

# НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

## NEUROINFORMATICS AND INTELLIGENT SYSTEMS

# НЕЙРОІНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 004.9

### ТЕХНОЛОГІЯ СОЦІАЛІЗАЦІЇ ОСОБИСТОСТЕЙ ЗА СПІЛЬНИМИ ІНТЕРЕСАМИ НА ОСНОВІ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА SEO-ТЕХНОЛОГІЙ

**Батюк Т. М.** – студент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

**Висоцька В. А.** – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

#### АНОТАЦІЯ

**Актуальність.** Соціалізація особистостей за спільними інтересами спричинено потребою більшості людей спростити частину життєвих моментів за рахунок зменшення часу на їх реалізацію. З швидкими темпами росту інформації, завантаженості людини в суспільстві та у зв'язку з останніми епідемічними світовими подіями людина стає ізольованою від можливості спілкуватися. А це однією із важливих потреб людської свідомості та самореалізації. Тому є актуальним попитом мати можливість отримувати рекомендований список подібних людей за спільними інтересами як результат інтелектуального пошуку множини релевантних користувачів соціальних мереж через аналіз фото людського обличчя на користувацьких фотографіях (на основі нейронних мереж) і аналіз користувацької інформації (на основі алгоритмів нечіткого пошуку та моделі Noisy Channel).

**Мета** – розроблення технології для соціалізації особистостей на основі SEO-технології та методу машинного навчання через використання згортової та сіамської нейронних мереж для ідентифікації користувачів та алгоритмів аналізу тексту для підбору релевантних користувачів майбутнього спілкування.

**Метод.** При реалізації SEO-технології обрано алгоритми нечіткого пошуку по словах на основі моделі Noisy Channel з алгоритмами ефективного розподілу текстової інформації. При реалізації машинного навчання розроблено згорткову нейронну мережу для ідентифікації користувачів системи.

**Результати.** Розроблено інтелектуальну систему соціалізації особистостей за спільними інтересами на основі SEO-технології та методи машинного навчання. Здійснено реалізацію роботи двох нейронних мереж: згортової та сіамської, що дозволило здійснити пошук людського обличчя, на завантажуваних користувачем фотографіях і порівняти знайдене обличчя з уже наявними в базі даних/Інтернет. Це дає можливість ефективно ідентифікувати справжність користувача та гарантувати, що цього користувача на даний момент нема в базі даних, відповідно він потенційно є реальним. За допомогою алгоритмів нечіткого пошуку, алгоритму Левенштейна та моделі Noisy Channel створено алгоритм аналізу та порівняння користувацької інформації, який для поточного користувача формує список наявних користувачів системи, посортований по спаданню відсоткового співвідношення подібності користувачів та вказує, наскільки інтереси в інших користувачів збігаються з інтересами поточного користувача.

**Висновки.** Виявлено, що реалізований в системі алгоритм для формування вибірки користувачів є ефективнішою та точнішою приблизно на 25–30% в порівнянні зі звичайним алгоритмом Левенштейна. Також реалізований алгоритм здійснює вибірку приблизно в 10 разів швидше, ніж звичайний алгоритм Левенштейна.

**КЛЮЧОВІ СЛОВА:** нечіткий пошук, алгоритм Левенштейна, модел Noisy Channel, згорткова нейронна мережа, сіамська нейронна мережа, фотоаналіз обличчя, алгоритм розширення вибірки, алгоритм N-грам.

#### АБРЕВІАТУРИ

БД – база даних;  
ІС – інформаційна система;  
ІТ – інформаційна технологія;  
МН – машинне навчання;  
ПЗ – програмне забезпечення;  
SEO – search engine optimization.

#### НОМЕНКЛАТУРА

$S$  – система соціалізації особистостей;  
 $I$  – множина вхідних даних;  
 $O$  – множина вихідних даних;  
 $R$  – основні правила опрацювання потоку вхідних даних в ІС соціалізації користувача;  
 $U$  – параметри опрацювання вхідних даних;  
 $N$  – нейронна мережа;

$\alpha$  – оператор скачування вхідних даних;  
 $\beta$  – оператор опрацювання вхідних даних;  
 $\gamma$  – оператор пошуку релевантних користувачів за аналізом профілів та фото;  
 $P$  – вдосконалена імітаційна модель пошуку релевантних користувачів;  
 $\mu$  – оператор ідентифікації користувача;  
 $\chi$  – оператор формування даних фотогалереї;  
 $\omega$  – оператор формування списку та даних релевантних користувачів;  
 $\lambda$  – оператор підтримки соціальних запитів;  
 $i_1$  – множина даних ідентифікації (фото, пароль, логін, відбиток пальця, голос);  
 $i_2$  – сховище даних фотогалереї обличч конкретної соціальної мережі;  
 $i_3$  – різні фото обличчя користувача;  
 $i_4$  – конкретний запит користувача;  
 $o_1$  – запити з ІС до конкретних релевантних користувачів за вимогою;  
 $o_2$  – оновлення для профілю користувача ІС в конкретній соціальній мережі;  
 $o_3$  – лайк за вимогою користувача ІС;  
 $r_1$  – правила алгоритму взаємодії;  
 $r_2$  – правила роботи згорткової нейронної мережі;  
 $r_3$  – правила алгоритму нечіткого пошуку;  
 $r_4$  – правила алгоритму автозбереження даних;  
 $u_1$  – множина рівнів доступу;  
 $u_2$  – множина вимог доступу;  
 $u_3$  – множина вимог додавання фото;  
 $u_4$  – множина вимог формування списку релевантних користувачів;  
 $u_5$  – множина вимог підтримки соціалізації;  
 $C_{AU}$  – контент авторизованого користувача.

## ВСТУП

На сьогодні соціалізація особистостей за спільними інтересами є надзвичайно важливим процесом під час ізоляції людей із-за подовженості світової епідемії COVID2019 [1–5]. Паралельно більшість людей завжди намагаються спростити та автоматизувати всі основні життєві процеси, які зазвичай займають багато вільного часу [6–9]. Це ж стосується і процесу соціалізації особистості. МН та SEO-технології на даний момент є надзвичайно важливими в контексті розроблення ІС опрацювання та аналізу великих даних [10–12]. Практично кожна популярна серед великої кількості людей ІС використовує відповідні механізми соціалізації [13–15]. Для ефективної реалізації ІТ соціалізації зазвичай оптимізують існуючі алгоритми і/або створюють нові алгоритми відповідно до поставлених вимог та розв'язку конкретної задачі [16–21]. Питання МН навчання, соціалізації та SEO-технології є досить популярними і висвітлені в низці статей [22–27].

Об'єктом дослідження є процес соціалізації особистостей, оскільки на сьогодні завдання соціалізації є дуже важливим і всі сучасні соціальні мережі намагаються максимально оптимізувати та

автоматизувати соціалізацію різних класів користувачів (за віком, статтю, вподобаннях, хобі тощо) з використанням усіх популярних сучасних ІТ, таких як нейронні мережі та алгоритми аналізу користувачьких текстових повідомлень. Для успішного створення ІС соціалізації особистостей за спільними інтересами найважливішим завданням є визначити конкретну мету соціалізації (наприклад, за якими інтересами/хобі, стилем життя тощо) та відповідно опрацювати/підтримувати процес соціалізації відповідного класу користувачів.

Предметом дослідження метод и та засоби ІТ соціалізації відповідно класу користувачів. Тому здійснюється аналіз/дослідження мети/особливостей користувача ІС, а саме визначення достовірності (справжності існування) конкретного користувача за допомогою пошуку людського обличчя на множині користувачьких фотографіях з використанням нейронних мереж і аналіз користувачької інформації з використанням алгоритмів нечіткого пошуку та моделі Noisy Channel.

**Метою дослідження** є розроблення ІТ для соціалізації особистостей, яка використовує SEO-технології та методи машинного навчання. Для досягнення мети були поставлені такі завдання:

– удосконалити імітаційну модель для пошуку множини релевантних користувачів в ІС соціалізації особистостей за допомогою пошуку людського обличчя на користувачьких фото з використанням нейронних мереж і аналіз користувачької інформації з використанням алгоритмів нечіткого пошуку та моделі Noisy Channel;

– удосконалити згорткову нейронну мережу, що дозволило ефективно здійснювати пошук людських обличч на фото, та перевіряти наявність вже існуючих людей в БД ІС;

– розробити алгоритму аналізу користувачької інформації та пошуку найбільш релевантних користувачів, відповідно до проаналізованого тексту на основі вже існуючих алгоритмів, таких як алгоритм Левенштейна, алгоритм розширення вибірки, алгоритм N-грам та моделі Noisy Channel;

– розробити та описати ПЗ візуальної пошуку множини релевантних користувачів для соціалізації особистостей за допомогою аналізу фото;

– здійснити аналіз результатів експериментальної апробації запропонованої ІТ пошуку множини релевантних користувачів соціалізації особистостей.

## 1 ПОСТАНОВКА ПРОБЛЕМИ

Систему соціалізації особистостей  $S$  подано імітаційною моделлю через кортежем:

$$S = \langle I, O, R, U, N, \alpha, \beta, \gamma \rangle,$$

$$\begin{aligned} I &= \{i_1, i_2, i_3, i_4\}, & O &= \{o_1, o_2, o_3\}, \\ R &= \{r_1, r_2, r_3, r_4\}, & U &= \{u_1, u_2, u_3, u_4, u_5\}. \end{aligned}$$

Основними процесами ІС соціалізації особистості є «Ідентифікація користувача», «Формування даних

фотогалереї», «Формування списку релевантних користувачів» та «Підтримка соціальних запитів».

Процес ідентифікації користувача ІС соціалізації особистості опишемо суперпозицією:

$$C_{AU} = \mu \circ \beta \circ \alpha,$$

$$C_{AU} = \mu(\beta(\alpha(i_1, i_2, i_4), r_1, u_1), u_2).$$

Процес формування даних фотогалереї користувача ІС соціалізації особистості опишемо суперпозицією:  $C_{CU} = \chi \circ \beta \circ \alpha$ , тобто

$$C_{CU} = \chi(\beta(\alpha(C_{AU}, i_2, i_3, i_4), r_1, u_3), r_2).$$

Процес формування списку релевантних користувачів опишемо суперпозицією:

$$C_{UL} = \omega \circ \gamma \circ \beta \circ \alpha,$$

$$C_{UL} = \omega(\gamma(\beta(\alpha(C_{CU}, i_2), i_3), u_4), r_3).$$

Процес підтримки соціальних запитів користувача ІС соціалізації особистості опишемо суперпозицією:

$$C_{US} = \lambda \circ \gamma \circ \beta \circ \alpha,$$

$$C_{US} = \lambda(\gamma(\beta(\alpha(C_{US}, i_2), i_4), u_5), r_4).$$

## 2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Під час створення обраної ІС важливим завданням є дослідження та аналіз вже існуючих аналогів для формування множини особливостей, характеристик, переваг та недоліків в існуючих ІС. Під час пошуку аналогів практично не знайдено подібних ІС соціалізації особистостей. Серед знайденого ПЗ виділимо 3 аналоги, які за функціоналом є найбільш наближеними до запропонованої ІС соціалізації, зокрема Tinder, Badoo та Chatous.

Tinder є найбільш популярною та старою ІС соціалізації особистостей. Серед основних переваг є підтримка кросплатформеності як для всіх операційних системах, так і для будь-якого смартфона та чрез браузер. Також перевагою є аналіз фотографії обличчя людини, пошук аналогів обличчя в Tinder та наявність обличчя на фото гарантує те, що ІС користуються лише реальні люди, і при знайомстві, чи під час переписки з користувачем можна бути впевненим, що це реальний користувач, а не шахрай. Ще до переваг можна виділити повну відсутність реклами, що дозволяє не відволікатися і повноцінно використовувати ПЗ. Серед недоліків в першу чергу варто виділити те, що ІС використовує лише базові фільтри пошуку користувачів, а саме вік, стать, та місцезнаходження, що дозволяє лише приблизно звизити область пошуку релевантних користувачів. Також ІС не має ніяких алгоритмів соціалізації та підбору користувачів за спільними інтересами. Ще одним недоліком є обмеження кількості користувачів, яких можна переглянути за день, обмеження відповідно можна зняти за платну підписку.

Badoo є досить популярною новою ІС соціалізації особистостей. Серед основних переваг можна виділити наявність даної програми на всіх

платформах. Badoo можна використовувати як програму для робочого столу на всіх операційних системах, як застосунок для смартфона та у браузерній версії. ІС є повністю безкоштовною, відповідно в ній немає обмежень на кількість користувачів, яких можна переглянути за день. Також має безлімітні вподобання, які можна виставляти користувачам та необмежену кількість повідомлень, які можна написати. Серед недоліків варто відзначити величезну кількість реклами, яку неможливо вимкнути, що не дозволяє повноцінно зручно використовувати програму. Badoo не має ніяких алгоритмів соціалізації за спільними інтересами та надає можливість використання лише базових фільтрів пошуку користувачів.

Chatous є найменш популярною ІС соціалізації з усіх розглянутих. Основною перевагою даної ІС є те, що вона повністю безкоштовна для використання, має необмежену кількість вподобань, повідомлень та переглядів користувачів протягом дня. Також в даній програмі повністю відсутня будь яка реклама. Серед переваг можна виділити аналіз фото користувача та підтвердження обличчя користувача на фото, що гарантує те, що ми спілкуємося з реальною людиною. Серед недоліків те, що дана ІС не доступна на всіх платформах, а лише у вигляді браузерної програми, що обмежує кількість можливих користувачів. Chatous має основний недолік, як і в інших аналогах – повністю відсутні алгоритми підбору користувачів по інтересах, що не дозволяє оптимально здійснювати пошук користувачів. Присутні лише базові алгоритми фільтрації основних параметрів користувачів, що є неефективним, так як це означає, що більшість користувачів не будуть відповідати очікуванням пошуку, так як не матимуть спільних інтересів.

З таблиці 1 можна зробити висновок, що важливим пунктом реалізації ІС є відсутність реклами, або її мінімальна кількість. Також перевагою порівняно з Tinder є підтримка необмеженого числа користувачів, їх вподобань та повідомлень, які можна написати протягом дня.

Таблиця 1 – Порівняння аналогів ІС соціалізації

Назва	Tinder	Badoo	Chatous
Відсутність реклами	+	-	+
Необмежені користувачі	-	+	+
Необмежені вподобання	-	+	+
Необмежені повідомлення	-	+	+
Платна версія	+	-	-
Крос платформеність	+	+	-
Аналіз фотографій	+	+	+
Базові фільтри пошуку	+	+	+
Алгоритми соціалізації	-	-	-

Також перевагами програм Tinder та Vadoo є доступність програм на всіх платформах, що значно розширює кількість потенційних користувачів.

### 3 МАТЕРІАЛИ ТА МЕТОДИ

Одним з найголовніших реалізованих алгоритмів в ІС соціалізації особистостей за спільними інтересами є реалізації нейронної мережі, яка здійснює пошук обличчя на фото користувачів та порівнює з вже наявними в поточній БД. Для цього завдання вирішено використовувати глибинні нейронні мережі, основна особливість яких в тому, що вони крім вхідного і вихідного шару також складаються з певної кількості прихованих шарів (рис. 1).

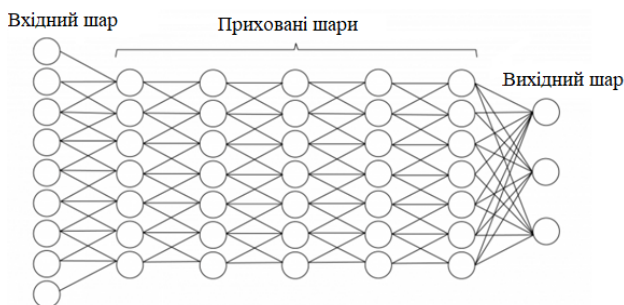


Рисунок 1 – Нейронна мережа з прихованими шарами

Для побудови глибинної нейронної мережі розпізнавання обличчя обрано концепцію розподілу шарів на дві повністю окремі підвибірки (згорткові шари та дискретизуючі шари). Основна особливість такого підходу в тому, що кожен нейрон відповідає окремій точці на фіксованому зображенні і також має зв'язок з рецептивним полем (областю зображення, яке подається на вхід). В кожному місці обраного шару можна розподілити певну кількість нейронів. Кожен є унікальним, так як має унікальний набір вхідних ваг, які сформовані за допомогою нейронів в попередньому прямокутному шарі перед поточним шаром опрацювання (рис. 2).



Рисунок 2 – Структура згорткової нейронної мережі

При програмній реалізації даної згорткової мережі перше, що треба зробити, це завантажити фотографію, або декілька фотографій в ІС, зберегти завантажені фотографії за допомогою асинхронної опрацювання в БД, після чого почати опрацювання.

Згорткова нейронна мережа складається з 5 основних шарів: вхідний шар, вихідний шар, та 3 внутрішні приховані шари (P-Net, R-Net та O-net). Кожен шар послідовно здійснює опрацювання завантаженого зображення. Спочатку здійснює опрацювання шар P-Net, який на зображенні виділяє дві квадратні рамки (рис. 3).

На рис. 3 червоне поле відображає ядро 24x24, яке змінилось розміром до початкового зображення. Обчислено ширину та висоту ядра:

$$1500 - 200 = 300, 1800 - 500 = 300.$$

Отримана ширина та висота – це ширина та висота ядра до початкового розміру. Потім здійснюється множення координатних рамок на 300:

$$0,4 \times 300 = 120, 0,2 \times 300 = 60,$$

$$0,9 \times 300 = 270, 0,7 \times 300 = 210.$$

Додається верхня ліва координата ядра, щоб отримати координати граничного поля:

$$(200 + 120, 500 + 60) \text{ і } (200 + 270, 500 + 210)$$

$$\text{або } (320, 560) \text{ та } (470, 710).$$

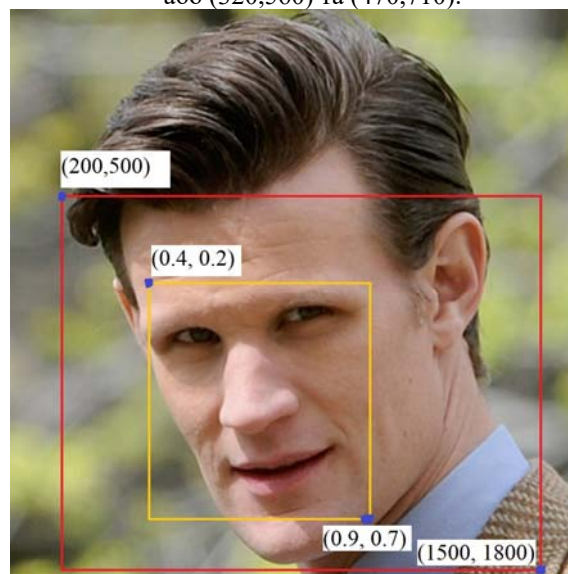


Рисунок 3 – Робота шару P-Net

Отримані дані зберігаються і далі починає працювати шар R-Net. Для кожного обмежувального вікна створюється масив однакового розміру та копіюються значення пікселів у новий масив. Іноді зображення може містити лише частину обличчя, що визирає з боку кадру (рис. 4). У такому випадку шар R-Net може повернути обмежувальний ящик, який частково знаходиться поза кадром.

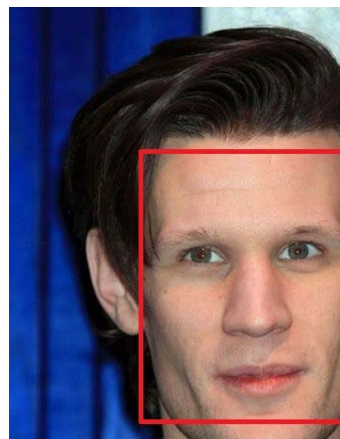


Рисунок 4 – Робота шару R-Net

Якщо обмежувальне поле не виходить за межі, копіюється частина зображення в обмежувальному полі в новий масив і заповнюється все інше 0. Цей процес заповнення масивів з 0 є padding. Після того, як прокладено масиви обмежувальної коробки, їх розмір змінюється до 24 x 24 пікселів і нормалізується до значень від -1 до 1. В даний час значення пікселів становлять від 0 до 255. Кожне значення пікселя віднімається на половину 255 (127,5) і ділиться на 127,5, після чого зберігається значення між -1 і 1.

Після стандартизації координат переставляються обмежувальні поля на квадрат, який слід передати O-Net. Отримані дані зберігаються і далі починає працювати шар O-Net. Виходи O-Net трохи відрізняються від результатів P-Net та R-Net. O-Net забезпечує 3 виходи: координати обмежувального поля, координати 5 орієнтирів обличчя та рівня довіри кожного поля. Позбавившись від ящиків із нижчим рівнем довіри, здійснюється стандартизація як координатних рамок, так і координат орієнтуру обличчя. Останнім кроком є упаковка всієї інформації у словник з трьома клавішами: «поле», «впевненість» та «ключові точки». «Поле» містить координати обмежувального поля, «впевненість» містить рівень довіри мережі для кожного вікна, а «ключові точки» містять координати кожного орієнтира обличчя (очі, ніс і кінцеві точки рота).

Наступним кроком є використання сіамської нейронної мережі, яка відповідає за те, щоб шукати аналогічні обличчя в вже існуючих фото в БД (рис. 5).

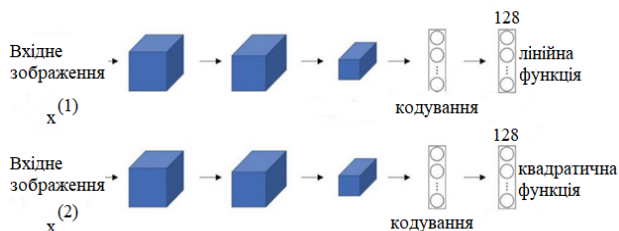


Рисунок 5 – Структура сіамської нейронної мережі

Сіамська нейронна мережа складається з 2 ідентичних шарів, кожен з яких мають однакові точні ваги. Кожен з шарів приймає одну і ту саму картинку в якості вхідних даних. Далі на виході з кожного шару є лінійна та квадратична функції, кожна з яких здійснює порівняння зображення з наявним зображенням всередині БД та формують певний коефіцієнт. Останнім кроком є вирахування різниці між коефіцієнтами та визначення наявності зображення.

#### 4 ЕКСПЕРИМЕНТИ

Головною функцією ІС соціалізації особистостей за спільними інтересами є пошук релевантних користувачів, тому основним завданням є написати оптимізований алгоритм, який максимально автоматизує процес соціалізації користувачів. В даному випадку створений спеціальний алгоритм на

основі таких алгоритмів, як алгоритм Левенштейна, розширення вибірки, N-грам та моделі Noisy Channel.

В першу чергу варта виділити вхідні дані за допомогою яких повинна бути сформована кінцева вибірка для користувача. Цими даними є параметри користувача системи, а саме опис користувача, інтереси користувача, та поле, в якому описано кого саме шукає користувач, саме ці частини будуть використовуватися для соціалізації.

З самого початку формується розширена вибірка на основі розформування поданого масиву даних на окремі елементи за допомогою алгоритму розширення вибірки. Цей алгоритм використовується найчастіше у всіх програмах порівняння слів, його особливість в тому, що він відкидає всі знайдені результати без співпадінь і відкладає їх для алгоритмів нечіткого пошуку, а сам здійснює пошук по значеннях отриманих елементів. Для кожного слова отриманого з початкової вибірки будується множина значень на основі якої і відбувається реалізація пошуку подальших слів. Цей алгоритм по суті розділяє вибірку на 2 підвибірки, одна з яких буде оброблюватися основними алгоритмами нечіткого пошуку, а інша буде оброблена моделлю Noisy Channel. Алгоритм додатково модифіковано для генерації проміжних варіантів опрацювання тексту з використанням спеціальних правил з відкиданням закінчень слів, щоб уникнути неправильного розподілу на підвибірки через помилку в структуруванні слова з самого початку аналізу тексту. Дана модифікація дозволяє значно оптимізувати роботу алгоритму, так як в цьому випадку можна уникнути використання повних колекцій даних для зберігання тексту, а динамічно подавати дані окремими елементами. Блок-схема алгоритму подана на рис. 6.

Розподіливши дані на 2 підвибірки береться перша підвибірка і для її обробки застосовується алгоритм N-грам. Сам по собі алгоритм є досить старий і використовується вже давно, так як він є досить простий у реалізації, легкий у модифікації відповідно до певних унікальних вимог, та досить швидко працює, тим самим він є найшвидшим алгоритмом нечіткого пошуку серед тих, який використовуються в інформаційній системі соціалізації особистостей за спільними інтересами.

Якщо описувати роботу алгоритму, то він бере сформований масив даних, порівнює його поелементно з порівнюваним масивом даних. Саме порівняння здійснюється за дуже простою формулою, якщо слово 1 співпадає зі словом 2 з врахуванням деяких помилок, то є великий шанс того, що в них буде спільний рядок довжиною N. Блок-схема алгоритму подана на рис. 7.

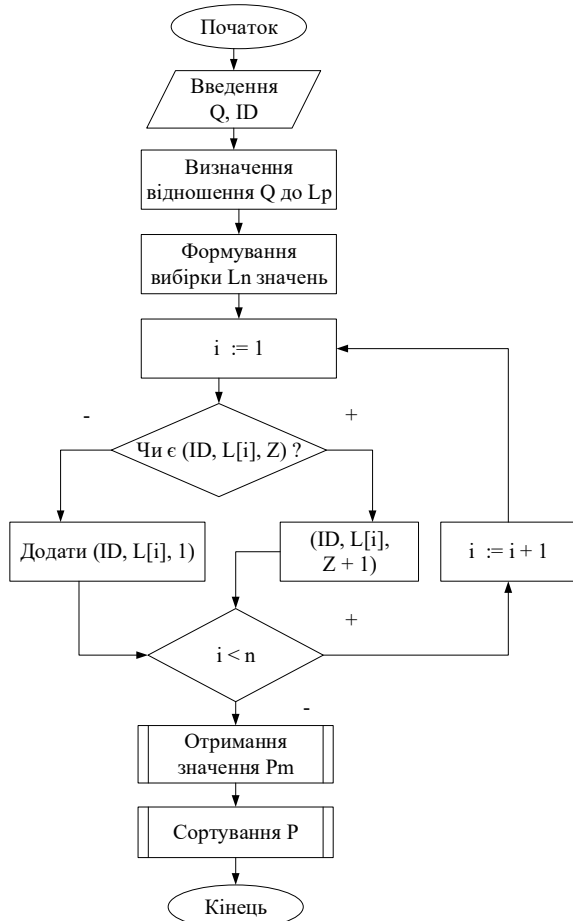


Рисунок 6 – Алгоритм розширення вибірки

Під час індексації потрібне слово розбивається на такі N-грами, щоб середній шанс схожості співпадіння слова був більше 50% і попадає в окремий список слів, найчастіше використовуються триграми, тобто рядки, які складаються з 3 букв. Також оптимізовано пошук і під час обробки даних, дані подаються по декілька елементів, тим самим не створюючи окремий список даних, а просто формуючи кінцевий список.

Також даний алгоритм має певні недоліки, наприклад при створенні всіх триграм рядків є безліч слів, які можуть випасти з вибірки, навіть якщо підходили по параметрам пошуку слова, так як може попасти слово, яке не відповідає поділу і буде помилково не вибрано триграмами, тому саме перед цим використано алгоритм розширення вибірки, щоб мінімізувати можливість наявності подібних слів в першій підвибірці.

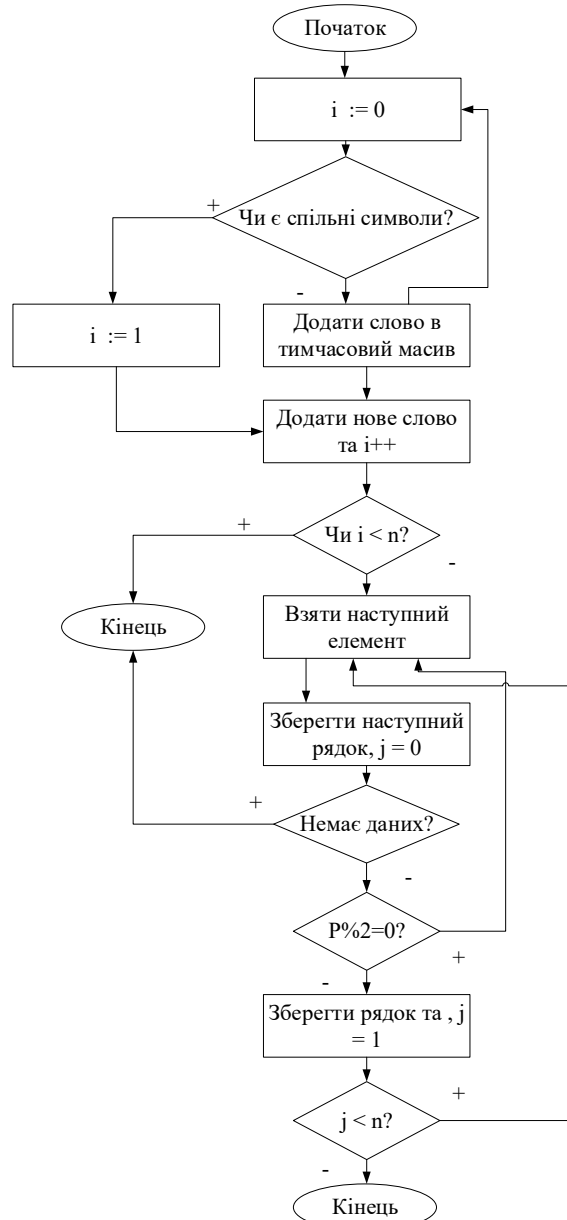


Рисунок 7 – Алгоритм N-грам

Вважається, що більшість таких слів попали саме в другу вибірку, тому відносно неї використовується модель Noisy Channel, її особливість в тому, що вона призначена для опрацювання всіх слів, які скоріше за все були певним чином спотворені, і розуміє під цим створення додаткового словника можливих значень, який і буде ще одним місцем для зберігання та порівняння слів з другої підвибірки, таким чином формується колекція можливих слів, які були спотворені. Сама модель Noisy Channel є дуже повільною у використанні через постійне формування проміжного словника під час обробки тексту, тому для її оптимізації і створено підвибірку слів, які скоріше за все спотворено, щоб не витратити зайві ресурси. Блок-схема моделі Noisy Channel подана на рис. 8.

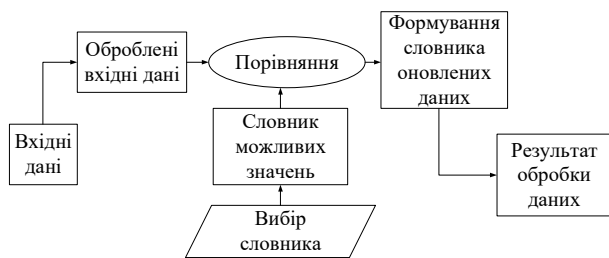


Рисунок 8 – Модель Noisy Channel

Робота моделі Noisy Channel показана на прикладі виправлення слова «actress», суть полягає в підстановці схожих слів за допомогою певних операцій обробки слова, при відсутності помилкових літер обрано правильне слово (табл. 2).

Таблиця 2 – Приклад роботи моделі Noisy Channel

Помилка	Можливе виправлення	Правильна літера	Помилкова літера	Тип
acress	actress	t	Немає	Видалення
acress	cress	–	a	Вставка
acress	caress	ca	ac	Транспозиція
acress	access	c	r	Віднімання
acress	across	o	e	Віднімання
acress	acres	–	s	Вставка
acress	acres	–	s	Вставка

Внутрішнім алгоритмом обробки повністю сформованих вибірок слів є алгоритм Левенштейна. Основна суть алгоритму в тому, що він визначає відстань між декількома послідовностями символів.

Таким чином алгоритм на виході визначає певне значення, а саме кількість необхідних замін для того, щоб одне слово співпало з порівнюваним словом у випадку знаходження відмінностей та формує відсоток, який відображає шанс однаковості слів, за допомогою якого можна визначати чи у вибірку попало два аналогічних слова і є співпадіння, чи два слова абсолютно різні і не попадають в кінцеву вибірку. Сам алгоритм оптимізовано за допомогою його побудови з використанням кінцевого автомату, таким чином всі проміжні дані в пам'яті видаляються зі зміною ітерації. Відповідно, спочатку відбувається формування 2 підвибірок слів з вхідних даних користувачів, кожна з двох отриманих підвибірок додатково опрацьовується для збільшення точності пошуку слів, після чого за допомогою алгоритму Левенштейна йде безпосереднє порівняння вибірок слів усіх користувачів системи, кожне слово поточного користувача порівнюється з відповідними параметрами досліджуваного користувача. Блок-схема алгоритму подана на рис. 9.

В ході порівняння формуються за кожне співпадіння формуються бали, за співпадіння інтересів – 3 бали, параметру пошуку – 2 бали, загальної інформації – 1 бал. Перед початком роботи алгоритму Левенштейна формується максимальний бал поточного користувача, так як всі зайві слова вже викинуті і можна вважати, що бал сформований на основі ключових слів. Робота алгоритму Левенштейна

на прикладі порівняння слів «elephant» та «relevant» подана в табл. 3.

Після роботи алгоритму при співпадінні кожного слова в користувачів нараховуються бали, які з кожною ітерацією сумуються.

Останнім кроком є порівняння сумарного балу поточного користувача та балів досліджуваного користувача, після чого формується відсоткове співвідношення схожості користувачів. Отримавши готову вибірку користувачів, можна здійснювати взаємодію з користувачами і інформаційна система соціалізації особистостей за спільними інтересами надає для цього всі необхідні функції, які інтуїтивно зрозумілі для звичайного користувача.

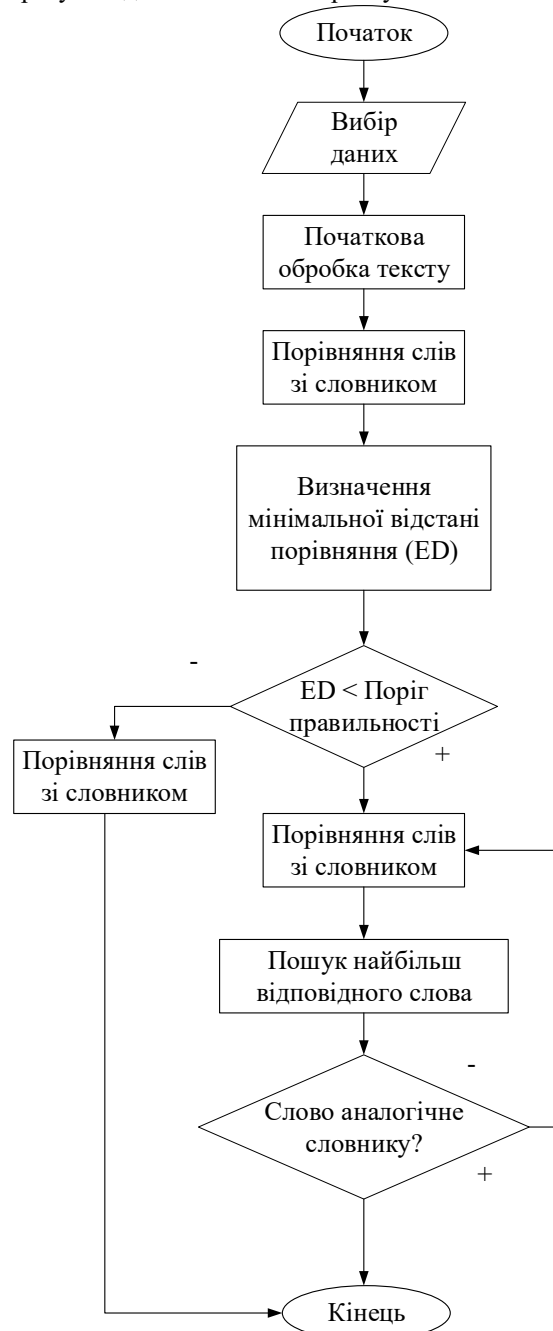


Рисунок 9 – Алгоритм Левенштейна

Таблиця 3 – Приклад роботи алгоритму Левенштейна

		E	L	E	P	H	A	N	T
	0	1	2	3	4	5	6	7	8
R	1	1	2	3	4	5	6	7	8
E	2	1	2	2	3	4	5	6	7
L	3	2	1	2	3	4	5	6	7
E	4	3	2	1	2	3	4	5	6
V	5	4	3	2	2	3	4	5	6
A	6	5	4	3	3	3	3	4	5
N	7	6	5	4	4	4	4	3	4
T	8	7	6	5	5	5	5	4	3

Так метод `GetUsers` призначений для виведення сформованого списку користувачів системи, спочатку здійснюється пошук поточного користувача, щоб його ж профіль не кидати у вибірку, після чого його `id` виключається з загальної вибірки.

За допомогою директиви `Response.AddPaging()` формується відповідь сервера на запит з вибірки користувачів. В відповідь передається поточна сторінка для відображення користувачів, розмір сторінки, кількість елементів та загальна кількість сторінок. Метод повертає сформовану вибірку користувачів відповідно до початкового параметра, відносно якого і формувалася вибірка з усіма сформованими даними сторінки.

Наступним йде метод `GetUser`, який повертає конкретного користувача для перегляду особистого профілю даного користувача. Метод приймає унікальний ідентифікатор користувача та повертає знайденого користувача у поточній базі даних, а повертає відповідь про те, що сталася помилка і користувача з таким ідентифікатором не існує.

Далі йде метод `UpdateUser`, який в якості параметрів приймає ідентифікатор користувача, дані про якого потрібно оновити, та об'єкт, який містить всі оновлені дані, надісланий з клієнта програми. В першу чергу перевіряється наявність поточного користувача в базі даних та перевірка, чи користувач авторизований в систему. Далі відбувається звертання до бази даних та асинхронно оновлюється інформація про користувача, у випадку помилки зберігання в базі даних викидається виняток, також код та причина помилки логуються всередині бази даних та зберігаються у відповідний список. Наступним йде метод `LikeUser`, який призначений для виставлення позначки вподобання даного користувача. Спочатку перевіряється чи авторизований користувач системи, якщо так, то перевіряється чи немає виставленої позначки поточний користувач, якщо має, то викидається повідомлення про помилку і неможливість здійснити цю саму дію ще раз. В іншому випадку, якщо помилок немає, створюється новий об'єкт `Like`, та додається в базу даних і відправляється запит користувачу, якому цю позначку виставили.

Наступним йде метод `CreateMessage` для надсилання повідомлень, він приймає 2 основні параметри, це ідентифікатор відправника і об'єкт самого повідомлення. Здійснюється пошук відправника в базі даних, далі в разі успішного

пошуку ідентифікатор відправника зберігається і шукається приймач повідомлення за допомогою ідентифікатора в базі даних. В разі успішного пошуку формується саме повідомлення та всі властивості повідомлення, які потрібно зберегти записуються в спеціальну змінну, яка є об'єктом повідомлення для зберігання даних. Спочатку повідомлення зберігається асинхронно в базу даних, далі формується анонімний об'єкт, який за допомогою механізму маршрутизації відправляється в діалог двох користувачів. Всі дані зберігаються в таблицях в реляційній базі даних, всі наявні дані пов'язані між собою за допомогою зв'язків. Схема поточної бази даних подана на рис. 10.

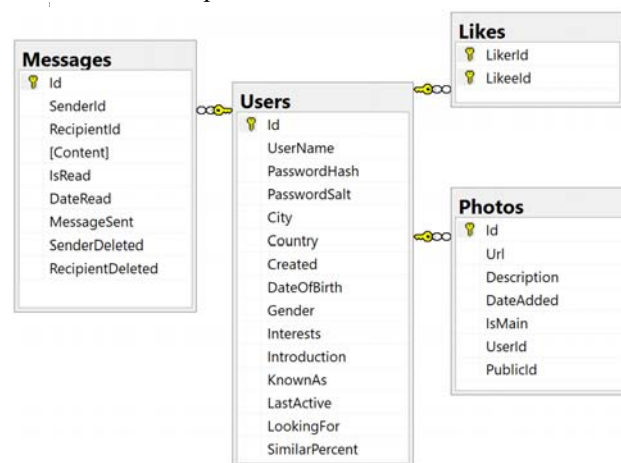


Рисунок 10 – Схема бази даних IC

В першу чергу варто описати ключову таблицю бази даних – `Users`, яка представляє поточного користувача системи, вона має такі основні поля: `Id` – унікальний ідентифікатор користувача, `UserName` логін користувача, `PasswordHash` – хеш паролю користувача, `PasswordSalt` – образ хешу користувача, `Gender` – стать користувача, `DateOfBirth` – дата народження, `KnownAs` – нікнейм користувача всередині системи, `Created` – дата реєстрації, `LastActive` – дата останніх відвідин користувача, `Introduction` – вступна інформація про користувача, `LookingFor` – основна інформація про те, що саме користувачу потрібно і що він шукає, `Interests` – опис інтересів користувача системи, `City` – місто проживання, `Country` – країна проживання, `SimilarPercent` – тимчасовий параметр схожості користувача. Також, як видно зі схеми бази даних, таблиця `Users` пов'язана з іншими таблицями: `Photos`, `Likes` та `Messages` за допомогою зв'язку «один до багатьох», відповідно `Users` є центральною таблицею.

Далі йде таблиця `Photos`, в якій є всі основні властивості фотографій користувачів інформаційної системи, а саме `Id` – унікальний ідентифікатор, `Url` – посилання на картинку, `Description` – опис, `DateAdded` – дата додавання, `IsMain` – булеве поле, яке визначає аватарку користувача, `PublicId` – ідентифікатор знаходження картинки в хмарному сервісі зберігання даних та `UserId` – зовнішній ключ, який означає що



картинка належить одному з користувачів. Далі таблиця Likes, в якій є ідентифікатори користувачів, які виставляли лайки, а саме LikerId – унікальний ідентифікатор користувача, який поставив лайк та LikeeId – унікальний ідентифікатор користувача, якому поставлено лайк в системі.

Далі йде таблиця Messages, в якій є всі основні властивості повідомлень надісланих користувачами інформаційної системи, а саме Id – унікальний ідентифікатор повідомлення, SenderId – унікальний ідентифікатор відправника, RecipientId – унікальний ідентифікатор отримувача, Content – зміст повідомлення, IsRead – позначка чи прочитане повідомлення, DateRead – дата прочитання, MessageSent – дата відправки, SenderDeleted – чи видалене відправником, RecipientDeleted – чи видалене отримувачем.

## 5 РЕЗУЛЬТАТИ

Контрольний приклад відображає основні функції і роботу створеної інформаційної системи, на рис. 11 подано головне вікно програми.

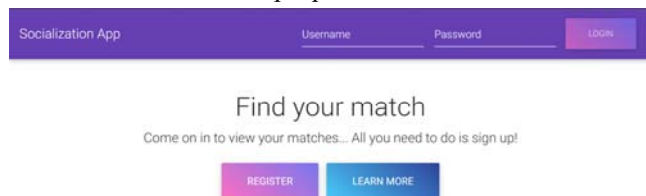


Рисунок 11 – Головне вікно програми

На рис. 12 подано кнопки головного вікна ІС.

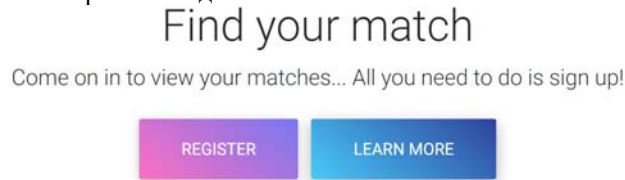


Рисунок 12 – Кнопки головного вікна програми

На рис. 13 подано форму реєстрації користувача.

Рисунок 13 – Форма для реєстрації користувача

На рис. 14 подано авторизацію користувача, введення логіну та паролю, на рис. 15 подано повідомлення про успішну авторизацію.



Рисунок 14 – Авторизація користувача

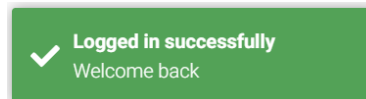


Рисунок 15 – Повідомлення про успішну авторизацію

На рис. 16 подано параметри профілю користувача, на рис. 17 подано заповнений профіль користувача.

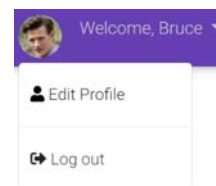


Рисунок 16 – Параметри користувацького профілю

### Your Page

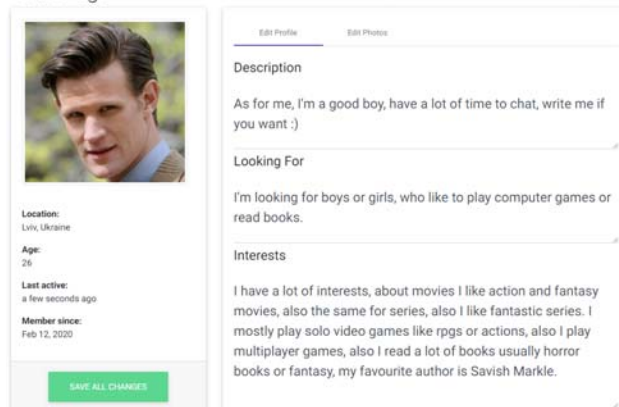


Рисунок 17 – Заповнений профіль користувача

На рис. 18 подано процес завантаження фотографій в систему, можна завантажувати одночасно 1 і більше фотографій перетягнувши їх вручну, або за допомогою провідника.



Рисунок 18 – Завантаження фотографій

На рис. 19 подано завантажені фотографії користувача, можна видалити всі фотографії, крім поточної головної фотографії та нейронні мережі обробили всі фотографії, і ті, на яких не знайдено лиця недоступні для виставлення основними фотографіями користувача.

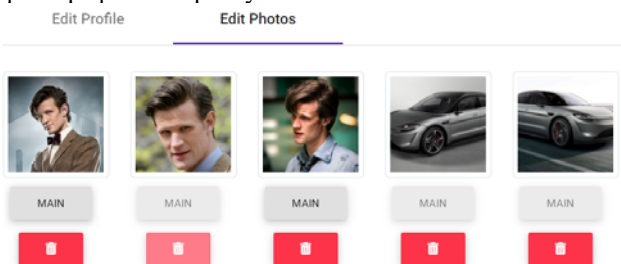


Рисунок 19 – Завантажені фотографії користувача

На рис. 20 подано сформований список користувачів за допомогою алгоритмів обробки тексту та посортований по спаданню відсоткового співвідношення схожості користувачів.

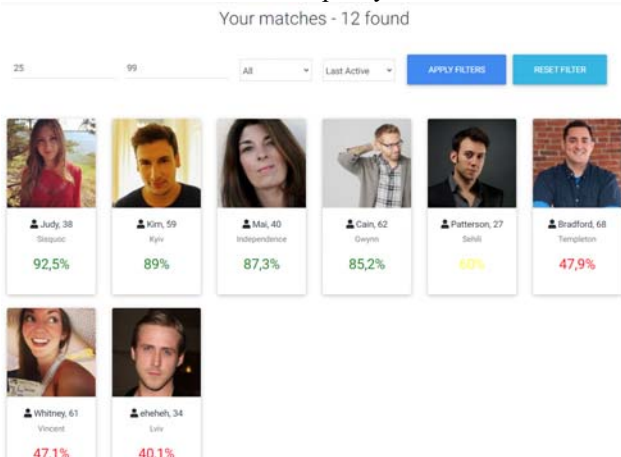


Рисунок 20 – Сформований список користувачів

На рис. 21 подано використання фільтрів для пошуку в вже сформованому списку.

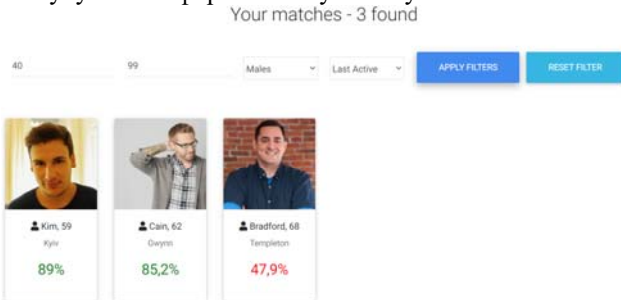


Рисунок 21 – Фільтрація списку

На рис. 22 подано вибір профілю користувача, видно можливість переглянути профіль користувача, поставити лайк і написати приватне повідомлення.

На рис. 23–24 подано перегляд вкладки з інформацією про вподобання користувачів, які вибрали нас, і яких вибрали ми.

На рис. 25–27 подано основну інформацію профілю вибраного користувача, вкладку з інтересами

користувача, та вкладку з усіма фотографіями користувача.

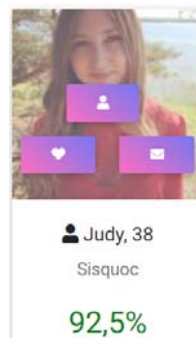


Рисунок 22 – Вибір користувача

Members who like me : 1

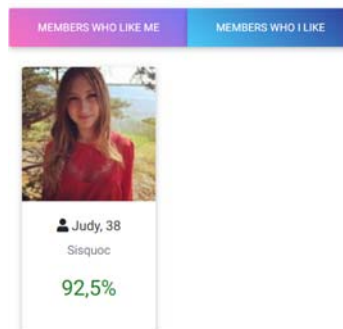


Рисунок 23 – Користувачі, які вибрали нас

Members who I've Liked : 3

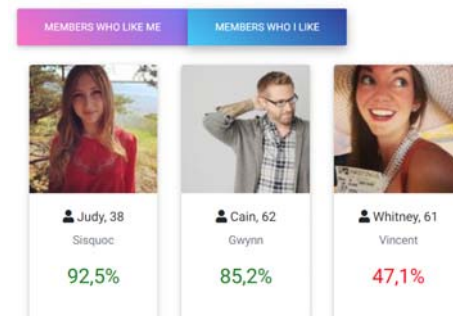


Рисунок 24 – Користувачі, яких вибрали ми

Judy's Page

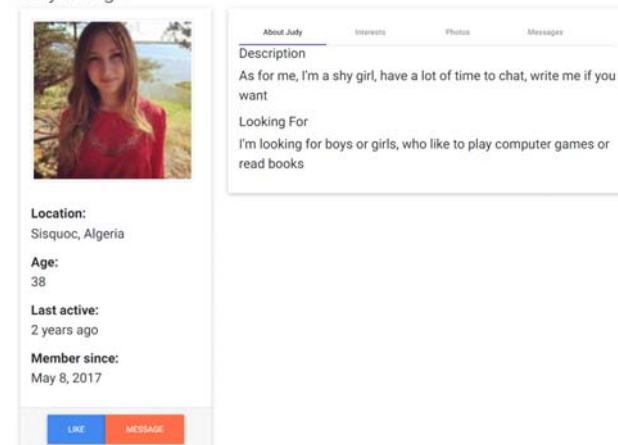


Рисунок 25 – Основна інформація про користувача



Рисунок 26 – Інтереси користувача

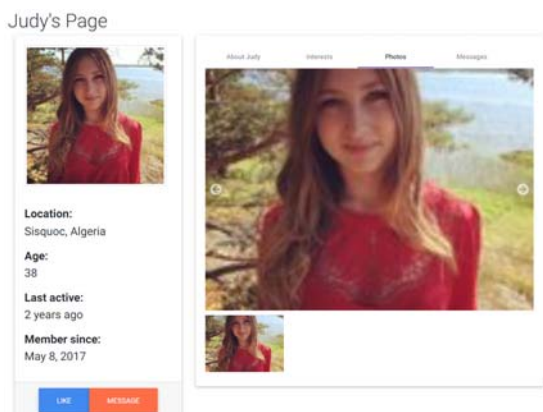


Рисунок 27 – Фотографії користувача

На рис. 28 подано вкладку з приватною перепискою з користувачем. В переписці подано ніки користувачів, фотографії, час надіслання і час прочитання повідомлень. На рис. 29–30 подано сторінку з інформацією про всі повідомлення, подано непрочитані, отримані та надіслані, можна здійснювати керування повідомленнями, а саме перегляд вибраного повідомлення за допомогою переходу в діалог з користувачем, або видалення вибраного свого повідомлення для всіх, або чужого повідомлення лише для себе.

На рис. 31–32 подано вхід з профілю іншого користувача, якого обрано першим користувачем системи та перегляд списку користувачів, які нас обрали, що дозволяє почати приватну переписку між двома користувачами. Що вибрали один одного.

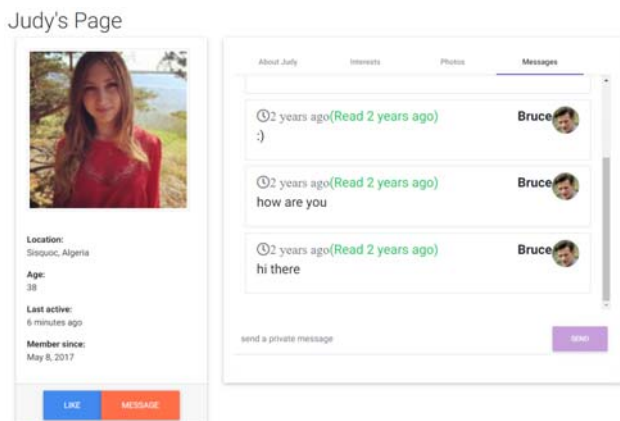


Рисунок 28 – Приватна переписка з користувачем

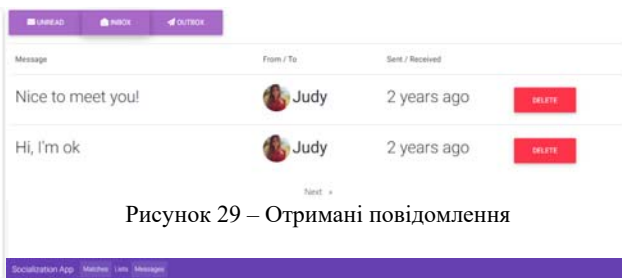


Рисунок 29 – Отримані повідомлення



Рисунок 30 – Надіслані повідомлення

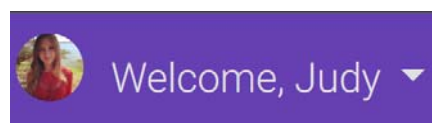


Рисунок 31 – Вхід іншого користувача в систему

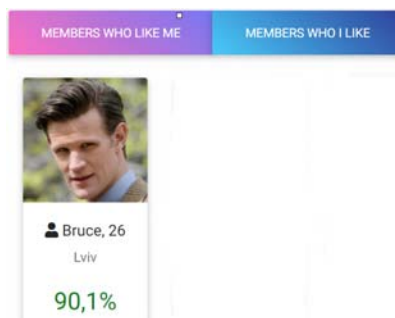


Рисунок 32 – Користувачі, що вибрали поточного користувача

На рис. 33 подано приватну переписку з початковим користувачем, від лица вибраного користувача системи.

Bruce's Page

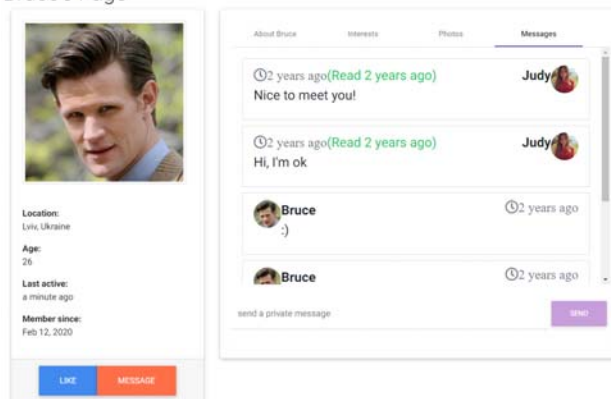


Рисунок 33 – Приватна переписка користувачів

## 6 ОБГОВОРЕННЯ

Отримавши реалізовану ІС здійснено статистичний аналіз 2 параметрів системи, а саме порівняння швидкості виконання формування вибірки

користувачів та точність отриманого відсоткового співвідношення. Порівнювались відповідно реалізована в інформаційній системі комбінація алгоритмів Левенштейна, N-грам, розширення вибірки, моделі Noisy Channel та звичайного алгоритму Левенштейна, який найчастіше використовують в схожих системах соціалізації особистостей. Спочатку проаналізовано ефективність формування вибірки користувачів, обрано 12 користувачів системи і кожному формувалась вибірка з використанням комбінації алгоритмів, та з використанням звичайного алгоритму Левенштейна, отримана діаграма подана на рис. 34, з якої можна зробити висновок, що реалізована в системі комбінація алгоритмів є ефективнішою та точнішою приблизно на 25–30% в порівнянні зі звичайним алгоритмом Левенштейна.



Рисунок 34 – Відсоткове співвідношення між користувачами

Далі проаналізовано швидкість формування вибірки користувачів, знову обрано 12 користувачів системи і кожному формувалась вибірка з використанням комбінації алгоритмів, та з використанням звичайного алгоритму Левенштейна, отримана діаграма подана на рис. 35, з якої можна зробити висновок, що реалізована в системі комбінація алгоритмів здійснює вибірку приблизно в 10 разів швидше, ніж звичайний алгоритм Левенштейна.



Рисунок 35 – Час формування вибірки даних

## ВИСНОВКИ

В наш час соціалізація особистостей за спільними інтересами є надзвичайно важливим процесом, так як більшість людей намагаються спростити та автоматизувати всі основні життєві процеси, які

завичай займають багато вільного часу, те саме стосується і процесу соціалізації, створена інформаційна система в цьому плані грає важливу роль. ІС розроблена з використанням алгоритмів нечіткого пошуку по словах з використанням моделі Noisy Channel з алгоритмами ефективного розподілу текстової інформації, та з використанням згорткової нейронної мережі для ідентифікації користувачів системи, так як на даний момент немає такої системи, яка б здійснювала аналіз вказаної про користувача інформації та формувала б список найбільш релевантних користувачів.

Створення ІС соціалізації особистостей є актуальним завданням, так як в сучасному світі люди намагаються оптимізувати всі життєві процеси для економії часу. Користувачі при пошуку тих чи інших програм в першу чергу вибирають ті, які економлять час, оптимізують роботу і достатньо автоматизовані, щоб виконувати більшість дій замість користувача системи. Дана інформаційна система об'єднує в собі зразу два важливих завдання: соціалізацію користувачів та максимально оптимізує і автоматизує сам процес соціалізації.

Якщо здійснити економічну оцінку, то розробка даного ПЗ є вигідною, оскільки створювана ІС пропонує унікальні можливості в плані пошуку та аналізу користувачів, відповідно крупні компанії з розробки програмної продукції будуть зацікавлені в покупці даного програмного продукту. Якщо провести маркетингову оцінку, то ІС даного типу є актуальною серед користувачів, оскільки в першу чергу вона надає можливості, які відсутні в схожих за функціоналом програмних продуктах, є легкою та зручною у використанні і пропонує повністю новий підхід до процесу пошуку, аналізу та соціалізації користувачів всередині системи.

Найважливішим кроком була практична реалізація інформаційної системи соціалізації особистостей за спільними інтересами, в першу чергу здійснено написання простого та надійного функціоналу реєстрації та подальшої авторизації користувача системи з використанням методів Identity та JWT токенів, що дозволило надійно зберігати паролі користувачів в базі даних та оптимізовано створювати сесію та надавати весь необхідний функціонал під час роботи користувача в системі. Далі здійснено реалізацію роботи двох нейронних мереж: згорткової та сіамської, що дозволило здійснити пошук людського лица, на фотографіях що завантажують користувач і порівняти знайдене лице з уже наявними в базі даних лицами, що дає можливість ефективно ідентифікувати справжність користувача та гарантувати, що цього користувача на даний момент нема в базі даних, відповідно він є реальним.

За допомогою алгоритмів нечіткого пошуку, алгоритму Левенштейна та моделі Noisy Channel створено алгоритм аналізу та порівняння користувачької інформації, який для поточного користувача формує список наявних користувачів

системи, посортований по спаданню відсоткового співвідношення схожості користувачів та вказує, наскільки інтереси в інших користувачів збігаються з інтересами поточного користувача.

В ході виконання роботи в більшості були досягнуті основні цілі відповідно до поставленої мети, існує ще багато речей, які можна удосконалити всередині створеної інформаційної системи та удосконалити, але той функціонал, який є доступним всередині інформаційної системи на даний момент відповідає поставленій меті створення інформаційної системи і є реалізована можливість усім користувачам системи пройти реєстрацію, вказати про себе всі необхідні дані, завантажити фотографії зі своїм лицем і після цього здійснювати ефективну соціалізацію, а саме переглядати користувачів зі сформованого списку, ставити відмітку про вподобання користувача та здійснювати приватну переписку всередині системи з вибраним користувачем.

До наукової новизни одержаних результатів варто віднести розроблення нового алгоритму аналізу користувацької інформації та пошуку найбільш релевантних користувачів ІС відповідно до проаналізованого тексту повідомлень профілю на основі вже існуючих алгоритмів Левенштейна, розширення вибірки, N-грам та моделі Noisy Channel. Для створення динамічної ІС соціалізації використано шаблон асинхронного програмування. Удосконалено згорткову нейронну мережу, що дозволило ефективно здійснювати пошук людських обличчя на фото та перевіряти наявність вже існуючих людей в БД ІС.

Практичне значення створеної ІС соціалізації особистостей є дуже важливим, оскільки вона виконує важливу функцію соціалізації особистостей згідно потреб сучасних користувачів соціальних мереж. Також ІС має важливе значення в плані інновацій, так як на даний момент не існує аналогічних систем для соціалізації особистостей, які б використовували розроблені алгоритми. Система дозволить ефективно та швидко здійснювати підбір, аналіз, опрацювання текстових даних та формування кінцевого результату. В системі використовуються SEO-технології для ефективного та якісного інтелектуального пошуку та опрацювання відповідних даних за потребою конкретного користувача. Нейронна мережа дозволяє ефективно здійснювати ідентифікацію користувача по його фото. Загалом використовувані алгоритми дозволяють створити зручну ІС соціалізації з використанням необхідних для цього алгоритмів.

Варто зазначити важливість оптимізації наявної в ІС, в першу чергу це повна асинхронність системи, що дозволить уникнути всіх довгих очікувань та важких в плані опрацювання та аналізу запитів, система дозволяє ефективно та динамічно працювати з різними обсягами великих даних, здійснювати їх аналіз, опрацювання та формування нових даних необхідних користувачам ІС. Також використовується хмарний сервіс, який дозволить здійснити розподіл

даних, відповідно можна буде зберігати всі найбільш важкі дані в хмарному середовищі і з використанням простого програмного інтерфейсу ІС за допомогою запитів здійснювати завантаження всіх необхідних даних. Таким чином, можна стверджувати, що створення даної ІС є важливим як і в соціальному плані, так і в плані реалізації всіх алгоритмів, які забезпечують необхідний функціонал ІС.

## ПОДЯКИ

Роботу виконано в рамках держбюджетної теми «Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій» (ID:839 2017-05-15 09:20:01 (2459-315)). Дослідження провадилося в межах спільних наукових досліджень кафедри інформаційних систем та мереж НУ «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства». Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ІСМ НУ «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

## ЛІТЕРАТУРА / LITERATURA

1. Chu S. C. Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter / S. C. Chu, Y. Sung // *Electronic Commerce Research and Applications*. – 2015. – Vol. 14(4). – P. 251–260. DOI: 10.1016/j.elerap.2015.04.002
2. De-Gregorio F. Understanding attitudes toward and behaviors in response to product placement: A Consumer Socialization Framework / F. De-Gregorio, Y. Sung // *Journal of Advertising*. – 2010. – Vol. 39(1). – P. 83–96. DOI: 10.2753/JOA0091-3367390106
3. Peer-based social media features in behavior change interventions: Systematic review / [S. M. R. A. Elaheebocus, M. Weal, L. Morrison, L. Yardley] // *Journal of Medical Internet Research*. – 2018. – Vol. 20(2). – P. 1–20. 10.2196/jmir.8342
4. Erkan I. The influence of e-WOM in social media on consumers' purchase intentions: An extended approach to information adoption / I. Erkan // *Computers in Human Behavior*. – 2016. – Vol. 4. – P. 47–55. DOI: 10.1016/j.chb.2016.03.003
5. Ferrara E. Online popularity and topical interests through the lens of Instagram / E. Ferrara, R. Interdonato, A. Tagarelli // *Hypertext and Social Media*. – 2014. – Vol. 2. – P. 24–23. DOI: 10.1145/2631775.2631808
6. Gao L. Online consumer behavior and its relationship to website atmospheric induced flow: Insights into online travel agencies in China / L. Gao, X. Bai // *Journal of Retailing and Consumer Services*. – 2014. – Vol. 21(4). – P. 653–655. DOI: 10.1016/j.jretconser.2014.01.001
7. Geurin-Eagleman A. N. Communicating via photographs: A gendered analysis of Olympic athletes' visual self – presentation on Instagram / A. N. Geurin-Eagleman,

- L. M. Burch // *Sport Management Review*. – 2016. – Vol. 19(2). – P. 133–145. DOI: 10.1016/j.smr.2015.03.002
8. From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies / [J. D. Guidry, M. Messner, Y. Jin, V. Medina-Messner] // *Corporate Communications: An International Journal*. – 2015. – Vol. 20(3). – P. 344–359. DOI: 10.1108/CCIJ-04-2014-0027
9. Hanna R. We're all connected: The power of the social media ecosystem / R. Hanna, A. Rohm, V. L. Crittenden // *Business Horizons*. – 2011. – Vol. 54(3). – P. 265–273. DOI: 10.1016/j.bushor.2011.01.007
10. Salganik M. J. Bit by bit: Social Research in the Digital Age / M. J. Salganik. – Princeton : Princeton University Press, 2019. – 448 p.
11. The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees / [S. Hudson, M. Roth, T. J. Madden, R. Hudson] // *Tourism Management*. – 2015. – Vol. 47. – P. 68–76. DOI: 10.1016/j.tourman.2014.09.001
12. Managing brand presence through social media: The case of UK football clubs / [M. Jeff, R. Jennifer, J. Catherine, P. Elke] // *Internet Research*. – 2014. – Vol. 24(2). – P. 181–204. DOI: 10.1108/IntR-08-2012-0154
13. Kim E. Brand followers' retweeting behaviour on Twitter: How brand relationship influence brand electronic word-of-mouth / E. Kim, Y. Sung, H. Kang // *Computers in Human Behavior*. – 2014. – Vol. 37. – P. 18–25. DOI: 10.1016/j.chb.2014.04.020
14. Kudeshia C. Spreading love through fan page liking: A perspective on small scale entrepreneurs / C. Kudeshia, P. Sikdar, A. Mittal // *Computers in Human Behavior*. – 2016. – Vol. 54. – P. 257–270. DOI: 10.1016/j.chb.2015.08.003
15. Lueg J. E. Interpersonal communication in the consumer socialization process: Scale development and validation / J. E. Lueg, R. Z. Finney // *Journal of Marketing Theory and Practice*. – 2007. – Vol. 15(1). – P. 25–39. DOI: 10.2753/MTP1069-6679150102
16. Mousavijad M. The effect of socialization factors on decision making of teenagers consumers in schools / M. Mousavijad, S. Payvandi // *Journal of School Administration*. – 2017. – Vol. 5(1). – P. 217–234.
17. Schnell R. Enhancing Surveys with Objective Measurements and Observer Ratings / R. Schnell // *Improving Survey Methods*. – London : Routledge, 2014. – P. 310–324.
18. Parry M. E. The effect of personal and virtual word-of-mouth on technology acceptance / M. E. Parry, T. Kawakami, K. Kishiya // *Journal of Product Innovation Management*. – 2012. – Vol. 29(6). – P. 952–966. DOI: 10.1111/j.1540-5885.2012.00972.x
19. Quan-Haase A. Introduction to the Handbook of Social Media Research Methods: Goals, Challenges and Innovations / A. Quan-Haase, L. Sloan // *The Sage Handbook of Social Media Research Methods*. – 2017. – Vol. 1. – P. 1–9.
20. Murphy S. T. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures / S. T. Murphy, R. B. Zajonc // *Journal of Personality and Social Psychology*. – 1993. – Vol. 64(5). – P. 723–739. DOI: 10.1037/0022-3514.64.5.723
21. Ngoma M. Word of mouth communication: A mediator of relationship marketing and customer loyalty / M. Ngoma, P. D. Ntale // *Cogent Business and Management*. – 2019. – Vol. 6(1). – A. 1580123. DOI: 10.1080/23311975.2019.1580123
22. Analyses of word-of-mouth communication and its effect on students' university preferences / [A. Ozdemir, B. Tozlu, E. Şen, A. Ateşoğlu] // *Procedia – Social and Behavioral Sciences*. – 2016. – Vol. 235. – P. 22–35. DOI: 10.1016/j.sbspro.2016.11.022
23. Park J. Style in the age of Instagram: Predicting success within the fashion industry using social media / J. Park, G. L. Ciampaglia, F. Ferrara // *Computer-Supported Cooperative Work & Social Computing : 19th ACM Conference, San Francisco, CA, USA, February 27 – March 2, 2016 : proceedings*. – ACM : Permissions@acm.org, 2019. – P. 64–73. DOI: 10.1145/2818048.2820065
24. The Intelligent System Development for Psychological Analysis of the Person's Condition / [O. Oborska, V. Andrunyk, L. Chyrun et al.] // *Computational Linguistics and Intelligent Systems (COLINS 2021) : 5th International Conference, Lviv, 22–23 April 2021 : CEUR workshop proceedings*. – Aachen: CEUR-WS.org, 2021. – Vol. 2870. – P. 1390–1419.
25. Ranjbaran B. A survey for identification of major factors influencing customers attitude toward machine made carpet brands / B. Ranjbaran, M. Jamshidian, Z. Dehghan // *Journal of Commercial Strategies*. – 2020. – Vol. 5(23). – P. 109–118.
26. Batiuk T. Intelligent System for Socialization by Personal Interests on the Basis of SEO-Technologies and Methods of Machine Learning / T. Batiuk, V. Vysotska, V. Lytvyn // *Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings*. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 1237–1250.
27. Vysotska V. Information Technology for Internet Resources Promotion in Search Systems Based on Content Analysis of Web-Page Keywords / V. Vysotska // *Radio Electronics, Computer Science, Control*. – 2021. – No 3. – P. 133–151. DOI: 10.15588/1607-3274-2021-3-12

Стаття надійшла до редакції 19.12.2021.

Після доробки 17.01.2022.

УДК 004.9

#### ТЕХНОЛОГИЯ СОЦИАЛИЗАЦИИ ЛИЧНОСТЕЙ ЗА ОБЩИМИ ИНТЕРЕСАМИ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ И SEO-ТЕХНОЛОГИЙ

**Батиук Т. М.** – студент кафедри «Інформаційні системи і мережі», Національний університет «Львівська політехніка», Україна.

**Висоцька В. А.** – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи і мережі», Національний університет «Львівська політехніка», Україна.

#### АННОТАЦИЯ

**Актуальность.** Социализация личностей по общим интересам вызвана потребностью большинства людей упростить часть жизненных моментов за счет уменьшения времени их реализации. С быстрыми темпами роста информации,

© Батиук Т. М., Висоцька В. А., 2022

DOI 10.15588/1607-3274-2022-2-6

загруженности человека в обществе и в связи с последними эпидемическими событиями человек становится изолированным от возможности общаться. А это одна из важных потребностей человеческого сознания и самореализации. Поэтому является актуальным спросом возможность получать рекомендованный список подобных людей по общим интересам как результат интеллектуального поиска множества релевантных пользователей социальных сетей через анализ фото человеческого лица на пользовательских фотографиях (на основе нейронных сетей) и анализ пользовательской информации (на основе алгоритмов нечеткого поиска и модели Noisy Channel).

**Целью** исследования является разработка технологии для социализации личностей на основе SEO технологии и метода машинного обучения через использование сверточной и сямской нейронных сетей для идентификации пользователей и алгоритмов анализа текста для подбора релевантных пользователей будущего общения.

**Метод.** При реализации SEO-технологий выбраны алгоритмы нечеткого поиска по словам на основе модели Noisy Channel с алгоритмами эффективного распределения текстовой информации. При реализации машинного обучения разработана сверточная нейронная сеть для идентификации пользователей системы.

**Результаты.** Разработана интеллектуальная система социализации личностей по общим интересам на основе SEO-технологии и методы машинного обучения. Осуществлена реализация работы двух нейронных сетей: сверточной и сямской, что позволило осуществить поиск человеческого лица, на загружаемых пользователем фотографиях и сравнить найденное лицо с уже имеющимися в базе данных/Интернет. Это позволяет эффективно идентифицировать подлинность пользователя и гарантировать, что этого пользователя на данный момент нет в базе данных, соответственно он потенциально реальный. С помощью алгоритмов нечеткого поиска, алгоритма Левенштейна и модели Noisy Channel создан алгоритм анализа и сравнения пользовательской информации, который для текущего пользователя формирует список имеющихся пользователей системы, рассортированный по убыванию процентного соотношения сходства пользователей и указывает, насколько интересы других пользователей совпадают с интересами текущего пользователя.

**Выводы.** Выявлено, что реализуемый в системе алгоритм для формирования выборки пользователей является более эффективным и точным примерно на 25–30% по сравнению с обычным алгоритмом Левенштейна. Также реализуемый алгоритм осуществляет выборку примерно в 10 раз быстрее, чем обычный алгоритм Левенштейна.

**КЛЮЧЕВЫЕ СЛОВА:** нечеткий поиск, алгоритм Левенштейна, модель Noisy Channel, сверточная нейронная сеть, сямская нейронная сеть, фотоанализ лица, алгоритм расширения выборки, алгоритм N-грамм.

UDC 004.9

#### TECHNOLOGY FOR PERSONALITIES SOCIALIZATION BY COMMON INTERESTS BASED ON MACHINE LEARNING METHODS AND SEO-TECHNOLOGIES

**Batiuk T.** – Student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

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

#### ABSTRACT

**Context.** The socialization of individuals with common interests is caused by the need of most people to simplify some of the moments of life by reducing the time for their implementation. With the rapid growth of information, the human workload in society and the recent epidemics of the world, people are becoming isolated from the opportunity to communicate. And this is one of the important needs of human consciousness and self-realization. Therefore, there is an urgent need to be able to obtain a recommended list of similar people of common interest as a result of intelligent search of many relevant users of social networks through analysis of human faces in user photos (based on neural networks) and analysis of user information based on fuzzy search algorithms and Noisy model. Channel).

**Objective** of the study is to develop technology for socialization of individuals based on SEO-technology and machine learning through the use of convolutional and Siamese neural networks to identify users and text analysis algorithms to select relevant users of future communication.

**Method.** In the implementation of SEO-technologies selected fuzzy word search algorithms based on the Noisy Channel model with algorithms for efficient distribution of textual information. During the implementation of machine learning, a convolutional neural network was developed to identify users of the system.

**Results.** An intelligent system of socialization of individuals by common interests based on SEO-technology and machine learning methods has been developed. The work of two neural networks was implemented: convolutional and Siamese, which allowed to search for a human face in photos uploaded by the user and compare the found face with those already available in the database / Internet. This makes it possible to effectively identify the authenticity of the user and ensure that this user is not currently in the database, so it is potentially real. Using fuzzy search algorithms, Levenstein's algorithm and the Noisy Channel model, an algorithm for analyzing and comparing user information was created, which for the current user forms a list of available users of the system, sorted by descending percentage of similarity and indicates how other users' interests coincide.

**Conclusions.** It was found that the algorithm implemented in the system for forming a sample of users is more efficient and accurate by about 25–30% compared to the usual Levenstein algorithm. Also, the implemented algorithm performs sampling approximately 10 times faster than the usual Levenstein algorithm.

**KEYWORDS:** fuzzy search, Levenstein algorithm, Noisy Channel model, convolutional neural network, Siamese neural network, facial photoanalysis, sample expansion algorithm, N-gram algorithm.

## REFERENCES

1. Chu S. C., Sung Y. Using a consumer socialization framework to understand electronic word-of-mouth (eWOM) group membership among brand followers on Twitter, *Electronic Commerce Research and Applications*, 2015, Vol. 14(4), pp. 251–260. DOI: 10.1016/j.elerap.2015.04.002
2. De-Gregorio F., Sung Y. Understanding attitudes toward and behaviors in response to product placement: A Consumer Socialization Framework, *Journal of Advertising*, 2010, Vol. 39(1), pp. 83–96. DOI: 10.2753/JOA0091-3367390106
3. Elaheebocus S. M. R. A., Weal M., Morrison L., Yardley L. Peer-based social media features in behavior change interventions: Systematic review, *Journal of Medical Internet Research*, 2018, Vol. 20(2), pp. 1–20. DOI: 10.2196/jmir.8342
4. 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, Vol. 4, pp. 47–55. DOI: 10.1016/j.chb.2016.03.003
5. Ferrara E., Interdonato R., Tagarelli A. Online popularity and topical interests through the lens of Instagram, *Hypertext and Social Media*, 2014, Vol. 2, pp. 24–23. DOI: 10.1145/2631775.2631808
6. Gao L., Bai X. 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, Vol. 21(4), pp. 653–655. DOI: 10.1016/j.jretconser.2014.01.001
7. Geurin-Eagleman A. N., Burch L. M. Communicating via photographs: A gendered analysis of Olympic athletes' visual self-presentation on Instagram, *Sport Management Review*, 2016, Vol. 19(2), pp. 133–145. DOI: 10.1016/j.smr.2015.03.002
8. Guidry J. D., Messner M., Jin Y., Medina-Messner V. From McDonalds fail to Dominos sucks: An analysis of Instagram images about the 10 largest fast food companies, *Corporate Communications: An International Journal*, 2015, Vol. 20(3), pp. 344–359. DOI: 10.1108/CCIJ-04-2014-0027
9. Hanna R., Rohm A., Crittenden V. L. We're all connected: The power of the social media ecosystem, *Business Horizons*, 2011, Vol. 54(3), pp. 265–273. DOI: 10.1016/j.bushor.2011.01.007
10. Salganik M. J. Bit by bit: Social Research in the Digital Age. Princeton, Princeton University Press, 2019, 448 p.
11. Hudson S., Roth M., Madden T. J., Hudson R. The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees, *Tourism Management*, 2015, Vol. 47, pp. 68–76. DOI: 10.1016/j.tourman.2014.09.001
12. Jeff M., Jennifer R., Catherine J., Elke P. Managing brand presence through social media: The case of UK football clubs, *Internet Research*, 2014, Vol. 24(2), pp. 181–204. DOI: 10.1108/IntR-08-2012-0154
13. Kim E., Sung Y., Kang H. Brand followers' retweeting behaviour on Twitter: How brand relationship influence brand electronic word-of-mouth, *Computers in Human Behavior*, 2014, Vol. 37, pp. 18–25. DOI: 10.1016/j.chb.2014.04.020
14. Kudeshia C., Sikdar P., Mittal A. Spreading love through fan page liking: A perspective on small scale entrepreneurs, *Computers in Human Behavior*, 2016, Vol. 54, pp. 257–270. DOI: 10.1016/j.chb.2015.08.003
15. Lueg J. E., Finney R. Z. Interpersonal communication in the consumer socialization process: Scale development and validation, *Journal of Marketing Theory and Practice*, 2007, Vol. 15(1), pp. 25–39. DOI: 10.2753/MTP1069-6679150102
16. Mousavijad M., Payvandi S. The effect of socialization factors on decision making of teenagers consumers in schools, *Journal of School Administration*, 2017, Vol. 5(1), P. 217–234.
17. Schnell R. Enhancing Surveys with Objective Measurements and Observer Ratings, *Improving Survey Methods*. London, Routledge, 2014, pp. 310–324.
18. 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, Vol. 29(6), pp. 952–966. DOI: 10.1111/j.1540-5885.2012.00972.x
19. Quan-Haase A., Sloan L. Introduction to the Handbook of Social Media Research Methods: Goals, Challenges and Innovations, *The Sage Handbook of Social Media Research Methods*, 2017, Vol. 1, pp. 1–9.
20. Murphy S. T., Zajonc R. B. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures, *Journal of Personality and Social Psychology*, 1993, Vol. 64(5), pp. 723–739. DOI: 10.1037/0022-3514.64.5.723
21. Ngoma M., Ntale P. D. Word of mouth communication: A mediator of relationship marketing and customer loyalty, *Cogent Business and Management*, 2019, Vol. 6(1), A. 1580123. DOI: 10.1080/23311975.2019.1580123
22. Ozdemir A., Tozlu B., Şen E., Ateşoğlu A. Analyses of word-of-mouth communication and its effect on students' university preferences, *Procedia – Social and Behavioral Sciences*, 2016, Vol. 235, pp. 22–35. DOI: 10.1016/j.sbspro.2016.11.022
23. Park J., Ciampaglia G. L., Ferrara F. Style in the age of Instagram: Predicting success within the fashion industry using social media, *Computer-Supported Cooperative Work & Social Computing, 19th ACM Conference, San Francisco, CA, USA, February 27 – March 2, 2016, proceedings*. ACM, Permissions@acm.org, 2019, pp. 64–73. DOI: 10.1145/2818048.2820065
24. Oborska O., Andrunyk V., Chyrun L., Hasko R., Vysotskyi A., Mushasta S., Petruchenko O., Shakleina I. The Intelligent System Development for Psychological Analysis of the Person's Condition, *Computational Linguistics and Intelligent Systems (COLINS 2021), 5th International Conference, Lviv, 22–23 April 2021, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2021, Vol. 2870, pp. 1390–1419.
25. Ranjbaran B., Jamshidian M., Dehghan Z. A survey for identification of major factors influencing customers attitude toward machine made carpet brands, *Journal of Commercial Strategies*, 2020, Vol. 5(23), pp. 109–118.
26. 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*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 1237–1250.
27. 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, No 3, pp. 133–151. DOI: 10.15588/1607-3274-2021-3-12