

МЕТОДИ ВЕРИФІКАЦІЇ ПЕРСОНАЛЬНИХ ДАНИХ НА ОСНОВІ ГЕНДЕРНОГО АНАЛІЗУ МОВИ КОРИСТУВАЧІВ ВЕБ-СПІЛЬНОТ

Дана стаття присвячена актуальній проблемі дослідження методів верифікації наданих особистих даних користувачів Веб-форуму на основі гендерних відмінностей мови написання повідомлень

Ключові слова – гендерні відмінності, Веб-форум

Данная статья посвящена актуальной проблеме исследования методов верификации предоставленных личных данных пользователей Веб-форума, основанных на гендерных различиях языка написания сообщений

Ключевые слова - гендерные различия, Веб-форум

This article considers current problem of investigation of verification personal data based on methods of gender differences of Web-forum member's messages

Keywords – gender differences, Web-forum

А.М. Пелещин

Доктор технічних наук, професор
Кафедра ІСМ*

Контактний тел.: 066-787-88-92

E-mail: apele@ridne.net

С.С. Федушко

Кафедра ПЛ*

Контактний тел.: 097-148-00-47

E-mail: felomia@gmail.ru

*Національний університет «Львівська Політехніка»
вул. Бандери, 12, м.Львів, Україна, 79013

1. Вступ

Дане дослідження присвячене розробці комп'ютерно-лінгвістичних методів верифікації персональних даних користувачів Веб-спільнот, які користувачі публікують в свої акаунти. На основі аналізу інформаційного наповнення Веб-форумів ("Львів. Форум Рідного Міста", "Тернопільський Форум", "Родинний форум "Малеча", "ROCK.LVIV.UA: Форум") ми дослідили комп'ютерно-лінгвістичні аспекти верифікації інформації, основані на класифікації категорій стилів написання повідомлень Веб-користувачами.

Актуальність дослідження зумовлена потребою розробки комп'ютерно-лінгвістичних методів верифікації наданої інформації користувачами, які полягають у формуванні категорій для ідентифікації користувачів Форумів.

2. Аналіз останніх досліджень

Аналіз сучасних праць науковців виявив, що незважаючи на велику кількість досліджень мови, вагомий вклад внесено вченими саме в дослідження гендерних відмінностей в мові. Результати дослідження гендер-

них відмінностей є основою для розробки комп'ютерно-лінгвістичних методів верифікації персональних даних користувачів Веб-спільнот Укрнету.

Гендерні відмінності розглянуті у дискусіях між користувачами Мережі. Дослідження стилів мови в онлайні показало [1], що характеристики стилів мови чоловіків і жінок в Інтернет-спілкуванні з порівнянні з реальною розмовою суттєво не змінюються. У дослідженні 3000 онлайн повідомлень, Вітмер і Кацман [2] виявили, що жінки частіше використовують графічні акценти, включаючи смайлики, щоб висловити свої емоції в дискурсі.

У використанні смайликів також існують вікові та гендерні відмінності. Дослідивши 3000 онлайн повідомлень, Діана Вітмер [3] виявила, що жінки використовують більше графічних акцентів, включаючи смайлики, щоб висловити свої емоції в дискусії, ніж чоловіки. Гендерні відмінності також є у виборі користувачами аватарів і Інтернет-імен.

Проаналізувавши проведені дослідження, ми зробили висновок, що вже виявлено відмінності у мові соціальних груп, проте ці відмінності не достатньо досліджені. Отже, гендерний фактор у мовленнєвому стилі потрібно досліджувати не поодинокі, а спільно з іншими соціальними факторами, такими, як вік, клас, етнічна приналежність, місце проживання та ін.

3. Цілі статті

Мета даного дослідження – розробка комп’ютерно-лінгвістичних методів для верифікації персональних даних користувачів на основі дописів на Веб-форумі. Цілями цієї статті є:

- Аналіз існуючих досліджень гендерних відмінностей мови користувачів;
- Встановлення особливостей використання Інтернет-імені, аватарів та смалів у Веб-комунікації, проаналізувавши інформаційне наповнення Веб-спільнот: “Тернопільський Форум”, “ROCK.LVIV.UA: Форум”;
- Визначення критеріїв для ідентифікації статі користувачів у Веб-спільнотах (“Львів. Форум Рідного Міста”, “Родинний форум “Малеча”).

4. Гендерна ідентифікація користувачів Веб-форуму за вибором ними Інтернет-імен та аватарів

У цьому дослідженні ми звернули основну увагу на методи та критерії вибору імен та аватарів користувачів Веб-спільнот (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA: Форум”). Ми дослідили використання Інтернет-імен, які вибрали для реєстрації учасники Веб-спільнот. Дорослі користувачів частіше подають свої справжні імена і прізвища в порівнянні з підлітками.

У дослідженні ми також виявили гендерну закономірності вибору Інтернет-імен користувачами чотирьох Веб-спільнот. Жінки надають перевагу іменам, які описують характер, емоційний стану; пов’язані з літературою, фільмами і відомими людьми. А чоловіки – іменам, пов’язаних з професією, політикою. З однаковою частотою як чоловіки, так і жінки вибирають імена, пов’язані з відомими людьми та з назвою природних явищ, тварини тощо.

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

Також, ми дослідили гендерну закономірність вибору аватарів підлітками. Дівчатка переважно обирають на аватарки картинку 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, картинку, зображення на яких вирізняється жорстокістю. Таким чином, Інтернет-імена і аватари надають первинну форму онлайн ідентичності і можливість ідентифікувати статі і вік користувача.

5. Графічні засоби вираження емоцій – графічні усмішки (смайлики)

Використання графічних і текстових усмішок користувачами у діалогах проаналізовано у цьому до-

слідженні. Оскільки, жінки традиційно більш емоційні, то ми очікували, що жінки у своїх повідомленнях використовуватимуть смайлики частіше, ніж чоловіки. Проте, всупереч прогнозам, досліджено тенденцію, що чоловіки використовують більше смайликів, ніж жінки (F=44%, M=56%). Чоловіки у своїх повідомленнях використовували більше *кокетливих та сумних* смайликів, а жінки – *радісних*. Отже, дослідження використання смайликів дає можливість визначити гендерну приналежність автора повідомлення.

6. Аналіз гендерних відмінностей мови Інтернет-користувачів на основі дописів на Веб-форумі “Львів. Форум Рідного Міста”

Результати наших досліджень гендерних відмінностей в мові користувачів Веб-форумів представлені в табл. 1. Ми обчислили ймовірність появи кожної з сформованих нами гендерних категорій у дискусіях Веб-форумів.

Таблиця 1

Критерії, які використовують користувачі в Інтернет-комунікаціях

Тип аналізу	Характерні категорії написання повідомлення	PF	PM
1	2	3	4
На рівні частин мови, конкретних слів та фраз	Велика кількість вибачень	0,024	0,0074
	Модальні конструкції	0,0239	0,0088
	Вказівні займенники – той, ті, того та ін.	0,025	0,042
	Згадка про емоції та почуття	0,036	0,0013
	Просторові посилання	0,003	0,0136
	Ухилення від відповіді	0,029	0,0021
	Фрази, які вказує на брак впевненості	0,044	0,006
	Прикметники (особиста оцінка, а не опис почуттів, речей)	0,062	0,0383
	Прийменники, які вказують на час і місце	0,0074	0,0279
	Географічні посилання	0,0018	0,024
	Стверджувальні фрази	0,0103	0,0286
	Слова розпоряджень	0,0069	0,0284
	Дієслова теперішнього часу	0,014	0,036
	Довгі слова	0,006	0,0091
	Використання ввічливих слів	0,052	0,019
	Акцент на словах – «отже», «дуже», «досить» та ін.	0,036	0,0109
	Лайливі та грубі форми слова	0,0089	0,041
	‘Порожні’ прикметники	0,0261	0,0045
	Лексика (родинні теми дискусій)	0,0306	0,0046
	Використання умовного способу	0,012	0,0035
Уникнення лайливих висловів	0,008	0,0033	
Підсилювальні слова	0,0224	0,0025	
Вказівка на розташування об’єктів	0,0099	0,048	

Продовження таблиці 1

1	2	3	4
На рівні частин мови, конкретних слів та фраз	Займенник першої особи однини (я, мене, моїх та ін.)	0,056	0,095
	Особовий займенник “ми”	0,025	0,009
	Пестливі слова	0,014	0,0054
	Лексика (спорт, технології, політика, гроші)	0,017	0,084
	Посилання на кількість, величину	0,0081	0,035
	Присяги і клятви	0,00063	0,0149
	Згадування негативних минулих подій	0,038	0,017
На рівні речень	Згадування минулих подій по відношенню до теперішнього часу	0,0021	0,039
	Окличне речення	0,015	0,011
	Вказівки	0,004	0,046
	Речення, які служать для вираження підтримки	0,0397	0,003
	Протиставлені речення	0,009	0,0023
	Речення, які служать для вираження згоди	0,049	0,065
	Перефразування речень	0,0012	0,0057
	Непрямі команди і прохання	0,0098	0,0047
	Підрядні речення обставини	0,098	0,0063
	Позначення вигуків	0,0046	0,013
	Безособові речення	0,0097	0,025
	Підрядні речення позбавлене смислу	0,0271	0,0056
	Підрядне речення, яке служить для уточнення	0,1041	0,0102
	Заперечні речення	0,0361	0,013
	Обговорення поточних проблем	0,012	0,044
	Запитання	0,0407	0,0037
	Розділові речення	0,0165	0,0056
	Прямі цитування	0,005	0,0013
Еліптичні речення	0,0091	0,0265	

PF – ймовірність появи критерію стилю мови жінки; **PM** – ймовірність появи критерію стилю мови чоловіка; **0,0п** – більша ймовірність появи категорії у повідомленні чоловіка/жінки.

7. Висновки

Результати нашого дослідження можуть суттєво допомогти в створенні ефективних методів верифікації особистої інформації, яку надав користувач. Нами проаналізовано інформаційне наповнення чотирьох Веб-форумів з різною тематикою і цільовими групами дописувачів (“Львів. Форум Рідного Міста”, “Тернопільський Форум”, “Родинний форум “Малеча”, “ROCK.LVIV.UA : Форум”). Проаналізувавши останні дослідження ми розробили основу для методики комп’ютерно-лінгвістичної перевірки достовірності наданої інформації користувачів Мережі у своїх аккаунтах та визначення ймовірності вживання характерних певному стилю мовлення категорій у Інтернет-спілкуванні. Цей лінгвістичний матеріал є доброю основою для розроблення спеціалізованого програмного забезпечення, яке за допомогою комп’ютерно-лінгвістичних методів верифікації наданої інформації користувачем, дає змогу ідентифікувати автора опублікованих повідомлень у Мережі.

Література

1. Mulac A. Female and male managers and professionals criticism giving. / Mulac A., Seibold D.R., Farris J.L. // Differences in language use and effects Journal of Language and Social Psychology, Vol.1, 2000. – P:389-415
2. Witmer D.F. Smile When You Say That: Graphic Accents as Gender Markers in Computer-Mediated Communication. – CA: Menlo Park, 1997.
3. Witmer D.F. On-line Smiles: Does Gender Make a Difference in the Use of Graphic Accents // Journal of Computer-Mediated Communication. – vol. 2, N4. – 1997.
4. Lakoff R. Language and Woman’s Place. / Lakoff R. – New York: Harper & Row, 1975.