

# ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ОБЩЕСТВА

УДК 001.811

DOI: 10.33186/1027-3689-2021-2-101-115

**Тициано Пикарди, Роберт Вест**

*Школа компьютерных и коммуникационных наук  
Федеральной политехнической школы Лозанны, Швейцария*

**Мириам Реди**

*Фонд Викимедия – Wikimedia Foundation, Франция*

**Джованни Колавицца**

*Лаборатория цифровых общественных наук  
Университета Амстердама, Нидерланды*

## **Количественные характеристики работы с цитатами в Википедии. (Часть 4)<sup>1</sup>**

**Аннотация:** Википедия – один из самых посещаемых сайтов интернета и самых распространённых источников информации для многих пользователей. В качестве энциклопедии Википедия задумывалась не как источник оригинальной (окончательной) научной информации, а, скорее, как ворота к более глубоким и точным источникам. В соответствии с базовыми принципами Википедии факты должны быть подкреплены надёжными источниками, которые отражают полный спектр всех мнений по рассматриваемой теме. Хотя цитаты лежат в основе функционирования Википедии, пока мало что известно о том, как пользователи работают с ними. Чтобы устранить этот пробел, авторы создали клиентские (пользовательские) инструменты для ведения записей (журналов) всех взаимодействий со ссылками из англоязычных статей Википедии на цитируемые ссылки в течение одного месяца и провели первый анализ взаимодействия читателей с цитатами.

Сопоставление факторов, связанных с переходами по ссылке, показывает, что переходы происходят чаще на страницах более коротких и относительно низкого качества, из чего можно предположить, что ссылки чаще всего требуются, когда Википедия не содержит информацию, которую ищет пользователь. Кроме того, авторы обратили внимание, что источники открытого доступа и ссылки о библиографических данных (рождения, смерти, браки и т.д.) особенно популярны.

---

<sup>1</sup> Части 1–3 см.: Науч. и техн. б-ки. – 2020. – № 9–11.

Собранные воедино, наши выводы углубляют понимание роли Википедии в глобальной информационной экономике, где надёжность становится всё менее определённой, а значение источников – всё более важным.

Исследование, проведённое авторами, представлено ими как доклад на конференции в Тайбэе (Тайвань) в апреле 2020 г. и размещено в системе Archive Корнельского университета (США) под лицензией Creative Commons Attribution 4.0 International (cc-BY 4.0).

Справочный формат АСМ для ссылок:

Тициано Пикарди, Мириам Реди, Джовани Колавицца и Роберт Вест, 2020.

Количественная оценка взаимодействия с цитатами в Википедии. В трудах: Веб-конференция 2020 (WWW'20), 20–24 апр. 2020 г., Тайбэй, Тайвань. АСМ, Нью-Йорк, США. 12 с.; <https://doi.org/10.1145/3366423.3380300>.

**Ключевые слова:** цитирование, гиперссылки, примечания, справки, Википедия, математическая статистика, поведение пользователей.

## 6. RQ3: Анализ цитирования на уровне гиперссылки

Завершающий вопрос нашего исследования: какие особенности конкретной ссылки могут предсказать, заинтересует ли это читателей. Обратите внимание, что такой подход отличается от того, что изучалось в разделе 5<sup>2</sup>, где мы работали на уровне страниц и не различали разные ссылки на одной и той же странице.

### 6.1. Предикторы ссылочных кликов

Начнём с регрессионного анализа, чтобы определить, какие факторы предсказывают, будет ли ссылка задействована. Мы отбрасываем внешние гиперссылки и исключаем множество мешающих факторов, отбирая пары активированных и неактивированных кликов из одного и того же просмотра страницы и таким образом избавляемся от ситуативных моментов – таких, как страница, пользователь, его информационные потребности и т.д. Как мы видели на рис. 5<sup>3</sup>, ссылки на верхнюю и

---

<sup>2</sup> См.: Науч. и техн. б-ки. – 2020. – № 10. – С. 68–75.

<sup>3</sup> Там же. – С. 71.

нижнюю части страниц априори более вероятны для того, чтобы их задействовали. Чтобы устранить помехи от положения ссылки на странице как вносящие путаницу и максимизировать вероятность того, что пользователь сможет видеть обе ссылки в паре, мы выбираем в качестве пары ссылки, наиболее близко расположенные друг к другу на странице, – активированную и неактивированную.

Чтобы убедиться в том, что мы исследуем образцы ссылок, связанных с предложением из основного текста, мы отбрасываем все сноски, расположенные в таблицах, в информационных блоках и изображениях, и оставляем только те, которые находятся в тексте статьи. Наконец, мы выбираем только одну пару за сеанс. После такого отбора мы получили 1,8 млн пар ссылок.

В качестве предикторов мы используем слова в предложении, которое цитирует соответствующая ссылка, а также слова в тексте ссылки (ср. раздел 3)<sup>4</sup>, представленные в виде бинарных показателей, указывающих для каждого из 1 тыс. наиболее часто встречающихся слов, если оно присутствует в предложении. Стоп-слова были удалены, а цифры (кроме четырёх, которые потенциально обозначают годы) были преобразованы в специальные числовые токены. Используя эти функции в качестве предикторов, мы «обучаем» логистическую регрессию предсказывать бинарный индикатор клика.

Мы выполняем этот анализ на представленном выше наборе данных, а также на подмножествах, состоящих только из просмотров страниц каждой из четырёх широких категорий (получены путём объединения 44 *WikiProjects* категорий из раздела 3.4: «Культура» – 1,3 млн пар; «Наука, технология, инженерия и математика – *Science, technology, engineering, and mathematics STEM*» – 436 тыс.; «География» – 530 тыс.; «История и общество» – 467 тыс.). Модель достигает уровня AUC-тестирования около 0,55 при работе на этих пяти множествах.

---

<sup>4</sup> См.: Науч. и техн. б-ки. – 2020. – № 9. – С. 102.

Таблица 1

Наиболее значимые слова – предикторы позитивного и негативного влияния для кликов (см. раздел 6.1) для пяти тематических групп: одна верхняя группа – все тематики вместе, далее вниз – категория «STEM», затем категории «Культура», «История и общество», ниже – «География». Слова сгруппированы по месту их появления: либо в тексте предложения (самой статьи), которое имеет аннотацию в виде сноски, либо же в тексте справки (примечания)

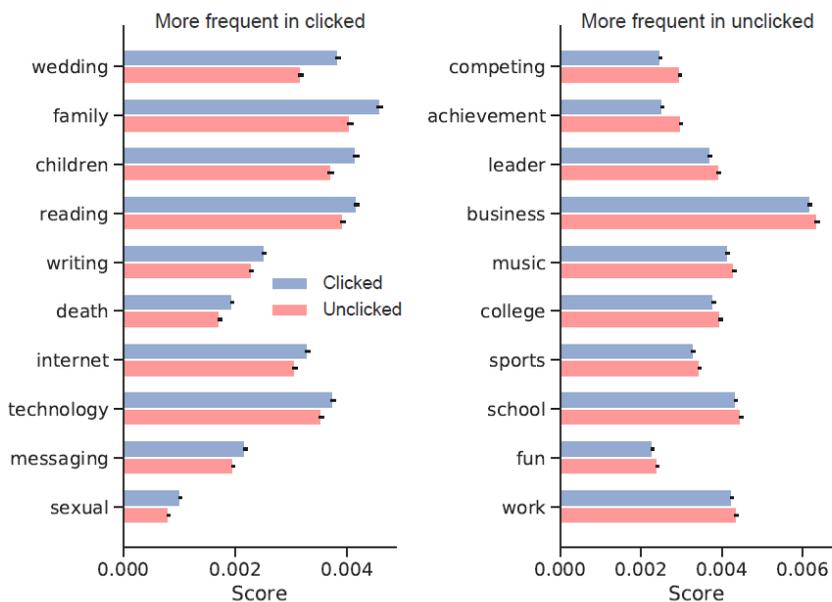
	Positive contribution				Negative contribution			
	In sentence		In reference		In sentence		In reference	
	Word	Coeff.	Word	Coeff.	Word	Coeff.	Word	Coeff.
All topics	greatest	0.36	know	0.25	debut	-0.25	awards	-0.33
	born	0.28	pmc	0.24	moved	-0.16	deadline	-0.32
	died	0.23	2019	0.21	worked	-0.16	billboard	-0.17
	website	0.23	website	0.21	awarded	-0.16	register	-0.17
	ranked	0.23	dies	0.20	joined	-0.13	link	-0.16
	known	0.20	former	0.19	began	-0.13	isbn	-0.15
	professional	0.19	family	0.16	appeared	-0.12	board	-0.14
	relationship	0.19	behind	0.15	score	-0.11	variety	-0.14
	rating	0.18	allmusic	0.15	festival	-0.11	next	-0.14
	article	0.18	story	0.15	attended	-0.11	archive	-0.13
	STEM	online	0.25	definition	0.30	requirements	-0.17	oclc
tests		0.23	2019	0.24	run	-0.24	best	-0.23
2019		0.23	free	0.22	rather	-0.16	jstor	-0.22
short		0.17	pmc	0.21	another	-0.15	evaluation	-0.16
known		0.17	website	0.20	said	-0.15	wiley	-0.16
algorithms		0.16	pdf	0.19	launched	-0.15	london	-0.15
published		0.16	overview	0.17	less	-0.14	isbn	-0.14
defined		0.15	methods	0.15	make	-0.12	internet	-0.14
programming		0.15	introduction	0.14	better	-0.12	industrial	-0.14
digital		0.15	years	0.13	popular	-0.12	source	-0.14
Culture		article	0.30	daughter	0.36	indicating	-0.42	awards
	born	0.28	obituary	0.31	premiered	-0.28	award	-0.33
	greatest	0.27	know	0.31	chart	-0.21	deadline	-0.28
	professional	0.27	instagram	0.29	debut	-0.21	cast	-0.22
	died	0.26	boy	0.28	moved	-0.20	global	-0.21
	known	0.25	sex	0.25	began	-0.17	next	-0.19
	ranked	0.24	wife	0.24	earned	-0.16	isbn	-0.18
	relationship	0.23	former	0.24	recorded	-0.16	drama	-0.18
	website	0.23	historic	0.24	alongside	-0.16	standard	-0.18
	sexual	0.23	2019	0.23	worked	-0.16	tour	-0.18
	History and Society	born	0.29	definition	0.43	came	-0.20	jstor
website		0.21	overview	0.22	award	-0.16	record	-0.21
2019		0.21	best	0.19	transportation	-0.13	link	-0.20
died		0.20	2019	0.19	protection	-0.12	2002	-0.17
currently		0.19	website	0.19	member	-0.12	election	-0.16
known		0.17	statistics	0.17	began	-0.11	1998	-0.15
referred		0.17	death	0.16	originally	-0.11	ed	-0.15
customers		0.16	last	0.16	specific	-0.11	isbn	-0.15
study		0.16	ship	0.15	awarded	-0.10	announces	-0.14
activities		0.15	top	0.15	addition	-0.10	board	-0.12
Geography		politician	0.50	woman	0.34	debut	-0.45	crime
	born	0.26	know	0.27	missing	-0.22	awards	-0.28
	magazine	0.25	dies	0.26	career	-0.21	register	-0.24
	believed	0.23	family	0.23	timothy	-0.20	link	-0.24
	married	0.23	website	0.20	executive	-0.19	interview	-0.19
	ranked	0.22	mail	0.19	episode	-0.17	2000	-0.17
	video	0.22	father	0.18	months	-0.17	culture	-0.17
	directed	0.18	son	0.18	close	-0.15	htm	-0.16
	crime	0.18	boy	0.18	case	-0.15	music	-0.15
	natural	0.18	biography	0.17	appointed	-0.15	paris	-0.15

Слова с наибольшим и наименьшим коэффициентами влияния отображены в табл. 1, где мы наблюдаем эти показатели для всех тем (за исключением категории «*STEM*»); многие положительные черты связаны с общественными и жизненными событиями и отношениями («умирает», «некролог», «женат», «жена», «отношения», «дочь», «семья» и т.п.). Распространённый образец вне указанного ряда – это «2019», а вот ссылки на карьеру («награды», «дебют» и т.д.) менее вероятны для задействия. Эти наблюдения обсудим в разделе 7. На страницах, связанных с категорией «*STEM*», ссылки на материалы открытого доступа получают больше кликов, чем другие, например, слова «бесплатно» и «*PDF*» находятся среди главных предсказателей, тогда как слова, связанные с традиционно закрытым доступом, такие как *JSTOR*, появляются среди отрицательных предикторов, в соответствии с предыдущими выводами [58].

## 6.2. Актуальные корреляты ссылочных кликов

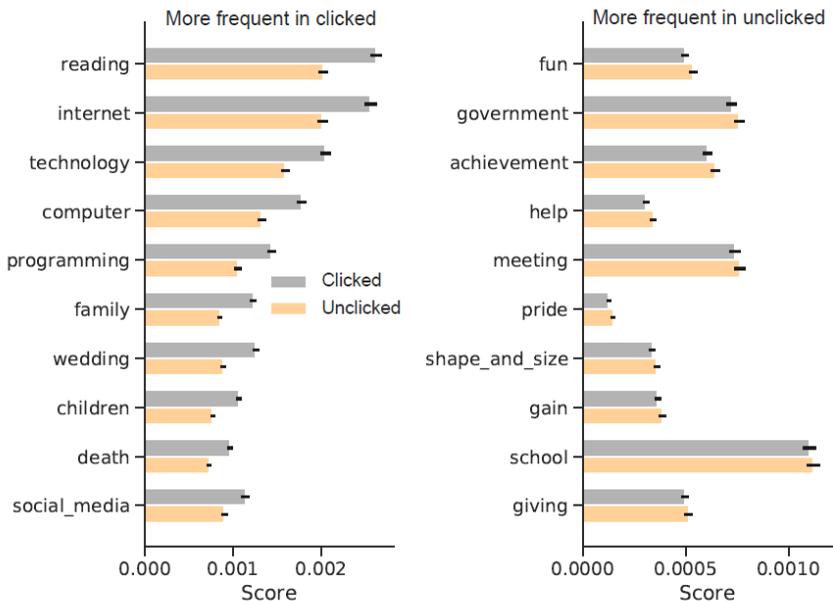
Для более широкого обследования мы проводим тематический анализ цитирования текста статьи и справочных текстов, отдельно для кликнутых и некликнутых ссылок из парного набора данных (см. раздел 6.1). Набор тематик *Empath* мы заимствуем из работы [14]; набор поставляется с предварительно обученной моделью для маркировки входного текста с распределением на более 200 широких тематик.

После применения процедуры используемой модели к каждой точке данных мы вычисляем среднее распределение тематик по кликнутым и некликнутым ссылкам и сортируем темы по разнице между их вероятностью для ссылок.



**Рис. 12а. Роль тематик из набора *Empath* в активации кликов в тексте статьи:** левая колонка – тематика, наиболее часто встречающаяся среди кликнутых предложений; сверху вниз: свадьба, семья, дети, чтение, смерть, интернет, послания, секс; правая колонка – тематика, для которых характерны некликуемые предложения; сверху вниз: конкуренция, достижения, лидер, бизнес, музыка, колледж, спорт, школа, забава, работа

Темы с наибольшим положительным и отрицательным отличием перечислены на рис. 12а и 12б соответственно для цитирования предложений из текста статьи и из ссылки. Результаты подтверждают ранее представленные результаты (раздел 6.1): человеческие факторы (свадьба, семья, секс, смерть) более заметны среди ссылок по кликам, в то время как связанные с карьерой темы (соревнования, конкуренция или достижения) менее востребованы.



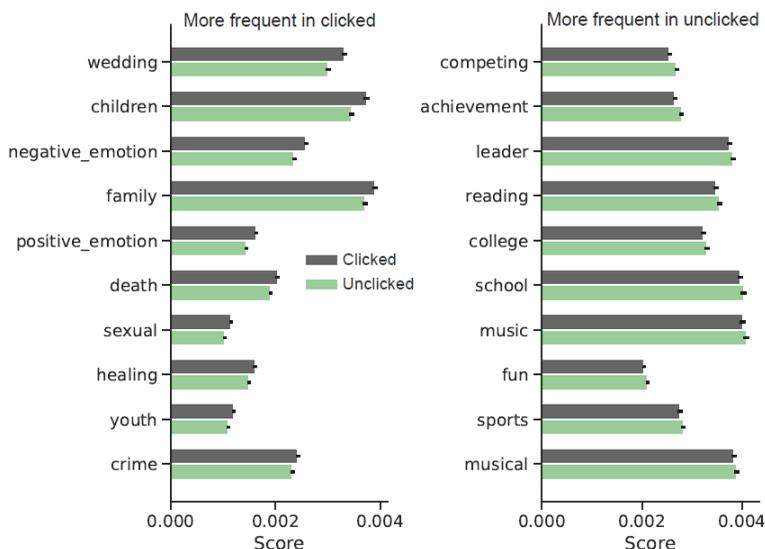
**Рис. 126. Роль тематик набора *Empath* в активации кликов в тексте примечания (справки):**

левая колонка – тематика, наиболее часто встречающаяся среди кликнутых примечаний: чтение, интернет, технология, компьютер, программирование, семья, свадьба, смерть, социальные медиа; правая колонка – тематики, для которых характерны некликнутые примечания: забава, правительство, достижения, помощь, встреча, гордость, размер и форма, выигрыш, школа, давать

Среди наиболее заметных тем для справочных текстов (рис. 126) также появляются связанные темы технологии и интернет.

### 6.3. Предикторы всплывающих сносок

Анализы в разделах 6.1 и 6.2 касались изучения работы со справочными ссылками. Как видно на рис. 4<sup>5</sup>, на настольных устройствах использование всплывающих ссылок-подсказок – это наиболее распространённый способ взаимодействия со ссылками. Поэтому мы повторили приведённый выше анализ с использованием события *fnHover* вместо события *refClick* (8,7 млн эталонных пар), с той разницей, что исключили из анализа слова из ссылочных текстов в качестве функций, так как пользователь вряд ли видел эти слова, прежде чем «завис» над сноской.



**Рис. 12в. Роль тематик набора *Empath* в тексте статьи, наиболее заметных в активации всплывающей подсказки (справки):**

левая колонка – тематика, наиболее часто встречающаяся среди кликнутых примечаний, – сверху вниз: свадьба, дети, негативные эмоции, семья, позитивные эмоции, смерть, секс, лечение, молодость, преступление; правая колонка – тематики, для которых характерны некликнутые примечания, – сверху вниз: конкуренция, достижения, лидер, чтение, колледж, школа, музыка, забава, спорт, музыкальный

<sup>5</sup> См.: Науч. и техн. б-ки. – 2020. – № 10. – С. 69.

Результаты повторяют выводы разделов 6.1 и 6.2, поэтому по причинам экономии места в статье мы не обсуждаем методику регрессионного анализа для всплывающей подсказки (см. раздел 6.1) и вместо этого сфокусируемся на тематическом анализе (см. раздел 6.2). На рис. 12в мы видим более сильную тенденцию событий *fnHover* по сравнению с событиями *refClick*, которые должны вызываться словами, связанными как с положительными, так и с отрицательными эмоциями.

#### **6.4. Предикторы эталонных кликов после зависания**

Когда пользователь наводит курсор на (*fnHover*), текст соответствующей ссылки раскрывается в так называемой справочной подсказке (рис. 1)<sup>6</sup>. В этот момент у пользователя есть выбор: кликнуть по ссылке URL (*refClick*) или остаться на странице статьи. В итоговом анализе этой статьи нам интересно понять, какие слова в тексте ссылки влияют на пользователя при принятии этого решения.

Мы создаём набор данных, отбирая события «загрузка страницы» как минимум с двумя событиями «всплывающая сноска», причём одна сноска реализуется в *refClick* (положительный результат), тогда как другая не реализуется (отрицательный результат). Как и в предыдущих исследованиях, мы выбирали не более одной случайной пары за сеанс, что привело к набору данных 440 тыс. пар событий всплывания.

Как и в исследовании, изложенном в разделе 6.1, мы представляем справочные тексты как 1К-мерные векторы индикаторов слова и используем их в качестве предикторов в логистической регрессии для прогнозирования событий *refClick* (величина тестовой *AUC* 0.54).

---

<sup>6</sup> Там же. – № 9. – С. 99.

**Наиболее значимые слова – предикторы (слева положительные, справа отрицательные) для справочных кликов (*reference click*), следующих за активацией всплывающих примечаний (*footnote hover*; раздел 6.4)**

Positive		Negative	
Word	Coeff.	Word	Coeff.
killer	0.16	oclc	-0.22
greatest	0.16	jason	-0.16
critic	0.15	episode	-0.15
things	0.15	die	-0.15
daughter	0.15	dictionary	-0.13
reveals	0.14	spanish	-0.12
baby	0.14	isbn	-0.12
instagram	0.13	le	-0.11
wife	0.13	board	-0.11
sheet	0.13	channel	-0.11

Самые высокие коэффициенты влияния приведены в табл. 2; картина согласуется с выводами предыдущих анализов: читатели, увидев предварительный показ ссылки во всплывающей подсказке, с большей вероятностью нажмут на цитируемую ссылку, когда в её тексте упоминаются социальные и жизненные аспекты («жена», «малыш», «инстаграм» и т.д.).

Отрицательные коэффициенты указывают, что читатели склонны не переходить к словарным статьям, книжным каталогам (*ISBN*, *OCLC*) и к информации на других языках, кроме английского: ручная проверка показала, что «испанский» в основном связан с примечанием «на испанском», «*le*» – французская статья, распространённая во французских названиях газет (например, «*Le Monde*»), а «*die*» – немецкая статья.

## 7. ОБСУЖДЕНИЕ И ВЫВОДЫ

Наше исследование даёт важную информацию о роли Википедии в качестве шлюза к информации в интернете. Мы обнаружили, что в большинстве случаев Википедия является конечным пунктом поиска информации: менее 1 из 300 просмотров страниц приводит к необходимости использовать цитирование. В нашем анализе мы сосредоточились на той части пользователей, которые обращаются к примечаниям и ссылкам, и

это служит одним из примеров того, каким образом Википедия используется в качестве шлюза к внешним знаниям. Представим наши выводы.

***Пользователь начинает работать со ссылками в Википедии, когда текст статьи не удовлетворяет его информационную потребность.***

В разделе 5<sup>7</sup> показано, что читатели с большей вероятностью активируют ссылки в более коротких и менее качественных статьях. Такой результат поначалу казался нелогичным, поскольку более качественные статьи содержат больше ссылок, которые потенциально могут быть использованы. Это соответствует ранее сделанному выводу: ссылки на источники, которые сообщают отдельные и разрозненные факты, обычно доступные в статьях Википедии (например, награды, карьера), также, как правило, менее привлекательны (раздел 6<sup>8</sup>). В совокупности эти результаты предполагают, что читатели склонны искать контент за пределами Википедии только в тех случаях, когда сама энциклопедия не удовлетворяет их информационные потребности<sup>9</sup>.

***Цитаты в менее привлекательных статьях более привлекательны.*** В разделе 5 мы отметили, что количество кликов на ссылки уменьшается с ростом популярности статьи. Впрочем, вывод можно было сделать и из предыдущего пункта, потому что длинные, высококачественные статьи, как правило, более популярны. Это также может означать, что менее популярные статьи посещаются с какой-то конкретной целью для поиска нужной информации. Ранее в предыдущих работах было показано, что популярные статьи будут с большей вероятностью просматриваться пользователями, которые, скорее всего, случайно зашли в Википедию [53].

***Мы работаем с контентом о жизни людей.*** Мы ясно увидели, что читательский интерес особенно высок к ссылкам на людей, их общественную и личную жизнь (раздел 6). Это особенно верно для всплывающих примечаний. Интерес к всплывающей ссылке чаще вызывается словами, которые связаны с эмоциями – как положительными, так и отрицательными.

---

<sup>7</sup> Там же. – № 10. – С. 68–75.

<sup>8</sup> Там же. – № 11. – С. 148–156.

<sup>9</sup> Там же. – № 10. – С. 68–75.

**Недавний (свежий) контент более интересен.** Мы обнаружили, что ссылки, отражающие недавние события (например, текст включает «2019»), более привлекательны, как для вызова всплывающей ссылки, так и для готовности напрямую кликнуть.

**Контент открытого доступа более привлекателен.** Наконец, мы увидели, что ссылки на страницах Википедии о науке и технике, особенно если они указывают на источники с открытым доступом (например, имеющие пометки «бесплатный» или «pdf» в ссылочном тексте), также, скорее всего, будут активированы (нажаты).

**Теоретические последствия.** Наши результаты дают новое понимание о пользователях сети и их информационных потребностях через призму работы с самой большой онлайн-энциклопедией. Впервые, характеризуя цитирование Википедии, мы можем количественно оценить значение Википедии как шлюза в более широкую сеть. Полученные нами выводы позволяют исследователям развивать новые теории об информационных потребностях читателей и возможных барьерах, разделяющих знания внутри и вне энциклопедии.

Наше исследование также может помочь сообществу веб-авторов и иных создателей материала в определении приоритетов усилий для повышения достоверности информации: мы обнаружили, что люди особенно легко полагаются на цитируемые источники при поиске информации о последних событиях или о биографических данных, а из этого следует необходимость особенно тщательно проверять подобные данные в Википедии. Наконец, тот факт, что читатели больше взаимодействуют со свободно доступными источниками, подчёркивает важность открытого доступа и открытых научных инициатив.

**Практические последствия.** Количественная оценка полноты статьи в Википедии оказалась нетривиальной задачей [45]. Важно понимание того, что категория полноты статьи тесно связана с деталями взаимодействия читателей с Википедией; работа со ссылками открывает идеи для создания новых приложений, чтобы помочь удовлетворить информационные потребности пользователей сети, включая модели, которые количественно определяют нехватку информации в статье путём изучения и контроля сигналов, связанных с темпом перехода по ссылкам. Наши выводы также помогут определить приоритетные области работы с содержанием, которая

будет проведена редакторами Википедии в отношении качества цитирования: в тех областях контента, где Википедия выступает в качестве основных ворот к знанию, качество и надёжность источников, посещаемых читателями, становится ещё более важным.

Наконец, данные, которые мы собрали, обуславливают возможность создания следующей модели: если в предложении отсутствует необходимая цитата (т.е. поставлен тег «требуется цитата»), то можно заранее количественно определить, насколько вероятно, что читатели будут заинтересованы в доступе к соответствующей информации и тем самым помочь редакторам Википедии выделить приоритеты в нерассмотренных делах о недостающих ссылках.

**Ограничения и будущая работа.** Общий низкий *AUC* (от 0,54 до 0,6) регрессионных моделей (разделы 5–6<sup>10</sup>) подчёркивает врождённую непредсказуемость поведения читателя.

Способность к предсказанию у опробованных моделей значительно выше случайной, что делает их полезными для анализа влияния различных предикторов, однако производительность этих моделей в настоящее время слишком низка, чтобы считать их полезными в качестве практических инструментов прогнозирования. Следовательно, будущая работа должна быть нацелена на создание более мощных последовательных моделей, которые могли бы повысить точность.

Сосредоточив внимание только на английской части Википедии, наше исследование сужает представление о взгляде на проект этой энциклопедии, которая доступна почти на 300 языках для пользователей по всему миру. Поэтому в будущем мы планируем произвести такое же изучение других языковых вариантов Википедии. До сих пор мы не учитывали персональные характеристики пользователя, такие как более глобальные поведенческие черты за пределами уровня просмотра страниц, а также географическую информацию, которая играет важную роль в поведении пользователей [32, 57]. Будущая работа должна включать изучение таких факторов.

---

<sup>10</sup> Там же. – № 10. – С. 68–75; № 11. – С. 148–156.

Мы также будем более внимательно исследовать намерения читателя. До сих пор только записи о кликах и всплывающих примечаниях дают информацию об интересах пользователей, они не могут рассказать нам о конкретных обстоятельствах, которые привели пользователя к применению кликов или всплывающих примечаний, а также об уровне удовлетворённости, достигнутом после прочтения ссылки. В будущем мы планируем лучше понять эти аспекты с помощью качественных методов, таких как опросы и интервью. Далее наш анализ будет сосредоточен на ссылках в разделе «Примечания»; нужно также изучить другие виды внешних ссылок для удовлетворения информационных потребностей читателей.

Наконец, поскольку внешние события сильно влияют на информационные потребности пользователей Википедии [53], будущая работа должна выходить за рамки понимания и изучения Википедии как изолированной платформы, и анализ цитирования следует проводить с учётом воздействия последних новостей и событий с не определённой заранее информацией. Это расширит и уточнит нашу картину Википедии как ворот к глобальной информации.

**Благодарность:** *Leila Zia, Michele Catasta, Dario Taraborelli* за их содействие; аналитику *Bahodir Mansurov*, за помощь в организации работы *event logging*; *James Evans* за интересные обсуждения; *Microsoft, Google, Facebook, SNSF* за поддержку.

Список литературы (70 позиций) представлен по адресу: <https://doi.org/10.1145/3366423.3380300>.

**Перевод А. И. Земскова,  
ГПНТБ России**

## Информация об авторах

**Тициано Пикарди** – Школа компьютерных и коммуникационных наук Федеральной политехнической школы Лозанны, Швейцария  
tiziano.piccardi@epfl.ch

**Роберт Вест** – доцент лаборатории научных данных Школы компьютерных и коммуникационных наук Федеральной политехнической школы Лозанны, Швейцария  
robert.west@epfl.ch

**Мириам Реди** – исследователь в научной группе Фонда Викимедия, Франция  
miriam@wikimedia.org

**Джованни Колавица** – доцент Лаборатории цифровых общественных наук Университета Амстердама, Нидерланды  
g.colavizza@uva.nl

