

забезпечення обчислювальних машин і систем” / А.С. Василюк. – Львів, 2008. – 20 с. 12. Овсяк О. Класи інформаційної системи генерування коду / О. Овсяк // Вісн. Терноп. держ. техн. ун-ту: “Тернопільський національний технічний університет імені Івана Пулюя”. – 2010. – № 1. – С. 171–176. 13. Овсяк О.В. Граматика опису функційних унітермів / О.В. Овсяк // Поліграфія і видавнича справа // Української академії друкарства: Зб. наук. праць. – 2009. – № 2(50),– С. 18–22. 14. Овсяк О.В. Модель абстрактної підсистеми комп’ютерної інформаційної системи генерування коду (у цьому “Віснику”). 15. Petzold С. Programowanie Windows w języku C#. – Warszawa: RM, 2002. – 1161 s. 16. Мэтью Мак-Дональд. Windows presentejtation foundation в .NET 3.5 с примерами на C# 2008. – М., СПб., К.: Apress, 2008. – 922 с.

УДК 004.738.5

А.М. Пелецишин, Ю.О. Сєров, С.С. Федушко\*  
Національний університет “Львівська політехніка”,  
кафедри інформаційних систем та мереж,  
\*кафедра прикладної лінгвістики

## РОЗРОБЛЕННЯ АЛГОРИТМУ РЕЄСТРАЦІЇ ТА ВАЛІДАЦІЇ ПЕРСОНАЛЬНИХ ДАНИХ УЧАСНИКІВ ВЕБ-СПІЛЬНОТИ

© Пелецишин А.М., Сєров Ю.О., Федушко С.С., 2010

**Запропоновано методи перевірки персональних даних учасників веб-форуму. Ці методи реалізовані в алгоритмі реєстрації та валідації персональних даних про учасника веб-спільноти.**

**Ключові слова:** учасник веб-спільноти, персональний профіль, достовірність даних.

**This paper deals with the methods of verification of Web-forum members’ personal data, which are implemented in the algorithm of registration and validation of personal data on Web-community member.**

**Keywords:** Web-community member, personal profile, data verification

### Постановка проблеми

Розроблення нового підходу до верифікації даних, які надає веб-користувач під час реєстрації, є актуальним питанням в управлінні та модеруванні веб-спільнот. Зі збільшенням чисельності веб-спільнот та їх користувачів у мережі виникла потреба в розробленні методу перевірки максимальної кількості даних про потенційного учасника веб-форуму та комп’ютерно-лінгвістичного методу перевірки інформаційного сліду учасника інтернет-спільноти. За допомогою розроблення методів реєстрації учасників веб-спільноти спрощується реєстрація потенційних користувачів.

Для спрощення управління та підвищення ефективності діяльності веб-спільноти ми розробили основу для комп’ютерно-лінгвістичного методу перевірки інформаційного сліду учасника веб-форуму, який ґрунтується на лінгвістичному аналізі інформаційного наповнення Веб-спільноти. За допомогою комп’ютерно-лінгвістичного аналізу адміністрація спільноти зможе позбутись небажаних учасників та покращити комунікативну поведінку користувачів веб-спільноти.

### Аналіз останніх досліджень

Веб все більше набуває ознак реальної глобальної спільноти людей, у якій неможливо обійтися без механізмів ефективної ідентифікації [2]; відкритої системи формування і моніторингу комунікативної поведінки учасника глобальної спільноти тощо. Як наслідок, все актуальнішою стає перевірка достовірності інформаційного наповнення Вебу та інформації веб-особистості – множини

даних, які стосуються конкретної особи і можуть стосуватись будь-якої категорії або будь-якої комбінації категорій даних, доступних у мережі Інтернет [6].

На відміну від попередніх досліджень, у цій роботі особливу увагу звернено на проблему верифікації даних та розроблення методів реєстрації в інтернет-спільнотах, що найбільше впливають на підвищення ефективності функціонування віртуальних спільнот, на рівень достовірності персональних даних профілів учасників веб-спільнот. Вирішення цих проблем можливе за допомогою комп'ютерно-лінгвістичного аналізу дописів учасників інтернет-спільнот. Також актуальною проблемою сучасного Вебу є "інформаційний шум" [2] – величезна кількість неточної, неповної, поверхневої та відверто неправдивої інформації, яку створюють у Вебі всі охочі ("культ аматора", "антиелітність") [7].

Велика кількість користувачів Вебу та проблеми, які виникли на цьому ґрунті, вимагають нових підходів до розроблення методів реєстрації, перевірки персональних даних учасників, інформаційного наповнення веб-спільнот та класифікації користувачів за достовірністю даних їхніх персональних профілів. Методи валідації даних та перевірки інформаційного сліду веб-користувачів частково полягають у гендерному та віковому комп'ютерно-лінгвістичному аналізі інформаційного наповнення веб-форуму, який ґрунтується на дослідженні дописів учасників веб-форуму. Наприклад, щоб перевірити вік та стать учасника, виконують аналіз інтернет-імені, аватару та стилю написання дописів у дискусіях веб-спільноти.

Досліджено, що гендерні та вікові закономірності є у виборі користувачами аватарів та інтернет-імен. Інтернет-ім'я (нікнейм) впливає на соціальну взаємодію. Позитивні прізвиська сприяють швидкій адаптації в середовищі однолітків [8]. Аватари – це різноманітні картинки, фото, які відображають характер людини, її світогляд, адже зображення на аватарі користувач вибирає сам. Аватар завжди розташований поруч з інтернет-іменем.

Недавні дослідження [9] стверджують, що використання мови також варіюється залежно від віку людини, а гендерні відмінності є у дітей різного віку. У зв'язку з цим виникла потреба дослідити інформаційний слід підлітків та дорослих користувачів веб-спільноти, що дасть змогу ідентифікувати вік та стать учасників спільноти і перевірити достовірність даних персонального профілю учасника веб-спільноти.

Втім, проаналізувавши дослідження, варто зауважити, що науковий інтерес зростає, проте кількість фундаментальних досліджень у цій сфері невелика.

### Цілі статті

Мета цього дослідження – розроблення методів реєстрації та валідації персональних даних учасників веб-спільнот та побудова алгоритму перевірки даних потенційного учасника та інформації персонального профілю користувачів веб-спільноти.

**Отже, цілями цієї статті є:**

- аналіз досліджень перевірки даних та методів реєстрації учасників у мережі;
- розроблення методів реєстрації учасників веб-спільнот;
- розроблення методу перевірки максимальної кількості даних про потенційного учасника веб-форуму;
- побудова алгоритму реєстрації та валідації даних про учасника веб-спільноти;
- встановлення основних категорій вибору інтернет-імен учасниками веб-спільнот;
- визначення рівня достовірності персональних даних профілю учасника веб-спільноти;
- класифікація користувачів веб-спільноти за достовірністю даних персональних профілів.

## 4. Основний матеріал

### 4.1. Методи реєстрації учасників веб-спільноти

Методи реєстрації учасників у веб-спільнотах поділяють на три види:

- без верифікації (анонімна);
- з верифікацією;
- за запрошенням.

**Без верифікації.** Це найпростіший метод участі у віртуальних спільнотах. Метод участі без верифікації передбачає два варіанти: добровільне надання учасниками свого імені/прізвища (нікнейму) або участь у комунікації без надання імені (анонімна участь). Недоліком такого методу верифікації є найнижчий рівень захищеності від зловмисників, перевагою – найвища ймовірність створення людиною повідомлення.

**З верифікацією.** Цей метод реєстрації вимагає від людини, що реєструється, надання певних даних, передбачених правилами спільноти. У найпростішому випадку верифікація може бути повністю автоматичною і потребувати лише правильного надання адреси електронної пошти. У складнішому варіанті реєстрація є частково автоматизованою і передбачає участь модератора та/чи адміністратора веб-спільноти. У такому разі процес верифікації має різні ступені складності.

**За запрошеннями.** Такий метод реєстрації не є дуже поширеним і використовується у спеціальних спільнотах закритого типу. У такому разі верифікація учасника відбувається ще до процесу реєстрації.

Залежно від визначеного сценарію розвитку веб-форуму її власники вибирають один з методів реєстрації учасників [10].

Найпоширенішим сьогодні є метод реєстрації з верифікацією. Залежно від побажань власників веб-форуму цей метод має різні ступені складності.

#### 4.2. Алгоритм реєстрації та валідації даних про учасника

Авторами розроблено метод перевірки максимальної кількості даних про потенційного учасника веб-форуму та основу для комп'ютерно-лінгвістичного методу перевірки інформаційного сліду учасника веб-форуму, оснований на лінгвістичному аналізі інформаційного наповнення веб-спільноти.

Алгоритм премодерованої реєстрації призначений для того, щоб запобігти потраплянню на веб-форум небажаних для спільноти учасників. Ідея алгоритму полягає у максимальній перевірці та аналізі інформації про учасника, що здійснив запит на реєстрацію.

Алгоритм реєстрації та валідації даних профілю учасника веб-спільноти (рис. 1)

1. **Перевірка правильності адреси електронної пошти.** Якщо адреса електронної пошти вказана неправильно, то процес реєстрації припиняється, оскільки неможливо надіслати пароль.

2. **Перевірка інтернет-імені згідно з правилами веб-спільноти.** Якщо адреса електронної пошти вказана правильно, то відбувається перевірка інтернет-імені учасника:

- на відповідність правилам іменування у веб-спільноті (наприклад, адміністрація Форуму може вимагати, щоб користувачі Форуму під час реєстрації для інтернет-імені використовували шаблон: "Ім'я Прізвище" або змістовні псевдоніми);
- на відсутність уже зареєстрованих учасників з таким іменем.

3. **Комп'ютерно-лінгвістична перевірка інтернет-імені.** Якщо ж інтернет-ім'я відповідає правилам і відсутні уже зареєстровані учасники з таким іменем, то відбувається його комп'ютерно-лінгвістичний аналіз [5]. Комп'ютерно-лінгвістичний аналіз полягає в гендерній та віковій перевірці інтернет-імені за допомогою сформованих нами категорій вибору інтернет-імені певного віку та статі користувачів веб-форумів.

Таблиця 1

#### Категорії вибору інтернет-імені дорослих користувачів веб-спільнот

Категорії вибору	Приклади реальних інтернет-імен
імена людей	Шапак Тетяна, Олена, Sonja
імена, пов'язані з політичними поглядами	націоналіст, комуніст
імена, пов'язані з назвами природних явищ, рослин, тварини тощо	Сова, Форест, fish, Вітер, осінь, зіронька, Panda, Crobak, root
імена, пов'язані з територіальною приналежністю	Східняк, львів'янин, Московит, Moscal
використання художніх засобів мови	світлий вечір, крижана посмішка, море сліз
імена, пов'язані з літературою, міфологією, фільмами, історією і відомими людьми	Seneka, Kallisto, Відьмочка, чарівник з міста Ромельн, козак
імена сексуального характеру	serceedka, LLLover
імена, які описують характер, емоційний стан	сумна, freewill, ReadHead, freak, рішучий, бравий
імена, пов'язані з професією	Соло, репортер, каунтер, artist

Формуючи ці категорії, ми виявили певні закономірності вибору інтернет-імен. А саме: жінки надають перевагу іменам, які описують характер, емоційний стан; пов'язані з літературою, фільмами і відомими людьми; використовують художні засоби мови у формулюванні імені. А чоловіки, своєю чергою, вибирають імена, пов'язані з професією, ті, які вказують на політичні погляди та імена, пов'язані з містом, країною.

З однаковою частотою як чоловіки, так і жінки вибирають імена, пов'язані з літературою, міфологією, фільмами, історією і відомими людьми та імена, пов'язані з назвою природних явищ, рослин, тварини тощо.

Також характерною ознакою використання інтернет-імен дорослих користувачів, є те, що, як правило, такі користувачі для реєстрації частіше подають свої справжні імена та прізвища порівняно з підлітками.

Таблиця 2

### Категорії вибору інтернет-імені дітей/підлітків

Критерії вибору	Приклади реальних інтернет-імен
імена	Сашко, Олег, Katrusja
прізвиська	21dimas, MarkX, Fredi
імена-назви тварин	хом'як, bunny, Wild-cat, Pig
імена, пов'язані з інструментами або виявами насилля	калашнікоф, меч, breather, breaker
імена, які описують емоції, особисті характеристики	самотність, бліда поганка, іржа, зараза, abnormal, скромна
імена, пов'язані з музикою та поп-культурою	Kurt, sound, Hardmetalgirl
імена, пов'язані з фантазією/міфологією, фільмами, казками	дитя місяця, Andriy Haus, незнайка, alien, femida, drakyla, русалка, PoKEMoN
імена, які допомагають самоствердитись	superman, egoman, Coolgirl
лагідні та пестливі імена	Lovely, kotenok, Angelbaby
імена, що символізують природні явища	зіронька, веселка, крижана зірка, sun
безглузді імена	уть ють ать ять, трамбулямбулямбус

Хлопчики, як правило, вибирають імена, пов'язані з насиллям, з музикою та поп-культурою, прізвиська, а дівчатка – імена, які символізують природні явища, лагідні та пестливі імена, імена, які описують емоції, особисті характеристики. Безглузді імена, імена, які пов'язані з фантазією/міфологією, фільмами, казками, назвами тварин, допомагають підняти самооцінку, символізують природні явища, вибирають діти незалежно від статі.

За допомогою цього аналізу зможемо встановити вік та стать учасників. Це полегшить роботу модераторів чи адміністраторів з перевірки наданої користувачем інформації щодо віку та статі.

Коли ж інтернет-ім'я вже використовує інший користувач чи воно не відповідає правилам веб-спільноти та комп'ютерно-лінгвістична перевірка інтернет-імені невдала, заявнику відправляється повідомлення з проханням змінити інтернет-ім'я.

**4. Перевірка інформації про учасника.** Якщо інтернет-ім'я потенційного учасника пройшло аналіз, то відбувається перевірка іншої інформації про учасника, яка потребує участі адміністратора.

Зокрема, перевіряється:

- ім'я учасника – на відповідність правилам іменування у спільноті, на відсутність уже зареєстрованих учасників з таким іменем;
- адреса електронної пошти та уся додаткова інформація перевіряється на відповідність імені та достовірність за допомогою засобів пошукових систем;
- достовірність вказання географічного розміщення перевіряють засобами визначення IP-адрес [7];
- інформаційний слід веб-особистості [1, 2, 3] потенційного учасника веб-форуму перевіряється за допомогою пошукових систем.

Коли інформації про учасника не знайдено, то дозволяється реєстрація і ставиться мітка з метою перманентного моніторингу комунікативної поведінки учасника веб-спільноти, щоб пересвідчитись, що учасник надав для реєстрації свої реальні дані.

Якщо надана інформація не компрометує учасника, а інформаційний слід веб-особистості не містить негативної інформації, то учасник отримує згоду на приєднання до спільноти та здійснюється комп'ютерно-лінгвістичний аналіз даних персонального профілю користувача [4]. В іншому разі йому відмовляють у реєстрації.

5. **Післяреєстраційний комп'ютерно-лінгвістичний аналіз аватара, юзербару, біографії, підпису та статусу веб-учасника.** Відбувається перевірка інформації, яку надав учасник веб-спільноти у своєму персональному профілі, на достовірність віку, статі, територіальної приналежності, біографії, професії тощо. Наприклад, є певна гендерна закономірність вибору аватарів підлітками: дівчатка переважно вибирають на аватарки картинки тварин, персонажів з мультфільмів, фото зі своїм другом, музичних кумирів, привабливих жінок або свої спокусливі фото; а ось хлопчики – фото відомих спортсменів, кіноакторів, музичних кумирів, сексуальних дівчат, картинки, зображення на яких вирізняється жорстокістю. Також частим явищем є надписи на самій фотографії чи малюнку, який функціонує як аватар користувача, підписи під дописом та стан (настрій) учасника веб-спільноти. На основі цієї інформації відбувається аналіз, в результаті якого отримуємо можливість ідентифікувати користувача за певними критеріями (вік, стать, біографія, професія, територіальна належність тощо).

Коли ж в результаті цього аналізу дані профілю користувача не є достовірними чи не відповідають правилам, то учаснику надсилається повідомлення з вимогою змінити інформацію.

Якщо учасник знехтував проханням адміністрації, не змінивши даних персонального профілю (віку, статі, країни/міста проживання, професії, даних біографії тощо), то у тому разі додаємо учасника в список з недостовірною інформацією і моніторимо всі його дії. Якщо учасник надав достовірні дані під час заповнення власного профілю чи змінив їх на вимогу адміністрації, то користувача додають до списку учасників з перевіреними даними та надалі моніторять всі дії учасника.

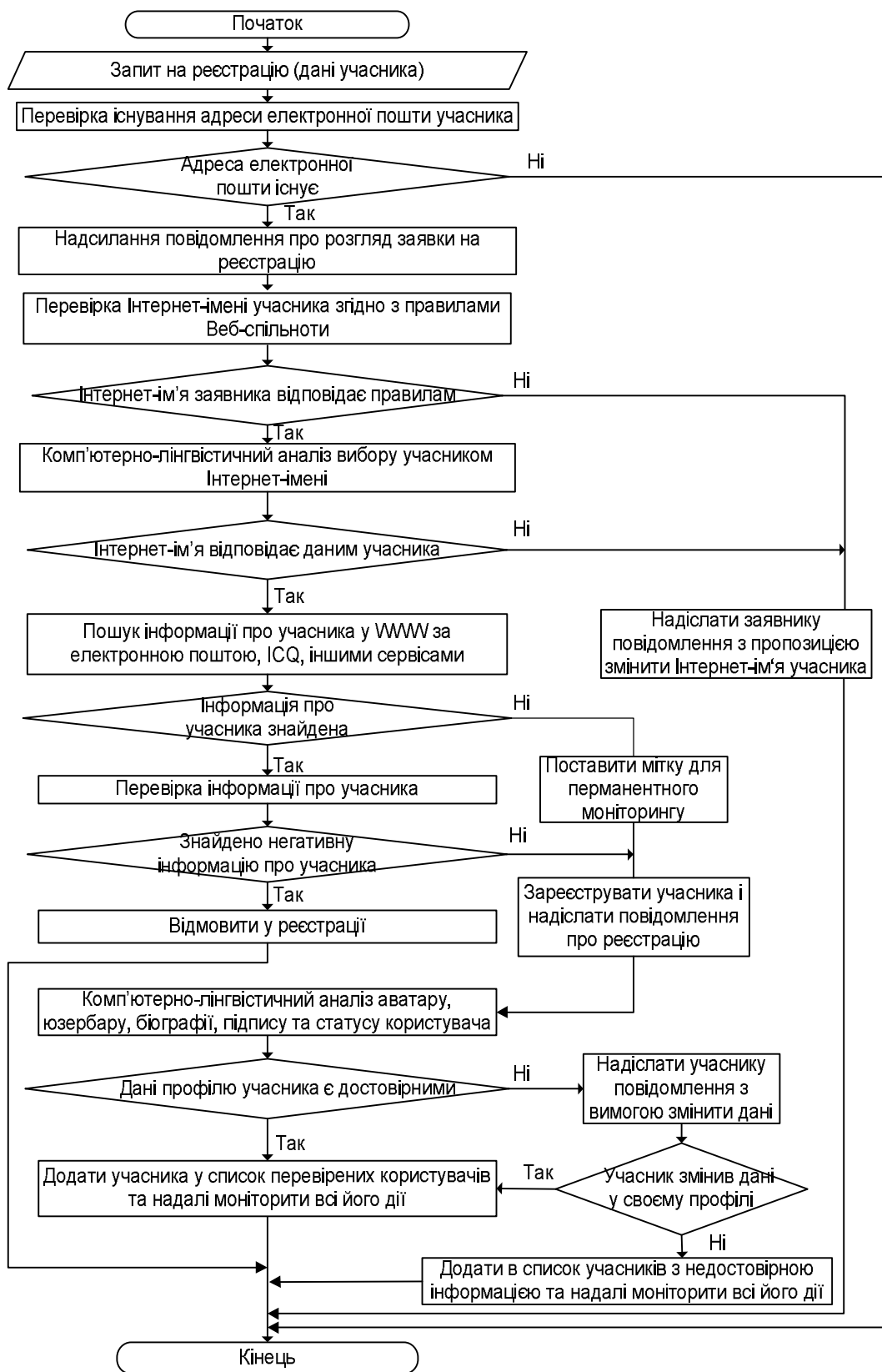
Аналіз реєстраційних та персональних даних користувачів, який відбуватиметься за алгоритмом реєстрації та валідації даних про учасника веб-спільноти, дає змогу виявити міру достовірності даних персональних профілів учасників веб-спільноти, виділити учасників з достовірними та недостовірними даними. За допомогою алгоритму реєстрації та валідації даних про учасника ми отримуємо можливість класифікувати учасників веб-спільноти за достовірністю наданої інформації та розділити всіх учасників за цією ознакою. В результаті одержимо класифікацію, яка допоможе нам структурніше та поступовіше моніторити давно та нещодавно зареєстрованих користувачів веб-спільноти.

Класифікація персональних профілів користувачів веб-спільнот за ознакою достовірності даних:

1. **Достовірні.** Особиста інформація, яку надав користувач у своєму профілі, достовірна або ж він змінив її на вимогу адміністрації веб-спільноти під час комп'ютерно-лінгвістичного аналізу і комунікативна поведінка учасника відповідає всім правилам спільноти.

2. **Під підозрою.** Не всі персональні дані веб-користувача відповідають реальності. Користувач частково або цілком проігнорував прохання змінити недостовірну інформацію. Всі зміни даних та комунікативну поведінку моніторить адміністрація веб-спільноти.

3. **Недостовірні (псевдоучасники).** Дані персонального профілю цілком недостовірні, учасник порушує правила веб-спільноти. Профіль учасника блокується/видаляється, забороняється доступ до веб-спільноти.



Алгоритм реєстрації та валідації даних про учасника веб-спільноти

### Висновки

Методи реєстрації та валідації персональних даних реалізовано в алгоритмі реєстрації та валідації даних про учасника веб-спільноти. Ідея алгоритму полягає у максимальній перевірці та аналізі інформації про учасника, що здійснив запит на реєстрацію, а також у післяреєстраційній

комп'ютерно-лінгвістичній верифікації даних персонального профілю веб-учасника та його інформаційного сліду. Цей алгоритм забезпечує також збільшення кількості учасників з повною (достовірною) інформацією та можливість їх класифікувати, покращити комунікативну поведінку користувачів та усунути з веб-спільноти псевдоучасників, спростивши управління та підвищивши ефективність діяльності веб-спільноти.

1. Березко О.Л. Використання вікі-засобів для персоніфікації інформаційного наповнення World Wide Web / О.Л. Березко // Вісн. Нац. ун-ту «Львівська політехніка». – 2008. – № 631: Інформаційні системи та мережі: – С. 11–18.
2. Березко О.Л. Каталог Веб-особистостей / О.Л. Березко // Вісн. Нац. ун-ту «Львівська політехніка». – 2008. – № 630: Комп'ютерні системи та мережі: – С. 12–16.
3. Березко О.Л. Особливості сучасних концепцій розвитку Вебу / О.Л. Березко, А.М. Пелецишин // Proc. of the First Intern. Conf. on Computer Science and Engineering «CSE-2006». – Lviv, 2006. – P. 9–12.
4. Пелецишин А.М. Гендерні та вікові відмінності мови у Інтернет-комунікації./ А.М. Пелецишин, С.С. Федушко // Людина. Комп'ютер. Комунікація: Зб. наук. праць / За ред. Ф.С. Бацевича. – Львів: Вид-во Нац. ун-ту "Львівська політехніка", 2010. – С. 88–90.
5. Пелецишин А.М. Методи верифікації персональних даних на основі гендерного аналізу мови користувачів Веб-спільнот / А.М. Пелецишин, С.С. Федушко // Східно-Європейський журнал передових технологій. – 2010. – № 3/4(45). – Харків, 2005. – С. 37–39.
6. Пелецишин А.М. Формальні аспекти ідентифікації особистості у WWW. Комп'ютерні системи та мережі / А.М. Пелецишин, О.Л. Березко // Вісн. Нац. ун-ту "Львівська політехніка" – 2005. – № 546.– С.126–131.
7. Carr N. The amorality of Web 2.0. Rough type, October 3, 2005 -[http://www.routhtype.com/archives/2005/10/the\\_amorality\\_o.php](http://www.routhtype.com/archives/2005/10/the_amorality_o.php)
8. Crawford M. Talking Difference: On Gender and Language / M. Crawford. – London: Sage Publication, 1995.
9. Croll A. Complete Web Monitoring: Watching Your Visitors, Performance, Communities, and Competitors / A. Croll, S. Power. – O'Reilly Media, 2009. – 672 p.
10. Mulac A. Male/female language differences and attributional consequences in children's television / A. Mulac, J.J. Bradac, S.K. Mann // Human Communication Research, 11(4), 1985. – P. 481–506.
11. Peleschyshyn A. Typical Ways of Web Communities Development / A. Peleschyshyn, Yu. Syerov // Proceedings of the International Conference of Computer Science and Information Technologies «CSIT-2006». – Lviv, 2006. – P. 56–58.