

УДК 53.082.539+53.082.54

## ПРИМЕНЕНИЕ КВАНТОВЫХ МЕТОДОВ ДЛЯ КОНТЕКСТНОГО ПОИСКА НА ФОТОНАХ С ИСПОЛЬЗОВАНИЕМ ТЕСТОВ БЕЛЛА

© 2021 г. С. Н. Андрианов<sup>1</sup>, Н. С. Андрианова<sup>2</sup>, Ф. М. Аблаев<sup>2</sup>,  
А. А. Калачев<sup>3</sup>, Ю. Ю. Кочнева<sup>1</sup>, А. В. Шкаликов<sup>3</sup>, \*

<sup>1</sup>Государственное научное бюджетное учреждение “Академия наук Республики Татарстан”,  
Институт прикладных исследований, Казань, Россия

<sup>2</sup>Федеральное государственное автономное образовательное учреждение высшего образования  
“Казанский (Приволжский) федеральный университет”, Казань, Россия

<sup>3</sup>Казанский физико-технический институт имени Е.К. Завойского – обособленное структурное подразделение  
Федерального государственного бюджетного учреждения науки “Федеральный исследовательский центр  
“Казанский научный центр Российской академии наук”, Казань, Россия

\*E-mail: andrei\_vs@rambler.ru

Поступила в редакцию 05.07.2020 г.

После доработки 26.07.2021 г.

Принята к публикации 27.08.2021 г.

Проанализированы возможности контекстного поиска с использованием квантовой гиперпространственной языковой модели. Разработан подход к установлению семантической связи между словами в текстах на основе общепринятой базовой формулы для параметра Белла с использованием его как параметра связи. Предложена модификация подхода, позволяющая определить гипонимическую изотопию и, тем самым, находить смысловую связь между понятиями. Обсуждаются перспективы использования квантовых состояний света для реализации данного подхода.

DOI: 10.31857/S0367676521120036

### ВВЕДЕНИЕ

В рамках познавательной деятельности человека информация структурируется с точки зрения ее значения и организации, постепенно превращаясь в знание. Формализуемые на естественном языке знания систематизируются в рамках лексической системы языка в толковых, семантических, идеографических словарях.

В когнитивистике и лингвистическом структурировании связь между объектами (понятиями, лексемами) строится на основе выделения общих признаков (сем). На основе последовательности из двух и более слов, содержащих общий семантический признак, и выявляется изотопия (т.е. семантическая связность) в высказывании или тексте, а также семантическая близость между словами (т.е. сходство слов, находящихся в семантических отношениях) [1–3]. По типу семантического признака выделяют родовые и видовые изотопии, в которых семантическая связь устанавливается между словами, обозначающими частное видовое понятие и являющимися гипонимами, и словами, обозначающими общее родовое понятие и являющимися гиперонимами (например, слова “ложка” и “нож” находятся в

отношении изотопии по признаку “столовый прибор” и являются согипонимами).

В области обработки естественных языков и машинного перевода степень семантической близости между словами вычисляется на основании их распределения (дистрибуции) в больших массивах лингвистических данных (текстовых корпусах). Оценка семантической близости происходит при помощи контекстных векторов, присваиваемых лексическим единицам (word2vec) [4], которые формируют семантическое пространство.

### КВАНТОВАЯ МОДЕЛЬ HAL И КОНТЕКСТНЫЙ ПОИСК

Кроме средства построения логической модели баз данных семантические пространства применяются в автоматизированных системах перевода и при контекстном поиске. В частности, используется подход, представленный языковой моделью HAL (Hyperspace Analogue to Language). В нем сначала вводится лексикон из того или иного набора слов. При построении модели HAL считается, что любое слово находится в контексте, а также учитывается близость слов друг другу в его линейной развертке. Речь идет о построении

базы данных связей слов с учетом их непосредственного словесного окружения [5].

В языковой модели HAL слова распределяются не так, как в словаре обычного языка последовательно, например, просто в алфавитном порядке, а в виде матрицы (таблицы), где слова распределены по вертикали в столбцах по алфавиту, как и в обычном словаре. В строках же каждое слово размещается в парных сочетаниях с другими словами с весом, обратным по удаленности друг от друга в тексте, что должно отражать какую-то связь этого отдельного слова с остальными словами, что делает эту модель более осмысленной, чем обычный словарь. Каждое слово-строка здесь может рассматриваться как вектор, а скалярные произведения этих векторов дают силу ассоциативной связи слов.

Таким образом, эмпирический подход языковой модели HAL позволяет выявить некую взаимосвязь слов. Это связь может быть чисто формальной, являясь следствием случайных совпадений. Но связь может быть и изотопической, то есть иметь смысловой характер. Такую связь между парой слов в том или ином тексте авторы работы [6] предложили искать, используя квантовые алгоритмы при записи слов в тексте при помощи языковой модели HAL. При этом квантовые алгоритмы рассматривались чисто математически с использованием известных в квантовой механике формул. Здесь мы изложим такой подход на более физическом языке, подразумевая использование многомерных фотонных состояний.

В квантовой модели HAL векторное состояние документа определяется как:  $|\Psi\rangle = \sum_i^N |w_{v_i}\rangle$ , где  $|w_{v_i}\rangle = \sum_j^M |w_{v_i}^{(j)}\rangle$  – квантовое состояние слова  $i$ , которое мы будем нумеровать частотой фотона  $v_i$ , разложенное по базовым состояниям  $|w_{v_i}^{(j)}\rangle$  языковой модели HAL, например, по состояниям орбитального углового момента  $l_j$ . Мы можем представить эти состояния как сумму проекций в гильбертовом пространстве нормированного состояния слова в фотоне  $|u_{v_i}\rangle$  на базовые состояния  $l_j$ :  $|u_{v_i}\rangle = \sum_j^M a_{v_i}^{(l_j)} |u_{l_j}\rangle$ .

Двум словам  $A$  и  $B$  из текста можно сопоставить векторные состояния  $|u_{v_A}\rangle$ ,  $|u_{v_B}\rangle$  и общую плоскость в гильбертовом пространстве, проходящую через эти векторы. Тогда векторные состояния документа  $|\psi_A\rangle$  и  $|\psi_B\rangle$  в базисе этих слов можно определить как проекцию состояния  $|\Psi\rangle$  на эту плоскость с последующей проекцией в плоскости слов на сам вектор слова в гильбертовом пространстве и ортогональную к нему ось в плоскости слов. Если сопоставить эти проекции с поля-

ризацией фотона при помощи соответствующего унитарного преобразования, то мы получим следующие состояния:

$$|\psi_A\rangle = \frac{1}{\sqrt{2}}(\alpha_{\sigma_+} |u_{v_A, \sigma_+}\rangle + \alpha_{\sigma_-} |u_{v_A, \sigma_-}\rangle), \quad (1)$$

$$|\psi_B\rangle = \frac{1}{\sqrt{2}}(\beta_{\sigma_+} |u_{v_B, \sigma_+}\rangle + \beta_{\sigma_-} |u_{v_B, \sigma_-}\rangle), \quad (2)$$

где индексы  $\sigma_{+,-}$  нумеруют поляризацию фотонов с частотами  $v_A$  и  $v_B$ , соответствующие словам  $A$  и  $B$ . Коэффициенты в выражениях (1), (2) можно записать как

$$\alpha_{\sigma_+} = \frac{\langle u_{v_A, \sigma_+} | \Psi \rangle}{\sqrt{\langle u_{v_A, \sigma_+} | \Psi \rangle^2 + \langle u_{v_A, \sigma_-} | \Psi \rangle^2}}, \quad (3)$$

$$\alpha_{\sigma_-} = \frac{\langle u_{v_A, \sigma_-} | \Psi \rangle}{\sqrt{\langle u_{v_A, \sigma_+} | \Psi \rangle^2 + \langle u_{v_A, \sigma_-} | \Psi \rangle^2}}, \quad (4)$$

$$\beta_{\sigma_+} = \frac{\langle u_{v_B, \sigma_+} | \Psi \rangle}{\sqrt{\langle u_{v_B, \sigma_+} | \Psi \rangle^2 + \langle u_{v_B, \sigma_-} | \Psi \rangle^2}}, \quad (5)$$

$$\beta_{\sigma_-} = \frac{\langle u_{v_B, \sigma_-} | \Psi \rangle}{\sqrt{\langle u_{v_B, \sigma_+} | \Psi \rangle^2 + \langle u_{v_B, \sigma_-} | \Psi \rangle^2}}. \quad (6)$$

Вид состояний (1), (2) позволяет связать каждое слово с тем или иным квантовым битом (кубитом) информации. Определим теперь операторы запроса контекстного поиска на фотонах с учетом того, что поляризации фотонов можно ассоциировать с его спиновыми состояниями. Оператор прямого значения слова соответствует оператору  $z$ -проекции спина:

$$\begin{aligned} \hat{A}|\psi_A\rangle &= \hat{S}_{Az}|\psi_A\rangle = \\ &= \frac{1}{\sqrt{2}}(\alpha_{\sigma_+} |u_{v_A, \sigma_+}\rangle - \alpha_{\sigma_-} |u_{v_A, \sigma_-}\rangle), \end{aligned} \quad (7)$$

$$\hat{B}|\psi_B\rangle = \hat{S}_{Bz}|\psi_B\rangle = \frac{1}{\sqrt{2}}(\beta_{\sigma_+} |u_{v_B, \sigma_+}\rangle - \beta_{\sigma_-} |u_{v_B, \sigma_-}\rangle). \quad (8)$$

Оператор запроса противоположного значения можно определить как

$$\begin{aligned} \hat{A}_x|\psi_A\rangle &= \hat{S}_{Ax}|\psi_A\rangle = \\ &= \frac{1}{\sqrt{2}}(\alpha_{\sigma_-} |u_{v_A, \sigma_+}\rangle + \alpha_{\sigma_+} |u_{v_A, \sigma_-}\rangle). \end{aligned} \quad (9)$$

Параметр поиска можно записать простой форме как на параметр, введенный Беллом для определения степени перепутанности состояний, белловское неравенство [7, 8] в трактовке работ [9, 10] может быть записано как

$$p(n_1) + p(m_1) \geq p(n_1, m_1) + p(n_2, m_1) + p(n_1, m_2) - p(n_2, m_2), \quad (10)$$

где  $p(n_i)$  и  $p(m_i)$  – вероятности того, что величина  $n_i$  или  $m_i$  равна 1, т.е. произошло то или иное со-

бытие, а  $p(n_i, m_j)$  – вероятность того, что величины  $n_i$  и  $m_j$  равны 1 одновременно, т.е. одновременно произошли два соответствующих события. Вероятность детектирования фотона за некоторый промежуток времени пропорциональна  $\langle \hat{a}^+ \hat{a} \rangle$ , где  $\hat{a}^+$  и  $\hat{a}$  – операторы рождения и уничтожения фотона в исследуемой моде. Вероятность одновременного детектирования фотонов в модах  $a$  и  $b$  пропорциональна величине  $\langle \hat{a}^+ \hat{a} \hat{b}^+ \hat{b} \rangle$ .

Используя нашу классификацию фотонов по частоте и поляризации, можно, с учетом приведенных выше выражений для вероятности детектирования, привести формулу (10) к виду

$$\langle \hat{a}_1^+ \hat{a}_1 \rangle + \langle \hat{b}_1^+ \hat{b}_1 \rangle \geq \langle \hat{a}_1^+ \hat{a}_1 \hat{b}_1^+ \hat{b}_1 \rangle + \langle \hat{a}_2^+ \hat{a}_2 \hat{b}_1^+ \hat{b}_1 \rangle + \langle \hat{a}_1^+ \hat{a}_1 \hat{b}_2^+ \hat{b}_2 \rangle - \langle \hat{a}_2^+ \hat{a}_2 \hat{b}_2^+ \hat{b}_2 \rangle, \quad (11)$$

где индексы нумеруют поляризацию фотонов  $a$  и  $b$ . Можно поляризационным состояниям фотонов сопоставить спиновые состояния и перейти от операторов числа фотонов к операторам их  $z$ -проекции спина по формулам

$$\hat{a}_1^+ \hat{a}_1 = \frac{1}{2}(\hat{A}_1 + 1), \quad (12)$$

$$\hat{b}_1^+ \hat{b}_1 = \frac{1}{2}(B_1 + 1). \quad (13)$$

Подстановка выражений (12) и (13) в неравенство (11) дает

$$\frac{1}{2}(\langle \hat{A}_1 \rangle + 1) + \frac{1}{2}(\langle \hat{B}_1 \rangle + 1) \geq \frac{1}{4}(\langle \hat{A}_1 + 1 \rangle)(\langle \hat{B}_1 + 1 \rangle) + \frac{1}{4}(\langle \hat{A}_2 + 1 \rangle)(\langle \hat{B}_1 + 1 \rangle) + \frac{1}{4}(\langle \hat{A}_1 + 1 \rangle)(\langle \hat{B}_2 + 1 \rangle) - \frac{1}{4}(\langle \hat{A}_2 + 1 \rangle)(\langle \hat{B}_2 + 1 \rangle), \quad (14)$$

откуда получаем

$$2 \geq \langle \hat{A}_1 \hat{B}_1 \rangle + \langle \hat{A}_2 \hat{B}_1 \rangle + \langle \hat{A}_1 \hat{B}_2 \rangle - \langle \hat{A}_2 \hat{B}_2 \rangle. \quad (15)$$

Следовательно, белловский параметр можно записать как

$$S_{query} = 2 \left| \langle \hat{A}_1 \hat{B}_1 \rangle + \langle \hat{A}_2 \hat{B}_1 \rangle + \langle \hat{A}_1 \hat{B}_2 \rangle - \langle \hat{A}_2 \hat{B}_2 \rangle \right|. \quad (16)$$

Наконец, используя волновые функции (4), (5), можно получить

$$S_{query} = 4 \left( \alpha_{\sigma_+}^2 - \alpha_{\sigma_-}^2 \right) \left( \beta_{\sigma_+}^2 - \beta_{\sigma_-}^2 \right). \quad (17)$$

Формула (17) отличается от формулы для параметра поиска работы [6]. Это связано с тем, что в работе [6] проводится семантическое уточнение первого слова по связи в тех или иных текстах со вторым словом, а формула (17) должна использоваться для симметричного смыслового поиска.

Применяя эту формулу, можно вычислить белловский параметр  $S_{query}$ , величина которого определяет степень перепутанности состояний документа по словам А и В, а значит и установить наличие смысловой связи между выбранными словами в этом документе. Такие вычисления можно провести на классическом компьютере с использованием обычных матриц HAL, т.е. классической информации, как это описано в работе [6]. Однако наиболее интересным представляется полностью квантовая реализация языковой модели HAL, когда анализируемому тексту сопоставляется многомерное квантовое состояние некоторой физической системы. При этом наиболее сложным этапом реализации является именно приготовление такого состояния. С этой точки зрения весьма перспективным является использование световых пучков, обладающих орбитальным угловым моментом [11]. Такие пучки можно приготовить в перепутанных состояниях, обладающих большими значениями орбитального момента (вплоть до  $10^3$  [12]). Запись и воспроизведение подобных состояний в устройствах оптической квантовой памяти позволит, в перспективе, осуществлять полностью оптическую обработку информации и вычисление значения параметра Белла.

## ЗАКЛЮЧЕНИЕ

Квантовая языковая модель HAL дает возможность нахождения смысла парных словосочетаний в рамках данного текста или набора текста, т.е. может позволить найти изотопию, т.е. общий семантический признак, связывающий понятия или, наоборот автоматизированным образом определить соответствует ли тот или иной текст смыслу, заданному парой слов.

При поисковом запросе установление связи (выявление изотопии) между понятиями может осуществляться в текстах различного характера (тексты, относящиеся к одной терминологической области; тексты разных тематических областей). Характер текстов, а также характер запроса пользователя (например, определение значения термина-неологизма посредством сравнения его с термином, относящимся к той же терминологии, или поиск двух явно несвязанных друг с другом понятий), влияет на определение изотопии (семантической связи) между этими понятиями: гипонимия, гиперонимия (родовидовые отношения), метафора, антонимия. Таким образом можно установить семантический признак, связывающий понятия. Полученные данные могут быть использованы для создания как толковых словарей, так и специализированных словарей (тезаурусов) в той или иной области в зависимости от характера использованных текстов. Они могут

применяется в системах автоматизированного перевода [13, 14] и системах поиска [15, 16].

В данной работе мы рассмотрели возможности физической реализации математического подхода работы [6] к решению вышеуказанных задач. При этом формулы, полученные нами на основе этого подхода, позволяют построить семантическую связь между словами в тексте. Это позволяет уточнить смысл одного слова по коннотации с другим словом. Таким путем можно строить семантические словари, являющимися плоскими по отношению к обычному линейному орфографическому словарю, так как в них смысл слов, расположенных в столбцах, раскрывается в строках путем коннотации с другими словами текста или набора текстов. Также мы предложили иной подход к установлению семантической связи между словами в текстах на основе общепринятой базовой формулы для параметра Белла с использованием его как параметра поиска. Этот подход позволяет определить гипонимическую изотопию, что позволяет находить смысловую связь между понятиями, которая может служить самостоятельным понятием. Таким путем можно построить семантический словарь, который будет объемлен, так как в дополнение к случаю языковой модели HAL будет содержать еще и смысловую вертикаль на основе значения параметра контекстного поиска по той или иной паре слов.

Исследование выполнено при финансовой поддержке РФФИ (проект № 18-29-20091 – Шкаликов А.В.) и в рамках темы государственного задания № АААА-А18-118030690040-8 (Калачев А.А.).

## СПИСОК ЛИТЕРАТУРЫ

1. Greimas A.J. *Sémantique structurale. Recherche de méthode.* Paris: Larousse, 1966. 262 p.
2. Rastier F. // Documents du Groupe de recherche en sémio-linguistique. V. 3. № 29. Paris, 1981. P. 5.
3. Величковский Б.М. Когнитивная наука: Основы психологии познания: в 2 т. Т. 2. М.: Смысл. Изд. центр “Академия”, 2006. 432 с.
4. <https://www.osp.ru/os/2017/02/13052229>.
5. Lund K., Burgess C. // Behav. Res. Meth. Instrum. Comp. 1996. V. 28. P. 203.
6. Barros J., Toffano Z., Meguebli Y. et al. // arXiv: 1304.6920v3. 2013.
7. Yariv A. // IEEE J. Quantum Electron. 1978. V. QE-14. No. 9. P. 650.
8. Dolev S., Fandina N., Rosen J. // Nat. Comput. 2015. V. 14. P. 433.
9. Clauser J.F., Horne M.A., Shimony A. et al. // Phys. Rev. Lett. 1969. V. 23. No. 15. P. 880.
10. Freedman S.J., Clauser J.F. // Phys. Rev. Lett. 1972. V. 28. P. 938.
11. Forbes A., Nape I. // AVS Quant. Sci. 2019. V. 1. Art. No. 011701.
12. Fickler R., Campbell G., Buchler B. et al. // PNAS. 2016. V. 113. Art. No. 13642.
13. Beltran L., Geriente S. // Found. Sci. 2019. V. 24. P. 227.
14. Бессмертный И.А., Васильев А.В., Королева Ю.А. и др. // Изв. вузов. Приборостр. 2019. Т. 62. No. 8. С. 702.
15. Wang C., Seneff S. // ACM Trans. Speech Lang. Process. 2006. V. 3. No. 2. P. 1.
16. Jia Ye, Weiss R.J., Biadys F. et al. // ArXiv: 1904.06037v2. 2019.

## Quantum approach to contextual query by using photons and Bell tests

S. N. Andrianov<sup>a</sup>, N. S. Andrianova<sup>b</sup>, F. M. Ablayev<sup>b</sup>, A. A. Kalachev<sup>c</sup>,  
Yu. Yu. Kochneva<sup>a</sup>, A. V. Shkalikov<sup>c,\*</sup>

<sup>a</sup> Institute of Applied Studies, Academy of Sciences of the Republic of Tatarstan, Kazan, 420111 Russia

<sup>b</sup> Kazan Federal University, Kazan, 420008 Russia

<sup>c</sup> Zavoisky Physical-Technical Institute, Kazan Scientific Center of the Russian Academy of Sciences, Kazan, 420029 Russia

\*e-mail: andrei\_vs@rambler.ru

Possibilities of contextual query using the Hyperspace Analogue Language (HAL) model are analyzed. An approach to establishing a semantic relationship between words in texts is developed on the basis of the generally accepted basic formula for the Bell parameter that is used as a connection parameter. A modification of the approach is proposed, which makes it possible to determine the hyponymic isotopy and, thereby, to find the semantic connection between the concepts. The prospects for using quantum states of light for the implementation of this approach are discussed.