

УДК 004.89;004.94

DOI: <https://doi.org/10.30837/ITSSI.2021.16.013>

Т. М. БАТЮК, В. А. ВИСОЦЬКА

РОЗРОБКА ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ПІДТРИМКИ СОЦІАЛІЗАЦІЇ КОРИСТУВАЧА ЗА ПОДІБНІСТЮ ІНТЕРЕСІВ

Об'єктом дослідження є процес соціалізації особистостей, оскільки в наш час завдання соціалізації є дуже важливим і всі сучасні соціальні мережі намагаються максимально оптимізувати та автоматизувати соціалізацію різноманітних користувачів з використанням популярних сучасних ІТ, таких як нейронні мережі та алгоритми аналізу користувацького контенту. **Предметом** дослідження є методи та технологічні засоби пошуку та формування списку релевантних користувачів за подібністю інтересів для здійснення соціалізації. Відповідно здійснюється аналіз профіля користувача системи, а саме визначення справжності користувача на основі пошуку людського обличчя на користувацьких фотографіях з використанням нейронних мереж і аналізу користувацького контенту через алгоритми нечіткого пошуку та моделі Noisy Channel. **Мета** роботи – створення інтелектуальної системи для соціалізації особистостей на основі нечіткого пошуку за ключовими словами, моделі Noisy Channel, алгоритму ефективного розподілу текстового контенту та згорткової нейронної мережі. В статті вирішуються наступні **завдання**: проаналізувати сучасні та найвідоміші підходи, методи, засоби та алгоритми розв'язування задач соціалізації особистостей за подібністю інтересів; розробка загальної структури типової інтелектуальної системи соціалізації особистостей за спільними інтересами; сформулювати функціональні вимоги до основних модулів структури типової інтелектуальної системи соціалізації особистостей за спільними інтересами; розробити інтелектуальну систему підтримки соціалізації користувача за подібністю інтересів на основі нейронних мереж, нечіткого пошуку та моделі Noisy Channel та провести експериментальну апробацію. Використовуються такі **методи**: метод Левенштейна; алгоритм N-грам; модель Noisy Channel; нечіткий пошук. Отримано наступні **результати**: побудовано та описано загальну структуру типової інтелектуальної системи соціалізації особистостей за спільними інтересами. До основного призначення створення системи в першу чергу варто віднести створення нового алгоритму аналізу користувацького контенту та пошуку найбільш релевантних за інтересами користувачів на основі таких алгоритмів як N-грам, розширення вибірки, Левенштейна, та моделі Noisy Channel. Також зазнав подальшого розвитку шаблон асинхронного створення програмного продукту, що дозволяє створити практично повністю динамічну систему. Варто удосконалити згорткову нейронну мережу, що дозволить ефективно та динамічно здійснювати пошук людських обличчя на фотографії, та перевіряти наявність вже існуючих людей в базі даних системи. **Висновки**: виявлено, що реалізований алгоритм здійснює вибірку приблизно в 10 разів швидше, ніж звичайний алгоритм Левенштейна. Також реалізований в системі алгоритм для формування вибірки користувачів є ефективнішим та точнішим приблизно на 25-30% в порівнянні зі звичайним алгоритмом Левенштейна.

Ключові слова: відстань Левенштейна; алгоритм N-грам; модель Noisy Channel; нечіткий пошук.

Вступ

Розроблення та впровадження інтелектуальної системи (ІС) соціалізації особистостей є актуальною та перспективною задачею, так як сьогодні користувачі соціальних мереж намагаються оптимізувати життєві процеси для економії часу і відповідно, для більш доцільної/ефективної його реалізації, зокрема, для автоматизації пошуку інших подібних до власних інтересів користувачів соціальних мереж. Така ІС вирішує дві важливі проблеми теперішнього часу інформатизації та глобалізації: автоматизація процесу соціалізації користувачів та максимально оптимізує економію часу для організації буденних потреб у спілкуванні та корисному проведенні часу з однодумцями. Створення ІС ефективного аналізу та підбору користувачів є важливим завданням, так як на даний момент не існує подібних систем. Система соціалізації дозволить оптимізувати процес пошуку та знайомства людей в інтернеті. Для реалізації даної ІС потрібно використати згорткову нейронну мережу для пошуку обличчя на фотографії. Також необхідно застосувати алгоритми нечіткого пошуку та модель Noisy Channel для аналізу інформації про користувача і формування списку релевантних користувачів. Для успішного створення ІС соціалізації особистостей за спільними інтересами найважливішим завданням є правильно зрозуміти та опрацювати процес соціалізації

користувачів.

Аналіз останніх досліджень і публікацій

В напряму дослідження та реалізації процесів соціалізації сьогодні існує досить велика кількість публікацій, зокрема, в [1-2] автори пропонують нову систему рангів серед користувачів соціальної мережі та адаптивну мережеву модель для зручної взаємодії між користувачами. Також автори пропонують удосконалити систему агрегації та сортування інформації в соціальних мережах відповідно до заданих параметрів. В [3-4] автори пропонують удосконалити можливості соціальних мереж в якості аналізу даних користувачів, формування характеристик роботи користувача та удосконалити можливості соціальних мереж через пошук релевантної подібності між користувачами та, в подальшому, визначення основних інтересів користувача і видача йому найбільш релевантних співпадінь при кожному наступному пошуку контенту в соціальній мережі. В [5] автори пропонують новий підхід для сортування/зберігання даних в хмарі та їх видачу при пошуку з використанням нової гнучкої моделі видобування даних в соціальних мережах. Основною перевагою подібної ІС є те, що опису практично аналогів або немає (Tinder та Badoo) [6], або це є закритими комерційними проектами, тому категорично стверджувати, що аналоги існують є не припустимим. Подібність ІС з Tinder та Badoo є в

застосуванні згорткової нейронної мережі, яка перевіряє наявність аналогічних користувачів та використовується для ідентифікації користувачів системи [7]. Основним недоліком Tinder та Badoo є те, що вони надають максимально обмежений механізм соціалізації, дозволяючи відфільтрувати користувачів по статі, віку та місцезнаходженню, що є неоптимізованим процесом і несуттєво економить витрачений час на соціалізацію [8].

Виділення невіршених раніше частин загальної проблеми. Мета роботи

Метою даної статті є створення ІС соціалізації особистостей на основі нечіткого пошуку за ключовими словами на базі моделі Noisy Channel з алгоритмами ефективного розподілу контенту [9] та згорткової нейронної мережі для ідентифікації користувачів. На сьогодні не існує подібної ІС аналізу користувацького контенту та формування списку найбільш подібних за інтересами користувачів. Об'єктом аналізу даних для соціалізації є персональні профілі потенційних релевантних користувачів [10] соціальних мереж. ІС має аналізувати інформацію про користувачів і на основі цього формувати посортований список користувачів за вирахуванням відсотком подібності [11] між користувачами ІС, що і є основною перевагою системи в порівнянні з її основними аналогами [12]. ІС повинна складатися з таких етапів: реєстрація/авторизація користувача, ознайомлення користувача з правилами користування системою [13], завантаження фотографії (на основі згорткової нейронної мережі перевірено на наявність людського обличчя на фото та у вже існуючій базі даних ІС) [14], додавання аватару користувача та заповнення альбому фотографіями. Наступним етапом є введення користувачем контенту про себе, аналіз вказаної інформації та її порівняння з контентом про вже існуючих в системі користувачів з використанням алгоритму Левенштейна, алгоритмів нечіткого пошуку та моделі Noisy Channel, після чого буде сформований посортований список користувачів за вирахуванням відсотком подібності між користувачами. Є можливість додатково відфільтрувати сформований список, змінити особисту інформацію в результаті чого буде наново сформований список, або вийти в головне меню системи. Також всі зміни, внесені користувачем, та переписки зберігатимуться в БД, а всі фотографії в хмарному сервісі збереження медіа-даних. Останнім кроком є дослідження профілю та інтересів [15] обраного з сформованої вибірки користувача та початок спілкування з обраним користувачем, вся переписка у вигляді окремого потоку повідомлень буде зберігатися в БД.

Якщо здійснити постановку та обґрунтування проблеми, то її можна поділити на дві основні частини, а саме реалізація роботи двох нейронних мереж: згорткової та сіамської, що дозволить здійснити пошук людського обличчя та відтворення алгоритмів SEO-технологій, а саме алгоритму

Левенштейна та моделі Noisy Channel, які необхідні для здійснення соціалізації користувачів за спільними інтересами всередині ІС. В наш час важливість використання машинного навчання в сучасних ІС соціалізації особистостей є безперечно важливою, так як всі соціальні мережі оперують великою кількістю даних [16], часто необхідно здійснити розподіл даних, що дозволить оптимізувати та пришвидшити роботу системних процесів, цього можна досягти здійснивши обробку частини даних з використанням машинного навчання. Важливо також не забути про особливості створення серверних програм, а саме оптимізацію алгоритмів взаємодії серверних програм, реалізувати алгоритм оптимізованої пакетної передачі даних різних розмірів. Необхідно реалізувати алгоритми аналізу текстових даних в соціальних мережах [17], створити нейронну мережу, яка порівнює фотографії користувача та текстову інформацію залишену користувачем і на основі проаналізованих даних формує вибірку. Важливо реалізувати оптимальне збереження даних на мобільних пристроях, сформувати основні шаблони збереження обробки та формування вибірки даних з використанням поточної вибірки. Потрібно оптимізувати алгоритми пошуку інформації та здійснити реалізацію механізмів видобування даних, створити нейронну мережу, яка аналізує на вході всю подану соціальну інформацію та з використанням вхідних параметрів відкидає нерелевантну інформацію, зберігаючи лише необхідну в певний момент вибірку даних. Також потрібно не забувати здійснювати оптимізацію процесів всередині системи, а саме зберігання даних на серверах системи та в локальній БД пристрою, та оптимізацію алгоритмів передачі та зберігання даних.

Функціонал системи

Одним з основних завдань є реалізація алгоритму Левенштейна, оскільки він виконує основну роботу ІС, зокрема, пошук відстаней між словами, тобто аналізує подібність слів в контенті. Основна суть алгоритму Левенштейна полягає в тому, що можна побудувати кінцевий автомат, який точно розпізнає набір рядків на заданій відстані Левенштейна до цільового слова. Потім можливо подати будь-яке слово, і автомат прийме або відхилить його залежно від того, чи є відстань Левенштейна до цільового слова щонайбільше тією відстанню, яка наперед визначена, що дозволяє використовувати цей алгоритм для великих обсягів текстового контенту. Крім того, оскільки створюється кінцевий автомат, він здійснює всю роботу за $O(n)$ часу з довжиною рядка, що тестується. Можна зробити висновок, що даний алгоритм ідеально підходить для роботи з середніми об'ємами тексту, оскільки виконується лінійно і швидко здійснює обробку підготовленої текстового контенту.

Ініціалізується матриця, яка вимірює (m, n) - комірки відстані Левенштейна між m -символьним префіксом одного з n -префіксом іншого слова. Матрицю заповнюють від верхнього лівого до

нижнього правого кута. Кожен стрибок по горизонталі або вертикалі відповідає вставці/видаленню відповідно. Вартість зазвичай встановлюється на 1 для кожної з операцій. Діагональний стрибок може коштувати 1-2 символи, коли i в рядку i в стовпці символи не збігаються, або 0, якщо збігаються. Кожна

клітинка завжди мінімізує вартість локально. Таким чином число в нижньому правому куті є отримана відстань Левенштейна між обома словами. В табл. 1 подано приклад аналізу двох слів та знаходження відстаней між ними, яка в результаті дорівнює значенню 4.

Таблиця 1. Приклад роботи алгоритму Левенштейна

| words | w1 | m | e | i | l | e | n | s | t | e | i | n |
|-------|----|----|----|---|---|---|---|---|---|---|----|----|
| w2 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| l | 1 | 1 | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| e | 2 | 2 | 1 | 2 | 3 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| v | 3 | 3 | 2 | 2 | 3 | 4 | 4 | 5 | 6 | 7 | 8 | 9 |
| e | 4 | 4 | 3 | 3 | 3 | 3 | 4 | 5 | 6 | 6 | 7 | 8 |
| n | 5 | 5 | 4 | 4 | 4 | 4 | 3 | 4 | 5 | 6 | 7 | 7 |
| s | 6 | 6 | 5 | 5 | 5 | 5 | 4 | 3 | 4 | 5 | 6 | 7 |
| h | 7 | 7 | 6 | 6 | 6 | 6 | 5 | 4 | 4 | 5 | 6 | 7 |
| t | 8 | 8 | 7 | 7 | 7 | 7 | 6 | 5 | 4 | 5 | 6 | 7 |
| e | 9 | 9 | 8 | 8 | 8 | 7 | 7 | 6 | 5 | 4 | 5 | 6 |
| i | 10 | 10 | 9 | 8 | 9 | 8 | 8 | 7 | 6 | 5 | 4 | 5 |
| n | 11 | 11 | 10 | 9 | 9 | 9 | 8 | 8 | 7 | 6 | 5 | 4 |

Модель Noisy Channel є алгоритмом для визначення правильного слова в ситуаціях, коли воно незрозуміле після виконання алгоритму Левенштейна. Шум у цьому контексті – це все, що приховує сигнали та дані. Модель зашумленого каналу отримала таку назву, тому що вихідний сигнал – призначене слово – приховане під час передачі, коли порушення або помилки створюють шум у каналі. У цьому випадку шум полягає в орфографічній помилці. Якщо слово не знайдено в словнику перевірки орфографії, воно ідентифікується як орфографічна помилка, і слова-кандидати пропонуються на основі їхньої ймовірності бути призначеним словом, як правило, в результаті того, наскільки ці слова близькі до орфографічної помилки. Як правило, найімовірніші кандидати включатимуть одну зміну, і є чотири різні типи помилок в співставленні літер: видалення, вставка, заміна та скасування. Найпростішим прикладом однієї зміни може бути: відсутні літери, наприклад, leter замість letter, написано зайві літери, наприклад, misstake, замість mistake, літери переставлені, наприклад, recieved замість received, замінені літери, наприклад, fimate замість finite.

Алгоритм N-грам широко використовується, є досить простий та вискоелективний. Ідея методу полягає в тому, що якщо два слова A та B збігаються з огляду на можливі помилки та неточності, вони, ймовірно, матимуть спільні n-грами, де n – довжина підрядка. Під час індексації слово чітко ділиться на N-грами, а потім це слово ввійде до словника для кожної N-грами. Під час пошуку, поточний запит також ділиться на N-грами, а для кожної з них виконується послідовний пошук списку слів, що містять такий підрядок. Найбільш часто використовуваним на практиці є 3-грами – підрядки з довжиною в 3 символи. Вибір більшого значення n

призводить до обмеження мінімальної довжини слова, при якій помилки вже можуть бути виявлені.

ІС соціалізації особистостей за спільними інтересами на основі SEO-технологій та методів машинного навчання має одну зовнішню сутність, а саме: *Користувач* і основні внутрішні сутності, до яких належать: *Браузер*, *Контролер системи*, *База даних*, *Сесія користувача*, *Сервер* та *Сервіс системи*. Під час функціонування ІС соціалізації особистостей здійснюється постійна взаємодія між зовнішніми і внутрішніми сутностями:

1. При ініціалізації роботи ІС зовнішня сутність *Користувач* входить в систему за допомогою внутрішньої сутності – *Браузер*.

2. За допомогою сутності *Браузер* та внутрішньої сутності *Контролер системи* здійснюється авторизація або реєстрація користувача в ІС.

3. Після успішної реєстрації/авторизації сутність *Контролер системи* надсилає сесійний токен в сутність *Браузер* та зберігає поточний токен всередині браузера на час існування сесії.

4. Сутність *Користувач* здійснює завантаження власної фотографії для підтвердження справжності за допомогою сутності *Браузер*, який додає фотографію за допомогою сутності *Контролер системи*.

5. Сутність *Контролер системи* на основі нейронних мереж перевіряє наявність обличчя на фотографії сутності *Користувач* та перевіряє на наявність вже аналогічних фотографій в БД ІС.

6. Сутність *Контролер системи* на основі сутності *Сервер* зберігає всі отримані дані про користувача та надсилає сформовані дані в іншу внутрішню сутність *База даних*.

7. Сутність *База даних* зберігає всю інформацію на допомогу внутрішньої сутності *Сервіс системи*.

8. *Сервіс системи* опрацьовує інформацію та створює список унікальних користувачів системи.

9. *Сервіс системи* на основі алгоритмів нечіткого пошуку здійснює аналіз даних користувача, який в даний момент працює всередині системи та інших зареєстрованих користувачів.

10. На основі результаті роботи алгоритмів нечіткого пошуку сутностей *Сервіс системи* та *Сесія користувача* створює список користувачів, сформований по спаданню відсоткової подібності між ними відповідно до даних поточного користувача.

11. Сутність *Сервіс системи* здійснює кінцеву перевірку отриманих даних на наявність пошкоджених та недостовірних даних та зберігає отриману інформацію за допомогою сутності *База даних*.

12. За допомогою користувацького запиту сформовані дані з сутності *База даних* надсилаються сутності *Користувач*.

13. *Користувач* здійснює перегляд наданої інформації та вибирає профіль користувача, який йому найбільше сподобався з використанням сутності *Сервіс системи*.

14. Сутність *Користувач* здійснює дію проставлення вподобання іншому користувачу, для того щоб інший користувач системи міг побачити, що його обрав поточний користувач.

15. Сутності *Сервіс системи* та *Контролер системи* починають переписку з іншим користувачем, формуючи діалог, який за допомогою сутностей *Контролер системи* та *База даних* зберігається та існує всередині ІС у вигляді потоку повідомлень конкретної сесії для подальшого динамічного відображення динаміки діалогу даних/повідомлень користувачів/сесій в конкретний момент часу.

16. Сутність *Сервіс системи* надсилає запит сутності *База даних* для перевірки наявності даних, оскільки одним з найважливіших аспектів системи є безпека користувачів та даних, відповідно час від час сутність *Сервіс системи* надсилає сутності *База даних* запит, щоб перевірити цілісність даних та надіслати повідомлення про помилку у разі певних проблем з обробкою даних, також необхідно забезпечити повністю асинхронну обробку та передавання даних заради максимально швидкої роботи ПЗ, що дозволить користувачам оптимально швидко користуватися ІС.

Внутрішня сутність *База даних* складається з основних елементів, які дозволять безпечно та швидко опрацювати інформацію користувача, а саме:

- Перевірити дані.
- Перевірити наявність бекап файлів.
- Зберегти дані.
- Зберегти діалог користувача.
- Сформувати пакет даних.
- Перевірити правильність запиту.
- Відправити дані на запит.

Після цього сутність *Контролер системи* надсилає запит сутності *Сервіс системи* та асинхронну обробку даних. Сутність *Сервіс системи* за допомогою використання динамічних подій постійно здійснює перевірку станів ІС, в яких вона знаходиться і відповідно реагує на всі зміни системи в конкретний момент часу. Сутність *Сервіс системи*

здійснює час від час в окремому потоці перевірку ІС на наявність помилок та винятків, які не були зареєстровані в журналі системи і намагається їх виправити або надіслати на обробку сутності *Контролер системи* на подальше опрацювання.

Сутність *Контролер системи* має такі функції:

- Опрацювання даних користувача.
- Опрацювання даних системи.
- Перевірка даних.
- Перевірка доступності сесії.
- Перевірка ідентифікаторів.
- Виведення повідомлення про помилку.
- Зберігання посилань системи.
- Перевірки токена користувача.
- Перевірка правильності оновлень.

Сутність *Сервіс системи* містить такі функції:

- Створення списку користувачів.
- Аналіз списку за допомогою алгоритмів.
- Створення відсоткового співвідношення.
- Обробка переписок користувачів.
- Створення потоку повідомлень.
- Обробка потоку повідомлень.
- Створення користувацьких токенів.

Сутність *Браузер*, хоч і не є основною, але містить такі функції:

- Авторизація/Реєстрація.
- Збереження токена.
- Перевірка сесії.
- Відображення повідомлень користувачу.
- Зчитування користувацьких повідомлень.
- Додавання фотографії користувача.
- Відправлення запитів користувача на сервер.
- Відображення помилок клієнта.

Після того як сутність *Сервіс системи* сформував запит потоку повідомлень, він відправляє запит сутності *Контролер системи*, який запитує токен діалогу в сутності *Сервіс системи*. Остання відправляє поточний токен, в якому містяться дані про поточну сесію. Далі сутність *Контролер системи* через сутність *Браузер* відображає сформований діалог сутності *Користувач*, після чого *Користувач* або продовжує роботу, або виходить з ІС.

Результати досліджень та їх обговорення

Створено діаграму переходів станів, яка відображає систему, що в заданий момент часу знаходиться точно в одному з скінченної множини станів. Діаграма складається з переходів, які містять шість основних етапів:

1. Вхід користувача в систему – користувач входить в ІС, здійснює реєстрацію і/або авторизацію.

2. Додавання фотографії – користувач додає власну фотографію, ІС перевіряє наявність обличчя на фотографії та справжність користувача, здійснює пошук по фотографіях в БД, в разі успішного додавання фотографії зберігає її в БД, а користувач отримує доступ до списків користувачів.

3. Формування списку користувачів – система здійснює аналіз всіх даних користувача та на їх основі здійснює за допомогою алгоритмів нечіткий пошук

по текстовій інформації інших користувачів, тим самим формує список релевантних користувачів по відсотковому співвідношенню, сортує список по спаданню і надсилає на перегляд користувачу.

4. Перегляд вибраного профілю – користувач здійснює перегляд обраного профілю, переглядає фотографії, загальну інформацію та інтереси вибраного користувача і в результаті успіху ставить в подобаючий профілю та розпочинає переписку.

5. Здійснення переписки – формується окремий потік повідомлень для переписки між двома користувачами, який динамічно оновлюється та зберігається в базі даних системи.

6. Збереження даних та вихід з ІС – після здійснення переписки здійснюється перевірка наявності токена та роботи сесії, зберігаються всі

сформовані дані, користувацька інформація, потік повідомлень та логування в БД системи, після чого користувач виходить в головне меню системи.

На рис. 1 подано композиційну схему ІС, за допомогою якої здійснено поділ об'єктів на класи та відображено залежність класів між собою. На рис. 2 подано діаграму станів, що описує зміни стану об'єкта протягом усього процесу роботи системи. На рис. 3 подано діаграму класифікації видів інформації, а також її підвидів та залежностей необхідних для роботи ІС. На рис. 4 подано діаграму сценарію роботи ІС, яка будує сценарій того, як за допомогою синхронних та асинхронних процесів, змін станів, форм даних та умов виконання запитів функціонує система.

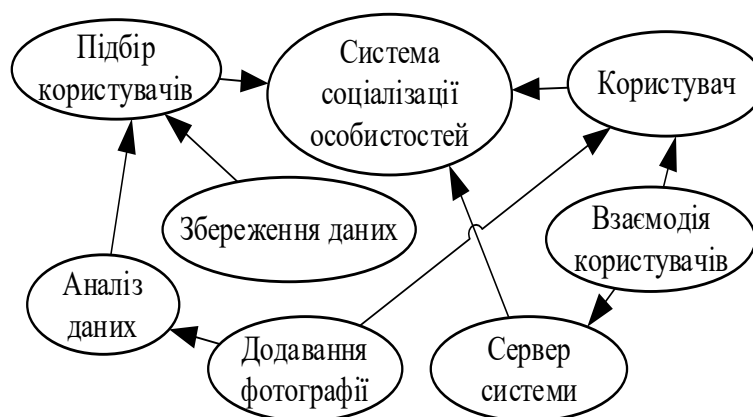


Рис. 1. Композиційна схема інтелектуальної системи

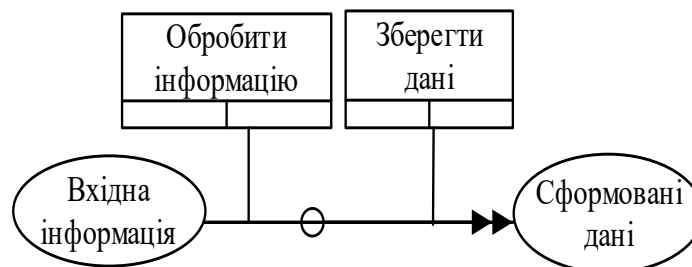


Рис. 2. Діаграма станів



Рис. 3. Діаграма класифікації видів інформації

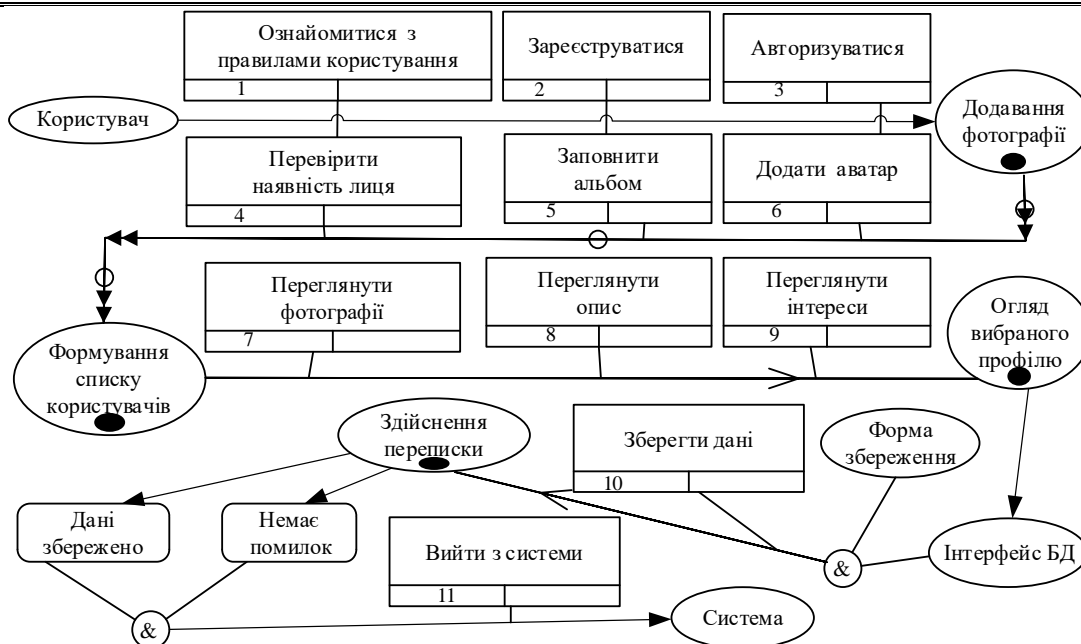


Рис. 4. Діаграма сценарію роботи інтелектуальної системи

Здійснено побудову діаграм сутність-зв'язок, система має наступні сутності:

1. *Користувач* – сутність, що відповідає за додавання особистої інформації, пошук відповідних користувачів ІС, додавання фотографії та здійснення діалогу з вибраним користувачем (рис. 5). *Користувач*

має 5 атрибутів: ID_користувача, Логін, Пароль, Персональні дані, Середній бал. Сутність *Користувач* має три зв'язки з сутністю *Сервер*, а саме: "Шукає користувачів", "Надсилає інформацію" та "Вибирає користувача зі списку".

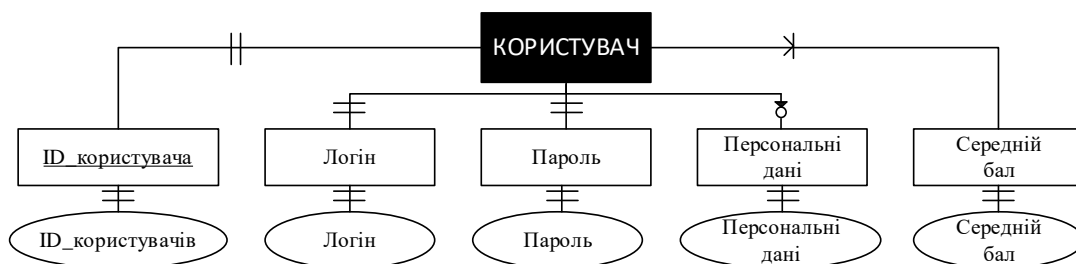


Рис. 5. Сутність "Користувач"

2. *Браузер* – сутність, що відповідає за підтримку інтерактивного діалогу між користувачем та ІС.

3. *База даних* – сутність, що відповідає за збереження даних користувача, підібраних списків релевантних користувачів та проміжних результатів функціонування ІС.

4. *Контролер системи* – сутність, що відповідає за контроль коректного функціонування ІС.

5. *Сервіс системи* – сутність, що відповідає за вибір операцій, здійснення відповідей за запит, виконання алгоритмів аналізу текстової інформації, перевірки справжності користувачів та формування посортованого списку користувачів по інтересах (рис. 6). Сутність *Сервіс системи* має 5 атрибутів: ID_серверного сеансу, Назва операції, Параметри запиту, Оцінка користувачів, Значення середнього параметру.

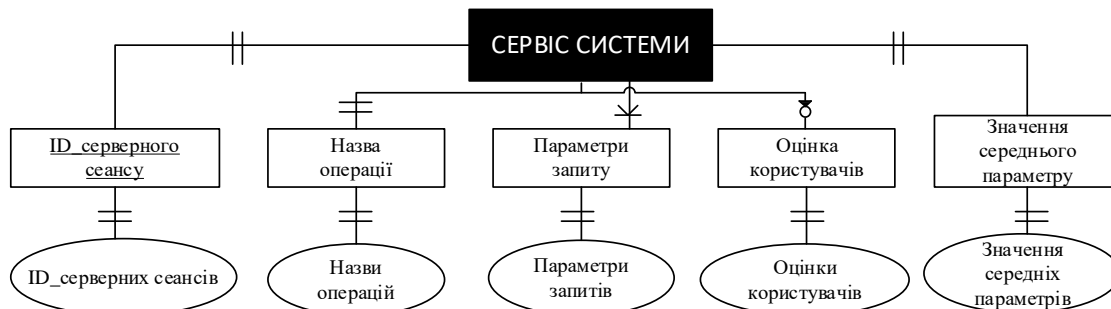


Рис. 6. Сутність "Сервіс системи"

6. *Сесія користувача* – слабка сутність, що відображає сесію поточного користувача та містить всі необхідні параметри для роботи користувача всередині сесії (рис. 7). Сутність *Сесія користувача* є слабкою сутністю, має 2 атрибути: ID_сеансу, Токен

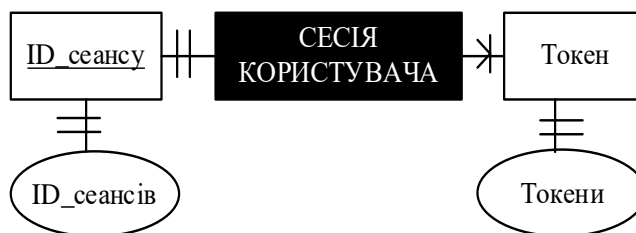


Рис. 7. Сутність "Сесія користувача"

7. *Сервер* – сутність, що сама по собі містить лише сформовані дані та ідентифікатори, але є дуже важливою, оскільки саме завдяки серверу здійснюється взаємодія між іншими сутностями та всі дані в результаті роботи зберігаються на сервері (рис. 8). Сутність *Сервер* має 2 атрибути: Пакет даних,

Ідентифікатор. Сутність *Сервер* має чотири зв'язки з сутністю *Сервіс системи*, а саме: "Зберігає проміжні параметри", "Надсилає запит користувачу", "Зберігає потік діалогу користувачів" та "Формує список користувачів".



Рис. 8. Сутність "Сервер"

Сутність *Користувач* та слабка сутність *Сесія користувача* мають слабкі зв'язки типу 1-N (рис. 9). Слабка сутність *Сесія користувача* має слабкий зв'язок з сутністю *Сервер* типу 1-N. Сутність

Користувач має зв'язки з сутністю *Сервер* типу 1-N. Сутність *Сервер* має зв'язки з сутністю *Сервіс системи*, а саме: типу 1-N та 1-1.

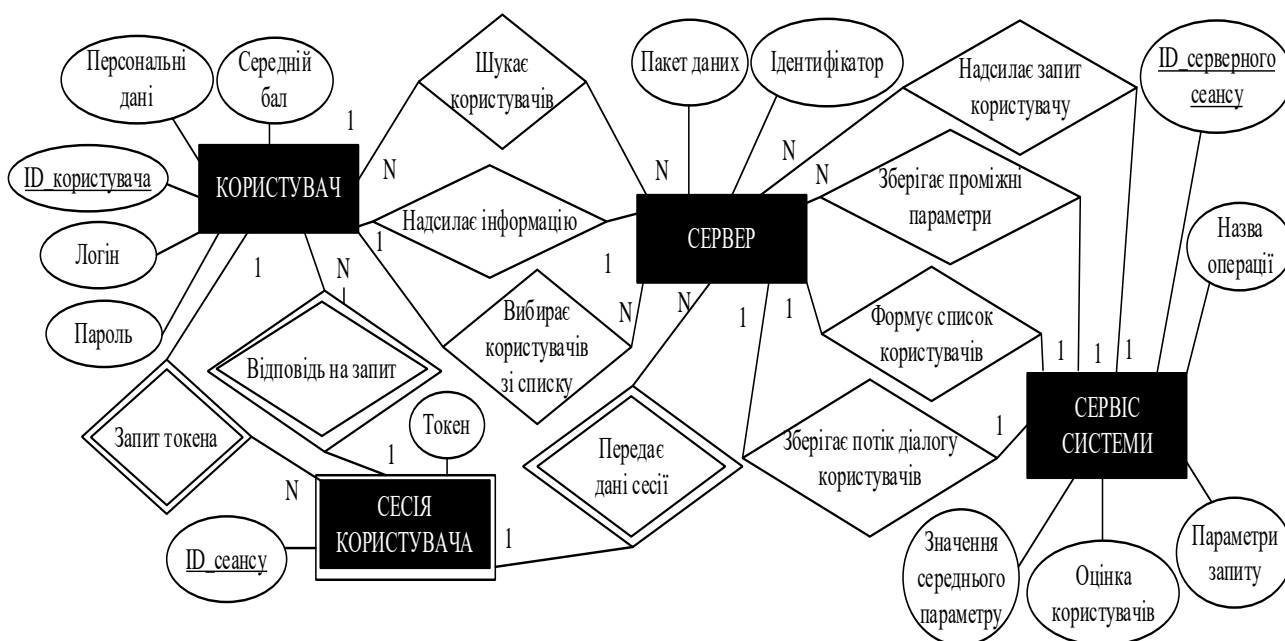


Рис. 9. Діаграма сутність-зв'язок

Створено діаграму переходів станів, яка відображає систему, що в заданий момент часу знаходиться точно в одному з скінченної множини

станів (рис. 10) та розроблено матрицю переходів (табл. 2).

Таблиця 2. Матриця переходів станів системи

| Поточний стан | Умова | Дія | Наступний стан |
|-----------------|--|--|----------------|
| ПОЧАТКОВИЙ СТАН | Вхід користувача в систему | Завантаження системи | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Завантажено систему | Реєстрація | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Зареєстровано | Авторизація користувача | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Авторизовано | Створення токена користувача | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Створено токен | Відправлення токена в браузер на збереження | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Збережено токен | Вхід користувача на головну сторінку | КІНЦЕВИЙ СТАН |
| Група 2 | | | |
| ПОЧАТКОВИЙ СТАН | Користувач у системі | Завантаження фотографії | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Завантажено фотографію | Пошук обличчя на фотографії | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Знайдено обличчя | Перевірка обличчя | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Обличчя перевірено | Порівняння обличчя з іншими в БД | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Перевірено обличчя користувача | Збереження фотографії | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Збережено фотографію | Створення альбому | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Додавання фотографії в альбом | Збереження альбому в БД | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Збережено альбом | Завершення роботи з фотографіями | КІНЦЕВИЙ СТАН |
| Група 3 | | | |
| ПОЧАТКОВИЙ СТАН | Введено інформацію | Обробка інформації | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Сформовані текстові дані | Обробка тексту | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Оброблено текстові дані | Початок нечіткого пошуку серед всіх користувачів | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Узагальнення нечіткого пошуку між користувачів | Формування списку | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Список сформовано | Сортування списку | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Список посортовано | Фільтрування списку по зовнішніх параметрах | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Відфільтровано список | Вивід інформації користувачу | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Інформація опрацьована | Завершення роботи | КІНЦЕВИЙ СТАН |
| Група 4 | | | |
| ПОЧАТКОВИЙ СТАН | Отримано список | Перегляд списку всіх користувачів | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Список переглянуто | Вибір користувача | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Користувача вибрано | Перегляд фотографій користувача | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Фотографії переглянуто | Перегляд опису користувача | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Опис переглянуто | Перегляд інтересів користувача | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Інтереси переглянуто | Виставлення вподобання користувачу | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Вподобання поставлено | Початок переписки з користувачем | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Почато переписку | Створення потоку повідомлень | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Потік створено | Збереження ідентифікатора в БД | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Ідентифікатор збережено | Збереження токена сесії | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Токен сесії збережено | Збереження потоку повідомлень в БД | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Потік збережено | Перевірка токена сесії | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Перевірено наявність токена | Збереження параметрів в БД | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Параметри збережено | Перевірка наявності помилок | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Помилки перевірені | Збереження логування в БД | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Логування збережено | Перевірка локальної БД | ОПРАЦЮВАННЯ |
| ОПРАЦЮВАННЯ | Локальну БД перевірено | Виведення даних користувачу | ОЧІКУВАННЯ |
| ОЧІКУВАННЯ | Виведена інформація | Вихід в головне меню | КІНЦЕВИЙ СТАН |

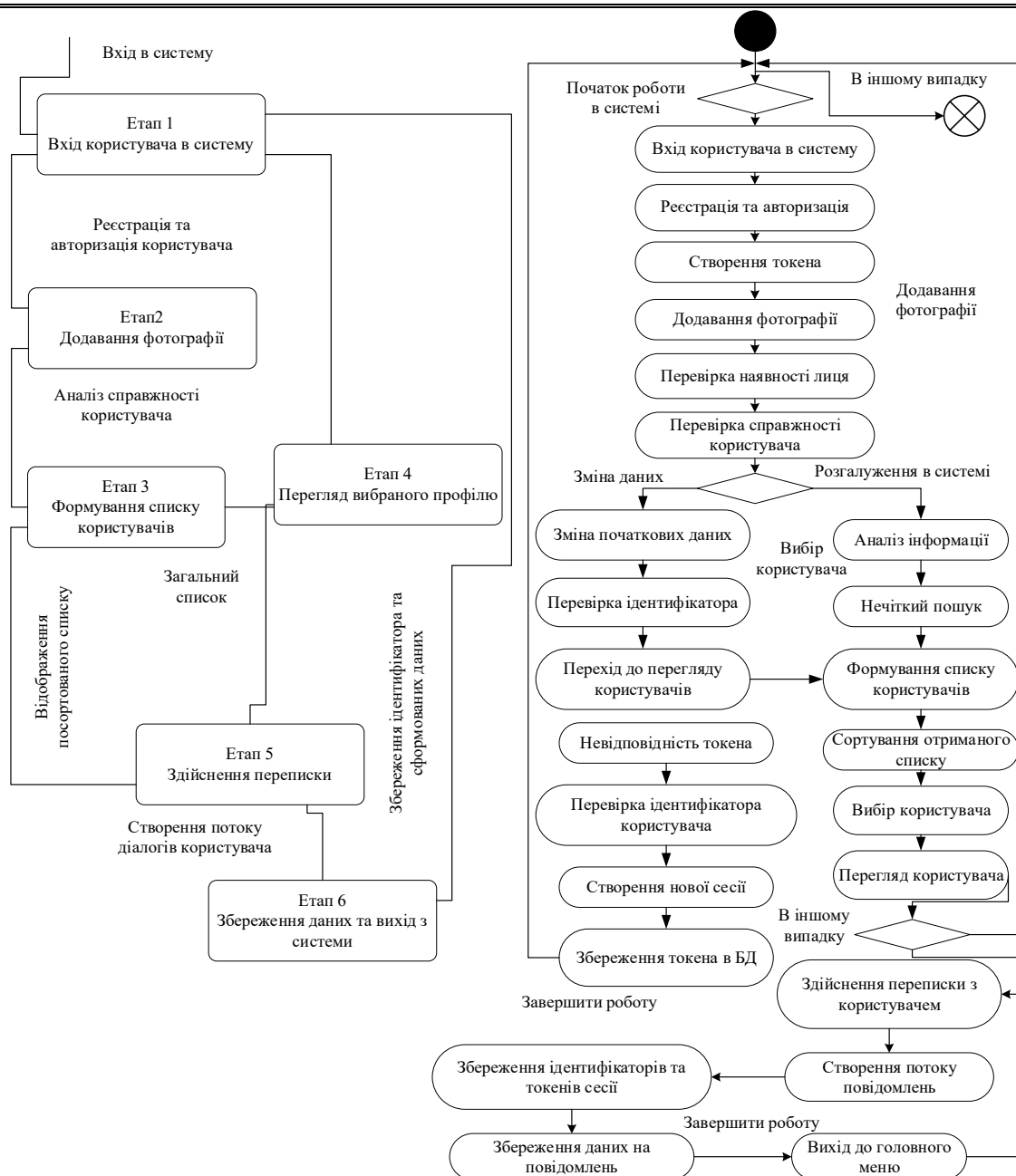


Рис. 10. Діаграма переходів станів

ІС повинна складатися з таких етапів: реєстрація і/або авторизація користувача, знайомство користувача з правилами користування ІС. Завантаження фотографії, яку з використанням згортової нейромережі перевірено на наявність людського обличчя, та на наявність цього обличчя в вже існуючій БД системи, додавання аватару користувача та заповнення альбому фотографіями. Наступним кроком є вказання користувачем інформації про себе, аналіз вказаної інформації та її порівняння з інформацією про вже існуючих в системі користувачів з використанням алгоритму Левенштейна, алгоритмів нечіткого пошуку та моделі Noisy Channel, після чого формується посортований список користувачів за вирахуванням відсотком схожості між користувачами. Є можливість додатково відфільтрувати сформований список, змінити

особисту інформацію в результаті чого буде наново сформований список, або вийти в головне меню системи. Також всі зміни внесені користувачем, та переписки зберігатимуться в БД, а всі фотографії в хмарному сервісі збереження медіа-даних. Останнім кроком є дослідження профілю та інтересів обраного з сформованої вибірки користувача та початок спілкування з обраним користувачем, вся переписка у вигляді окремого потоку повідомлень зберігається в БД. Загальний алгоритм роботи ІС поданий на рис. 11 у вигляді use case діаграми та на рис. 12 у вигляді activity діаграми. Далі подано основні скріншоти роботи ІС, на рис. 13 подано процес додавання інформації про користувача, а саме опис, інтереси та загальна інформація про те, що та кого шукає користувач.

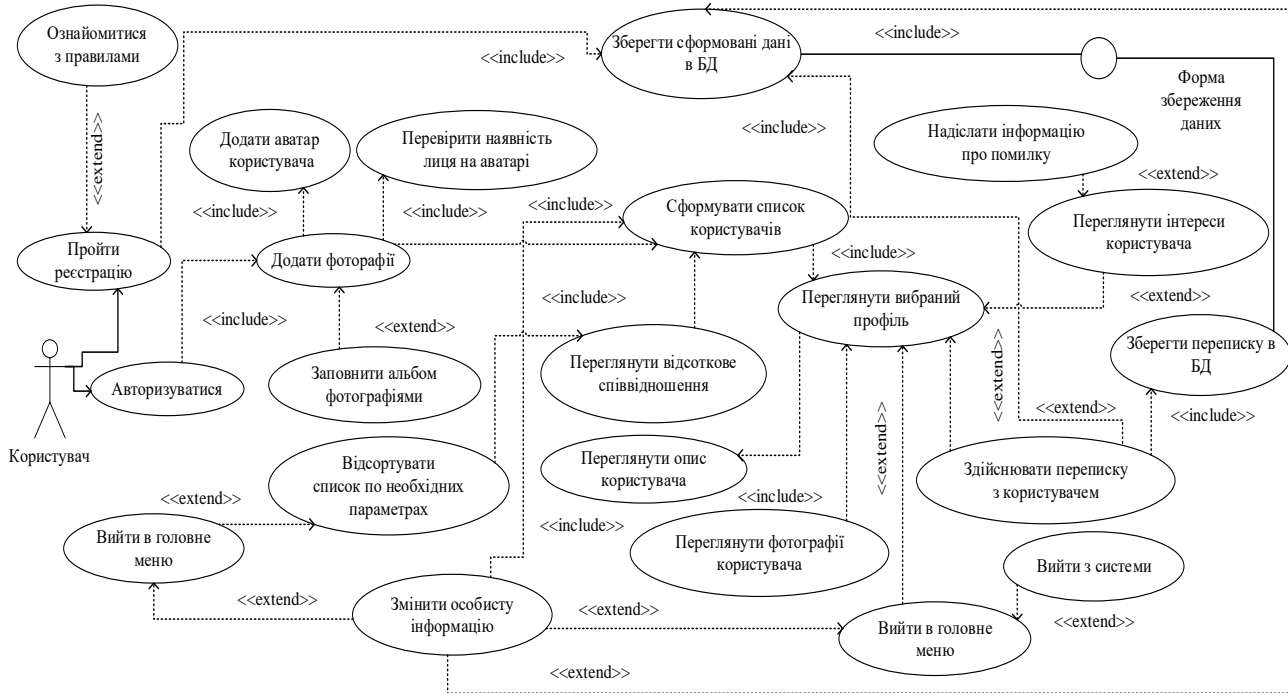


Рис. 11. Use Case діаграма

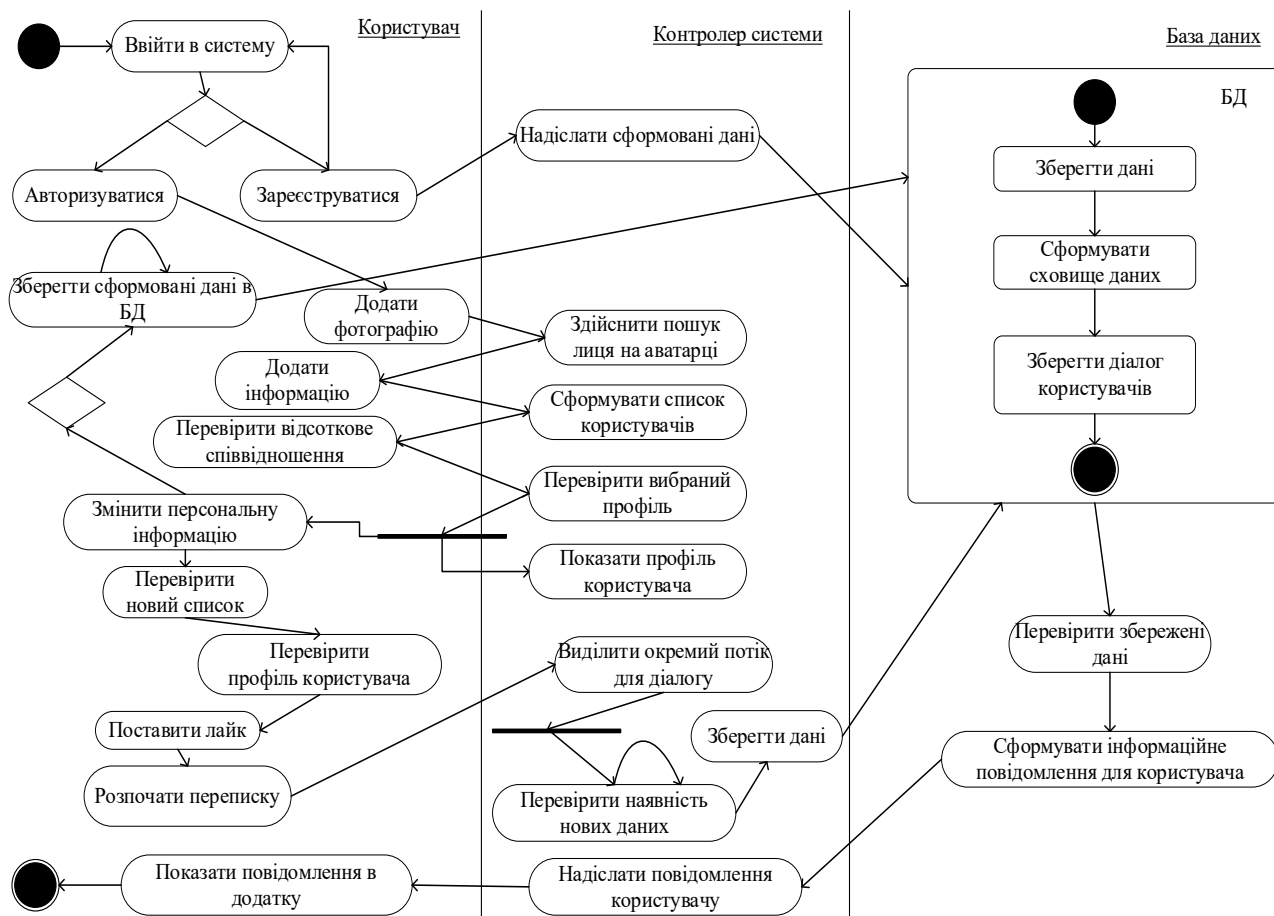


Рис. 12. Activity діаграма

На рис. 14 подано поточні додані фотографії та використання згорткової нейронної мережі. Згорткова нейронна мережа складається з шарів двох типів: згорткових шарів і subsampling шарів або шарів

підвибірки. У ній кожен шар має топографічну структуру, тобто кожен нейрон пов'язаний з фіксованою точкою вихідного зображення, а також з областю вхідного зображення, яка опрацьовується

даними нейроном. У кожному місці кожного шару існує цілий ряд різних нейронів, кожен зі своїм набором вхідних ваг, пов'язаних з нейронами в прямокутному фрагменті попереднього шару. Згортова нейронна мережа дозволяє визначити чи є такий користувач вже в БД, та чи зображене обличчя на фотографії, як видно на виділеному фрагменті скріншоту – 2 картинки система не дозволяє поставити на аватар, оскільки на них немає обличчя.

На рис. 15 подано сформований список користувачів з використанням нечіткого пошуку посортований за вирахуванням відсотком подібності між користувачами. Серед алгоритмів нечіткого пошуку вибрано лінійний пошук – просте послідовне застосування метрики (в даному випадку метрики Левенштейна) до слів з вхідного тексту. Щоб обчислювати оптимальну відстань між словами в

тексті, потрібно модифікувати алгоритм знаходження звичайної відстані Левенштейна наступним чином: зберігати не дві, а три останні рядки матриці, а також додати відповідну додаткову умову – в разі виявлення транспозиції при розрахунку відстані також враховувати її вартість. Таким чином модифікувавши алгоритм використано модель Noisy Channel, що дозволило оптимізувати процес аналізу інформацію, спочатку створивши матрицю параметрів користувачів, потім асинхронно з використанням лінійного пошуку порівнювати поточні параметри і формувати до кожного параметру відсоток подібності. Останнім кроком є пошук середнього відсотку подібності користувачів. На рис. 16-17 подано вибір користувача, перегляд профілю та фотографій обраного користувача та початок переписки.

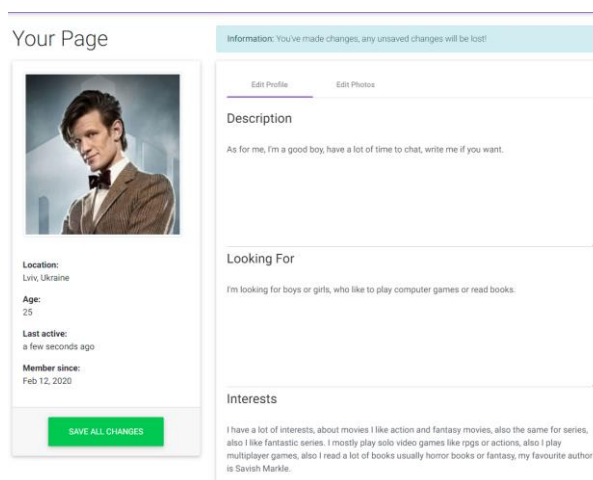


Рис. 13. Інформація користувача

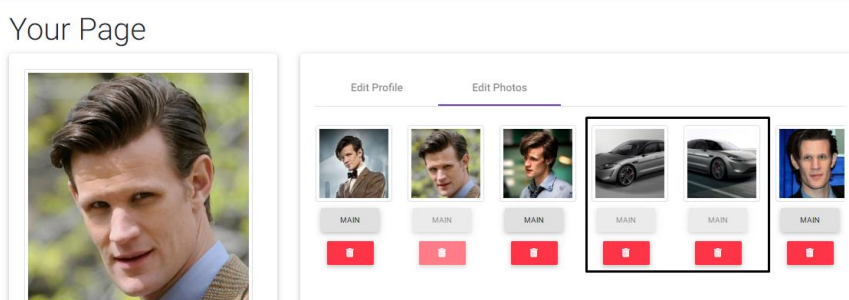


Рис. 14. Вибір аватару користувача

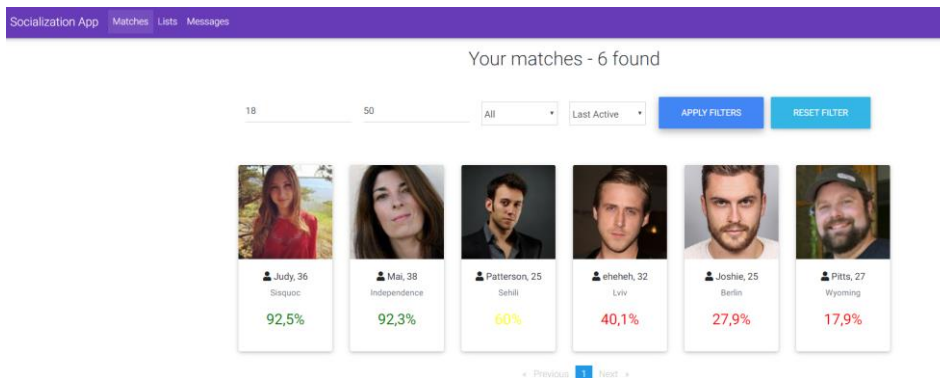


Рис. 15. Сформований список користувачів



Рис. 16. Вибір користувача

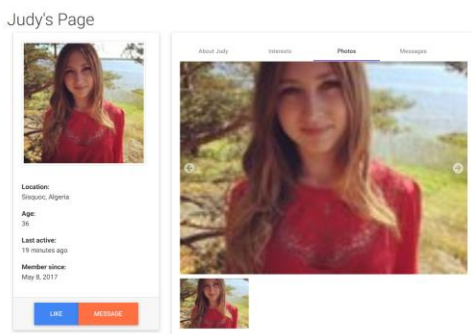


Рис. 17. Перегляд фотографій користувача

Реалізований в ІС алгоритм формування вибірки подібних за інтересами користувачів є ефективнішою та точнішою на 25–30% на відмінну від звичайного алгоритму Левенштейна. Одночасно з цим реалізований алгоритм здійснює вибірку в 10 разів швидше за алгоритм Левенштейна.

Висновки та перспективи подальшого розвитку

Здійснено аналітичний огляд літературних джерел, серед яких коротко описано про всі аспекти

Список літератури

1. Parry M. E., Kawakami T., Kishiya K. The effect of personal and virtual word-of-mouth on technology acceptance. *Journal of Product Innovation Management*. 2012. № 29 (6). P. 952–966. DOI: <http://doi.org/10.1111/j.1540-5885.2012.00972.x>
2. Schivinski B., Dąbrowski D. The effect of social-media communication on consumer perceptions of brands. *Journal of Marketing Communications*. 2014. № 22 (2). P. 189–214. DOI: <http://doi.org/10.1080/13527266.2013.871323>
3. Ranjbaran B., Jamshidian M., Dehghan Z. A survey of identification of major factors influencing customers attitude toward machine made carpet brands. *Journal of Business Strategies*. 2007. № 5 (23). P. 109–118.
4. Schmäh M., Wilke T., Rossmann A. Electronic word of mouth: A systematic literature analysis. *Digital Enterprise Computing*. 2017. P. 147–158.
5. Wang X., Yu C., Wei Y. Social media peer communication and impacts on purchase intentions: A consumer socialization framework. *Journal of Interactive Marketing*. 2012. № 26 (4). P. 198–208. DOI: <http://doi.org/10.1016/j.intmar.2011.11.004>
6. Vysotska V. Information Technology for Internet Resources Promotion in Search Systems Based on Content Analysis of Web-Page Keywords. *Radio Electronics, Computer Science, Control*. 2021. № 3. P. 133–151.
7. Batiuk T., Vysotska V., Lytvyn V. Intelligent System for Socialization by Personal Interests on the Basis of SEO-Technologies and Methods of Machine Learning. *Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23-24 April 2020 : CEUR workshop proceedings*. 2020. № 2604. P. 1237–1250.
8. Hudson S., Roth M., Madden T. J. The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management*. 2015. № 2 (8). P. 68–76. DOI: <http://doi.org/10.1016/j.tourman.2014.09.001>
9. Hanna R., Rohm A., Crittenden V. L. We're all connected: The power of the social media ecosystem. *Business Horizons*. 2011. № 54 (3). P. 265–273. DOI: <http://doi.org/10.1016/j.bushor.2011.01.007>
10. Guidry J. D., Messner M., Jin Y. From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies. *Corporate Communications: An International Journal*. 2015. № 20 (3). P. 344–359.
11. Gao L. Online consumer behavior and its relationship to website atmospheric induced flow: Insights into online travel agencies in China. *Journal of Retailing and Consumer Services*. 2014. № 21 (4). P. 653–655.

що відносяться до сучасної соціалізації особистостей, а саме інформація про нейронні мережі з розпізнавання обличчя та алгоритми нечіткого пошуку для опрацювання текстової інформації. Описано основні призначення спроектованої та розробленої ІС, мету її створення, які основні проблеми вирішуються при створенні такого типу систем. Проаналізовано причини та фактори, що є важливими для значення актуальності створення даної ІС. Вказані існуючі аналоги системи, їх переваги та недоліки. Здійснено опис основних алгоритмів роботи ІС, а саме алгоритм Левенштейна, модель Noisy Channel та алгоритм N-грам. Наведено приклад роботи алгоритму Левенштейна з пошуком відстані наявної різниці між двома словами. Здійснено побудову необхідних IDEF5 діаграм, в яких описано основні класи системи як об'єкти, створені на основі цих класів та процеси взаємодії об'єктів між собою. Розроблено діаграму переходів станів та відповідно до неї побудовано матрицю переходів станів ІС, де відображено умови, дії та залежності станів системи. Створено діаграму сутність-зв'язок, на якій відображені окремі сутності системи та їх взаємодія в ході виконання системи. Побудовано діаграму варіантів використання та діаграму активностей системи, що дозволило розглянути систему з боку об'єктно-орієнтованого підходу та повноцінно описати загальні вимоги до типової структури створюваної ІС соціалізації особистостей за спільними інтересами, що дає змогу здійснювати подальшу реалізацію відповідного ПЗ. Експериментальна апробація реалізованої ІС соціалізації особистостей на основі розробленого алгоритму формування вибірки спільних подібних за інтересами користувачів продемонструвала точність пошуку релевантних користувачів на 25–30% та швидкість в 10 разів більша при отриманні результату на відмінну від звичайного алгоритму Левенштейна.

12. Ferrara E., Interdonato R., Tagarelli A. Online popularity and topical interests through the lens of Instagram. *Hypertext and Social Media*. 2014. № 2. P. 24–23. DOI: <http://doi.org/10.1145/2631775.2631808>
13. Erkan I. The influence of e-WOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*. 2016. № 4. P. 47–55.
14. Elaheebocus S. M., Weal M., Morrison L. Peer-based social media features in behavior change interventions: Systematic review. *Journal of Medical Internet Research*. 2018. № 20 (2). P. 1–20. DOI: <http://doi.org/10.2196/jmir.8342>
15. De-Gregorio F., Sung Y. Understanding attitudes toward and behaviors in response to product placement. *Journal of Advertising*. 2010. № 39 (1). P. 83–96. DOI: <http://doi.org/10.2753/JOA0091-3367390106>
16. Geurin-Eagleman A. N. Communicating via photographs: A gendered analysis of Olympic athletes' visual self -presentation on Instagram. *Sport Management Review*. 2015. № 19 (2). P. 133–145. DOI: <http://doi.org/10.1016/j.smr.2015.03.002>
17. Chu S. C. Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter. *Electronic Commerce Research and Applications*. 2016. № 14 (4). P. 251–260.

References

1. Parry, M. E., Kawakami, T. (2012), Kishiya, K. "The effect of personal and virtual word-of-mouth on technology acceptance", *Journal of Product Innovation Management*, No. 29 (6), P. 952–966. DOI: <http://doi.org/10.1111/j.1540-5885.2012.00972.x>
2. Schivinski, B., Dąbrowski, D. (2014), "The effect of social-media communication on consumer perceptions of brands", *Journal of Marketing Communications*, No. 22 (2), P. 189–214. DOI: <http://doi.org/10.1080/13527266.2013.871323>
3. Ranjbaran, B., Jamshidian, M., Dehghan, Z. (2007), "A survey of identification of major factors influencing customers attitude toward machine made carpet brands", *Journal of Business Strategies*, No. 5 (23), P. 109–118.
4. Schmäh, M., Wilke, T., Rossmann, A. (2017), "Electronic word of mouth: A systematic literature analysis", *Digital Enterprise Computing*, P. 147–158.
5. Wang, X., Yu, C., Wei, Y. (2012), "Social media peer communication and impacts on purchase intentions: A consumer socialization framework", *Journal of Interactive Marketing*, No. 26 (4), P. 198–208. DOI: <http://doi.org/10.1016/j.intmar.2011.11.004>
6. Vysotska, V. (2021), "Information Technology for Internet Resources Promotion in Search Systems Based on Content Analysis of Web-Page Keywords", *Radio Electronics, Computer Science, Control*, No. 3, P. 133–151.
7. Batiuk, T., Vysotska, V., Lytvyn, V. (2020), "Intelligent System for Socialization by Personal Interests on the Basis of SEO-Technologies and Methods of Machine Learning", *Computational Linguistics and Intelligent Systems (COLINS 2020): 4th International Conference, Lviv, 23-24 April 2020 : CEUR workshop proceedings*, No. 2604, P. 1237–1250.
8. Hudson, S., Roth, M., Madden, T. J. (2015), "The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees", *Tourism Management*, No. 2 (8), P. 68–76. DOI: <http://doi.org/10.1016/j.tourman.2014.09.001>
9. Hanna, R., Rohm, A., Crittenden, V. L. (2011), "We're all connected: The power of the social media ecosystem", *Business Horizons*, No. 54 (3), P. 265–273. DOI: <http://doi.org/10.1016/j.bushor.2011.01.007>
10. Guidry, J. D., Messner, M., Jin, Y. (2015), "From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies", *Corporate Communications: An International Journal*, No. 20 (3), P. 344–359.
11. Gao, L. (2014), "Online consumer behavior and its relationship to website atmospheric induced flow: Insights into online travel agencies in China", *Journal of Retailing and Consumer Services*, No. 21 (4), P. 653–655.
12. Ferrara E., Interdonato R., Tagarelli A. (2014), "Online popularity and topical interests through the lens of Instagram", *Hypertext and Social Media*, No. 2, P. 24–23. DOI: <http://doi.org/10.1145/2631775.2631808>
13. Erkan, I. (2016), "The influence of e-WOM in social media on consumers' purchase intentions: An extended approach to information adoption", *Computers in Human Behavior*, No. 4, P. 47–55.
14. Elaheebocus, S. M., Weal, M., Morrison, L. (2018), "Peer-based social media features in behavior change interventions: Systematic review", *Journal of Medical Internet Research*, No. 20 (2), P. 1–20. DOI: <http://doi.org/10.2196/jmir.8342>
15. De-Gregorio, F., Sung, Y. (2010), "Understanding attitudes toward and behaviors in response to product placement", *Journal of Advertising*, No. 39 (1), P. 83–96. DOI: <http://doi.org/10.2753/JOA0091-3367390106>
16. Geurin-Eagleman, A. N. (2015), "Communicating via photographs: A gendered analysis of Olympic athletes' visual self -presentation on Instagram", *Sport Management Review*, No. 19 (2), P. 133–145. DOI: <http://doi.org/10.1016/j.smr.2015.03.002>
17. Chu, S. C. (2016), "Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter", *Electronic Commerce Research and Applications*, No. 14 (4), P. 251–260.

Надійшла (Received) 25.12.2021

Відомості про авторів / Сведения об авторах / About the Authors

Батюк Тарас Миронович – магістр кафедри "Інформаційні системи та мережі", Національний університет "Львівська політехніка", вул. Степана Бандери 12, Львів, Україна; e-mail: taras.batiuk.mnsa.2020@lpnu.ua; ORCID ID: <https://orcid.org/0000-0001-5797-594X>

Батюк Тарас Миронович – магістр кафедри "Информационные системы и сети", Национальный университет "Львовская политехника", Украина.

Taras Batiuk – PhD student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Висоцька Вікторія Анатоліївна – кандидат технічних наук, доцент, доцент кафедри "Інформаційні системи та мережі", Національний університет "Львівська політехніка", вул. Степана Бандери 12, Львів, Україна; e-mail: Victoria.A.Vysotska@lpnu.ua; ORCID ID: <https://orcid.org/0000-0001-6417-3689>

Висоцкая Виктория Анатольевна – к.т.н., доцент, доцент кафедри «Информационные системы и сети», Национальный университет "Львовская политехника", Украина.

Victoria Vysotska – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ СОЦИАЛИЗАЦИИ ПОЛЬЗОВАТЕЛЯ ПО СХОЖЕСТИ ИНТЕРЕСОВ

Объектом исследования является процесс социализации личностей, поскольку в наше время задача социализации очень важна и все современные социальные сети пытаются максимально оптимизировать и автоматизировать социализацию разнообразных пользователей с использованием всех популярных современных технологий, таких как нейронные сети и алгоритмы анализа пользовательского текста. **Предметом** исследования является методы и технологические средства поиска и формирования списка релевантных пользователей по сходству интересов для осуществления социализации. Соответственно, осуществляется анализ профиля пользователя системы, а именно определение подлинности пользователя, посредством поиска человеческого лица на пользовательских фотографиях с использованием нейронных сетей и анализ пользовательской информации с использованием алгоритмов нечеткого поиска и модели Noisy Channel. **Цель** работы – создание интеллектуальной системы для социализации личностей на основе нечеткого поиска по словам с использованием модели Noisy Channel с алгоритмами эффективного распределения текстовой информации и сверточной нейронной сети для идентификации пользователей системы. В статье решаются следующие **задачи**: 1. Проанализировать современные и известные подходы, методы, средства и алгоритмы решения задач социализации личностей по сходству интересов. 2. Разработка общей структуры типовой интеллектуальной системы социализации личностей по общим интересам. 3. Сформировать функциональные требования к основным модулям структуры типовой интеллектуальной системы социализации личностей по общим интересам. 4. Разработать интеллектуальную систему поддержки социализации пользователя по сходству интересов на основе нейронных сетей, нечеткого поиска и модели Noisy Channel и провести экспериментальную апробацию. Используются следующие **методы**: метод Левенштейна; модель Noisy Channel; алгоритм N-грам; нечеткий поиск. Получены следующие **результаты**: построена и описана общая структура типовой интеллектуальной системы социализации личностей по общим интересам. К основному назначению создания системы в первую очередь следует отнести создание нового алгоритма анализа пользовательской информации и поиска наиболее подходящих пользователей, в соответствии с проанализированным текстом на основе уже существующих алгоритмов, таких как алгоритм Левенштейна, алгоритм расширения выборки, алгоритм N-грамм и модели Noisy Channel. Также испытан дальнейшее развитие шаблон асинхронного создания программного продукта, что позволит создать практически полностью динамическую систему. Следует усовершенствовать сверточную нейронную сеть, что позволит эффективно и динамично осуществлять поиск человеческих лиц на фотографии и проверять наличие уже существующих людей в базе данных системы. **Выводы**: Выявлено, что реализуемый алгоритм осуществляет выборку примерно в 10 раз быстрее, чем обычный алгоритм Левенштейна. Также реализуемый в системе алгоритм для формирования выборки пользователей является более эффективным и точным примерно на 25-30% по сравнению с обычным алгоритмом Левенштейна.

Ключевые слова: расстояние Левенштейна; модель Noisy Channel; алгоритм N-грам; нечеткий поиск.

INTELLECTUAL SYSTEM DEVELOPMENT FOR USER SOCIALIZATION SUPPORT BY INTERESTS SIMILARITY

The **object** of research is the process of socialization of individuals, because nowadays the task of socialization is very important and all modern social networks try to optimize and automate the socialization of various users using all popular modern technologies such as neural networks and user text analysis algorithms. The **subject** matter of the study is the methods and technologies for the search and formation of a list of relevant users by similarity of interests for socialization. Accordingly, the system user profile analysis is studied, namely the identification of the user by searching the human face in user photos using neural networks and analysing user information using fuzzy search algorithms and the Noisy Channel model. The **goal** of the work is to create an intelligent system for socialization of individuals based on fuzzy word search using the Noisy Channel model with algorithms for efficient distribution of textual information, and a convolutional neural network to identify users of the system. The following **tasks** were solved in the article: 1. Analyse modern and most well-known approaches, methods, tools and algorithms for solving problems of socialization of individuals by similar interests. 2. To development the general structure of a typical intellectual system of socialization of individuals by common interests. 3. To form functional requirements to the basic modules of structure of typical intellectual system of socialization of persons on common interests. 4. Develop an intelligent system of support for user socialization by similarity of interests based on neural networks, fuzzy search and Noisy Channel model and conduct experimental testing. The following **methods** are used: Levenstein's method; Noisy Channel model; N-gram algorithm; fuzzy search. The following **results** were obtained: the general structure of a typical intellectual system of socialization of individuals by common interests was built and described. The main purpose of the system is to create a new algorithm for analysing user information and finding the most suitable users, according to the analysed text based on existing algorithms such as Levenstein's algorithm, sampling algorithm, N-gram algorithm and Noisy Channel model. The template of asynchronous creation of a software product which will allow to create almost completely dynamic system also underwent further development. It is necessary to improve the convolutional neural network, which will allow efficient and dynamic search of human faces in the photo, and check the presence of existing people in the database of the system. **Conclusions**: It was found that the implemented algorithm performs sampling approximately 10 times faster than the usual Levenstein algorithm. Also, the implemented in the system algorithm for forming a sample of users is more efficient and accurate by about 25-30% compared to the usual Levenstein algorithm.

Keywords: Levenshtein distance; Noisy Channel Model; N-gram algorithm; Fuzzy search.

Бібліографічні описи / Bibliographic descriptions

Батиук Т. М., Висоцька В. А. Розробка інтелектуальної системи підтримки соціалізації користувача за подібністю інтересів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 1 (19). С. 13–26. DOI: <https://doi.org/10.30837/ITSSI.2022.19.013>

Batiuk, T., Vysotska, V. (2022), "Intellectual system development for user socialization support by interests similarity", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (19), P. 13–26. DOI: <https://doi.org/10.30837/ITSSI.2022.19.013>