithm called the Kaczmarz dynamic algorithm is considered, which uses some a priori drift model. It should be noted, however, that errors in the specification of the law of variation of parameters can lead to the loss of convergence properties of the algorithm.

Works of Sh.G. Lelashvili [5] gave impetus to the development of multi-step projection algorithms. In a number of works, the recurrent form of these algorithms was proposed, and the properties of random pseudoinverse matrices and projection matrices were established, which allowed to determine the rate of convergence of these algorithms and conclude that taking into account the information on L previous algorithms in these algorithms steps in terms of speed of convergence reduce the dimension of the original space N by L. Thus, the use of multi-step projection algorithms can significantly accelerate the process of identification and is enough effect injective when estimating nonstationary parameters.

In foreign literature, the Kaczmarz algorithm, better known as the normalized least squares algorithm (NLMS - normalized least-mean-square), has found wide application not only in the identification systems of stationary and non-stationary objects, but it is also widely used to solve filtering tasks. It should be noted that in a number of works, the Markov model of the first order was used to describe non-stationary parameters.

The goal of this paper is to study the issues of convergence of the one-step Kaczmarz regularized algorithm in the problem of identifying non-stationary parameters described by the Markov first-order model, and determining the parameters of the algorithm ensuring its maximum rate of convergence.

### **Regularized Kaczmarz algorithm**

Regularized Kaczmarz algorithm, having the form

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma \frac{e_{n+1} x_{n+1}}{\|x_{n+1}\|^2 + \delta}, \quad (2)$$

Where

$$e_{n+1} = y_{n+1} - \hat{y}_{n+1} = y_{n+1} - \hat{\theta}_{n+1}^T x_{n+1}, \quad (3)$$

$\hat{y}(k+1)$ - model output signal;  $\hat{\theta}(k) = (\hat{\theta}_1(k), \hat{\theta}_2(k), \dots, \hat{\theta}_N(k))^T$  - parameter estimation vector  $N \times 1$ ;  $\gamma$  - some parameter affecting the rate of convergence of the algorithm;  $\delta > 0$  – regularization parameter;  $\|\bullet\|$  - the Euclidean norm, is one of the most used both in solving the problem of identification, and in signal processing tasks (interference cancellation, echo cancellation, etc.).

It is assumed that the estimated non-stationary parameters, which are stationary random processes, can be described by the Markov model of the first order

$$\theta_{n+1} = \theta_n + S_{n+1} \quad (4)$$

Where

$S_{n+1} = (S_{n+1,1}, S_{n+1,2}, \dots, S_{n+1,N})^T$  - vector of random sequence  $N \times 1$ ;  $S_i \sim N(0, \sigma_s^2)$ .

Given the presence of interference  $\xi_{n+1}$  and  $\varsigma_{n+1}$ , the output signal of the object can be represented as follows:

$$y_{n+1} = \theta_n^T x_{n+1} + s_{n+1} x_{n+1} + \theta_n^T \varsigma_{n+1} - \theta_n^T \varsigma_{n+1} + \xi_{n+1}.$$

We introduce an evaluation error

$$\tilde{\theta}_{n+1} = \theta_{n+1} - \hat{\theta}_{n+1}. \quad (5)$$

Taking into account (4) the expression for can be written as follows:

$$e_{n+1} = \tilde{\theta}_n^T x_{n+1} - \theta_n^T \varsigma_{n+1} + S_{n+1}^T x_{n+1} + \xi_{n+1} \quad (6)$$

It is assumed that the interference is distributed according to the normal law with  $\xi_{n+1} \sim N(0, \sigma_\xi^2)$ ,  $\varsigma_{i,n+1} \sim N(0, \sigma_\varsigma^2)$ .

Taking into account the accepted type of nonstationarity (4), the Kaczmarz algorithm with respect to identification errors can be written as

$$\begin{aligned} \tilde{\theta}_{n+1} &= \tilde{\theta}_n + S_{n+1} - \gamma \frac{\tilde{x}_{n+1} \tilde{x}_{n+1}^T}{\|\tilde{x}_{n+1}\| + \delta} \tilde{\theta}_n + \gamma \frac{\tilde{x}_{n+1} \varsigma_{n+1}^T}{\|\tilde{x}_{n+1}\| + \delta} \theta_n - \gamma \frac{\tilde{x}_{n+1} x_{n+1}^T}{\|\tilde{x}_{n+1}\| + \delta} S_{n+1} - \frac{\tilde{x}_{n+1}}{\|\tilde{x}_{n+1}\|^2 + \delta} \xi_{n+1} = \\ &= \left( 1 - \gamma \frac{\tilde{x}_{n+1} \tilde{x}_{n+1}^T}{\|\tilde{x}_{n+1}\|^2 + \delta} \right) (\tilde{\theta}_n + S_{n+1}) + \gamma \frac{\tilde{x}_{n+1} \varsigma_{n+1}^T}{\|\tilde{x}_{n+1}\| + \delta} \theta_n - \gamma \frac{\tilde{x}_{n+1}}{\|\tilde{x}_{n+1}\|^2 + \delta} \xi_{n+1}. \quad (7) \end{aligned}$$

It is assumed that the components of the estimation error vector  $\tilde{\theta}_n$  obey the normal distribution law with  $\tilde{\theta}_{i,n} \triangleq M \{ \tilde{\theta}_{i,n} \}$  and variance  $\sigma_{i,n}^2$

**Analysis of the effect of interference  $\zeta_{n+1}$  on the statistical properties of the assessment**

Calculate  $M\{\tilde{\theta}_{n+1}\}$

$$M\{\tilde{\theta}_{n+1}\} = M\left\{\left(I - \gamma \frac{\tilde{x}_{n+1}\tilde{x}_{n+1}^T}{\|\tilde{x}_{n+1}\|^2 + \delta}\right)(\tilde{\theta}_n + S_{n+1})\right\} + \gamma M\left\{\frac{\zeta_{n+1}\zeta_{n+1}^T}{\|\tilde{x}_{n+1}\|^2 + \delta}\theta_n\right\} - \gamma M\left\{\frac{\tilde{x}_{n+1}}{\|\tilde{x}_{n+1}\|^2 + \delta}\xi_{n+1}\right\}$$

Here  $I$  is the identity matrix  $N \times N$ .

Considering the statistical independence and the fact that [4]

$$M\left\{\frac{x_{n+1}x_{n+1}^T}{\|x_{n+1}\|^2 + \delta}\right\} = \frac{1}{N}I - M\left\{\frac{\delta}{\|x_{n+1}\|^2 + \delta}\right\}I; \quad (8)$$

$$M\left\{\frac{1}{\|x_{n+1}\|^2 + \delta}\right\} = \frac{1}{(N-2)\sigma_x^2 + \delta} \left(1 - \frac{2\delta\sigma_x^2}{((N-2)\sigma_x^2 + \delta)^2}\right). \quad (9)$$

Taking into account the statistical properties of signals and interference, we have

$$M\{\tilde{\theta}_{n+1}\} = \left(1 - \frac{\gamma}{N} \left(\frac{((N-2)\sigma_x^2 + \delta)^3 - \delta N((N-2)\sigma_x^2 + \delta)^2 + 2\delta N\sigma_x^2}{((N-2)\sigma_x^2 + \delta)^3}\right)\right) M\{\tilde{\theta}_n + S_{n+1}\} + \frac{\gamma}{N} \frac{\sigma_x^2 - N\sigma_\xi^2 + \delta}{(\sigma_x^2 + \delta)} \theta_n \quad (10)$$

From the obtained expression (10) it follows that for the convergence of algorithm (2) it is necessary to satisfy the inequality

$$\left|1 - \frac{\gamma}{N} \left(\frac{((N-2)\sigma_x^2 + \delta)^3 - \delta N((N-2)\sigma_x^2 + \delta)^2 + 2\delta N\sigma_x^2}{((N-2)\sigma_x^2 + \delta)^3}\right)\right| < 1, \quad (11)$$

where it comes from

$$0 < \gamma < \frac{(\sigma_x^2 - N\sigma_\xi^2 + \delta)((N-2)\sigma_x^2 + \delta)^3}{((N-2)\sigma_x^2 + \delta)^3 - \delta N((N-2)\sigma_x^2 + \delta)^2 + 2N\delta\sigma_x^2} \quad (12)$$

Thus, when executing (12), algorithm (2) converges on average, i.e. evaluation will be offset  $\lim_{n \rightarrow \infty} M\{\tilde{\theta}_{n+1}\} \neq 0$ .

In other words, the vector of estimates will be associated with the vector of the desired parameters by the ratio

$$M\{\hat{\theta}_{n+1}\} = \frac{(\sigma_x^2 - N\sigma_\xi^2 + \delta)((N-2)\sigma_x^2 + \delta)^3}{((N-2)\sigma_x^2 + \delta)^3 - \delta N((N-2)\sigma_x^2 + \delta)^2 + 2N\delta\sigma_x^2} \theta_{n+1}. \quad (13)$$

Note that this formula does not contain the characteristics of non-stationarity and the statistical characteristics of the interference  $\xi_{n+1}$ . This is due to the type of adopted model of nonstationarity and properties of  $\xi_{n+1}$ .

Using the above formulas, it is easy to obtain that for the regularized Kaczmarz algorithm

$$M\{\hat{\theta}_{n+1}\} = \frac{\sigma_x^2 + \delta}{\sigma_x^2 + \sigma_\xi^2 + \delta} \theta_{n+1},$$

and for the classic Kaczmarz algorithm with  $\delta = 0$

$$M\{\hat{\theta}_{n+1}\} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\xi^2} \theta_{n+1}.$$

If there is no interference in the input signal, the estimate will be unbiased.

To eliminate the bias caused by the presence of interference  $\xi_{n+1}$ , it is necessary in some way to evaluate  $\sigma_\xi^2$  and adjust the resulting estimate  $\hat{\theta}_{n+1}$ . This can be done either in real time using procedures like stochastic approximation, i.e. correct the estimate of the desired parameter at each step of the identification process, or obtain an estimate  $\sigma_\xi^2$  based on a certain amount of information and adjust  $\hat{\theta}_{n+1}$  after all the calculations. At the same time, however, the issues of ensuring the convergence and achievement of the required rate of convergence of the corresponding estimation procedure for recurrent estimation  $\sigma_\xi^2$  and determining the adequacy of the resulting estimate  $\sigma_\xi^2$  of the real variance remain unsolved.

## Analysis of the effect of interference $\xi_{n+1}$ on the statistical properties of the assessment

As follows from the above results, the presence of interference  $\xi_{n+1}$  does not lead to a bias in the estimate. Consider how it affects the variance of the estimate.

To study convergence in mean square multiply (7) from the left to  $\tilde{\theta}_{n+1}^T$

$$\begin{aligned} \|\tilde{\theta}_{n+1}\|^2 &= \|\tilde{\theta}_n\|^2 + \|S_{n+1}\|^2 - 2\gamma \frac{\tilde{\theta}_n^T x_{n+1}}{\|x_{n+1}\|^2 + \delta} \left[ \xi_{n+1} + \tilde{\theta}_n^T x_{n+1} \right] - \\ &- 2\gamma \frac{x_{n+1}^T S_{n+1}}{\|x_{n+1}\|^2 + \delta} \left[ \xi_{n+1} + \tilde{\theta}_n^T x_{n+1} \right] + \gamma^2 \frac{\|x_{n+1}\|^2}{\left(\|x_{n+1}\|^2 + \delta\right)^2} \left[ \xi_{n+1} + \tilde{\theta}_n^T x_{n+1} \right]^2 \end{aligned}$$

Calculate  $M \left\{ \|\tilde{\theta}_{n+1}\|^2 \right\}$

$$\begin{aligned} M \left\{ \|\tilde{\theta}_{n+1}\|^2 \right\} &= M \left\{ \|\tilde{\theta}_n\|^2 \right\} + M \left\{ \|S_{n+1}\|^2 \right\} - 2\gamma M \left\{ \frac{\theta_n^T x_{n+1} x_{n+1}^T \theta_n}{\|x_{n+1}\| + \delta} \right\} - 2\gamma M \left\{ \frac{S_{n+1}^T x_{n+1} x_{n+1}^T S_{n+1}}{\left(\|x_{n+1}\|^2 + \delta\right)^2} \|x_{n+1}\|^2 \right\} + \\ &+ \gamma^2 M \left\{ \frac{\xi_{n+1}^2 \|x_{n+1}\|^2}{\left(\|x_{n+1}\|^2 + \delta\right)^2} \right\} \end{aligned}$$

Considering (8), (9) and the fact that [4]

$$\begin{aligned} M \left\{ \frac{(\tilde{\theta}_n^T x_{n+1})^2}{\|x_{n+1}\|^2 + \delta} \right\} &= \left[ \frac{1}{N} I - \frac{\delta}{(N-2)\sigma_x^2 + \delta} \left( 1 - \frac{2\sigma_x^2 \delta}{\left((N-2)\sigma_x^2 + \delta\right)^2} \right) \right] M \left\{ \|\tilde{\theta}_n\|^2 \right\}; \\ M \left\{ \frac{\|x_{n+1}\|^2}{\left(\|x_{n+1}\|^2 + \delta\right)^2} \xi_{n+1}^2 \right\} &= \frac{(N-2)\sigma_x^2 \sigma_\xi^2}{\left((N-2)\sigma_x^2 + \delta\right)^2}; \\ M \left\{ \frac{\|x_{n+1}\|^4}{\left(\|x_{n+1}\|^2 + \delta\right)^2} \right\} &= -2\delta \left( \frac{1}{(N-2)\sigma_x^2 + \delta} \left( 1 - \frac{2\sigma_x^2 \delta}{\left((N-2)\sigma_x^2 + \delta\right)^2} \right) - \frac{\delta^2}{\left((N-2)\sigma_x^2 + \delta\right)^2} \right) = \\ &= \frac{\delta^2 \left( (N-2)\sigma_x^2 + \delta \right) - 2\delta \left( (N-2)\sigma_x^2 + \delta \right)^2 + 4\delta^2 \sigma_x^2}{\left( (N-2)\sigma_x^2 + \delta \right)^3}, \end{aligned}$$

after simple transformations we get

$$M \left\{ \|\tilde{\theta}_{n+1}\|^2 \right\} =$$



$$\left\{ 1 - \frac{\gamma}{N} \left[ (2 - \gamma) - \frac{2(\gamma - 1)(\delta(N - 2)\sigma_x^2 + \delta)^2 + 2\sigma_x^2\delta}{((N - 2)\sigma_x^2 + \delta)^3} - \frac{\gamma^2\delta^2}{((N - 2)\sigma_x^2 + \delta)^2} \right] \right\} M \{ \|\tilde{\theta}_n\|^2 + \|S_{n+1}\|^2 \} + \gamma^2 \frac{(N - 2)\sigma_x^2\sigma_\xi^2}{((N - 2)\sigma_x^2 + \delta)^2},$$

whence it follows that for the convergence of the algorithm (2) in the root mean square, it is necessary to fulfill the condition

$$\left| 1 - \frac{\gamma}{N} \left[ (2 - \gamma) - \frac{2(\gamma - 1)(\delta(N - 2)\sigma_x^2 + \delta)^2 + 2\sigma_x^2\delta}{((N - 2)\sigma_x^2 + \delta)^3} - \frac{\gamma^2\delta^2}{((N - 2)\sigma_x^2 + \delta)^2} \right] \right| < 1 \quad (14)$$

The optimal value of the parameter that provides the maximum rate of convergence of the algorithm can be obtained by solving the equation

$$\frac{\partial M \{ \|\theta_{n+1}\|^2 \}}{\partial \gamma} = 0. \quad (15)$$

Omitting simple transformations, we get that

$$\gamma^{opt} = \frac{[A^3 + \delta A^2 + \delta \sigma_x^2] M \{ \|\theta_n\|^2 + \|S_{n+1}\|^2 \}}{[A^3 + 2\delta A^2 + \delta^2 A] M \{ \|\theta_n\|^2 + \|S_{n+1}\|^2 \} + N(N - 2)\sigma_x^2\sigma_\xi^2 A} \quad (16)$$

Where  $A = (N - 2)\sigma_x^2 + \delta$ .

It should be noted that the obtained relations generalize the known results. In our case, for the classical Kaczmarz algorithm with  $\delta = 0$ , (12) implies the well-known condition for average convergence

$$0 < \gamma < 2N, \quad (17)$$

and from (14) are the conditions for convergence in the mean square

- for non-stationary case

$$\begin{aligned} \gamma^{opt} &= \frac{((N - 2)^3 \sigma_x^6 (M \{ \|\theta_n\|^2 \} + \sigma_S^2))}{((N - 2)\sigma_x^6 M \{ \|\theta_n\|^2 \} + \sigma_S^2) + N(N - 2)^2 \sigma_x^4 \sigma_\xi^2} = \\ &= 1 - \frac{N\sigma_\xi^2}{(N - 2)\sigma_x^2 M \{ \|\theta_n\|^2 + \sigma_S^2 \}} \end{aligned} \quad (18)$$

$$\text{and } \gamma^{opt} = 1 \quad (19)$$

for stationary case and no interference.

**Some notes on the choice of the regularization parameter  $\delta$**

The purpose of using the regularization parameter is to increase the computational stability of the estimation algorithm. However, this raises the problem of effectively choosing this parameter. Fundamentally in the literature, this issue has hardly been investigated. This is apparently due to the fact that its solution requires a sufficiently large amount of information about the statistical properties of useful signals and interference affecting the properties of the algorithm. Some recommendations are given in a number of works. It was proposed to determine, based on the maximization of the signal-to-noise ratio

$$ENR = \frac{\sigma_y^2}{\sigma_\xi^2} = \frac{\theta^T R_x \theta}{\sigma_\xi^2},$$

where is  $\sigma_y^2 = M \{y_n^2\}$ ;  $\sigma_\xi^2 = M \{\xi_n^2\}$ ;  $R_x = M \{x_n x_n^T\}$ .

The expression proposed in this paper for determining the regularization parameter is

$$\delta = \frac{N(1 + \sqrt{1 + ENR})}{ENR} \sigma_x^2. \tag{20}$$

Thus, the value of the parameter  $\delta$  depends on the dimension of the model  $N$ , the dispersion of the input signal  $\sigma_x^2$  and  $ENR$ . If the value  $N$  is known, and  $\sigma_x^2$  can be determined or estimated, the value  $ENR$  is often not known.

In a number of works, the possibility of determining the optimal value of a parameter  $\delta$  from the conditions for maximizing the rate of convergence of the algorithm was considered. This approach allowed us to obtain the following expression for  $\delta_{n+1}^{opt}$

$$\delta_{n+1}^{opt} = \frac{N \sigma_\xi^2}{M \left\{ \left\| \tilde{\theta}_n \right\|^2 \right\}} \dots \dots \dots \tag{21}$$

Expression (21) contains, in addition to an unknown quantity  $\sigma_\xi^2$ , an unknown quantity  $M \left\{ \left\| \tilde{\theta}_n \right\|^2 \right\}$ . Therefore, these recommendations are more of theoretical interest. Their analysis allows to conclude that the value  $\delta_{n+1}^{opt}$  should be selected variable. At the same time, however, when  $ENR \rightarrow 0$  and  $M \left\{ \left\| \tilde{\theta}_n \right\|^2 \right\} \rightarrow 0$

$\delta_{n+1}^{opt} \rightarrow \infty$ , the influence of interference  $\xi_n$  is eliminated, however, the algorithm will not track the nonstationarity of parameters.

### Conclusion

As the results of research have shown, the use of a regularizing additive in identification algorithms, improving the stability of algorithms, leads to a certain slowing down of the model building process. The conditions for convergence of the regularized Kaczmarz algorithm are determined when estimating stationary parameters and the presence of measurement interference. The presence of interference in the input signal leads to a shift in the estimates obtained, which can be eliminated by using the procedure for estimating the variance of the interference and the corresponding correction of the estimate of the identified parameter. The presence of interference in the output signal does not lead to additional bias, however, it increases the variance of the identification error.

The estimates obtained are fairly general and depend both on the degree of nonstationarity of the object, and on the statistical characteristics of interference. In addition, expressions for the optimal values of the parameters of the algorithms are determined, ensuring their maximum convergence rate. Since these expressions contain a number of unknown parameters (estimation error  $\|\tilde{\theta}_n\|^2$ , the degree of non-stationarity of the object  $\sigma_s^2$ ), for their practical application one should use any recurrent procedure for estimating these parameters and use the resulting estimates to refine the parameters included in the algorithms.

### Literature

1. Цыпкин Я.З. Алгоритмы адаптации и обучения в нестационарных условиях / Я.З. Цыпкин, А.И. Каплинский, К.А. Ларионов / Изв. АН СССР. Техн. Кибернетика.-1970.-5.-С.9-21
2. Kaczmarz S. Angenäherte Auflösung von Systemen linearer Gleichungen / S. Kaczmarz / Bull. Int. Acad. Polon. Sci. Lett., C 1, Sci. Math. Nat., Ser. A- 1937. – S. 355-357.
3. English translation: Kaczmarz S. Approximate solution of systems of linear equations / S. Kaczmarz / Int. J. of Control, 1993. – 57. – Pp. 1269-1271.
4. Райбман Н.С. Адаптивные модели в системах управления / Н.С. Райбман, В.М. Чадеев. – М.: Сов. Радио. – 1966. – 156 с.
5. Либероль Б.Д. Исследование сходимости одношаговых адаптивных алгоритмов идентификации / Б.Д. Либероль, О.Г. Руденко, А.А. Бессонов / Проблемы управления и информатики.-2018.-5.-С.19-32.
6. Лелашвили Ш. Г. (1965) Применение одного итерационного метода для анализа многомерных автоматических систем./ В сб.«Схемы автоматического управления», С.19-33.«Мецниереба».

## ГЛАВА 8

### КЛАСИФІКАЦІЯ ЕЛЕКТРОННИХ НАУКОВО-ТЕХНІЧНИХ ТЕКСТІВ В ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМАХ

#### Вступ та постановка задачі

В області автоматичної обробки електронних текстів склався ряд відносно самостійних напрямків: витяг об'єктів і ознак, реферування, класифікація, кластеризація, інтелектуальний пошук, семантичний аналіз тощо. [1, 2]. Зокрема, завдання організації ефективного доступу до неструктурованої тематичної інформації безпосередньо пов'язано із завданням класифікації електронних текстів, які надходять з ресурсів мережі Інтернет або електронних бібліотек. Для вирішення останнього розроблено чимало ефективних методів, деякі з яких характеризуються якістю класифікації, що можна порівняти з результатами класифікації, виконуваної кваліфікованими експертами. До найбільш використовуваних методів слід віднести, зокрема, метод опорних векторів, метод логістичної лінійної регресії, метод k-найближчих сусідів, нейромережевої класифікації тощо. У той же час слід зазначити, що дані методи характеризуються високою складністю розробки та високою обчислювальною складністю.

Завдання класифікації тематичних текстових документів в загальному вигляді можна сформулювати наступним чином.

На вхід класифікатора надходить множина електронних текстів на природній мові (масив тематичних текстових документів). Передбачається, що існує множина класів (тематичних груп) заданої кількості, на які можна розбити даний масив документів. Тоді задача класифікації тематичних текстових документів зводиться до пошуку множини класів таким чином, щоб підсумкове безліч було оптимальним відповідно до деякого критерію якості розбиття документів. Строго кажучи, вхідними даними задачі класифікації є не самі тексти на природній мові, а їх векторні уявлення, що характеризують смисловий зміст вихідних документів. Ознаки документів автоматично витягаються з текстів відповідно до обраного способу представлення тексту і в найпоширенішому випадку є поодинокими словами [3].

Сучасні методи векторного представлення текстової інформації є розвитком моделі, заснованої на понятті векторного простору VSM (Vector Space Model). У такій моделі компоненти текстів, такі як слова, словосполучення, фрагменти текстів, цілі документи, представлені

векторами. Елементи векторів є значення деякої функції від частоти спільної зустрічальності компонентів текстів та їх контекстів. Ступінь подібності між компонентами текстів  $S_i$  і  $S_j$ , визначається величиною подібності між їх векторами. Зазвичай використовується деяка монотонна функція кута між векторами, наприклад косинус  $\cos(S_i, S_j)$ , який для нормованих векторів збігається зі скалярним добутком  $(S_i, S_j)$ . У моделі VSM використовуються слова, які насправді часто семантично залежні. Для подолання цього недоліку була запропонована альтернатива у вигляді методу GVSM (Generalized Vector Space Model) – узагальненої моделі векторного простору

Особливо актуальною задачею класифікації документів для систем інформаційного пошуку та обробки науково-технічних тематичних документів (статей, дисертаційних робіт, авторефератів тощо).

Метою статті є розробка та програмна реалізація метода класифікації колекції науково-технічних документів, що полягає у наявності фіксованої кількості деяких еталонних текстів та необхідності оперативного пошуку в цій колекції множин документів, які формуються згідно з розробленим методом. Кінцевою метою такої класифікації є пошук у колекції класів документів, які є близькими до заданих еталонів за запропонованими критеріями.

### **Загальна схема запропонованого алгоритму автоматичної класифікації електронних документів**

Введемо поняття міри близькості  $m(S_i, S_j)$  між двома електронними текстовими документами аналізованого корпусу  $S_i$  і  $S_j$  як деяку функцію, що залежить від ряду ознак, які характеризують ступінь близькості цих документів

$$m(S_i, S_j) = F(S_i \cap_p S_j), \quad (1)$$

де  $p$  – кількість ознак.

У розглянутій постановці електронний документ визначається значеннями компонент вектора ознак, причому ступінь впливу цих ознак на загальну оцінку міри (1) може бути різною. Відповідно до цього задамо відстань між документами  $S_i$  і  $S_j$  в такий спосіб:

$$K_H(S_i, S_j) = \sqrt{H_1 \left( \frac{x_{i1} - x_{j1}}{\bar{x}_{M1}} \right)^2 + \dots + H_n \left( \frac{x_{in} - x_{jn}}{\bar{x}_{Mn}} \right)^2}, \quad (2)$$

де  $H_k$  – міра невизначеності, що характеризує виключення  $k$ -ї ознаки з моделі оцінки документа;  $x_{ik}$ ,  $x_{jk}$  – значення  $k$ -ї ознаки для  $i$ -го та  $j$ -го документів відповідно;  $\bar{x}_{Mk}$  – середні значення  $k$ -ї ознаки для аналізованого корпусу документів.

Визначимо міру близькості  $m(S_i, S_j)$  між документами  $S_i$  і  $S_j$ , як величину, зворотну відстані (2):

$$m(S_i, S_j) = \frac{1}{K_H(S_i, S_j)}. \quad (3)$$

Під мірою близькості  $m(S_\alpha, D)$  довільного документа до кінцевої множини документів  $D$  будемо розуміти таке відношення:

$$m(S_\alpha, D) = \frac{1}{N_D} \sum_{S_j \in D} m(S_\alpha, S_j), \quad (4)$$

де  $N_D$  – кількість елементів множини  $D$ .

Виділимо з початкової множини документів  $M_0 \{S_1, S_2, \dots, S_n\}$  дві довільні кінцеві множини  $C$  та  $D$  і розглянемо функцію  $K(C, D)$ , що задає міру близькості між  $C$  та  $D$ :

$$K(C, D) = \frac{1}{N_C N_D} \sum_{S_i \in C} \sum_{S_j \in D} m(S_i, S_j), \quad (5)$$

де  $N_C$ ,  $N_D$  – кількість елементів множин  $C$  та  $D$  відповідно.

Визначимо двоетапну алгоритмічну послідовність агрегування початкової множини документів.

На першому етапі множині  $M_0 \{S_1, S_2, \dots, S_n\}$  ставляться у відповідність дві послідовності  $M \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_n\}$  та  $N \{\tilde{K}_1, \tilde{K}_2, \dots, \tilde{K}_n\}$  за таким правилом:

– з множини  $M_0$  довільно вибираємо будь-яку ситуацію  $S_j \in M_0$ , яка становить перший елемент  $\tilde{S}_1$  послідовності  $M$ , що формується. Документ  $S_j$  виключається з множини  $M_0$ ;

– серед решти  $(n - 1)$  документів шукаємо найближчий до  $\tilde{S}_1$  в сенсі (2) документ  $\tilde{S}_2$  і включаємо його другим в послідовність  $M$ , що формується. Першим елементом послідовності  $N$  буде величина  $m(\tilde{S}_1, \tilde{S}_2)$ ;

– серед решти  $(n - 2)$  документів шукаємо документ  $\tilde{S}_3$ , найближчий в сенсі (4) до перших двох документів, включаємо його третім в послідовність, що формується і так далі.

При цьому кожному документу  $\tilde{S}_i \in M$  ставимо у відповідність величину  $\tilde{K}_i$  послідовності  $N$ , що формується. Очевидно, що  $\tilde{K}_i$  є мірою близькості документа, що обирається на  $i$ -му кроці, до множини вже відібраних до  $i$ -го кроку документів  $\{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_{i-1}\}$ :

$$K_i = \max(\tilde{S}_i, \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_{i-1}\}), \quad (6)$$

де  $\tilde{S}_i \in \{S_1, S_2, \dots, S_n\} \setminus \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_{i-1}\}$ .

Якщо документ  $\tilde{S}_1 \in A$ , то спочатку відберемо всі документи, що належать множині  $A$ . Логічно припустити, що на межі переходу між множиною  $A$  і частиною множини  $M_0$ , що залишилася, величина  $\tilde{K}$  буде стрибкоподібно зменшуватися.

На другому етапі відбувається розбиття початкової послідовності на множини  $A_1, A_2, \dots, A_r$ . Спочатку формується множина  $A_1$  за таким правилом. Вважаємо, що  $\tilde{S}_1 \in A_1$ . Нехай далі до  $A_1$  віднесені документи  $\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_j$ . Документ  $\tilde{S}_{j+1}$  віднесемо до  $A_1$ , якщо виконується наступна нерівність:

$$\tilde{K}_j - \tilde{K}_{j+1} < \Delta, \quad (7)$$

де  $\Delta$  – пороговий параметр.

В іншому випадку формування множини  $A_1$  закінчується. Відзначимо, що кількість множин  $R$  залежить від величини параметра  $\Delta$ .

Необхідно вибрати критерій, що дозволяє, варіюючи цей параметр, вибрати з множини можливих класифікацій одну в певному сенсі найкращу. Очевидно, що класифікація тим краще, ніж ближче один до одного будуть

класи, що формуються в результаті класифікації. Разом з тим, класифікація тим краще, чим далі один від одного в сенсі (5) будуть розташовані сформовані класи. Визначимо міру близькості документів всередині класу, як власний потенціал класу:

$$K(A, A) = \frac{2}{N_A(N_A - 1)} \sum_{i,j,i>j}^{N_A} m(S_i, S_j), \quad (8)$$

де  $N_A$  – кількість документів в класі.

При цьому множник  $\frac{2}{N_A(N_A - 1)}$  відповідає величині, зворотній кількості членів суми.

Введемо поняття середнього власного потенціалу даної класифікації:

$$J_1 = \frac{1}{R} \sum_{i=1}^R K(A, A), \quad (9)$$

де  $R$  – кількість класів в класифікації.

Якщо порівнювати між собою кілька різних класифікацій, то величина  $J_1$  буде максимальною для тієї з них, в якій документи класу в середньому розташовані найбільш тісно.

Другою складовою шуканого критерію повинна бути величина, що характеризує близькість класів між собою. В якості такої величини приймемо наступну критеріальну складову:

$$J_2 = \frac{1}{R(R - 1)} \sum_{i,j,i>j}^R K(A_i, A_j), \quad (10)$$

де  $K(A_i, A_j)$  – визначена вище міра (5) близькості між класами.

Якщо порівнювати між собою кілька різних класифікацій, то величина  $J_2$  буде, очевидно, мінімальною для тієї з них, де класи в середньому розташовані далі один від одного.

Шуканий критерій  $J$  повинен бути таким, щоб найкраща класифікація відповідала відносно високому значенню  $J_1$  та в той же час відносно малому значенню  $J_2$ . Виділимо два таких можливих критерії:



$$J = J_1 - J_2, \quad (10)$$

$$J = \frac{J_1 - J_2}{J_1 + J_2}. \quad (11)$$

Варіюючи значення порогового параметра  $\Delta$ , визначимо класифікацію, для якої величина  $P$  є максимальною.

Сформувавши за наведеним алгоритмом класи, необхідно для кожного з них визначити еталонний документ, найбільш близький в сенсі (4) до всіх документів класу. В  $r$ -му класі еталонним документом  $S_{er}$  буде той, для якого виконується наступна умова:

$$m(S_{er}, \{M_r \setminus S_{er}\}) = \max(S_{ir}, \{M_r \setminus S_{ir}\}), \quad i = 1, N_r, \quad (12)$$

де  $M_r$  – множина документів  $r$ -го класу;  $N_r$  – кількість документів  $r$ -го класу.

### **Автоматична класифікація електронних документів за зразком**

Розглянемо завдання класифікації колекції документів  $M_0$ , що полягає у наявності  $r$  деяких еталонних текстів та необхідності оперативного пошуку в цій колекції множин документів, які формуються згідно з розглянутим методом. Кінцевою метою такої класифікації є пошук у колекції класів документів, які є близькими до заданих еталонів за розглянутими вище критеріями.

Це завдання має деякі особливості. Першим елементом  $\tilde{S}_1$  послідовності  $\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_j$ , що формується для побудови першого класу  $A_1$ , має бути перший з еталонних документів. В разі виконання умови (7)  $j$ -му кроці формування  $A_1$  завершується та алгоритм починає формувати другий клас  $A_2$ . Останнім класом, що буде сформовано таким чином, буде клас  $A_r$ . Документи, що не потрапили до жодного з класів послідовності  $M$ , вважаються далекими від заданих еталонів та надсилаються в спам-клас  $A_{r+1}$ .

Очевидно, що така постановка потребує попереднього виконання деяких стандартних процедур, насамперед, формування простору ознак, за якими будуть порівнюватися документи, та визначення порогового

параметра  $\Delta$ .

Простим прикладом ознаки документа є терм, тобто слово в деякій початковій формі, семантика якого допомагає описати основний зміст документа. Два документа вважаються близькими, якщо вони використовують одні і ті ж терми. Формування простору ознак полягає в вилученні з текстових документів ознак відповідно до обраного алгоритму. Можна виділити чотири класи алгоритмів вилучення ознак з текстів, що розрізняються підходами до визначення ознаки документа:

- алгоритми виділення унікальних одиночних слів. В даному випадку ознакою є унікальне одиночне слово, що зустрічається в тексті одного або більше документів в масиві документів. Ознаки цього виду називаються простими ознаками, або термами. Даний підхід заснований на інтуїтивному понятті, що родинні документи повинні використовувати однакові слова. Однак такий підхід теоретично вважається найбільш уразливим з точки зору вирішення таких проблем, як синонімія, омонімія тощо. У зв'язку з цим робляться спроби формування складових ознак;

- алгоритми виділення  $n$ -грам. Одним з варіантів формування складових ознак є формування множин пов'язаних термів, заснованих на припущенні про те, що слова, які з високою частотою зустрічаються в масиві документів спільно, насправді пов'язані один з одним. Такий процес називають статистичними алгоритмом виділення фраз. Такі фрази отримують більшу вагу за рахунок обліку зворотної документної частоти і найкращим чином «розрізняють» документи один від одного. Послідовності слів (елементів) довжини  $n$  називають  $n$ -грамами. Залежно від додатків в якості елементів можуть виступати фонемі, склади, букви, слова або словосполучення. Найбільш ефективним є використання біграм або діграм ( $n$ -грам розмірності 2). Однак, незважаючи на теоретичний потенціал застосування біграм експерименти показують, що представлення текстових документів в такій формі не завжди підвищує якість роботи програмного додатку (наприклад, програмного класифікатора). Підвищення точності отриманих в результаті роботи програми результатів може бути досягнуто при використанні біграм на додаток до уніграм ( $n$ -грами розмірності 1). Це можна досягти тільки в разі, якщо приділяється особлива увага розробкам процедур ретельного відбору «якісних» фраз або біграм, що пов'язано з великими обчислювальними витратами;

- алгоритми угруповання слів з використанням готових словників. Іншим варіантом може бути отримання інформації про взаємозв'язок термів з

електронних словників та слів, наприклад, тезаурусів, які можуть бути використані для розширення документа додаванням слів-синонімів до початкових терм.

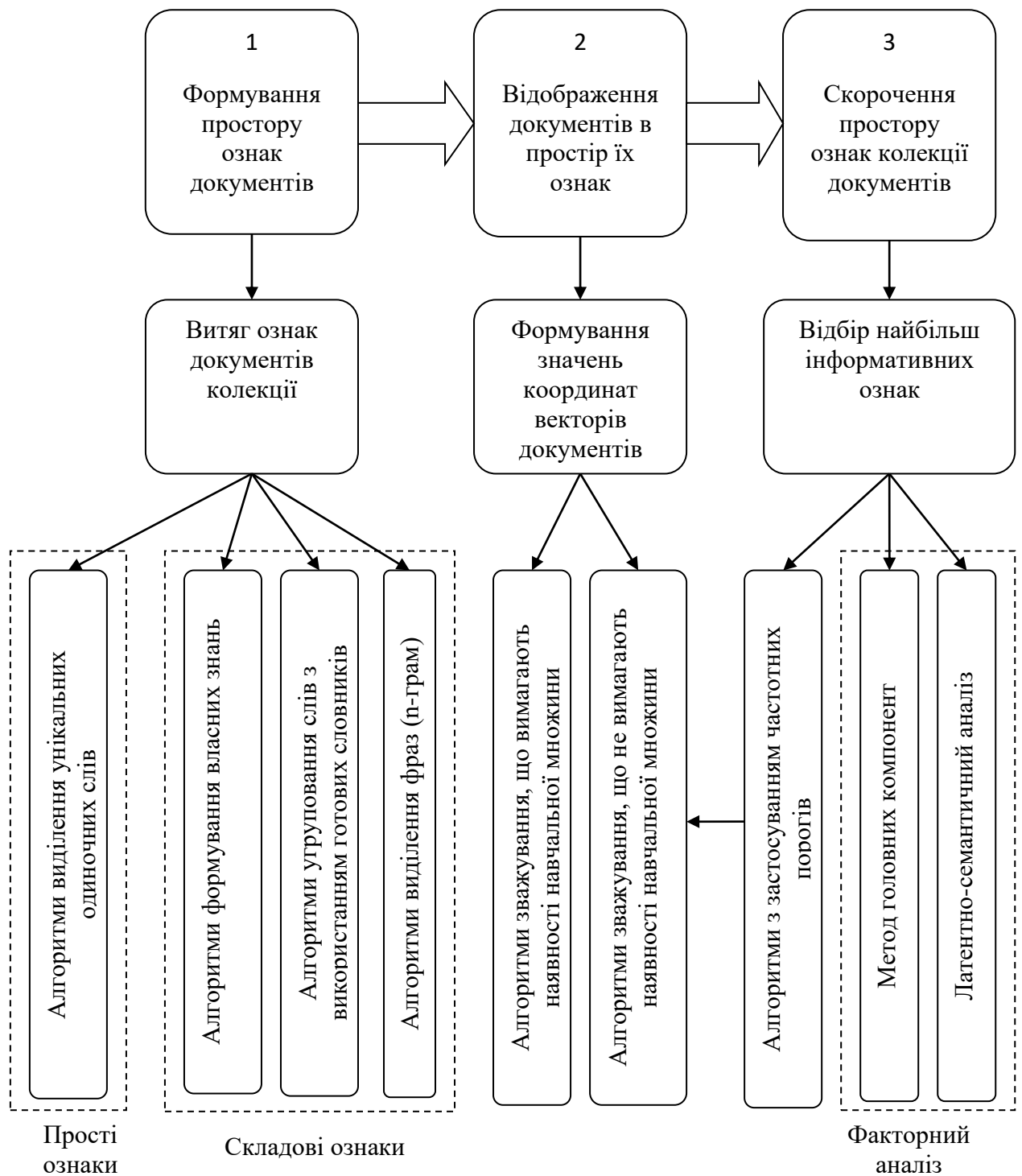
На рис 1. наведено етапи проведення попередньої обробки аналізованої послідовності електронних документів

Координатами вектора текстового документа в просторі ознак є величини, що характеризують ступінь значущості кожного вимірювання для визначення вмісту даного документа. Іншими словами, координати вектора документа є вагами ознак, що характеризують значимість ознак в даному документі.

Останнім часом набуло поширення використання ваг ознак, пропорційних частоті їх появи в окремих документах (term frequencies, або tf), та зворотної документної частоти (inverse document frequency, або idf), а також комбінації цих двох підходів [4].

$$tf \times idf = tf_{ij} \times \log_2 \left( \frac{|D|}{df_i} \right), \quad (13)$$

де  $tf_{ij}$  – частота появи  $i$ -го терма в  $j$ -му документі,  $|D|$  – потужність навчальної множини,  $df_i$  – кількість документів, в яких зустрічається  $i$ -й терм.



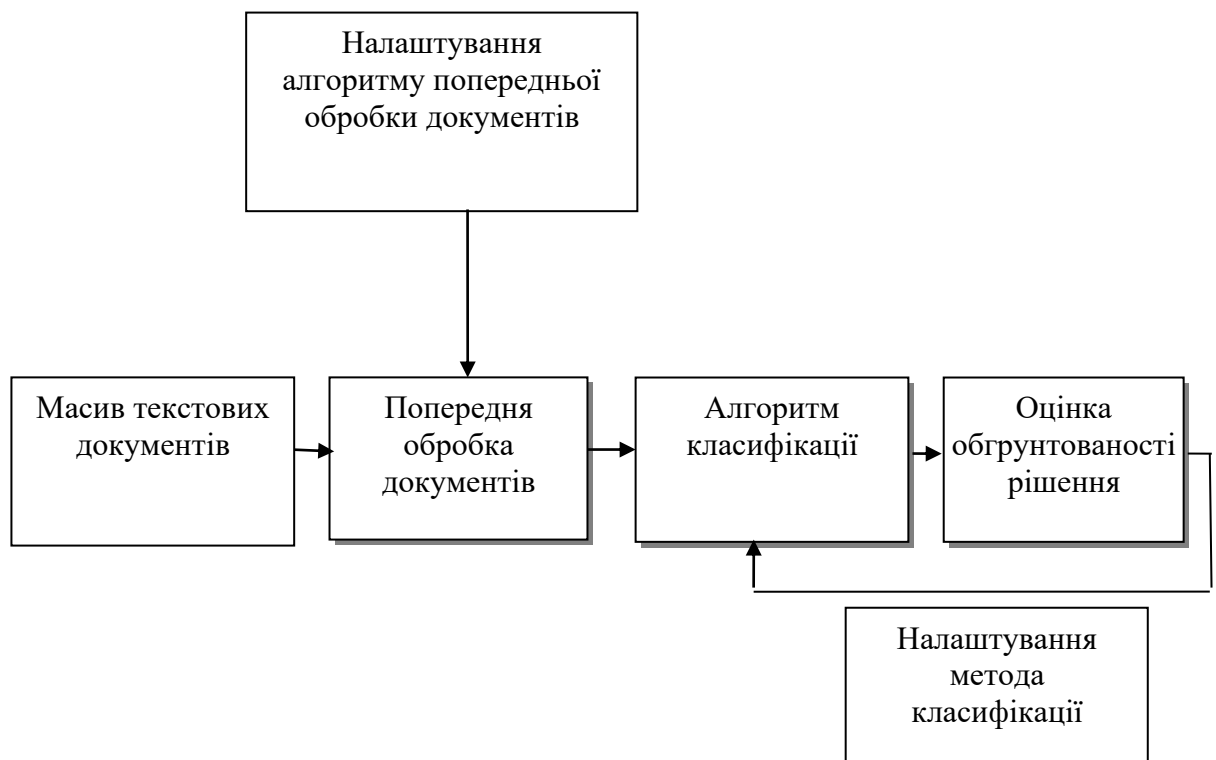
**Рис. 1. Етапи попередньої обробки текстових документів за статистичним підходом.**

Даний спосіб обчислення ваг може бути з успіхом застосований і для інших типів ознак документів. Функція tf-idf реалізує інтуїтивне уявлення про те, що чим частіше терм зустрічається в документі, тим точніше він описує його зміст, і чим в більшій кількості документів терм зустрічається,

тим менше розпізнавальних ознак він має. Отримані за схемою tf-idf ваги  $w_{ij}$  нормалізуються таким чином, щоб  $0 \leq w_{ij} \leq 1$ :

$$w_{ij} = \frac{\text{tfidf}(t_i, d_j)}{\sqrt{\sum_{k=1}^{N_p} (\text{tfidf}(t_k, d_j))^2}} \quad (14)$$

На рис. 2. представлена функціональна схема процесу класифікації масивів текстових документів в загальному вигляді.



**Рис. 2. Схема процесу класифікації масивів текстових документів**

Оцінка якості розбиття документів на класи різними алгоритмами класифікації виконується шляхом обчислення значень мір якості отриманого результату класифікації, серед яких виділяють:

- зовнішні міри: порівняння отриманого розбиття даних з залученою ззовні «еталонною» розбивкою цих же даних. «Еталонне» розбиття може бути отримано, наприклад, на основі експертних оцінок;
- внутрішні міри: автоматичний аналіз внутрішніх властивостей, властивих конкретному набору даних.

Для обчислення зовнішніх мір якості класифікації використовується поняття інтра-пари документів. Будь-яка пара документів є інтра-парою, якщо обидва документи належать одному і тому ж класу в даному наборі класів, інакше дана пара документів не є інтра-парою. Ключовим поняттям в порівнянні еталонного і автоматично отриманого результату класифікації є аналіз подібності прогнозів експертів і прогнозів системи щодо належності кожної пари документів одному або різним класам. Іншими словами, для кожної пари документів  $(S_j, S_i)$  на основі знання про два розбиття (одне отримано від експертів, інше – алгоритмом класифікації).

Обчислення зовнішніх мір засноване на аналізі результату класифікації відповідно до категорій. Відомими зовнішніми мірами є повнота і точність. В даному випадку повнота – це відношення коректних інтра-пар документів в автоматично отриманому розбитті до загальної кількості інтра-пар документів в еталонному розбитті. Точність – це відношення коректних інтра-пар документів в отриманому системою розбитті до загальної кількості інтра-пар документів в автоматично отриманому розбитті. Таким чином, повнота і точність обчислюються за такими формулами:

$$\text{Recall} = \frac{a}{a+b}, \text{Precision} = \frac{a}{a+c}, \quad (15)$$

Де Recall – це міра повноти; Precision – це міра точності;

Крім повноти і точності також використовуються: Accuracy – відношення правильно прийнятих системою рішень до загальної кількості рішень; Error – похибка, тобто відношення неправильно прийнятих системою до загальної кількості рішень.

При побудові системи класифікації важливо прагнути до підвищення показників як повноти, так і точності. У зв'язку з цим використовують міри, що дозволяють отримати сукупну оцінку як повноти, так і точності результату. Прикладом такої міри є F-міра:

$$F_{\beta} = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \quad (16)$$

де  $\beta (0 \leq \beta < +\infty)$  може розглядатися як відносна ступінь важливості

показників повноти і точності. Якщо  $\beta = 0$ , то значення  $F_{\beta}$  співпадає зі значенням Recall. Зазвичай значення  $\beta$  приймається рівним 1 (для рівної участі повноти і точності).

Головним недоліком підходу з використанням зовнішніх мір є суб'єктивність еталонного розбиття, отриманого на основі експертних висновків фахівців, а також необхідність витрат на отримання та обробку цих висновків.

### **Програмна реалізація алгоритму класифікації електронних текстових документів за зразком**

Запропонований алгоритм класифікації текстів (українськомовних, російськомовних та англомовних) за зразком включено до складу комплексного програмного модуля Booker V, що дозволяє здійснювати класифікацію великих обсягів текстової інформації по категоріям, враховуючи можливість розподіленого зберігання масивів інформації [5].

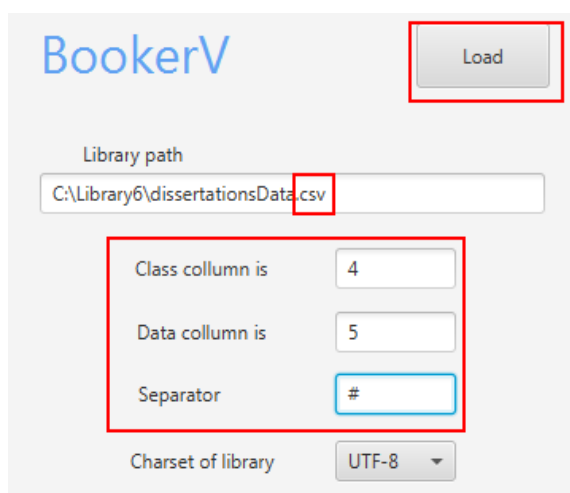
Оболонка програми складається з чотирьох основних структур блоків:

- блок завантаження колекції документів;
- блок системних налаштувань;
- блок введення еталонних документів на класифікацію;
- блок виведення інформації.

Завантаження вхідних даних з бібліотеки проводиться через один шлюз, з можливістю додаткових налаштувань.

При представленні текстів за допомогою файлу (LIB\_NAME.data / struct.csv) необхідно зробити попередні налаштування для аналізу колекції / бібліотеки. Приклад таких налаштувань наведено на рис. 3.

Для спрощення взаємодії з системою був реалізований алгоритм визначення типу вхідної інформації перед її класифікацією, що дозволяє вводити інформацію для класифікації у такий спосіб: посилання на файли в розширеннях .txt і .csv; списки посилань, які сприймаються через підрядник; текст в один рядок; текст багатостроковий; текстові csv-подібні структури.



**Рис. 3. Налаштування модуля Booker V (для cvs подібних структур).**

Для спрощення взаємодії з системою був реалізований алгоритм визначення типу вхідної інформації перед її класифікацією, що дозволяє вводити інформацію для класифікації у такий спосіб:

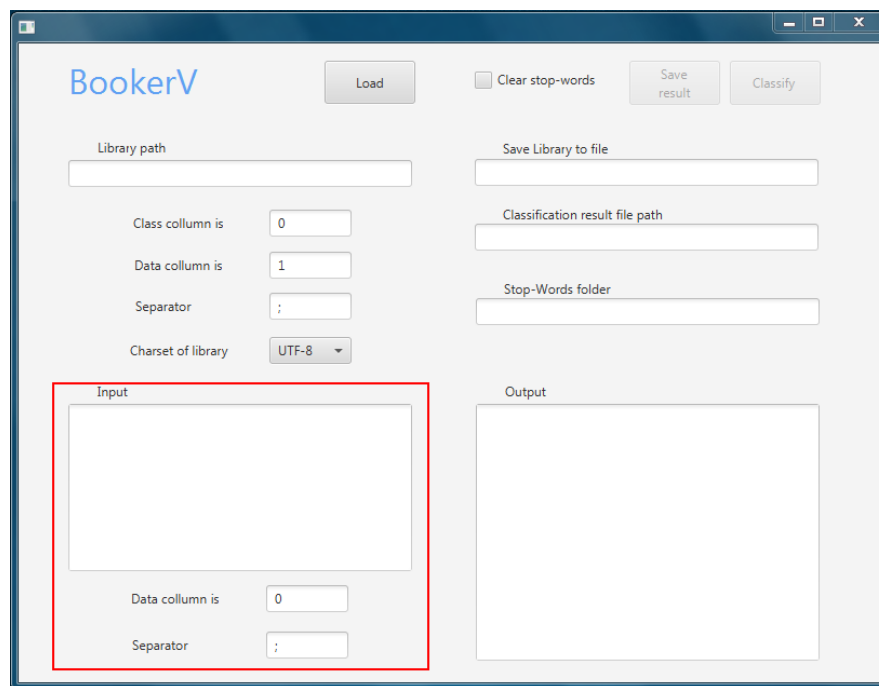
- посилання на файли в розширеннях .txt і .cvs;
- списки посилань сприймається через підрядник;
- текст в один рядок;
- текст багаторядковий (для цього необхідно стерти роздільник);
- текстові cvs-подібні структури.

Запуск класифікатора натисненням клавіши Classify, яка стає доступною, якщо в активній пам'яті програми є дані про бібліотеку (рис. 4).

При дослідженні можливостей класифікатора ВМ були використані два методи тестування вибірок, програмно реалізовані в модулі BookerV: testModel і crossValidation. Дані функції викликаються після навчання класифікатора на навчальній вибірці під час його перевірки на тестових даних. Це дозволяє: скоротити обчислювальний час завдяки скороченню числа слів, що враховуються в кожному класі; зменшити можливість переналаштування; знизити вплив викидів і шумів на результат класифікації.

На різних вибірках бібліотек формувалися персональні набори дескрипторів трьох видів: аббревіатури (наприклад, HDL, IP, JPEG, MIMO, VLS, ВПС), зв'язкові дескриптори (наприклад, «використання в конкретній функціональній задачі дозволяє розширити - функції адміністрування», «модель актуалізації оперативних даних - дозволяє автоматизувати - функція обробки документів») та загальні ключові слова (наприклад, центральний процесор, оцінка швидкодії, програмне забезпечення тощо).



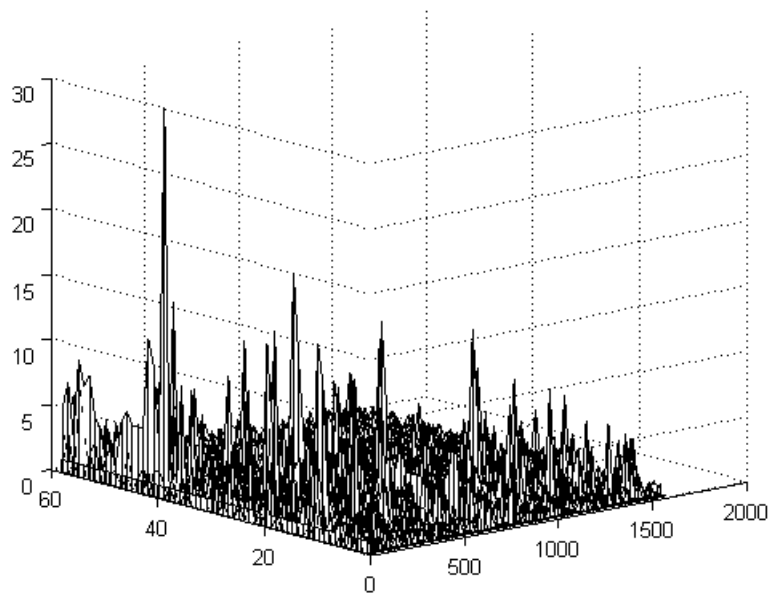


**Рис. 4. Блок введення інформації для класифікації.**

Новинний корпус Reuters містить 21578 документів, в тому числі політематичних, які відносяться до 135 категорій. Після попередньої обробки даний корпус містить 18933 терма (ознаки) і являє собою набір tf-idf-векторів для кожного документа корпусу. На рис. 5 графічно представлені tf-idf-вектори документів колекції, що відносяться до двох класів.

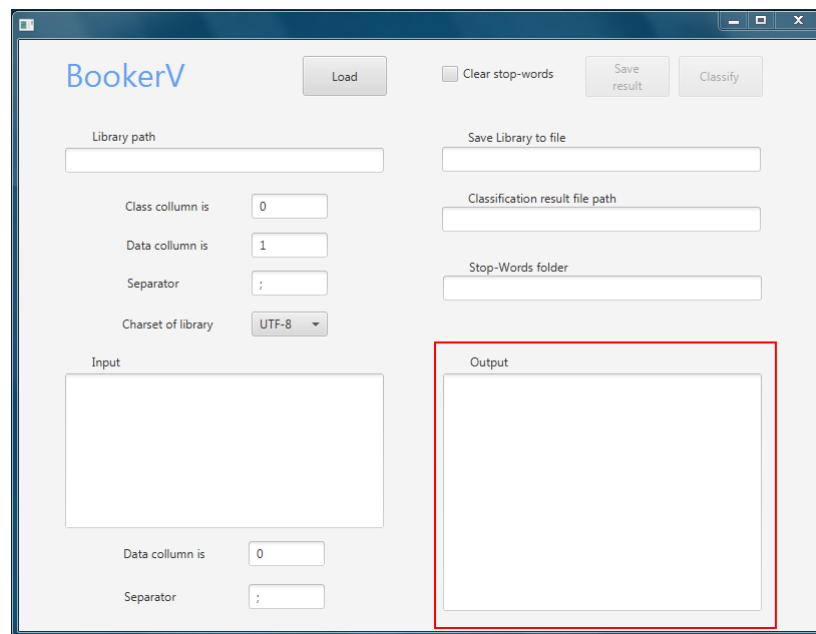
При тестуванні класифікатора також були використані: частина архіву вибірки електронних документів, наведеного в ресурсі <http://archive.ics.uci.edu/ml/dataset/SMS+Spam+Collectio>; частина електронної бібліотека «ІТАР» обсягом 21 ГБ і колекція анотацій авторефератів дисертаційних робіт обсягом 2 ГБ.

Оскільки в класифікаторі реалізована можливість для користувача створення або редагування особистих списків стоп-слів через шлях зберігання текстових файлів з їх наборами, кожен окремий файл вважався окремою групою і прохід по групах для користувача стоп-слів відбувався в порядку їх зберігання в директорії.



**Рис.5. Вектор tf-idf для документа из корпуса Reuters-21578.**

В блоці виведення відображаються результати класифікації (рис. 6).



**Рис.6. Блок виведення результатів класифікації.**

За результатами тестування точність класифікації (Precision) досягала для більшості текстових вибірок 97%.

## Висновки та перспективи подальших досліджень

Запропонований в статті метод дозволяє здійснювати класифікацію колекції науково-технічних документів для завдання, що полягає у наявності фіксованої кількості деяких еталонних текстів та необхідності оперативного пошуку в цій колекції множин документів, які формуються згідно з розробленим методом. Кінцевою метою такої класифікації є пошук у колекції класів документів, які є близькими до заданих еталонів за запропонованими критеріями.

Для спрощення взаємодії з системою був реалізований алгоритм визначення типу вхідної інформації перед її класифікацією, що дозволяє вводити інформацію для класифікації у такий спосіб: посилання на файли в розширеннях .txt і .cvs; списки посилань, які сприймаються через підрядник; текст в один рядок; текст багатостроковий; текстові cvs-подібні структури. Перспективним розвитком розглянутого підходу є розширення його функціональності на основі комплексного застосування методів формування простору ознак (зокрема, використання лінгвістичних дескрипторів та словосполучень, врахування зображень в текстах тощо).

## Література

- 1 Sebastiani F. Machine learning in automated text categorization/ Sebastiani F. // ACM Computing Surveys, 34(1), 2002. – pp. 1-47.
2. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В./ – М.: МИЭМ, 2011. – 272 с.
3. Епрев А.С. Автоматическая классификация текстовых документов / А.С. Епрев // Математические структуры и моделирование. – 2010 – Вып.21. – С. 65-81.
4. Чалая, Л. Э. Оценивание пертинентности лингвистических дескрипторов в системах информационного поиска документов / Л.Э. Чалая, Ю.Ю. Харитонова // Восточно-европейский журнал передовых технологий. – 2015. – № 1/9(73). – С. 46–53
5. Чалая Л.Э. Метод двухэтапной классификации электронных текстов // Л.Э. Чалая, С.Г. Удовенко, Е.С. Кушвид // Бионіка інтелекту. – 2016. – № 2 (87). – С.16 – 23.

## ГЛАВА 9

### АНАЛИЗ БЕЗОПАСНОСТИ ПРОТОКОЛОВ ПЕРЕДАЧИ СООБЩЕНИЙ. В MESSENGER TELEGRAM

#### **Введение**

Одним из самых популярных средств персонального обмена информацией в современном обществе стали мессенджеры – программы для мгновенного обмена сообщениями и файлами с мобильных устройств и настольных компьютеров. Большая распространенность персональных средств интерактивного обмена текстовой информацией и файлами в настоящее время сочетается с недостаточной защищенностью, используемых этими программными средствами, соединений между участниками диалога определяет актуальность исследования этой темы. Создатели мессенджеров борются за право называть свою продукцию самыми защищенными. Важным аспектом защиты передаваемых с использованием мессенджеров данных является способ шифрования и аутентификации этих данных.

Лучшим на сегодняшний день считается сквозное защищенное соединение. Общей чертой такого способа обмена сообщениями в мессенджерах является использование сервера только в роли «почтового ящика». Сервер не получает доступа к содержанию отправляемых сообщений. Для безопасной передачи сообщений между собеседниками в Telegram разработчиками предусмотрен специальный тип соединений – секретные чаты, которые обеспечивают сквозное шифрование и аутентификацию сообщений. Секретные чаты предназначены для общения только двух собеседников, в групповых чатах сквозное шифрование не используется. Протокол обмена сообщениями в секретных чатах Telegram называется MTProto, он обеспечивает так называемое сквозное шифрование (End-to-End Encryption).

#### **9.1. Регистрация устройства**

После первой установки приложения обмена сообщениями Telegram на новое устройство пользователю будет предложено ввести свой номер телефона, который используется для аутентификации пользователя для будущих входов в систему вместо имени пользователя/электронной почты и пароля. После ввода номера телефона пользователь получает 5-ти цифровой код по SMS для подтверждения номера. После ввода этого кода в приложение Telegram запускается протокол авторизации устройства.

Для клиента  $C$  регистрирующегося на сервере  $S$  протокол работает следующим образом:

1.  $C$  отправляет запрос на сервер  $S$  со случайным одноразовым значением битовой строки *nonce*.

2.  $S$  отвечает другой случайной битовой строкой *server\_nonce*, представляющей собой составной номер  $n$  и хеш открытого ключа алгоритма RSA.

3.  $C$  разбивает  $n$  на простые числа  $p$  и  $q$  такие, что  $p < q$ , что будет служить как доказательство стойкости.  $C$  имеет список открытых ключей, хранящихся локально, и выбирает ключ  $server_{pk}$ , который соответствует полученному хешу.

4.  $C$  выбирает другую случайную битовую строку *new\_nonce*, которая отличается от *nonce* и *server\_nonce* и не будет отправлена в виде открытого текста.  $C$  создает сообщение  $m_{npq}$  (набор сеансовых общесистемных параметров) из трех чисел  $n$ ,  $p$  и  $q$ . Это сообщение вместе с его хешем, шифруется с использованием RSA на ключе  $server_{pk}$  и отправляется на  $S$ .

5.  $S$  отвечает общесистемными параметрами протокола распределения ключей Diffie-Hellman (DH) сообщением набором чисел  $g$ ,  $p$ ,  $g_a$ , зашифрованными с помощью AES256 в режиме IGE, используя временный ключ AES и вектор инициализации AES IV, полученные из *server\_nonce* и *new\_nonce*.

6.  $C$  выбирает случайное секретное значение  $b$  и вычисляет  $g_b \leftarrow g^b \bmod p$  и  $auth\_key \leftarrow (g_a)^b \bmod p$ .  $g_b$  отправляется на сервер, зашифрованное с помощью AES-256 в режиме IGE с использованием временного ключа и вектора инициализации IV.

Случайные 128-битные строки *nonce* и *server\_nonce* генерируются случайно соответственно для клиента и сервера. Число *new\_nonce* представляет собой 256-битную строку которая вбирается клиентом  $C$  случайным образом.

$n$  представляет собой составное 64-битное целое число, которое является произведением двух нечетных простых чисел  $p$  и  $q$ , и хеш открытого ключа RSA (64 младших значащих бита хеша открытого ключа сервера полученного и с использованием SHA1).

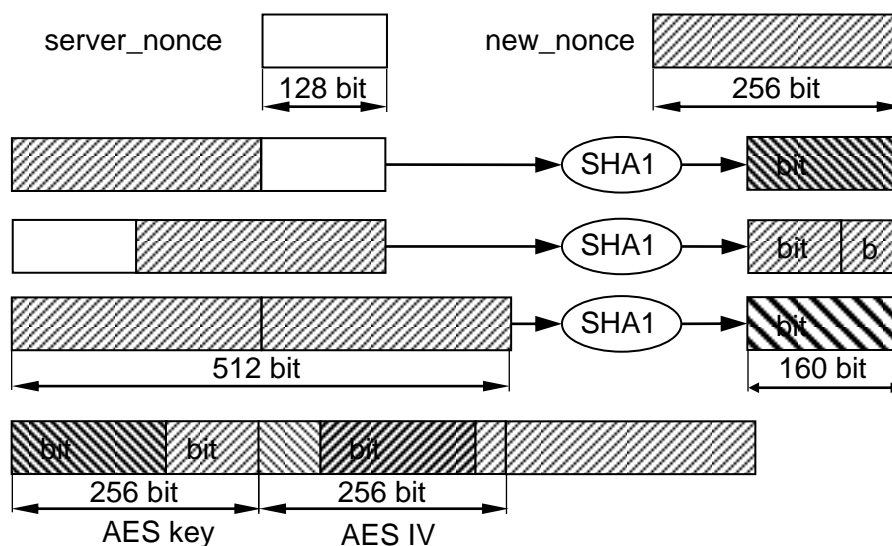
Шифрование RSA выполняется с использованием Java RSA (интерфейс `javax.crypto.Cipher`) с крипто-провайдером по умолчанию. Крипто-провайдер использует стандарт PKCS#1, в котором дополнение сообщения представляет собой псевдослучайные данные согласно Internet Engineering Task Force.

Чтобы расшифровать ответ сервера и получить параметры ДН, клиент должен получить 256-битный ключ AES и вектор инициализации IV, как показано на рисунке 1. Расшифрованный ответ содержит *nonce*, *server\_nonce*, параметры ДН  $g$ ,  $p$ ,  $g_a$  и временную метку сервера, используемую для синхронизации, а также включает 160 битный хеш (SHA1) содержимого. Ответ дополнен случайной последовательностью байт, которая отбрасывается без проверки, кроме длины, которая должна быть 0-15 байтов.

Клиент проверяет, что содержимое ответа соответствует хешу, который был отправленный с ним,  $p$  должен быть большим простым числом, удовлетворяющим  $2^{2047} < p < 2^{2048}$  и  $q = (p - 1)/2$ , а  $g$  равен 2, 3, 4, 5, 6 или 7 и должен порождать циклическую подгруппу простого порядка  $(p - 1)/2$ , а также  $1 < g_a < p - 1$ . Если какая-либо из этих проверок не пройдена, протокол прерывается.

Секретное значение  $b$  является случайно сгенерированной 2048-битной битовой строкой. Запрос, отправляемый на сервер, содержит данные *nonce*, *server\_nonce* и  $g_b$ , а также 160-битный хеш (SHA1) этих данных. Запрос дополняется и шифруется с использованием временного ключа AES и IV.

Клиент также вычисляет общий долгосрочный (мастер) ключ  $auth\_key \leftarrow (g_a)^b \bmod p$ .



**Рис. 1. Формирование ключевых данных для AES IGE и IV при регистрации.**

После получения  $g_b$ , сервер делает то же самое и отправляет ответ в "OK" или "FAILURE" (прерывает протокол), если *nonce*, *server\_nonce* или любые параметры ДН неверны.

Клиент и сервер теперь имеют общий секретный ключ, который используется только для передачи сообщений между клиентом и сервером и

обычных (не end-to-end point) чатов. Клиентское устройство теперь зарегистрировано и авторизовано. Приложение Telegram будет загружать телефонные контакты клиента на сервер, чтобы проверить, что они уже зарегистрированы, и, если, это так, он автоматически добавляет их как Telegram-контакты.

Если пользователь пытается войти с нового, неавторизованного устройства, новый код будет отправлен сервером клиенту. Сначала через приложение Telegram на авторизованное устройство, или через SMS, если первое не удастся, то вышеуказанный протокол повторяется.

## 9.2. Обмен ключами

Рассмотрим, как два пользователя иницируют сквозной зашифрованный секретный чат. Согласно веб-сайту Telegrams [3], стандартный Diffie-Hellman (DH) обмен ключами выполняется для того, чтобы пользователи сформировали общий секретный ключ, который позже будет использоваться для получения сеансового ключа.

Два пользователя  $A$  и  $B$  хотят начать секретный чат:

1. Пользователь  $A$  связывается с сервером, чтобы получить параметры DH, простое число  $p$  и генератор  $g$ .  $A$  также получает "salt", чтобы помочь своему клиенту (авторизованному устройству) создать случайное секретное значение  $a$ .

2.  $A$  вычисляет общедоступное значение DH  $g_a \leftarrow g^a \bmod p$ , и отправляет  $g_a$  второму пользователю  $B$ .

3.  $B$  получает запрос чата и принимает его (запрос) только на одном из авторизованных устройств. Он связывается с сервером, чтобы получить последние параметры DH и генерирует его секретное значение  $a$ .

4.  $B$  вычисляет публичную ключ  $g_b \leftarrow g^b \bmod p$  и отправляет его  $A$ .

5. Теперь оба пользователя вычисляют общий секретный ключ  $auth\_key \leftarrow (g_a)^b \bmod p = (g_b)^a \bmod p$  и обмен ключами завершен.

Клиенты должны проверить, что параметры DH, возвращаемые сервером, удовлетворяют следующим условиям:  $p$  должен быть большим простым числом, удовлетворяющим  $2^{2047} < p < 2^{2048}$  и  $q = (p - 1)/2$ , а  $g$  равен 2, 3, 4, 5, 6 или 7 и должен порождать циклическую подгруппу простого порядка  $(p - 1)/2$ . Следует обратить внимание, что DH параметры, полученные с сервера, могут обновляться между версиями приложения.

Секретные значения  $a$ ,  $b$  генерируются следующим образом:

$$a = r_{clent} \oplus r_{server} ,$$

где  $r_{client}$  – 2048 случайных битов, сгенерированных клиентом,  
 $r_{server}$  – 2048 случайные биты, генерируемых сервером.

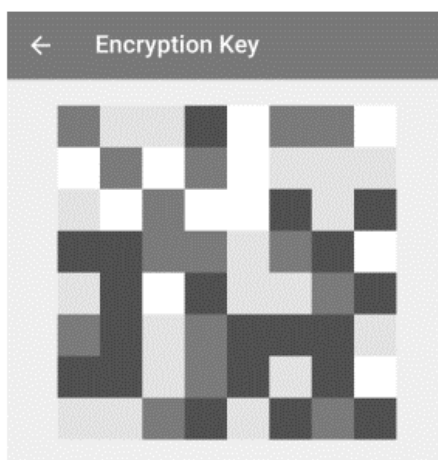
Причина в том, чтобы помочь клиентам со слабыми генераторами случайных чисел [1].

Клиенты должны проверить, что  $1 < g_a, g_b < p - 1$ , и  $2^{2048 - 64} < q_a, q_b < p - 2^{2048 - 64}$ . Вероятность этого примерно  $2^{-19}$ , что достаточно для обеспечения необходимой стойкости. Если  $g_a$  или  $g_b$  длиной менее 2048 бит, они дополняются нулями.

Обмен ключами Diffie-Hellman основан на проблеме дискретного логарифма, и гарантирует, что ни один пассивный криптоаналитик не сможет узнать  $auth\_key$ . Однако ДН восприимчив к активной атаке «человек посередине» так как не реализует услугу проверки подлинности.

Для гарантии подлинности Telegram, предоставляет каждому пользователю визуализацию, так называемый визуальный код аутентификации  $auth\_key$  для сравнения по аудио каналу.

Хеш общего секретного ключа – это 128 бит младших значащих битов SHA1 ( $auth\_key$ ). Это визуализируется в приложении Telegram как сетка 8x8, каждая ячейка которой имеет один из четырех цветов, как показано на рисунке 2. Для противодействия атаке «человек посередине» пользователям необходимо встретиться лично, или передать изображение по другому каналу связи, и сравнить визуальное представление хеша общего секретного ключа, убедившись, что нет различий. При условии, что пользователи используют для передачи изображения обычный чат с несекретным протоколом, их сеанс end-to-end point чата можно считать аутентифицированным и безопасным от третьих лиц.



This image is a visualization of the encryption

**Рис. 2. Визуализация хеша общего секретного ключа.**



### 9.3. Шифрование сообщений

В Telegram реализована собственная схема шифрования, MTProto, которая оптимизирована для мобильных устройств по вычислительным затратам без ущерба по стойкости. Алгоритмы шифрования используют стандартные примитивы и объединяют их таким образом, чтобы известные атаки были перекрыты.

Вначале рассмотрим содержание сообщения, которое будем называть полезной нагрузкой (рис. 3):



Рис.3. Структура полезного сообщения для шифрования.

Ключевые слова содержимого полезной нагрузки:

*Length*: 32-разрядное целое число, определяющее длину полезной нагрузки (не считая отступы или саму длину).

*Header*: каждая полезная нагрузка содержит три фиксированных заголовка, связанных с версией протокола и последний с указанием типа присоединенного media, который будем считать заголовком *media\_empty*.

*Random\_bits*: 120 случайных битов, сгенерированных отправляющим клиентом и 8 битов, используемых для указания длины *random\_bits* в байтах.

*Layer*: 32-разрядное целое число, указывающее версию протокола.

*seq\_in*: 32-битный счетчик для сообщений, отправляемых инициатору чата.

*seq\_out*: 32-битный счетчик для сообщений, отправленных инициатором чата.

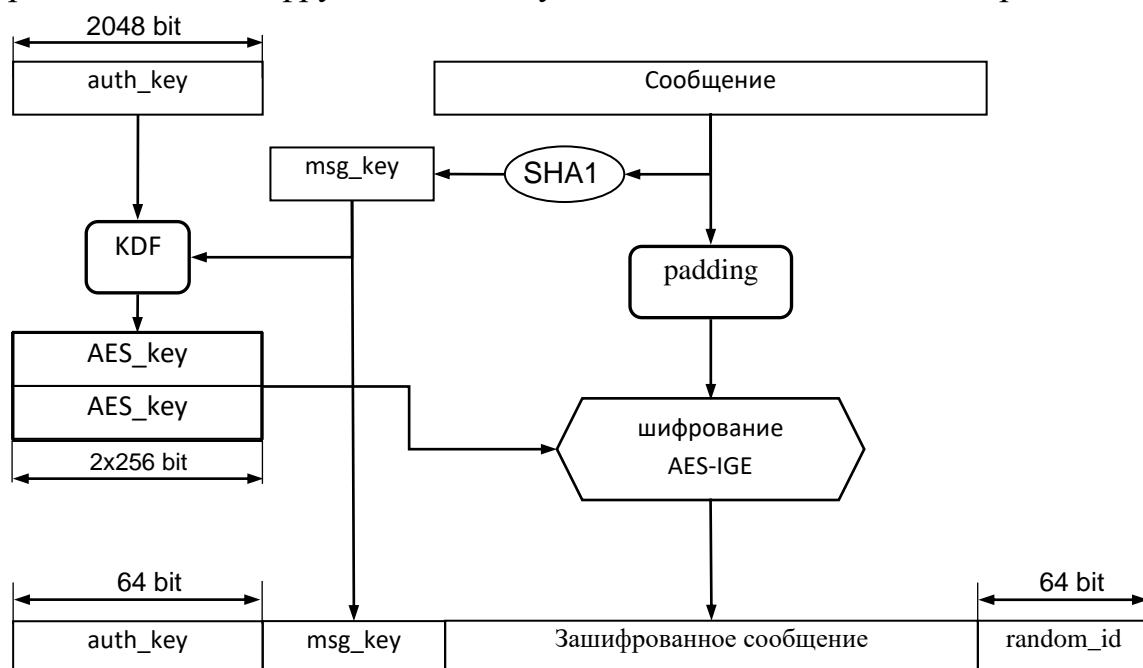
*Random\_id*: 64-битное случайное число, которое отправляется в виде открытого текста, генерируется отправляющим клиентом.

*ttl*: время жизни, 32-разрядное целое число, указывающее количество секунд в течении которых принимающему пользователю разрешено просматривать сообщение до его удаления.

*Message*: сообщение переменной длины, вводимое пользователем.

*Padding*: не входит в полезную нагрузку, заполнение добавляется непосредственно перед шифрованием.

Начиная с версии 2.7.0, содержимое сообщения, отправляемое в секретные чаты, шифруется используя MTProto, как показано на рис. 4.



**Рис.4. Схема шифрования MTProto версии 2.7.0.**

*auth\_key*: 2048-битный мастер-секрет, см. в разделе XX.2. Только 1024 самых значимых бита используются для получения ключа.

*Payload*: структура сообщения, показанная на рисунке XX.3.

*msg\_key*: усеченные до 128 младших значащих битов хэш SHA1 контент для шифрования. Используется для проверки целостности.

*Padding*: 0-96 случайных битов, генерируемых отправляющим клиентом, добавляются для обеспечения размера блока в 128 бит для шифра AES.

*AES key* и *IV*: 256-битный ключ и 256-битный вектор инициализации IV, требуемые в AES в режим IGE. Получено с использованием функции KDF, рассматриваемой в разделе XX.4.

*auth\_key\_id*: 64 младших бит хэша (SHA1) мастер-секрета *auth\_key*. В случае коллизии *auth\_key\_id* общий ключ восстанавливается.

Как показано на рисунке 4, проверка целостности *msg\_key* вычисляется для всего сообщения нагрузки, но без учета *padding*. В более поздних версиях протокола телеграмм *padding* включен в структуру сообщения, где заполнение включено в *msg\_key* и составляет 0-15 байт (0-120 бит) после чего зашифровывается.

Фрагмент кода на рис.5 взят из секретных чатов [2]. Данный код хэширует полезную нагрузку и усекает ее до ключа сообщения, а затем добавляет *padding*.

В данной версии протокола MTProto заполнение на самом деле составляет всего 0-96 бит, а не 0-120 бит как заявлено.

```
708 byte[] messageKeyFull =
        Utilities.computeSHA1(toEncrypt.buffer);
709 byte[] messageKey = new byte [16];
710 System.arraycopy(messageKeyFull,
        messageKeyFull.length-16, messageKey, 0,
        16);
711
712 MessageKeyData keyData =
        Utilities.generateMessageKeyData(chat.auth_key,
        messageKey,
        false);
713
714 len = toEncrypt.length();
715 int extraLen = len % 16 != 0 ? 16 - len % 16 : 0;
716 ByteBufferDesc dataForEncryption =
        BuffersStorage.getInstance()
        .getFreeBuffer(len+extraLen );
717 toEncrypt.position(0);
718 dataForEncryption.writeRaw(toEncrypt);
719 if (extraLen != 0) {
720     byte [] b = new byte[extraLen];
721     Utilities.random.nextBytes(b);
722     dataForEncryption.writeRaw(b);
723 }
```

**Рис.5. Фрагмент кода хэширующий сообщение без учета *padding*.**

### 3.1. Порядковые номера сообщений

Атака воспроизведения – это атака, при которой злоумышленник записывает отправленное сообщение и отправляет его получателю еще раз, но позже. Эта атака вводит с чат сообщение без ведома оригинального отправителя. Данная атака перекрывается путем включения счетчика

сообщений в каждое сообщение и отслеживания порядка получения отправленных и полученных сообщений.

Аналогично, атака зеркального отражения – это когда злоумышленник записывает отправленное сообщение, и отправляет его обратно отправителю. Если никакие меры предосторожности не приняты, сообщение будет принято в чат, и пользователь будет думать, что другой клиент отправил сообщение идентичное тому, которое он только что отправил. В этом случае, для перекрытия атаки, использование только счетчика сообщений может оказаться не достаточным. Пользователь не может быть уверен, в том, что другой клиент отправил идентичное сообщение в тот же момент времени, в этом случае счетчики сообщений будут одинаковыми.

Для перекрытия данной атаки MTProto использует два счетчика сообщений [5], один для входящих и один для исходящих сообщений (с точки зрения инициатора чата), и один бит каждого используется для указания какой пользователь отправил данное сообщение. Предположим, у нас есть два пользователя *A* и *B*. *A* является инициатором чата:

- каждый секретный чат содержит два счетчика, *seq\_out* и *seq\_in*;
- они инициализируются в (0,0);
- оба увеличиваются на 1 после каждого отправленного сообщения;
- они трансформируются следующим образом:

$$2 * seq\_no + x,$$

где *x* вычисляется по следующему правилу:

	<i>seq_in</i>	<i>seq_out</i>
сообщение отправленное <i>A</i>	0	1
сообщение отправленное <i>B</i>	1	0

Оба клиента поддерживают необработанные (не преобразованные) счетчики и два преобразованных счетчика, которые включены в каждое отправленное сообщение.

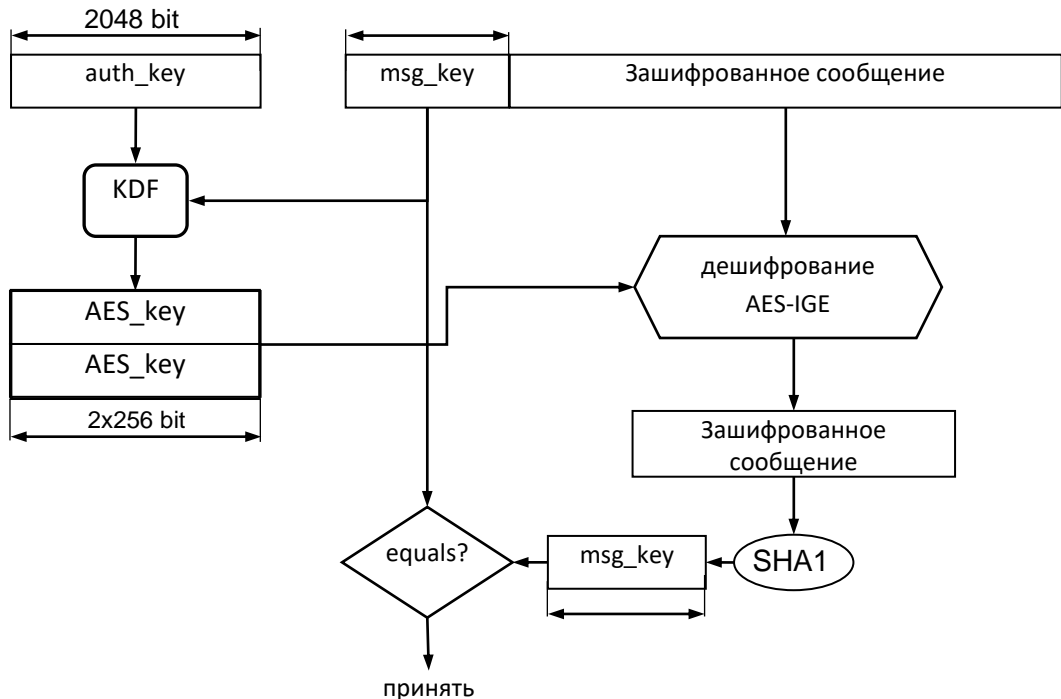
Теперь, если злоумышленник попытается реализовать атаку воспроизведения, получатель сообщения увидит, что он уже получил сообщение с таким же порядковым номером, и отменит сообщение.

Если злоумышленник предпримет попытку зеркальной атаки на пользователя *A*, например, *A* будет видеть, что сообщение является исходящим и не предназначено для него, и сможет отменить сообщение.

### 3.2. Расшифровка сообщения

Расшифровка сообщений в MTProto аналогична процессу шифрования, как показано на рис. 6.

На рис. 6 не показано проверка первого полученного *auth\_key\_id* на соответствие с ассоциированным секретным чатом на клиенте. Далее в алгоритме 1 рассмотрим, как параметры проверяются при расшифровке.



**Рис.6. Схема дешифрования MTProto версии 2.7.0.**

Псевдокод в алгоритме на рис.7 принимает в качестве входных данных *auth\_key\_id'*, *msg\_key'*, *encrypted\_data*. Он использует методы *SHA1(input)* и *readBits(input; skip; length)*, где *skip* – количество пропускаемых бит и *length* – количество считанных бит из *input*, метод KDF рассматривается в разделе 4.

В строке 11 полезная нагрузка уменьшена по длине, это сделано для удаления заполнения до вычисления *msg\_key*.

Если в любой момент проверка не пройдена, полученное сообщение отбрасывается полностью без сообщения об ошибке.

**Input:** *auth\_key\_id'*; *msg\_key'*; *encrypted\_data*

```
1: auth_key_id ← computeAuthKeyID(auth_key);
2: if (auth_key_id' ≠ auth_key_id) {
3:   throw "Отклонить сообщение";
4: }
5: AES_key; AES_IV ← KDF(auth_key; msg_key');
6: message ← AES-IGE_Decrypt(encrypted_data; AESkey; AESiv);
7: length ← readLength(message);
8: if (length удовлетворяет ограничениям) {
9:   throw "Отклонить сообщение";
10: }
11: message ← removePadding(message);
12: msg_key ← computeMsgKey(message);
13: if (msg_key' ≠ msg_key) {
14:   throw "Отклонить сообщение";
15: }
16: seq_in; seq_out ← readMessageCounters(message);
17: if (seq_in; seq_out ≠ local seq_in; seq_out) {
18:   throw "Отклонить сообщение";
19: }
20: return message;
```

**Рис.7. Псевдокод алгоритма шифрования MTProto**

### 3.3. Подлинность зашифрованных сообщений

MTProto обеспечивает конфиденциальность с использованием шифрования AES и целостность открытого текста на основе *msg\_key*. Кроме того, режим IGE обеспечивает целостность зашифрованного текста, искажая сообщение, если зашифрованный текст подделан. MTProto явно не использует конструкцию encrypt-then-authenticate, фактически он вообще не использует MAC.

Вместо этого получатель сообщения проверяет подлинность, на основе соответствия полученного *msg\_key* хешу дешифрованного сообщения. Если эта проверка проходит, тогда ключи шифрования и дешифрования должны быть получены из того же общего секрета, и, следовательно, он должен быть зашифрован и отправлен вторым пользователем в секретном чате, который аутентифицируется путем сравнения хеша *auth\_key*.

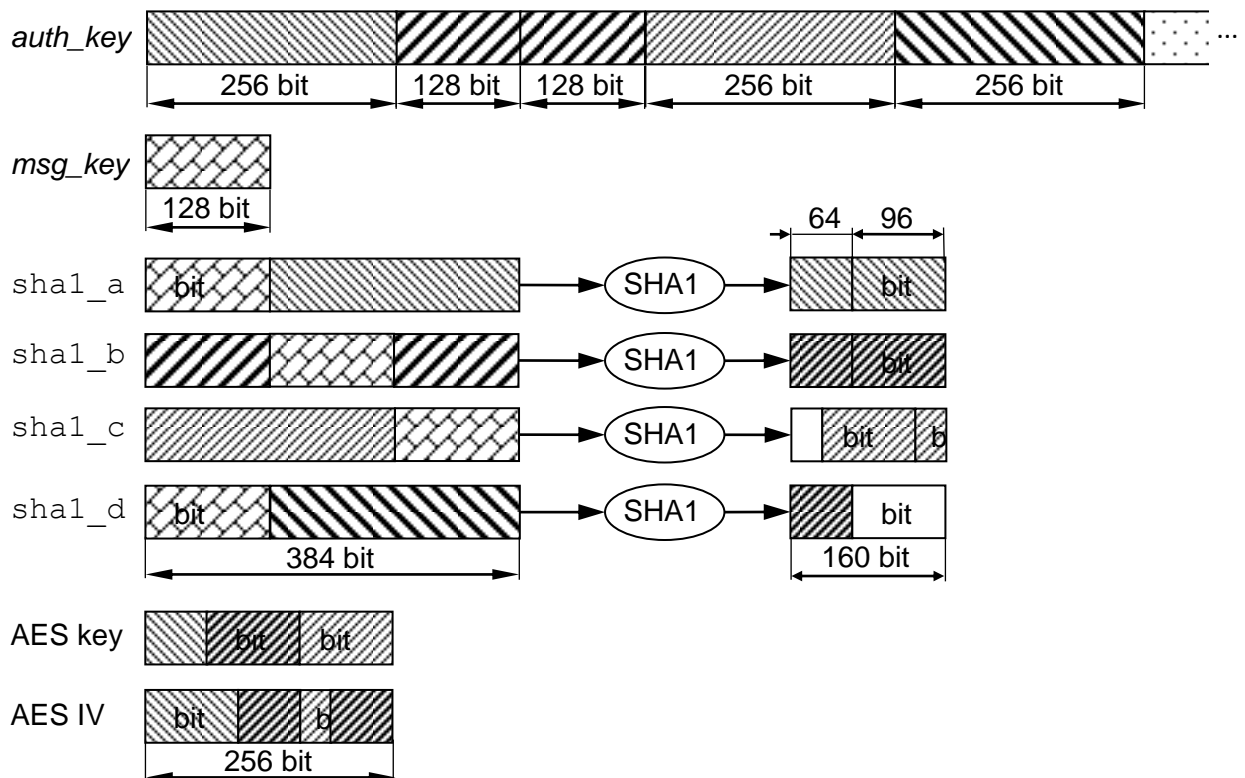
## 9.4. Функция генерации ключа

Рассмотрим, как MTProto вычисляет ключи шифрования, которые используются для AES-IGE. MTProto использует AES в режиме IGE с 256-битными ключами. IGE требует вектор инициализации, в два раза

превышающий размер одного блока AES, 256 бит. Пара новый ключ и вектор инициализации IV генерируется для каждого сообщения, и они получены из общего секрета *auth\_key* и сообщения. Таким образом только другой пользователь имеющий *auth\_key* может расшифровать сообщение, а ключ будет расшифровывать только одно сообщение.

Ключ и вектор инициализации IV получают путем использования нескольких хэшей SHA1, так как, один хэш SHA1 генерирует только 160 бит из необходимых 512 биты. С точки зрения современного уровня безопасности более целесообразно использовать более новую хеш-функцию, такую как SHA-512, но MTPROTO использует SHA1.

Как показано на рис. 8, генерируются четыре хэш-значения SHA1, каждый из которых получает 256-битный ключ *auth\_key* и 128-битный ключ *msg\_key* (усеченный хэш SHA1 сообщения). В общей сложности используется 1024 значимых бит *auth\_key*. Следует обратить внимание, что положение *msg\_key* на входе хеша смещается каждый раз, предположительно, чтобы ввести большую степень неопределенности ("случайности").



**Рис.8. Генерация хеш-значений SHA1 для функции KDF в MTPROTO.**

Затем формируются ключ AES и вектор инициализации IV путем взятия подстрок на каждом из четырех выходов SHA1 и объединением их в две 256-битные строки.

Функция генерации ключа описана более точно в алгоритме псевдокода рис.9, используя *readBits(input, skip, length)*, где *skip* – количество пропускаемых битов и *length* – количество считываемых битов на входе.

```
1 sha1_a <- SHA1(msg_key + readBits(auth_key, 0, 256))
2     sha1_b <- SHA1(readBits(auth_key,256, 128) +
3     msg_key +
4     readBits(auth_key, 384, 128))
5     sha1_c <- SHA1(readBits(auth_key, 512, 256) +
6     msg_key)
7 sha1_d <- SHA1(msg_key + readBits(auth_key, 768, 256))
8
9     AES_key <- readBits(sha1_a, 0, 64) +
10    readBits(sha1_b, 64, 96) + readBits(sha1_c, 32, 96)
11    AES_IV <- readBits(sha1_a, 64, 96) +
12    readBits(sha1_b, 0, 64) + readBits(sha1_c, 128, 32) +
13    readBits(sha1_d, 0, 64)
```

**Рис.9. Функция генерации ключа KDF в алгоритме MTProto.**

Совершенная прямая секретность (Perfect forward secrecy, PFS) – свойство некоторых протоколов согласования ключа, которое гарантирует, что сессионные ключи, полученные при помощи набора ключей долговременного использования, не будут скомпрометированы при компрометации одного из долговременных ключей.

В MTProto [4] это соответствует ключам AES, полученным из *auth\_key*, который фиксирован для секретного чата. Если *auth\_key* скомпрометирован, то злоумышленник получит доступ к каждому отправленному сообщению. Для обеспечения PFS в MTProto версии 20 введен счетчик *key\_use\_count*, отслеживающий сколько раз ключ *auth\_key* используется для получения ключа AES для шифрования/дешифрования. Если этот счетчик достигает 100 или секретный ключ использовался более одной недели он будет заменен на новый ключ *auth\_key*, а старый уничтожен.

Новый обмен ключами повторно использует параметры ДН, которые были получены от сервера во время первоначального обмена ключами, и случайный вектор не добавляется при генерации новых секретных значений приватных ключей *a* и *b*. Один пользователь будет вычислять его открытый ключ  $g_a$  и отправит его пользователю *B* через уже установленный чат. *B* отправляет *A* его новым открытый ключ  $g_b$ , и оба пользователя вычисляют и обновляют *auth\_key* для существующего сеанса, аналогично тому, как это



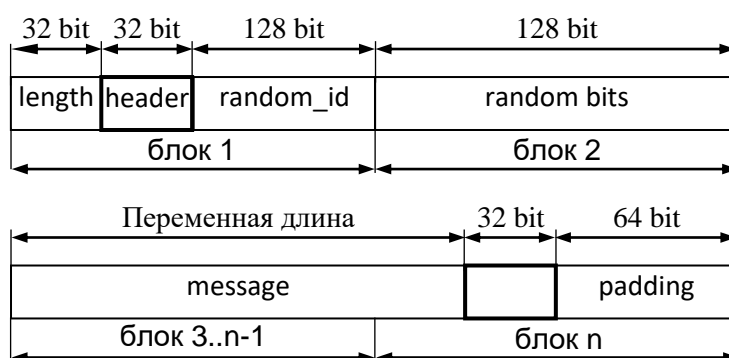
было сделано в разделе 2. Все старые сообщения останутся расшифрованными и читаемыми на обоих клиентах.

Хеш *auth\_key* используемый для аутентификации клиентов, не будет изменен. Это означает, что, если противник завладеет старым ключом *auth\_key*, он может получить ключи AES и прочитать 100 зашифрованных сообщений, но он не сможет быть в состоянии построить следующий ключ из открытых значений ДН.

За последнее время MTProto получил несколько обновлений, и в результате нет гарантии, что все клиенты Telegram используют последнюю версию. Чтобы клиенты могли общаться через разные версии, известные как слои, обновленные слои сделаны обратно совместимыми со старыми.

Официальный клиент Telegram выполняет согласование версии слоя, как только обмен ключами в секретном чате завершен. Ключ *auth\_key* хранится в объекте *encryptedChat* вместе с версией слоя, инициализированным 0. Инициатор чата *A* отправляет свой слой в зашифрованном сообщении *B*, а *B* отвечает его версией слоя. Оба клиента хранят наименьший из двух уровней в объекте *encryptedChat*, чтобы соответствовать более старой версии.

PFS была введена с 20-го слоя, кроме этого, никаких серьезных изменений не произошло с уровня 17. В слое 16 по 8 включительно (версия, которая ввела секретные чаты), порядковые номера сообщений обрабатывались очень по-разному. Вместо того, чтобы быть включенным в каждое зашифрованное сообщение, они хранились и поддерживались на соответствующем сервере, и структура сообщений выглядела как представлена на рис.10.



**Рис.10. Структура сообщения до слоя 17.**

Это означает, что зашифрованные сообщения не содержали никаких указаний на порядковый номер сообщения, и клиенты должны были доверять серверу.

Сами сообщения не содержали никакого механизма, чтобы остановить несанкционированный доступ третьих лиц для прослушивания отправленных сообщений, и выполнения атак воспроизведения или зеркального отражения, как описано в разделе 3.1, которые смягчаются последующим шифрованием сообщений на ключе авторизации клиент-сервер из 1.

### **Заключение**

Проведено исследование протокола MTProto мессенджера Telegram версии 2.7.0, а именно регистрация устройства, обмен ключами, шифрования сообщения, функции генерации ключа.

Анализ показывает, что данному протоколу присущ ряд недостатков:

- использование устаревшего алгоритма хеширования;
- использование не симметричных алгоритмов ключей с низкой производительностью.

Данный протокол не применим для организации защищенных групповых чатов.

### **Литература**

1. Klyubin, Alex. "Some SecureRandom Thoughts." android-developers.googleblog.com. August 14, 2013. <https://android-developers.googleblog.com/2013/08/some-securerandom-thoughts.html>.

2. github.com. "Telegram team. Telegram Android source code." April 9, 2015  
<https://github.com/DrKLO/Telegram/tree/9d113ccb7a109954f73260ecfbe5cd8467e2f7a8>.

3. core.telegram.org. "Telegram. Secret chats, end-to-end encryption." Accessed November 28, 2018. [https://core.telegram.org/api/end-to-end\\_v1](https://core.telegram.org/api/end-to-end_v1).

4. core.telegram.org. "Telegram. Perfect Forward Secrecy." Accessed December 10, 2018. <https://core.telegram.org/api/end-to-end/pfs>.

5. core.telegram.org. "Telegram. Sequence numbers in Secret Chats." Accessed December 10, 2018. [https://core.telegram.org/api/end-to-end/seq\\_no](https://core.telegram.org/api/end-to-end/seq_no).

## ГЛАВА 10

### МЕТОДОЛОГІЧНА ОСНОВА ПРОЕКТУВАННЯ ПРОБЛЕМНО-ОРІЄНТОВАНИХ ВЕБ-СЕРВІСІВ

**Вступ та постановка задачі.** Загальний розвиток існуючих універсальних комп'ютерних систем (УКС) дозволяє розв'язувати багато задач наукового, виробничо-технічного та іншого характеру. Однак існують надзвичайно важливі задачі, що пов'язані зі спілкуванням, пересуванням та прийняттям рішень людиною, в яких необхідна оперативна реакція в режимі реального часу на виникаючі перетворення. Для вирішення таких задач можливостей сучасних УКС не вистачає. Все це вимагає взаємодії системних, прикладних програмістів і кінцевих користувачів та заохочує до пошуку і впровадження нових рішень для створення нових проблемно-орієнтованих систем.

Причинами створення спеціалізованих комп'ютерних систем (СКС) є, по-перше, суперечності між методами постановки і вирішенням завдань, що важко формалізуються, з одного боку, і технічними можливостями УКС, з іншого боку. Застосування нейронних мереж (НМ) має практичне значення в будь-якій предметній області та характеризується наступним рядом переваг перед традиційними математичними методами: коли завдання не може бути формалізоване традиційними математичними методами; коли завдання формалізоване, але на теперішній час відсутній апарат для його рішення; коли для добре формалізованого завдання існує відповідний математичний апарат, але реалізація обчислень із його допомогою на базі наявних обчислювальних систем не задовольняє вимогам одержання рішень за часом, розміром, вагою, енергоспоживанням й ін. Однак, зі збільшенням розмірності вхідних сигналів структура нейронної мережі стає більш складною, що у свою чергу веде до більш тривалого часу навчання та до вповільнення швидкості збіжності. У той саме час, нейронна мережа є високопродуктивним обчислювачем, тому алгоритми, призначені для роботи з НМ, є паралельними.

Другою причиною є змістовна сторона задачі, перша особливість якої безпосередньо пов'язана із прийняттям продуманих рішень. Для прийняття рішень існує величезна різноманітність усіляких методів з різним рівнем складності. Забезпечити погоджені рішення дозволяє обчислювальна парадигма з використанням мультиагентних систем (МАС). У різних сферах діяльності мотивація застосування МАС різна, однак можна виділити основні

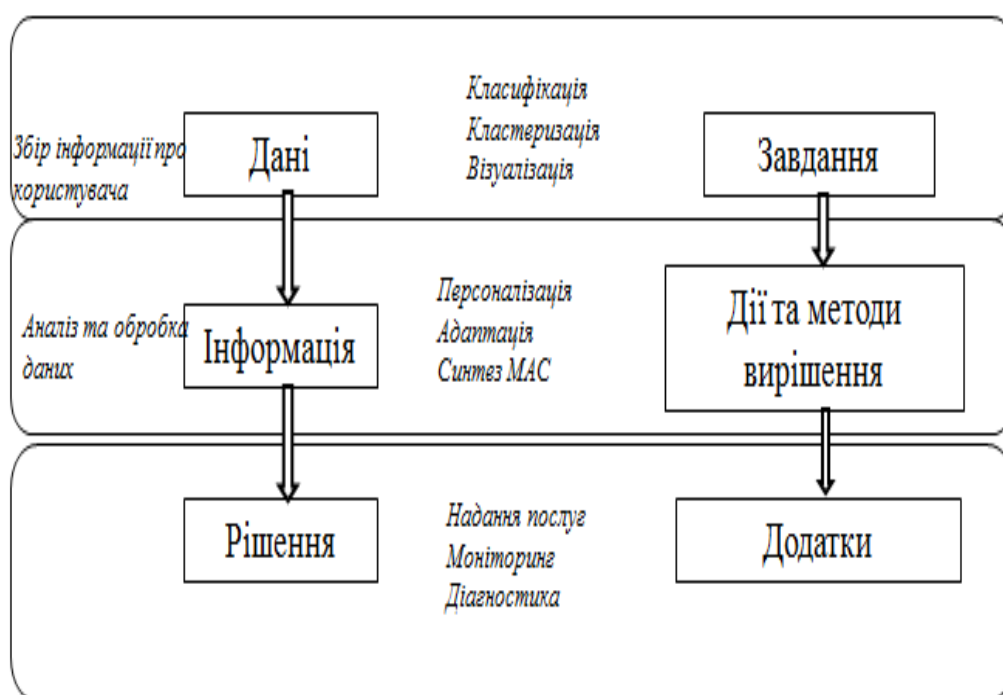
переваги застосування мультиагентних систем: суб'єкти беруть до уваги особливості додатка й навколишнього середовища; є можливість моделювання й дослідження взаємодії між суб'єктами; окремі компоненти системи моделюються на різних рівнях. Інша особливість зумовлена нагромадженням й обробкою величезної кількості інформації, що у свою чергу стосується концепції великих даних. Проблеми обробки великих даних, пов'язані з їх різноманіттям, зі складнощами збору, зберігання, управління та аналізу, об'ємом пам'яті та швидкістю обчислень. За допомогою візуалізації багатовимірних даних можна одержати ряд корисних речей, таких як витяг корисних шаблонів, виявлення шахрайства, керування факторами ризику, зниження витрат на охорону здоров'я та ін. Кількісна інформація, що отримана за результатами аналізу зображень великих розмірів, може використовуватися для прийняття рішень, наприклад, в діагностичних системах.

У зв'язку з розвитком Інтернету і мобільних пристроїв, набирає популярність використання об'єктів, що підключаються та взаємодіють між собою, із метою надання різних сервісів. Сервіс-орієнтовані веб-портالي відрізняє можливість багаторазового надання сервісів з певними інтерфейсами. Сервіси, що викликаються різними способами, не мають уявлення про додатки, які їх викликають, а додатки не мають інформації про те, яким чином послуга, що викликається, буде реалізована. Для користування віддаленими сервісами велику перевагу має служба хмарних обчислень. Проте, за відсутності Інтернет-з'єднання, доступ до таких послуг стає неможливим. В якості ефективного способу вирішення цієї проблеми пропонується ієрархія розподіленого комп'ютингу типу «cloud-fog-dew» для надання сервісу медичним веб-порталом.

Метою роботи є побудова багатоагентної системи оперативного реагування на основі розподіленої нейромережевої обробки великих даних, що дозволяє шляхом дистанційного моніторингу екстрено приймати рішення, пов'язані з критичною зміною стану досліджуваного об'єкту. Для досягнення поставленої мети потрібно розробити методологічний підхід для вирішення завдань екстреного реагування в розподіленому середовищі, які є трудомісткими та погано формалізуються.

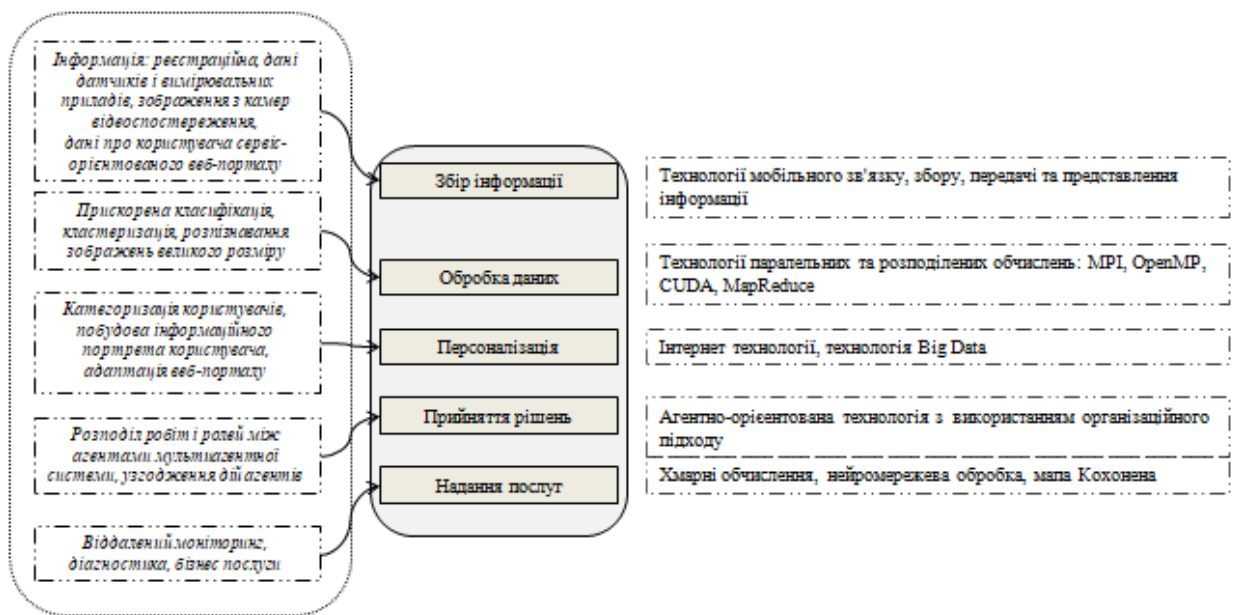
## Розробка адаптивного сервіс-орієнтованого розподіленого середовища

Для організації розподілених обчислювальних середовищ особливо значущою проблемою є розміщення у них проблемно-орієнтованих програмних комплексів, забезпечення віддаленого високорівневого доступу предметних користувачів до цих комплексів та управління ними в процесі обчислень. Цей процес є циклічним. Прийняття рішень вимагає інформації, яка заснована на даних. Дані забезпечують інформацію, яка підтримує рішення. На рис. 1 наведено зв'язок понять «різномірних даних», «обробка інформації» та «прийняття рішення», який виникає в процесі надання послуг.



**Рис. 1. Зв'язок понять для побудови проблемно-орієнтованих програмних комплексів**

Методологічний підхід [1] для вирішення трудомістких та погано формалізованих задач з екстреним реагуванням на критичну зміну стану досліджуваного об'єкта у розподіленому середовищі будується на використанні ряду технологій: технології мобільного зв'язку, збору, передачі та представлення інформації; веб-технології; технології Big Data; технології паралельних та розподілених обчислень; нейромережевої технології; агентно-орієнтованої технології та хмарних обчислень (рис. 2).



**Рис. 2. Технології організації адаптивного сервіс-орієнтованого розподіленого середовища**

Такий підхід забезпечує ефективний доступ до сервіс-орієнтованих обчислювальних середовищ, які надають сервіси найбільшому числу користувачів з використанням різноманітних датчиків та мобільних пристроїв, та дозволяє створити мультиагентну систему оперативного реагування на критичну зміну стану досліджуваного об'єкту [2].

Користувачам і додаткам надається повний набір послуг з доступу до сервісів системи (рис. 3). Користувачі веб-ресурсів виступають в ролі споживачів товарів і послуг, інформація про які розміщена на сайті, і тому для дослідження цих користувачів використовуються відповідні методи і моделі. Однак, особливості поведінки людини в Інтернеті, характер зібраних про нього даних і природа самої мережі змушують розробляти більш ефективні шляхи дослідження відвідувачів веб-сайтів, більш точного аналізу і прогнозування їх поведінки з метою підвищення ефективності планування бізнесу і просування товарів і послуг через Інтернет.

Нехай  $\epsilon$ :

– реєстраційна інформація досліджуваного об'єкту, а також дані датчиків і вимірювальних приладів  $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_r\}$ ,



**Рис. 3. Структура інтерфейсів та функцій багатоагентної системи розподіленої нейромережевої обробки великих даних**

– зображення з камер відеоспостереження, в якого значення яскравості елементів позначаються  $h_{kj}$  ( $h_{kj} = \overline{0,255}$ ,  $k=1, \overline{m}$ ,  $j=1, \overline{n}$ );

– доступна інформація про  $U(i)$ -го ( $i=1, \overline{m}$ ) користувача сервіс-орієнтованого веб-порталу Res (ім'я браузера, номер версії, мова, платформа, вбудовані розширення, адреса попередньої сторінки, часовий пояс, час відвідування сторінки, інформація про монітор і тому подібне)  $C(i) = [C^1(i), C^2(i), \dots, C^k(i)]$ .

Тоді вирішення проблеми оперативного реагування на критичну зміну стану досліджуваного об'єкту на основі аналізу зображень і показників різноманітних датчиків полягає в розробці методологічного підходу, що включає сукупність наступних завдань:

1. Рішення задачі надання сервісу  $D_H$  полягає в розробці:

– моделі процесу нейромережевої діагностики та моніторингу  $\Xi$ , її складових компонентів і їх взаємодії;

– методів обробки зображень досліджуваного об'єкту як у стані спокою, так і в стані руху.

2. Рішення задачі розподілу робіт і ролей між агентами багатоагентної системи оперативного реагування полягає в розробці:

– багатоагентної організаційної структури компонентів узагальненої моделі процесу нейромережевої діагностики і моніторингу (з описом їх взаємодії);

– методів планування і координування групових дій агентів, проведення переговорів і вибору оптимального рішення для досягнення поставленої мети.

3. Рішення задачі персоналізації  $D_p$  веб-ресурсу  $Res$  при обмеженій кількості образів веб-інтерфейсів  $\Omega$  полягає в розробці:

– моделі персоналізації  $B_R$ , її складових компонентів і їх взаємодії;

– об'єктної моделі документів Інтернет-ресурсу у вигляді графа  $Res$ ;

– методу персоналізації веб-порталу, що дозволяє для кожної  $R(n)$ -ої категорії користувачів ( $n = \overline{1, L}$ ) адаптувати інформаційне наповнення сторінок  $\overline{X}_k \in \overline{Res}$  і  $\overline{Res} \subset Res (k < \Delta)$ , де  $\Delta$  – кількість вершин графу  $Res$ .

4. Рішення завдань прискореної обробки великих даних полягає в розробці:

– моделей паралельної нейромережевої обробки даних;

– методів прискореного навчання і функціонування багатошарової нейронної мережі прямого поширення в комп'ютерних кластерах з різними топологіями передачі даних;

– методів рівномірного розподілу нейромережевої обробки даних в гетерогенному обчислювальному середовищі;

– cloud-fog-dew архітектури для сервіс-орієнтованих систем, що персоналізуються.

У загальному вигляді процес оперативного реагування на зміни стану зареєстрованого користувача веб-ресурсу, що надає сервісні послуги, описується моделлю (рис. 4), яка включає наступні компоненти [3]: веб-інтерфейс, процедурний блок, агентський блок, а також блок накопичення і аналізу досвіду.

**Кооперація агентів.** Мультиагентна система представляється трійкою  $MAS = \{A, E, R\}$ , де  $A = \{A_{user}, A_{expert}, A_{data}\}$  – множина агентів, які функціонують у середовищі, що представляє собою програмну платформу для виконання агентів і надає функціональність для створення/знищення агентів, для застосування інтелектуальних методів для забезпечення сервісів і для приймання/передавання повідомлень;  $E$  – множина станів середовища;  $R$  – веб-портал, що надає бізнес послуги, побудований на основі організаційної моделі для взаємодії агентів [4].





**Рис. 4. Зв'язок компонентів моделі системи оперативного реагування на зміну стану зареєстрованого користувача веб-ресурсу, що надає сервісні послуги**

Аспектами для керування інцидентом будуть: моніторинг стану здоров'я, постановка діагнозу й медична допомога. Для моніторингу використовується механізм «градації» стану пацієнта, що визначає діапазони вимірюваних значень, таких як: кардіограма, кров'яний тиск, рівень цукру, температура тіла тощо, які розділяються на наступні рівні ваги: критичний, високий, середній, низький.

Зобов'язання агентів  $Com(pr, us, an, re)$  означають, що при виникненні антецедента результат буде отриманий, коли виробник обслужить користувача ( $pr$  – постачальник послуг,  $us$  – користувач,  $an$  – антецедент,  $re$  – результат). Взаємодія між агентами може бути витлумачена з погляду впливу на їхні зобов'язання. Наприклад, пропозиція про прийняття ліків для зниження тиску від постачальника послуг  $A_{doctor}$  для споживача  $A_{user}$  може бути витлумачена як  $Com(A_{doctor}, A_{user}, \text{високий\_тиск}, \text{прийняти\_ліки\_для\_зниження\_тиску})$ .

Інакше кажучи, якщо в пацієнта підвищився тиск, агент доктора  $A_{doctor}$  посилає повідомлення агенту пацієнта  $A_{user}$  про прийняття ліків для зниження тиску.

Узгодження дій агентів  $A_i$  означає, що всі вони виконують зобов'язання на основі своїх спостережень [5]. Під виконанням багатоагентної системи

розуміється перехід системи з одного стану в інший. Щоразу, коли агент відправляє або одержує повідомлення, система переходить у новий стан. Стани системи виражаються через стійкість і цілісність: система перебуває в стійкому стані, якщо повідомлення не посилаються і не приймаються.

**Засіб спілкування.** Агенти обмінюються повідомленнями. Модель комунікації складається з наступних припущень:

1. Зв'язок «один-до-одного».  $A_i \cdot \text{send}(A_j, m)$  – агент  $A_i$  відправляє повідомлення  $m$  агентові  $A_j$  ( $i, j = \{\text{user, doctor, nurse, data}\}$ ).

2. Агент спостерігає тільки ті повідомлення, які він відправляє або одержує.

3. Повідомлення достовірні. Повідомлення не створюються й не руйнуються інфраструктурою.

4. Повідомлення впорядковані. Упорядковані спостереження відправляються у вигляді повідомлень. Одержувач приймає повідомлення за порядком проходження.

Спостереження агента  $A_i$  в конкретному виконанні MAS описуються послідовністю повідомлень  $\langle m_0, m_1, \dots, m_n \rangle_{A_i}$ , спостереження мультиагентної системи описуються вектором  $\text{Obs} = [\text{Obs}_{A_{\text{doctor}}}, \text{Obs}_{A_{\text{user}}}, \text{Obs}_{A_{\text{nurse}}}, \text{Obs}_{A_{\text{data}}}]$ , де  $\text{Obs}_{A_i}$  – послідовність спостережень агента  $A_i$ .

Спостереження  $\text{Obs}$  агентів  $\forall A_i, A_j \in \text{MAS}$  слухні в таких випадках:

– якщо повідомлення  $A_i \cdot \text{send}(A_j, m_k)$  належить  $\text{Obs}_{A_j}$ , тоді повідомлення  $A_i \cdot \text{send}(A_j, m_k)$  належить і  $\text{Obs}_{A_i}$  ( $k = \overline{0, n}$ );

– якщо повідомлення  $A_i \cdot \text{send}(A_j, m_1)$  належить  $\text{Obs}_{A_j}$  і повідомлення  $A_i \cdot \text{send}(A_j, m_0)$  передуює повідомленню  $A_i \cdot \text{send}(A_j, m_1)$  в  $\text{Obs}_{A_j}$ , тоді повідомлення  $A_i \cdot \text{send}(A_j, m_0)$  передуює повідомленню  $A_i \cdot \text{send}(A_j, m_1)$  в  $\text{Obs}_{A_i}$ . Таким чином, спостереження  $\text{Obs}_{\text{MAS}}$  являє собою множину усіх станів системи MAS.

**Стійкість.** Система перебуває в стійкому стані, якщо ніякі повідомлення не приймаються і не передаються. Спостереження перебувають у стійкому стані, якщо  $\forall A_i, A_j \in \text{MAS}$  повідомлення  $A_i \cdot \text{send}(A_j, m_k)$  належить  $\text{Obs}_{A_j}$ , тоді повідомлення  $A_i \cdot \text{send}(A_j, m_k)$  належить  $\text{Obs}_{A_i}$  ( $k = \overline{0, n}$ );

**Цілісність.** Нехай  $Obs_{A_j}$  послідовність повідомлень виду  $\langle m_0, m_1, \dots, m_n \rangle_{A_i}$ . Тоді для будь-якого повідомлення  $m'$  форма  $\langle m_0, \dots, m_n, m' \rangle_{A_i}$  є конкатенацією  $Obs_{A_i}$  з  $m'$ . Множину пропозицій  $Seq(Obs_{A_i})$ , яку можна вивести з множини  $Obs_{A_i}$ , являє собою стани  $A_i$  після спостереження повідомлень в  $Obs_{A_i}$ . Обмеження цілісності виглядає в такий спосіб:  $[m[Ant2:Ant1]m']_{A_i}$ , де  $Ant2$  й  $Ant1$  перед- і післяумови впливу на  $m$ ,  $m'$  – ефект впливу на  $m$  за умови виконання перед- і післяумов.

**Узгодженість.** Узгодженість в багатоагентній системі виглядає в такий спосіб:  $[\langle MAS \rangle]$ , якщо  $\forall Obs_{A_i} \in Obs_{MAS}$  спостереження  $Obs_{A_i}$  стійкі й повноцінні щодо обмежень цілісності:

$$\forall A_i, A_j \in MAS : Com(pr, us, an, re) \in Seq(Obs_{A_i}) \Rightarrow Com(pr, us, an, re) \in Seq(Obs_{A_j})$$

Таким чином, спостереження того самого антецеденту постачальником послуг і користувачем приводить до здійснення однієї дії для досягнення бажаного результату.

Кожний агент описується наступним набором  $A_i = \{A_i^a, E_i^e, Feaa, A_i^s, Fase, Fasa a\}$ , де  $A_i^a$  – множина дій агентів:  $i=user$ : «Агент користувача»  $A_{user}$  збирає інформацію про користувача веб-ресурсу й повертає результат (наприклад, при обробці зображення новоутворень шкіри: попередній діагноз – невус);  $i=expert$ : «Агент експерта»  $A_{expert}$  звертається за наданням послуг (аналіз медичних зображень, віддалений моніторинг стану здоров'я людини тощо);  $i=data$ : «Агент обробки даних»  $A_{data}$  передає/приймає інформацію для нагромадження й аналізу інформації;  $E_i^e$  – множина станів зовнішнього середовища;  $Feaa : A_i^a \times E_i^e \rightarrow 2^{A_i^a}$  – функції поведінки зовнішнього середовища;  $A_i^s$  – множина внутрішніх станів агентів (передає, одержує або очікує повідомлення);  $Fase : A_i^s \times E_i^e \rightarrow A_i^s$  – функції відновлення стану агента;  $Fasa a : A_i^s \rightarrow E_i^e$  – функції прийняття рішень для здійснення дій агентом за поточним внутрішнім станом:

– відображення  $\Pi : P \rightarrow \Lambda$  дозволяє для кожного користувача веб-ресурсу одержати таке  $\lambda_j \in \Lambda$  ( $j = \overline{1, \Omega}$ ), яке є рішенням завдання адаптації веб-ресурсу, ( $P$  – множина інформаційних портретів Інтернет-Користувачів,  $\Lambda$  – множина інтерфейсів веб-ресурсу,  $\Omega$  – кількість кластерів);

– функція  $\Phi : C \rightarrow S$ , що відображає, для кінцевої множини доступних даних про користувача формує  $S_n \in S$  ( $n = \overline{1, N}$ ), яке є рішенням завдання автоматичного породження гіпотез про причини наявності або відсутності певних властивостей Інтернет-Користувача ( $C$  – множина доступних даних про користувача,  $S$  – гіпотези, що є причиною наявності цільової властивості,  $N$  – кількість цільових властивостей);

– відображення  $\Gamma : G \rightarrow \mathfrak{R}$  дозволяє для будь-якого Інтернет-Користувача одержати таке  $\pi_i \in \mathfrak{R}$  ( $\pi_i$  – стаціонарна ймовірність розподілу процесів  $X = \{X_t, t \geq 0\}$ ,  $X_t$  – сторінка, яку користувач відвідав під час  $t$ , ( $t > 0$ )), який є рішенням завдання аналізу поведінки користувача ( $G$  – граф веб-серфінгу користувача) тощо.

Планувати дії агент може у такий спосіб:  $AP = \{A_i^a, P_i^{ap}, I_i^{ap}, \rho_{ap}, \iota_{ap,0}\}$ , де  $P_i^{ap}$  – множина сприйняття агентом станів зовнішнього середовища;  $I_i^{ap}$  – підмножина внутрішніх станів агента, що є частиною множини внутрішніх станів агента  $I = I_i^{ap} \times I'$ ;  $\rho_{ap} \subseteq P_i^{ap} \times I_i^{ap} \times I_i^{ap} \times A_i^a$  – відношення переходів, що визначає по сприйняттю  $p_i^{ap}$  поточного сприйняття зовнішнього середовища й поточному внутрішньому стану плану  $\iota_i^{ap}$ ;  $\iota_{ap,0} \in I_i^{ap}$  – початковий стан агента.

Таким чином, результатом планування буде множина ланцюжків переходів агентів з початкового стану в кінцевий, що реалізує поставлену мету.

**Методи та засоби забезпечення інформаційної безпеки.** В даній медичній системі обов'язково повинен бути реалізований ряд заходів щодо забезпечення безпеки як інформації, так і інформаційної системи в цілому. Будь-який користувач, що одержує доступ до системи, несе повну відповідальність за забезпечення конфіденційності інформації, яку він вносить, використовує або передає іншим користувачам. Таким чином, для використання системи необхідно вирішення низки програмних, організаційно-технічних питань. Більш конкретно під безпекою інформації будемо розуміти стан захищеності інформації, що обробляється засобами обчислювальної техніки, від внутрішніх або зовнішніх загроз, а також захист інформації від витоку, розкрадання, втрати, несанкціонованого знищення, викривлення, модифікації (підробки), копіювання і блокування.

На рис. 5 представлена модель взаємодії технічних, організаційних і програмних компонент, які забезпечують інформаційну безпеку в системі.

Робота в системі ведеться в розрахованому на багато користувачів режимі з розмежуванням прав доступу за рахунок ідентифікації суб'єктів.



**Рис. 5. Модель взаємодії компонентів інформаційної безпеки**

При вході в систему і видачу запитів на доступ проводиться аутентифікація користувачів, яка має в своєму розпорядженні необхідні дані для ідентифікації, аутентифікації, а також перешкоджає несанкціонованому доступу до ресурсів. Щоб забезпечити безпеку даних, система захисту обмежує доступ до інформації навіть всередині однієї групи користувачів. Так, лікар має доступ до своїх історій хвороби і до архівних історій хвороби тих пацієнтів, яких він курирував; завідувач відділенням – тільки до історій хвороби пацієнтів свого відділення; лікар-консультант отримує доступ до історій хвороби заданих (адміністратором безпеки) відділень і тільки при наявності відповідного напрямку в електронній історії хвороби.

Особливість архітектури системи в тому, що дані медичного закладу зберігаються на центральному сервері, який розміщується в захищеному приміщенні. На комп'ютерах медичного персоналу дані не зберігаються. Серверне обладнання і робочі місця медичного персоналу об'єднані в локальну обчислювальну мережу і забезпечують надійну роботу інформаційної системи. Всі спроби входу в систему фіксуються в

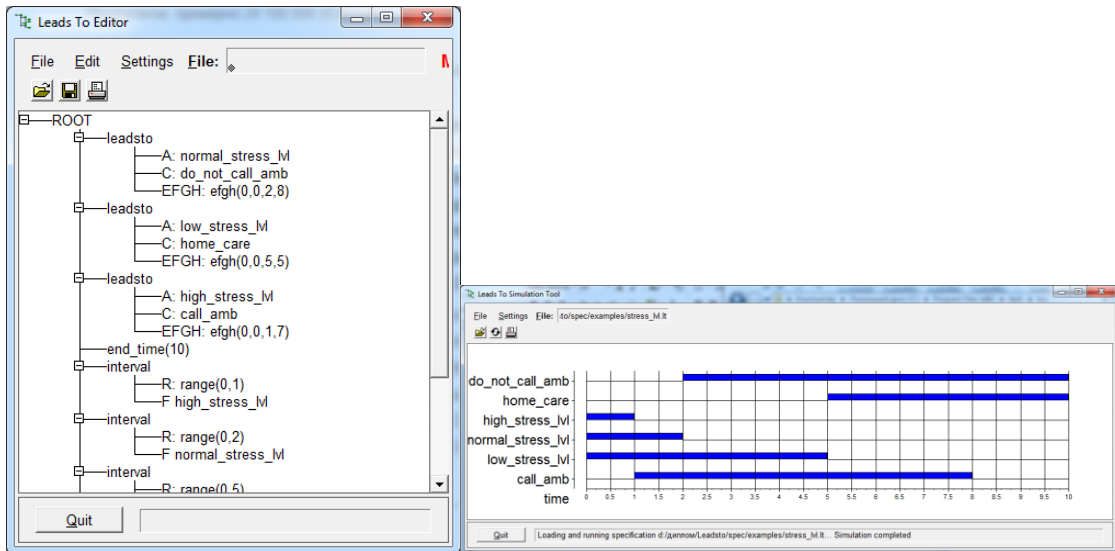
електронних журналах.

Захист інформації від несанкціонованого доступу забезпечено блокуванням доступу до інформації: для СУБД – з боку як персоналу, так і тих завдань системи, яким дана інформація не потрібна в силу функціонального призначення; на робочому місці – з боку користувачів, що не володіють відповідними повноваженнями на доступ до різних інформаційних ресурсів; по каналах зв'язку – з боку мережевих користувачів і тих завдань системи, яким дана інформація не потрібно в силу функціонального призначення.

**Моделювання.** Виконано моделювання, засноване на спостереженнях за пацієнтом за показниками датчиків [6]. Адаптаційні процеси моделювалися як характеристики нижчого рівня (наприклад, рольові характеристики), які були переведені в підмножину TTL, називану leads-to. Для цього використана мова LEADSTO, а також TTL Editor для перевірки правильності моделювання. Спрощений формат leads-to дозволяє моделювати прямі тимчасові залежності між двома характеристиками стану в такий спосіб. Нехай  $\alpha$  і  $\beta$  – характеристики стану «кон'юнкції літералів» (де літерал – це атом), і  $e, f, g, h$  – невід'ємні дійсні числа. У мові LEADSTO вираз  $\alpha \rightarrow \rightarrow_{e,f,g,h} \beta$  означає наступне: якщо характеристика стану  $\alpha$  знаходиться на певному інтервалі часу із тривалістю  $g$ , то після деякої затримки (між  $e$  і  $f$ ) характеристика стану  $\beta$  буде перебувати на певному інтервалі часу довжиною  $h$ .

Реалізована окрема постановка завдання – нагляд за рівнем артеріального тиску пацієнта та прийняття мір для збереження здоров'я.

У завданні, що пов'язане з артеріальним тиском, змінні пов'язані саме з тиском: `high_stress_lvl`, `normal_stress_lvl`, `low_stress_lvl`. Також, змінні для позначення того, що саме потрібно робити з показниками тиску: `home_care`, `call_amb`, `do_not_call_amb`. Задано час для кінцевого моделювання та додано інтервали для кожної зі змінних артеріального тиску (наприклад 10 одиниць), а також початок сліду та кінець. Кінець сліду позначає початок наслідку для даної змінної. Змінна `high_stress_lvl` сигналізує, що потрібно викликати швидку й надавати негайну допомогу пацієнту. Усіма часовими відмітками пояснене, що якщо вираз буде носити правдивий характер з тривалістю  $g$ , то весь вираз буде мати значення тривалістю  $h$  із затримкою 0.0 (рис. 6).



**Рис. 6. Часові інтервали для усіх Leads-to-умов**

Для перевірки коректної роботи специфікації запускається LEADSTO Simulation Tool, і якщо під час введення специфікації не виникло ніяких помилок, то буде показано результат на зразок рис. 7.

```

SWI-Prolog (Multi-threaded, version 5.5.26)
File Edit Settings Run Debug Help

LEADSTO version 1.27

1 ?- Loading and running specification d:/диплом/Leadsto/сpec/examples/stress_lvl.lt...
% d:/диплом/Leadsto/сpec/examples/stress_lvl.lt compiled into spec 0.00 sec, 2,604 bytes
HANDLED TIME -> 8

HANDLED TIME -> 9

HANDLED TIME -> 10

SAVED TRACE IN trace.tr
TIME:0.00399995 CPUTIME:0.0

Simulation completed

```

**Рис. 7. Успішно виконана специфікація**

Розроблена експериментальна система добре узгоджується із запропонованою архітектурою cloud-fog-dew для сервіс-орієнтованих систем. Ця система включає чотири типи програмного забезпечення.

1. Персональне програмне забезпечення встановлюється на особистому цифровому пристрої пацієнта. Воно взаємодіє з медичною апаратурою (датчики, вимірювальні пристрої тощо), яка збирає інформацію про пацієнта та відправляє оцифровану інформацію Master-серверу.

2. Програмне забезпечення обслуговуючого персоналу з'єднується з Master-сервером і отримує дані відповідно до вимог користувача. Це

програмне забезпечення дозволяє взаємодіяти з даними про стан пацієнта незалежно від місця розташування спостерігача.

3. Серверне програмне забезпечення створене із застосуванням об'єктно-орієнтованої мови програмування C# для платформи .NET Framework та встановлено на Master-сервері. Воно обробляє і зберігає дані про пацієнтів. Для зберігання інформації про користувачів обрана база даних MS SQL Server. Підключення до бази даних здійснюється за допомогою ORM (Object-Relational Mapping).

4. Резервне копіювання даних і їх відтворення – це механізм, що забезпечує безпеку даних при збої в роботі серверів.

Веб-інтерфейс системи реалізований за допомогою мови гіпертекстової розмітки HTML, каскадних таблиць стилів CSS і скриптової мови JavaScript. Мова JavaScript дозволила отримати максимально повну інформацію про користувача, яка застосовувалася для автоматичної локалізації сторінки, а також про розміри дисплея, що дозволило автоматично перебудовувати ресурс в залежності від них.

Реалізована окрема постановка задачі первинної діагностики гіпертонії на підставі показань цифрового вимірювача артеріального тиску. У процесі моніторингу пацієнт віддалено відповідає на питання, наведені в анкеті. На підставі відповідей формується вектор опитування, який разом з результатом обробки тонометра надходить на вхід нейронної мережі для постановки первинного діагнозу. Залежно від отриманих результатів мобільний агент  $A_{\text{doctor}}$  посилає рекомендації про лікування, дає розпорядження обслуговуючому персоналу або здійснює дзвінок у швидку допомогу.

В ході експериментів лікарів спочатку пропонувалося зробити діагноз самостійно, потім з використанням розробленої системи (табл. 1).

Критерієм ефективності розробки є задоволення вимог  $k: \forall (D_i \in D_H)[(\tau^a < \tau^{\max}) \& (\tau^p < \tau^{\max}) \& (\tau^r < \tau^{\max}) \& (\varpi > \varpi^{\min})] \Rightarrow Y_D$ , де  $D_i$  – подзадача спільної справи надання сервісу  $D_H$ ;  $\tau^a$  – час адаптації локального сайту,  $\tau^p$  – час обробки медико-діагностичних показників,  $\tau^r$  – час реакції системи;  $\varpi$  – ступінь релевантності відображеної інформації.



### Результати тестування розробленої системи віддаленого моніторингу стану здоров'я пацієнта

Клас	Кількість прикладів	Розподіл результатів по класах					
		без використання системи			з використанням системи		
		норма	домашній догляд	виклик швидкої допомоги	норма	домашній догляд	виклик швидкої допомоги
норма	100	90	7	3	95	3	2
домашній догляд	100	3	91	6	2	96	2
виклик швидкої допомоги	100	0	9	91	0	5	95

На рис. 8(а) показана імітація процесу реєстрації пацієнта на локальному комп'ютері. Скріншот інтерфейсу пацієнтів для користувачів з мобільним пристроєм показано на рис. 8(б).



Рис. 8. Імітація процесу реєстрації пацієнта

Запропонована система допоможе пацієнтам, які перебувають на обліку в медичному закладі, своєчасно отримати медичну допомогу за рахунок віддаленого автоматичного контролю стану здоров'я.

#### Висновок.

В роботі розглядається методологічний підхід до побудови сервіс-орієнтованого середовища, який містить складові, що враховують різноманітність даних великого обсягу, засоби подолання проблем інтеграції існуючих і розроблених програмних модулів в єдиний комплекс. Запропоновані методи і моделі побудови проблемно-орієнтованих сервісів

забезпечують високорівневий доступ зареєстрованих користувачів до спеціалізованих програмних комплексів, що дозволяє оперативно реагувати на зміни, які відбуваються у реальному часі. Запропоновано модель забезпечення інформаційної безпеки, що передбачає створення системи захисту медичної інформації та компонентів інформаційної системи за рахунок попередження несанкціонованого отримання інформації, її зміни та утримання.

### Література

1. Н. Г. Аксак, “Концепція побудови мультиагентних систем розподіленої нейромережевої обробки великих даних”, Вісник ХНТУ, т.1, № 3(66), с. 205-212, 2018.
2. N. Axak, D. Rosinskiy, O. Barkovska, I. Novoseltsev, “Cloud-fog-dew architecture for personalized service-oriented systems”, in Proc. 9th IEEE Int. Conf. on Dependable Systems, Services and Technologies, DESSERT’2018, Kyiv, Ukraine, 2018, pp. 80-84
3. N. Axak, “Development of multi-agent system of neural network diagnostics and remote monitoring of patient”, Eastern-European Journal of Enterprise Technologies, vol. 4 No 9(82), pp. 4-11, 2016.
4. Н. Г. Аксак, Н. М. Кораблев, “Дистанционный мониторинг пациента на основе концепции агент-группа-роль”, Системні дослідження та інформаційні технології, № 3, с. 7-18, 2018.
5. Н. Г. Аксак, Н. М. Кораблев “Согласование многоагентных обязательств в системе удаленного мониторинга состояния здоровья человека”, Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: сборник научных трудов, Херсон: Видавництво “ПП Вишемирський В.С.”, с. 139-141, 2018.
6. Н. Г. Аксак, И. В. Новосельцев, Д. Н. Росинский “Система предоставления медицинских услуг на основе архитектуры cloud-fog-dew”, Наука та виробництво: міжвуз. темат. зб. наук. пр., Вип. 19, ДВНЗ “ПДТУ”, Маріуполь, ПДТУ, с. 150-159, 2018.

## ГЛАВА 11

### НЕЧІТКО-МНОЖИННА МЕТОДИКА ОЦІНКИ КОНКУРЕНТОСПРОМОЖНОСТІ ПІДПРИЄМСТВА

**Вступ і постановка задачі.** Поряд з теоретичними дослідженнями сутності ринкових відносин, в сучасній літературі обговорюється проблема практичної оцінки конкурентоспроможності суб'єктів економічної діяльності. Можна констатувати, що в питаннях оцінки конкурентоспроможності продукції на сьогоднішній день досягнуто певних успіхів, розроблені цілком прийнятні методики оцінки показників конкуренції ідентичних товарів і послуг. Складніша справа з оцінкою конкурентоспроможності підприємств. Незважаючи на те, що певні кроки в цьому напрямку робилися і робляться, універсальної і загальновизнаної методики комплексної оцінки конкурентоспроможності підприємства економістами в даний час не вироблено. У той же час, потреба в оцінці перспектив розвитку того чи іншого підприємства існує, оскільки в умовах ринкової економіки оцінка своїх конкурентних позицій є невід'ємним елементом діяльності будь-якого господарюючого суб'єкта [1].

Вивчення конкурентів і умов риночних відносин в галузі потрібно підприємству в першу чергу для того, щоб визначити, в чому його переваги і недоліки перед ними, і зробити висновки для вироблення підприємством власної успішної стратегії і підтримки його переваги. Визначення конкурентоспроможності підприємства є невід'ємним елементом діяльності будь-якого господарюючого суб'єкта. Зокрема, оцінка подальшої ефективної діяльності господарюючого суб'єкта необхідна в цілях: розробки заходів по підвищенню конкурентоспроможності; вибору контрагентів для спільної діяльності; складання програми виходу підприємства на нові ринки збуту; здійснення інвестиційної діяльності; здійснення державного регулювання економіки на основі висновку про стан відповідного ринку.

У будь-якому випадку здійснення оцінки конкурентоспроможності підприємства має на меті визначити положення підприємства на ринку. Основним завданням кожного спеціаліста, який вивчає проблему конкурентоспроможності підприємств, є визначення ефективного методу оцінювання та критеріїв, що характеризують діяльність суб'єкта господарювання.

В економічній літературі розглядається велика кількість підходів до методів оцінки конкурентоспроможності підприємства, їх зміст наведено в табл.1 [1-3].

**Таблиця 1**

**Зміст найбільш поширених методів оцінювання конкурентоспроможності підприємства**

Метод	Зміст методу
1	2
Матриця „Бостонської консалтингової групи” (БКГ)	Суть оцінки полягає в аналізі матриці, побудованої за принципом системи координат: по горизонталі – темпи зростання/скорочення кількості продажів в лінійному масштабі; по вертикалі – відносна частка товару (послуги) на ринку. Найбільш конкурентоздатними вважаються підприємства, які займають значну частку на швидкозростаючому ринку.
Матриця McKinsey	На відміну від матриці БКГ є багатофакторною матрицею і більш детальним варіантом. Вона складається уже не з чотирьох, а з дев'яти квадрантів і характеризує довгострокову привабливість галузі (ринку) та конкурентну позицію підприємства в ній.
Метод «профільнів»	Полягає в ідентифікації та кількісній оцінці характеристик, які визначають ступінь лояльності споживачів до підприємства. Профіль являє собою сукупність специфічних параметрів, які характеризують об'єкт аналізу і завдяки яким він відомий цільовій групі споживачів. Після того дається їхня графічна інтерпретація, яка дозволяє наочно побачити позиції, по яких мається відставання від конкурента або випередження його результатів.
SWOT-аналіз	Базується на оцінці сильних і слабких сторін підприємства з одного боку, і можливостей і загроз зовнішнього середовища. На основі аналізу формулюється стратегія підприємства, що дозволяє використовувати можливості підприємства з урахуванням впливу факторів зовнішнього середовища.
Багатокутник конкурентоспроможності	Побудова багатокутників конкурентоспроможності здійснюється наступним чином: коло ділять радіальними оціночними шкалами на рівні сектори, кількість яких дорівнює кількості обраних критеріїв; шкали на радіальних прямих градуюють так, щоб всі значення критеріїв знаходилися всередині оціночного кола; значення критеріїв збільшуються по мірі віддалення від центра кола; на кожній вісі з використанням відповідного масштабу вимірювання позначаються точки, що відповідають значенням критеріїв; по точках відповідно для кожного з підприємств необхідно провести ламану лінію, яка і формує багатокутник.
Метод, що базується на теорії рівноваги фірми	Оцінка рівня конкурентоспроможності здійснюється шляхом співставлення не лише виробничих витрат, а і обсягів та норми прибутку, та/або обсягів продажів, та/або ринкових часток. Більш високий рівень показника при співставленні вважається за достатню умову для того, щоб оцінити підприємство як більш конкурентоспроможне.

## Закінчення таблиці 1

1	2
Бенчмаркінг	Успішно застосовується у практичній діяльності японських, американських, західноєвропейських та інших компаній. Особливості бенчмаркінга: загальносистемний характер (обхват всіх аспектів діяльності компанії, включаючи її корінні зміни як системи в цілому); чітка спрямованість на досягнення кращого рівня; прикладна орієнтація на досягнення конкурентних переваг. Бенчмаркінг – це безперервний, систематичний пошук вивчення кращої практики конкурентів і організацій з суміжних галузей, а також постійне порівняння бажаних змін і результатів бізнесу із створеною еталонною моделлю.
PEST- аналіз	Основним завданням є мінімізація впливу негативних факторів на можливості підприємства. Аналіз здійснюється за схемою «чинник - підприємство». Результати аналізу оформлюються у вигляді матриці, на горизонтальній осі якої визначаються чинники макросередовища, на вертикальній - сила їхнього впливу в балах, рангах або інших одиницях виміру.
Метод, що базується на оцінці конкурентоспроможності продукції підприємства	Передбачає оцінювання рівня конкурентоспроможності підприємства виходячи із споживчої цінності виробленої продукції. Згідно даного методу, конкурентоспроможність виробника є тим вищою, чим вищою є конкурентоспроможність його продукції. Для оцінки конкурентоспроможності товару або послуги підприємства, використовується співвідношення двох характеристик: якості і ціни. Чим вищою є різниця між споживчою вартістю товару для покупця і ціною, котру він за нього сплачує, тим вищим є запас конкурентоспроможності.
Метод, заснований на теорії ефективної конкуренції	Використовується співставлення показників стану підприємства з показниками підприємств - конкурентів та з середньогалузевими показниками. Згідно теорії ефективної конкуренції, найбільш конкурентоспроможними вважаються ті підприємства, де найкращим чином організована робота всіх підрозділів та служб з точки зору ефективності використання ними ресурсів. В основі методу лежить оцінка чотирьох групових показників чи критеріїв конкурентоспроможності із подальшим розрахунком інтегрального показника.
Метод інтегральної оцінки	Ґрунтується на використанні під час оцінки конкурентоспроможності підприємства інтегрального показника, який включає в себе два елементи: критерій, що відображає ступінь задоволення потреб споживача – характеризує відносну конкурентоспроможність товару; критерій ефективності виробництва. Використовується рентабельність активів, рентабельність власного капіталу або середня норма рентабельності за певний період.

Аналізуючи зазначені методи оцінки конкурентоспроможності підприємств можна зробити наступні висновки:

значна кількість методик не заснована на методах прогнозування, оптимізації, нормування;

більшість наведених методик не є комплексними, тобто вони не охоплюють технічні, економічні, соціальні, управлінські та інші аспекти управління;

склад об'єктів оцінки конкурентоспроможності не відповідає їх цілісності, серед них відсутня оцінка зовнішніх факторів конкурентоспроможності: оподаткування, грошово-кредитної політики тощо;

методики орієнтовані не в майбутнє, а в минуле (сучасний стан оцінюється за минулими даними), оцінюється фактична конкурентна позиція, а не потенційна можливість до її збільшення;

переважна кількість методів включає експертні оцінки, що вносить до оцінки значну частку суб'єктивізму і ставить залежність адекватної оцінки конкурентоспроможності від професіоналізму експертів.

Наведені в табл.1 методи оцінювання конкурентоспроможності підприємства досить різні в силу багатоаспектності, складності та наукової дискусійності поняття конкурентоспроможності підприємства. Для обрання певного методу оцінювання конкурентоспроможності підприємства доцільно, по-перше, звернути увагу на переваги та недоліки кожного з них.

Зауважимо, що жоден з існуючих підходів до оцінювання конкурентоспроможності підприємств не знайшов широкого застосування в практиці економічного аналізу [1]. Це дозволяє зробити висновок про те, що універсальної методики комплексної оцінки конкурентоспроможності підприємства в даний час не існує. Це обумовлено тим, що запропоновані підходи мають цілу низку недоліків:

система чинників конкурентоспроможності підприємства є відкритою, а безліч елементів цієї системи – нечітким;

кількість факторів конкурентоспроможності практично нескінченна, і який би не був повний їх перелік, він все одно не буде вичерпним, а значить, і заснована на такому неповному переліку оцінка конкурентоспроможності підприємства буде неадекватною;

використання ж значної кількості факторів дає при розрахунках значний рівень кореляції показників, що нівелює оцінку конкурентоспроможності.

Для оцінки виявлених дослідниками чинників конкурентоспроможності, а також визначення ряду інших показників використовуються експертні методи, які відрізняє істотна суб'єктивність і умовність. У ряді випадків уникнути такого підходу неможливо, однак використання подібних оцінок призводить до дуже слабкого математичного зв'язку вихідних факторів з оцінюваним показником конкурентоспроможності.

В комплексних оцінках дослідники вводять коефіцієнти, що визначають вагове значення кожного з оцінюваних чинників, однак застосовувані коефіцієнти в більшості випадків досить умовні, що тягне за собою неадекватність оцінки впливу тих чи інших факторів на конкурентоспроможність підприємства [2]. Окрім цього, різні економічні фактори в кожній конкретній економічній ситуації в різній мірі впливають на конкурентоспроможність різних підприємств, тому неадекватним є явне встановлення єдиних вагових коефіцієнтів для оцінки конкурентоспроможності різних господарюючих суб'єктів.

Більшість методик передбачає порівняння практично ідентичних підприємств, що виробляють схожі товари і послуги в схожих економічних умовах. Проте наразі, розвиток товарно-грошових відносин призводить до все більшої диверсифікації підприємств та більшої диференціації товарів і послуг, все більш посилюються відмінності в економічних умовах діяльності підприємств. Це тягне за собою не адаптованість подібних методик оцінки конкурентоспроможності підприємств [3].

Таким чином, в даний час не існує універсальної методики оцінки конкурентоспроможності підприємства. Це пояснюється величезною кількістю факторів, що мають вплив на конкурентоспроможність підприємства, а також складністю збору, обробки та об'єктивної оцінки багатьох з них. Крім того, лише два методи враховують вплив зовнішніх факторів, в тому числі оподаткування, на конкурентоспроможність підприємства: SWOT-аналіз та PEST-аналіз. Але, враховуючи те, що вказані методи не передбачають кількісної оцінки рівня конкурентоспроможності підприємства, постає необхідність в розробці методичного підходу до оцінювання рівня конкурентоспроможності підприємства.

### **Основна частина.**

На даний момент, теорія нечітких множин знайшла досить широке застосування і в техніці, і в економіці. Однак у вітчизняній практиці аналізу конкурентоспроможності підприємства ці методи використовуються або вкрай рідко, або ж взагалі не використовуються.

Метод	Переваги методу	Недоліки методу
Матриця БКГ	Простота, доступність, наочність, використання об'єктивних критеріїв привабливості та конкуренто-спроможності, зменшення рівня суб'єктивізму. Дає змогу дослідити розвиток процесів конкуренції в динаміці та про обсяги реалізації, дозволяє	Надмірна спрощеність, труднощі з оцінкою і визначенням масштабів ринку, ринкової частки підприємства і темпів зростання ринку. Неможливість проведення аналізу причин проблем, внаслідок чого ускладнюється вироблення рішень. Не забезпечує необхідного
Матриця McKinsey	Глибина (не призводить до спрощених висновків), гнучкість (згадані чинники вибираються відповідно до конкретної ситуації, що склалася на ринку), широта	Труднощі і суб'єктивізм при визначенні ваги певного чинника, рекомендації мають загальний характер і потребують уточнень. Не забезпечує необхідного системного погляду на
Багатокутник конкурентоспроможності	Простота, доступність, наочність	Не дає змоги встановити узагальненого критерію конкурентоспроможності. Не забезпечує системного погляду
Метод, що базується на оцінці конкурентоспроможності продукції підприємства	Враховує найбільш важливий критерій, що впливає на конкурентоспроможність підприємства – конкурентоспроможність товару	Дозволяє отримати обмежене уявлення щодо переваг і недоліків у роботі підприємства, конкурентоспроможність підприємства ототожнюється
Метод, що базується на теорії рівноваги фірми	Простота, доступність	Не дає змоги зробити цілісні висновки про рівень конкурентоспроможності
Метод, заснований на теорії ефективної конкуренції	Допомагає виявити сильні і слабкі сторони одного підприємства стосовно іншого, оцінити розмір	Досить складно зібрати всю необхідну інформацію; для достовірної оцінки необхідно здійснювати дуже складні та

**Рис. 1. Переваги та недоліки найбільш поширених методів оцінювання конкурентоспроможності підприємства**



Метод	Переваги методу	Недоліки методу
Метод інтегральної оцінки	Простота, наочність, дає змогу отримати однозначні оцінки конкурентних позицій	Може бути застосованим лише для оцінки конкурентоспроможності монопродуктових підприємств, нереально знайти для порівняння два чи більше підприємства - конкуренти, абсолютно
Бенчмаркінг	Зіставлення показників ефективності дає можливість зрозуміти уразливі і раціональні сторони діяльності компанії, порівняно з конкурентами і аналогічними підприємствами	Адекватна оцінка стає неможливою за відсутності достовірної інформації про конкурентів
SWOT-аналіз	Метод забезпечує високу наочність при використанні достовірної інформації, детальний аналіз зовнішнього середовища	Відсутність будь-яких кількісних або оцінюють показників, суб'єктивізм, не дозволяє оперативно
PEST-аналіз	Забезпечення цілісної картини зовнішнього оточення, детальний аналіз зовнішнього середовища, стимулювання керівника до дії і обліку зовнішніх факторів	Не враховує «внутрішні» фактори, підходить тільки для аналізу довгострокових цілей і не дозволяє оперативно реагувати на зміни

**Рис. 1. Переваги та недоліки найбільш поширених методів оцінювання конкурентоспроможності підприємства (закінчення)**

Для того, щоб втілити методи нечітких множин, необхідно пройти декілька етапів [4]:

визначення набору окремих показників, які впливають на оцінку конкурентоспроможності підприємства;

встановлення рівня впливу кожного показника на оцінку конкурентоспроможності підприємства;

побудова рівнів класифікації значень кожного показника;

оцінка поточного рівня показників;

визначення нечіткості поточних значень кожного показника;  
встановлення інтегрального показника конкурентоспроможності підприємства;

визначення поточного значення інтегрального показника та лінгвістичне розпізнавання рівня конкурентоспроможності підприємства.

Вибір показників для кількісного аналізу конкурентоспроможності підприємств відповідає опису різних сторін функціонування підприємств на ринку, зокрема – зовнішню і внутрішню. Для опису методики розглянемо тільки внутрішню сторону функціонування підприємств.

Побудуємо набір окремих показників  $X = \{x_i\}$  загальним числом  $N$ , які, на думку експерта-аналітика, з одного боку, впливають на оцінку конкурентоспроможності підприємства, а з іншого боку, оцінюють різні підприємства:  $x_1$  – обсяг реалізації;  $x_2$  – чисельність працюючих;  $x_3$  – власний капітал підприємства;  $x_4$  – сукупні зобов'язання;  $x_5$  – чистий прибуток;  $x_6$  – частка ринку;  $x_7$  – конкурентоспроможність продукції.

Зіставимо кожному показнику  $X_i$  рівень його значущості. При цьому будемо враховувати що всі показники мають однаковий вплив на конкурентоспроможність підприємства.

Побудуємо рівні класифікації значень кожного показника. Для довільного окремого показника  $x_i$  повна безліч його значень розбивається на чотири підмножини:

$b_1$  – підмножина "високий рівень показника  $x_i$ ";

$b_2$  – підмножина "підвищений рівень показника  $x_i$ ";

$b_3$  – підмножина "середній рівень показника  $x_i$ ";

$b_4$  – підмножина "низький рівень показника  $x_i$ ".

Причому тут і далі за замовчуванням припускаємо, що зменшення окремого показника  $x_i$  пов'язано зі зниженням конкурентоспроможності підприємства.

У постановці задач багатокритеріальної оптимізації центральне місце займають питання порівняння приватних критеріїв і завдання вимог до можливим співвідношенням їх значень в точці оптимуму, тобто вимог до оптимальності рішення, на підставі яких формується глобальний критерій якості. Ці питання розглядаються без урахування відмінностей між показниками та їх впливу на критерії та обмеження, які однаково чиним беруть участь у формуванні глобального показника якості в разі, якщо вони описуються функціями приналежності. Для об'єктивного аналізу конкурентоспроможності підприємств необхідно враховувати «вагу»

кожного з показників та ступінь їх впливу на глобальний критерій. Однак це питання є напрямом до окремого дослідження, яке повинно проводитися в конкретних економічних умовах, що склалися. В даному випадку обмежимося припущенням, що всі показники повинні бути стандартизованими.

Рівень конкурентоспроможності підприємства за  $i$ -м показником враховують відносно деякого ідеального значення цього показника по промисловості [3]:

$$X_i = x_i / C_i, \quad (1)$$

де  $X_i$  – відносний  $i$ -й показник конкурентоспроможності підприємства;  
 $C_i$  – ідеальний  $i$ -й показник конкурентоспроможності підприємства.

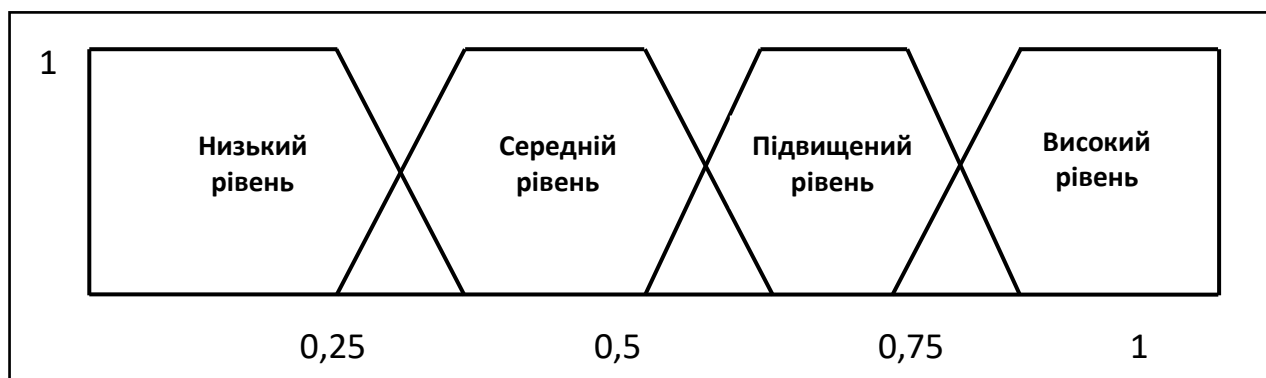
Виходячи з визначення  $X_i$ , його величина завжди задовольняє нерівності  $X_i < 1$ . На підставі цього визначення сформуємо універсальні нечіткі рівні класифікації значень кожного показника (табл.2).

Таблиця 2

**Нечіткі рівні класифікації значень показників**

Рівні класифікації	Найменування підмножини
$0.7 < b_1 < 1$	Високий рівень показника
$0.5 < b_2 < 0.75$	Підвищений рівень показника
$0.2 < b_3 < 0.6$	Середній рівень показника
$b_4 < 0.35$	Низький рівень показника

Такі нечіткі рівні класифікації значень показників задають тільки інтервали для визначення їх відповідних станів. Функції приналежності  $\mu_i(x)$  кожному з цих станів можуть мати різноманітний характер, наприклад трикутними, трапецевідними або біль складні. Виходячи зі значень обраних нечітких рівнів класифікації показників визначимо як трапецевідні (рис.2).



**Рис. 2. Функції приналежності показників нечітким рівням класифікації їх станів**

Визначимо нечіткості поточних значень кожного показника. Для цього необхідно виконати аналіз кожного з показників та ступінь їх можливої зміни в залежності від суб'єктивних факторів. Однак це питання є напрямом до окремого дослідження, яке повинно проводитися в конкретних економічних умовах, що склалися. В даному випадку обмежимося припущенням, що всі показники  $x_i$  мають стандартні інтервали.

Стан показника можна характеризувати належністю відповідному нечіткій множині. Згідно [4] аналізуючи досліджуваний показник, можна зробити висновок про ймовірність його приналежності якої-небудь нечіткій множині. При цьому, якщо показник (змінна)  $X$  визначений на інтервалі  $[x_0, x_k]$  і кожне з можливих чітких його значень може належати двом станам  $M_1$  і  $M_2$  з функціями належності  $\mu_1(x)$  і  $\mu_2(x)$  відповідно. В загальному випадку, якщо дійсний значення показника визначається не інтервалом, а функцією приналежності  $\mu(x)$ , середні ймовірності приналежності до можливих станів показника визначаються за формулами [6]:

$$P_{M_1}(x \in [x_0, x_k]) = \frac{1}{x_0 - x_k} \int_{x_0}^{x_k} \mu_1(x) dx;$$

$$P_{M_2}(x \in [x_0, x_k]) = \frac{1}{x_0 - x_k} \int_{x_0}^{x_k} \mu_2(x) dx.$$

(2)

де  $P_{M_1}, P_{M_2}$  – чіткі ймовірності приналежності показника  $x = x_i$  нечітким множинам  $M_1$  і  $M_2$ ;

$\mu(x)$  – функція приналежності показника можливим поточним значенням;

$x_0$  – мінімальне значення показника на інтервалі  $[x_0, x_k]$ ;

$x_k$  – максимальне значення показника на інтервалі  $[x_0, x_k]$ ;

Оцінювати конкурентоспроможність підприємства виходячи з формальних нечітких знань та імовірнісних характеристик (2) на основі лінгвістичних змінних можна двома способами [4].

При використанні першого способу спочатку визначається безліч атомарних висловлювань для кожного значення лінгвістичної змінної заданого  $i$ -го показника за величинами ймовірності приналежності того чи іншого стану. Наприклад, якщо лінгвістична змінна має три можливих стани ( $N = 3$ ), та визначено ймовірності приналежності  $i$ -го показника цим значенням  $P_1 = 0.1, P_2 = 0.45, P_3 = 0.45$ , то справедливий вислів, що критерій **ЙМОВІРНО НЕ НАЛЕЖИТЬ** до першого з них, а також **ЙМОВІРНО НАЛЕЖИТЬ** до другого і третього стану. Після визначення атомарних

висловлювань формуються формальні знання, які дозволяють приймати рішення про загальний стан системи.

Другий спосіб оцінки конкурентоспроможності підприємства на основі імовірнісного нечітко-множинного підходу полягає в першу чергу формуванні формальних знань. Після чого визначається ймовірність виникнення кожного з них.

Обчислення характеристик атомарних знань базується на ймовірності приналежності кожного параметра значенням відповідного нечіткої множини і виконується по формулі (2). Використовуючи аксіоматику незалежних подій теорії ймовірностей отримаємо ймовірність і-го атомарного знання [4]:

$$P_i = \sum_{k=1}^K \prod_{j=1}^N p(I_j = \Omega_r), i = 1 \dots Z, \quad (3)$$

де  $K$  – кількість сум в вираженні, яке визначається логікою формування формальних знань;  $I_j$  –  $j$ -й показник (критерій);  $\Omega_r$  –  $r$ -та нечітка множина  $r=1 \dots z_r$ .

Таким чином, для кожного з можливих станів конкурентоздатності підприємства, які характеризуються відповідними нечіткими підмножинами, визначається імовірнісний показник. При цьому необхідно визначити кількість можливих нечітких підмножин  $Z$  і правила обчислення ймовірностей приналежності конкурентоздатності підприємства цим підмножини  $P_i$  відповідно до виразу (3).

Розглянемо процес визначення конкурентоспроможності підприємств на прикладі наступних виробництв: ПАТ «Роменська гардинно-тюлева фабрика»; ВАТ «Тернопільська швейна фабрика «Галія»»; ВАТ «Херсонський бавовняний комбінат». Відносні показники конкурентоспроможності цих підприємств наведені в табл..3.

Для визначення ймовірностей приналежності показників нечітким множинам необхідно встановити функції приналежності кожного показника можливим поточним значенням  $\mu(x)$ . Формування цих функцій при повній невизначеності статистичних даних можливо з допомогою інтервалів [5]. При цьому, інтервали можуть бути як чіткі, так і нечіткі. В даному прикладі будемо вважати, що поточні значення показників існують в чітких інтервалах  $[x_i-0,05; x_i+0,05]$  та рівномірно розподілені на цих інтервалах  $\mu(x_i) = 1$ . Таке припущення спрощує розрахунки та не оказує суттєвий вплив на використання методики оцінювання конкурентоспроможності підприємства.

**Таблиця 3**

**Відносні показники конкурентоспроможності підприємств України  
в 2012-2016 рр**

Підприємства		Нормалізовані значення параметрів конкурентоспроможності						
		X1	X2	X3	X4	X5	X6	X7
ПАТ «Роменська гардинно-тюлева фабрика»	2012	0,55	0,40	0,80	0,02	0	0,35	0,09
	2013	0,48	0,30	0,70	0,01	0	0,40	0,09
	2014	0,59	0,27	0,57	0,01	0,1	0	0,15
	2015	0,76	0,37	0,79	0,01	0,1	0	0,23
	2016	0,29	0,26	0,76	0,01	0	0	0,12
ВАТ «Тернопільська швейна фабрика «Галія»»	2012	0,27	0,27	0,70	0,02	0	0,27	0,08
	2013	0,24	0,21	0,60	0,02	0	0,26	0,08
	2014	0,26	0,19	0,52	0,02	0,4	0	0,12
	2015	0,36	0,26	0,66	0,01	0,2	0	0,20
	2016	0,18	0,18	0,52	0,02	0,3	0	0,11
ВАТ «Херсонський бавовняний комбінат»	2012	0,19	0,18	0,03	0,88	0	0,16	0,06
	2013	0,17	0,14	0,03	0,87	0	0,18	0,06
	2014	0,19	0,12	0,02	0,04	0,4	0,00	0,07
	2015	0,30	0,18	0,03	0,01	0,2	0,00	0,12
	2016	0,14	0,12	0,06	0,01	0,5	0,00	0,07

Відповідно формулі (2) та на основі даних табл.3 визначимо ймовірності приналежності показників конкурентоспроможності підприємств. чотирьом можливим станам ( $b_1$ – $b_4$ ). Результати обчислення таких ймовірностей для ПАТ «Роменська гардинно-тюлева фабрика» в 2012-2016р. наведені в табл.4, 5.

**Таблиця 4**

**Ймовірності приналежності показників конкурентоспроможності  
ПАТ «Роменська гардинно-тюлева фабрика» в 2012-2014 рр**

Показ- ник	2012				2013				2014			
	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>
X <sub>1</sub>	0	0,5	0,5	0	0	0,47	0,53	0	0	0,84	0,16	0
X <sub>2</sub>	0	0	1	0	0	0	0,15	0,85	0	0	0,13	0,8
X <sub>3</sub>	0,85	0,15	0	0	0,5	0,5	0	0	0	0,77	0,23	0
X <sub>4</sub>	0	0	0	1	0	0	0	1	0	0	0	1
X <sub>5</sub>	0	0	0	1	0	0	0	1	0	0	0	1
X <sub>6</sub>	0	0	0,85	0,15	0	0	1	0	0	0	0	1
X <sub>7</sub>	0	0	0,85	0,15	0	0	1	0	0	0	0	1

Таблиця 5

**Ймовірності приналежності показників конкурентоспроможності ПАТ «Роменська гардинно-тюлева фабрика» в 2015-2016 рр**

Показ- ник	2015				2016			
	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>
X <sub>1</sub>	0,53	0,47	0	0	0	0	0,47	0,53
X <sub>2</sub>	0	0	0,87	0,13	0	0	0,09	0,91
X <sub>3</sub>	0,83	0,17	0	0	0,53	0,47	0	0
X <sub>4</sub>	0	0	0	1	0	0	0	1
X <sub>5</sub>	0	0	0	1	0	0	0	1
X <sub>6</sub>	0	0	0	1	0	0	1	0
X <sub>7</sub>	0	0	0,09	0,91	0	0	0	1

Встановимо лінгвистичний інтегральний показник (рівень конкурентоспроможності підприємства), який має три нечіткі підмножини: висока, задовільна та незадовільна конкурентоспроможність. Результати обчислення ймовірності приналежності інтегрального показника нечітким підмножинам відповідно виразу (3) для трьох підприємств наведені в табл.6.

Таблиця 6

**Ймовірності приналежності інтегрального показника нечітким підмножинам**

Рівень конкурентоспроможності	ПАТ «Роменська гардинно-тюлева фабрика»				
	2012	2013	2014	2015	2016
Високий	0	0	0	0	0
Задовільний	0,244	0,212	0,154	0,136	0,126
Незадовільний	0,756	0,788	0,846	0,864	0,874
Рівень конкурентоспроможності	ВАТ «Тернопільська швейна фабрика «Галія»»				
	2012	2013	2014	2015	2016
Високий	0	0	0	0	0
Задовільний	0,156	0,141	0,123	0,13	0,1
Незадовільний	0,844	0,859	0,877	0,87	0,9
Рівень конкурентоспроможності	ВАТ «Херсонський бавовняний комбінат»				
	2012	2013	2014	2015	2016
Високий	0	0	0	0	0
Задовільний	0,114	0,102	0,084	0,103	0,082
Незадовільний	0,886	0,898	0,916	0,897	0,918

За період, що аналізується показники конкурентоспроможності підприємств мали значні коливання. Порівняно з показниками 2012 р., показники конкурентоспроможності всіх підприємств в 2016 р. мають тенденцію до зниження. Конкурентоспроможність ВАТ «Херсонський бавовняний комбінат», порівняно з показниками конкурентів, має поступову тенденцію до зниження, но воно не таке значне, як у конкурентів в абсолютному (на 0,032) та відносному значенні (0,035).

Розрахунок рівня конкурентоспроможності підприємств показав, що на ринку виробництва тканини України є один лідер. Так, починаючи з 2012 року ПАТ «Роменська гардинно-тюлева фабрика» має стабільно найкращий показник серед інших підприємств. Це підприємство спеціалізується на випуску найякісніших тканин та має найбільший обсяг реалізації і чисельність працюючих. Як результат, воно утримує найбільшу частку вітчизняного ринку гардинно-тюлевих виробів.

ВАТ «Тернопільська швейна фабрика «Галія»» поступається лідеру за показниками конкурентоспроможності продукції та обсягами реалізації, але показник прибутку має вище значення за обраний період. При аналізі значущості характеристик видно, що з 2012 року для підприємства в цілому помітно знижується вплив конкурентоспроможності продукції на конкурентоспроможність підприємства, але зростає вплив обсягу реалізації.

Таким чином, оцінивши конкурентоспроможність наведених підприємств, можна дійти висновку, що загальна тенденція їх участі на ринку носить спадний характер, що зумовлює зниження частки ринку, яку утримує кожне підприємство.

Наведений приклад дає деяку оцінку на основі спрощених припущень при визначенні функцій приналежності показників конкурентоспроможності підприємства відповідним станам. Ці функції можуть бути визначені шляхом статистичних досліджень, методом експертних оцінок або адміністративними засобами. Для отримання більш об'єктивних результатів в умовах невизначеності необхідно задати нечіткий характер функціям приналежності  $\mu_i(x)$  з допомогою інтервального підходу [5]. В цьому випадку ймовірності приналежності високому, задовільному або незадовільному рівню конкурентоспроможності підприємства також будуть мати нечітких характер.



## Висновки

Запропонована методика оцінки конкурентоспроможності підприємств на основі нечітко-множинного інтервального аналізу показників різної природи дозволяє уникнути неоднозначності їх впливу на результат оцінювання. Методика не вимагає визначення вагових значень кожного з показників конкурентоспроможності підприємств, при цьому глобальний критерій визначається на основі правил формування формальних знань, які, в свою чергу можуть носити нечіткий характер. Результати порівняльної оцінки стійкі до такого роду невизначеності, а нечітко-множинна модель дозволяє враховувати вплив як внутрішніх, так і зовнішніх факторів на конкурентоспроможність підприємств.

Розроблена методика оцінки конкурентоспроможності підприємств дає можливість виконати не тільки порівняльний аналіз ефективності ведення бізнесу, збільшення/зменшення впливу досліджуваного виробництва на ринок в сфері своєї діяльності, але зробити висновок про стан відповідного ринку в країні.

## ЛІТЕРАТУРА

1. Сахно І. В. Аналіз основних методичних підходів до оцінки конкурентоспроможності підприємств / І. В. Сахно // Збірник наукових праць Таврійського державного агротехнологічного університету. Економічні науки. – 2012. – № 2. – Ч. 2. – С. 385-390.
2. Левицька А. О. Методи оцінки конкурентоспроможності підприємства: вітчизняні та закордонні підходи до класифікації / А. О. Левицька // Механізм регулювання економіки. – 2013. – № 4. – С. 155-163.
3. Литвинова В.А. Повышение корректности сравнительных оценок конкурентоспособности продукции на основе таксономического метода / В.А. Литвинова // Інноваційна економіка. – 2012. – № 5 (31). – С. 97 – 101.
4. Лосєв М.Ю. Нечітко-множинна оцінка стану параметрів техніко-економічних систем / М.Ю. Лосєв, Ю.М. Малишко // Системи обробки інформації: збірник наукових праць. – Х.: ХУПС, 2015. – Вип. 4(129). – С. 33-38.
5. Дилигенский Н.В. Нечеткое моделирование и многокритериальная оптимизация производственных систем в условиях неопределенности: технология, экономика, экология / Н.В. Дилигенский, Л.Г. Дымова, П.В. Севастьянов. – М.: «Издательство Машиностроение – 1», 2004. – 397 с.

**DEVELOPMENT OF A MULTIMEDIA INTERACTIVE ELECTRONIC  
EDUCATIONAL EDITION USING THE MODULES OF VARIOUS  
FORMAT**

**Introduction and problem statement**

The transition to paperless technologies in various fields of science and education has necessitated the development of appropriate tools and environments. Modern information technologies make it possible to achieve a new quality of education through the use of audio and video tools created in the educational process using modern methods and technologies for processing audio and video.

Creating a logical and comprehensive system of thematic interactive electronic learning tools using multimedia visual and audit tools (sound, multi-image and video, etc.). Highly-qualified university teachers will promote both the quality of students' knowledge of the students and the orientation of teachers to modern methods and teaching methods.

Studies show that the educational material presented in the form of a multimedia publication is learned much better than the same material taken by the teacher. The matter is here, on the one hand, in the previous thorough preparation of materials, and on the other - in clarity, since the use of such a form of education allows for a more accurate, effective, to present the subject matter in a fairly short time [3].

Electronic textbooks created by university lecturers and posted on the Simon Kuznets Kharkiv National University of Economics personal education system website can be used by students in the classroom for self-study and at home for distance learning. The use of multimedia manuals is especially relevant when explaining new material, when it is very difficult to confine to traditional means of knowledge transfer.

Multimedia tutorials are computer training programs and video files designed to organize the learning process on a student's computer.

Educational and methodological developments are implemented using multimedia and hypertext technologies with a large number of interactive simulators, which provide the creation of tasks, step-by-step control of their implementation, the detection of errors in solving problems and virtual laboratory work. All multimedia manuals are delivered on a data carrier and verified in a real learning process [2, 3].

To date, there are a number of software packages for creating audio and video of various types. The relevance of the use of such benefits is constantly growing, due to the action of several reasons:

when studying a material in a video-audio format, a student can set up a video view, study the video material from the required place, thereby dwelling on details that are not clear to him;

the effectiveness of audio and video recordings is increasing due to the fact that the user perceives the material provided to him by two senses – vision and hearing;

the study of video materials will allow the user to see the theoretical and practical part of the lesson, thereby understanding and consolidating the demonstrated material;

training can be done at any time and place convenient for student.

The disadvantage of audio-video recording is that when studying a complex material an audio-video course user may have additional questions to which he will not be able to find the answer in the presented materials [1].

In IT-targeting disciplines, the most common are classes such as lectures and laboratory work. Moreover, the second type has a more significant impact on the formation of competences in the field of computer science. Therefore, when developing the training material, it is necessary to ensure the various stages of mastering the discipline: gaining knowledge, acquiring skills and controlling the level of academic achievement. The listed stages are realized by various didactic means. Attempts to submit them to one widely used computer application failed to find a successful solution. The paper discusses the approach to the creation of multimedia interactive e-learning manual is considered, which combines the advantages of conventional computer tools in the areas of their optimal use.

### **Subject area**

Consider the development and use of multimedia interactive electronic edition using modules of various formats on the example of the training manual [1]. This tutorial is intended primarily for students of the branch of knowledge 12 “Information technologies” and is focused on the study and consolidation of theoretical material in disciplines related to data storage and processing.

This tutorial is designed to develop students' competencies that are provided by obtaining conceptual modern knowledge and skills to solve complex unforeseen tasks, creating students' ability to transfer their own experience to specialists, as well as taking responsibility for making decisions in various settings.

The tutorial is written in accordance with the program of the discipline "Technology Database" and covers the following topics:

ADO.NET architecture. Connection to databases.

Perform operations in a connected environment.

Perform operations in a disconnected environment.

Implementing transactions using ADO.NET.

Mastering these classic topics allows students to acquire skills within the following competencies: the ability to use connected and disconnected environment modes to access data in business applications.

Learning outcomes:

Possess the principles of building the architecture of modern business applications.

Apply the basic concepts of organizing connections to data stores in software products.

Choose the most appropriate data access environment.

Use the core components of ADO.NET and JDBC in projects.

Build algorithms for creating a connection with a data source in software projects.

Design connection management tools with data storage in software products.

Organize offline work with the database.

Create programs to perform data operations in a connected and disconnected environment.

Bind data to user interface in software products.

### **Development**

To analyze tools for implementing a multimedia interactive e-learning tutorial, such programs are used to create multimedia files such as Microsoft Office Word, Camtasia Studio and Adobe Captivate.

The simplest and most accessible instrumentation for general-purpose teacher is Word, included in the Microsoft Office software package. MS Word is a specialized automation tool for creating and developing lecture texts and instructions for performing laboratory work. The program provides for the development of electronic documents that provide the ability to include complex multimedia content and special playback capabilities. With this program, even a novice can easily create didactic materials at a fairly high professional level. The interface is understandable at the level of intuition and allows you to use graphics and use formatting functions

At the current moment, Microsoft has prepared and released several major versions of the program: Microsoft Word 2003, 2007, 2010, 2013, and 2016. The differences between the versions, of course, are that they have some improvements and additions the need for which became apparent over time and causes the need for further elaboration.

In our case, the main advantages of Microsoft Word are:

- a step-by-step method for creating educational materials;

- a basic set of tools for working with graphics, color correction and formatting, the presence of artistic effects;

- the ability to apply complex photo effects without the use of specialized software (the ability to change the color saturation, adjust the contrast and brightness of the picture, and use artistic filters such as blur, paintbrush and watercolor);

- simplifying the addition of video materials to the text, the ability to edit them using a wide range of high-quality special effects;

- the ability to navigate content using hyperlinks and the special Navigation panel;

- the implementation of the Backstage submission mode, which replaces the traditional menu, allowing you to save the ready-made material, publish them on the Internet or print them, as well as provide access to the materials without additional efforts;

- the ability to save the document in pdf format.

The latter circumstance is essential for the dissemination of didactic materials in the educational process.

Significant functionality for creating high-quality videos is the Camtasia Studio program, created by TechSmith [4]. This software is designed to capture video from the screen. This circumstance is extremely important in the visual explaining the techniques of creating software applications when performing laboratory work on discipline.

The program has a convenient control panel with which you can record. Camtasia Record, included in Camtasia Studio, can record not only from the screen or screen sections, but also from a webcam and audio devices (microphone). Before recording a video, a countdown window (3 seconds) is displayed, during which the user can tune in to video recording. During video recording, the control panel is reduced to the tray panel. The user can control the recording using hotkeys: Record / Pause - F9; Stop - F10; Marker - Ctrl + M; Highlight Ctrl + Shift + D on the screen.

The Camtasia Studio package has a video player and its own editor with which you can edit both the entire video and individual frames. The Camtasia Studio editor supports the following formats:

- 1) audio formats: wav, mp3, wma;
- 2) video formats: avi, mp4, mpeg, wmv, mov, swf, camrec (Camtasia's own format)
- 3) image formats: bmp, gif, jpg, png.

After recording the video, it reappears in the Preview browser, where you can view the video, transfer it to the editor, delete it. The Camtasia program editor allows the user to import the above formats and merge them into one movie.

The Camtasia editor interface resembles the standard Windows interface in Movie Maker. The editor has a huge amount of functionality: frame transition effects, markers, panoramic scaling, creating footnotes, signatures, creating polls and polls.

In order to create high-quality video materials, it is advisable to use the emphasis tools that are in the Camtasia editor:

1) marker - used when it is necessary to indicate an important part in the Marker video, such properties as appearance / fading, styles, size, appearance, possibility to choose its type (16 types) are characteristic.

2) signature - with its help you can sign conceptual ideas that are depicted on the video in a certain period of time, there is the possibility of text overlay on the video.

3) panoramic scaling - used when you need to pay attention only to a specific area in the video. This effect reduces the scope of the video, highlighting only the area on which you want to focus.

Camtasia Studio software contains Menu Maker, which is designed to create menus. The program creates links to the files specified by it (audio, video, images, programs, documents). When you save the existing menu, the program creates a folder in which the startup menu file (exe), autoload file (autorun.inf) and media folder in which all files that are used in the menu are located.

Camtasia Studio has an intuitive interface that allows the user to easily control the functionality of this package.

The main key features and advantages of Camtasia Studio are:

- 1) a clear record of what is happening on the screen, including the recording of a webcam, sound, system sounds, etc.;
- 2) countdown time to record;
- 3) convenient recording control panel (on the display panel)

- 4) a recording from a PowerPoint presentation;
- 5) built-in video-audio editor (supports video effects)
- 6) built-in focus;
- 7) built-in program for creating menus for video discs;
- 8) save the finished video in AVI, SWF, FLV, MOV, WMV, RM, GIF and CAMV formats.

The main disadvantages of Camtasia Studio are:

- 1) shareware license;
- 2) use of codecs when recording video;
- 3) significant use of processor resources.

Thus, creating a presentation in this program will allow you to convey the essence of the subject as fully as possible, since Camasia Studio has all the necessary tools.

Adobe Captivate (formerly known as RoboDemo) is a powerful modern tool for creating demo video presentations. Adobe Captivate is an e-learning program for Microsoft Windows that can be used to demonstrate video recording software, create a simulation program, create instructional presentations, and various tests in the .swf format [5].

The main benefit of Adobe Captivate is that users do not need programming skills to develop learning resources. Thanks to a simple and intuitive interface, mastering the program is quite simple.

The main advantages of the program are:

- 1) support for recording from multiple displays;
- 2) time saving function (calculation of the shortest path from event to event)
- 3) the presence of a powerful editor for audio / video processing;
- 4) the ability to create multi-level and branched e-learning courses;
- 5) the ability to convert slides to swf;
- 6) support for additional controls (buttons for text fields, lists, etc.);
- 7) support for Mp3, WAV, AVI, SWF, FLA and FLV formats;
- 8) a built-in program for creating and styling a video menu;
- 9) the possibility of introducing a test control. The following types of questions are supported: multiple choice; well no; fill in the gaps in the text; short answer; conformity; select region; sequence; ranking.

The disadvantages of Adobe Captivate are:

- 1) support only one SWF video format;
- 2) significant CPU load;
- 3) lack of support for recording from a webcam;

- 4) paid program;
- 5) lack of support for the Ukrainian language;
- 6) work on supporting additional audio and video codecs [8].

To organize the performance of laboratory work, an important condition is the possibility of simultaneously placing on the monitor screen instructions for performing and the software environment in which the student works (in our case, Visual Studio). Adobe Captivate delivers content only as fixed-size slides. This circumstance is unacceptable for instructions for performing laboratory work, since it significantly increases the psychological burden on the student, requiring that the content be remembered when the monitor is switched from instruction to the environment. And vice versa, this circumstance is an advantage for the stage of control of the level of academic achievements.

Didactic materials, which are prepared by means of Adobe Captivate, are presented in the form of projects. Tests can be included as an integral part of an e-learning project, or developed as a separate project. The authors used both types of tests. When preparing the electronic textbook on the “Classical means of data access” module, tests were included in its component. They were placed at the end of each section. Such a complete complex of didactic materials allowed to fully ensure the independent study of the module. The test as a separate project was applied to the study of the module, which allowed us to conduct classroom testing after the completion of each laboratory work.

You can work with Captivate tests in two modes: designer and testing. A test consisting of a number of tasks (designer mode) is added to the project. Here, in the Adobe Captivate environment, you can check its work. Formally, you can check the performance of such structural units of the entire project; from the current slide to the end of the project and the next N slides. It does not matter if the test slides or not. The project obtained in the Captivate environment can be exported to one of the following formats: SWF (HTML + SWF), HTML5, PDF (only one page with embedded flash), FLV, EXE, MP4 and APP (for MAC). So you can build tests for various devices (computer, tablet, smartphone) and various operating systems (Windows, Android and IOS).

In Captivate, two types of test tasks are implemented: open and closed. Open tasks include the following:

Fill-In-The-Blank – instead of a blank, a specific word should be entered in the text;

Short Answer – to answer you need to enter a word or phrase;



Rating Scale (Likert) – you need to specify the degree of your agreement with the above phrases.

Closed tasks are represented by the following types:

Multiple Choice – you need to choose one or several correct answers from the proposed ones;

True / False – you must answer "Yes" or "No" to the test questions;

Matching – it is necessary to establish the correspondence between the semantic units given in the question window and the answer window;

Hot Spot – should correctly identify areas on the slide by clicking them;

Sequence - the answer must determine the correct order of objects, actions, and the like.

Thus, in Captivate you can use all possible types of test tasks, and the most popular types of tasks with one and several correct answers are presented as multiple choice (the difference between them is marked as an option).

For individual tasks, you can set a response time limit. This is important for independent tests conducted in the classroom for the implementation of intermediate and final control of students' knowledge.

Weight can be set for each task. For this, the task is labeled Graded or Survey. In the first case, its weight (importance) can be noted, in the second - not. The test can include tasks with or without weight. For the result, the calculation of the correct and incorrect answers and the sum of the accumulated weights and the possible maximum amount are performed separately. Moreover, for the wrong answer to a specific question, you can assume a penalty that is subtracted from the points scored. All this allows you to determine the knowledge of the student with maximum accuracy.

For a test session, tasks are randomly selected. The selection results can be:  
all tasks;

N-tasks that are randomly selected from the so-called task pool;  
combination of previous options.

For tests embedded in an electronic workbook, the first option was used when the student himself was interested in an objective test of his knowledge. For testing in the audience is used the second option, which is deprived of the possibility of cheating. At the same time, the power of the task pool significantly exceeded the number of session tasks.

The test result is displayed on the screen. However, in the test settings there is an option Enable reporting for this project, the inclusion of which allows you to export or save the report in one of the main international standards of distance

learning systems. This feature allows you to automate the analysis of educational achievements as an individual student, and the entire group.

### **Using**

The training material from the textbook should be in the following order: first read the material presented in the section, then try to answer the questions and complete the tasks set at the end of the section. After getting acquainted with the section of theoretical material, you need to perform laboratory work. During work should pay attention to the choice of tasks. They must be doable for students to understand. To do this, each student chooses the level of complexity from the following:


- 1) frontal;
- 2) individual;
- 3) competence [3].

When choosing a front-level student performs the tasks of the basic level, which are described in detail in the instructions. After its completion, the student receives 60 points on the one hundred rating system.

To test his strength and improve his grades, a student can independently complete several other tasks, some of which are of a reproductive nature, and others of a creative type. For their correct decision up to ten points is added. A student can add 10 points to the sum of points obtained if he independently proposes and solves the initial problem on the topic being studied. This assignment must be from the subject area of study or the student's future profession. The total score for this level does not exceed 80 points.

In the case of choosing an individual level, the student gets acquainted with the instructions for completing the basic level task and solves a similar task from the subject area defined by the individual task. When performing such an individual task, the student receives 70 points. He can get another 10 points if he adapts to the problem domain from the chosen option presented in the instruction and solves them. As at the frontal level, a student can add 12 more to the sum of points obtained if he formulates and solves the initial problem on the topic being studied. The total score for this level does not exceed 92 points.


At the level of competence, the student demonstrates the ability to independently set and solve problems on the topic being studied from the subject area of study or future profession. First, he formulates and solves a problem similar to the base (74 points), then - similarly to additional tasks (11 more points) and at the end of the original problem (up to 9 points). The total score at this level can reach one hundred points.

When performing laboratory work, you can use either a step-by-step description of each task, or the corresponding video. The link to it is presented as an icon  at the beginning of the "Run" section of each task.


Each section contains tests. They can be used in two ways - for self-monitoring and for determining educational achievements in the field of laboratory protection.

If tests are used for self-examination, the student can pass them both before the start of the laboratory work, in order to assess the level of his theoretical knowledge on the topic being studied, and after they are completed, to consolidate the skills acquired during laboratory studies. Such a double self-test is recommended to determine the growth of professional skills and improve student self-esteem.

After completing the laboratory work, the student has the opportunity to take a test in the classroom, protecting the laboratory work. The results obtained by him are summarized in determining the final grade for laboratory work - up to ten points in the one hundred point rating system.

Such multiple testing is protected against fraud, since the test for each topic contains about one hundred tasks, and for each test session, 10 tasks are randomly selected from the total set. To go to the test, you need to click on the icon , which is located in the subsection "Test" before the conclusions for each section.

At the end of each section there is a "Questions and Tasks" heading. Relevant questions are formulated that allow you to test and consolidate the knowledge gained in the reading section. The last is an interactive exercise. It is presented in the form of a test task with three attempts to answer. Performing such an exercise is considered as a means of operational self-monitoring of the material under study.

To go to the interactive exercise, click the icon ,

The tutorial can be used in two modes - without an Internet connection (offline) and with connection to it (online mode). Next, you should consider the features of working with a textbook in each of these modes.

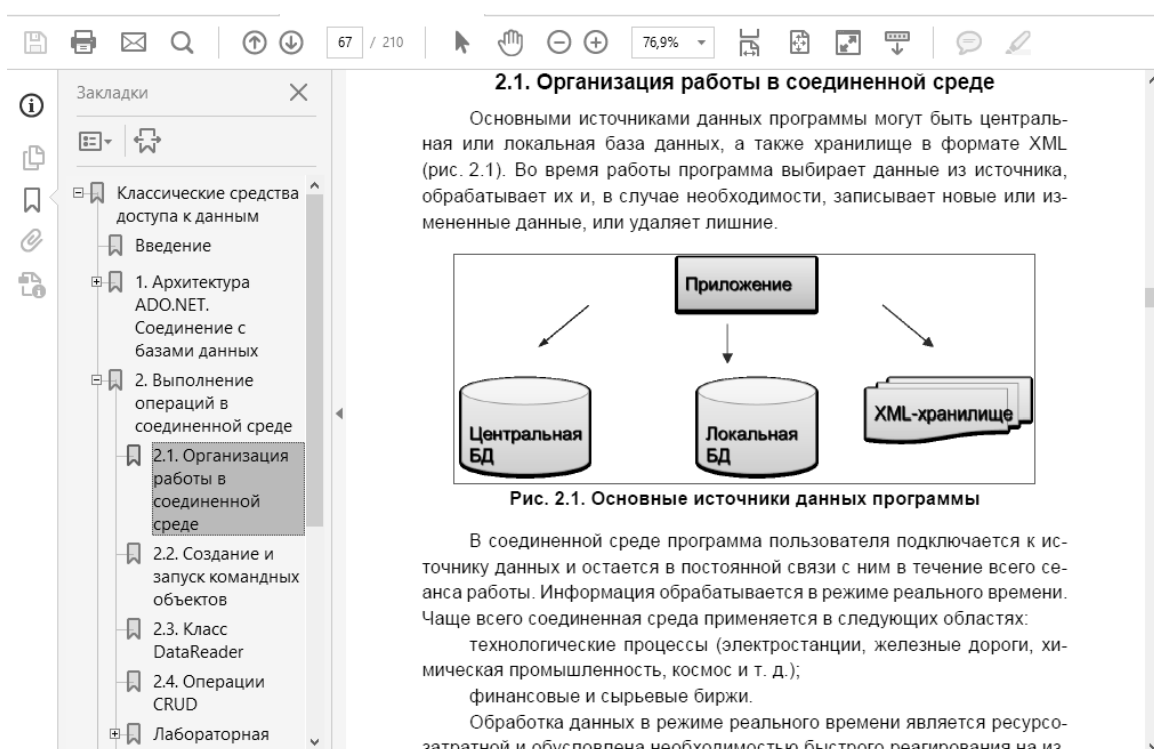
In offline mode, study materials are best viewed using the free Adobe Acrobat Reader.

For the convenience of navigating through the material in the tutorial, bookmarks, hyperlinks, page numerator are provided. In the Adobe Acrobat Reader program, the Bookmarks button is located in the navigation area of the open file window of this program (Fig. 1).



**Fig. 1. Bookmarks button in the program navigation area  
Adobe Acrobat Reader**

The first click on the “Bookmarks” button in the navigation area displays the root node of the tutorial, which can be opened to the desired depth to display the name of the required theoretical section or element describing laboratory work (Fig. 2).



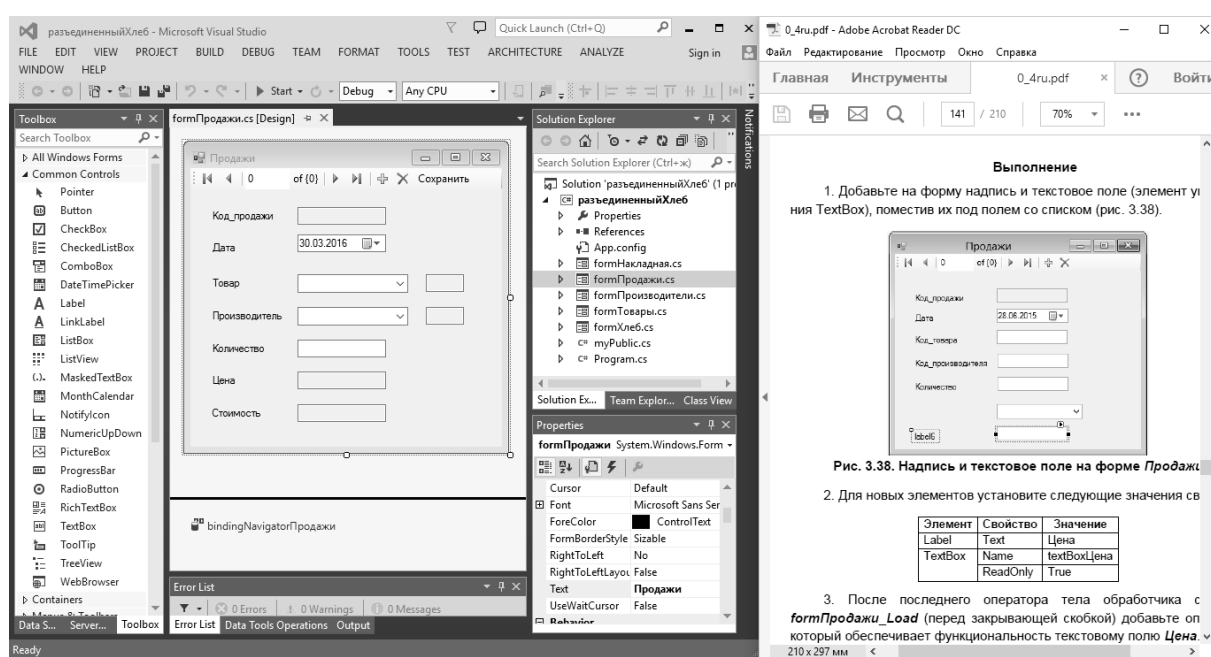
**Fig. 2. Navigate to the desired section using the Bookmarks area**

The tutorial widely uses hyperlinks to go to reference sites on the Internet or to a specific place in the tutorial. For example, <http://www.connectionstrings.com/> is used to get the format of a connection string with various databases. Local hyperlink is used, for example, when describing a task in the text

Write the code of Example 1.5 using the using block.

In the case of a local link, you need to worry about returning to the place where the transition occurred. To do this, you can either remember the source page number, and then use the Adobe Acrobat Reader window pager, or execute the command to switch to the previous view using the Alt + Left arrow key combination.

During the execution of the laboratory work, it is important to create a comfortable environment so that you can simultaneously see on the monitor a step-by-step description and Visual Studio. Usually, the step-by-step description is placed on the right side of the monitor, taking about a third of it (65% scale), Visual Studio is on the left of the remaining area (Fig. 3).



**Fig. 3. Placement of the step-by-step description and Visual Studio**

In the instructions for the implementation of laboratory work there are texts of program fragments. They can be copied to the clipboard and then pasted into the program editor in Visual Studio. This technique is recommended to use when the code is absolutely clear. If you need to understand this, it is better to enter the text of the program yourself. When you enter, as a rule, the algorithm presented in the code becomes clear.

To work with tests and interactive exercises you need a program for viewing swf-files. In order not to install an additional program, you can use a web browser that supports this feature. Please note that one of the most modern Edge browsers does not support it. For these purposes, it is recommended to use the Google Chrome web browser. The following describes how to configure this program in the Windows 10 operating system.

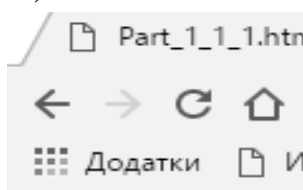
To set up Google Chrome to work with tests and interactive exercises do the following:

1. Click the “Settings” button on the Start panel.
2. Click the “System” button in the “Windows Settings” window.
3. Select “Default applications” in the left list of the “Settings” window, and on the right select standard applications for file types.
4. Find the .htm element in the left list of the "Select standard applications for file types" window, click the item on the right and set the Google Chrome application icon.

After completing the above setup, tests and interactive exercises will open in the Google Chrome web browser. At the end of the answers to your questions, close the Google Chrome program window to return to working with the tutorial in the Adobe Acrobat Reader window.

Working with the online tutorial has the following differences.

1. It is recommended to install Google Chrome as your default web browser.
2. When viewing materials online (through a browser), the Bookmarks area is not available. To navigate through the sections of the tutorial, use the "Content" section. When you click on the name of the desired section, it goes to the beginning of the selected section. To quickly navigate to the "Content" section, use the "To Content" hyperlink located at the top of each page (except for the cover page).
3. Tests and interactive exercises open in the same web browser window as the main material. Therefore, after completing the answers to the questions asked, press the “Back” button in the web browser to return to working with the main material of the tutorial tutorial (Fig. 4).



**Fig. 4. “Back” button in the web browser window**

The methodological techniques described above for setting up an environment for performing laboratory work can significantly improve a student’s productivity for acquiring computer skills.

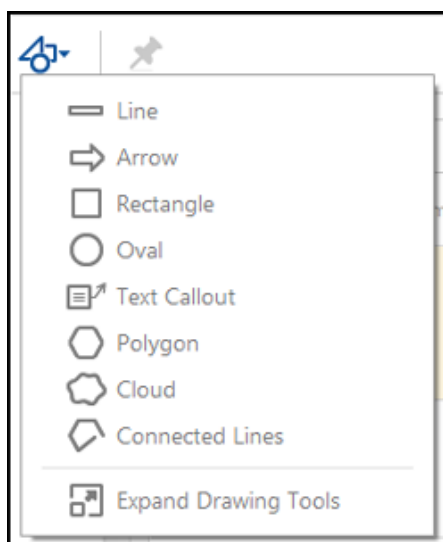
In addition to good navigation, working with pdf files offline has another advantage. Free Adobe Acrobat Reader DC allows you to add comments to a pdf document. Comments are used to highlight interesting or incomprehensible text fragments, text annotations, and an arbitrary picture. All means of creating

comments are in the Comments toolbar (Fig. 5). It is invoked by clicking the same button on the common toolbar.



**Fig. 5. Comment toolbar**

If you want to draw any shape, click the “Drawing Tools” button on the toolbar “Comment” and select the desired tool from the drop-down menu “Drawing Tools” (Fig. 6).



**Fig. 6. Drawing Tools menu**

Added to the text of the electronic tutorial can be saved and used when re-opening the tutorial. Thus, it is possible to organize work with the electronic textbook in the style of studying the material on the paper textbook.

## **Conclusions**

The article discusses the possibilities of authoring tools for creating a multimedia interactive electronic textbook in the field of informatics using modules of various formats. Such tools as the implementation of a multimedia interactive electronic tutorial, programs such as Microsoft Office Word, Camtasia Studio and Adobe Captivate are analyzed, and the most rational spheres of their use are identified. In particular, it was determined that Microsoft Office Word is best suited for creating texts of theoretical material and instructions for performing laboratory work, Camtasia Studio for creating video demonstrating steps for performing laboratory tasks in Visual Studio, and Adobe Captivate for creating tests that a didactic tool to support the stage of monitoring educational attainment.

The study leads to the following conclusions:

1. Many programs, consisting of Microsoft Office Word, Camtasia Studio and Adobe Captivate, are sufficient as a tool for creating a multimedia interactive electronic tutorial in the field of computer science.

2. Significant is the fact that Microsoft Office Word allows you to create pdf-files, which on the one hand are a link to the entire tutorial, and on the other hand, have good navigation properties, and also regulate the size and scale of the material.

3. Using the pdf-format as the main one allows for the phased development of the manual - an electronic edition is first created as a pdf-file, and then a multimedia edition with elements of interactivity is built on its basis.

4. Testing in the educational process showed that methodological developments developed by the described technology are perceived by students as natural and contribute to improving the quality of the educational process both in the classroom and during independent work.

### **Literature**

1. Федько В. В. Классические средства доступа к данным: учебное пособие по учебной дисциплине «Базы данных» для иностранных студентов : Мультимедийное интерактивное электронное издание комбинированного использования / В. В. Федько, А. В. Тарасов, М. Ю. Лосев. – Х. : Изд. ХНЭУ им. С. Кузнеця, 2016. – 218 с.

2. Тарасов О. В. Особливості побудови електронного навчального посібника для вивчення профільної дисципліни за напрямком "Комп'ютерні науки" / О. В. Тарасов, В. В. Федько. – Системи обробки інформації. – 2016. – № 5 (142) – С. 219 – 223

3. Федько В. В. Captivate-реалізація тестів з дисципліни "Організація баз даних та знань» / В. В. Федько, О. В. Тарасов. – Системи обробки інформації. – 2016. – № 4 (141). – С. 192 – 196

4. Screen Recorder and Video Editor [Electronic resource]. – Access mode: <https://www.techsmith.com/video-editor.html>.

5. Unlock the future of smart eLearning design [Electronic resource]. – Access mode: <https://www.adobe.com/products/captivate.html>.



## ГЛАВА 13

### КЛАССИФИКАЦИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ИММУННОЙ МОДЕЛИ КЛОНАЛЬНОГО ОТБОРА

**Введение и постановка задачи.** При решении задачи классификации большое распространение получили системы моделирования искусственного интеллекта, функционирующие на основе биологических принципов организации вычислений [1-3]: искусственные нейронные сети, генетические и эволюционные алгоритмы, системы моделирования коллективного разума и др. Основным преимуществом перечисленных систем является возможность самообучения и гибкость управления процессом группировки объектов.

При решении практических задач возникает необходимость разработки модели классификации, позволяющей изменять способ группировки исходных данных в процессе работы. При этом модель должна быть адаптивной и сочетать особенности различных способов обучения. Аппарат искусственных иммунных систем (ИИС) позволяет проводить группировку данных различными способами, а также изменять не только структуру сети иммунных объектов, но и методы решения поставленной задачи [4, 5]. Использование адаптивной иммунной модели позволяет группировать данные либо на основе обучающей выборки (ОВ), либо в процессе иммунного обучения (ИМО).

Классификация данных на основе иммунного подхода может быть выполнена с контролируемым и неконтролируемым обучением, а также путем автоматической классификации [6-8]. Входными данными для проведения классификации являются набор объектов классификации  $OBJ\{obj_1; \dots; obj_n\}$ ,  $n = \overline{1, N}$ , каждый из которых описывается матрицей признаков, и набор объектов ОВ  $SMP\{smp_1; \dots; smp_m\}$ ,  $m = \overline{1, N}$ , представляющих исходные классы. В связи с тем, что в процессе определения класса для множества группируемых объектов происходит ИМО, то применительно к теории ИИС исходные объекты представляются популяцией антител  $AB\{ab_1; \dots; ab_n\}$ , а объекты ОВ – популяцией антигенов  $AG\{ag_1; \dots; ag_n\}$ .

Наиболее распространёнными моделями ИИС, используемыми для классификации данных, являются модели клонального отбора и искусственной иммунной сети [4, 5]. Решению задач классификации и кластеризации данных на основе ИИС посвящены работы [6, 8]. В [9]

рассмотрено решение задачи классификации данных на основе модели искусственной иммунной сети, позволяющей изменять как способ группировки исходных данных, так и структуру сети иммунных объектов. Рассмотрены особенности процесса ИМО на основе модели искусственной иммунной сети, и разработаны алгоритмы классификации данных с контролируемым обучением, с неконтролируемым обучением (кластеризация), а также автоматическая классификация данных.

В данной работе ставится задача классификации данных на основе использования иммунной модели клонального отбора. С этой целью необходимо рассмотреть особенности процесса ИМО на основе иммунной модели клонального отбора, и разработать алгоритмы классификация данных с контролируемым обучением, с неконтролируемым обучением (кластеризация), а также автоматической классификации данных.

### **Классификация на основе иммунной модели клонального отбора.**

Модель клонального отбора является одной из наиболее простых в реализации иммунных моделей. Поэтому она часто используется для решения различных практических задач. Наиболее распространёнными методами, реализующими данную модель, являются CLONALG и ВСА [4, 5]. Основное различие между данными методами заключается в способе обработки антител в процессе ИМО, проведении первичного отбора и особенностях клонирования антител. Благодаря некоторым особенностям работы метод ВСА больше подходит для решения задачи классификации, однако CLONALG характеризуется более высоким быстродействием [4, 5]. Поэтому для решения задачи классификации на основе клонального отбора базовым методом был выбран CLONALG.

**1. Особенности иммунного обучения.** Процесс ИМО при использовании модели клонального отбора, исключает возможность взаимодействия внутри популяции антител, либо между антителами и множеством клонов. Поэтому все внутривидовые взаимодействия между антителами производятся либо перед началом этапа ИМО, либо после его завершения.

Постановка задачи классификации не оказывает влияние на организацию процесса обучения. Поэтому при классификации, кластеризации и автоматической классификации объектов на основе модели клонального отбора, используется один и тот же подход к проведению ИМО. Изменения производятся только в популяциях антител и антигенов ОБ. Для решения проблемы избыточных вычислений, характерной для

популяционной обработки антител, используется последовательная обработка антител [10]. При этом для каждого клонируемого антитела формируется отдельное множество клонов, которые подвергаются мутации и редактированию. Процесс ИМО при классификации на основе модели клонального отбора на уровне иммунных операторов (ИО) представляется следующим образом:

$$\text{Learning}(AB, AG, T, n) = \left. \begin{array}{l} \text{Cloning}(ab'_i, CL_i, n) \rightarrow \\ \rightarrow \text{Mutation}(CL_i) \rightarrow \\ \rightarrow \text{Presentation}(CL_i, Ts_i) \rightarrow \\ \rightarrow \text{Selection}(CL_i, cl') \rightarrow \\ \rightarrow \text{Ageing}(ab'_i, cl') \end{array} \right\}^T \quad (1)$$

где  $AB$  – популяция классифицируемых антител;  $AG$  – популяция антигенов  $OB$ ;  $T$  – предельное количество популяций, формируемых при ИМО;  $n$  – минимальное количество клонов;  $\text{Cloning}(ab'_i, CL_i, n)$  – ИО пропорционального клонирования антител с ограничением минимального количества клонов;  $\text{Mutation}(CL_i)$  – ИО обратно-пропорциональной мутации;  $\text{Presentation}(CL_i, Ts_i)$  – ИО представления целевых объектов  $i$ -го антитела  $Ts_i$  множеству мутировавших клонов;  $\text{Selection}(CL_i, cl')$  – ИО конкурентно целевого отбора клонов;  $\text{Ageing}(ab'_i, cl')$  – ИО старения антител.

Поскольку модель клонального отбора исключает возможность взаимодействия между антителами в процессе ИМО, то множество целевых объектов  $Ts_i$  для каждого  $i$ -го классифицируемого антитела формируется исключительно из популяции антигенов  $OB$ . При этом в случае классификации допускается возможность пополнения популяции антигенов путем копирования (клонирования) антител, которые были классифицированы на начальном этапе. В случае кластеризации, в условиях отсутствия априорной информации о классах, популяция антигенов формируется из антител, являющихся начальными центрами формируемых кластеров путем клонирования.

В выражении (1) используется значение  $T$ , определяющее предельное количество популяций антител, формируемых в процессе ИМО. Данное значение является критерием останова. В качестве данного критерия может использоваться достижение состояния полной специфичности классифицируемых антител целевым объектам. Таким образом, при организации ИМО на основе модели клонального отбора, используются все перечисленные модификации ИО, а также подходы, применяемые для повышения скорости обучения.

**2. Классификация с контролируемым обучением.** Для решения задачи классификации с контролируемым обучением на основе модели клонального отбора сформирован метод CLONALGm, который является модификацией метода CLONALG. Основными особенностями CLOANLGm являются:

- последовательная обработка антител в процессе ИМО;
- использование целевых объектов для повышения скорости обучения;
- пропорциональное клонирование с ограничением минимального количества клонов;
- обратно-пропорциональная мутация на основе аффинности дальнего предка клонированного антитела;
- использование конкурентно-целевого отбора клонов;
- возможность классификации антител без достижения состояния полной специфичности некоторому целевому объекту.

Процесс ИМО в CLONALGm организован в соответствии с (1). Кроме того, в нем поддерживается возможность классификации антител без достижения полной специфичности какому-либо целевому объекту. При этом целевыми объектами для классифицируемых антител могут быть только антигены ОВ, поскольку модель клонального отбора не использует возможность сетевого взаимодействия внутри популяции антител или клонов. Для классификации и достижения состояния специфичности целевому объекту достаточным является удовлетворение следующего условия:

$$\forall ab_i \in C_j : \text{aff}_{iAG_j} \geq \frac{1}{\lambda} \text{NAT}_j, \quad (2)$$

где  $ab_i$  –  $i$ -е классифицируемое антитело;  $C_j$  –  $j$ -й исходный класс;  $\text{aff}_{iAG_j}$  – аффинность между антителом и всеми антигенами класса;  $\text{NAT}_j$  – значение средней аффинности между антигенами  $j$ -го класса;  $\lambda$  – коэффициент роста классов.

При выполнении данного условия антитело определяет принадлежность к одному из исходных классов, и не участвует в процессе ИМО. Проверка возможности классификации данного антитела происходит после завершения работы оператора старения. Для решения данной задачи в CLONALGm на этапе ИМО используется оператор классификации неспецифичного антитела, применяемый после завершения работы оператора старения. При этом иммунному объекту представляется ограниченное количество классов. Это обуславливается тем, что при определении принадлежности неспецифичного

антитела некоторому классу происходит представление всех его антигенов, даже если они не являются целевыми объектами классифицируемого антитела. Для сокращения количества вычислительных операций антителу представляются антигены только тех классов, которым принадлежат его целевые объекты. Количество таких антигенов должно удовлетворять следующему условию:

$$\frac{N_{T_{ij}}}{N_{T_i}} \geq \frac{N_j}{N}, \quad (3)$$

где  $N_{T_{ij}}$  – количество целевых антигенов  $i$ -го антитела  $j$ -го класса;  $N_{T_i}$  – общее количество целевых антигенов  $i$ -го антитела;  $N_j$  – количество антигенов  $j$ -го класса;  $N$  – общее количество антигенов в ОБ.

Использование оператора классификации неспецифичных антител в процессе ИМО позволяет сократить количество популяций антител, формируемых при обучении и повысить быстродействие алгоритма.

В соответствии с иммунной моделью классификации с контролируемым обучением работа иммунного алгоритма CLONALGm на уровне ИО представляется следующим образом:

$$\text{CLONALGm}(AB, AG, T, \lambda, n) = \left[ \begin{array}{l} \text{NatCalculation}(AG) \rightarrow \\ \rightarrow \text{Presentation}(AB, AG) \rightarrow \\ \rightarrow \text{ClassDetection}(AB', \lambda) \rightarrow \\ \rightarrow \text{TrgSelection}(AB'') \end{array} \right]^{\text{PRP}} \rightarrow \left[ \begin{array}{l} \rightarrow \text{Learning}(AB, AG, T, \lambda, n) \rightarrow \\ \rightarrow \left[ \begin{array}{l} \text{ClassDetection}(AB'', \lambda) \rightarrow \\ \rightarrow \text{Reconstruction}(AB, AB') \end{array} \right]^{\text{CLS}} \end{array} \right] \quad (4)$$

где PRP – подготовительный этап; CLS – этап уточнения границ классов; NatCalculation(AG) – ИО определения значений NAT; Presentation(AB, AG) – ИО представления антигенов антителам; ClassDetection(AB', λ) – оператор отбора и классификации без ИМО; TrgSelection(AB'') – оператор определения целевых антигенов для антител, которые классифицируются в процессе ИМО; Learning(AB, AG, T, λ, n) – этап ИМО, представленный в (1); ClassDetection(AB'', λ) – оператор уточнения границ классов и группировки антител, которые не были классифицированы в процессе ИМО; Reconstruction(AB, AB') – оператор восстановления исходных объектов AB и определения их принадлежности к исходным классам по результатам классификации изменяемой в процессе ИМО популяции антител.

Определение значений пороговых аффинностей  $NAT$  происходит благодаря взаимодействию между антигенами  $OB$ . При этом на основании аффинностей между ними происходит определение значения  $NAT_{AG}$  для всей популяции антигенов, и значения  $NAT_{AG_i}$  для антигенов, принадлежащих каждому  $i$ -му исходному классу. После представления антител обучающим антигенам происходит их начальная классификация.

В результате классификации части антител без проведения ИМО популяция обрабатываемых объектов сокращается, и для каждого антитела данной популяции происходит определение множества целевых антигенов, с которыми оно (и его клоны) будет взаимодействовать. После этого начинается процесс ИМО с возможностью классификации неспецифичных антител. В случае завершения ИМО в результате формирования предельно допустимого количества популяций иммунных объектов, в популяции антител допускается существование не классифицированных объектов. Такие антитела определяются для проведения дальнейшей классификации на этапе уточнения границ исходных классов. При классификации данных объектов из всего множества его целевых объектов отбирается антиген, характеризующийся максимальным значением аффинности с данным антителом, который определяет классификацию антитела. В соответствии с этим, принадлежность антитела к одному из исходных классов определяется по максимальному значению аффинности между данным антителом и одним из его целевых антигенов [6]. После классификации всего множества антител происходит процесс восстановления объектов. При этом происходит загрузка исходного набора антител, характеризующихся уникальными идентификаторами, благодаря использованию которых определяется связь между классифицированным антителом и исходным объектом. В соответствии с этим, для каждого исходного антитела устанавливается принадлежность к одному из исходных классов по результатам классификации популяции антител.

Для оценки работы  $CLONALGm$  и сравнения его с другими алгоритмами и методами, используемыми для решения задачи классификации с контролируемым обучением, использовались наборы данных, представленные в табл. 1. Каждый набор характеризуется количеством объектов и размерностью матриц признаков. При этом, если размерность матрицы признаков характеризуется значением  $3 \times 3$ , то каждый классифицируемый объект и объект  $OB$  описывается девятью признаками.

Таблица 1

## Характеристики наборов исходных данных

Наборы	Классифицируемые объекты	Обучающая выборка	Количество классов	Матрица признаков
Набор 1	500	100	3	3 × 3
Набор 2	500	100	3	5 × 5
Набор 3	1000	200	5	5 × 5
Набор 4	2000	400	5	5 × 5
Набор 5	5000	400	5	5 × 5
Набор 6	5000	400	5	10 × 10
Набор 7	10 000	500	10	5 × 5
Набор 8	15 000	500	10	5 × 5
Набор 9	20 000	500	10	5 × 5
Набор 10	20 000	500	10	10 × 10

Для определения эффективности CLONALGm при классификации с контролируемым обучением было произведено сравнение результатов работы данного алгоритма с рядом наиболее распространённых алгоритмов классификации, функционирующих, в том числе, на основе иммунного подхода. Данными методами являются [1-5]: kNN, CART, SVM, BCA, CLONALG, aiNET. При оценке эффективности перечисленных алгоритмов учитывались следующие параметры: точность классификации, быстродействие, устойчивость классификации. Результаты классификации приведенных в табл. 1 наборов данных представлены в табл. 2.

Таблица 2

## Результаты классификации с контролируемым обучением

Методы		Классифицируемые наборы данных									
		Н 1	Н 2	Н 3	Н 4	Н 5	Н 6	Н 7	Н 8	Н 9	Н 10
kNN (k = 5)	T	11.7	12.1	12.5	12.6	12.5	14.2	14.8	14.7	14.7	16.0
	A	96.2	96.3	93.8	91.6	90.5	90.5	87.2	86.1	84.3	84.4
	S	99.8	99.7	98.6	98.7	97.5	97.4	95.3	95.4	95.4	95.3
CART	T	38.2	39.3	39.8	39.7	39.9	41.5	42.4	42.6	42.5	43.9
	A	97.3	97.4	97.2	97.4	96.1	96.2	96.3	96.2	96.3	96.5
	S	99.4	99.3	99.4	98.9	98.8	98.8	98.6	98.7	98.6	98.1
SVM	T	51.3	52.4	55.2	54.9	54.8	55.7	59.8	59.9	59.8	60.1
	A	99.9	99.9	99.8	99.9	99.8	99.9	99.9	99.8	99.9	99.8
	S	99.9	99.9	99.8	99.8	99.9	99.8	99.9	99.9	99.8	99.9
BCA	T	100	100	100	100	100	100	100	100	100	100
	A	99.8	99.7	99.7	99.6	99.6	99.4	99.3	99.4	99.3	99.4
	S	99.2	99.1	98.8	98.3	97.6	97.5	96.9	96.8	96.8	96.1
CLONALG	T	54.6	54.8	55.7	55.8	56.1	61.2	61.1	61.3	61.2	65.5
	A	65.8	65.6	65.2	63.8	62.6	62.4	61.3	60.9	60.8	60.0
	S	85.6	85.5	84.3	84.2	84.1	84.0	83.7	83.7	83.6	83.5
aiNET	T	55.1	55.2	55.9	55.9	56.0	62.0	61.9	62.1	62.2	66.8
	A	65.4	64.9	64.5	61.7	60.9	60.3	59.8	59.4	59.0	59.0
	S	88.3	88.3	88.0	88.0	87.9	87.9	87.3	87.2	87.2	87.0
CLONALGm	T	39.5	39.6	39.8	39.9	39.8	41.3	41.5	41.4	41.5	43.4
	A	98.2	98.3	98.2	98.2	98.1	98.1	97.9	97.8	97.8	97.7
	S	99.8	99.7	99.2	99.2	99.1	99.1	98.8	98.8	98.7	98.7

В табл. 2. используется несколько условных обозначений: «Т» характеризует быстродействие алгоритма классификации (time); «А» используется для отображения точности группировки (accuracy); «S» характеризует устойчивость алгоритма (stability). По результатам классификации различных наборов данных (табл. 1) метод ВСА характеризуется наихудшим быстродействием, поэтому все оценки быстродействия приводятся относительно данного алгоритма. Для характеристик «А» и «S» наиболее точным и устойчивым методом классификации с контролируемым обучением является SVM, по отношению к которому определены характеристики других алгоритмов. Наилучшим быстродействием характеризуется метод kNN, который по сравнению с ВСА позволяет классифицировать объекты практически в восемь раз быстрее.

Из результатов классификации следует, что иммунные методы классификации значительно уступают методам SVM, CART и kNN по быстродействию, что является следствием большого количества избыточных вычислений на этапе ИМО. Методы aiNET и CLONALG уступают не иммунным алгоритмам по точности и устойчивости классификации. Однако среди иммунных методов выделяется ВСА, который по точности и устойчивости классификации сопоставим с методом SVM, однако значительно уступает ему в быстродействии. Метод CLONALGm превосходит методы CLONALG и aiNET по всем перечисленным характеристикам. При этом по точности классификации CLONALGm превосходит данные методы на 30-35%, по устойчивости группировки и быстродействию – на 10-15%, что является следствием сокращения количества избыточных вычислений. При сравнении методов CLONALGm и ВСА, он уступает ВСА по точности классификации на 1-2%, сопоставим с ним по устойчивости группировки, а при обработке больших наборов объектов превосходит его на 1-1.5% и значительно превосходит его по быстродействию (на 60%). При сравнении CLONALGm с не иммунными методами, он уступает SVM по точности и устойчивости на 1-2%, но превосходит его по быстродействию на 12-18%, и сопоставим с алгоритмом CART, превосходя его по точности и устойчивости на 1-2%.

Управление работой CLONALGm в соответствии с (4) осуществляется путем изменения двух входных параметров: предельного количества популяций антител  $T$ , формируемых в процессе ИМО, и минимального количества объектов  $n$ , создаваемых при клонировании антител. Увеличение предельно допустимого количества популяций антител приводит к снижению



быстродействия алгоритма и повышению точности группировки. При повышении минимального количества клонов происходит увеличение времени обработки множества клонируемых антител. Но при этом сокращается количество популяций антител, поскольку при увеличении количества клонов происходит снижение времени, затрачиваемого на достижение состояния специфичности между мутировавшими клонами и целевыми антигенами. Таким образом, посредством изменения данных параметров происходит управление не только быстродействием алгоритма, но и точностью группировки классифицируемых объектов.

**3. Классификация с неконтролируемым обучением.** Основной особенностью задачи кластеризации является отсутствие объектов ОВ, следствием чего является необходимость формирования кластеров для набора группируемых объектов. Для решения данной задачи на основе модели клонального отбора был использован метод CLONALGmc, который является модификацией CLONALGm. В [8] приводятся несколько методов кластеризации данных на основе CLONALG, благодаря которым сформирован метод CLONALGmc. Особенностью данного метода является формирование ОВ антигенов перед группировкой исходного набора объектов. При организации ИМО в CLONALGmc используются те же подходы к проведению клонирования, мутации и редактирования популяции иммунных объектов, что и в методе CLONALGm. Отличие заключается в необходимости формирования популяции антигенов посредством клонирования антител, являющихся начальными центрами кластеров. Для определения начальных центров кластеров используется значение средней аффинности между классифицируемыми антителами NAT.

На уровне ИО CLONALGmc представляется следующим образом:

$$\begin{aligned}
 \text{CLONALGmc}(AB, T, \lambda, n) = & \left[ \begin{array}{l} \text{NatCalculation}(AB) \rightarrow \\ \rightarrow \text{AgGeneration}(AB) \rightarrow \\ \rightarrow \text{Presentation}(AB, AG) \rightarrow \\ \rightarrow \text{ClusterDetection}(AG, AB, \lambda) \rightarrow \\ \rightarrow \text{TrgSelection}(AB'') \end{array} \right]^{\text{PRP}} \rightarrow \\
 & \rightarrow \text{Learning}(AB, AG, T, \lambda, n) \rightarrow \\
 & \rightarrow \left[ \begin{array}{l} \text{ClusterDetection}(AB'', AG) \rightarrow \\ \rightarrow \text{Reconstruction}(AB, AB') \end{array} \right]^{\text{CLS}}
 \end{aligned} \tag{5}$$

где  $\lambda$  – коэффициент роста кластеров;  $\text{Learning}(AB, AG, T, \lambda, n)$  – этап ИМО; CLS – этап завершения кластеризации;  $\text{NatCalculation}(AB)$  – ИО

определения значения аффинности NAT; AgGeneration(AB) – ИО определения начальных центров кластеров и формирования популяции антигенов AG; Presentation(AB,AG) – ИО представления популяции целевых антигенов группируемым антителам; ClusterDetection(AG,AB, $\lambda$ ) – ИО кластеризации антител без ИМО; TrgSelection(AB'') – ИО определения целевых антигенов для антител, участвующих в процессе ИМО; ClusterDetection(AB'',AG, $\lambda$ ) – ИО определения кластеров для неклассифицированных антител; Reconstruction(AB,AB') – ИО восстановления исходных объектов и определения их принадлежностей к выделенным кластерам.

Таким образом, работа CLONALGmc начинается с определения значения NAT для популяции антител, при использовании которого происходит определение начальных центров кластеров. Поскольку модель клонального отбора исключает возможность взаимодействия между антителами в процессе ИМО, происходит клонирование выделенных начальных центров кластеров для создания популяции обучающих антигенов. После этого происходит определение возможности кластеризации без ИМО. При этом для каждого антитела происходит определение возможности кластеризации в соответствии со следующим выражением:

$$\forall ab_i \in C_j : \frac{1}{\lambda} \text{aff}_{ij} \geq \text{NAT}, \quad (6)$$

где  $ab_i$  –  $i$ -е антитело, для которого проводится проверка возможности кластеризации без ИМО;  $C_j$  –  $j$ -й кластер, характеризующийся антигеном-центром;  $\lambda$  – коэффициент роста классов и кластеров;  $\text{aff}_{ij}$  – аффинность между  $i$ -м антителом и центром  $j$ -го кластера; NAT – значение средней аффинности между антителами.

После проведения кластеризации без ИМО для множества свободных антител происходит определение целевых объектов аналогично тому, как это описано в методе CLONALGm. Следует отметить, что действию ИО на этапе ИМО подвергаются только те антитела, которые не специфичны антигенам и не были отнесены ни к одному кластеру на этапе подготовительном этапе. После завершения процесса ИМО происходит уточнение границ выделенных кластеров, а также формирование новых кластеров из множества объектов, которые не могут быть отнесены ни к одному из существующих кластеров.

Для оценки работы метода CLONALGmc использовался набор данных,

представленные в табл. 1. При кластеризации данных наборов объектов количество формируемых кластеров не учитывалось, и использовалось только в случае необходимости (для работы алгоритмов k-means и MST). Таким образом, количество формируемых кластеров определяется методов CLONALGmc в ходе работы и не является входным параметром.

При определении эффективности метода CLONALGmc было произведено сравнение результатов кластеризации с результатами работы наиболее распространённых методов кластеризации. Данными методами являются k-means, MST, BCA, CLONALG, aiNET. Результаты классификации с помощью перечисленных методов кластеризации представлены в табл. 3.

Таблица 3

**Результаты кластеризации с неконтролируемым обучением**

Методы		Группируемые наборы данных									
		Н 1	Н 2	Н 3	Н 4	Н 5	Н 6	Н 7	Н 8	Н 9	Н 10
k-means	T	24.2	24.1	24.2	27.7	27.8	27.8	30.2	30.4	30.5	30.5
	A	99.8	99.7	99.7	99.6	99.5	99.4	99.5	99.4	99.4	99.3
	S	99.9	99.8	99.8	99.7	99.7	99.7	99.7	99.7	99.7	99.8
MST	T	15.3	15.3	16.0	16.4	16.5	16.5	17.0	17.1	17.3	17.3
	A	99.7	99.7	99.5	99.3	99.0	99.0	99.2	99.1	99.0	99.1
	S	99.8	99.8	99.8	99.8	99.7	99.7	99.7	99.8	99.7	99.7
BCA	T	100	100	100	100	100	100	100	100	100	100
	A	95.6	95.5	95.0	94.9	94.9	93.7	93.6	93.6	93.5	93.4
	S	97.8	97.7	97.7	97.5	97.4	97.2	97.2	97.3	97.2	97.1
CLONALG	T	37.9	37.9	38.1	38.2	38.6	38.7	40.1	40.2	40.2	40.3
	A	64.2	64.2	63.8	63.8	63.9	63.1	63.1	63.0	63.1	63.0
	S	90.2	90.1	90.0	89.9	89.9	89.7	89.8	89.7	89.7	89.6
aiNET	T	36.3	36.4	36.8	37.0	37.2	37.2	39.0	39.1	39.2	39.2
	A	65.8	65.7	65.4	65.3	65.3	65.0	64.9	64.9	64.8	64.8
	S	89.6	89.6	89.6	89.5	89.6	89.3	89.0	89.0	88.9	89.0
CLONALGmc	T	25.5	25.5	25.7	25.8	25.9	26.2	26.8	26.9	27.0	27.1
	A	94.8	94.8	94.6	94.7	94.6	94.1	94.0	94.0	94.1	94.0
	S	98.1	98.0	97.9	97.9	97.8	97.6	97.5	97.5	97.5	97.4

По результатам кластеризации различных наборов данных BCA характеризуется наихудшим быстродействием, но наибольшей точностью среди других иммунных методов. Методы k-means и MST характеризуются наибольшей точностью и устойчивостью группировки, при этом MST превосходит k-means по скорости классификации на 9-11% и характеризуется наилучшим быстродействием среди приведенных алгоритмов. Для данных

методов обязательным условием является наличие информации о количестве выделяемых кластеров. Несмотря на то, что иммунные методы классификации уступают методам MST и k-means, они способны к автоматическому выделению кластеров без использования какой-либо априорной информации об их количестве. Из результатов кластеризации следует, что модифицированный метод CLONALGmc превосходит методы CLONALG и aiNET и сопоставим по точности и устойчивости группировки с методом VCA, превосходя его по быстрдействию. При этом CLONALGm превосходит по быстрдействию метод k-means на 4-5%, уступая ему по устойчивости и точности на 2-3%.

Управление работой метода CLONALGmc осуществляется путем изменения трех входных параметров: предельного количества популяций антител  $T$ , минимального количества объектов  $n$ , создаваемых при клонировании антител и коэффициентом роста кластеров  $\lambda$ . Основное отличие между CLONALGm и CLONALGmc на уровне входных параметров заключается в использовании коэффициента роста кластеров. Этот параметр является вещественным значением и определяется в диапазоне  $[0.01; 5.00]$ . В случае, если он принимает значение меньше единицы, происходит сокращение количества выделяемых кластеров. В противном случае происходит пропорциональное увеличение количества кластеров. Если данный коэффициент принимает максимальное значение – возможно снижение точности группировки объектов.

**4. Автоматическая классификация объектов.** Поскольку задача автоматической классификации объектов подразумевает не только распределение группируемых объектов между множеством исходных классов, но и выделение кластеров для объектов, которые не могут быть классифицированы, методы CLONALGm и CLONALGmc не могут использоваться для решения данной задачи. Поэтому на их основе был сформирован метод классификации CLONALGma, реализующий основные принципы клонального отбора в процессе ИМО. Таким образом, в методе CLONALGma используются основные принципы организации ИМО CLONALGm и CLONALGmc. Основной особенностью данного метода является возможность разделения исходного множества классифицируемых объектов на популяции антител, для которых классификация, и соответственно ИМО, проходит различными способами. При этом после выделения популяций классифицируемых объектов, в процессе ИМО участвует все множество классифицируемых объектов, т.е. обучение не

проводится отдельно для популяции классифицируемых антител, и популяции антител, для которых необходимо проведение кластеризации. В соответствии с этим, работа метода CLONALGma на уровне ИО представляется следующим образом:

$$\begin{aligned}
 \text{CLONALGma} (AB, AG, T, \lambda, n) = & \left[ \begin{array}{l} \text{NatCalculation} (AG, AB) \rightarrow \\ \rightarrow \text{Presentation} (AB, AG) \rightarrow \\ \rightarrow \text{AgGeneration} (AB) \rightarrow \\ \rightarrow \text{ClassDetection} (AB, \lambda) \rightarrow \\ \rightarrow \text{ClusterDetection} (AG, AB'', \lambda) \rightarrow \\ \rightarrow \text{TrgSelection} (AB'', AB''') \end{array} \right]^{\text{PRP}} \rightarrow \\
 & \left[ \begin{array}{l} \left. \begin{array}{l} \text{Cloning} (ab_i, CL_i, n) \rightarrow \\ \rightarrow \text{Mutation} (CL_i) \rightarrow \\ \rightarrow \text{Presentation} (CL_i, Trg_i) \rightarrow \\ \rightarrow \text{Selection} (CL_i, cl') \rightarrow \\ \rightarrow \text{Ageing} (ab_i, cl') \rightarrow \\ \rightarrow \text{ClassDetection} (ab_i, Trg_i, \lambda) \end{array} \right\} \begin{array}{l} T \\ \text{LRN} \end{array} \\ \rightarrow \left[ \begin{array}{l} \text{ClassDetection} (AB'', \lambda) \rightarrow \\ \rightarrow \text{ClusterDetection} (AB''', AG) \rightarrow \\ \rightarrow \text{Reconstruction} (AB, AB') \end{array} \right]^{\text{CLS}} \end{array} \right] \rightarrow \quad (7)
 \end{aligned}$$

где  $\text{ClassDetection}(AB, \lambda)$  – ИО отбора и классификации антител;  $\text{ClusterDetection}(AG, AB'', \lambda)$  – ИО проверки возможности кластеризации антител без ИМО;  $\text{TrgSelection}(AB'', AB''')$  – ИО определения целевых антигенов;  $\text{Presentation}(CL_i, Trg_i)$  – ИО представления целевых объектов клонам;  $\text{Selection}(CL_i, cl')$  – ИО конкурентно-целевого клонального отбора;  $\text{Ageing}(ab_i, cl')$  – ИО старения клонированного антитела;  $\text{ClassDetection}(ab_i, Trg_i, \lambda)$ , – ИО определения класса для антитела;  $\text{ClassDetection}(AB'', \lambda)$ ,  $\text{ClusterDetection}(AB''', AG)$  – ИО определения кластеров для неспецифичных антител множества;  $\text{Reconstruction}(AB, AB')$  – оператор восстановления исходных объектов и определение их исходным принадлежности классам и кластерам.

Таким образом, метод CLONALGma сочетает особенности работы методов CLONALGm и CLONALGmc, используемых для проведения

классификации с различным способом обучения. В соответствии с этим, в случае отсутствия ОБ, CLONALG<sub>ma</sub> проводит классификацию с неконтролируемым обучением, в случае использования антигенов ОБ, данный алгоритм проводит классификацию с контролируемым обучением и возможностью формирования новых кластеров для антител, которые в силу особенностей их признаков не могут быть отнесены ни к одному исходному классу.

Управление работой CLONALG<sub>ma</sub> в соответствии с (7) осуществляется аналогично управлению методом CLONALG<sub>mc</sub>, путем изменения входных параметров, определяющих предельное количество популяций антител  $T$ , минимальное количество клонов  $n$  и коэффициента роста классов и кластеров  $\lambda$ . Следует отметить, что роль коэффициента  $\lambda$  в CLONALG<sub>ma</sub> выше, чем в CLONALG<sub>mc</sub>, поскольку от величины данного параметра зависит не только количество формируемых кластеров, но и точность классификации при использовании исходного набора классов.

**Заключение.** Использование ИИС позволяет проводить группировку данных различными способами, а также изменять не только структуру сети иммунных объектов, но и методы решения поставленной задачи. При этом решение задачи классификации на основе иммунного подхода сводится либо к решению задачи разделения данных между заранее известными группами, либо формированию новых групп из исходного множества данных.

В работе предлагается решение задачи классификации данных на основе использования иммунной модели клонального отбора, которая является одной из наиболее простых в реализации иммунных моделей и в которой в результате взаимодействия антител с антигенами остаются антитела, которые характеризуются высокими аффинностями к распознаваемым антигенам. Рассмотрены особенности процесса ИМО на основе иммунной модели клонального отбора и разработаны алгоритмы классификация данных с контролируемым обучением, с неконтролируемым обучением (кластеризация), а также автоматическая классификация данных.

Проведены сравнительные экспериментальные исследования предложенных алгоритмов классификация данных с существующими на различных наборах данных, которые показали высокую эффективность использования для этих целей иммунной модели клонального отбора.

## Литература

1. Duda R.O. Pattern classification / R.O. Duda, P.E. Hart, D.G. Stork. – Wiley & Sons, 2010. – 738 p.
2. Самаркин М.Е. Классификация данных телекоммуникационной компании с помощью технологии data-mining / М.Е. Самаркин, В.Н. Тарасов // Инфокоммуникационные технологии. – ПГУТИ, Самара, 2017. –Том 14. – № 3. – С. 258-263.
3. Ершов К.С. Анализ и классификация алгоритмов кластеризации / К.С. Ершов, Т.Н. Романова // Новые информационные технологии в информационных системах, 2016. – № 1. – С.274-279.
4. Dasgupta D. Immunological computation, Theory and applications / D. Dasgupta, L.F. Nino – Taylor & Francis Group, 2009. – 278 p.
5. Dasgupta D. Recent Advanced in Artificial Immune Systems: Models and Applications / D. Dasgupta, S. Yu, F. Nino // Applied Soft Computing. Elsevier, 2011. – P. 1574-1587.
6. Кораблев Н.М. Классификация объектов на основе искусственных иммунных систем / Н.М. Кораблев, А.А. Фомичёв // Системи обробки інформації. – 2010. – Вип. 6 (87). – С. 13–17.
7. Кораблев Н.М. Автоматическая классификация данных на основе иммунного похода / Н.М. Кораблев, А.А. Фомичев // Бионика интеллекта: науч.-техн. журнал. – 2014. – № 2 (83). – С. 83-90.
8. Кораблев Н.М. Кластеризация данных на основе искусственных иммунных систем / Н.М. Кораблев, А.А. Фомичев // Системи обробки інформації. – 2009. – Вип. 4 (78). – С. 77–82.
9. Кораблев Н.М. Классификация данных с использованием модели искусственной иммунной сети / Н.М. Кораблев, А.А. Фомичев // Інформаційні технології: сучасний стан та перспективи: монографія. За заг. ред. В.С. Пономаренка. – Х.: ТОВ «ДІСА ПЛЮС», 2018. – С. 86-101.
10. Кораблев, Н.М. Исследование иммунных операторов в задаче кластеризации объектов / Н.М. Кораблёв, А.А. Фомичёв // Бионика интеллекта. – 2010. – №1(72). – С. 70–74.

## **ДОДАТОК**



## ВІДОМОСТІ ПРО АВТОРІВ ТА АНОТАЦІЇ

### ГЛАВА 1

## ВИКОРИСТАННЯ МОДЕЛЕЙ ЕКСПЕРТНОГО АНАЛІЗУ ДЛЯ ОЦІНЮВАННЯ СТАНУ ІТ-ГАЛУЗІ УКРАЇНИ

*Н. О. Бринза, О. В. Вільхівська*

**Анотація.** *Сформовано тривимірний набір показників, що відображає стан ІТ-індустрії України та включає інвестиційну привабливість регіонів, внутрішній стан ІТ-ринку та зарплату. Побудовано математичну модель стану ІТ-індустрії України, яка представлена як комплексний показник. Розрахунки для визначення оцінки показників стану ІТ-індустрії виконуються на основі побудованої моделі. За допомогою кореляційно-регресійного аналізу визначено силу взаємозв'язку між обраними факторами та комплексним показником, а також значимість впливу факторів на комплексний показник. За допомогою критерію Фішера було виявлено, що запропонована математична модель адекватна описаним процесам. Проведено класифікацію регіонів з використанням кластерного аналізу.*

**Ключові слова:** *ІТ-індустрія, багатокритеріальність, оптимізація, нормалізація, інвестиційна привабливість, тривимірний кортеж, узагальнена скалярна оцінка якості можливих рішень, комплексний показник, коефіцієнт кореляції, коефіцієнт детермінації, кореляційний і регресійний аналіз, множинне регресійне рівняння, кластерний аналіз.*

**Abstract.** *The article formed a three-dimensional tuple of indicators that reflects the state of the IT industry in Ukraine, and includes the investment attractiveness of the regions, the internal state of the IT market, and salary. A mathematical model of the state of the IT industry in Ukraine has been built, which is presented as a complex indicator. The calculations for determination the assessment of indicators of the state of the IT industry are performed on the basis of the constructed model. Using correlation and regression analysis, the strength of the relationship between the selected factors and the complex indicator were determined, as well as the significance of the influence of factors on the complex indicator. Using the Fisher criterion, it was found that the proposed mathematical model is adequate to the processes that it describes. The classification of regions with using cluster analysis.*

**Keywords:** *IT-industry, multicriteriality, optimization, normalization, investment attractiveness, three-dimensional tuple, generalized scalar assessment of the*

*quality of feasible solutions, a complex indicator, coefficient of correlation, coefficient of determination, correlation and regression analysis, multiple regression equation, cluster analysis.*

## **ГЛАВА 2**

### **ТЕХНОЛОГІЯ СТВОРЕННЯ СПІЛЬНОГО КОНТЕНТУ В ВЕБ- БАЗОВАНОМУ НАВЧАЛЬНОМУ КАБІНЕТІ**

*О.І. Пушкар, Є. М. Грабовський*

***Анотація.** На основі використання методу експертного опитування визначено основні елементи веб-базованого навчального кабінету. Визначено складові зовнішнього оформлення сайту веб-базованого кабінету. Розглянуті основні типи складності основних видів технологій створення спільного контенту. Аналізуються основні кроки реалізації педагогічної моделі формування спільного контенту.*

***Ключові слова:** Навчальний кабінет, спільний контент, експертне опитування, веб-базована технологія.*

***Abstract.** Based on the use of the expert survey method, the main elements of the web-based classroom are identified. The components of the external design of a web-based cabinet site are determined. The main types of complexity of the main kinds of technologies of creation of the joint content are considered. The main steps of the pedagogical model of the formation of common content are analyzed.*

***Keywords:** Study room, shared content, expert survey, web-based technology.*

## **ГЛАВА 3**

### **INTEGRATION CHAT-BOTA @RIBS\_KARKAS\_BOT WITH EXPERT SYSTEM**

*V. P. Burdaev*

***Анотація.** В роботі представлені результати інтегрування чат-бота @ribs\_karkas\_bot з експертною системою для організації консультування в режимі онлайн. Дано опис архітектури і реалізація імплементації чат-бота месенджера TELEGRAM в експертну систему на базі системи "КАРКАС" - інструментальний засіб для побудови моделей баз знань. Розглянуто структуру взаємодії чат-бота і агентів експертної системи в онлайн режимі. Наведено приклад онлайн консультації експертної системи в медичній предметної області.*

**Ключові слова:** чат-боти, агенти, повідомлення, база знань, експертна система.

**Abstract.** The paper presents the results of integrating the chatbot @ribs\_karkas\_bot with an expert system for organizing online consultation. A description of the architecture and implementation of the chat bot messenger TELEGRAM in an expert system based on the system "KARKAS" – a tool for building models of knowledge bases. The structure of the interaction of chatbot and agents of the expert system in the online mode is considered. An example of an online consultation of an expert system in the medical subject area is given.

**Keywords:** chatbots, agents, messages, knowledge base, expert system.

## ГЛАВА 4

### МОДЕЛІ ОЦІНЮВАННЯ ЕФЕКТУ ВИКОРИСТАННЯ СИСТЕМИ ПІДТРИМКИ ЕЛЕКТРОННОГО НАВЧАННЯ

**В. В. Браткевич, В. Є. Климнюк**

**Анотація.** Розглянута множина критеріїв для оцінювання ефекту застосування альтернативних варіантів систем електронного навчання. Пропонується покрокова методика побудови моделі, яка дозволяє надати початкову безліч критеріїв оцінювання навчального ефекту застосування системи електронного навчання у вигляді відповідних за рівнем впливу рангів. Розроблена холархічна модель взаємозв'язку критеріїв оцінювання альтернативних варіантів систем електронного навчання.

**Ключові слова:** учбовий ефект, критерії оцінювання, ієрархія, холархія, рангова модель.

**Abstract.** Considered a set of criteria for assessing the effect of the use of alternative variants of e-learning systems. The step-by-step method of constructing a model is proposed, which allows to provide an initial set of criteria for evaluating the educational effect of using the system of electronic education in the form of the corresponding level of influence of the ranks. The holar model of the interconnection of criteria of evaluation of alternative variants of e-learning systems is developed.

**Keywords:** educational effect, evaluation criteria, hierarchy, holharhii, rank model.

## ГЛАВА 5

### КРАУДСОРСИНГОВА ПЛАТФОРМА ЯК ІТ-ІНСТРУМЕНТ ДЛЯ ЗАБЕЗПЕЧЕННЯ ГЕНДЕРНОЇ РІВНОСТІ В УНІВЕРСИТЕТІ

*Г. О. Плеханова, О. Б. Плоха*

***Анотація.** Узагальнено досвід розробки та використання краудсорсингової онлайн-платформи в якості ІТ-інструменту для підтримки процесів гендерного аудиту університету, виявлення та подолання проблеми щодо гендерного паритету в освітніх установах. Розглянуто передумови розробки краудсорсингової онлайн-платформи, категорії користувачів, механізми та результати функціонування платформи, статистика її використання, джерела інформації та механізми розробки Плану гендерної рівності на прикладі Харківського національного економічного університету імені Семена Кузнеця загалом та кафедри інформаційних систем зокрема.*

***Ключові слова:** краудсорсингова онлайн-платформа, План гендерної рівності, гендерний аудит університетів.*

***Abstract.** The experience of the development and use of the crowdsourcing online platform as an IT tool for supporting the processes of gender audit of the university, identifying and overcoming the problem of gender parity in educational institutions is summarized. The paper considered the preconditions of development of the crowdsourcing online platform, user categories, its usage statistics, mechanisms and results of the platform's operation, information sources and mechanisms for developing the Gender Equality Plan on the example of the Simon Kuznets Kharkiv National University of Economics in general and the Information Systems Department in particular.*

***Keywords:** crowdsourcing online platform, Gender Equality Plan, gender audit of universities.*

## ГЛАВА 6

### FORECAST ERROR MEASURES: CRITICAL REVIEW AND PRACTICAL RECOMMENDATIONS

*O. G. Rudenko, A. A. Bezsonov*

***Анотація.** Аналізуються існуючі і досить поширені статистичні (RMSE, MAPE, MAE та інші) і деякі інформаційні критерії (AIC, KIC, HQ тощо)(метрики) помилок прогнозування, їх переваги та недоліки. На основі*

результатів аналізу розроблено рекомендації щодо вибору критеріїв прогнозування при вирішенні практичних задач. У зв'язку з неможливістю виділити один універсальний критерій рекомендується використовувати певний набір метрик помилки.

**Ключові слова:** прогнозування, часовий ряд, критерії оцінювання, моделювання, помилка прогнозу

**Annotation.** Existing and fairly common statistical (RMSE, MARE, MAE and others) and some information criteria (AIC, KIC, HQ, etc.) (metrics) of forecasting errors, their advantages and disadvantages are analyzed. Based on the results of the analysis, recommendations on the choice of forecast criterion for solving practical problems were developed. Due to the inability to isolate one universal criterion, it is recommended to use a specific set of error metrics.

**Keywords:** forecasting, time series, estimation criterion, modeling, forecast error

## ГЛАВА 7

### ANALYSIS OF THE INFLUENCE OF INTERFERENCE ON THE STATISTICAL PROPERTIES OF ESTIMATES OF NON-STATIONARY PARAMETERS

**O. G. Rudenko, A. S. Romanyk**

**Анотація.** Розглядається задача ідентифікації нестационарного об'єкту з допомогою найбільш простого в обчислювальному відношенні однокрокового алгоритму модифікованого Качмажа. Модифікація полягає у використанні його регуляризації алгоритму з метою поліпшення його обчислювальних властивостей.

Показано, що використання додатку, який регуляризує алгоритм ідентифікації, покращує стійкість алгоритмів, але призводить до деякого уповільнення процесу побудови моделі. Визначено умови збіжності регуляризованого алгоритму Качмажа при оцінюванні нестационарних параметрів, які описано Марківською моделлю першого порядку, при наявності завад вимірів. Отримані неасимптотичні оцінки є досить загальними і залежать як від ступеня нестационарності об'єкту, так і від статистичних характеристик завад. Якщо ж ці параметри не відомі, слід скористатися будь-якої рекурентною процедурою їх оцінювання і використовувати одержувані оцінки для уточнення параметрів, що входять в алгоритми.

**Ключові слова:** нестационарний об'єкт, ідентифікація, алгоритм, оцінка, регуляризація, нестационарні параметри, статистичні характеристики

**Annotation.** The problem of identifying a non-stationary object is considered with the help of the simplest one-speed algorithm of the modified Kaczmarz in the computational way. Modification is to use its regularization algorithm to improve its computing properties.

It is shown that the use of the application that regularises the identification algorithm improves the stability of the algorithms, but leads to some slowdown in the process of constructing the model. The conditions of convergence of the regularized Kaczmarz algorithm are determined in the estimation of nonstationary parameters, which are described by the Markov model of the first order, in the presence of disturbances of measurements. The resulting non-asymptotic estimates are quite general and depend on both the degree of non-stationary object and the statistical characteristics of the obstacles. If these parameters are not known, one should use any recurring procedure for their evaluation and use the obtained estimates to clarify the parameters included in the algorithms.

**Keywords:** non-stationary object, identification, algorithms, estimates, regularization, nonstationary parameters, statistical characteristics

## ГЛАВА 8

### КЛАСИФІКАЦІЯ ЕЛЕКТРОННИХ НАУКОВО-ТЕХНІЧНИХ ТЕКСТІВ В ІНФОРМАЦІЙНО-ПОШУКОВИХ СИСТЕМАХ

**С. Г. Удовенко, Л. Е. Чала**

**Анотація.** Запропоновано та програмно реалізовано метод класифікації колекції науково-технічних документів, що полягає у наявності фіксованої кількості деяких еталонних текстів та необхідності оперативного пошуку в цій колекції множин документів, які формуються згідно з розробленим методом. Кінцевою метою такої класифікації є пошук у колекції класів документів, які є близькими до заданих еталонів за запропонованими критеріями.

**Ключові слова:** класифікація текстів, програмний модуль, точність класифікації, формування простору ознак, міра близькості

**Abstract.** *The method of classification of a collection of scientific and technical documents is proposed and programmed, which consists in the availability of a fixed number of certain reference texts and the necessity of an operational search in this collection of sets of documents, which are formed according to the developed method. The ultimate goal of such a classification is to search in a collection of classes of documents that are close to the given standards according to the proposed criteria.*

**Keywords:** *classification of texts, program module, accuracy of classification, formation of space of signs, degree of proximity*

## ГЛАВА 9

### АНАЛИЗ БЕЗОПАСНОСТИ ПРОТОКОЛОВ ПЕРЕДАЧИ СООБЩЕНИЙ В MESSENGER TELEGRAM

*А. А. Поляков, В. Н. Федорченко*

**Аннотация.** *Рассмотрены используемые протоколы и алгоритмы передачи сообщений в Telegram и более конкретно схемы шифрования в MTProto. Проанализировано первичная регистрация клиента на сервере, процедура установления безопасного прозрачного зашифрованное соединения с другим узлом, а также процедура генерации ключевых данных.*

*Этот анализ основан на официальном исходном коде клиента Telegram для версии 2.7.0.*

**Ключевые слова:** *мессенджер, криптографический протокол, RSA, SHA1, Диффи-Хелман протокол распределения ключей, AES IGE, вектор инициализации, сеансовый ключ, общий секретный ключ, атака воспроизведения, атака зеркального отражения, счетчики сообщений.*

**Annotation.** *The used protocols and message transfer algorithms in Telegram and more specifically the encryption scheme in MTProto are considered. The primary registration of the client on the server, the procedure for establishing a secure, transparent, encrypted connection with another node, and the procedure for generating key data were analyzed.*

*This analysis is based on the official Telegram client source code for version 2.7.0.*

**Key words:** *messenger, cryptographic protocol, RSA, SHA1, Diffie-Hellman key distribution protocol, AES IGE, vector initialization, session key, shared secret key, replay attack, mirroring attack, message counters.*

## ГЛАВА 10

### МЕТОДОЛОГІЧНА ОСНОВА ПРОЕКТУВАННЯ ПРОБЛЕМНО-ОРІЄНТОВАНИХ ВЕБ-СЕРВІСІВ

*Н. Г. Аксак, Д. М. Росінський*

***Анотація.** Запропоновано методологічну основу для розв'язання наукомістких проблем, яка ґрунтується на комплексному використанні різних технологій, методів та моделей для синтезу, інтелектуалізації, підвищення ефективності та застосування розподілених спеціалізованих комп'ютерних систем у рамках єдиної технології створення проблемно-орієнтованих систем. Розроблено модель інформаційної безпеки для захисту медичної інформації та компонентів інформаційної системи.*

***Ключеві слова:** розподілені обчислення, інформаційна безпека, технології обробки великих даних, cloud-fog-dew.*

***Abstract.** A methodological basis for solving high technology issues, based on integrated use of various technologies, methods and models for the synthesis, intellectualization, increase efficiency and the use of specialized computer systems distributed within a single technology of problem-oriented systems, is proposed. Information safety model for the protection of medical information and information system components is proposed.*

***Keywords:** distributed computing, information safety, big data processing, cloud-fog-dew.*

## ГЛАВА 11

### НЕЧІТКО-МНОЖИННА МЕТОДИКА ОЦІНКИ КОНКУРЕНТОСПРОМОЖНОСТІ ПІДПРИЄМСТВА

*М. Ю. Лосев*

***Abstract.** The paper discusses the methodology for assessing the competitiveness of enterprises on the basis of a fuzzy-plural interval analysis of indicators of various nature. The methodology allows to avoid the ambiguity of their influence on the assessment result. The methodology does not require determining the weight values of each of the indicators, the results of the comparative assessment are resistant to this kind of uncertainty. The fuzzy-plural model allows to take into account the influence of both internal and external factors affecting the competitiveness of enterprises.*



**Keywords:** *enterprise, competitiveness, fuzzy set, multi-criteria evaluation, membership function, linguistic variable, uncertainty, probability of belonging.*

## ГЛАВА 12

### DEVELOPMENT OF A MULTIMEDIA INTERACTIVE ELECTRONIC EDUCATIONAL EDITION USING THE MODULES OF VARIOUS FORMAT

**V. V. Fedko**

**Анотація.** *Розглянуто питання побудови та використання мультимедійного інтерактивного електронного навчального посібника в галузі знань «Інформаційні технології». Проаналізовано інструментальні засоби для створення мультимедійних файлів такі програми, як Microsoft Office Word, Camtasia Studio і Adobe Captivate, визначені найбільш раціональні сфери їх застосування. Зокрема зазначено, що Microsoft Office Word краще підходить для створення текстів викладу теоретичного матеріалу та інструкцій щодо виконання лабораторних робіт, Camtasia Studio – для створення відео, в яких демонструються кроки виконання завдань лабораторних робіт в середовищі Visual Studio, а Adobe Captivate – для створення тестів, які є дидактичним засобом підтримки етапу контролю рівня знань. Розглянуто переваги використання pdf-формату як первинного для реалізації електронного навчального посібника.*

**Ключові слова:** *мультимедійний інтерактивний електронний навчальний посібник, інформаційні технології, технології баз даних.*

**Abstract.** *The issues of construction and use of multimedia interactive electronic tutorial in the field of knowledge "Information Technologies" are considered. Analyzed tools for creating multimedia files, such programs as Microsoft Office Word, Camtasia Studio and Adobe Captivate, identified the most rational areas of their use. In particular, it states that Microsoft Office Word is better suited for creating texts of theoretical material and instructions for performing laboratory work, Camtasia Studio for creating video demonstrating steps for performing laboratory tasks in Visual Studio, and Adobe Captivate for creating tests, which are didactic means of supporting the stage of knowledge control. The advantages of using pdf-format as the main one for the implementation of an electronic tutorial are considered.*

**Keywords:** *multimedia interactive e-learning tutorial, information technology, database technology.*

## ГЛАВА 13

### КЛАССИФИКАЦИЯ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ИММУННОЙ МОДЕЛИ КЛОНАЛЬНОГО ОТБОРА

*Н. М. Кораблев, А. А. Фомичев*

**Аннотация.** *Рассмотрено решение задачи классификации данных на основе иммунной модели клонального отбора, в которой в результате взаимодействия антител с антигенами остаются антитела, которые характеризуются высокими аффинностями к распознаваемым антигенам. Классификация данных на основе иммунной модели клонального отбора выполнена с контролируемым и неконтролируемым обучением, а также путем автоматической классификации. Проведены экспериментальные исследования, подтверждающие эффективность предложенных алгоритмов.*

**Ключевые слова:** *классификация, иммунная модель клонального отбора, обучающая выборка, иммунное обучения, аффинность, антитело, антиген.*

**Abstract.** *The solution of the problem of classifying data based on the clonal selection immune model, in which as a result of the interaction of antibodies with antigens, antibodies remain, which are characterized by high affinities to recognizable antigens, is considered. Classification of data based on the clonal selection immune model is carried out with controlled and uncontrolled learning, as well as by automatic classification. Experimental studies confirming the effectiveness of the proposed algorithms have been out.*

**Keywords:** *classification, clonal selection immune model, training sample, immune training, affinity, antibody, antigen.*

НАУКОВЕ ВИДАННЯ

**ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ**

**Монографія**

*під ред. д.е.н., проф. Пономаренка В. С.*

Відповідальний за випуск **Руденко О. Г.**

Відповідальний редактор **Грабовський Є. М.**

Підп. до друку 2.05.19    Формат 60 × 90 1/16.    Папір.    Друк офсетний.  
Ум.-друк. арк. 13,25    Обл.-вид. арк.    Тираж 150 екз.    Зам. №  
*Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи*  
Видавець і виготівник – ПП «Стиль-іздат»