

УДК 004.775:625.75

Валько Н. В., Запорожченко А. В.

Херсонський державний університет, Херсон, Україна

**ПРОЕКТУВАННЯ СИСТЕМИ АВТОМАТИЧНОГО РОЗПОДІЛУ РЕЦЕНЗУВАННЯ
НА ОСНОВІ АЛГОРИТМУ ГЕЙЛА-ШЕПЛІ**

DOI: 10.14308/ite000737

У статті наведено приклад реалізації алгоритму Гейла-Шеплі у системі автоматичного розподілу статей для рецензування. Розглянуто алгоритм, його властивості та особливості застосування. Зазначено вимоги та представлено структуру бази даних, що складається з таблиць «articles», «reviewers», «solution_reviewers», «keywords», «reviewer_keywords», «article_keywords». Звернено увагу на фреймворки, середовище розробки та сервери. Описано функціонал інформаційної системи, протестований на реальних даних наукового видання. Деталізовано функції, пов'язані з додаванням, редагуванням і видаленням статей, роботою з ключовими словами та рецензентами. Встановлено, що результати розподілу будуть відрізнятися в залежності від сторони ініціації («рецензент» чи «тема статті»). Також встановлено, що розподіл буде більш вигідним для сторони ініціації. У даному випадку це забезпечить вибір найбільш компетентного рецензента для публікації. Розроблено програмний продукт на основі цього алгоритму. Він співвідносить за ключовими словами варіанти вибору рецензентів та надає пріоритетність їх призначення. Також є можливість ручного корегування зробленого розподілу. Недоліком цієї системи є те, що вона розрахована на масовість та великі об'єми вхідної інформації. Перевагою є те, що процес призначення рецензента за великої кількості вхідних повідомлень можна частково автоматизувати. Продемонстровано, як модель реалізації алгоритму Гейла-Шеплі для задачі розподілу статей по рецензентах покладена в основу системи, що оптимізує роботу видання, пришвидшує і робить більш ефективними окремі процеси, зокрема процес рецензування статей із урахуванням ключових параметрів.

Ключові слова: алгоритм Гейла-Шеплі, система розподілу, рецензування публікацій, база даних

Вступ

Постановка проблеми. Від часу винайдення алгоритму Гейла-Шеплі він і його модифікації знайшли своє застосування у багатьох сферах людської діяльності. На сьогодні для успішного функціонування різного роду фірм, організацій та підприємств необхідно мати розвинуту інформаційну систему, яка дозволяє автоматизувати збір і обробку даних. Збільшення попиту на публікації в рецензованих збірниках викликає потребу у створенні системи, яка б автоматично пропонувала рецензентів до кожної статті. Для цієї задачі може бути використано вказаний алгоритм, який буде допомагати розподілити статті серед рецензентів на основі модифікації алгоритму Гейла-Шеплі. Це допоможе збільшити швидкість розподілення статей серед робітників видання, ураховуючи зайнятість рецензента на момент запуску системи, його спеціальність та інші показники.

Аналіз останніх досліджень і публікацій. Алгоритм уперше було представлено в роботі [1], як набір правил, які приводять до утворення стабільних пар. Проте, у залежності від порядку обробки множин, результат буде найкращим для однієї і не найкращим для іншої. Виявлені недоліки методу привели до появи різновидів цього алгоритму. Наприклад, у роботі [2] показано, що не потрібно робити всі порівняння. У



Валько Н. В., Запорожченко А. В.

роботах [3, 4, 5, 6] показано, як утворюються пари, якщо не вся інформація достовірна, кількість елементів двох множин не однакова, або існує квота при відношенні «багато до одного».

Мета і завдання статті. Метою статті є розробка моделі системи розподілу за алгоритмом Гейла-Шеплі та створення системи розподілу статей рецензентам у збірнику журналу. Для досягнення цієї мети були поставлені такі завдання: зробити постановку задачі і визначити вимоги до моделі розподілу; побудувати модель роботи системи на базі алгоритму Гейла-Шеплі, розробити систему, яка використовує одну з модифікацій алгоритму.

Результати дослідження

На початку роботи нашого алгоритму ми маємо список усіх об'єктів статей з їх характеристиками та список усіх існуючих об'єктів рецензентів, які мають свою квоту, тобто кількість максимально можливих оброблюваних статей і списки із ключовими словами, мовою, які виконують роль пріоритету. Дані списки – це вподобання рецензентів. У реалізованому нами алгоритму списки «відсікають» ті статті, які не мають нічого спільного з рецензентом та слугують пріоритетом для відбору статей для рецензування.

Перед початком роботи алгоритму ми маємо список статей, яким у процесі роботи алгоритму будуть пропонуватися рецензенти. Водночас ми маємо список об'єктів рецензентів, які будуть як можна точніше підбиратися для статей.

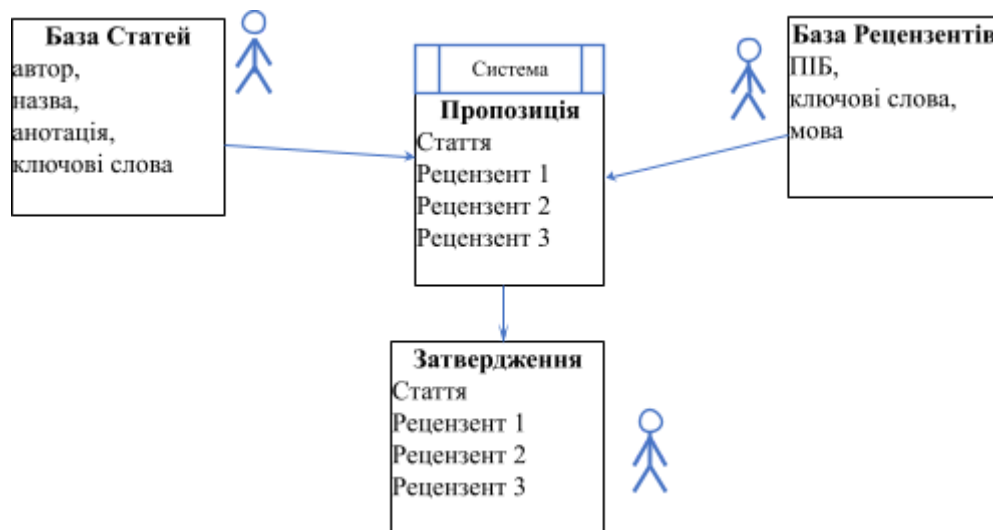


Рис. 1. Схема роботи системи

Побудова моделі роботи системи. Нехай існує дві кінцевих непересічних множини рецензентів і статей.

Нехай $A = \{a_1, a_2, \dots, a_n\}$ – множина статей, а $C = \{C_1, C_2, \dots, C_m\}$ – множина рецензентів.

У рецензентів є переваги серед статей – список $P(C_j)$ з елементів множини $A \supset \{C_j\}$. Кожна стаття може бути зарахована лише до одного рецензента, але кожен рецензент може прийняти кілька статей. Нехай $q_j > 0$ позначає максимальну кількість (квоту) статей, яку може обробити рецензент C_j .

У цій моделі переваги рецензентів на групах статей повністю визначаються списком властивостей, таких як мова статті, галузь знань рецензента, кількість статей на поточному рецензуванні тощо. Щоб одна й та ж стаття не потрапила на розгляд двічі до рецензента,

визначимо умови: для будь-якої підмножини статей $S \subset A$, що складається з не більше ніж $q_j - 1$ стаття, і для будь-яких двох статей $a_1, a_2 \notin S$,

$$S \supset \{a_1\} >_{cj} S \supset \{a_2\} \Leftrightarrow \{a_1\} >_{cj} \{a_2\}, \quad S \supset \{a_1\} >_{cj} S \supset \{C_j\} \Leftrightarrow \{a_1\} >_{cj} \{C_j\}.$$

За наявності місця рецензент C_j завжди вважатиме за краще взяти статтю, яку він вважає за кращу в індивідуальному порядку, незалежно від того, які статті потрапили в C_j .

Ключове поняття – стабільність розподілу. Розподіл статей по рецензентах стабільний, якщо не існує блокуючого індивідуального агента або пари, що складається з рецензента і групи статей. Існує стабільний розподіл статей по рецензентах, який може бути знайдено конструктивно, у результаті роботи модифікованого алгоритму Гейла-Шеплі. У версії алгоритму, у якій пропозиції роблять статті, єдина відмінність полягає в тому, що рецензенти можуть тимчасово приймати більше однієї пропозиції, але в межах квоти. А у версії алгоритму, в якій пропозиції роблять рецензенти, останні можуть посилати кілька пропозицій одночасно, але в межах квоти.

На першому етапі роботи системи ми створюємо список, де ключ – це ідентифікаційний номер статті, а значення – список із усіх відповідних рецензентів, які мають у списку своїх переваг мову і ключові слова статті, незважаючи на якому пріоритетному місці знаходиться мова і кількість ключових слів статті для рецензента.

На другому етапі роботи алгоритму ми отримуємо створений на першому кроці список статей із усіма можливими відповідними до умов рецензентів. Тобто до кожної статті був створений набір із рецензентів, які можуть рецензувати статтю, розуміючи мову статті і маючи вказану деяку кількість ключових слів статті. Далі, згідно з послідовністю пріоритетів для призначення рецензента, а саме мова, ключові слова, кількість робіт, отриманий список сортуємо за мовою у порядку спадання. Тобто для кожної статті створюємо список із рецензентів у порядку спадання за мовою, де на першому місці буде знаходитися рецензент, у якого мова статті на більш пріоритетному місці від інших. Після цього ми усікаємо наш список рецензентів для статті до трьох – у такий спосіб отримуємо трьох кандидатів (якщо їх більше трьох) у порядку спадання за пріоритетністю мови, які водночас у собі мають ключові слова даної статті.

Оскільки на попередньому етапі був отриманий список із людьми, у яких мова статті на найбільш пріоритетному місці стосовно інших і при цьому включають у своїх вподобаннях стосовно ключових слів ключові слова статті, сортуємо наш масив даних за кількістю цих слів. Як результат, на виході отримуємо для кожної статті у порядку найбільшого вподобання трьох рецензентів.

Відповідно до предметної області система будується з урахуванням таких особливостей:

- Стаття написана одним автором.
- Стаття рецензується трьома рецензентами.
- Рецензент може рецензувати декілька статей.

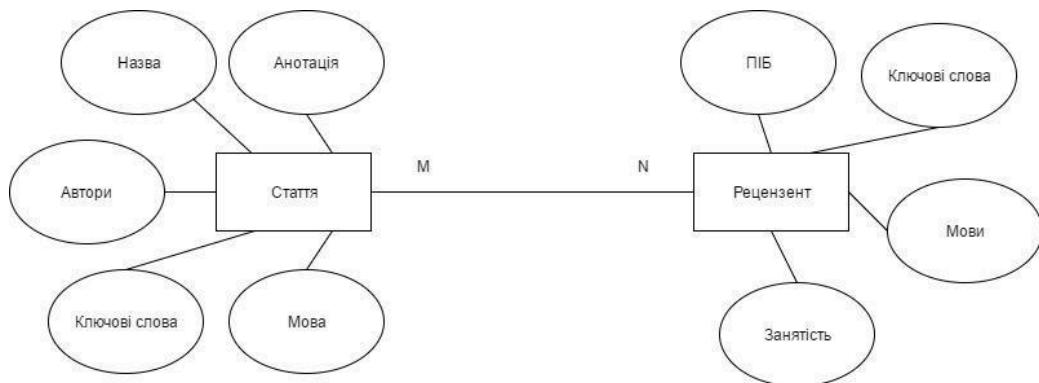


Рис. 2. ER-діаграма системи рецензування

Систематизація вимог до бази даних. Сформулюємо вимоги до бази даних цього проекту. Повинен бути спосіб:

- зберігати інформацію про статті;
- зберігати інформацію про рецензентів;
- зберігати інформацію про затверджені статті до рецензентів;
- додавати, редагувати, видаляти статтю;
- додавати, редагувати, видаляти рецензента.

Таблиця 1.

Словник даних про статтю

Елемент даних	Опис
Автор	ПІБ автора статті
Назва	Назва статті
Анотація	Анотація статті
Ключові слова	Ключові слова статті
Мова	Мова статті

Таблиця 2.

Словник даних про рецензента

Елемент даних	Опис
ПІБ	ПІБ рецензента
Ключові слова	Список ключових слів
Мова	Список мов
Зайнятість	Кількість оброблюваних статей

База даних повинна бути спроектована так, щоб запити оброблялися якомога швидше. Досягти цього можна за рахунок використання реляційної моделі зв'язків, тобто поєднання декількох таблиць через зовнішній ключ.

Для оптимальної роботи системи було вирішено побудувати таку базу даних, яка міститиме таблиці: `articles`, `reviewers`, `solution_reviewers`, `keywords`, `article_keywords`, `reviewer_keywords`.

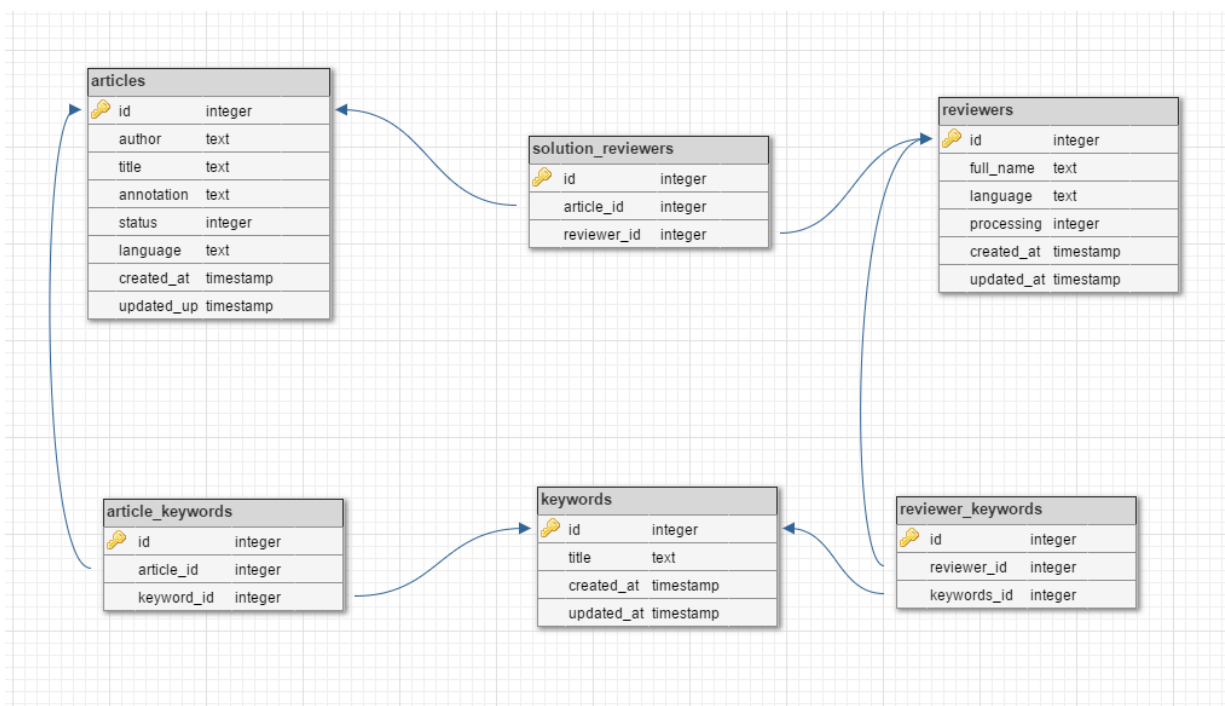


Рис. 3. Структура бази даних

У таблиці «articles» зберігаються статті. Кожна стаття має ідентифікаційний номер, який являється первинним ключем, ПІБ авторів, назву, анотацію, мову, статус, який відповідає за рецензованість статті, дату додавання і останнього редагування.

Таблиця «reviewers» містить інформацію про рецензентів. Вона має такі поля, як ідентифікаційний номер (первинний ключ), ПІБ рецензента, список мов за пріоритетністю, кількість оброблюваних статей на даний момент часу, дату додавання і дату останнього редагування.

Таблиця «solution_reviewers» слугує для утворення зв'язків. Вона складається із ідентифікаційного номеру, а також зовнішнього ключа article_id, який зв'язаний із первинним ключем таблиці «articles», і зовнішнього ключа reviewer_id – зв'язок із таблицею «reviewers».

Для зберігання ключових слів слугує таблиця «keywords», яка має ідентифікаційний номер, назву ключового слова (або строчки), дату додавання і редагування.

Рецензенти пов'язані з ключовими словами за допомогою таблиці «reviewer_keywords», у якій знаходяться ідентифікаційні номери ключових слів у порядку пріоритетності їх для рецензента.

У свою чергу було вирішено створити таблицю «article_keywords» для зв'язку статті із ключовими словами, які зберігаються у порядку пріоритетності для статті.

Для покращення роботи над системою був обраний фреймворк Laravel, тому що він має:

- Bundles (Модулі) – є репозиторій з великою кількістю модулів.
- The Eloquent ORM – ActiveRecordORM, вмє будувати зв'язки (many to many, one to many, one to one).
- Migrations – правила хорошого тону.
- Фабрики – допоможуть заповнити базу даних тестовими даними.

У якості середовища розробки було обрано NuSphere PhpED. NuSphere PhpED – це програмний продукт для формування додатків із застосуванням можливостей баз даних і кодів класу PHP, HTML, XML, CSS. Мовою програмування обрали PHP. Для графічного оформлення була обрана зв'язка HTML/CSS. Однак існують спеціальні CSS-фреймворки

для поліпшення роботи web-дизайнера. Для розробки зовнішнього вигляду нашого мікро сервісу ми надали перевагу фреймворку Bootstrap. Створений як набір інструментів для створення сайтів та web-додатків компанією Twitter, Bootstrap зайняв високу позицію серед розробників завдяки своїй адаптивності, простоті та сучасності, адже Bootstrap використовує найновіші розробки в області HTML и CSS. Дизайн повинен бути сучасним, у той же час зрозумілим і функціональним. За допомогою Bootstrap можна без проблем втілити це в життя, використавши один із готових шаблонів, чи написавши власний.

Серед списку серверів – Denwer, Xampp, Vertrigo, OpenServer – нами був обраний останній, адже OpenServer – це портативний локальний WAMP / WNMP сервер, який має багатфункціональну керуючу програму і великий вибір компонентів, що підключаються.

Систему перевірено на основі справжніх даних, а саме статей одного з журналів Херсонського державного університету, які знаходяться за адресою «<http://ite.kspu.edu>».

Функціонал інформаційної системи. Список усіх наявних статей у журналі можна побачити, натиснувши в меню «Усі статті». За замовчування в системі обрана ця вкладка.

●Додавання статті. Потрапити у форму додавання статті можливо, натиснувши в меню пункт «Додати статтю». Форма додавання представлена на рисунку 4.

Автор:
Тарас Шевченко

Назва статті:
Цікаві факти

Анотація:
1. Тарас Шевченко протягом життя був «душею компанії», любив сміятися і вмів розвеселити навіть незнайомих людей.
2. Основна освіта Тараса Шевченка тривала лише два роки. Її він здобув у церковно-приходській школі.
3. Тарас Шевченко завжди з любов'ю згадував той випадок, коли малим пішов шукати «залізні стовпи», що «підпирають» небо, і заблукав у полі. Чумаки, зустрівши хлопця, забрали його з собою і ввечері привезли до Кирилівки

Ключові слова:
* односторонні функції * узагальнені матриці Галуа * Марія Пентиліук

Мова:
Українська

Додати

Рис. 4. Вікно додавання статті

●Введення ключових слів відбувається так: натиснути поле «ключові слова»; у випадяючому списку відобразяться всі ключові слова системи, виберіть потрібний (за необхідністю декілька). Реалізовано пошук ключових слів у списку опцій – для цього натисніть на поле ключових слів і почніть вводити слово (рис.5).

Ключові слова:

від

- відношення ізоморфізму
- функції від матриць
- відео лекція
- відкрита освіта
- мережні технології відкритих систем
- відставання у навчанні

Рис. 5. Пошук ключових слів

Після введення даних натиснути кнопку «Додати».

Про вдале додавання сповістить повідомлення зі змістом «Стаття додана».

●Редагування статті. Потрапити у форму редагування можливо зі сторінки «Усі статті», натиснувши кнопку «Редагувати» навпроти необхідної статті. Редагування статті

має ідентичний інтерфейс до додавання з тією лише різницею, що замість кнопки «Додати» – «Редагувати». Про вдале редагування сповістить відповідне повідомлення «Стаття оновлена».

●Видалення статті. Щоб видалити статтю, необхідно натиснути навпроти необхідної статті кнопку «Видалити» на сторінці «Усі статті». Про вдале видалення сповістить повідомлення «Стаття видалена».

●Рецензенти. Список усіх наявних рецензентів у журналі можна побачити натиснувши в меню «Усі рецензенти». Потрапити у форму додавання рецензента можливо, натиснувши в меню пункт «Додати рецензента». Після введення даних натиснути кнопку «Додати». Про вдале додавання сповістить повідомлення зі змістом «Рецензента додано» (рис.6).

ПІБ:

Іванов Іван Іванович

Ключові слова:

✕ односторонні функції прото

алгебраїчна атака на протокол обміну ключами шифрування.

протокол обміну ключами

Додати

Рис. 6. Додавання рецензента

Щоб запустити алгоритм розподілу статей по рецензентах, необхідно перейти на вкладку «Робота системи». На цій сторінці будуть відображені всі статті, які на даний момент часу не рецензовані (рис.7).

Статті для рецензування			Запустити
☐ Назва	Автор	Мова	
☐ Назва статті 1	Автор 1	Англійська	
☒ Назва статті 2	Автор 2	Українська	
☐ Назва статті 3	Автор 3	Українська	
☒ Назва статті 4	Автор 4	Українська	
☒ Назва статті 5	Автор 5	Українська	
☐ Назва статті 6	Автор 6	Українська	

Рис. 7. Обрання статей для роботи системи

Необхідно помітити необхідні статті і натиснути кнопку «Запустити» в правому верхньому кутку.

Після завершення роботи алгоритму система видає пропонувані варіанти вибору рецензента для статті, залишаючи можливість остаточного затвердження адміністратору із запропонованих кандидатів (рис.8).

Пропонований результат

Підтвердити

Назва статті	Автор	Рецензент
Назва статті 1	Автор 1	Прізвище, ім'я
Назва статті 2	Автор 2	Прізвище, ім'я
Назва статті 3	Автор 3	Прізвище, ім'я Прізвище, ім'я Прізвище, ім'я

Рис. 8. Пропонований результат

Підтвердивши свій вибір, система помітить статті, яким був призначений рецензент, збільшить кількість робіт у рецензента і повідомить про успішність повідомленням «Рецензенти затверджені!», перенаправивши на сторінку «Робота системи».

Для перегляду рецензованих статей необхідно в боковій панелі меню натиснути пункт «Результати». На сторінці «Результати» відображається список із назви статті, її автора і рецензента, який був затверджений за статтею. ПІБ рецензента є посиланням, натиснувши на яке можна побачити всі статті, затверджені за ним (рис.9).

Результат

Назва статті	Автор	Рецензент
Назва статті 1	Автор 1	Рецензент 1
Назва статті 2	Автор 2	Рецензент 2
Назва статті 3	Автор 3	Рецензент 3

Рис. 9. Статті, затверджені за рецензентами

Висновок

У процесі дослідження побудовано модель реалізації алгоритму Гейла-Шеплі для задачі розподілу статей по рецензентах журналу збірника. За цією моделлю була спроектована і створена система розподілу статей по рецензентах журналу збірника, яка передбачає такі пункти, як: операції додавання/редагування/видалення зі статтями, рецензентами, призначення рецензентів. Використання такої системи дасть можливість збільшити швидкість обробки інформації про статті та розподіл їх серед рецензентів видання, з урахуванням розподілу параметрів, таких як зайнятість, спеціальність, мова.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 9–15.
2. Roth, Alvin E. & Peranson, Elliott (1990). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89, 748–780.
3. Dubins, L. E., & Freedman, D. A. (1981). Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly*, 88(7), 485. doi:10.2307/2321753
4. Roth, Alvin E. (2003). The Origins, History, & Design of the Resident Match. *Journal of American Medical Association*, (289), 909–912.
5. Wallis, C. Bradley, Kannan P. Samy, Alvin E. Roth, & Michael A. Rees (2011). Kidney Paired Donation. *Nephrology Dialysis Transplantation*, 26(7), 2091–2099.

6. Abdulkadiroglu, Atila & Tayfun Sonmez (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93, 729–747.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1), 9–15.
2. Roth, Alvin E. & Peranson, Elliott (1990). The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design. *American Economic Review*, 89, 748–780.
3. Dubins, L. E., & Freedman, D. A. (1981). Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly*, 88(7), 485. doi:10.2307/2321753
4. Roth, Alvin E. (2003). The Origins, History, & Design of the Resident Match. *Journal of American Medical Association*, (289), 909–912.
5. Wallis, C. Bradley, Kannan P. Samy, Alvin E. Roth, & Michael A. Rees (2011). Kidney Paired Donation. *Nephrology Dialysis Transplantation*, 26(7), 2091–2099
6. Abdulkadiroglu, Atila & Tayfun Sonmez (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93, 729–747.

Nataliia Valko, Artem Zaporozhchenko
Kherson State University, Kherson, Ukraine

DESIGN AND REALIZATION OF AUTOMATIC DISTRIBUTION REVIEW SYSTEMS BASED ON GAIL-SHAPLE ALGORITHM

The article provides an example of implementing the Gale-Shapley algorithm in an automatic article distribution system for peer review. The algorithm, its properties, and application specifics are discussed. The requirements are specified and the database structure is presented, consisting of the tables "articles", "reviewers", "solution_reviewers", "keywords", "reviewer_keywords", "article_keywords". Attention is paid to frameworks, development environments and servers. The functionality of the information system, tested on real data from a scientific publication, is described. Functions related to adding, editing, and deleting articles, working with keywords, and reviewers are detailed. It is noted that the distribution results will vary depending on the initiating party ("reviewer" or "article topic"). It is also established that the distribution will be more advantageous for the initiating party, in this case, ensuring the selection of the most competent reviewer for publication. Additionally, a software product based on this algorithm has been developed. It associates options for selecting reviewers with keywords and prioritizes their assignment. Manual adjustments to the distribution are also possible. A drawback of this system is its focus on mass and large volumes of input information. An advantage is that the reviewer assignment process can be partially automated, especially with a large number of incoming messages. It is demonstrated how the implementation model of the Gale-Shapley algorithm for the problem of distributing articles by reviewers is the basis of a system that optimizes the work of the publication, speeds up and makes individual processes more efficient, in particular the process of reviewing articles taking into account key parameters.

Key words: gale-shapley algorithm, distribution system, publication review, database

Стаття надійшла до редакції 29.01.2021
The article was received 29 January 2021