УДК 519.765:004.89

Б.М. ПАВЛИШЕНКО

АНАЛІЗ СЕМАНТИЧНИХ ОБРАЗІВ У МАСИВАХ ТЕКСТОВИХ ОБ'ЄКТІВ ЗА ДОПОМОГОЮ КВАНТОВИХ ОБЧИСЛЕНЬ

Анотація. *Проведено аналіз семантичних образів у масивах текстових об'єктів із використанням елементів квантового алгоритму Гровера. Показано, що реалізація квантових алгоритмів для деякого класу задач такого аналізу дає можливість експоненційно зменшити об'єм необхідної пам'яті та поліноміально зменшити час виконання алгоритму у порівнянні із класичними алгоритмами внаслідок реалізації квантового паралелізму.*
Ключові слова: *квантові обчислення, алгоритм Гровера, аналіз текстів, семантичі поля.*

Аннотация. *Проведен анализ семантических образов в массивах текстовых объектов с использованием элементов квантового алгоритма Гровера. Показано, что реализация квантовых алгоритмов для некоторого класса задач такого анализа дает возможность экспоненциально уменьшить объем требуемой памяти и полиномиально уменьшить время выполнения алгоритма по сравнению с классическими алгоритмами в результате реализации квантового параллелизма.*
Ключевые слова: *квантовые вычисления, алгоритм Гровера, анализ текстов, семантические поля.*

Abstract. *The analysis of semantic patterns in the text objects arrays using Grover algorithms elements has been performed. It is shown that implementation of quantum algorithms for some classes of problems gives the ability to decrease required memory exponentially and decrease the time of algorithm performance polynomial in comparison with the classical algorithms due to quantum parallelism.*
Keywords: *quantum calculations, Grover's quantum algorithm, texts analysis, semantic fields.*

1. Вступ

Пошук нових алгоритмів аналізу слабоструктурованих даних, зокрема, текстового типу, є одним із актуальних напрямів сучасних інформаційних технологій. Якісно нові підходи до такого аналізу можливі при використанні квантових обчислень, теорія яких активно розвивається останнім часом [1, 2]. В інтелектуальному аналізі текстових масивів можна виділити деякий клас задач, на якому можна показати ефективність реалізації квантових алгоритмів. Вони дають можливість суттєво пришвидшити розв'язок внаслідок реалізації квантового паралелізму та заплутаності квантових станів. Одним із відомих квантових алгоритмів є алгоритм Гровера для пошуку даних у неструктурованій базі даних [3, 4]. У порівнянні із класичним алгоритмом, в якому пошук здійснюється за час $O(N)$, алгоритм Гровера дає квадратичне прискорення розв'язку, що є актуальним при великих значеннях N. Особливістю квантового алгоритму Гровера є те, що він являється ймовірнісним алгоритмом, тобто дає правильний розв'язок із заданою ймовірністю, яка може бути збільшена шляхом повторного використання алгоритму. Алгоритм Гровера може бути складовою частиною інших квантових алгоритмів, зокрема, в роботі [5] він використовується для еволюційного аналізу кліткових автоматів. У сучасних алгоритмах аналізу текстів часто використовують модель векторного простору. Основна ідея цієї моделі полягає у представленні кожного текстового документа у вигляді вектора у деякому векторному просторі [6–8]. Вважають, що точки, які є близькі між собою у цьому просторі, відображають семантично близькі документи і навпаки – семантично близькі документи відображаються близькими точками у фазовому просторі. В основі використання векторної моделі лежить статистична гіпотеза, яка полягає в тому, що статистичні характеристики використання слів відображають ті поняття, які відображаються у цих текстах [6].

2. Постановка задачі

Розглянемо деякі моделі задач, які виявляють ефективність квантових обчислень. Нехай існує відображення масиву текстових документів, в якому тексти представлені у вигляді частотних спектрів лексем словникового складу або у вигляді характеристик лексемних об'єднань. Такими об'єднаннями можуть бути, наприклад, семантичні поля. Сукупність частот семантичних полів текстового об'єкта можна розглядати як семантичний образ, який характеризує текстовий об'єкт. Нехай завдання полягає у виявленні існування текстових об'єктів із заданими властивостями. Розглянемо семантичну векторну модель масиву текстових об'єктів. Проаналізуємо базові елементи квантових обчислень, які можуть бути використані в алгоритмах аналізу текстів. Створимо модель квантового запису масиву семантичного вектора текстового об'єкта. Розглянемо пошук заданих семантичних образів у квантовій базі даних текстових об'єктів, використовуючи елементи алгоритму Гровера. Проаналізуємо ефективність квантового алгоритму пошуку семантичних образів текстових об'єктів у порівнянні з класичними алгоритмами.

3. Семантична векторна модель текстових об'єктів

Під текстовими об'єктами будемо розглядати текстові документи у вигляді сукупності логічно пов'язаних лексем, об'єднаних у масиві речень. Розглянемо представлення текстових об'єктів у вигляді векторів семантичного простору. В основі такого представлення лежить статистична семантична гіпотеза [6], яка припускає, що статистичні характеристики вживання лексем можуть бути використані для визначення змісту сказаного. Якщо деякі частини тексту мають подібні вектори в частотних матрицях, тоді ці частини мають подібні значення. Сукупність текстових об'єктів опишемо такою множиною:

$$T = \{ t_j \mid j = 1,2...,N_t \}, \tag{1}$$

де N_t – кількість об'єктів. Упорядкований за алфавітом словник текстового об'єкта t_j розглянемо як мультимножину W_j^t над множиною словника W:

$$W_j^t = \{ n_{ij}^{wt}(w_i) \mid w_i \in t_j, i = 1,2...,N_w \}, \tag{2}$$

де n_{ij}^{wt} – кількість входжень лексеми w_i із словника W у множину лексем текстового документа t_j. Введемо множину семантичних полів:

$$S = \{ s_k \mid k = 1,2..., N_s \}. \tag{3}$$

Під семантичним полем розуміють множину лексем, об'єднаних деяким спільним поняттям [9, 10]. Прикладом семантичних полів можуть бути поле руху, поле комунікації, поле сприйняття та ін. Відображення лексемного складу словника W на множину семантичних полів S (3) задамо таблицею, яка визначається експертним лексикографічним аналізом. Лексемний склад семантичного поля s_k визначимо як

$$W_k^s = \left\{ w_i \mid w_i \xrightarrow{U_{ws}} s_k, i = 1,2..., N_w \right\}. \tag{4}$$

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

$$S_f = \{ n_k^s(s_k) \mid k = 1,2..., N_s \}, \tag{5}$$

де n_k^s – кількість лексем словника W, які відносять до семантичного поля s_k. Введемо мультимножину образів відображення U_{ws} семантичних полів для окремого текстового об'єкта t_j:

$$S_j^t = \left\{ n_{kj}^{st}(s_k) \,|\, k = 1,2 \ldots, N_s \right\}, \tag{6}$$

де n_{kj}^{st} – кількість лексем семантичного поля s_k в лексемному складі документа t_j. Аналогічно до [8] введемо оператор відображення семантичного складу S_j^t текстового об'єкта t_j на множину квантитативних ознак:

$$U_{sd} : s_k \to p_{kj}^{st},\, k = 1,2,\ldots N_s,\; j = 1,2\ldots N_t. \tag{7}$$

Величина p_{kj}^{st} визначає структурну частоту лексем семантичного поля s_k у текстовому документі d_j. Сукупність значень p_{kj}^{st} утворюють матрицю ознака-об'єкт, у якій ознаками виступають частоти семантичних полів в об'єктах:

$$M_{sd} = \left(p_{kj}^{st} \right)_{k=1,j=1}^{N_s, N_d}. \tag{8}$$

Вектор

$$V_j^s = \left(p_{1j}^{st}, p_{2j}^{st}, \ldots, p_{N_s j}^{st} \right) \tag{9}$$

відображає текстовий об'єкт t_j в N_s-мірному просторі текстових об'єктів. Використання векторного представлення (9) дає можливість пошуку подібних об'єктів у векторному просторі з базисом, утвореним частотними характеристиками семантичних полів. Цей базис має суттєво меншу розмірність у порівнянні з базисом, утвореним частотними характеристиками лексем словника текстових масивів. Це дає можливість зменшити кількість необхідних обчислень в алгоритмах аналізу текстів.

Складові семантичного вектора (9) є дійсними величинами, які можуть набувати значень у проміжку [0,1]. Однак можна знайти відображення цього вектора на множину бінарних значень:

$$V_j^s \to V_j^{sb},$$

де

$$V_j^{sb} = \left(p_{1j}^{stb}, p_{2j}^{stb}, \ldots, p_{N_s j}^{stb} \right), \quad p_{ij}^{stb} \in \left\{ 0,1 \right\}. \tag{10}$$

У найпростішому випадку відображення (10) може здійснюватись так:

$$p_{ij}^{stb} = \begin{cases} 0, \, p_{ij}^{st} < (p_i^{st})_{th}, \\ 1, \, p_{ij}^{st} \geq (p_i^{st})_{th}, \end{cases} \tag{11}$$

де $(p_i^{st})_{th}$ – деякі порогові значення частот семантичних полів, вибрані експериментальним шляхом. У загальному випадку, якщо потрібно враховувати різні значення частот, можна використати декілька двійкових розрядів для кодування. Наприклад, якщо збільшити розмір бінарного вектора V_j^{sb} у два рази, тоді можна буде кодувати кожну частоту двома двійковими розрядами, що буде відповідати чотирьом інтервалам значень частот.

4. Базові елементи квантових обчислень

Розглянемо базові квантові принципи реалізації квантових алгоритмів [1, 2]. Основною відмінністю квантового біта – кубіта від класичного біта є те, що кубіт, крім станів $|0\rangle$ і $|1\rangle$, може також знаходитись у суперпозиції цих станів:

$$|\psi\rangle = a|0\rangle + b|1\rangle . \tag{12}$$

Стани $|0\rangle$ і $|1\rangle$ є базисними векторами, які можна представити у матричному вигляді:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{13}$$

Регістр із n кубітів $|x_1, x_2, ... x_n\rangle$ утворює суперпозицію із $N = 2^n$ станів, яку можна записати так:

$$|\psi\rangle = \sum_{i=0}^{N-1} a_i |i\rangle . \tag{14}$$

Ортонормований базис $\{ |0\rangle, |1\rangle, ... |2^n - 1\rangle \}$ називають обчислювальним базисом.

Розглянемо однокубітні квантові вентилі. Розрахунки у квантових алгоритмах здійснюють за допомогою унітарних перетворень, які можна розглядати як повороти комплексного векторного простору. Розглянемо базові операції над кубітами. Оператор тотожного перетворення не змінює значення кубітів і має вигляд

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{15}$$

Операцію заперечення використовують для реалізації інверсії значень кубітів і визначають так:

$$X = |0\rangle\langle 1| + |1\rangle\langle 0| . \tag{16}$$

У спінорному зображенні оператор заперечення має вигляд матриці:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{17}$$

Одним із важливих елементів є «контрольоване НЕ», яке здійснюється над двома кубітами і змінює значення другого кубіта на протилежне, якщо значення першого кубіта рівне 1. Цей логічний елемент може бути визначений як

$$U_{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X , \tag{18}$$

а матриця оператора унітарного перетворення «контрольоване НЕ» має вигляд

$$U_{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{19}$$

Дію вентиля «контрольоване НЕ» можна зобразити так:

$$U_{CNOT} : |a,b\rangle \to |a, a \oplus b\rangle ,\tag{20}$$

де $\oplus$ означає сумування за модулем 2. Ще одним важливим логічним елементом є вентиль Тоффолі, який діє на три кубіти і змінює значення третього кубіта на протилежне, якщо значення першого та другого кубітів рівне 1. Від логічного елемента «контрольоване НЕ» вентиль Тоффолі відрізняється наявністю ще одного додаткового керуючого кубіта. Цей вентиль можна визначити так:

$$T = |0\rangle\langle 0| \otimes I \otimes I + |1\rangle\langle 1| \otimes U_{CNOT} .\tag{21}$$

Перетворення Тоффолі можна зообразити в такому вигляді:

$$T : |a,b,c\rangle \to |a,b,c \oplus ab\rangle .\tag{22}$$

Вентиль Тоффолі є універсальним квантовим логічним елементом, на основі якого можна побудувати оборотну квантову машину Тюрінга. В подальших дослідженнях будемо використовувати вентиль Тоффолі з n керуючими кубітами. В такому вентилі відбувається інверсія керованого кубіта при умові, що n керуючих кубітів мають значення $|1\rangle$.

5. Представлення семантичного вектора текстового об'єкта у квантовій пам'яті

Нехай існує деякий масив текстових об'єктів (1). Множину семантичних полів розглянемо у вигляді (3). Кожне семантичне поле можна закодувати його номером i за допомогою відображення

$$U_s : \ S_i \to i .\tag{23}$$

Для простоти розгляду будемо вважати, що розмір множини масиву текстових об'єктів рівний $N_t = 2^{(nt)}$, а розмір множини семантичних полів рівний $N_w = 2^{(ns)}$. Тоді, для кодування положення об'єкта в масиві, потрібно nt двійкових елементів, а для кодування семантичного образу потрібно ns двійкових елементів. Нехай квантовий регістр складається із двох частин $|qt\rangle$ і $|qs\rangle$:

$$|qt\rangle \otimes |qs\rangle,\ |qt\rangle = |q_1^t, q_1^t, \ldots, q_{nt}^t\rangle,\ |qs\rangle = |q_1^s, q_1^s, \ldots, q_{ns}^s\rangle.\tag{24}$$

Регістр $|qt\rangle$ описує положення текстового об'єкта у масиві, а регістр $|qs\rangle$ описує семантичний образ, який складається із набору семантичних полів. Введемо додатковий кубіт $|f\rangle$, який буде відображати наявність семантичного поля із заданим номером у заданому текстовому об'єкті. Якщо дане поле є присутнє в заданому текстовому об'єкті, тобто, якщо його бінарна частота p_{ij}^{stb} (11) рівна 1, то значення цього кубіта рівне $|1\rangle$, в іншому випадку рівне $|0\rangle$. Тоді весь масив текстових об'єктів можна буде записати у вигляді такої системи кубітів:

$$|qr\rangle = |q_1^t, q_1^t, \ldots, q_{nt}^t\rangle \otimes |q_1^s, q_1^s, \ldots, q_{ns}^s\rangle \otimes |f\rangle .\tag{25}$$

Для запису в квантову пам'ять масиву текстових об'єктів розміром N_t, який містить набір семантичних полів розміром N_s, достатньо $nt + ns + 1 = log_2(2 \cdot N_t \cdot N_s)$ кубітів. Така експоненційна економія квантової пам'яті у порівнянні із класичною пам'яттю можлива внаслідок реалізації квантового паралелізму. Наприклад, якщо набір семантичних полів містить 2^5 елементів, а масив текстових об’єктів – 10^{15} елементів, то для запису такої інформації необхідно лише 5+15+1=21 кубітів.

Запис масиву текстових об’єктів у квантову пам’ять розглянемо феноменологічно за допомогою квантового оракула. В теорії квантових обчислень показано, що на основі однокубітних та двокубітних квантових унітарних вентилів можна побудувати еквівалентні алгоритми класичної машини Тюрінга. Під оракулом будемо розуміти деяке формалізоване унітарне перетворення, за допомогою якого реалізуються наперед задані обчислення. Елементи масиву текстових об’єктів визначаються квантовими станами складеного регістру кубітів (25). Суперпозиція цих станів утворює вектор у комплексному Гільбертовому просторі. Цей вектор є квантовим еквівалентним відображенням текстових об’єктів. Розглянемо послідовність квантового запису масиву текстових об’єктів. Кубіт $|f\rangle$ візьмемо в початковому стані $|0\rangle$, а регістри $|qt\rangle$, $|qs\rangle$ – в початкових станах $|0\rangle^{\otimes(nt)}$ та $|0\rangle^{\otimes(ns)}$ відповідно. Застосуємо однокубітні унітарні перетворення Адамара до регістрів $|qt\rangle\otimes|qs\rangle\otimes|f\rangle$:

$$|\psi\rangle = H^{\otimes(nt)} \otimes H^{\otimes(nw)} \otimes I\left(|qt\rangle\otimes|qs\rangle\otimes|f\rangle\right). \tag{26}$$

В результаті отримаємо

$$|\psi\rangle = \frac{1}{\sqrt{2^{nt+ns}}} \sum_{i=0,j=0}^{N_t,N_s} |i\rangle\otimes|j\rangle\otimes|0\rangle . \tag{27}$$

Суперпозиція $|\psi\rangle$ містить базисні ортогональні стани, кожен з яких відповідає одному запису семантичного вектора текстового об’єкта. У процесі вимірювання відбувається редукція суперпозиції до одного базового стану, який відповідає одному текстовому об’єкту. Отже, та кількість памяті, яка в класичному випадку необхідна для запису семантичного вектора одного текстового об’єкта, в квантовому випадку є достатня для запису цілого масиву текстового об’єкта. Нехай наявність заданого семантичного вектора текстового об’єкта описується функцією $f_q(qt,qs)$, де індекс qt описує текстовий об’єкт, а індекс qs описує спектр семантичних полів текстового об’єкта. У випадку наявності заданого спектра семантичних полів у заданому текстовому об’єкті функція набуває значення 1, інакше 0. Процес запису текста у квантову пам’ять опишемо унітарним перетворенням U_F, яке визначається квантовим оракулом:

$$U_F : |qt\rangle\otimes|qs\rangle\otimes|f\rangle \rightarrow |qt\rangle\otimes|qs\rangle\otimes|f \oplus f_q(qt,qs)\rangle, \tag{28}$$

де $\oplus$ означає сумування за модулем 2. Враховуючи, що кубіт $|f\rangle$ є в початковому стані $|0\rangle$, отримаємо

$$U_F : |qt\rangle\otimes|qs\rangle\otimes|0\rangle \rightarrow |qt\rangle\otimes|qs\rangle\otimes|f_q(qt,qs)\rangle. \tag{29}$$

6. Пошук заданих семантичних образів текстових об’єктів у квантовій базі даних

Завдання полягає в пошуку деякого ключового семантичного образу, який може бути закодований у вигляді квантового стану:

$$|qk\rangle = |q_1^k, q_1^k, \ldots, q_{ns}^k\rangle . \tag{30}$$

Використаємо вентиль Тоффолі для знаходження квантових станів, в яких закодовані ключові семантичні образи. Введемо в систему кубітів (25) додатковий кубіт – анцилу $|z\rangle$, отримаємо

$$|qr\rangle = |qt\rangle \otimes |qs\rangle \otimes |f_q(qt,qs)\rangle \otimes |z\rangle. \tag{31}$$

Подіємо на кубіт $|z\rangle$ у стані $|1\rangle$ оператором Адамара:

$$|z\rangle = H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{32}$$

Допоміжний кубіт $|z\rangle$ буде керуватись за допомогою $ns+1$-мірного елемента Тоффолі T_{ns+1}, в якому керуючими кубітами виступають ns кубітів регістру $|qs\rangle$ та кубіт $|f(qt,qs)\rangle$. Значення кубіта $|z\rangle$ у квантовому стані може змінитись під впливом вентиля Тоффолі у випадку, коли всі кубіти регістру $|qs\rangle$ та кубіт $|f(qt,qs)\rangle$ будуть рівні одиниці. Щоб перевести значення кубітів у значення $|1\rangle$ для квантових станів, які описують ключовий семантичний образ $|qk\rangle$, розглянемо унітарний оператор, який є тензорним добутком однокубітних операторів:

$$S_T^q = I^{\otimes(nt)} \otimes \left(\bigotimes_{i=1}^{ns} S_i^q \right) \otimes I \otimes I, \tag{33}$$

де $S_i^q = \begin{cases} I, \; q_i^k = 1, \\ X, \; q_i^k = 0. \end{cases}$

Оператор S_T^q переводить квантові базисні стани регістру $|qr\rangle$ в стани, в яких значення регістру $|qs\rangle$ рівні 1, якщо співпадають кодування семантичного вектора текстового об’єкта в тексті та ключового семантичного образа $|qs\rangle = |qk\rangle$, тобто

$$S_T|qr\rangle = |qt\rangle \otimes |1,1,...1\rangle_{ns} \otimes |f(qt,qs)\rangle \otimes |z\rangle, \text{ якщо } |qs\rangle = |qk\rangle. \tag{34}$$

Розглянемо оператор

$$U_T = (S_T^q) \cdot (I^{\otimes(nt)} \otimes T_{ns+1}) \cdot (S_T^q). \tag{35}$$

Подіємо цим оператором на систему регістрів кубітів $|qr\rangle$. Перша справа група операторів, виділених дужками, переводить регістр $|qs\rangle$ у значення $|1,1,...1\rangle_{nw}$ для станів, які відповідають шуканим ключовим семантичним образам, друга група операторів реалізує інверсію керованого кубіта $|z\rangle$ для шуканих станів семантичних образів, третя група повертає змінені першою групою стани у стан перед застосуванням оператора U_T. В результаті дії оператора U_T отримаємо

$$\begin{aligned} |\psi\rangle_T &= U_T \left(\frac{1}{\sqrt{2^{nt+ns}}} \sum_{x=1}^{x=2^{nt+ns}} |x\rangle \otimes |f(x)\rangle \otimes |z\rangle \right) = \frac{1}{\sqrt{2^{nt+ns}}} \times \\ &\times \left(\sum_{x \notin X_k} |x\rangle \otimes |f(x)\rangle - \sum_{x \in X_k} |x\rangle \otimes |f(x)\rangle \right) \otimes |z\rangle, \quad |x\rangle = |qt\rangle \otimes |qs\rangle, \end{aligned} \tag{36}$$

де X_k – множина станів суперпозиції, які відповідають кодуванню шуканих ключових семантичних образів. Допоміжний керований кубіт $|z\rangle$ не змінив свого значення під дією оператора U_T і знаходиться у стані (32), в якому він знаходився перед дією оператора U_T, тому його можна винести за дужки та вилучити із подальшого розгляду. Це зумовлено тим, що кубіт $|z\rangle$ був переведений в новий базисний стан (32) за допомогою оператора Адамара. Інверсія стану у цьому базисі є рівнозначна інверсії знаку амплітуди підсистеми квантових станів, в яких закодовані ключові слова. Роль цього кубіта полягала в тому, щоб під дією вентиля Тофолі він змінював свій знак на протилежний у квантових станах, які відповідають умові пошуку. Умова рівності регістрів семантичного вектора текстового об'єкта та регістру ключового семантичного образу враховується в операторі S_T^q.

Наступним кроком алгоритму є переведення додаткового кубіта $|f\rangle$ в початковий базисний стан $|0\rangle$ та вилучення його із подальшого розгляду. Це можна зробити за допомогою унітарного оператора U_F^{-1}, який є оберненим до оператора U_F (29). В результаті отримаємо

$$|\psi\rangle_{F^{-1}} = U_F^{-1}|\psi\rangle_T = \frac{1}{\sqrt{2^{nt+nw}}}\left(\sum_{x\notin X_k}|x\rangle - \sum_{x\in X_k}|x\rangle\right),\ |x\rangle = |qt\rangle\otimes|qs\rangle. \tag{37}$$

Отже, в результаті дії складеного оператора $U_F^{-1}U_T U_F$ отримаємо суперпозицію базисних рівноймовірнісних станів (27), в якій стани, що кодують наявні ключові семантичні образи, мають від'ємну амплітуду.

Для того, щоб при вимірюванні виявити квантові стани, в яких закодовані ключові семантичні образи, необхідно підсилити амплітуди цих станів. Таке підсилення амплітуд станів, які мають задані властивості, можна здійснити за допомогою оператора інверсії відносно середнього, який використовується в алгоритмі Гровера для пошуку в неструктурованій квантовій базі даних [3, 4]. Оператор інверсії розглянемо у вигляді

$$U_G = 2|\psi_c\rangle\langle\psi_c| - I, \tag{38}$$

де

$$|\psi_c\rangle = H^{\otimes n}|0\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}}\sum_{i=0}^{i=2^n-1}|i\rangle. \tag{39}$$

У геометричній інтерпретації оператор U_G здійснює в Гільбертовому просторі дзеркальне відображення деякого вектора відносно осі, яка визначається вектором $|\psi_c\rangle$. Оператор інверсії можна представити сукупністю однокубітних операторів Адамара та операторів інверсії стану кубіта відносно базисного вектора $|0\rangle$:

$$U_G = H^{\otimes n}(2|0\rangle\langle 0| - I)^{\otimes n}H^{\otimes n}. \tag{40}$$

Підсилення амплітуд станів з інверсними знаками амплітуд відбувається внаслідок дії оператора інверсії U_G аналогічно до механізму, описаного в алгоритмі Гровера [3,4]. Враховуючи визначення операторів (28), (35), (38), розглянемо ітерацію алгоритму Гровера у загальному вигляді складеного оператора:

$$U_I = U_G U_F^{-1} U_T U_F. \tag{41}$$

Можна показати, що внаслідок реалізації ітерації U_I можна підсилити амплітуди заданих станів у 3 рази. Якщо шуканий ключовий семантичний образ зустрічається в масиві текстових об'єктів лише один раз, то, аналогічно алгоритму Гровера [3, 4], оптимальна кількість ітерацій U_I буде

$$k_U \approx \frac{\pi}{4}\sqrt{N}, N = 2^{nt+ns}, \tag{42}$$

де N – кількість квантових станів рівна $N = N_t N_s$. Отже, складність алгоритму в даному випадку буде $O(\sqrt{N})$. Якщо кількість шуканих об'єктів із ключовим семантичним образом є рівною l, тоді, згідно із [3, 4], кількість необхідних ітерацій буде рівна

$$k_U \approx \frac{\pi}{4}\sqrt{\frac{N}{l}}. \tag{43}$$

Однак наперед невідомо, скільки текстових об'єктів відповідають ключовому семантичному образу. В такому випадку можна провести серію реалізацій алгоритму Гровера з кількостями ітерацій U_I, які утворюють прогресію

$$k_U = 1,2,4,8,...l_U, \tag{44}$$

де l_U – деяке максимальне значення кількості ітерацій U_I. Тобто спочатку реалізується алгоритм з однією ітерацією, потім з двома і т.д. Якщо в цій серії реалізацій алгоритму при вимірюванні виявлено шукані квантові стани, тоді можна прийняти рішення про наявність шуканого семантичного образу у текстових об'єктах аналізованого масиву, а також оцінити кількість текстових об'єктів із заданим семантичним образом у цьому масиві. Можна показати, що складність алгоритму в такому випадку є також $O(\sqrt{N})$. У класичному алгоритмі складність пошуку семантичних образів у неструктурованому масиві текстових об'єктів буде $O(N)$.

7. Висновки

У роботі розглянута модель квантового представлення семантичних векторів масиву текстових об'єктів, яка дає можливість експоненційно зменшити об'єм необхідної квантової пам'яті у порівнянні із класичним записом. Запропоновано квантовий алгоритм пошуку ключових семантичних образів у масивах текстових об'єктів. Реалізація цього алгоритму здійснюється на основі квантових логічних елементів, зокрема, з використанням вентиля Тоффолі. Ітерація Гровера використовується для підсилення амплітуд квантових станів, які описують семантичні вектори текстових об'єктів. Показано, що реалізація квантових алгоритмів аналізу семантичних образів текстових об'єктів для деякого класу задач дає можливість поліноміально зменшити час виконання алгоритму у порівнянні із класичними алгоритмами внаслідок реалізації квантового паралелізму.

СПИСОК ЛІТЕРАТУРИ

1. Крохмальський Т. Квантові комп'ютери: основи й алгоритми (короткий огляд) / Т. Крохмальський // Журнал фізичних досліджень. – 2004. – Т. 8, № 4. – С. 1 – 15.
2. Китаев А. Классические и квантовые вычисления / Китаев А., Шень А., Вялый М. – М.: МЦНМО, ЧеРо, 1999. – 192 с.
3. Grover L.K. Quantum Mechanics helps in searching for a needle in haystack / L.K. Grover // Phys.Rev. Lett. – 1997. – N 79 (2). – P. 325 – 328.

4. Zalka C. Grover's quantum searching algorithm is optimal / C. Zalka // Phys. Rev. A. – 1999. – N 60 (4). – P. 2746 – 2751.
5. Павлишенко Б.М. Квантовий алгоритм еволюційного аналізу одновимірних кліткових автоматів / Б.М. Павлишенко // Журнал фізичних досліджень. – 2011. – Т. 15, № 3. – С. 1 – 6.
6. Pantel P. From Frequency to Meaning: Vector Space Models of Semantics / P. Pantel, P.D. Turney // Journal of Artificial Intelligence Research. – 2010. – Vol. 37. – P. 141 – 188.
7. Павлишенко Б.М. Сингулярна декомпозиція матриці семантичних ознак в алгоритмі ієрархічної кластеризації текстових масивів / Б.М. Павлишенко // Математичні машини і системи. – 2012. – № 1. – С. 69 – 76.
8. Павлишенко Б.М. Використання концепції семантичного поля у векторній моделі текстових документів / Б.М. Павлишенко // Східно-Європейський журнал передових технологій. – 2011. – № 6/2 (54). – С. 7 – 11.
9. Вердиева З.Н. Семантические поля в современном английском языке / Вердиева З.Н. – М.: Высшая школа, 1986. – 120 с.
10. Левицкий В.В. Экспериментальные методы в семасиологии / В.В. Левицкий, И.А. Стернин. – Воронеж: Изд-во ВГУ, 1989. – 192 с.

Стаття надійшла до редакції 30.05.2012