

**Т.М. Заболотня, к.т.н., доц.
А.В. Соколовська, магістрант**

*Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»*

Метод автоматизованого визначення наявності образливого вмісту текстових повідомлень у соціальних мережах

Проведено аналіз специфіки спілкування Інтернет-користувачів. Особливу увагу приділено спілкуванню в соціальних мережах та, зокрема, проблемі виявлення образливого вмісту в повідомленнях користувачів. Визначено основні характерні риси спілкування в соціальних мережах, а також проведено аналіз чинників, що можуть погіршувати точність автоматизованої детекції образливого вмісту. Проаналізовано етапи обробки природномовних текстових даних, зокрема, визначення образливого контенту, на предмет можливості їх модифікації для врахування особливостей спілкування в соціальних мережах та інших факторів, що можуть вплинути на кінцевий результат класифікації текстових даних. Визначено, що характерні риси повідомлень та особливості спілкування в соціальних мережах, такі як наявність символів та цифр, що можуть призвести до заплутування сприйняття повідомлення, наявність емодзі, контекст (на рівні зв'язків користувачів у соцмережах) та контекст (на рівні повідомлень у соціальних мережах), можуть бути враховані під час автоматизованого виявлення образливого вмісту в текстових повідомленнях у соціальних мережах. Визначено відповідні модифікації етапів передоброблення тексту та аналізу результатів класифікації. Запропоновано метод автоматизованого визначення образливого вмісту текстових повідомлень у соціальних мережах, що базується на засадах машинного навчання з модифікованим підходом до передоброблення природномовних текстових даних. Запропоновано формат вхідних даних для ефективного визначення в них образливого вмісту.

Ключові слова: образливий вміст; оброблення текстових даних; соціальні мережі; машинне навчання; передоброблення природномовних текстових даних.

Постановка проблеми. Протягом останніх років у світі значно зросла кількість людей, які постійно користуються Інтернет [1]. Можливості всесвітньої мережі широко використовуються у різних сферах діяльності людини, а особливої популярності віднедавна здобули соціальні мережі.

Сьогодні соціальні мережі застосовуються для особистого спілкування, ведення блогів, реклами та, навіть, ведення бізнесу. Майже кожна компанія, від маленьких стартап-проектів та невеличких крафтових виробництв до величезних корпорацій та лідерів індустрії, має свою сторінку у соціальній мережі. Соціальні мережі стали місцем спілкування та об'єднання людей за інтересами. Проте там, де є велике скупчення людей та панує плюралізм думок, можливе виникнення палкої дискусії та поява образливого контенту.

Зважаючи на той факт, що сучасні соціальні мережі – це не лише майданчик для спілкування, а й платформа для ведення бізнесу, поява образливого контенту, який жодним чином не модерується, може призвести до втрати рекламодавців та зниження прибутків самої соціальної мережі. Рекламодавці ж не бажають, щоб їх продукт асоціювався з негативними емоціями чи скандалами, які можуть вплинути на подальшу популярність товарів та прибутки. Саме тому соціальні мережі та рекламодавці зацікавлені у модератії даних повідомлень.

Формулювання проблеми. Такі гіганти ринку як Facebook, Twitter, YouTube та інші подібні платформи впроваджують відповідні політики безпеки, щоб захистити користувачів від різноманітних ризиків, наприклад, цькування, дискримінації, проявів агресії [2]. Проте в останні роки стрімке зростання чисельності Інтернет-аудиторії та відповідне збільшення кількості користувачів соціальних мереж, які продукують тисячі нових коментарів та постів щохвилино, відчутно ускладнило задачу модератії цих повідомлень, призвело до збільшення фінансових та часових витрат на їх оброблення, а часом і взагалі унеможливило ручний аналіз їх вмісту. Таким чином, розроблення методів, алгоритмів, а також програмних засобів автоматизованого визначення образливого вмісту текстових повідомлень у соцмережах є актуальною задачею сьогодні.

Аналіз досліджень та публікацій. У більшості випадків відвідувачі соціальних мереж обмінюються особистими повідомленнями, пишуть пости на своїх приватних чи публічних сторінках, коментують щось на сторінках інших користувачів чи різноманітних спільнот. Перший пункт із наведеного вище переліку в контексті виявлення образливого вмісту розглядати не доцільно, адже приватні повідомлення

не мають перевірятися, бо це порушує конфіденційність розмови. У всіх інших випадках модерація коментарів та постів є доцільною.

Варто зазначити, що існуючі сьогодні підходи до модерації повідомлень та відповідні методи є недостатньо ефективними за критеріями точності та швидкості оброблення даних, а також вимагають значних фінансових затрат. Розглянемо їх більш детально.

По-перше, наразі активно використовується ручна модерація. Недоліками даного підходу є: низька швидкість оброблення даних, висока вартість даного процесу та незадовільна якість отримуваних результатів (різні люди можуть по-різному трактувати як образливий вміст, так і правила, за якими має здійснюватися модерація). Ще одним із підходів так званої «ручної модерації» є система скарг інших користувачів. Даний підхід має ключовий недолік – недобросовісні скарги (користувач може поскаржитися на будь-що, навіть на цілком нейтральний та необразливий пост чи коментар). Цим активно користуються зловмисники, наприклад, здійснюючи замовні атаки на сторінки відомих людей.

По-друге, сьогодні існують методи та способи автоматизованого визначення образливого вмісту текстових даних [3]. Недоліками даних розробок є те, що вони дають дуже загальну оцінку повідомленню щодо наявності в ньому образливого вмісту (текст образливий чи нейтральний) та не беруть до уваги специфіку спілкування у соціальних мережах. Якщо не враховувати дану специфіку, соціальні мережі з місця вільного висловлення думок можуть перетворитися на місця з обмеженням свободи слова та жорсткою цензурою, а відповідно, і втратять прихильність більшості аудиторії.

Виділення невіршених раніше частин загальної проблеми. Таким чином, на основі результатів проведеного дослідження існуючих підходів до модерації текстових повідомлень у соціальних мережах, а також відповідних методів та способів автоматизованого визначення образливого вмісту такого роду контенту можна зробити висновок, що підвищення ефективності процесу детекції образи у тому чи іншому повідомленні Інтернет-користувачів має базуватися, в першу чергу, на детальному вивченні специфіки публікацій Інтернет-користувачів у соціальних мережах та формулюванні переліку їх характеристик, які можуть бути враховані під час автоматизованого визначення наявності образливого вмісту, а також на обґрунтованому визначенні місця та способу врахування характеристик повідомлень у соцмережах, а також формату стосунків між різними типами та групами відвідувачів соціальних мереж у загальному процесі детекції образливого контенту.

Мета дослідження – підвищення точності визначення образливого вмісту текстових повідомлень у соціальних мережах шляхом розроблення нового методу визначення образливого вмісту в природномовних текстових даних, з урахуванням специфіки спілкування в соцмережах.

Викладення основного матеріалу. Розглянемо спочатку *особливості спілкування Інтернет-користувачів у соціальних мережах*, адже текстові повідомлення користувачів соцмереж мають певні особливості, що відрізняють їх серед інших видів природномовних текстових даних та ускладнюють задачу їх автоматизованого оброблення в цілому і визначення образливого вмісту зокрема.

Соціальна мережа – це структура, утворена індивідами або організаціями, яка відображає різноманітні зв'язки між ними через різні соціальні взаємовідносини, починаючи з випадкових знайомств і закінчуючи тісними родинними зв'язками. З 2001 року в Інтернет стали з'являтися сайти, в яких використовується технологія під назвою «Коло друзів». Ця форма соціальних мереж набула популярності і сьогодні широко використовується у віртуальних спільнотах (наразі існує більш ніж 200 сайтів з можливостями організації соціальних мереж). У 2004 році була створена найбільша на сьогодні соціальна мережа у світі Facebook [4].

Спілкування у соціальних мережах найчастіше можна зарахувати до розмовного стилю мовлення, в якому переважають ознаки побутового спілкування. Саме тому особливості побутового спілкування є релевантними і для спілкування в соціальних мережах, наприклад [5]:

1. Мова найчастіше є спонтанною, непередбаченою, відображає хвилинний настрій, відчуття. У ній виражена індивідуальність кожного суб'єкта.

2. Переважає діалог, хоча наявний полілог. Окремі репліки можуть видатись не зрозумілими поза текстом (характеризується терміном «ситуативна мова»).

3. Емоційний рівень спілкування надзвичайно високий.

4. Наявні умовчання, натяки, алюзії, посилання на події, відомі лише співбесідникам.

5. У широкому вжитку сленгові слова та скорочення.

Зупинимось на *характерних рисах повідомлень Інтернет-користувачів*, що утруднюють виявлення в них образливого вмісту.

По-перше, варто зазначити, що мова спілкування у соціальних мережах характеризується використанням постійно змінюваного сленгу та в цілому не обмежена жорсткими рамками, завдяки чому постійно поповнюється неологізмами та аббревіатурами. Також характерним є використання нецензурних слів, які у комбінації з різними допоміжними словами деколи можуть надавати словосполученню чи фразі кардинально нового відтінку та змісту (інколи навіть гостро-позитивного [6]). Крім того, деякі образи, що є неприпустимими для однієї групи людей, можуть сприйматися не як образа іншою групою.

По-друге, інколи використання тих чи інших слів та фраз може суперечити політиці безпеки соціальних мереж, тому для уникнення блокування недобросовісні користувачі соціальних мереж вдаються до навмисного заплутування слів і фраз. Такі вирази як «пi9 9er» чи відсутність пробілів у образливому реченні ускладнюють успішне виявлення образи, а інколи роблять вирішення даної задачі взагалі неможливим [3].

Ще однією проблемою при визначенні образливого вмісту в повідомленнях Інтернет-користувачів є врахування контексту, тобто де та чи інша потенційно образлива фраза була сказана. До так званої ситуативної мови можуть бути зараховані речення «Ця свиня товста.» та «Ти – товста свиня.», які мають абсолютно різне забарвлення (перше – констатація факту, друге – відверта образа) проте майже не відрізняються лексично. Також речення «Ми зобов'язані їх знищити!» теж не можна визначити як образливе, не маючи на руках контексту всього повідомлення (може йтися про міжнародну ворожнечу або ж про смертельний вірус, який закликають знищити). З точки зору специфіки повідомлень у соціальних мережах контекст може не обмежуватися лише одним повідомленням: це може бути низка повідомлень, що об'єднані спільною ідеєю, і лише при розгляді всього набору таких повідомлень можна визначити образливий вміст, позаяк окремо один від одного повідомлення можуть нести нейтральне забарвлення.

Ще однією особливістю спілкування у соціальних мережах та одночасно суттєвою перепоною при обробленні природномовних текстів в цілому і визначенні образливого вмісту зокрема є сарказм. Деякі користувачі публікують саркастичні коментарі в тому ж стилі, що і люди, які створюють образливий контент. Або навпаки пишуть все завуальовано і приємно, хоча мають на меті образу. Дуже важко, як для людей, так і для машин отримати правильне розуміння написаного, оскільки це вимагає знання сленгу та культури спільноти і, можливо, навіть самих учасників, які створюють даний контент.

Наведений перелік рис, характерних для спілкування Інтернет-користувачів у соціальних мережах, був би неповним, якби ми не згадали про емодзі – новий спосіб висловлення думок та надання повідомленням емоційного забарвлення. Емодзі (походить з японської мови від слів «картинка», «знак», «символ») – особлива мова ідеограм і смайлів, що широко використовують в електронних повідомленнях та на сторінках сайтів, яка отримала неймовірну популярність за останні роки та стала новою мовою для комп'ютерних комунікацій (Computer-Mediated Communication або СМС). Враховуючи те, що користувачі генерують великий обсяг емодзі, нещодавно були проведені дослідження для того, щоб зрозуміти способи використання емодзі через додатки [8], на різних платформах [9] та навіть у різних культурах [7]. Порівняно з традиційними інформаційними представленнями, такими як текстові повідомлення, картинки або навіть смайлики, емодзі вважаються більш жвакими, виразнішими та більш семантично насиченими, а саме тому цінуються користувачами Інтернету.

Сьогодні емодзі жодним чином не враховуються при аналізі природномовних текстових даних. Але зважаючи на популярність даного способу обміну повідомленнями в мережі Інтернет в цілому, та соціальних мережах зокрема, нехтувати емодзі при класифікації текстів є неправильним.

Також важливо наголосити на тому, що зазвичай при визначенні образливого вмісту повідомлень користувачів соцмереж опрацьовують лише сам текст повідомлення і не аналізують його контексту на рівні зв'язків користувачів у соціальних мережах [10], хоча, на думку авторів, врахування такого роду даних може сприяти зменшенню кількості необґрунтованих блокувань повідомлень на приватних сторінках, тим самим запобігаючи жорсткому цензуруванню повідомлень у соціальних мережах.

Звісно, взяти до уваги всі згадані вище особливості повідомлень Інтернет-користувачів під час визначення в них образливого вмісту за допомогою автоматизованого інструментарію просто неможливо (наприклад, сарказм), проте врахування частини з наведених вище ознак може підвищити якість визначення наявності образливого вмісту. Автори вважають за доцільне враховувати такі риси:

1. *Наявність символів та цифр, що можуть призвести до заплутування сприйняття повідомлення* – символи та цифри, що візуально сприймаються за існуючі літери, можуть негативно вплинути на коректність визначення наявності образливого вмісту повідомлення.

2. *Контекст (на рівні повідомлень у соціальних мережах)* – автори вважають, що врахування контексту сукупності повідомлень може значно погіршити часові показники роботи методу з визначення образливого вмісту. Проте, абсолютне ігнорування контексту може спричинити зниження якості роботи методу. Саме тому пропонується врахувати контекст окремих повідомлень шляхом використання підходів до автоматизованого визначення образливого вмісту текстових даних, які працюють не з окремими словами чи фразами (наприклад, підходи, що базуються на пошуку ключових слів), а з вектором слів.

3. *Наявність емодзі* – з точки зору автоматизованого оброблення текстових даних емодзі – це символи, що закодовані в юнікод. Наприклад, U+1F601 – це код смайлу емодзі. Звичайний класифікатор відкинув би ці дані на етапі передоброблення тексту як непотрібні символи, хоча насправді дані символи впливають на сприйняття користувачем повідомлень, шляхом зміни їх емоційного забарвлення.

Зважаючи на дані дослідження [9], автори вважають за доцільне додати можливість врахування емодзі при визначенні образливого вмісту повідомлень Інтернет-користувачів у соціальних мережах.

4. *Контекст (на рівні зв'язків користувачів у соцмережах)* – щоб врахувати зв'язки між користувачами при аналізі вмісту їх повідомлень, пропонується ввести такий формат вхідних даних:

```
{
  author: user
  complainer: user
  publication: private closed/public page,
               another user's page,
               community page
  marked people: list of users
  text: text for detection
}
```

Такий формат даних дозволяє врахувати інформацію про автора повідомлення; користувача, що скаржиться; тип сторінки: приватна закрита чи приватна відкрита сторінка; повідомлення, написане на сторінці іншого користувача чи публічній сторінці; список згаданих людей у повідомленні і, відповідно, сам текст повідомлення.

Врахування списку згаданих у повідомленні людей дозволяє збільшити вагомість скарги користувача соцмережі в процесі автоматичного прийняття рішення щодо образливості контенту (якщо скаржник згаданий у повідомленні, що визначається як повідомлення з образливим вмістом, то такий пост потрібно заблокувати).

Визначимо тепер, на якому етапі *процесу детекції образливого вмісту текстових даних* можна врахувати зазначені вище характерні риси повідомлень Інтернет-користувачів.

Умовно даний процес можна розділити на три етапи: передоброблення тексту, класифікацію підготовленого тексту та аналіз результатів класифікації. Автори вважають за доцільне здійснити модифікацію етапу передоброблення тексту, щоб врахувати спецсимволи та емодзі, а також ввести новий формат вхідних даних, який було описано вище, використання якого дозволить враховувати контекст та стосунки між користувачами соціальних мереж.

Передоброблення тексту пропонується проводити безпосередньо над текстовим повідомленням за допомогою класичних методів, які модифікуються шляхом додавання двох допоміжних кроків.

По-перше, вхідний текст має бути перевіреном на наявність в ньому символів, що можуть бути використаними для заплутування написання слів з метою обходу визначення образливого вмісту класифікатором. Наприклад, такі символи як «99» мають бути замінені на «gg», «@» на «a», «\$\$» на «ss» тощо. Даний крок ускладнить вживання образливих слів та дозволить покращити точність роботи класифікатора.

По-друге, автори вважають за доцільне перетворювати емодзі-символи на текстові дані, що дозволить не втрачати смислового та емоційного навантаження на текст. Існують стандартизовані таблиці емодзі-символів з кодами та значеннями, що містять у собі той чи інший код емодзі. Використання тлумачень кодів емодзі дозволить легко перетворити їх на слова або словосполучення та використати при класифікації, на відміну від звичайного передоброблення, де дані символи було б відкинуто.

Визначення образливого вмісту текстових повідомлень, що пройшли передоброблення, пропонується здійснювати на основі класичних методів визначення образливого вмісту текстових даних (наприклад, за допомогою Наївного баєсівського класифікатора). Жодних змін тут не потрібно.

Результат, наданий класифікатором (ймовірність зарахування до класу «образливий» текст), буде коригуватися на фінальному етапі, з урахуванням зв'язків між користувачами в соціальній мережі та місцем публікації. Мається на увазі, якщо текст є образливим, але він був опублікований на приватній сторінці і не має посилань на інших користувачів, то дане повідомлення не буде блокуватися. Якщо на ньому було відмічено певних людей, тоді текст має бути додатково перевіреном (наприклад, людиною-модератором). Проте, якщо даний текст опубліковано на публічній сторінці, то його має бути заблоковано. Даний підхід дозволяє убезпечити користувачів від образливого контенту, а з іншого боку – зменшує рівень цензури та дозволяє вільно висловлюватися.

Отже, у запропонованому авторами методі можна виділити такі етапи визначення образливого вмісту повідомлень користувачів соціальних мереж:

1) передоброблення тексту:

- із заміною спецсимволів на літери, які візуально схожі між собою;
- з перетворенням емодзі на текстові дані;

2) визначення образливого вмісту відібраного повідомлення за допомогою класичних алгоритмів визначення образливого вмісту природномовних текстових даних;

3) формування результатів, з урахуванням оцінки класифікатора та зв'язків Інтернет-користувачів у соціальній мережі.

Для практичного дослідження ефективності запропонованого методу авторами розроблено програмне забезпечення, яке реалізує даний метод автоматизованого визначення образливого вмісту повідомлень Інтернет-користувачів. На власноруч сформованому та розміченому наборі тестових даних з соціальної мережі Твіттер було проведено навчання та тестування. За часовими характеристиками запропонований метод показав відчутно гірші показники, ніж класичний метод, що спричинено доданою конвертацією емодзі в слова. Проте якість визначення образливого вмісту покращилася: класичний метод показує точність у 86 %, запропонований – 88 %. Таким чином, перспективним напрямком подальших досліджень даної проблематики автори вважають оптимізацію алгоритму інтерпретації емодзі символів у відповідні слова та фрази для зменшення часових показників виконання запропонованого методу.

Також варто зазначити, що запропонований метод дозволяє визначити образливий вміст у текстах, які класичний метод в принципі не може правильно опрацювати. Наприклад, повідомлення «jews are f@t and stlnky pls» визначається класичним методом як образливе, з ймовірністю 47 %, запропонований же метод визначає текст образливим, з ймовірністю 97 %. Отже, запропонований метод дозволяє досягти значного покращення якості визначення образливого вмісту у реченнях, що мають на меті заплутування слів і фраз.

Наразі авторами проводиться підготовка великого набору текстових повідомлень із соціальних мереж, що відображають наведені вище особливості та представлені у запропонованому форматі.

Висновки. В межах дослідження, описаного в даній статті, проведено вивчення специфіки публікацій Інтернет-користувачів у соціальних мережах та сформовано перелік їх характеристик, які можуть бути враховані під час визначення образливого вмісту, а саме: наявність символів, що можуть призвести до заплутування сприйняття слів і фраз; контекст (на рівні повідомлень у соціальних мережах); контекст (на рівні зв'язків користувачів у соціальних мережах); наявність емодзі.

Проведено аналіз процесу детекції образливого вмісту текстових даних з точки зору можливості врахування в ньому специфіки повідомлень користувачів соціальних мереж, а саме запропоновано модифікацію етапів передоброблення тексту та аналізу результатів класифікації. Розроблено метод визначення образливого вмісту повідомлень Інтернет-користувачів у соціальних мережах на основі комплексного врахування характеристик повідомлень у соціальних мережах та специфіки стосунків між різними групами відвідувачів соціальних мереж. Істотно новим в роботі є запровадження специфічного формату вхідних даних для класифікатора та врахування емодзі для визначення образливого вмісту. На базі даних досліджень розроблено програмну бібліотеку методів визначення образливого вмісту повідомлень Інтернет-користувачів, які реалізують описаний вище метод.

Список використаної літератури:

1. Internet World Stats [Електронний ресурс]. – Режим доступу : <https://www.internetworldstats.com/stats>.
2. Cyber Bullying Law and Legal Definition [Електронний ресурс]. – Режим доступу : <https://definitions.uslegal.com/c/cyber-bullying/>.
3. Визначення зловживань мовами в Інтернет-контент користувача / Ч.Новата, Дж.Тетро, А.Томас, Ю.Мехдад, Ю.Чанг // 25 Міжнародна конференція на веб-сторінках. – Монреаль, Квебек (Канада). – 2016. – С. 145–153.
4. Вікіпедія – вільна енциклопедія [Електронний ресурс]. – Режим доступу : https://uk.wikipedia.org/wiki/Соціальна_мережа.
5. Навчальні матеріали онлайн [Електронний ресурс]. – Режим доступу : http://pidruchniki.com/18421120/ritorika/pobutove_spilkuvannya.
6. Заболотня Т.М. Програмна бібліотека методів аналізу тональності відгуків Інтернет-користувачів / Т.М. Заболотня, А.Ю. Бартков'як // Informatics and Computer Technics Problems : Тези доповідей. – Чернівці. – 21–24 травня, 2016. – С. 85–87.
7. Вчимося з повсюдної мови: емпіричний аналіз використання смайликів користувачем смайлів / Х.Лу, Ш.Аї, Х.Лю, Ц.Лі, Н.Ванг, Г.Хуан, Ц.Мей // Процедури Міжнародної спільної конференції АСМ з повсюдного та повсякденного обчислення. – 2016. – С. 770–780.
8. Тауч К. Ролі «ем» у повідомленні мобільних телефонів / С.Тауч, Е.Каньо // Міжнародна спільна конференція з повсякденного та повсякденного обчислення. – UbiComp Adjunct. – 2016. – С. 1560–1565.
9. «Щасливо щасливі» або «готові до боротьби»: різне тлумачення смайликів / Н.Міллер, Ж.Тобальт-Спікер, С.Чанг, І.І. Джонсон, Л.Г. Тервен, В.Хехт // 10 Міжнародна конференція Web і соціальних медіа. – ICWSM 2016. – 2016. – С. 259–268.
10. Заболотня Т.М. Підхід до визначення образливого вмісту текстових повідомлень в соціальних мережах / Т.М. Заболотня, А.В. Соколовська // Прикладна математика та комп'ютинг. ПМК 2018 : тези доповідей. – Київ. – 21–23 березня, 2018. – С. 243–249.

References:

1. «Internet World Stats» [On-line], available at: <https://www.internetworldstats.com/stats>
2. «Cyber Bullying Law and Legal Definition» [On-line], available at: <https://definitions.uslegal.com/c/cyber-bullying/>
3. Novata, Ch., Tetro, Dzh., Tomas, A., Mehdad, Ju. and Chang, Ju. (2016), «Vyznachennja zlovzhyvan' movamy v Internet-kontent korystuvacha», 25 *Mizhnarodna konferencija na veb-storinkah*, Monreal', Kvebek (Kanada), pp. 145–153.
4. «Vikipedija – vil'na encyklopedija» [On-line], available at: https://uk.wikipedia.org/wiki/Social'na_merezha
5. «Navchal'ni materialy on-lajn» [On-line], available at: http://pidruchniki.com/18421120/ritorika/pobutove_spilkuvannya
6. Zabolotnja, T.M. and Bartkov'jak, A.Ju. (2016), «Programna biblioteka metodiv analizu tonal'nosti vidgukiv Internet-korystuvachiv», *Informatics and Computer Technics Problems*, tezy dopovidej, vid 21–24 travnja, Chernivci, pp. 85–87.
7. Lu, X., Ai, W., Lju, X., Li, Q., Vang, N., Huan, G. and Mei, Q. (2016), «Vchymosja z povsjudnoi' movy: empirychnyj analiz vykorystannja smajlykiv korystuvachem smajliv», *Procedury Mizhnarodnoi' spil'noi' konferencii' ACM z povsjudnogo ta povsjakdenного obchyslennja*, pp. 770–780.
8. Tauch, K. and Kanjo, E. (2016), «Roli «em» u povidomlenni mobil'nyh telefoniv», *Mizhnarodna spil'na konferencija z povsjudnogo ta povsjakdenного obchyslennja*, UbiComp Adjunct, pp. 1560–1565.
9. Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I.L., Terveen, L.G. and Hecht, B. (2016), «Shhaslyvo shhaslyvi» abo «gotovi do borot'by»: rizne tlumachennja smajlykiv», 10 *Mizhnarodna konferencija Web i social'nyh media*, ICWSM-2016, pp. 259–268.
10. Zabolotnja, T.M. and Sokolovs'ka, A.V. (2018), «Pidhid do vyznachennja obrazlyvogo vmistu tekstovyh povidomlen' v social'nyh mrezhah», *Prykladna matematika ta komp'jutyng. PMK-2018*, tezy dopovidej, vid 21–23 bereznja, Kyi'v, pp. 243–249.

Заболотня Тетяна Миколаївна – кандидат технічних наук, доцент кафедри програмного забезпечення комп'ютерних систем НТУУ «КПІ ім.І.Сікорського».

Наукові інтереси:

- обробка природомовних текстів;
- програмування.

E-mail: tetiana.zabolotnia@gmail.com.

Соколовська Анна Віталіївна – магістрант кафедри програмного забезпечення комп'ютерних систем НТУУ «КПІ ім.І.Сікорського».

Наукові інтереси:

- обробка природомовних текстів;
- програмування.

E-mail: anetkanfetka@gmail.com.

Стаття надійшла до редакції 18.05.2018.