

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В НАУКОВИХ ДОСЛІДЖЕННЯХ

УДК 519.765:519.767:004.93

ІЄРАРХІЧНА КЛАСТЕРИЗАЦІЯ ТЕКСТОВИХ ДОКУМЕНТІВ У ВЕКТОРНОМУ ПРОСТОРІ СЕМАНТИЧНИХ ПОЛІВ

Б. Павлишенко

*Львівський національний університет імені Івана Франка
вул. Драгоманова, 50, 79005 Львів, Україна
pavlsh@yahoo.com*

Запропоновано алгоритм ієрархічної кластеризації текстових документів у векторному просторі, утвореному частотними характеристиками семантичних полів. Доведено ефективність використання низькорозмірних матриць семантичних характеристик документів в алгоритмах кластерного аналізу текстових масивів.

Ключові слова: інтелектуальний аналіз текстів, кластерний аналіз, частотний словник, семантичні поля.

Розвиток методів аналізу текстових документів є перспективним напрямом сучасних інформаційних технологій. Один із таких методів – кластерний аналіз, який дає змогу виявити групи об'єктів, які подібні між собою за певними критеріями [1, 2, 3]. Кластерний аналіз ефективний у разі вивчення структури текстових масивів [4, 5]. Для зображення текстових документів часто використовують модель векторного простору [5]. У цій моделі кожний документ відображається як вектор у багатовимірному просторі, кожний вимір якого відповідає квантитативній характеристиці лексеми зі словників аналізованих текстових масивів. Текстовий масив можна зобразити у вигляді матриці слів та документів, у якій стовпці визначають документи, а рядки – частоти лексем у цих документах. Тоді кожен стовпець є вектором частот лексем для заданого документа, який задають номером стовпця. Мірою відстані між двома документами може бути кут між векторами цих документів в утвореному векторному просторі. Такий підхід має також ряд проблем, зокрема, розмірність аналізованого простору є великою, оскільки зумовлена розміром словника. Документи також можуть бути квантитативно близькими не тільки за частотами окремих лексем, а також за характеристиками заданих лексемних об'єднань, наприклад, семантичних полів [6, 7]. Пошук комплексних характеристик текстових документів є важливим, зокрема в разі аналізування авторства текстів, оскільки лексемний частотний спектр творів може бути однаковим, проте відрізнятися за характеристиками комбінованих лексемних груп.

Формулювання задачі. Для виявлення нових підмножин метаданих, які будуть ефективними в алгоритмах аналізу текстових масивів, розглянемо структурний поділ лексемного складу за семантичними полями. Щоб виконати аналіз семантичного

простору, побудуємо теоретико-множинну модель семантичних полів. На підставі цієї моделі реалізуємо алгоритм агломеративної кластеризації вибірки текстових документів. Проаналізуємо отримані результати з погляду ефективності кластеризації в просторі семантичних полів для тестової вибірки документів.

Теоретико-множинна модель семантичних полів. Розглянемо модель на підставі теорії множин, яка описує сукупність текстових документів, лексемний склад та семантичні поля. Нехай існує певний словник лексем, які трапляються у текстових масивах. Опишемо цей словник як упорядковану множину

$$W = \{ w_i \mid i = 1, 2, \dots, N_w \}. \quad (1)$$

Сукупність текстових документів опишемо такою множиною:

$$D = \{ d_j \mid j = 0, 1, 2, \dots, N_d \}. \quad (2)$$

Під документом з $j=0$ будемо вважати документ з нейтральним текстом, який відповідає лінгвостатистичні нормі. Введемо множину семантичних полів

$$S = \{ s_k \mid k = 1, 2, \dots, N_s \}. \quad (3)$$

Під семантичним полем розуміють таку множину лексем, які об'єднані певним спільним поняттям [6, 7]. Прикладом семантичних полів може бути поле руху, поле комунікації, поле сприйняття та ін.

Документ d_j з множини текстових документів D можна уявити як упорядковану множину слів, порядок елементів якої відповідає порядку слів у цьому документі:

$$T_j^d = \{ t_{lj} \mid l = 1, 2, \dots, N_j^d \}. \quad (4)$$

Упорядкований за алфавітом словник текстового документа d_j розглянемо як мультимножину W_j^d над множиною словника W :

$$W_j^d = \{ n_{ij}^{wd} (w_i) \mid w_i \in d_j, i = 1, 2, \dots, N_w \}, \quad (5)$$

де n_{ij}^{wd} – кількість входжень лексеми w_i зі словника W в множину лексем текстового документа d_j , яку можна визначити як

$$n_{ij}^{wd} = \sum_{l=1}^{N_j^d} f_{wd}(t_{lj}, w_i), \quad (6)$$

$$\text{де } f_{wd}(t_{lj}, w_i) = \begin{cases} 1, & t_{lj} = w_i, \\ 0, & w_{lj}^d \neq w_i. \end{cases} \quad (7)$$

Уведемо відображення лексемного складу словника W на множину семантичних полів S за допомогою деякого оператора U_{ws} :

$$U_{ws} : w_i \rightarrow s_k, \quad i = 1, 2, \dots, N_w; k = 1, 2, \dots, N_s. \quad (8)$$

Оператор U_{ws} задамо таблицею, яка визначена експертним лексикографічним аналізом [6, 7]. Лексемний склад семантичного поля s_k визначимо як

$$W_k^s = \left\{ w_i \mid w_i \xrightarrow{U_{ws}} s_k, i = 1, 2, \dots, N_w \right\}. \quad (9)$$

Формування семантичних полів відбувається на підставі лексикографічного аналізу, і часто трапляється, що одна лексема може належати декільком семантичним полям. Це приводить до непустиї множини перерізу множин лексемних складів семантичних полів

$$W_p^s \cap W_q^s \neq \emptyset, \quad p \neq q; \quad p, q \in (1, 2, \dots, N_s). \quad (10)$$

Множину образів відображення U_{ws} розглянемо як мультимножину над множиною семантичних полів S :

$$S_f = \{ n_k^s(s_k) \mid k = 1, 2, \dots, N_s \}, \quad (11)$$

де n_k^s – кількість лексем словника W , які відносять до семантичного поля S_k :

$$n_k^s = \sum_{i=1}^{N_w} f_s(w_i, s_k), \quad (12)$$

$$\text{де } f_s(w_i, s_k) = \begin{cases} 1, & w_i \in W_k^s, \\ 0, & w_i \notin W_k^s. \end{cases} \quad (13)$$

Уведемо мультимножину образів відображення U_{ws} семантичних полів для окремого документа d_j ,

$$S_j^d = \{ n_{kj}^{sd}(s_k) \mid k = 1, 2, \dots, N_s \}, \quad (14)$$

де n_{kj}^{sd} – кількість лексем семантичного поля s_k в лексемному складі документа d_j

$$n_{kj}^{sd} = \sum_{l=1}^{N_j^t} f_s(t_{lj}, s_k), \quad (15)$$

$$\text{де } f_s(t_{lj}, s_k) = \begin{cases} 1, & t_{lj} \in W_k^s, \\ 0, & t_{lj} \notin W_k^s. \end{cases} \quad (16)$$

Уведемо деяку множину P квантитативних ознак, за допомогою яких можна порівнювати характеристики текстових документів. Також уведемо оператор відображення лексемного словника W на множину квантитативних ознак у масиві документів

$$U_{wd} : w_i \rightarrow p_{ij}^{wd}, \quad i = 1, 2, \dots, N_w, \quad j = 1, 2, \dots, N_d. \quad (17)$$

У загальному випадку величина p_{ij}^{wd} може мати довільне походження квантитативної характеристики. Надалі будемо розглядати цю величину як текстову частоту лексеми w_i в текстовому документі d_j , яка визначена такою функціональною залежністю:

$$p_{ij}^{wd} = \frac{n_{ij}^{wd}}{N_j^t}. \quad (18)$$

Аналогічно введемо оператор відображення семантичного складу S_j^d текстового документа d_j на множину квантитативних ознак:

$$U_{sd} : s_k \rightarrow p_{kj}^{sd}, \quad k = 1, 2, \dots, N_s, \quad j = 1, 2, \dots, N_d. \quad (19)$$

Величина p_{kj}^{sd} визначає структурну частоту лексем семантичного поля s_k у текстовому документі d_j . Визначимо p_{kj}^{sd} за такою формулою:

$$p_{kj}^{sd} = \sum_{i=1}^{N_w} p_{ij}^{wd} f_s(w_i, s_k), \quad (20)$$

$$\text{де } f_s(w_i, s_k) = \begin{cases} 1, & w_i \in W_k^s, \\ 0, & w_i \notin W_k^s. \end{cases} \quad (21)$$

Сукупність значень p_{ij}^{wd} утворює матрицю типу ознака - документ

$$M_{wd} = \left(p_{ij}^{wd} \right)_{i=1, j=1}^{N_w, N_d}. \quad (22)$$

У матриці M_{wd} роль ознаки відіграє текстова частота лексеми. Введемо вектор

$$V_j^w = \left(p_{1j}^{wd}, p_{2j}^{wd}, \dots, p_{N_w j}^{wd} \right). \quad (23)$$

Такий вектор відображає документ d_j в N_w -вимірному просторі текстових документів. Сукупність значень p_{kj}^{sd} утворює іншу матрицю ознака - документ, у якій ознаками є частоти семантичних полів у документах:

$$M_{sd} = \left(p_{kj}^{sd} \right)_{k=1, j=1}^{N_s, N_d}. \quad (24)$$

Вектор

$$V_j^s = \left(p_{1j}^{sd}, p_{2j}^{sd}, \dots, p_{N_s j}^{sd} \right) \quad (25)$$

відображає документ d_j в N_s -вимірному просторі текстових документів.

Запропонована модель дає змогу визначити матриці типу ознака - документ для частотних характеристик лексемного та семантичного складу текстових документів на підставі подібних теоретичних підходів. Матриці частоти лексем - документ та частоти семантичних полів - документ описують масив текстових документів у різних фазових просторах із суттєво різною розмірністю. Перехід до простору семантичних полів не тільки зменшує розмірність задачі аналізу текстів, а також уводить новий базис для текстових характеристик. У семантичному базисі можна спостерігати якісно нові групування текстових документів. Розгляд таких групувань може бути ефективним в алгоритмах комплексного аналізу текстів.

Відображення (8) лексемного складу словника на множину семантичних полів, яке задають табличним оператором U_{ws} , відповідає правилу підстановки для безконтекстної граматики

$$A \rightarrow \beta, \quad (26)$$

де A – символ нетермінального алфавіту, β – символ термінального алфавіту. З цих позицій множину семантичних полів можна розглядати як підмножину нетермінального алфавіту деякої формальної граматики

$$G = (V_N, V_T, P, S_G), \quad (27)$$

де V_N – множина нетермінальних символів; V_T – множина термінальних символів; P – множина правил підстановки; S_G – початковий кореневий символ. Отже, множину семантичних полів можна розглядати як підмножину множини нетермінальних символів формальної граматики:

$$S \subset V_N \quad (28)$$

Такий підхід дає змогу вводити семантичні поля в алгоритми структурного розпізнавання текстових образів за допомогою теорії скінченних автоматів.

Належність елементів множини словника з різною кількістю входжень до множини лексем деякого документа утворює структуру текстових документів. Відображення множин лексем кожного з документів на множину квантитативних ознак утворює структурний квантитативний поділ множини документів. Такий поділ можна виявити за допомогою алгоритмів кластеризації документів у вибраному просторі. Очевидно поділ множини словника за допомогою відображення в множині семантичних полів утворює новий структурний семантичний поділ текстових документів. Це вносить нові можливості аналізування текстових масивів за семантичними ознаками.

Модель кластеризації текстових документів у семантичному просторі. Частоти семантичних полів утворюють семантичний векторний простір, у якому кожен документ можна зобразити за допомогою вектора V_j^s , який описують виразом (25). Очевидно, розмірність семантичного векторного простору є суттєво меншою від розмірності простору, утвореного лексемами частотного словника, оскільки $N_s < N_w$. Крім того, семантичний простір відображає додаткову комбіновану складову характеристику текстових документів, яка може утворювати іншу кластерну структуру порівняно зі структурою в лексемному просторі.

Розглянемо групування документів за семантичними ознаками за допомогою алгоритму ієрархічної кластеризації. Нехай ϵ – множина текстових документів D , яку описує вираз (2), та множина кластерів

$$C = \{ c_m \mid m = 0, 1, 2, \dots, N_c \}. \quad (29)$$

Необхідно побудувати відображення множини документів на множину кластерів:

$$U_{DC} : D \rightarrow C. \quad (30)$$

Відображення U_{DC} задає модель даних, яка є розв'язком задачі кластеризації [1, 2, 4]. Кожний елемент c_m множини кластерів C складається з підмножини текстових документів, які подібні між собою відповідно до деякої кількісної міри подібності r :

$$c_m = \{ d_i, d_j \mid d_i \in D, d_j \in D, r(d_i, d_j) < \epsilon \}, \quad (31)$$

де ϵ визначає деякий поріг для включення документів у кластер. Величина $r(d_i, d_j)$ є відстанню між елементами d_i та d_j , для якої виконуються такі умови:

$$\begin{aligned} r(d_i, d_j) &\geq 0, \forall i, j, \\ r(d_i, d_j) &= 0, i = j, \\ r(d_i, d_j) &= r(d_j, d_i), \\ r(d_i, d_j) &\leq r(d_i, d_k) + r(d_k, d_j). \end{aligned} \quad (32)$$

Якщо виконується умова

$$r(d_i, d_j) < \epsilon, \quad (33)$$

то елементи вибірки вважають подібними і належними до спільного кластера. В іншому випадку елементи містяться в різних кластерах. Матриця

$$R = \{ r_{ij} = r(d_i, d_j) \} \quad (34)$$

є матрицею відмінностей в алгоритмі кластеризації. Очевидно діагональні елементи цієї матриці дорівнюють нулю. Оскільки на множині текстових документів введено поняття відстані, то кожен документ уявляють у вигляді точки в N_s -вимірному просторі R^{N_s} семантичних полів. Є декілька методів обчислення мір близькості точок в N_s -вимірному просторі. Евклідова відстань

$$r_e(d_i, d_j) = \sqrt{\sum_{k=1}^{N_s} (p_{ki}^{sd} - p_{kj}^{sd})^2} \quad (35)$$

Відстань Немінга обчислюють за такою формулою:

$$r_n(d_i, d_j) = \sum_{k=1}^{N_s} |p_{ki}^{sd} - p_{kj}^{sd}| \quad (36)$$

Подібність між двома текстовими документами в N_s -вимірному просторі також визначена кутом між векторами цих документів, і за кількісну міру можна взяти косинус цього кута. Оскільки відстань між двома однаковими документами повинна дорівнювати нулю, то подібність між векторами документів можна визначити так:

$$r_c(d_i, d_j) = 1 - \frac{\sum_{k=1}^{N_s} (p_{ki}^{sd} \cdot p_{kj}^{sd})}{\sqrt{\sum_{k=1}^{N_s} (p_{ki}^{sd})^2} \cdot \sqrt{\sum_{k=1}^{N_s} (p_{kj}^{sd})^2}} \quad (37)$$

Розглянемо ієрархічний метод агломеративної кластеризації. На першому кроці всю множину текстових документів розглядають як множину кластерів:

$$c_1 = \{ d_1 \}, c_2 = \{ d_2 \}, \dots, c_{Nd} = \{ d_{Nd} \}. \quad (38)$$

На наступному кроці два близькі один до одного документи (наприклад d_p і d_q) об'єднують в один спільний кластер, нова множина на цьому кроці вже складається з $N_d - 1$ кластерів і має вигляд

$$c_1 = \{ d_1 \}, c_2 = \{ d_2 \}, \dots, c_p = \{ d_p, d_q \}, \dots, c_{Nd-1} = \{ d_{Nd-1} \}. \quad (39)$$

Повторюючи кроки, на яких об'єднують кластери, отримаємо множину з N_c кластерів. Процес об'єднання кластерів завершується на тому кроці алгоритму, коли жодна пара кластерів не відповідає порогу об'єднання для міри близькості елементів. На кожній ітерації алгоритму необхідно робити перерахунок між кластерами. Оскільки кластери можуть складатися з декількох об'єктів, то існують різні методи формування та об'єднання кластерів на підставі відстаней між об'єктами всередині кластера. Метод найближчого сусіда полягає у виборі найменшої відстані між двома кластерами p і q :

$$r(p, q) = \min \{ r(d_{pi}, d_{qj}), i \in (1, 2, \dots, N_p), j \in (1, 2, \dots, N_q) \}. \quad (40)$$

У методі дальніх сусідів вибирають відстань між двома найдалішими сусідами

$$r(p, q) = \max \{ r(d_{pi}, d_{qj}), i \in (1, 2, \dots, N_p), j \in (1, 2, \dots, N_q) \}. \quad (41)$$

Середню відстань між кластерами розраховують за формулою

$$r(p, q) = \frac{1}{N_p N_q} \sum_{i=1}^{N_p} \sum_{j=1}^{N_q} r(d_{pi}, d_{qj}). \quad (42)$$

У центроїдному методі розраховують Евклідову відстань між центрами двох кластерів

$$r(p, q) = \|\bar{x}_p - \bar{x}_q\|_2, \bar{x}_p = \frac{1}{N_p} \sum_{i=1}^{N_p} x_{pi}, \bar{x}_q = \frac{1}{N_q} \sum_{i=1}^{N_q} x_{qi}, \quad (43)$$

де $\|\cdot\|_2$ – Евклідова відстань; x_{pi} – вектор відображення i -го документа в p -му кластері у векторному семантичному просторі, x_{qi} – вектор відображення i -го документа в q -му кластері. У методі Варда обчислюють квадрати евклідових відстаней від окремих документів до центра кожного кластера. Далі ці відстані підсумовують. У новий кластер об'єднують ті кластери, у разі об'єднання яких виходить найменший приріст суми квадратів відстаней.

Графічним зображенням результату ієрархічної кластеризації є дендрограма, на якій відображають процес агломеративного об'єднання кластерів. По осі абсцис відкладають номери кластерів, а по осі ординат – відстані між кластерами. За певних значень відстаней починається об'єднання кластерів. Зі збільшенням міжкластерної відстані кластери об'єднуються аж до повного злиття в один кластер. Тому для отримання інформативної кластерної структури вибирають деякий поріг міжкластерної відстані, за якого утворюється оптимальна з погляду аналізу текстових масивів кластерна структура. Наприклад, у ході дослідження можливості кластеризації текстових документів за авторами доцільно взяти таке порогове значення міжкластерної відстані, за якого утворюється кількість кластерів, що дорівнює кількості аналізованих авторів. У дослідженнях ми обмежили кластерну структуру рівнем з 20 кластерами.

Експериментальна частина. Для аналізу ефективності розглянутих алгоритмів кластеризації взято текстову вибірку 155 художніх творів анломовної класики чотирьох відомих авторів (Ч. Діккенс, Джек Лондон, В. Скотт, М. Твен). Для утворення семантичного простору сформовано 15 семантичних полів, у які входить близько 5 000 неозначених форм дієслова. Деталізація літературних та лексикографічних характеристик вхідних даних не є суттєвою для аналізу можливості кластерного структурування даних, тому для подальшого аналізу будемо розглядати лише статистичні характеристики текстових документів. Для кожного документа розраховано частотні словники, на підставі яких розраховано частотні спектри семантичних полів документів. Отже, кожний документ розглядають як вектор в 15-вимірному семантичному просторі. На рис. 1 показано дендрограму ієрархічної кластеризації центроїдним методом, а на рис. 2 – дендрограму ієрархічної кластеризації методом Варда. По осі абсцис відкладено номери кластерів, а по осі ординат – міжкластерні відстані. Наведені дендрограми обмежені 20 кластерами. Як випливає з рисунків, метод формування кластерів суттєво впливає на кластерну структуру.

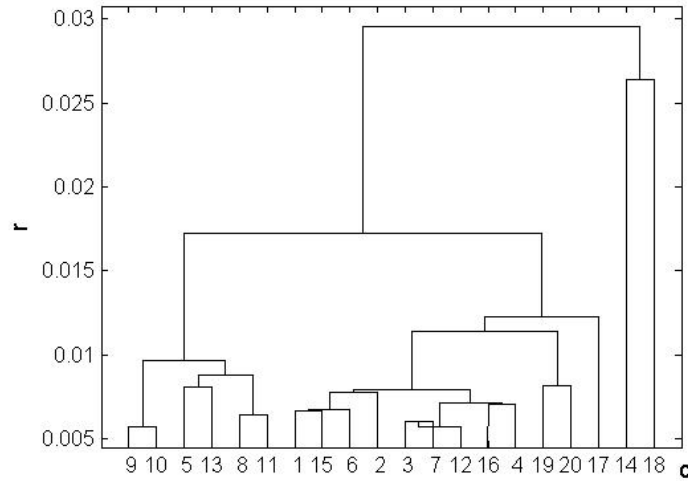


Рис. 1. Дендрограма кластеризації масиву текстових документів центроїдним методом.

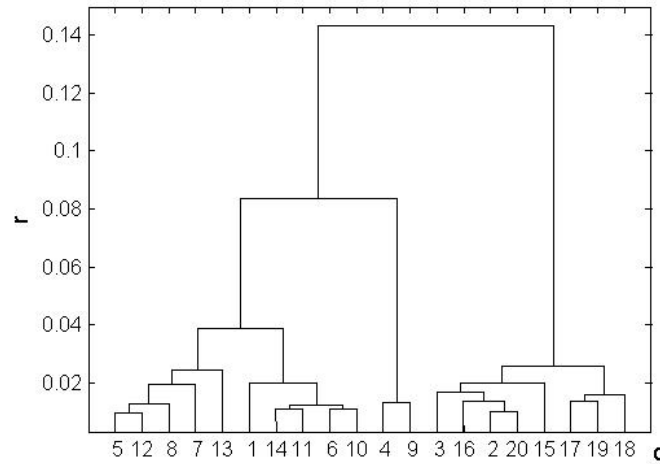


Рис. 2. Дендрограма кластеризації масиву текстових документів методом Варда.

Для вибору оптимального методу кластеризації необхідно зіставляти утворену кластерну структуру з іншими класифікаціями текстових документів. Проаналізуємо класифікацію текстових документів за авторами. Виберемо таку порогову міжкластерну відстань, за якої утворюється кількість кластерів, що дорівнює кількості авторів текстів у досліджуваній вибірці. В аналізованому випадку це чотири кластери. На рис. 3 показано розподіл кількості текстових документів за чотирма кластерами, утвореними методом Варда.

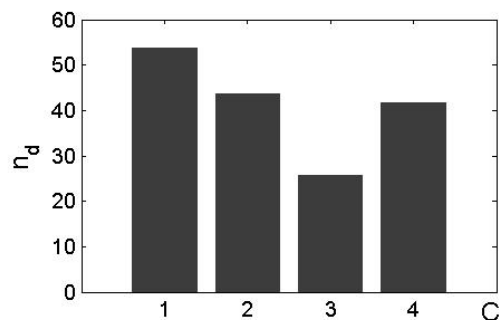


Рис. 3. Розподіл кількості текстових документів за кластерами.

На рис. 4 показано розподіл текстів за авторами (1–Ч. Дікенс, 2–Д. Лондон, 3–В. Скотт, 4–М. Твен) в кожному з чотирьох кластерів, утворених методом Варда. Як випливає з наведених даних, текстів автора № 3 нема в кластері 1 і вони максимально сконцентровані в кластері 4. Кластер 3 складається переважно із текстів автора № 2 і т.д. Такий нерівномірний розподіл текстів за авторами в кластерах свідчить про те, що кластерна структура документів у семантичному просторі відображає класифікаційну структуру документів за авторами.

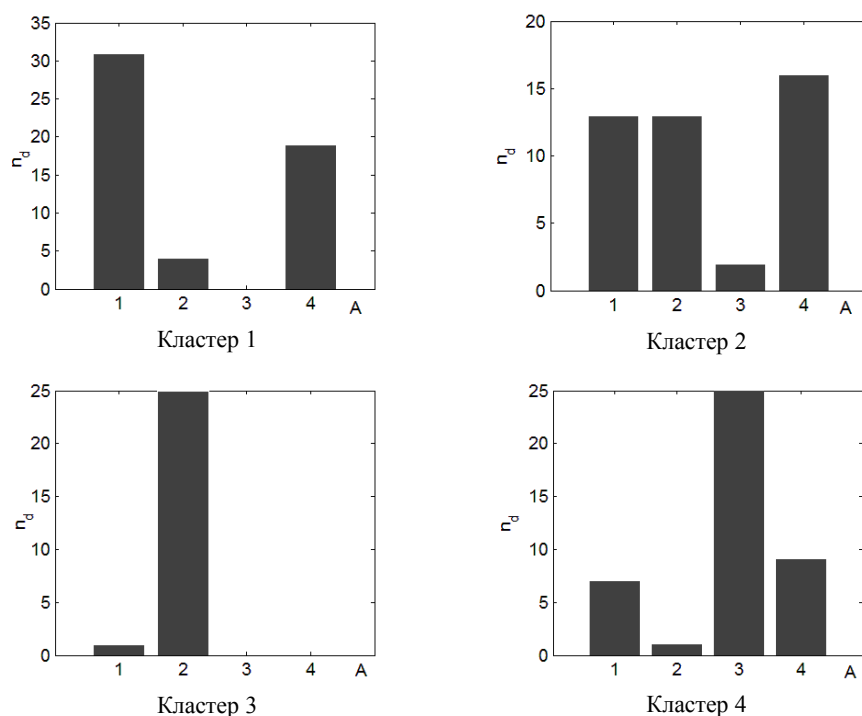


Рис. 4. Розподіл кількості текстових документів за авторами в досліджуваних кластерах.

Отже, запропонована модель кластеризації текстових документів у семантичному просторі дає змогу отримувати новий структурний поділ документів за семантичними ознаками в просторі суттєво меншої розмірності, ніж простір, утворений лексемним складом текстової вибірки. Такий структурний поділ відображає класифікацію документів за новими ознаками документів, зокрема, за авторством текстів. Текстові вибірки деяких авторів можуть мати свої чіткі області в семантичному просторі. Це дає змогу вивчати авторство текстових документів через аналіз належності семантичних векторів цих документів до заданих областей простору семантичних полів.

1. *Ким Д.О.* Факторный, дискриминантный и кластерный анализ / Д.О. Ким, Ч.У. Мьюллер, У.Р. Клекка – М.: Финансы и статистика, 1989. – 215 с.
2. *Жамбю М.* Иерархический кластер-анализ и соответствия / М. Жамбю – пер. с фр. – М.: Финансы и статистика, 1988. – 342 с.
3. *Павлишенко Б.М.* Векторизація кластерів на растрових зображеннях електронної мікроскопії // Б.М. Павлишенко – Вісн. Львів. ун-ту. Сер. фіз. – 2007. – Вип. 40. – С. 117–121.
4. *Брасегян А.А.* Анализ данных и процессов: учеб. Пособие // А.А. Брасегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров – СПб.:БХВ-Петербург, 2009. – 512 с.
5. *Pantel Patrick, Turney Peter D.* From Frequency to Meaning: Vector Space Models of Semantics [Електронний ресурс] // arXiv:1003.1141, 2010, <http://arxiv.org/abs/1003.1141>.
6. *Вердиева З.Н.* Семантические поля в современном английском языке / З.Н. Вердиева – М.: Высшая школа, 1986. – 120 с.
7. *Левицкий В.В.* Экспериментальные методы в семасиологии / В.В. Левицкий, И.А. Стернин – Воронеж: Изд-во Воронеж. ун-та, 1989. – 192 с.

HIERARCHICAL CLUSTERIZATION OF TEXT DOCUMENTS IN THE VECTOR SPACE CREATED BY SEMANTIC FIELDS

B. Pavlyshenko

*Ivan Franko National University of L'viv,
Dragomanov Str., 50, UA-79005 Lviv, Ukraine.
pavlsh@yahoo.com*

The algorithm of hierarchical clusterization of text documents in the vector space created by frequencies characteristics of semantic fields has been proposed. The effectiveness of using low dimension matrix of documents' semantic characteristics in the algorithms of text arrays cluster analysis has been shown.

Key words: text mining, cluster analysis, frequency dictionary, semantic fields.

**ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ ТЕКСТОВЫХ ДОКУМЕНТОВ
В ВЕКТОРНОМ ПРОСТРАНСТВЕ СЕМАНТИЧЕСКИХ ПОЛЕЙ****Б. Павлышенко**

*Львовский национальный университет имени Ивана Франко,
ул. Драгоманова, 50, 79005 Львов, Украина.
pavlsh@yahoo.com*

Предложен алгоритм иерархической кластеризации текстовых документов в векторном пространстве образованном частотными характеристиками семантических полей. Показана эффективность использования низкоразмерных матриц семантических характеристик документов в алгоритмах кластерного анализа текстовых массивов.

Ключевые слова: интеллектуальный анализ текстов, кластерный анализ, частотный словарь, семантические поля.

Стаття надійшла до редколегії 10.01.2011
Прийнята до друку 29.03.2011