

Обробка інформації в складних організаційних системах

УДК 004.9

С.В. Голуб, О.В. Константиновська, М.С. Голуб

Черкаський національний університет імені Богдана Хмельницького, Черкаси

ВІДОБРАЖЕННЯ ВЛАСТИВОСТЕЙ АВТОРА ТЕКСТУ В СТРУКТУРІ БАГАТОПАРАМЕТРИЧНОЇ МОДЕЛІ

Запропоновано результати досліджень процесу відображення властивостей автора тексту в структурі багатопараметричної індуктивної моделі. Описано процес використання нового методу формування інформативних характеристик тексту та вибору вигляду опорних функцій в технології інформаційного моніторингу з багаторівневим перетворенням вигляду даних. Виявлено, що різноманітності багаторівневого алгоритму МГУА достатньо для отримання адекватних моделей авторів друкованих текстів.

Ключові слова: *текстове повідомлення, інформативні ознаки, модель автора, ідентифікація, моніторинг.*

Вступ

Останнім часом актуальність задачі визначення властивостей автора за його текстовими повідомленнями зростає. Це пов'язано не тільки із необхідністю автоматизації процесу атрибуції тексту [1] – встановлення авторства анонімних повідомлень, чи до цього часу невідомих авторів літературних творів. Цікавими у науковому плані є дослідження тексту з метою автоматизації процесу виявлення додаткової інформації про стан здоров'я автора, виявлення його прихильності до вибору певних рішень. Розв'язання цієї задачі є важливою складовою процесу консолідації інформації [2], отриманої із різнорівневих джерел, з метою виявлення механізмів впливовості зовнішніх факторів на досліджувані процеси в криміналістиці, історичній інформатиці, соціології та інших галузях людської діяльності, де вимагається багаторівневе перетворення даних у інформацію. Інформаційні технології аналізу текстових повідомлень можуть бути успішно використані також в процесі адаптивного позиціонування сайтів в глобальному інформаційному середовищі [3].

У попередній статті [4] ми дали загальну характеристику питанню ідентифікації авторства тексту та привели результати своїх досліджень, запропонували свій метод вибору інформативних ознак тексту на прикладі довільних вибірок із авторських творів О. Гончара, І. Франка, М. Коцюбинського. Вибір цих авторів на початку досліджень був зумовлений тим, що творчість майстрів дозволяє якісніше виявити необхідні ознаки (мається на увазі професіоналізм авторів, що виявляється у стійкості синтаксичних характеристик тексту, відсутність копіювання "чужого" стилю).

Огляд методів та засобів розв'язку даної задачі. Питання вибору інформативних ознак та таких супутніх тем, як об'єм вибірок, їх кількість у процесі розпізнавання стоїть на першому плані у проблемі ідентифікації. Як ці питання вирішують деякі сучасні дослідники?

Вважається доведеним [5], що текст може служити засобом діагностики індивідуальних особливостей його автора. Поширеним є поняття профілювання [6] – виявлення характеристик автора за його текстовим повідомленням. Воно ґрунтується на доведеній гіпотезі про те, що слова, що використовуються людиною, несуть змістове повідомлення як частина речення, та психологічну інформацію як засіб відображення особистості. Для ідентифікації автора тексту використовуються аналіз послідовності використання букв, будувати функції розподілу їх використання відомими авторами та порівнювати ці характеристики тесту невідомих авторів. У випадку, коли відхилення побудованих статистичних ознак текстів не перевищує наперед задану норму, то ці тексти вважаються написаними одним автором [7].

Д. Хмельов пропонує програму "Аналізатор", яка створена на базі прозових творів російської фантастики [8]. Вона визначає трьох найближчих авторів запропонованого їй твору, користуючись базою даних, що налічує більше 120 авторів та 1350 творів. Об'єм пам'яті вибірок, які обробляла програма, становить від 12 Кб (приблизно 1000 слів) до 100 Кб (20000 – 30000 слів). Алгоритм роботи програми складається з використання ланцюгів А. Маркова, основними інформаційними ознаками є морфологічні структури (дослідження службових частин мови, морфемного складу слова, попарного використання літер без пробілів та пунктуаційних знаків).

Дослідники В.П. Фоменко та Т.Г. Фоменко висловлюють думку про стійкість зустрічаємості службових частин мови у кожного з авторів (т.зв. авторський інваріант). Кількість досліджуваних авторів становить 20 осіб, а запропонований об'єм вибірки – 16000 слів [9].

Л.І. Бородкін пропонує метод побудови графів, основні ознаки тексту – слова різних граматичних класів та їх характеристики [10].

В цьому напрямку дослідження внесла свій не малий вклад І.П. Севбо, висуваючи ідею аналізу графів синтаксичних зв'язків речення [11].

Як бачимо, дослідники не минули жодної з областей науки мовознавства, щоб визначити, яку саме роль кожна з них відіграє у процесі розпізнавання авторських текстів. Кожна з цих праць окремо та всі взагалі – титанічний труд, кропіткий пошук, а ідеї та їх реалізація відтворені у статтях та програмах.

Звертає увагу на себе той факт, що не виявлено досліджень, направлених на підвищення технологічності аналізу текстових повідомлень. Окремо розробляються методи ідентифікації текстових повідомлень, інші підходи використовуються для виявлення психо-фізичних особливостей автора. На сьогодні є потреба в розробці методів аналізу текстів, які дозволять поєднати в єдину технологію процеси ідентифікації автора тексту та виявлення характеристик його фізичного та психологічного станів.

Метою даної роботи є створення технології аналізу текстових повідомлень, яка здатна розв'язувати комплекс задач із ідентифікації автора тестового повідомлення та визначення характеристик його стану. Ми пропонуємо свій метод побудови ефективних моделей ідентифікації автора друкованих творів.

Технологія інформаційного моніторингу

Для досягнення поставленої мети – ідентифікації автора – використовується *технологія розпізнавання образів*. “Розпізнавання передбачає завдання перетворення вхідної інформації, за яку треба вважати деякі параметри, що є ознаками образів розпізнавання, у вихідну інформацію, за яку треба вважати висновок про те, до якого класу відноситься образ розпізнавання. В системі розпізнавання працюють сумісно колектив кваліфікованих спеціалістів та сукупність технічних інформаційно-обчислювальних засобів отримання і перетворення інформації, тому вона дуже складна та призначена для розв'язання завдань розпізнавання окремих образів чи явищ” [12 – С. 7, 9]. Власне технологія розпізнавання образів складається з декількох поступових етапів, що гарантують потрібний кінцевий результат. На кожному з цих етапів дослідники використовують певні інструменти (інформаційні системи), кількість яких залежить

від загальної кількості етапів дослідження та рівня забезпечення цих систем. Бажано, щоб дослідник мав можливість користуватися однією інформаційною системою протягом усіх етапів дослідження. Але винахід цієї системи – складний великий процес, беручи до уваги кількість операцій, що необхідно виконати для обробки вхідної інформації до моменту її закладення в саму програму.

Для розпізнавання характеристик автора друкованого тексту в даній роботі використовується технологія інформаційного моніторингу із багаторівневим перетворенням даних [13]. Вона містить такі етапи:

- 1) визначення вимог до інформації, яку необхідно отримати в результаті моніторингу;
- 2) декомпозиція задачі перетворення чисельних характеристик стану первинного опису в інформацію у вигляді характеристик моделей, визначивши кількість рівнів та задачі на кожному рівні;
- 3) визначення переліку станів та набору ознак, що характеризують ці стани об'єктів моніторингу;
- 4) визначення чисельних значень цих ознак та підготовка до подальшого моделювання;
- 5) формування структури глобальної функціональної залежності у вигляді ієрархічного поєднання багатопараметричних моделей об'єкту моніторингу шляхом їх висхідного синтезу;
- 6) випробування глобальної функціональної залежності на придатність до її використання;
- 7) застосування глобальної функціональної залежності для визначення поточного стану об'єктів моніторингу.

Метод синтезу моделей

Для вирішення завдань розпізнавання образів найчастіше використовуються детерміновані або статичні (імовірнісні) методи. Всі вони належать до т. зв. *детерміністичного напрямку* [14 – С. 3, 140, 143], що розглядає причинно-наслідкові зв'язки та розшукує нові знання за допомогою переліку всіх можливих варіантів рішень. Мінусами таких методів є їх суб'єктивізм, потреба брати великі вибірки, витрата великої кількості часу, залучення до дослідження багатьох спеціалістів, а також великих грошових коштів. Все це зумовлює недостатню якість (точність) розпізнавання. Детерміновані методи також непридатні для точного розпізнавання за таких причин: а) можлива повторюваність цих ознак у декількох авторів; б) велика кількість ознак.

Саме ці причини та наявність малої кількості вихідних даних зумовлює застосування методу групового врахування аргументів (МГУА) [14]. Пріоритет МГУА полягає в тому, що для синтезу вихідної функції потрібен відносно невеликий об'єм даних, також його використання не вимагає якихось специфічних фахових знань. Достатньо вивчити можливості керування цим інструментом

та застосувати свої професійні знання, щоб вимірювати та отримати задовільні результати (як не треба знати складові мікросхем телевізора, щоб його дивитись).

Масив вхідних даних було сформовано за методом, поданим в [4]. Окремий рядок масиву формувався як точка спостереження, що відображена в багатовимірному просторі характеристик тексту.

Масив точок спостереження було розділено на три частини: послідовність А – для навчання моделей, послідовність Б – для їх випробування, послідовність С – екзамени отриманих моделей. Точки спостереження послідовності С не використовувались в процесі синтезу моделей. Навчання моделей відбувалось шляхом розрахунку значень невідомих коефіцієнтів опорних моделей вигляду, що наведені в [14] та комбінованої моделі іншого вигляду, що подані в табл. 1.

Таблиця 1

Опорні моделі

№ з/п	Назва опорної функції	Вигляд опорної функції
1	Лінійна	$f = a_0 + a_1 \cdot y_1 + a_2 \cdot y_2$
2	З добутком многочленів	$f = a_0 + a_1 \cdot y_1 + a_2 \cdot y_2 + a_3 \cdot y_1 \cdot y_2$
3	Неповна квадратична	$f = a_0 + a_1 \cdot y_1 + a_2 \cdot y_2 + a_3 \cdot y_1^2 + a_4 \cdot y_2^2$
4	Квадратична	$f = a_0 + a_1 \cdot y_1 + a_2 \cdot y_2 + a_3 \cdot y_1 \cdot y_2 + a_4 \cdot y_1^2 + a_5 \cdot y_2^2$
5	Комбінована	$f = a \cdot \sqrt{y_1 y_2}$

Селекція моделей проводилась за критерієм регулярності (1) – значенням середньої квадратичної похибки на послідовності Б:

$$\delta = \sqrt{\frac{1}{n} \sum_{i=1}^n (\varphi_i - \varphi_i^*)^2} \quad (1)$$

Результати випробувань сконструйованих моделей оцінювались на послідовності С за допомогою цільової функції-класифікатора вигляду:

$$\varphi(\chi_i, \chi_j) = \begin{cases} 1, & \chi_i = \chi_j; \\ 0, & \chi_i \neq \chi_j, \end{cases} \quad (2)$$

де χ_i – дійсне значення інформативної ознаки даного тексту, χ_j – результат моделювання цієї ознаки.

Всі досліджувані автори були поділені на 5 класів, що об'єднують усі належні їм вибірки тексту. Для розпізнавання цих класів в якості вирішальної функції використовувався вираз (2), що має два значення: "1", якщо вибірка належить до даного класу (тобто "свій") та "0", якщо вона не належить до нього (тобто "чужий").

Результати моделювання мають дисперсію (розкид значень у визначених межах), тому вони можуть не дорівнювати точно "1" або "0", але найбільший показник відноситься до даного класу [15]. Щоб покращити якість розпізнавання, клас повинен включати не одну, а декілька вибірок, таким чином при його ідентифікації відбувається також визначення індивідуальності автора.

Метою застосування методу МГУА було виявити, чи придатний він до такого складного завдання як ідентифікація авторства, як змінюються результати в залежності від кількості вибірок та їх об'єму, від кількості "чужих" авторів, що беруть участь у розпізнаванні. Тобто перевірити результати на стійкість.

Характеристика предметної області

Об'єктом нашого дослідження є використання технології багаторівневого перетворення форми інформації при вирішенні завдань комп'ютерної лінгвістики, предметом – знаходження найефективнішого методу розпізнавання образів для ідентифікації авторства та генерація пробних моделей на прикладі вибірок авторських текстів письменників-новелістів кін. XIX – поч. XX ст.

До дослідження ми залучили 5 авторів (класів), з кожного класу взято по 6 вибірок об'ємом а) 1000 знаків тексту (літери, пунктуаційні знаки, пробіли); б) від 1797 до 2525 знаків; в) від 3055 до 4946 знаків.

Вибірki кожного класу мають випадковий характер. Всі вибірки взяті зі збірника [16]. Використовувалась перша тисяча знаків на сторінці.

В табл. 2 подані характеристики текстів, що використані в процесі досліджень. Формувались три масиви даних а) – для синтезу моделей; б) – для визначення характеристики функції, що розділяє класи; в) – для випробування моделей.

Таблиця 2

Характеристики текстів

Клас (автор)	Вибірка	Назва твору	Сторінка
1	2	3	4
А (Дніпрова Чайка)	А 1	Дівчина-чайка	а) с. 214; б) с. 214; в) с. 214, 215
А (Дніпрова Чайка)	А 2	Дівчина-чайка	а) с. 216; б) с. 216; в) с. 216, 217
А (Дніпрова Чайка)	А 3	Мара	а) с. 221; б) с. 221; в) с. 220, 221
А (Дніпрова Чайка)	А 4	Плавні горять	а) с. 226; б) с. 226; в) с. 225, 226
А (Дніпрова Чайка)	А 5	У школі	а) с. 230; б) с. 230; в) с. 230, 231
А (Дніпрова Чайка)	А 6	У школі	а) с. 234; б) с. 234; в) с. 233, 234
Б (Микола Яцків)	Б 1	Архітвір	а) с. 435; б) с. 435; в) с. 434, 435
Б (Микола Яцків)	Б 2	В казармі	а) с. 418; б) с. 418; в) с. 418, 419

1	2	3	4
Б (Микола Яцків)	Б 3	Де правда:	а) с. 432; б) с. 432; в) с. 431, 432
Б (Микола Яцків)	Б 4	Дівчина на чорнім коні	а) с. 426; б) с. 426; в) с. 425, 426
Б (Микола Яцків)	Б 5	За горою	а) с. 422; б) с. 422; в) с. 422, 423
Б (Микола Яцків)	Б 6	Лісовий дзвін	а) с. 420; б) с. 420; в) с. 420, 421
В (Модест Левицький)	В 1	Куди соколи літають, туди ворон не пускають	а) с. 268; б) с. 268; в) с. 268, 269
В (Модест Левицький)	В 2	Куди соколи літають, туди ворон не пускають	а) с. 272; б) с. 272; в) с. 272, 273
В (Модест Левицький)	В 3	Ніобея	а) с. 285; б) с. 285; в) с. 284, 285
В (Модест Левицький)	В 4	Перша льгота	а) с. 254; б) с. 254; в) с. 254, 255
В (Модест Левицький)	В 5	Тяжка дорога	а) с. 277; б) с. 277; в) с. 276, 277
В (Модест Левицький)	В 6	Щастя Пейсаха Лейдермана	а) с. 263; б) с. 263; в) с. 262, 263
Г (Наталія Кобринська)	Г 1	Брати	а) с. 115; б) с. 115; в) с. 114, 115
Г (Наталія Кобринська)	Г 2	Брати	а) с. 117; б) с. 117; в) с. 116, 117
Г (Наталія Кобринська)	Г 3	Душа	а) с. 106; б) с. 106; в) с. 106, 107
Г (Наталія Кобринська)	Г 4	Душа	а) с. 99; б) с. 99; в) с. 98, 99
Г (Наталія Кобринська)	Г 5	Перша вчителька	а) с. 81; б) с. 81; в) с. 80, 81
Г (Наталія Кобринська)	Г 6	Чортище	а) с. 110; б) с. 110; в) с. 110, 111
Д (Олексій Плющ)	Д 1	Огляд торби	а) с. 588; б) с. 588; в) с. 588, 589
Д (Олексій Плющ)	Д 2	Палкий мисленник і учитель	а) с. 592; б) с. 592; в) с. 592, 593
Д (Олексій Плющ)	Д 3	Палкий мисленник і учитель	а) с. 594; б) с. 594; в) с. 593, 594
Д (Олексій Плющ)	Д 4	Страшна помилка	а) с. 598; б) с. 598; в) с. 598, 599
Д (Олексій Плющ)	Д 5	Страшна помилка	а) с. 603; б) с. 603; в) с. 602, 603
Д (Олексій Плющ)	Д 6	Страшна помилка	а) с. 611; б) с. 611; в) с. 610, 611

Вибір предмету нашого подальшого дослідження (на відміну від попереднього [4]) зумовлений подібністю характеристик: всі автори належать до однієї історико-культурної епохи, мають приблизно однаковий професійний рівень, а їх твори відносяться до одного жанру (новела). Кожному з класів обов'язково будуть притаманні ознаки, так чи інакше тотожні з ознаками інших класів. Процес ідентифікації ускладнюється.

На першому етапі досліджень проводилась попередня обробка даних та їх підготовка до синтезу моделей. Обробка даних проведена за допомогою програми "Analiz", розробленої на кафедрі Інтелектуальних систем прийняття рішень Черкаського національного університету імені Богдана Хмельницького за такою послідовністю дій:

- 1) обчислення загальної кількості ознак тексту (1190) [4];
- 2) розрахунок середнього значення кількості ознак на 1000 знаків тексту за формулою:

$$x^* = \frac{x}{X} \cdot 1000, \quad (3)$$

де x^* – середнє значення на 1000 знаків тексту; x – числове значення ознаки; X – загальна кількість знаків тексту вибірки;

- 3) визначення інформативних ознак [4];
- 4) ранжування ознак за критерієм інформативності.

Серед множини досліджених критеріїв інформативності найкращий результат отриманий після кореляційного ранжування ознак. Ранжування числових значень ознак за коефіцієнтом кореляції до-

зволяє відкинути „шкідливі” для розпізнавання ознаки. Наприклад, погано корельовані ознаки (з коефіцієнтом нижче 0,4) дають слабкий результат.

При опрацюванні даних за допомогою багаторядного алгоритму МГУА [14], наприклад, при вирішальній функції $f = a\sqrt{y_1 y_2}$ на 1 ряді селекції маємо:

$$y_1 = a_1 \sqrt{x_1 x_2}, \quad y_1 = a_2 \sqrt{x_2 x_3},$$

на 2 ряді

$$z = a_3 \sqrt{y_1 y_2}.$$

Якщо під коренем опиниться хоч би одне нульове значення, це непогано для „чужих”, тому що дає нульове кінцеве значення, але для „свого” це буде означати великий відсоток похибки.

Аналіз результатів

У результаті багатьох етапів досліджень, випробування різних комбінацій було виявлено, що величина об'ємів вибірок прямо пропорційно впливає на якість результатів, але це – аксіома математичного моделювання. Найкращі результати ми отримали при синтезі моделі за опорною функцією вигляду 5 (табл. 1). На рис. 1 подані результати випробування моделей авторів текстів, синтез яких проводився на основі опорних функцій цього вигляду. На від моделі подався масив вхідних даних, сформованих із характеристик вибірок текстів 5 авторів послідовності С. На всіх моделях маємо досить стійкі результати розпізнавання авторів тексту незалежно від об'єму вибірок.

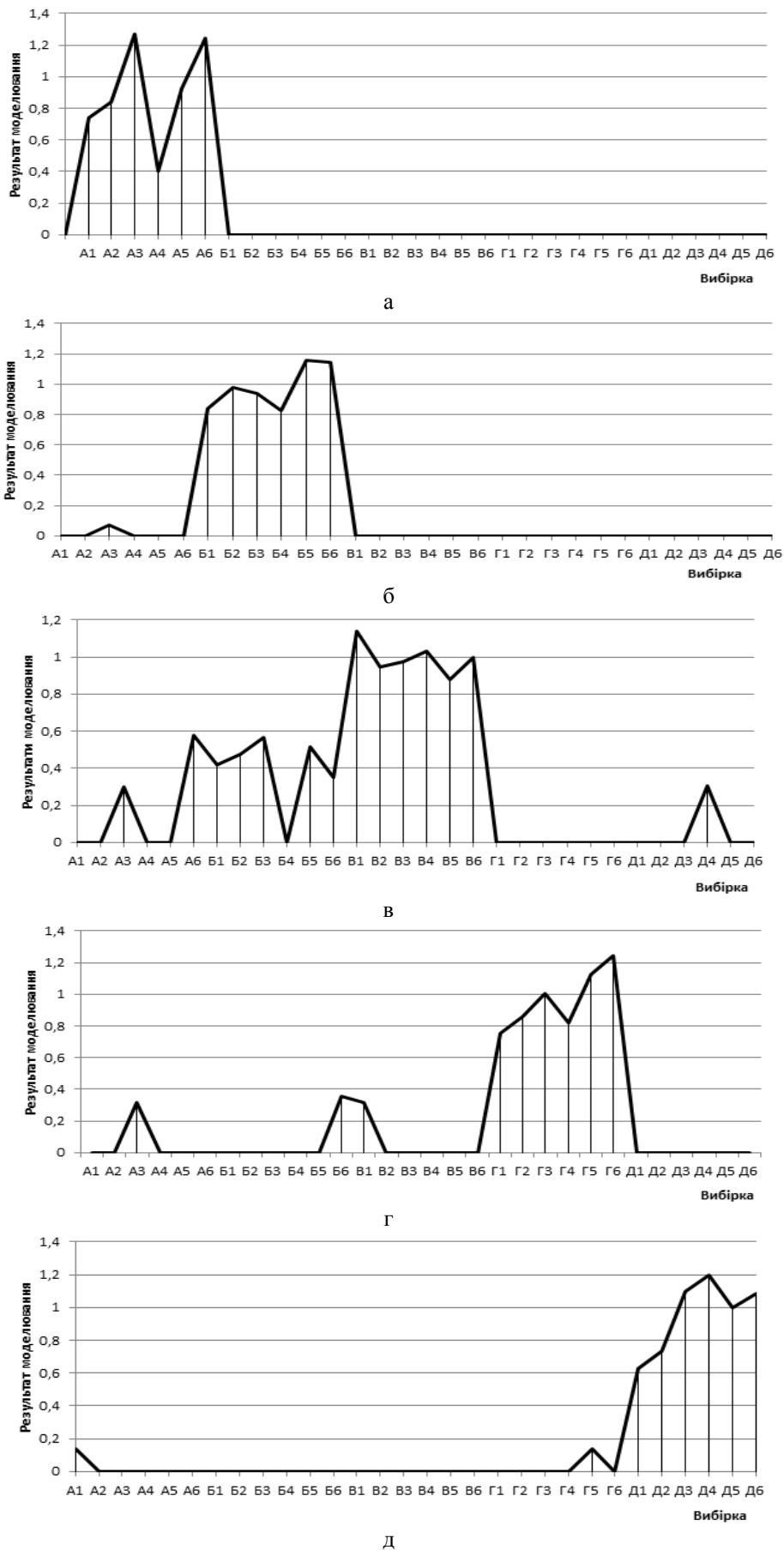


Рис. 1. Результати випробування моделей авторів текстів: а – класу А (Дніпрова Чайка); б – класу Б (Микола Яцків); в – класу В (Модест Левицький); г – класу Г (Наталія Кобринська); д – класу Д (Олексій Плющ)

Висновки

Таким чином, експериментально доведено ефективність використання технології інформаційного моніторингу з багаторівневим перетворенням форми інформації для відображення властивостей автора в структурі моделі текстового повідомлення. Описано використання нового методу формування інформаційних ознак текстових повідомлень в структурі технології інформаційного моніторингу характеристик автора текстового повідомлення, який поєднує в своїй структурі процеси формування масиву інформативних ознак, їх попередню обробку, синтез моделей, їх випробування та використання. Доведено, що різноманітності багаторядного алгоритму МГУА, в поєднанні із новим методом формування масиву інформативних ознак достатньо для отримання адекватних моделей авторів текстових повідомлень. Наступні дослідження доцільно проводити в напрямку адаптування нового методу при виявленні таких характеристик автора тексту, як вік, стать, місце проживання, показники його психологічного та фізичного стану.

Список літератури

1. Луценко Е.В. Атрибуция текстов как обобщение задач идентификации и прогнозирования / Е.В. Луценко // Научный журнал КубГАУ. – 2003. – № 2 (2). – С. 19.
2. Saracevic Tefko. A Course in information consolidation: a handbook for education and training in analysis, synthesis and repackaging of information (preliminary version) / prepared by Tefko Saracevic [for the J General Information Programme and UNISIST]. – Paris: Unesco, 1986. – ix, 128 p.; 30 cm. – (PGI-86/WS/14).
3. Пелецишин А.М. Позичування сайтів у глобальному інформаційному середовищі / А.М. Пелецишин. – Львів: Вид-во Національного університету "Львівська політехніка", 2007. – 258 с.
4. Голуб С.В. Формування показників масиву вхідних даних для ідентифікації авторства текстових повідомлень / С.В. Голуб, О.В. Константиновська, М.С. Голуб // Системи обробки інформації: збірник наукових праць. – Х.: Харківський університет Повітряних сил імені Івана Кожедуба, 2014. – Вип. 2 (118). – С. 89-92.

5. Фомина Н.А. Свойства личности и особенности речевой деятельности / Н.А. Фомина. – Рязань: Узоречье, 2002. – 362 с.

6. Pennebaker J.W. Secret life of pronouns: what our words say about us / J.W. Pennebaker. – N.Y.: Blumberry Press, 2011.

7. Борисов Л.А. Идентификация автора текста по распределению частот буквосочетаний / Л.А. Борисов, Ю.Н. Орлов, К.П. Осминин // Препринты ИПМ им. М.В. Келдыша. – 2013. – № 27. – 26 с. [Электронный ресурс]. – Режим доступа до ресурсу: URL: <http://library.keldysh.ru/preprint.asp?id=2013-27>.

8. ЛингвоАнализатор [Электронный ресурс]. – Режим доступа : URL: <http://www.rusf.ru/books/analysis>.

9. Фоменко В.П. Тихий Дон. Нерешенная загадка русской литературы XX века / В.П. Фоменко, Т.Г. Фоменко. [Электронный ресурс]. – Режим доступа до ресурсу: URL: <http://www.philol.msu.ru/~lex/td/?pid=012281&oid=01228>.

10. Бородин Л.И. Математические методы и компьютер в задачах атрибуции текстов. [Электронный ресурс] / Л.И. Бородин. От Нестора до Фонвизина. Новые методы определения авторства; под ред. Л.В. Милова. – М.: Прогресс, 1994. – Режим доступа до ресурсу: URL: <http://www.textology.ru/library/book.aspx?bookId=11&textId=13>.

11. Севбо И.П. Графическое представление синтаксических структур / И.П. Севбо. – К., 1981. – 192 с.

12. Горелик А.Л. Методы распознавания / А.Л. Горелик, В.А. Скрипкин. – М.: Высшая школа, 1984. – 208 с.

13. Голуб С.В. Багаторівневе моделювання в технологіях моніторингу оточуючого середовища / С.В. Голуб. – Черкаси: Вид. від. ЧНУ імені Богдана Хмельницького, 2007. – 220 с.

14. Ивахненко А.Г. Принятие решений на основе самоорганизации / А.Г. Ивахненко, Ю.П. Зайченко, В.Д. Димитров. – М.: Сов. радио, 1976. – 280 с.

15. Степаненко В.Е. Хроматографическая идентификация индивидуальных соединений методом распознавания образов / В.Е. Степаненко // Журнал аналитической химии. – 1982. – Т. 37, Вып. 12. – С. 2230-2234.

16. Українська новелістика кінця XIX – початку XX ст. Під ред. І.О. Дзевєріна. Бібліотека української літератури. – К: Наукова думка, 1989. – 690 с.

Надійшла до редколегії 6.10.2014

Рецензент: д-р техн. наук, проф. В.М. Рудницький, Черкаський державний технологічний університет, Черкаси.

ОТРАЖЕНИЕ СВОЙСТВ АВТОРА ТЕКСТА
В СТРУКТУРЕ МНОГОПАРАМЕТРИЧЕСКОЙ МОДЕЛИ

С.В. Голуб, А.В. Константиновская, М.С. Голуб

Предложены результаты исследований процесса отражения свойств автора текста в структуре многопараметрической индуктивной модели. Описан процесс использования нового метода формирования информативных характеристик текста и выбора вида опорных функций в технологии информационного мониторинга с многоуровневым превращением вида данных. Выявлено, что разнообразие многоуровневого алгоритма МГУА достаточно для получения адекватных моделей авторов печатных текстов.

Ключевые слова: текстовое сообщение, информативные признаки, модель автора, идентификация, мониторинг.

A REFLECTION OF PROPERTIES OF AUTHOR OF TEXT IS IN THE STRUCTURE OF MULTIPARAMETER MODEL

S.V. Holub, A.V. Constantinovscaya, M.S. Holub

The results of researches of process of reflection of properties of author of text offer in the structure of multiparameter inductive model. The process of the use of new method of forming of informing descriptions of text and choice of type of supporting functions is described in technology of the informative monitoring with multilevel transformation of type of data. It is educed that varieties of multilevel algorithm of GMDH sufficiently for the receipt of adequate authors of the printed texts.

Keywords: text message, informing signs, model of author, authentication, monitoring.