

Дипломний проект «Модель адаптивного електронного навчання на базі алгоритму кластеризацій CLOPE»

Виконав: студент групи КА-44м Крюков О.О.
Керівник проекту: ст. вик. каф. ММСА, к.т.н. Діденко Д.Г.

Київ - 2016

Об'єкт та предмет дослідження

- ▶ Об'єкт дослідження: адаптивні системи навчання.
- ▶ Предмет дослідження: застосування алгоритму кластеризації категорійних даних для побудови моделі адаптивного навчання.

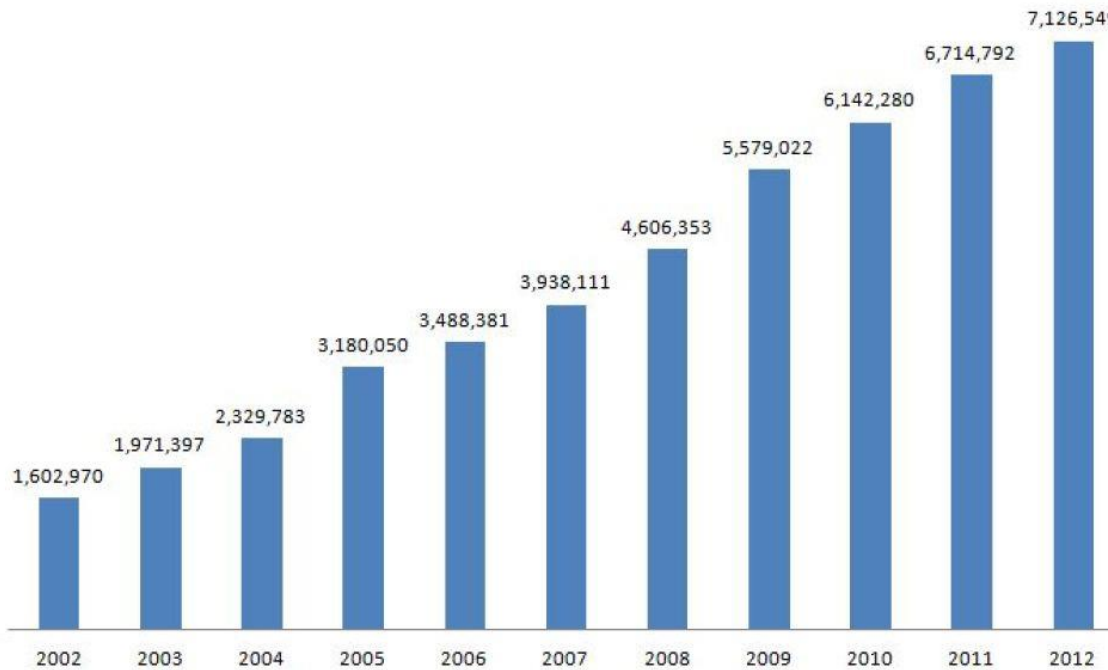
Цілі та задачі

- ▶ Ціллю даної роботи є розробка та впровадження гнучкої моделі адаптивного навчання, що дозволяє структурувати доменну модель з мінімальним залученням експерта.
- ▶ Основні задачі:
 - автоматизація побудови структури доменної моделі;
 - побудова адаптивної моделі навчання;
 - дослідження залежностей на базі побудованої моделі.

Онлайн освіта

Online Education Enrollment, 2002-2012

■ Students Taking at Least One Online Course



Source: Babson Survey Research Group



duolingo



Підходи до побудови адаптивних систем навчання.

- ▶ дерева рішень;
- ▶ продукційні правила;
- ▶ байесівські мережі довіри;
- ▶ ланцюги Маркова;
- ▶ мережі Петрі.

Недоліки існуючих моделей

- ▶ велика трудоемність побудови структури доменної моделі;
- ▶ проблема «холодного старту»;
- ▶ залежність від даних про результати навчання інших користувачів;
- ▶ швидкість роботи.

Платформа “Quizful”

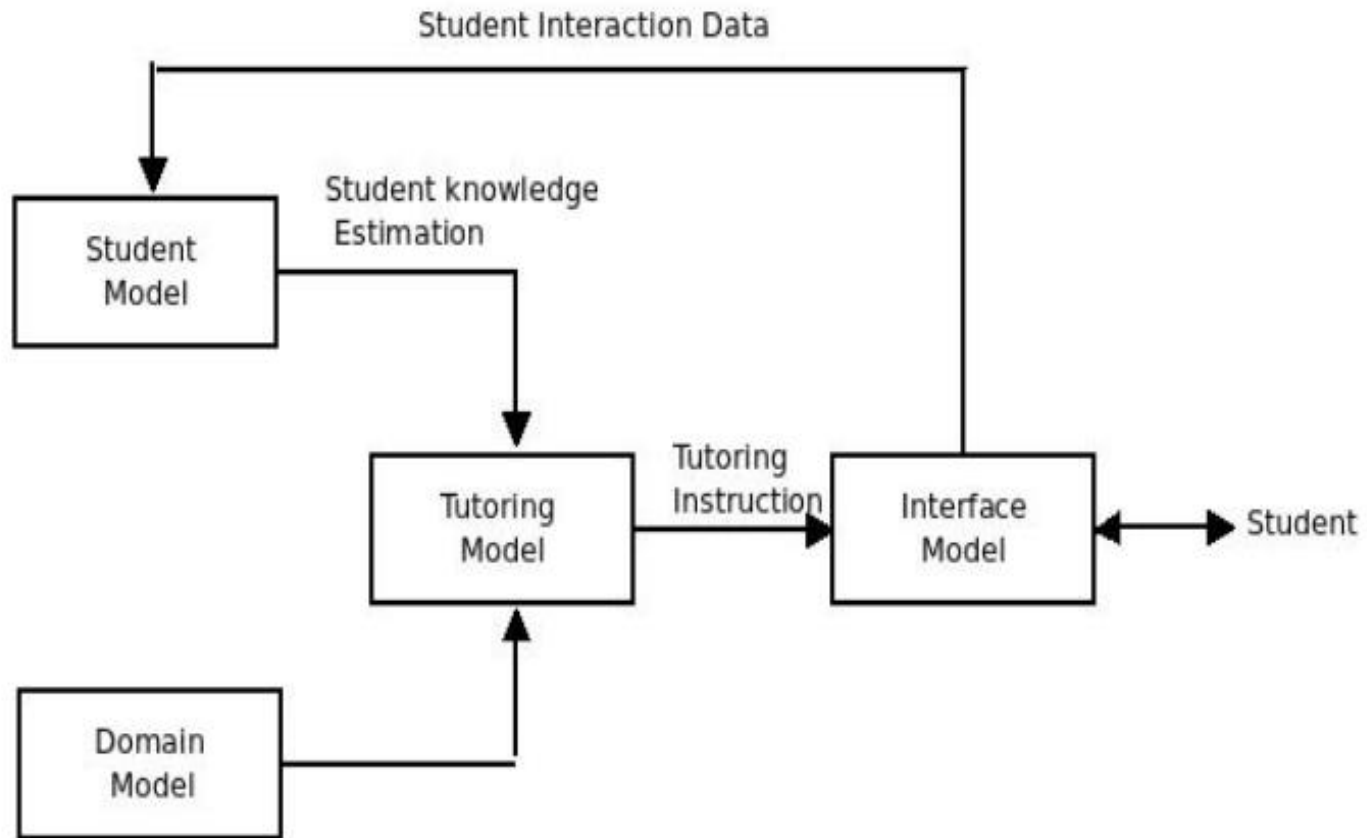
The screenshot displays the Quizful platform interface. At the top, there is a navigation bar with a hamburger menu, the text "Math Equations...", and a "Get Full Access" button. On the right side of the navigation bar, there are icons for a document, a chat bubble, and a user profile with a red "Alpha" badge.

The main content area shows a math problem: "Solve the equation: $(x - 1) + (x + 2) = 3$ ". Below the equation, there are four multiple-choice options: $x = 2$, $x = -1$, $x = -2$, and $x = 1$. The option $x = 1$ is marked as correct with a green checkmark. Below the options, there is an "Explanation" section with the following text: "Reveal brackets and reduce similar terms: $2x + 1 = 3$. Add -1 to both sides of the equation: $2x + 1 - 1 = 3 - 1$. $2x = 2$. Divide both sides of the equation by 2: $x = 1$ ".

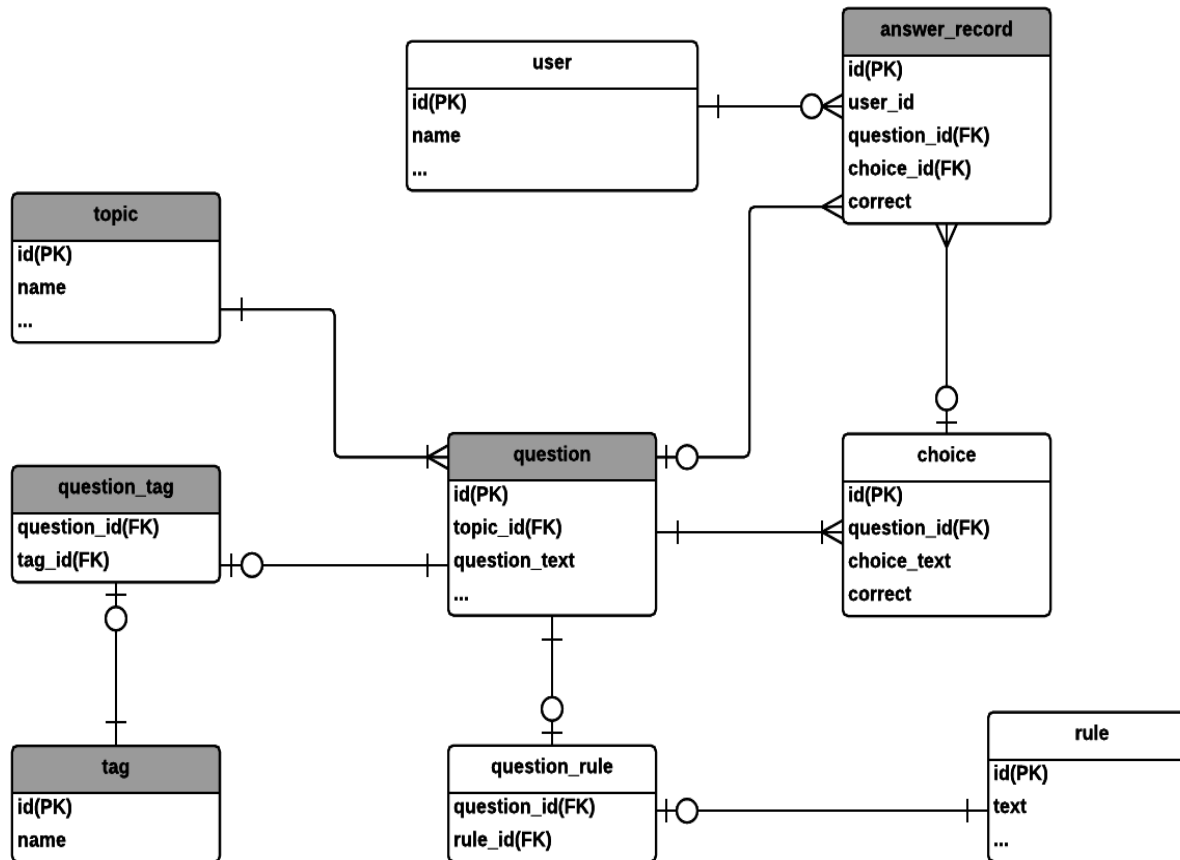
At the top and bottom of the main content area, there are pink banners with the text "Oops! This is the wrong answer.".

On the right side, there is a "Module Statistics" overlay. It shows a progress bar and two buttons: "Right (x)" and "Wrong (x)". Below this, there is a yellow "Update your username" dialog box with a text input field containing "test_user", a green checkmark indicating "Username is available.", and "Update" and "Cancel" buttons. Below the dialog box, there is a "Follow ExplainMe" section with Twitter and Facebook icons, and a "Send Feedback" button.

Загальна схема адаптивної системи навчання



Доменная модель в рамках платформы “Quizful”



Алгоритм CLOPE

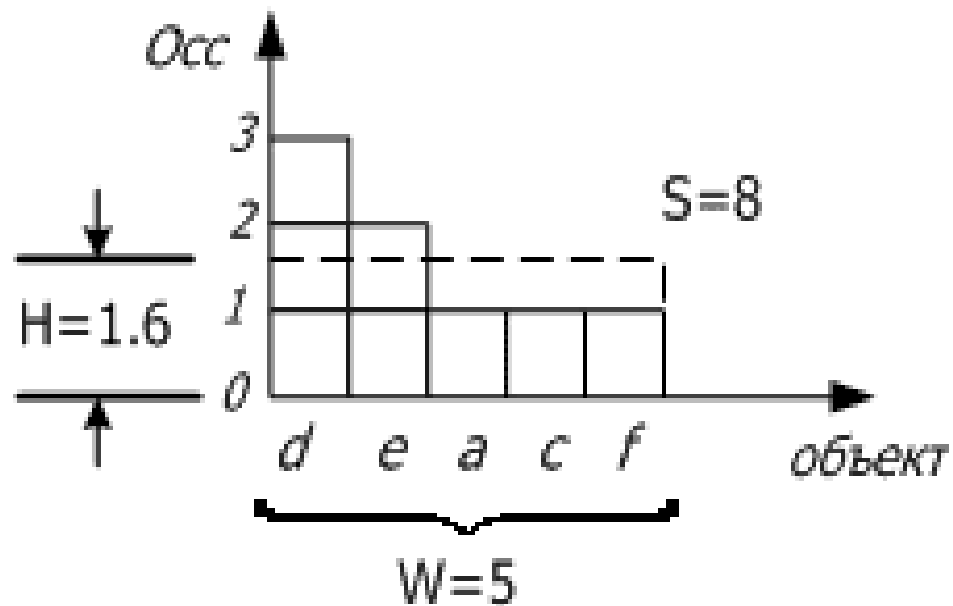
Алгоритм CLOPE - алгоритм кластеризації категорійних даних, запропонований в 2002 році групою китайських вчених. Він забезпечує більш високу продуктивність і кращу якість кластеризації в порівнянні з багатьма алгоритмами кластеризації категорійних даних.

В основі алгоритму кластеризації CLOPE лежить ідея максимізації глобальної функції корисності, яка підвищує близькість транзакцій в кластерах за допомогою збільшення параметра кластерної гістограми.

Глобальна функція корисності обчислюється як:

$$\text{Profit}(C) = \frac{\sum_{i=1}^k G(C_i) \times |C_i|}{\sum_{i=1}^k |C_i|} = \frac{\sum_{i=1}^k \frac{S(C_i)}{W(C_i)^r} \times |C_i|}{\sum_{i=1}^k |C_i|}$$

Приклад гістограми кластера CLOPE



Проблема розподілу питань на теми

- ▶ Базовий алгоритм не дає можливості задавати кількість кластерів або елементів в кластері. Проте в рамках даної задачі необхідно сформулювати рівномірне розбиття усіх питань на кластери із заданою кількістю елементів.

Обмеження верхньої границі кластера в модифікованому алгоритмі CLOPE

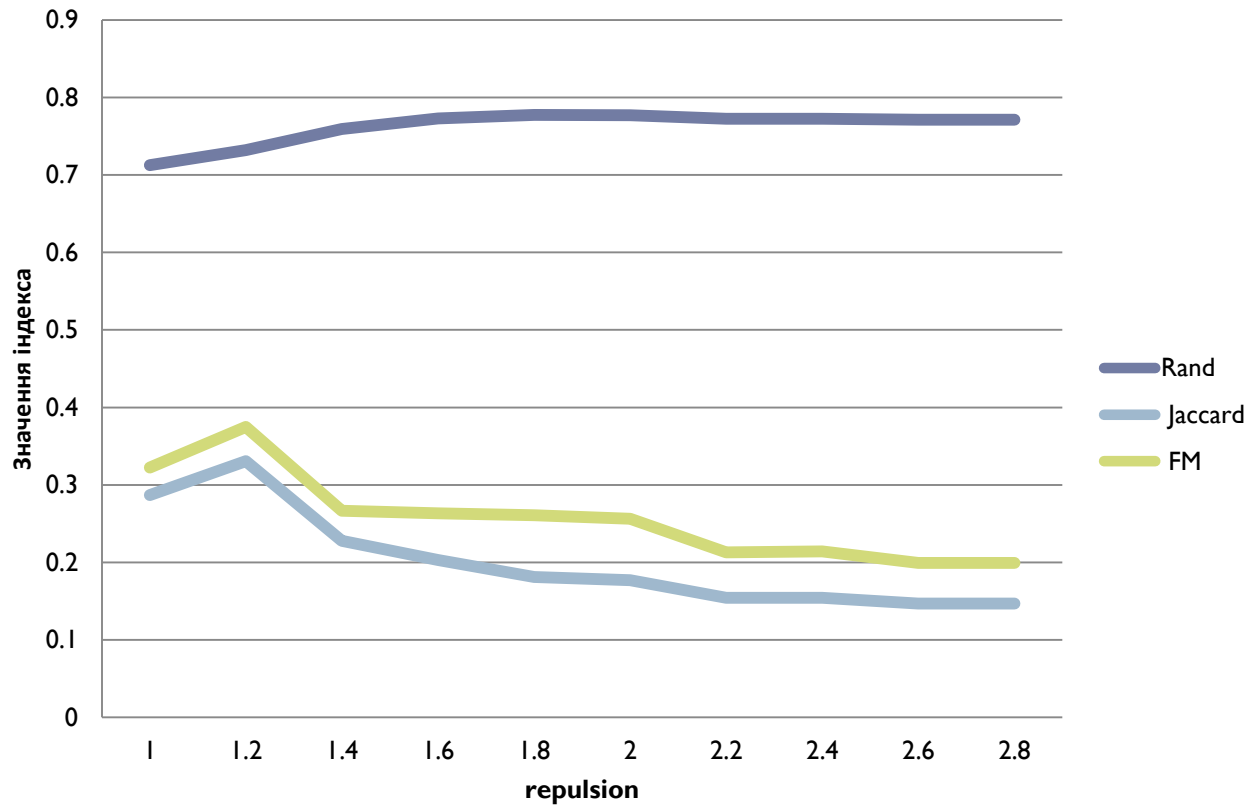
$$\begin{aligned} & Profit(C) \\ &= \frac{\sum_{i=1}^k \frac{S(C_i)}{W(C_i)^r} \times \log_b(|C_i| + 1)}{\sum_{i=1}^k |C_i|} \end{aligned}$$

Обмеження нижньої границі кластера в модифікованому алгоритмі CLOPE

Додатковий етап алгоритму кластеризації:

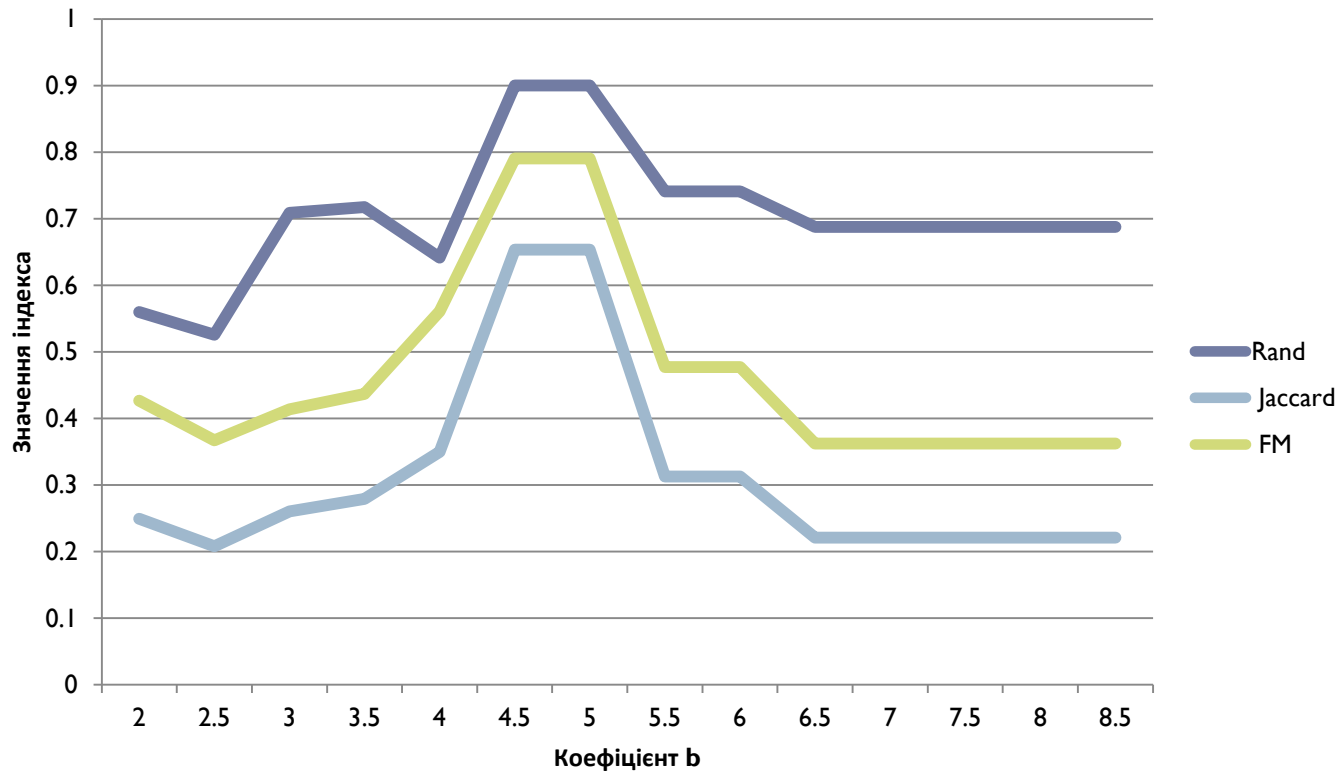
1. Сортування кластерів за кількістю елементів за зростанням;
2. Вибираємо перший кластер із сортованого списку. Якщо кількість елементів в ньому менше заданого порогу `minClusterSize`, то на крок 3, інакше - завершення алгоритму.
3. Ітеративний перерозподіл елементів кластера в інші кластери (для кожного елемента вибираємо кластер, який дає найбільший приріст функції корисності). Видаляємо кластер зі списку. Переходимо на крок 1.

Оцінка розбиття базовим алгоритмом



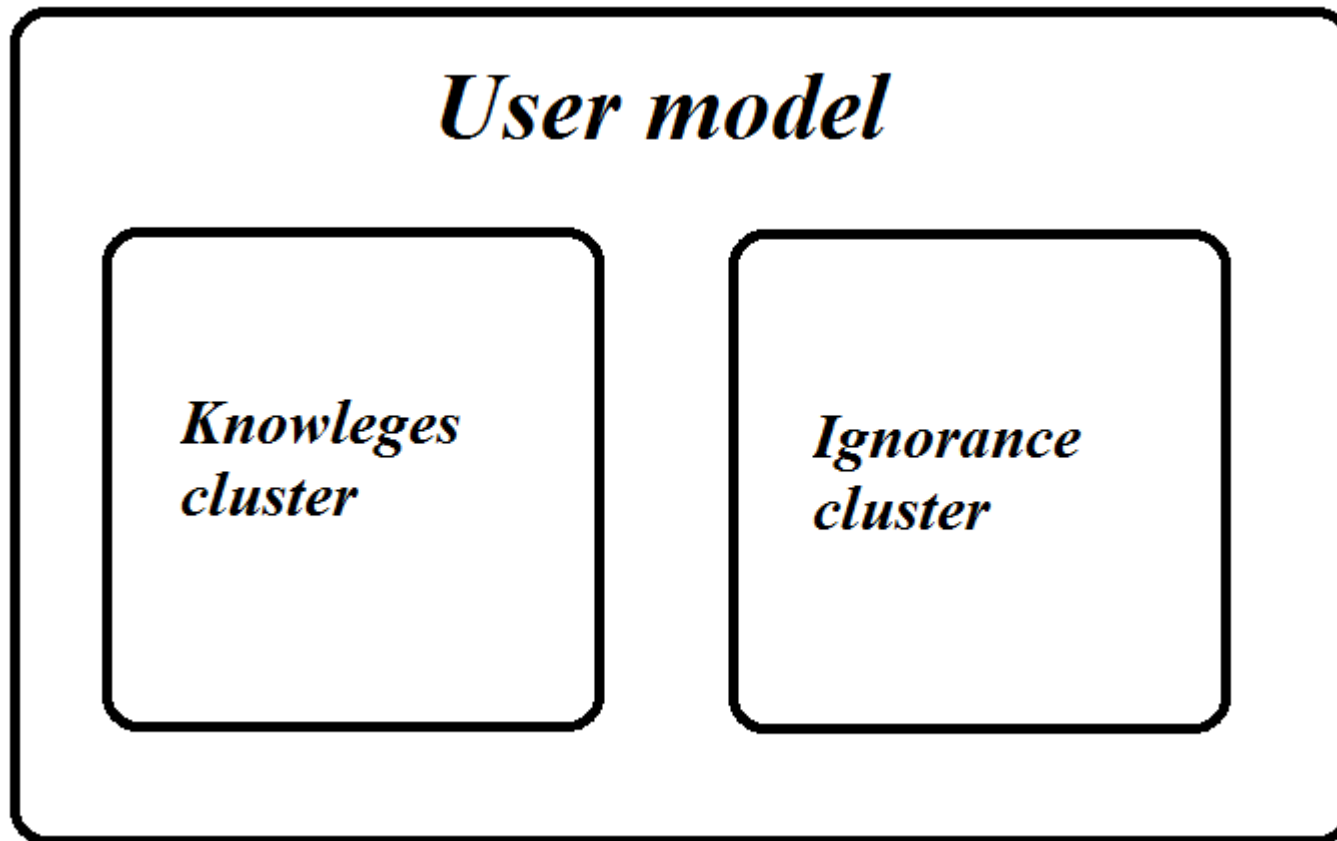
Графік 1. Залежність значень коефіцієнтів Rand, Jaccard, FM від значення repulsion.

Оцінка розвитку модифікованим алгоритмом



Графік 2. Залежність значень коефіцієнтів Rand, Jaccard, FM від значення b .

Модель учня



Алгоритм підбору питання

- ▶ В процесі навчання основною метою є збільшення глобальної функції корисності, що уособлює модель учня:

$$Pr\ ofit = \frac{S(C^{know})}{W(C^{know})^r} \times |C^{know}| \times k - \frac{S(C^{ign})}{W(C^{ign})^r} \times |C^{ign}| \times (1-k)$$

Алгоритм підбору питання

- ▶ Таким чином вибираємо наступне питання з максимальним значенням $\Delta Profit$:

$$\Delta Profit = \Delta Profit_{know}(q) - \Delta Profit_{ign}(q) =$$
$$\left(\frac{(S(C_{old}^{know}) + |q|) \times (|C_{old}^{know}| + 1)}{W(C_{new}^{know})^r} - \frac{(S(C_{old}^{know})) \times (|C_{old}^{know}|)}{W(C_{old}^{know})^r} \right) \times k$$
$$- \left(\frac{(S(C_{old}^{ign}) + |q|) \times (|C_{old}^{ign}| + 1)}{W(C_{new}^{ign})^r} - \frac{(S(C_{old}^{ign})) \times (|C_{old}^{ign}|)}{W(C_{old}^{ign})^r} \right) \times (1 - k)$$

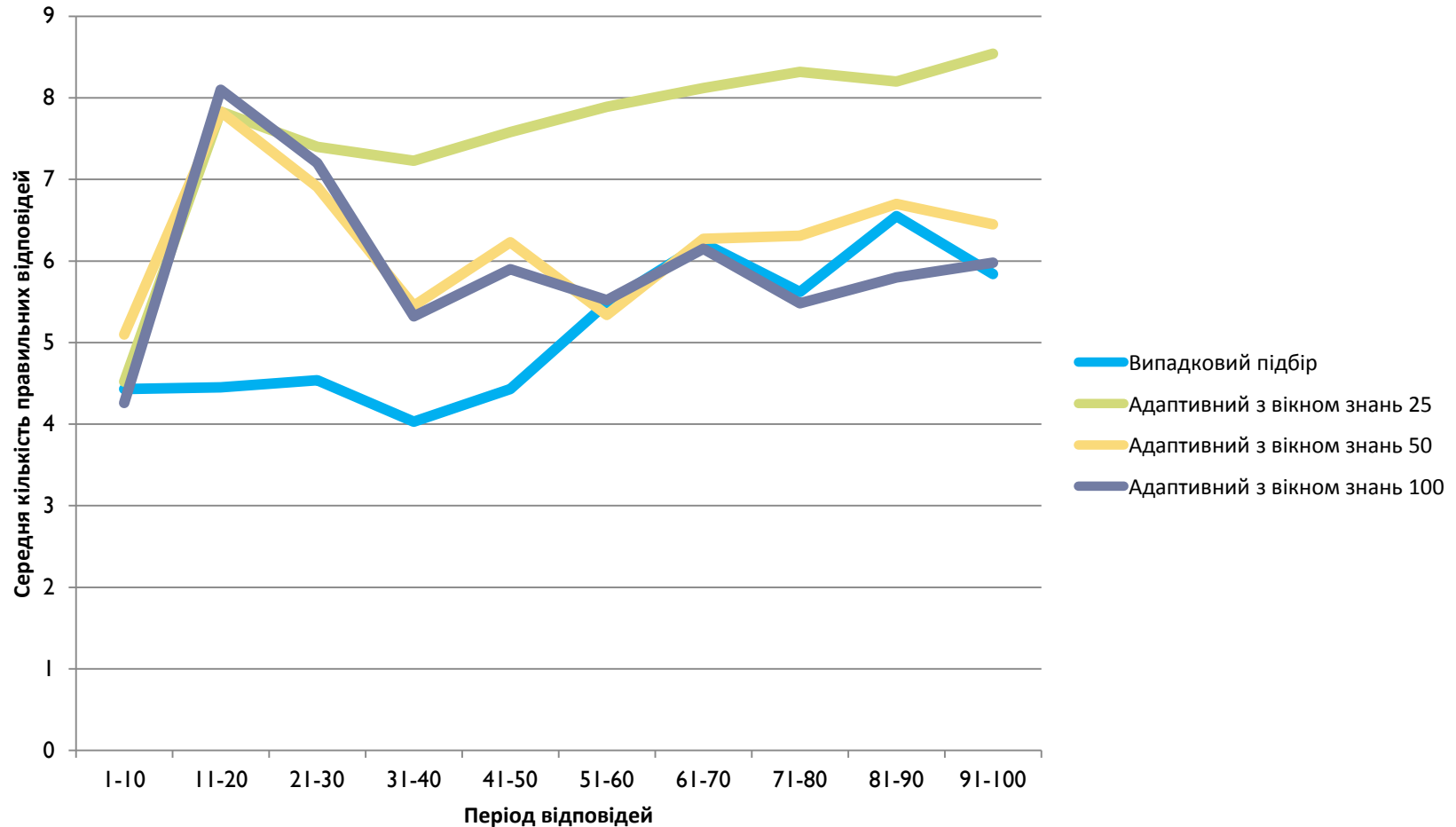
Алгоритм підбору питань

В загальному вигляді під час кожної користувацької сесії алгоритм дій виглядає так:

1. Формуємо поточний стан моделі користувача на основі записів його відповідей і відповідного заданого вікна знань.
2. Формуємо пул питань на основі вибраного користувачем курсу та виключаємо питання, що містяться в його моделі.
3. Вибираємо питання з найбільшим значенням `deltaProfit`. Видаляємо задане питання з доступного пулу питань.

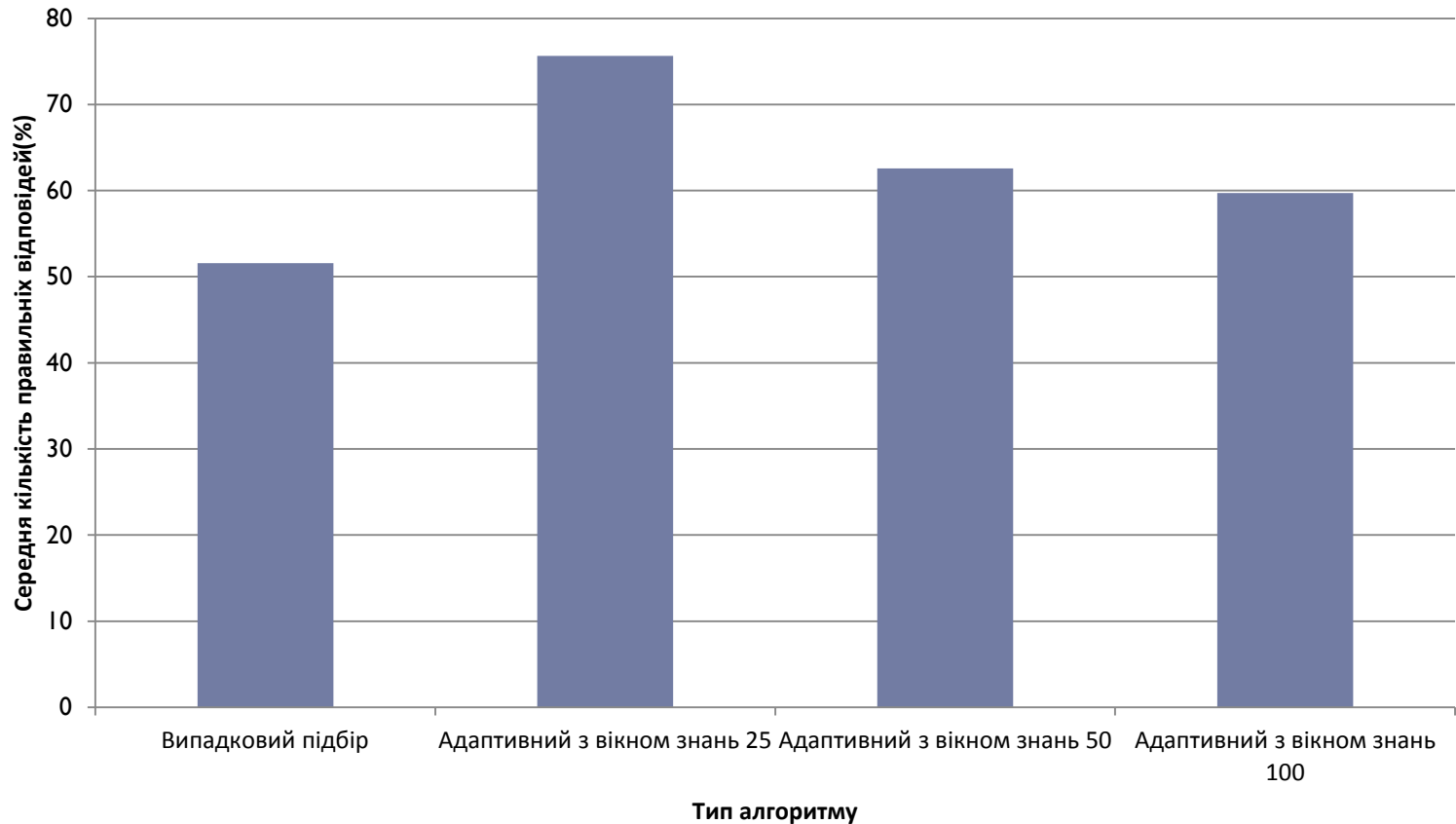


Результати роботи



Графік 3. Залежність середньої кількості правильних відповідей від періоду відповідей

Результати роботи



Діаграма 1. Середня кількість правильних відповідей

Переваги побудованої моделі

- ▶ втручання експерта мінімальне;
- ▶ легке масштабування (немає необхідності перебудови всієї структури кластерів при додаванні нових питань);
- ▶ алгоритм вибору питань не залежить від даних про результати інших користувачів;
- ▶ можливе використання попередніх даних про результати користувача при впровадженні моделі в систему навчання;
- ▶ швидкість роботи.

Наукові досягнення

- ▶ Діденко Д.Г., Крюков О.О. Модель адаптивного навчання на базі алгоритму кластери-зації категорійних даних CLOPE. // 18-а Міжнародна науково-технічна конференція «Системний аналіз та інформаційні технології (САІТ-2016)», м. Київ. - 2016. - С.346-347;
- ▶ Крюков О.О., Козлов К.В., Діденко Д.Г. Використання алгоритму CLOPE для структурування доменної моделі адаптивної системи навчання, електронний науковий журнал ІПСА «Системні науки та кібернетика».

Дякую за увагу