

АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ ПОВТОРЮВАНИХ ЗАПИТАНЬ ПРИ ОРГАНІЗАЦІЇ ІНТЕЛЕКТУАЛЬНИХ ІГОР

Вінницький національний технічний університет

Анотація

У даній роботі розглядаються методи виявлення повторюваних запитань при організації інтелектуальних ігор, проводиться їх порівняльний аналіз.

Ключові слова: інтелектуальна гра, організація, інформація, плагіат, виявлення плагіату.

Abstract

This article observes methods of duplicate questions detection on intellectual games organization and provides analysis of these methods.

Keywords: intellectual game, organization, information, plagiarism, plagiarism detection.

Вступ

Інтелектуальна гра «Що? Де? Коли?» (рос. «Что? Где? Когда?») була започаткована у 1975 році у телевізійному форматі [1]. Популярність телегри призвела до того, що у 1989 році з'явилась «спортивна» версія з дещо відмінним форматом [2]. Телевізійна версія гри полягає у тому, що «команда знавців» у складі шести гравців змагається проти «команди телеглядачів» – авторів запитань – до шести очок. У «спортивній» версії змагаються декілька команд, намагаючись відповісти на якомога більшу кількість запитань. Інтелектуальна гра «Що? Де? Коли?» набула неабиякої популярності у багатьох країнах світу – найбільше на території колишнього Радянського Союзу (рисунок 1) – про що свідчить статистика: на сайті міжнародного рейтингу «Що? Де? Коли?» зареєстровано 34 тисячі команд [3] та більше 148 тисяч гравців [4].

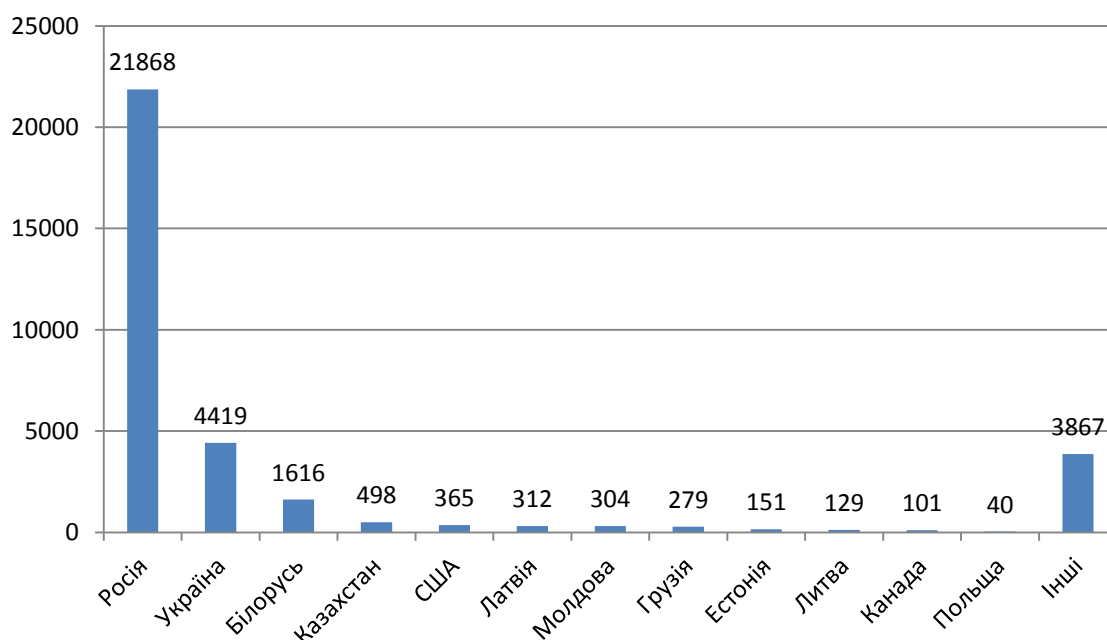


Рисунок 1 – Розподіл кількості команд за країнами

При проведенні будь-яких змагань – не лише інтелектуальних – особливо важливим є забезпечення об'єктивності результатів. Основним критерієм об'єктивності є забезпечення рівності

умов між гравцями та командами. Організатори зобов'язані мінімізувати ризик необ'єктивності результатів гри – наприклад, виключити можливість використання гравцями зовнішніх інформаційних ресурсів (Інтернет), забезпечити відсутність підказок, унеможливити підслуховування обговорення іншої команди тощо. Одним з факторів, що здатні внести дисбаланс у гру, є наявність повторюваних, або «засвічених» запитань («свічок») – запитань, на які частина команд або гравців можуть відповісти з тієї причини, що вже зустрічали таке запитання під час іншої гри. Таке явище може трапитись у різних випадках: для гри використовується пакет запитань, який розігрувався на іншій грі, у якій брали участь деякі учасники поточної гри; автор запитання надіслав його декільком редакторам для різних пакетів запитань; пакет запитань для поточної гри формувався організаторами і являє собою компіляцію різних пакетів запитань, деякі з яких розігрувалися на інших іграх за участю деяких учасників поточної гри; гравець чи декілька гравців є авторами чи тестувальниками наявного пакету запитань. Таким чином, актуальною задачею є розробка методу виявлення повторюваних запитань, що дозволить підвищити об'єктивність результатів гри.

Методи аналізу на виявлення «засвічених» запитань

Задача аналізу на виявлення «свічок», по суті, є одним з випадків задачі виявлення плагіату, тому у якості методів аналізу на «засвічені» запитання розглянемо методи виявлення плагіату.

Вирізняють наступні форми плагіату [5]:

1. Сору & paste плагіат – використання частини тексту джерела без внесення змін.
2. Замаскований плагіат – використання частини тексту джерела з внесенням маскуючих змін.
3. Переказ – переформулювання тексту джерела.
4. Переклад – плагіат іншомовного джерела.
5. Плагіат ідей – самостійна розробка тексту, що базується на неоригінальній ідеї.

«Засвічені» запитання можна віднести до форми плагіату сору & paste.

Існують такі методи виявлення плагіату (рисунок 2):

1. Дактилоскопія – метод що полягає у виборі з ряду документів набору підрядків – «відбитків». Текст, що досліджується, порівнюється з цими відбитками.
2. Перевірка дослівним перекриттям – просте порівняння рядків.
3. Аналіз «множини слів» – текст представляється, як невпорядкований набір слів, за яким і визначається схожість.
4. Цитування – визначає спільні цитати. Шаблон цитат являє собою підпоследовності, що містять не лише спільні цитати для двох документів, але і подібний порядок і наближеність цитат в тексті.
5. Стилometрія – статистичний метод для виявлення авторства анонімних документів і для комп'ютерної перевірки на плагіат.

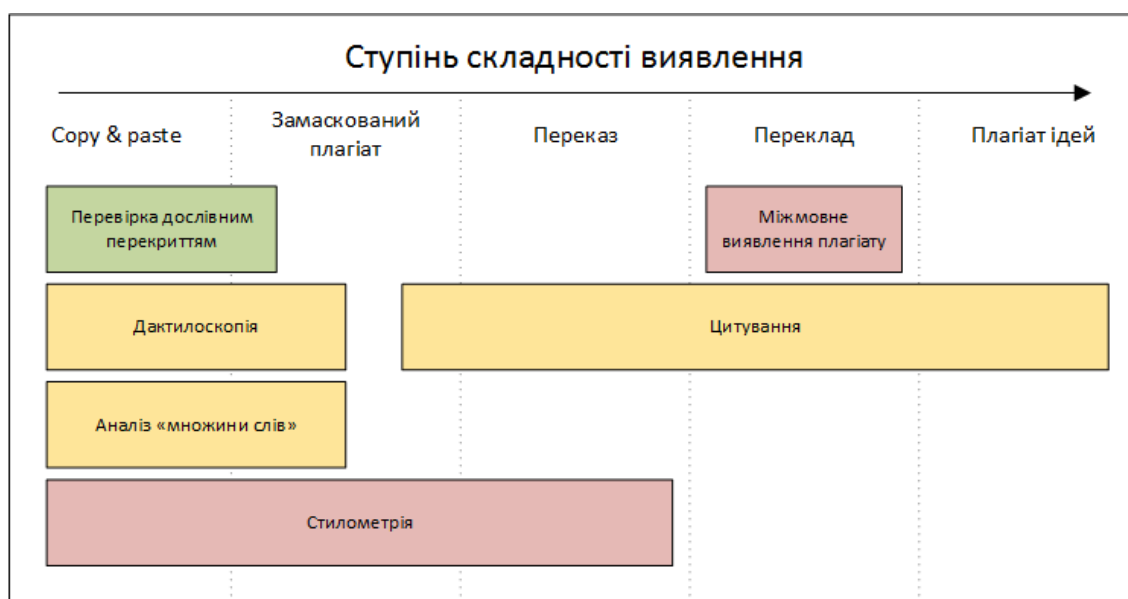


Рисунок 2 – Способи виявлення плагіату залежно від форми

Розглянемо перспективи застосування вищенаведених методів для виявлення повторюваних запитань.

Метод перевірки дослівним перекриттям рядків характеризується високим ступенем надійності. Такий його недолік, як низька швидкодія при аналізі великої кількості інформації, є несуттєвим при виконанні перевірки на повторюваність запитань, оскільки розмір вхідних даних є незначним. Іншим його недоліком є неспроможність виявляти плагіат ідей.

Дактилоскопія є модифікацією методу перевірки дослівним перекриттям рядків, спрямованою на покращення швидкодії при аналізі великих обсягів даних. Утім, при пошуку повторюваних запитань розмір вхідних даних недостатній для застосування цього методу.

Аналіз «множини слів», на відміну від дактилоскопії, може бути застосований для виявлення «свічок», проте він є менш точним, ніж метод перевірки дослівним перекриттям рядків.

Метод цитування потенційно здатен забезпечити виявлення плагіату ідей, проте його практичному застосуванню для виявлення «засвічених» запитань заважає малий розмір вхідних даних.

Стилометрія – метод, що застосовується для встановлення авторства тексту, а його ефективність при пошуку плагіату ідей є вкрай низькою.

Висновки

Внаслідок проведеного аналізу існуючих методів виявлення плагіату було встановлено, що серед розглянутих методів найефективнішим для застосування у задачі аналізу виявлення повторюваних запитань є метод перевірки дослівним перекриттям. Утім, для досягнення абсолютної об'єктивності при проведенні інтелектуальних ігор необхідно розробити новий метод аналізу, що виявлятиме такий вид плагіату, як плагіат ідей.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Что? Где? Когда?: материал из Википедии – свободной энциклопедии: [Електронний ресурс]. – Режим доступу: https://ru.wikipedia.org/wiki/Что%3F_Где%3F_Когда%3F
2. Что? Где? Когда? (спортивная версия): материал из Википедии – свободной энциклопедии: [Електронний ресурс]. – Режим доступу: [https://ru.wikipedia.org/wiki/Что%3F_Где%3F_Когда%3F_\(спортивная_версия\)](https://ru.wikipedia.org/wiki/Что%3F_Где%3F_Когда%3F_(спортивная_версия))
3. Команды – Спортивное «Что? Где? Когда?» Официальный рейтинг МАК: [Електронний ресурс]. – Режим доступу: <http://rating.chgk.info/teams.php?release=1328>
4. Игроки – Спортивное «Что? Где? Когда?» Официальный рейтинг МАК: [Електронний ресурс]. – Режим доступу: <http://rating.chgk.info/players.php?release=1328>
5. Plagiarism detection – Wikipedia: [Електронний ресурс]. – Режим доступу: https://en.wikipedia.org/wiki/Plagiarism_detection

Кавка Олексій Олександрович – студент групи ІПІ-14б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: oleksi.life@gmail.com

Романюк Оксана Володимирівна – канд. техн. наук, доцент кафедри програмного забезпечення, Вінницький національний технічний університет, Вінниця, e-mail: romaniukoksnav@gmail.com

Oleksii O. Kavka – Student of Department of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: oleksi.life@gmail.com

Oksana V. Romaniuk – Cand. Sc. (Eng), Assistant Professor of the Software Chair, Vinnytsia National Technical University, Vinnytsia, e-mail: romaniukoksnav@gmail.com