

Гуреев В.Н., Мазов Н.А.
**Институт нефтегазовой геологии
и геофизики СО РАН**
**Государственная публичная
научно-техническая библиотека
СО РАН**

**РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ ПО
ОБНАРУЖЕНИЮ ЗАИМСТВОВАНИЙ
С ИСПОЛЬЗОВАНИЕМ АНАЛИЗА
ЦИТИРОВАНИЙ**

Определение

Переводной плагиат – перевод текста с одного языка на другой без отсылки к первоисточнику и его последующее опубликование. Является наиболее сложным для выявления типом плагиата.

Отличительные особенности

- распространен в научной сфере, где рецензирование строже и простой плагиат быстро выявляется;
- распространен в неанглоязычных странах, где ученые:
 - ✓ переводят англоязычные тексты и публикуют их на своих языках;
 - ✓ переводят собственные работы с родного языка на английский (самоплагиат);
- исследователи отмечают, что росту популярности способствуют успехи машинного перевода.

- **Научная экспертиза** – требует привлечения большого числа ученых, что затратно по времени и оплате; неприменима для массовой обработки подозрительных текстов.
- Выявление плагиата с применением **анализа цитирования**, при котором два текста разных авторов со схожими списками литературы дают основание предполагать наличие плагиата.
- Автоматизированные **системы прямого сличения текстов на разных языках** только начинают появляться.

Обоснование метода

При переводе текста на родной язык авторы сохраняют всю или значительную долю ссылок, а нередко и их последовательность.

Обнаружив в двух текстах идентичные списки процитированной литературы, можно предположить, что более поздняя публикация является плагиатом.

Обоснование метода

Преимущества использования цитирований перед прямым сличением текстов в том, что цитирования не зависят от языков и в то же время являются очень емкими поисковыми образами (Ю. Гарфилд), доказывающими семантическую близость двух текстов.

Метод основан на понятии библиографического сочетания (bibliographic coupling) – работы М. Кесслера начала 1960-х гг.: две работы имеют осмысленное отношение друг к другу и тематически связаны, если у них есть одна и более общих ссылок в пристатейных списках литературы.

Обоснование метода

ScienceDirect

Journals

Books

Vadim Gureyev



Outline



Download

Export

Search ScienceDirect



Advanced



Journal of Virological Methods

Volume 236, October 2016, Pages 231-236

Multiplex method for initial complex testing of antibodies to blood transmitted diseases agents

Alexander G. Poltavchenko, Oleg V. Nechtaylo, Pavel V. Filatov, Anna V. Ersh, Vadim N. Gureyev

Show more

<https://doi.org/10.1016/j.jviromet.2016.09.003>

Recommended articles

Array-in-well platform-based multiple...

Journal of Immunological Methods, Volu...

Download PDF

View details

Empirically optimized flow cytometric...

Analytical Biochemistry, Volume 401, Iss...

Download PDF

View details

Evolving surgical prosthetic rehabilita...

Web of Science

Clarivate Analytics

Поиск

Возврат к результатам поиска

Мои инструменты

История поиска

Список отмеченных публикаций

Параметры полного текста

Найти полный текст



Сохранить в EndNote online

Добавить в список отмеченных публикаций

1 из 1

Multiplex method for initial complex testing of antibodies to blood transmitted diseases agents

Автор: Poltavchenko, AG (Poltavchenko, Alexander G); Nechtaylo, OV (Nechtaylo, Oleg V); Filatov, PV (Filatov, Pavel V); Ersh, AV (Ersh, Anna V); Gureyev, VN (Gureyev, Vadim N)

Показать ResearcherID и ORCID

JOURNAL OF VIROLOGICAL METHODS

Том: 236 Стр.: 231-236

DOI: 10.1016/j.jviromet.2016.09.003

Сеть цитирований

0 цитирований

Просмотр Related Records

Создать оповещение о цитировании

(данные из Web of Science Core Collection)

Метод активно используется для поиска рекомендательной литературы, но может быть применен к поиску плагиата.

Основная проблема заключается не столько в самом сличении цитирований двух текстов, сколько в поиске возможного оригинала, имеющего схожий список литературы с анализируемой публикацией.

Оригинал для переведенного текста может быть найден:

- в полнотекстовой базе данных методом семантического анализа (система тестируется в «Антиплагиате»)



- дороговизна подписки электронных ресурсов разных издательств;
- зависимость от полных текстов;
- невозможность охватить все источники.

- в полнотекстовой базе данных методом анализа цитирования (коллектив под руководством Б. Гиппа)



- в библиографической базе данных методом анализа цитирования (наша разработка)



- доступность библиографических БД во всех российских организациях;
- дешевизна фактологической базы;
- очень широкий охват источников.

Преимущества:

- анализ как набора ссылок, так и их последовательности;
- возможность проведения семантического анализа.

Недостатки:

- отсутствие всеохватной базы данных с полными текстами;
- дороговизна подписки из-за платного доступа к большинству полнотекстовых баз данных.

Преимущества:

- использование библиографических БД дает возможность поиска оригинала в очень широком диапазоне научных текстов;
- возможна автоматизированная генерация запроса в БД на основе цитируемых в подозрительном тексте источников;
- имеющийся инструментарий БД дает широкие возможности поиска по пристатейным спискам литературы.

Недостатки:

- невозможен анализ последовательности ссылок в случае приведения ссылок в оригинале по алфавиту, т.к. нужен доступ к полному тексту;
- невозможность семантического анализа текстов.

Текущие ограничения анализа цитирования:

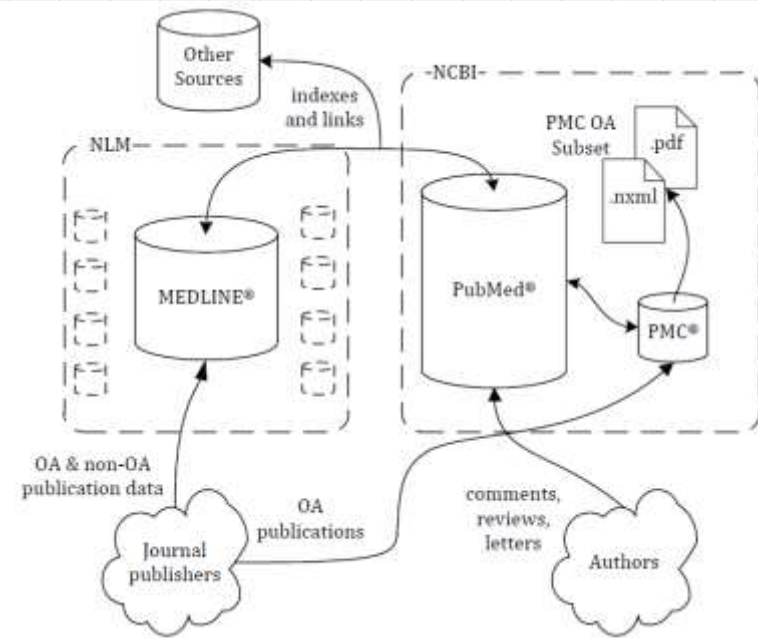
- малое число ссылок;
- цитирование литературы, плохо индексируемой в базах данных (характерно для общественных и гуманитарных наук).

Ограничения, которые уже нашли разрешение:

- перемешивание ссылок;
- масштабирование ссылок (большее или меньшее число использований в сравнении с оригиналом);
- пропуск ссылок;
- добавление собственных ссылок.

Апробация метода в полнотекстовых БД

Масштабная проверка метода с применением анализа цитирования проводилась коллективом Б. Гиппа с использованием полнотекстовой базы данных PubMed Central Open Access Subset, где была проведена полная проверка каждого текста со всей совокупностью других текстов (более 185 тыс., из которых выявлено несколько десятков случаев плагиата).



Прототип программы (www.citeplag.org) на основе анализа цитирования группы Б. Гиппа позволяет сравнивать лишь два имеющихся текста.

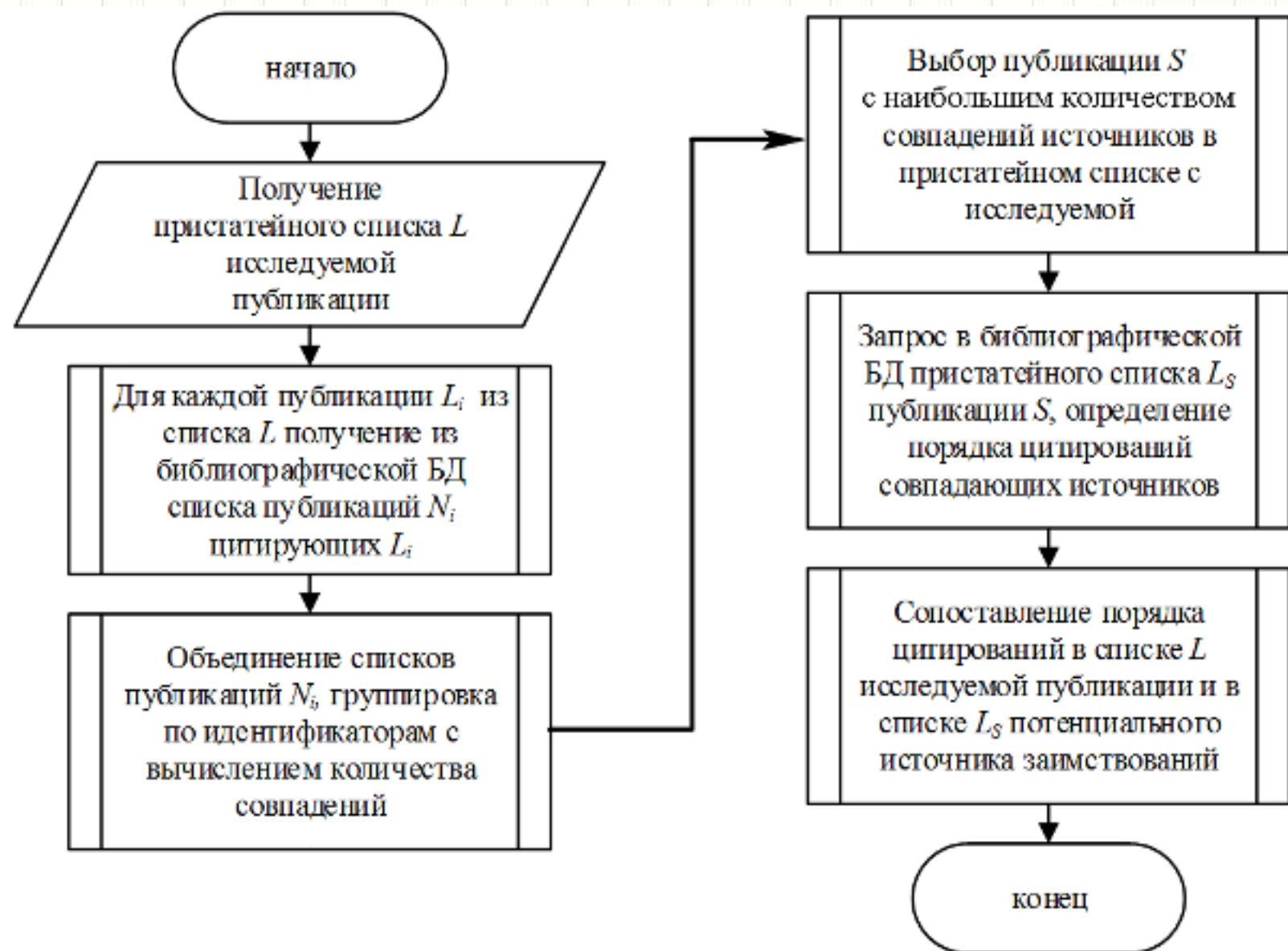
Возможность самого поиска оригинала в настоящий момент не реализована.

В режиме создания и адресации запросов в мультидисциплинарные БД Scopus и Web of Science нашим коллективом было подтверждено несколько случаев плагиата в области информатики и биологических наук.

Алгоритм поиска оригинала

1. для каждого источника в списке литературы подозрительной публикации создавался запрос в библиографическую БД;
2. выявлялась группа публикаций, цитирующих тот же источник;
3. выгружался список идентификаторов публикаций, цитирующих те же источники, что были в списке литературы подозрительной публикации;
4. напротив идентификаторов проставлялись порядковые номера источников из списка литературы подозрительной публикации.

Схема алгоритма



Пример

ID статьи в WoS,
где цитируется
тот же источник,
что и в
подозрительной
публикации

Номер цитируемой статьи
в списке литературы
подозрительной
публикации

000078603800016***58
000078603800055***18
000078603800077***12***18***19
000078614200002***43***44
000078640400003***43

Совпадение двух или
трех источников не
вызывает подозрений

000278518900002***38***45***46***47***48***49***50***51***5
2***53***54***55***56***57***58***59***60***61***62***63

ID искомого оригинала, где
последовательно цитируются
сразу 18 источников из
списка литературы
подозрительной статьи

Русскоязычная статья

Для запуска адаптивного иммунного ответа против ВГ необходимы профессиональные антиген-презентирующие клетки (АПК), такие как дендритные клетки (DC). Уменьшение численности DC в легких в процессе инфекции приводило к снижению уровня ИФН 1-го типа и CD8⁺-клеток [53]. Альвеолярные макрофаги и ДК одними из первых активируются инфекцией ВГ типа А [54]. Во время заражения альвеолярные макрофаги производят большое количество воспалительных цитокинов, главным образом интерлейкина-6 (ИЛ-6) и фактор некроза опухоли-α (ФНО-α). Кроме того, инфицированные эпителиальные клетки легких выделяют в больших дозах моноцитарный хемотрактантный белок 1 (MCP-1), мобилирующий моноциты, которые могут видоизменяться в DC и свободные макрофаги [55]. Во время активации DC созревают и мигрируют в дренирующие лимфатические узлы легких, где они представляют пептидный антиген ВГ в комплексе с молекулами MHC класса I предшественникам CD8⁺-Т-лимфоцитам. Это вызывает пролиферацию и дифференцировку Т-клеток в эффекторные клетки (цитотоксические CD8⁺-Т-лимфоциты).

В процессе эволюции ВГ приобрел механизмы, которые способны подавлять иммунитет хозяина. В частности показано, что белок NS1 ограничивает продукцию ИФН 1-го типа, а также других цитокинов и хемокинов во время заражения вирусом гриппа А [56]. Более того, ВГ способен предотвращать созревание DC. Зараженные DC имеют сниженную экспрессию молекул MHC класса I и таких костимулирующих рецепторов, как CD80 и CD86, что препятствует представлению эндогенно экспрессируемых вирусных антигенов CD8⁺-Т-клеткам [57].

Список литературы русскоязычной статьи

53. McGill J, Herold JM, Legge KL. Innate immune control and regulation of influenza virus infections. J Leukoc Biol. 2009;86:603-12.

54. Nichols JM, Chan MC, Chao WY, Wang HK, Cheung CY, Kwong DL, et al. Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract. Nat Med. 2007;13:147-9.

55. Herold S, von Wuffen W, Steinhilber M, Pleschka S, Kuzel WA, Mack M, et al. Alveolar epithelial cells direct monocyte transendothelial migration upon influenza virus infection: impact of chemokines and adhesion molecules. J Immunol. 2006;177:1017-24.

56. Hays K, Burmanens S, Morat T, Garcia-Sastre A, Fernandez-Sesma A. The NS1 protein of a human influenza virus inhibits type I interferon production and the

induction of antiviral responses in primary human dendritic and respiratory epithelial cells. J Virol. 2009;83:6848-62.

57. Young LJ, Wilson NS, Schreiber P, Mount A, Lundie RJ, La Gruta ML, et al. Dendritic cell precursors impair MHC class II presentation of vaccines and endogenous viral antigens. Proc Natl Acad Sci USA. 2007;104:17753-8.

The quality of the initial innate response against IAVs has profound consequences for the following adaptive response. Triggering the adaptive immune response against IAVs requires professional APC-like DCs. Depletion of lung resident DCs was shown to result in reduced recruitment of pDCs (reviewed in [Nichols et al., 2009a](#)) which are the main producers of type I IFN, and CD8⁺ cells upon IAV infection.

Residential alveolar macrophages and dendritic cells are among the first cells activated by IAV infection. Depending on the virus strain a significant proportion of infected cells are alveolar macrophages ([Nichols et al., 2007](#)). Upon infection with IAVs alveolar macrophages produce significant amounts of inflammatory cytokines like IL-6 and TNF-α and become highly phagocytic. Additionally, infected lung epithelial cells secrete high levels of monocyte chemoattractant protein 1 (MCP-1), recruiting monocytes, which can differentiate into monocyte derived DCs and inflammatory macrophages ([Herold et al., 2006](#)).

Upon activation, DCs undergo maturation and migrate to the draining lymph nodes of the lung where they present peptide antigens to IAV-specific naive T cells. This results in proliferation and differentiation of these T cells into effector cells. As shown for other cell types, NS1 limits the production of type I IFN as well as that of cytokines and chemokines upon IAV infection ([Fernandez-Sesma et al., 2008](#); [Hays et al., 2009](#)). Moreover, IAVs are capable of interfering with the maturation of DCs. Infected DCs show reduced upregulation of MHC class II and costimulatory receptors like CD80 and CD86, essential for T cell priming. Young and colleagues demonstrated that infection of mature DCs by influenza virus interferes with the presentation of an endogenous viral antigen and can thus impact the adaptive immune response ([Young et al., 2007](#)).

Англоязычный оригинал

Наиболее интересный и сложный случай – отчет в один из грантовых фондов

- 202 ссылки в списке литературы;
- 194 источника нашлись в Scopus;
- на эти 194 источника также ссылались 29971 раз из 17228 публикаций.

Распределение ссылок:

- по 1 совпадению – 12711 публикаций;
- 2–5 совпадений – 4440 публикаций;
- 6–10 совпадений – 442 публикации;
- 11–20 совпадений – 108 публикаций;
- 21–30 совпадений – 18 публикаций;
- 31–40 совпадений – 6 публикаций.

Анализ результатов:

отчет составлен из переводов трех статей с числом совпадений ссылок 41, 65 и 82.

Текущее состояние проекта

- Проводятся работы по автоматизации генерации строк запросов в БД и обработки выгруженных из БД в текстовый формат списков литературы.
- Высокая степень автоматизации может быть достигнута взаимодействием с веб-интерфейсами БД с имитацией деятельности пользователя (например, программным интерфейсом Scopus Search API).
- Публичный доступ к сервису поиска потенциальных источников заимствований по приведенному алгоритму с использованием API будет нарушать политику использования программных интерфейсов правообладателя и потребует специального согласования, которое может быть достигнуто крупными системами выявления плагиата.

В 2016 г. наша разработка модели выявления плагиата на основе анализа цитирований получила поддержку РФФИ (проект 16-07-00652). Проект продолжается.

Спасибо за внимание!

**Гуреев Вадим Николаевич,
с.н.с. ИАЦ ИНГГ СО РАН;
с.н.с. ГПНТБ СО РАН**

**Мазов Николай Алексеевич,
зав. ИАЦ ИНГГ СО РАН;
с.н.с. ГПНТБ СО РАН**

**E-mail: GureyevVN@ipgg.sbras.ru
MazovNA@ipgg.sbras.ru**