

УДК 37.02

DOI 10.30914/2072-6783-2025-19-1-24-29

## СЕМАНТИЧЕСКИЕ РАССТОЯНИЯ МЕЖДУ ПОНЯТИЯМИ МИКРОМИРА В ФИЗИЧЕСКИХ И АСТРОНОМИЧЕСКИХ ТЕКСТАХ

*Р. В. Майер*

*Глазовский инженерно-педагогический университет им. В. Г. Короленко, г. Глазов, Российская Федерация*

**Аннотация. Введение.** Статья посвящена проблеме изучения семантического пространства текстов по физике и астрономии с помощью компьютерных технологий. **Цель статьи** заключается в: 1) определении семантических расстояний между ключевыми понятиями микромира в физических и астрономических учебных текстах; 2) построении графов, визуализирующих семантические пространства этих текстов, и их сравнении. **Материалы и методы.** Для анализа использовались тексты по физике и астрономии из школьных и вузовских учебников, а также статьи Википедии (всего более 100 тысяч слов). Обработка текста и вычисление косинусной меры близости понятий осуществлялись с помощью компьютерных программ, написанных в ABCPascal и электронных таблиц Excel; для построения графов использовался онлайн-ресурс Интернета. **Результаты исследования.** Получены два файла с текстами по физике и астрономии, выявлены ключевые понятия физики микромира, и для них определены косинусные меры близости и семантические расстояния. Построены графы, вершины которых соответствуют понятиям, а ребра характеризуют смысловые связи между ними. В результате их сравнения обнаружено, что семантические связи между ключевыми понятиями в физических и астрономических текстах имеют принципиальные различия. Для каждого текста может быть вычислена средняя косинусная близость некоторого понятия (например, «атом») с ключевыми понятиями «электрон», «протон», «нейтрон», «ядро» и т. д.; она является одной из характеристик учебного текста. **Заключение.** Предложенный метод позволяет получить объективные оценки степени близости ключевых понятий в учебных текстах, изучить и сравнить семантические пространства.

**Ключевые слова:** астрономия, микромир, понятие, семантическое пространство, текст, физика, учебник

Автор заявляет об отсутствии конфликта интересов.

**Для цитирования:** Майер Р. В. Семантические расстояния между понятиями микромира в физических и астрономических текстах // Вестник Марийского государственного университета. 2025. Т. 19. № 1. С. 24–29. DOI: <https://doi.org/10.30914/2072-6783-2025-19-1-24-29>

## SEMANTIC DISTANCES BETWEEN MICROCOSM CONCEPTS IN PHYSICAL AND ASTRONOMICAL TEXTS

*R. V. Mayer*

*Glazov State University of Engineering and Pedagogics named after V. G. Korolenko, Glazov, Russian Federation*

**Abstract. Introduction.** The paper is devoted to the problem of studying the semantic space of texts on physics and astronomy using computer technology. **The paper purpose** is to: 1) determining the semantic distances between the key concepts of the microcosm in physical and astronomical educational texts; 2) constructing graphs visualizing the semantic spaces of these texts and comparing them. **Materials and methods.** The analysis used texts on physics and astronomy from school and university textbooks, as well as Wikipedia articles (total volume is 100 thousand words). Text processing and calculation of the cosine measure of concept proximity is carried out using computer programs written in ABCPascal and Excel spreadsheets; an online Internet resource is used to build graphs. **Research results.** Two files with texts on physics and astronomy are obtained, key concepts of microcosm physics are identified, and cosine proximity measures and semantic distances are determined for them. Graphs are constructed; their vertices correspond to concepts, and the edges characterize the semantic connections between them. As a result of their comparison, it was found that the semantic links between key concepts in physical and astronomical texts have fundamental differences. For each text, the average cosine proximity of some concept (for example, “atom”) with the key concepts “electron”, “proton”, “neutron”, “nucleus”, etc. can be calculated; it is one of the educational text characteristics. **Conclusion.** The proposed method allows to obtain objective estimates of the proximity degree of key concepts in educational texts, to study and compare semantic spaces.

**Keywords:** astronomy, microcosm, concept, semantic space, text, physics, textbook

The author declares no conflict of interest.

**For citation:** Mayer R. V. Semantic distances between microcosm concepts in physical and astronomical texts. *Vestnik of the Mari State University*, 2025, vol. 19, no. 1, pp. 24–29. (In Russ.). DOI: <https://doi.org/10.30914/2072-6783-2025-19-1-24-29>

## Введение

Проблема изучения семантического пространства учебного текста (УТ) является актуальной [1; 2], так как определение семантической близости между ключевыми понятиями помогает упростить учебный материал, выявить важнейшие логические блоки, визуализировать связи между понятиями и теориями с помощью таблиц и рисунков, сделать материал более доступным для понимания, а также способствует развитию у обучаемых когнитивных навыков, таких как критическое мышление, анализ и синтез информации. Все это позволяет улучшить методику преподавания, так как помогает сгруппировать похожие идеи и теории, эффективнее использовать ассоциативную память, фокусировать внимание на ключевых понятиях, отсеивая второстепенные, создавать ментальные карты [3], облегчающие понимание и запоминание.

Установление степени семантической связанности понятий П1 и П2 – непростая задача [4], для решения которой используются следующие методы: 1) лексический анализ: анализируют определения П1 и П2, находят отношение числа общих слов к их суммарному количеству, – чем выше результат, тем сильнее связь; 2) структурный анализ: сравнивают структуры определений П1 и П2, если они схожи, то понятия близки; 3) семантический анализ: сравнивают графы онтологических баз знаний, учитывают типы и длины связей с другими понятиями; переходят к векторным представлениям определений, находят косинусную меру близости между ними; используют латентный семантический анализ для выявления скрытых связей в большом корпусе текстов.

Семантический анализ опирается на дистрибутивную гипотезу, согласно которой слова (термины), часто встречающиеся в одних и тех же фрагментах текста, связаны смысловыми связями. Например, если рядом с термином «протон» часто встречается слово «нейтрон», а термины «молекула» или «кристалл» – реже, понятия «протон» и «нейтрон» семантически

ближе друг к другу. Чем сильнее семантическая связанность понятий, тем больше похожи их распределения в тексте, тем чаще эти понятия встречаются в одном абзаце.

**Цель исследования** состоит в: 1) определении семантических расстояний между ключевыми понятиями (концептами) микромира в физических и астрономических учебных текстах; 2) построении графов, визуализирующих семантические пространства этих текстов, и их сравнении.

## Материалы и методы исследования

Для достижения поставленной цели были проанализированы тексты:

– по астрономии: 1) школьный учебник (А. В. Засов, В. Г. Сурдин *Астрономия: 10–11 кл.*, 2019), главы: 5. Солнце и звезды; 6. Строение и эволюция Вселенной; 2) школьный учебник (И. В. Галузо, В. А. Голубев, А. А. Шимбаев *Астрономия: 11-го кл.*, 2009), главы: 6. Солнце – дневная звезда; 7. Звезды; 8. Строение и эволюция Вселенной; 3) учебник (Б. А. Воронцов-Вельяминов, Е. К. Страут *Астрономия 11 кл.*, 2003), главы: 6. Солнце и звезды; 7. Галактики; 8. Эволюция Вселенной; 4) статья из Википедии ([ru.wikipedia.org](http://ru.wikipedia.org)) «Звезда».

– по физике: 1) вузовский учебник (Т. И. Трофимова, 2001), параграфы 202–275 за исключением параграфов 206, 216–222, 229–250, 260; 2) вузовский учебник (С. И. Кузнецов, 2015; С. 114–264); 3) статьи из Википедии «атом», «молекула», «электрон», «протон», «нейтрон», «ион», «кварк».

**Методологической основой** исследования являются работы следующих ученых: С. В. Ракитина [1], А. В. Крюкова [2], С. С. Яковлева [3], С. D. Manning, P. Raghavan и H. Schütze [4], Т. В. Батура [5], С. А. Громцев [6], К. В. Крюков, Л. А. Панкова, В. А. Пронина, В. С. Суховеров и Л. Б. Шипилина [7], В. А. Яцко [8], Р. Солсо [9],



5. С помощью онлайн-ресурса или вручную рисуют граф, визуализирующий семантические связи между понятиями. Для этого находят центральное

понятие, с которым тесно связаны все остальные. Короткие ребра считаются жесткими, а длинные – эластичными и изображаются кривыми линиями.

Таблица / Table

$K(A, B)$  и  $L(A, B)$  между понятиями в физических УТ /  $K(A, B)$  and  $L(A, B)$  between concepts in physical UTs

| Понятие  | N    | вещ-во | частица | молекула | атом  | ион   | фотон  | электрон | протон | нейтрон | нуклон | кварк  | ядро    |
|----------|------|--------|---------|----------|-------|-------|--------|----------|--------|---------|--------|--------|---------|
| вещество | 160  | 1      | 6,85    | 7,21     | 13,04 | 14,48 | 26,69  | 16,36    | 48,48  | 13,75   | 40,48  | ---    | 30,84   |
| частица  | 1064 | 0,127  | 1       | 24,90    | 10,26 | 7,45  | 9,55   | 7,61     | 7,16   | 7,14    | 20,88  | 13,00  | 8,37    |
| молекула | 211  | 0,122  | 0,039   | 1        | 5,81  | 15,31 | 472,93 | 13,81    | 50,26  | 49,74   | 70,38  | ---    | 78,30   |
| атом     | 912  | 0,071  | 0,089   | 0,147    | 1     | 5,24  | 25,31  | 2,00     | 7,55   | 12,33   | 30,84  | 292,26 | 4,86    |
| ион      | 553  | 0,065  | 0,118   | 0,061    | 0,160 | 1     | 18,96  | 6,36     | 21,93  | 37,15   | 15,28  | 7,08   | 31,99   |
| фотон    | 248  | 0,036  | 0,095   | 0,002    | 0,038 | 0,050 | 1      | 7,44     | 87,42  | 96,94   | 117,91 | 331,23 | 119,34  |
| электрон | 1246 | 0,058  | 0,116   | 0,068    | 0,334 | 0,136 | 0,119  | 1        | 8,24   | 13,43   | 79,58  | 368,00 | 5,07    |
| протон   | 340  | 0,020  | 0,123   | 0,020    | 0,117 | 0,044 | 0,011  | 0,108    | 1      | 1,14    | 7,33   | 10,40  | 3,04    |
| нейтрон  | 342  | 0,068  | 0,123   | 0,020    | 0,075 | 0,026 | 0,010  | 0,069    | 0,467  | 1       | 9,99   | 17,90  | 3,25    |
| нуклон   | 103  | 0,024  | 0,046   | 0,014    | 0,031 | 0,061 | 0,008  | 0,012    | 0,120  | 0,091   | 1      | 276,01 | 1,77    |
| кварк    | 435  | 0,000  | 0,071   | 0,000    | 0,003 | 0,124 | 0,003  | 0,003    | 0,088  | 0,053   | 0,004  | 1      | 1959,78 |
| ядро     | 480  | 0,031  | 0,107   | 0,013    | 0,171 | 0,030 | 0,008  | 0,165    | 0,247  | 0,235   | 0,361  | 0,001  | 1       |

В таблице представлены результаты вычисления косинусной меры близости (слева, под диагональю) и семантических расстояний (справа, над диагональю) между некоторыми понятиями в анализируемых текстах по физике. Аналогичная матрица близости и расстояний получена для текстов по астрономии. На рисунках 2 и 3 изображены графы, визуализирующие проекции многомерных семантических пространств этих текстов на некоторую плоскость. Они похожи на ментальные карты [3]. Для них не выполняются

теоремы евклидовой геометрии, одна сторона треугольника может превышать сумму двух других сторон.

Полученные матрицы близости и графы отражают семантические связи между понятиями в анализируемых текстах. Видно, что семантические расстояния отличаются в десятки и сотни раз. Круги большого радиуса соответствуют часто используемым терминам. Сила связи с редко используемыми понятиями определяется с погрешностью, такие связи изображены пунктиром.

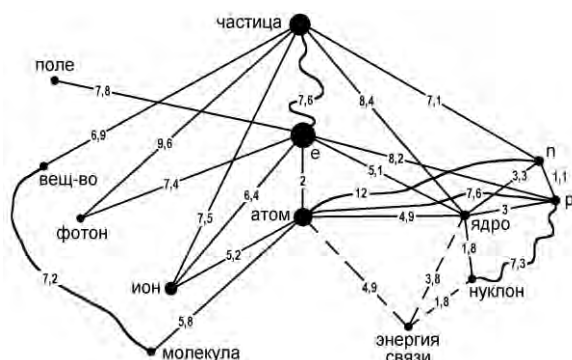


Рис. 2. Семантическое пространство физических текстов / Fig. 2. The semantic space of physical texts

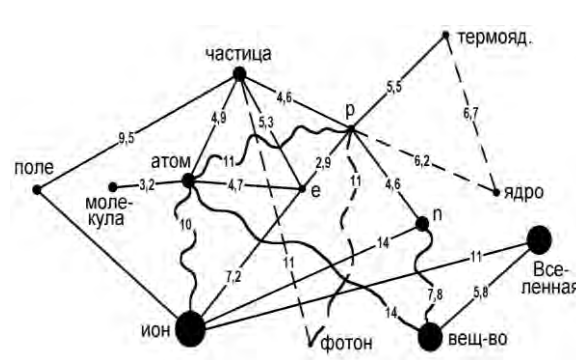


Рис. 3. Семантическое пространство астрономических текстов / Fig. 3. Semantic space of astronomical texts

Сравнивая полученные результаты, можно обнаружить, что в физических текстах к часто используемым терминам относятся: «электрон» (доля употребления среди других терминов 0,19), «частица» (0,17), «атом» (0,14), а в астрономических: «ион» (0,22), «Вселенная» (0,16), «вещество»

(0,15). УТ можно охарактеризовать средним коэффициентом связи некоторого понятия  $\Pi$  к выбранным понятиям-маркерам (например, электрон, протон, нейтрон, ядро):

$$K_{\text{ср}}(\Pi) = (K(\Pi, \text{электрон}) + K(\Pi, \text{протон}) + K(\Pi, \text{нейтрон}) + K(\Pi, \text{ядро})) / 4.$$



Для физических УТ  $K'_{\text{cp}}(\text{Вселенная}) = 0,00253$ ,  $K'_{\text{cp}}(\text{атом}) = 0,174075$ , а для астрономических  $K'_{\text{cp}}(\text{Вселенная}) = 0,017$ ,  $K'_{\text{cp}}(\text{атом}) = 0,095$ .

Ключевые понятия микромира в физических текстах группируются вокруг терминов: «ядро» (средний коэффициент связи с другими понятиями 0,1), «протон» (0,09), «атом» (0,08), «электрон» (0,08), «нейтрон» (0,08), «нуклон» (0,08), а в астрономических текстах – вокруг терминов: «протон» (0,07), «атом» (0,07), «электрон» (0,07), «частица» (0,06), «ион» (0,05), «вещество» (0,05).

В физических текстах «Вселенная» упоминается редко,  $L(\text{Вселенная}, \text{вещество}) = 27$ ,  $L(\text{Вселенная}, \text{частица}) = 39$ ,  $L(\text{Вселенная}, \text{ион}) = 53$ ,  $L(\text{Вселенная}, \text{электрон}) = 98$ ,  $L(\text{Вселенная}, \text{атом}) = 134$ , понятие «Вселенная» встречается  $n_{\text{ф}}(\text{Вселенная}) = 22$  раза. В астрономических текстах обсуждается возникновение и эволюция Вселенной, образование элементарных частиц и т. д., поэтому  $L(\text{Вселенная}, \text{вещество}) = 5,8$ ,  $L(\text{Вселенная}, \text{ион}) = 11$ ,  $L(\text{Вселенная}, \text{атом}) = 30$ ,

$L(\text{Вселенная}, \text{нейтрон}) = 50$ ,  $L(\text{Вселенная}, \text{молекула}) = 92$ ,  $n_{\text{А}}(\text{Вс}) = 224$ .

На рисунке 4 представлены графы, показывающие связи понятия «поле» (гравитационное, электромагнитное и т. д.) с другими близко расположенными к нему понятиями в физических и астрономических УТ. Видно, что множества ближайших понятий и семантические расстояния до них сильно отличаются, что объясняется особенностями учебного материала. В физических текстах концепт «поле» связано с понятием «электрон» и «плазма», а в астрономических – сильно связано с понятиями «ион», «плазма», «частица». Последнее можно записать так: поле  $\propto$  (ион – 8; плазма – 8; частица – 10). Числа показывают расстояния  $L$  до понятия «поле». Другой пример, в физических УТ – атом  $\propto$  (электрон – 2; ядро – 4,9; ион – 5,2; протон – 7,6), а в астрономических УТ – атом  $\propto$  (молекула – 3,2; электрон – 4,7; частица – 4,9; вещество – 9,9; ион – 9,9).



Рис. 4. Связи с понятиями, которые близки к концепту «поле» / Fig. 4. Links to concepts that are close to the concept “field”

### Заключение

В статье рассмотрена проблема изучения семантического пространства текстов по физике и астрономии с помощью компьютерных технологий. В результате анализа школьных и вузовских учебников по физике и астрономии, а также статей Википедии определены семантические расстояния между ключевыми концептами микромира. Построены графы, характеризующие смысловые связи между ними, осуществлено их

сравнение. Показано, что семантические связи между ключевыми понятиями в физических и астрономических текстах имеют принципиальные различия. Для каждого текста может быть вычислена средняя косинусная близость, допустим, понятия «атом» к ключевым понятиям «электрон», «протон», «нейтрон», «ядро» и т. д., которая может рассматриваться как одна из характеристик УТ. Предлагаемый подход открывает новые возможности изучения учебных текстов.

1. Ракитина С. В. Концептосфера и семантическое пространство научного текста // Альманах современной науки и образования. 2009. № 8–1. С. 125–126. URL: <https://elibrary.ru/owdptl> (дата обращения 30.08.2024).

2. Крюкова А. В. Определение семантической близости текстов с использованием инструмента DKPro Similarity // Компьютерная лингвистика и вычислительные онтологии. 2017. № 1. С. 87–97. URL: <https://www.elibrary.ru/ywioyk> (дата обращения 30.08.2024).

3. Яковлева С. С. Использование ментальных карт в обучении студентов вуза // Научное обозрение. Педагогические науки. 2019. № 4–1. С. 134–137. URL: <https://www.elibrary.ru/nedxel> (дата обращения 30.08.2024).

4. Manning C. D., Raghavan P., Schütze H. An Introduction to Information Retrieval. Cambridge University Press, 2008. 569 p.

5. Батура Т. В. Семантический анализ и способы представления смысла текста в компьютерной лингвистике // Программные продукты и системы. 2016. № 4. С. 45–57. URL: <https://www.elibrary.ru/xvigav> (дата обращения 30.08.2024).

6. Громцев С. А. Использование семантического конструктора для формирования учебных заданий // Информатика и образование. 2017. № 7 (286). С. 45–47. URL: <https://www.elibrary.ru/mgmsap> (дата обращения 30.08.2024).
7. Меры семантической близости в онтологии / К. В. Крюков, Л. А. Панкова, В. А. Пронина, В. С. Суховеров, Л. Б. Шипилина // Проблемы управления. 2010. № 5. С. 2–14. URL: <https://www.elibrary.ru/muvnsp> (дата обращения 30.08.2024).
8. Яцко В. А. Эффективность применения косинусной метрики для определения смысловой близости документов // Грани познания. 2020. № 4 (69). С. 3–6. URL: <https://www.elibrary.ru/ftiasg> (дата обращения 30.08.2024).
9. Солсо Р. Когнитивная психология. СПб. : Питер, 2006. 589 с.
10. Майер Р. В. Сложность учебных понятий и текстов : монография. Глазов : ГИПУ, 2024. 132 с. URL: <https://elibrary.ru/item.asp?id=59763871> (дата обращения: 30.08.2024).

*Статья поступила в редакцию 10.01.2025 г.; одобрена после рецензирования 12.02.2025 г.; принята к публикации 10.03.2025 г.*

## Об авторе

### Майер Роберт Валерьевич

доктор педагогических наук, доцент, профессор кафедры физики и дидактики физики, Глазовский инженерно-педагогический университет им. В. Г. Короленко (427621, Российская Федерация, г. Глазов, ул. Первомайская, д. 25), ORCID: <https://orcid.org/0000-0001-8166-9299>, [robert\\_maier@mail.ru](mailto:robert_maier@mail.ru)

*Автор прочитал и одобрил окончательный вариант рукописи.*

1. Rakitina S. V. Kontseptosfera i semanticheskoe prostranstvo nauchnogo teksta [Conceptosphere and semantic space of a scientific text]. *Al'manakh sovremennoi nauki i obrazo-vaniya* = Almanac of modern science and education, 2009, no. 8–1, pp. 125–126. Available at: <https://elibrary.ru/owdptl> (accessed 30.08.2024). (In Russ.).
2. Kryukova A. V. Opredelenie semanticheskoi blizosti tekstov s ispol'zovaniem instrumenta DKPro Similarity [Computing semantic similarity of Russian texts by means of DKPro Similarity tool]. *Komp'yuternaya lingvistika i vychislitel'nye ontologii* = Computational linguistics and computational ontologies, 2017, no. 1, pp. 87–97. Available at: <https://www.elibrary.ru/ywioyk> (accessed 30.08.2024). (In Russ.).
3. Yakovleva S. S. Ispol'zovanie mental'nykh kart v obuchenii studentov vuza [The use of mental maps in teaching students of the university]. *Nauchnoe obozrenie. Pedagogicheskie nauki* = Scientific Review, 2019, no. 4–1, pp. 134–137. Available at: <https://www.elibrary.ru/nedxcl> (accessed 30.08.2024). (In Russ.).
4. Manning C. D., Raghavan P., Schütze H. An Introduction to Information Retrieval. Cambridge University Press, 2008, 569 p. (In Eng.).
5. Batura T. V. Semanticheskii analiz i sposoby predstavleniya smysla teksta v komp'yuternoi lingvistike [Semantic analysis and ways of representing the meaning of text in computational linguistics]. *Programmnye produkty i sistemy* = Software and systems, 2016, no. 4, pp. 45–57. Available at: <https://www.elibrary.ru/xvigav> (accessed 30.08.2024). (In Russ.).
6. Gromtsev S. A. Ispol'zovanie semanticheskogo konstruktora dlya formirovaniya uchebnykh zadaniy [Use of semantic constructor for the formation of training tasks]. *Informatika i obrazovanie* = Informatics and Education, 2017, no. 7 (286), pp. 45–47. Available at: <https://www.elibrary.ru/mgmsap> (accessed 30.08.2024). (In Russ.).
7. Kryukov K. V., Pankova L. A., Pronina V. A., Sukhoverov V. S., Shipilina L. B. Mery semanticheskoi blizosti v ontologii [Measures of semantic proximity in ontology]. *Problemy upravleniya* = Control sciences, 2010, no. 5, pp. 2–14. Available at: <https://www.elibrary.ru/muvnsp> (accessed 30.08.2024). (In Russ.).
8. Yatsko V. A. Effektivnost' primeneniya kosinusnoi metрики dlya opredeleniya smyslovoi blizosti dokumentov [Efficiency of the use of cosine measure to determine the degree of document similarity]. *Grani poznaniya* = Facets of cognition, 2020, no. 4 (69), pp. 3–6. Available at: <https://www.elibrary.ru/ftiasg> (accessed 30.08.2024). (In Russ.).
9. Solso R. Kognitivnaya psikhologiya [Cognitive psychology]. SPb., Piter Publ., 2006, 589 p. (In Russ.).
10. Maier R. V. Slozhnost' uchebnykh ponyatii i tekstov : monografiya [The complexity of educational concepts and texts : monograph]. Glazov, GIPU Publishing House, 2024, 132 p. Available at: <https://elibrary.ru/item.asp?id=59763871> (accessed 30.08.2024). (In Russ.).

*The article was submitted 10.01.2025; approved after reviewing 12.02.2025; accepted for publication 10.03.2025.*

## About the author

### Robert V. Mayer

Dr. Sci. (Pedagogy), Associate Professor, Professor of the Department of Physics and Didactics of Physics, Glazov State University of Engineering and Pedagogics named after V. G. Korolenko (25 Pervomayskaya St., Glazov 427621, Russian Federation), ORCID: <https://orcid.org/0000-0001-8166-9299>, [robert\\_maier@mail.ru](mailto:robert_maier@mail.ru)

*The author has read and approved the final manuscript.*