

**В.Я. ГЕЛЬМАН,**Северо-Западный государственный медицинский университет им. И.И. Мечникова  
(Санкт-Петербург, Российская Федерация; e-mail: [Viktor.Gelman@szgmu.ru](mailto:Viktor.Gelman@szgmu.ru))

## ПРОБЛЕМЫ ФОРМАЛЬНО-МЕХАНИСТИЧЕСКОГО ПОДХОДА К ВЫЯВЛЕНИЮ ПЛАГИАТА В НАУЧНЫХ РАБОТАХ

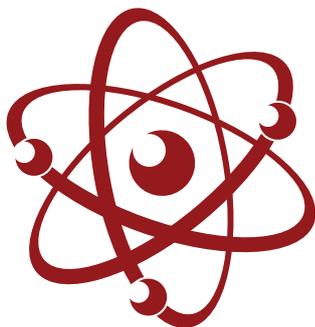
УДК: 001

<https://doi.org/10.22394/2410-132X-2020-6-3-180-185>

**Аннотация:** Оценка результативности научных коллективов и ученых в значительной мере осуществляется по их публикационной активности. Одним из показателей качества научных работ является отсутствие в них плагиата. Целью настоящей статьи являлось рассмотрение проблем, возникающих при формально-механистическом подходе к выявлению плагиата в научных работах. Рассмотрены вопросы соотношения смысла и плагиата; цитирования собственных работ; цитирования и неточного цитирования общепринятых выражений, законов, закономерностей, определений; необходимости доработки систем антиплагиата. Показано, что в связи с тем, что системы антиплагиата не анализируют смысл текста, происходит «гипердиагностика». В результате не критичного подхода к выводам таких систем существенно усложняется написание научных работ. Отмечена необходимость выработки четких критериев для определения того, что является плагиатом в научных работах. Высказаны некоторые предложения по снижению недостатков формального подхода.

**Ключевые слова:** плагиат, научные работы, системы антиплагиата, формальный подход, ценность научной работы, смысловое содержание

**Для цитирования:** Гельман В.Я. Проблемы формально-механистического подхода к выявлению плагиата в научных работах. *Экономика науки*. 2020; 6(3): 180–185. <https://doi.org/10.22394/2410-132X-2020-6-3-180-185>



### ВВЕДЕНИЕ

Оценка результативности научных коллективов и ученых в значительной мере осуществляется по их публикационной активности, количеству выпускаемых статей и диссертационных исследований [1].

Возрастанию эффективности науки и образования в России способствует также повышение качества публикуемых материалов. Одним из показателей качества научных работ является отсутствие в них плагиата [2, 3].

Обычно под плагиатом в научных работах понимают некорректное цитирование чужого текста. Это узкое понимание термина «плагиат». В более широком смысле под плагиатом следует понимать умышленно совершаемое физическим лицом незаконное использование результатов чужого творческого труда, включая идеи, результаты и смысловое содержание, представленное под своим именем [4, 5].

Однако, формально, заимствование идеи, смыслового содержания (т.е. перефразирование существенных частей чужих работ) плагиатом не является, поскольку сами идеи не могут быть объектом авторского права [2, 6].

С начала XXI в. в научном мире начал вызывать беспокойство рост объема некорректного цитирования заимствованного

материала в научных работах (особенно в диссертационных работах) без соответствующих ссылок [7]. Многолетняя успешная деятельность сообщества «Диссернет» привлекла внимание научной общественности к проблеме некорректного цитирования и в нашей стране (отметим, что «Диссернет» не интересуется самими диссертациями, как научными работами [7]).

В то же время, благодаря развитию интернета, выявление плагиата в узком смысле существенно облегчилось. Появились общедоступные системы антиплагиата для выявления заимствований («Антиплагиат», «Advego Plagiat», «eTht Antiplagiat», «Диссерорубка», ИС КОП СГУ, ряд метапоисковых систем и др. [8, 9]), которые находят текстовые совпадения проверяемого документа с учитываемыми (проиндексированными) в системе текстовыми источниками.

Системы антиплагиата выявляют следующие элементы текста [10]:

– Заимствования, т.е. долю всех найденных текстовых пересечений, за исключением тех, которые система отнесла к цитированиям, по отношению к общему объему документа.

– Цитирования – долю текстовых пересечений, которые не являются авторскими, но система посчитала их использование корректным, по отношению к общему объему документа. Сюда относятся оформленные по ГОСТу цитаты; общеупотребительные выражения; фрагменты текста, найденные в источниках из коллекций нормативно-правовой документации.

– Текстовые пересечения, т.е. фрагменты текста проверяемого документа, совпадающие или почти совпадающие с фрагментом текста источника. При этом источником считается документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.

Оценивается степень оригинальности, т.е. доли фрагментов текста проверяемого документа, не обнаруженные ни в одном источнике, по которым шла проверка, по отношению к общему объему документа. В сумме заимствования, цитирования и оригинальность, являясь отдельными показателями, дают 100%, что соответствует всему тексту проверяемого документа».

В результате в последнее время начата активная борьба с плагиатом в научных работах. Это нашло свое отражение в требованиях к отсутствию плагиата и необходимости проведения соответствующей проверки в положении ВАК о защите диссертаций [11] и правилах для авторов всех отечественных научных журналов. В вузах жесткие требования по отсутствию плагиата предъявляют к курсовым [12] и выпускным квалификационным работам [13].

Безусловно, это важное и нужное направление повышения качества научных работ. Однако здесь существуют некоторые проблемы. При механическом применении компьютерных систем антиплагиата всегда будет выявлена определенная доля плагиата, т.к. в любой сфере науки есть устоявшиеся обороты, общепринятые выражения, широко известные законы, закономерности и определения.

Легкость использования и возможность введения формальных критериев сделали выявление плагиата в узком смысле удобным инструментом при оценке качества научной работы [14]. В результате наличие плагиата заняло одно из первых мест среди фактически применяемых характеристик научной работы.

При этом формально-механистический подход и некритичное доверие системам антиплагиата приводит к значительным превышениям, «гипердиагностике» и перегибам при оценке степени плагиата. Хотя даже в отчетах систем по поиску плагиата указывается, что «система является вспомогательным инструментом, определение корректности и правомерности заимствований или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего» [10].

Таким образом, если в прошлом веке существовала «гиподиагностика» плагиата, когда допускались к публикации работы с практически любым объемом плагиата, то теперь осуществляется «гипердиагностика», когда практически любые совпадения частей текста относят к плагиату.

Целью настоящей работы является рассмотрение проблем, возникающих при формально-механистическом подходе к выявлению плагиата в научных работах.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Роль плагиата, в узком смысле совпадения фрагментов текста, отнюдь не самая главная при оценке научного качества работы.

Что является главным в научной работе? Новое знание, новое понимание, новизна, решение проблемы, задачи, достоверность и добротность исходных данных... Обобщая: самое важное – это смысловое содержание работы. И здесь, совпадение фрагментов текста является второстепенным вопросом [15].

### *Соотношение смысла и плагиата.*

Действительно, является ли фраза «Казнить нельзя помиловать» при разных местах запятой плагиатом? Формально-механически, руководствуясь системой антиплагиата, это, безусловно, плагиат: слова полностью повторяются. Однако, по содержанию и смыслу, это совершенно противоположные высказывания, которые не могут быть признаны плагиатом. Таким образом, в достаточно часто встречаемых случаях, когда незначительными изменениями тексту придается совершенно другой смысл, системами антиплагиата этот текст рассматривается как заимствование, что, несомненно, является «ложной тревогой».

Широко известны случаи плагиата, когда в диссертациях заменяется название объекта исследования, а остальной текст остается неизменным (например, контекстной заменой объекта «шоколад» на слово «говядина» [7]). Но если в двух диссертациях, помимо замены объекта исследования, заменены все исходные данные и расчеты, это плагиат? Это, безусловно, новое знание и работа имеет очевидную научную ценность, хотя, конечно, использовались те же методы исследования и способ описания, что и в противопоставляемой работе.

### *Цитирование собственного текста.*

Часто бывают ситуации, когда исследователь, работающий в определенном научном направлении, публикуя очередную статью, для обеспечения целостности ее восприятия, вынужден воспроизводить отдельные части своих предыдущих работ или работ, выполненных в соавторстве. Здесь возникает проблема самоцитирования [16]. Это при современных

формальных правилах проверки – плагиат, но это также результат новых итераций работы, все результаты которой и не должны быть новыми.

Более того, законченная научная работа, как правило, предполагает промежуточные публикации, в которых содержатся ее основные положения, полученные в процессе её выполнения. На практике часто содержание промежуточных публикаций полностью перекрывает основную работу.

Возникают следующие вопросы. Есть ли необходимость в изложении требуемого по смыслу, но ранее описанного материала другими словами [17]? Как в этом случае избежать обвинений в плагиате? Прямое цитирование, обычно, невозможно, т.к. требуются согласования и изменения в соответствии с планом изложения новой статьи. Ссылка на свою работу в списке литературы не отменяет указание системы проверки на то, что это заимствование. Кроме того, в некоторых журналах не допускаются (или, чаще, количественно ограничиваются) ссылки на собственные работы. Поэтому представляется целесообразным выявленные текстовые совпадения с предыдущими собственными (или выполненными в соавторстве) работами менее некоторого объема (например, 30%) не считать плагиатом.

### *Цитирование и неточное цитирование.*

Существующие правила требуют цитирование чужого текста выделять кавычками и давать ссылку на источник. В тех же случаях, когда требовалось сослаться на неточно процитированный материал (по ходу статьи в цитате требовалось внести какие-либо изменения, согласовать ее с остальным текстом, поэтому кавычки не ставились), обычно в конце фразы ставили ссылку на источник. Теперь, как правило, системы выявления плагиата квалифицируют это как «заимствования». Поэтому в настоящее время крайне нежелательно использовать косвенную речь и не выделенное кавычками неточное (частичное) цитирование, что существенно снижает возможности изложения материала.

*Общепринятые выражения, широко известные законы, закономерности и определения.*

В связи с регулярным выявлением этих объектов системами антиплагиата возникают вопросы:

на какие известные утверждения можно не давать ссылок? Теорема Пифагора, коэффициент корреляции...? Надо ли переписывать общепринятые определения, давать свои собственные? Например, часто в студенческих и аспирантских работах идет уже не пересказ учебников и статей своими словами, а бессмысленный перебор различных формулировок одной и той же мысли с проверкой, какая пройдет через систему антиплагиата [17]. В результате сейчас студенческие работы нередко представляют собой необычный набор словосочетаний, потому что нормальные формулировки на грамотном русском языке уже кто-то однажды использовал [18].

*Выявление совпадений отдельных частей предложений.*

Нельзя не отметить частые ложные срабатывания систем антиплагиата по распространенным стилистическим и речевым оборотам. Особенно нежелательно срабатывание системы по стилистическим оборотам, характерным для научной речи. Это связано с тем, что системы антиплагиата не анализируют смысл текста. В связи с этим, возникают вопросы: где нижняя граница плагиата? Какова допустимая длина последовательности совпадающих слов? Представляется целесообразным последовательности менее некоторого количества слов подряд (например, 7–10), совпадающие с не многократно повторяющимся источником, исключать из рассмотрения системами антиплагиата.

*Высокие требования к работам.*

Нет установившихся общих правил оценки того, что конкретно следует считать недопустимым уровнем заимствований. В разных журналах устанавливаются различные допустимые границы объема заимствований (текстовых пересечений) – от 10% до 40%.

Наиболее часто называемый минимально допустимый процент «оригинальности» работы – 80% (20% заимствований и цитирований для системы «Антиплагиат»). По опыту проверок, это очень высокая величина, которая может достигаться, как правило, только искусственным путем. Действительно, научная работа ведется по определенным принципам, полагающим, что в основе конкретной работы лежит идея,

концепция, эксперимент, работа, кем-то когда-то где-то уже высказанная или выполненная, что в обязательном порядке предполагает ее цитату, пересказ или упоминание. Большинство уникальных научных работ, например, по информатике, покажет меньшую величину «оригинальности» [8].

Поэтому имеет смысл допускать определенный уровень механически обнаруженного плагиата (чистых заимствований, определяемых системой антиплагиата), например, в 15–20%. С другой стороны, недопустимо без ссылок цитировать большие фрагменты чужого текста. Здесь тоже, по-видимому, целесообразно ввести ограничение на максимальный размер совпадающего отрывка текста, например, в 300–400 знаков с учетом пробелов.

*Доработка систем антиплагиата.*

Используемый алгоритм применяемой системы антиплагиата или процедура проверки с помощью данной системы должны учитывать вид исследуемого материала, особенности научной отрасли, к которой относится работа, для более адекватного отражения границ правомерного заимствования.

Поэтому необходимо доработать системы антиплагиата, введя дополнительные настройки, позволяющие пользователям самим устанавливать границы, того, что считать плагиатом. Например, если заимствования из собственных работ составляют менее 30%, то это не считать плагиатом. Если менее двух третей фразы единично (не в большом заимствованном фрагменте) совпадает с другим источником – это не плагиат, то же, если менее, например, семи – десяти слов подряд и т.п.

*Возможности апелляции.*

Фактически, автору невозможно оспорить заключение системы антиплагиата. Поэтому использование системы автоматизированной проверки на заимствования должно сопровождаться организацией и разработкой процедур по апелляции и повторной экспертизе уже экспертом или экспертным советом.

Таким образом, необходимы четкие критерии определения, что является плагиатом, а что – нет. Более того, наличие плагиата – это

лишь одна из характеристик работы, причем, не самая важная. Поэтому не стоит преувеличивать его значение. Главное, чтобы в научной работе был новый смысл, а не новые слова. При этом, формально-механистический подход к использованию систем антиплагиата, существенно усложняет без реальной необходимости написание научных работ.

## ЗАКЛЮЧЕНИЕ

В статье рассмотрены проблемы, возникающие при формально-механистическом

подходе к выявлению плагиата в научных работах. Показано, что в связи с тем, что системы антиплагиата не анализируют смысл текста, происходит «гипердиагностика». В результате не критичного подхода к выводам таких систем существенно усложняется написание научных работ. Отмечена необходимость выработки чётких критериев для определения того, что является плагиатом в научных работах. Высказаны некоторые предложения по снижению недостатков формального подхода.

## ЛИТЕРАТУРА

1. Masic I. (2012) Plagiarism in scientific publishing // *Acta Informatica Medica*. 20(4): 208–213. DOI: 10.5455/aim.2012.20.208–213.
2. Yadav S., Rawal G., Baxi M. (2016) Plagiarism – a serious scientific misconduct // *International Journal of Health Sciences*. 6(2):364–366.
3. Masic I. (2014) Plagiarism in scientific research and publications and how to prevent it // *Mater Sociomed*. 26(2):141–146.
4. Habibzadeh F., Shashok K. (2011) Plagiarism in scientific writing: words or ideas // *Croatian Medical Journal*. 52(4):576–577. DOI: 10.3325/cmj.2011.52.576.
5. Витко В.С. (2018) О признаках понятия «плагиат» в авторском праве / М.: «Статут». 142 с.
6. Чернявская В.Е. (2011) Плагиат как социокультурный феномен // *Известия Санкт-Петербургского государственного экономического университета*. 3: 26–31.
7. Ростовцев А.А. (2014) Диссернет. Экспериментальная социология // *Вестник общественного мнения. Данные. Анализ. Дискуссии*. 3–4:127–138. DOI: 10.24411/2070-5107-2014-00005.
8. Чиркин Е.С. (2013) Системы автоматизированной проверки на правомерные заимствования // *Вестник Тамбовского университета*. Серия: Гуманитарные науки. 12(128):164–174.
9. Шинкаренко В.И., Куропятник Е.С. (2017) Проблемы выявления плагиата и анализ инструментального программного обеспечения для их решения // *Наука и прогресс транспорта*. Вестник Днепропетровского национального университета железнодорожного транспорта. 1(67): 131–142. DOI: 10.15802/stp2017/94034.
10. Антиплагиат – первая и лучшая российская система обнаружения текстовых заимствований (2019) / *Antiplagiat*. <https://www.antiplagiat.ru>.
11. Постановление Правительства РФ от 24.09.2013 г. № 842 (2013) О порядке присуждения ученых степеней / Официальный интернет-портал правовой информации. <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102167993>.
12. Begovic E. (2014) My View on Plagiarism // *Acta Informatica Medica*. 22(1):145–146. DOI: 10.5455/aim.2014.22.145–146.
13. Гельман В.Я. (2019) Оценивание выпускных квалификационных работ государственной аттестационной комиссией // *Alma mater*. 2:83–86. DOI: 10.20339/AM.02–19.083.
14. Alzahrani S., Palade V., Salim N., Abraham A. (2012) Using structural information and citation evidence to detect significant plagiarism cases in scientific publications // *Journal of the American Society for Information Science and Technology*. 63(2): 286–312.
15. Bouville M. (2008) Plagiarism: Words and Ideas // *Science and Engineering Ethics*. 14(3):311–322.
16. Broome M.E. (2004) Self-plagiarism: Oxymoron, fair use, or scientific misconduct? // *Nursing Outlook*. 52(6): 273–274.
17. Гореликов С.Ю. (2018) «Антиплагиат» – фильтр или тормоз? // *Наука и школа*. 6: 201–204.
18. Кичерова М.Н., Кыров Д.Н., Смыкова П.Н., Пилипушко С.А. (2013) Плагиат в студенческих работах: анализ сущности проблемы // *Наукосведение*. Интернет-журнал. № 4. 8 с.

## Информация об авторе

Гельман Виктор Яковлевич – доктор технических наук, профессор кафедры медицинской информатики и физики, Северо-Западный государственный медицинский университет им. И.И. Мечникова; ORCID: 0000-0003-3546-3278 (Российская Федерация, 191015, г. Санкт-Петербург, ул. Кирочная, д. 41; e-mail: Viktor.Gelman@szgmu.ru).

**V.YA. GELMAN,**

Medical Informatics and Physics North-West State Medical University named after I.I. Mechnikov (St. Petersburg, Russian Federation; e-mail: Viktor.Gelman@szgmu.ru)

**PROBLEMS OF FORMAL-MECHANISTIC APPROACH TO IDENTIFICATION OF PLAGIARISM IN SCIENTIFIC WORKS**

UDC: 001

<https://doi.org/10.22394/2410-132X-2020-6-3-180-185>

**Abstract:** The effectiveness of research teams and scientists is largely estimated by their publication activity. One indicator of the quality of scientific work is the absence of plagiarism. The purpose of this paper was to consider the problems that arise with the formal-mechanistic approach to the detection of plagiarism in scientific papers. The questions of the correlation of meaning and plagiarism; quoting your own work; quoting and inaccurate quoting; the use of generally accepted expressions, laws, definitions; the need to refine anti-plagiarism systems are considered. It is shown that due to the fact that anti-plagiarism systems do not analyze the meaning of the text, “overdiagnosis” occurs. As a result of an uncritical approach to the conclusions of such systems, the writing of scientific papers becomes more complicated. The necessity of developing more accurate criteria for determining what is plagiarized in scientific papers is noted. Some suggestions were made to reduce the shortcomings of the formal approach.

**Keywords:** plagiarism, scientific work, anti-plagiarism systems, formal approach, the value of scientific work, semantic content

**For citation:** Gelman V.Ya. Problems of Formal-mechanistic Approach to Identification of Plagiarism in Scientific Works. *The Economics of Sciencе*. 2020; 6(3):180–185. (In Russ.) <https://doi.org/10.22394/2410-132X-2020-6-3-180-185>

**REFERENCES**

1. Masic I. (2012) Plagiarism in scientific publishing // *Acta Informatica Medica*. 20(4): 208–213. DOI: 10.5455/aim.2012.20.208–213.
2. Yadav S., Rawal G., Baxi M. (2016) Plagiarism – a serious scientific misconduct // *International Journal of Health Sciences*. 6(2): 364–366.
3. Masic I. (2014) Plagiarism in scientific research and publications and how to prevent it // *Mater Sociomed*. 26(2): 141–146.
4. Habibzadeh F., Shashok K. (2011) Plagiarism in scientific writing: words or ideas // *Croatian Medical Journal*. 52(4): 576–577. DOI: 10.3325/cmj.2011.52.576.
5. Vitko V.S (2018) On the signs of the concept of “plagiarism” in copyright / Moscow: Statute. 142 p. (In Russ.)
6. Chernyavskaya V.E. (2011) Plagiarism as a sociocultural phenomenon // *Bulletin of the St. Petersburg State University of Economics*. 3: 26–31. (In Russ.)
7. Rostovtsev A.A. (2014) Dissernet. Experimental sociology // *Bulletin of public opinion. Data Analysis. Discussions*. 3–4: 127–138. DOI: 10.24411/2070-5107-2014-00005. (In Russ.)
8. Chirkin E.S. (2013) Automated verification systems for illegal borrowing // *Bulletin of the Tambov University. Series: Humanities*. 12(128): 164–174. (In Russ.)
9. Shinkarenko V.I., Kuropyatnik E.S. (2017) Problems of plagiarism detection and analysis of instrumental software for their solution // *Science and Transport Progress. Bulletin of the Dnipropetrovsk National University of Railway Transport*. 1(67): 131–142. DOI: 10.15802/stp2017/94034. (In Russ.)
10. Anti-plagiarism – the first and best Russian system for detecting text borrowings (2019) / *Antiplagiat*. <https://www.antiplagiat.ru>. (In Russ.)
11. Resolution of the Government of the Russian Federation dates 24.09.2013 № 842 (2013) On the procedure for awarding academic degrees / Official Internet portal of legal information. <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102167993>. (In Russ.)
12. Begovic E. (2014) My View on Plagiarism // *Acta Informatica Medica*. 22(1): 145–146. DOI: 10.5455/aim.2014.22.145–146.
13. Gelman V.Ya. (2019) Assessment of graduation qualifications by the state certification commission // *Alma mater*. 2:83–86. DOI: 10.20339/AM.02–19.083. (In Russ.)
14. Alzahrani S., Palade V., Salim N., Abraham A. (2012) Using structural information and citation evidence to detect significant plagiarism cases in scientific publications // *Journal of the American Society for Information Science and Technology*. 63(2): 286–312.
15. Bouville M. (2008) Plagiarism: Words and Ideas // *Science and Engineering Ethics*. 14(3):311–322.
16. Broome M.E. (2004) Self-plagiarism: Oxymoron, fair use, or scientific misconduct? // *Nursing Outlook*. 52(6):273–274.
17. Gorelikov S. Yu. (2018) “Anti-plagiarism” – a filter or a brake? // *Science and school*. 6: 201–204. (In Russ.)
18. Kicherova M.N., Kyrov D.N., Smykova P.N., Pilipushko S.A. (2013) Plagiarism in student work: analysis of the essence of the problem // *Science of Science. Internet magazine*. № 4. 8 p. (In Russ.)

**Author**

*Gelman Viktor Yakovlevich* – Doctor of Technical Sciences, Professor of the Department of Medical Informatics and Physics, North-West State Medical University named after I.I. Mechnikov; ORCID: 0000-0003-3546-3278 (Russian Federation, 191015, Saint-Petersburg, Kirochnaya street, 41; e-mail: Viktor.Gelman@szgmu.ru).