

УДК 81'373.47

**ВИКИПЕДИЯ КАК ИНСТРУМЕНТ
ЛЕКСИКОГРАФИЧЕСКИХ ИССЛЕДОВАНИЙ
(НА МАТЕРИАЛЕ РУССКОЯЗЫЧНОГО КОРПУСА СТАТЕЙ)**

¹Демченков С.А., ²Федяева Н.Д.

¹ФГБОУ ВПО «Омский государственный университет им. Ф.М. Достоевского»,
Омск, e-mail: DemchenkovSA@omsu.ru;

²ФГБОУ ВПО «Омский государственный педагогический университет»,
Омск, e-mail: ndfed@yandex.ru

В статье анализируются возможности Википедии как аналитического инструмента при проведении лексикографических исследований. В настоящее время «народная» лексикография активно вступает в конкурентные отношения с научной, что обуславливает необходимость её изучения. Благодаря большим объёмам метаинформации, накапливаемой в «свободной энциклопедии» применительно к каждому документу, появляется возможность отслеживать историю правок произвольно взятой статьи с момента её создания, изучать субъективную мотивацию осуществлённых редакторами изменений, детально отслеживать лексикографическую активность пользователей. Заложенные в Википедию аналитические возможности позволяют существенно повысить эффективность исследований в области «народной» лексикографии. На примере конкретной энциклопедической статьи («Селфи») показано, каким образом можно использовать метаданные Википедии в работе филолога. Сделан вывод о том, что статьи Википедии эволюционируют в результате конфликта разговорно-бытового и научного дискурсов.

Ключевые слова: Википедия, народная лексикография, лексикографическое исследование, энциклопедия, словарная статья, селфи

**WIKIPEDIA AS A TOOL OF LEXICOGRAPHIC RESEARCH
(ON THE BASIS OF RUSSIAN ARTICLES CORPUS STUDIES)**

¹Demchenkov S.A., ²Fedyayeva N.D.

¹Omsk F.M. Dostoevsky State University, Omsk, e-mail: DemchenkovSA@omsu.ru;

²Omsk State Pedagogical University, Omsk, e-mail: ndfed@yandex.ru

The article analyzes the possibilities of Wikipedia as an analytical tool in conducting lexicographical research. Currently folk lexicography actively competes with scientific lexicography, necessitating its study. Due to the large amounts of meta-information accumulated in the «free encyclopedia» with respect to each document, it is possible to trace the editing history of any article from the moment of its creation, to study subjective motivation of editors' alterations, to monitor thoroughly lexicographic activity of encyclopedia users. Analytical capabilities of Wikipedia make it possible to increase significantly the effectiveness of research in the field of folk lexicography. On the example of a particular encyclopedia article («Selfie») we show how you can use the Wikipedia metadata in philology. Draw the conclusion that Wikipedia articles evolve as a result of the conflict spoken and scientific discourse.

Keywords: Wikipedia, folk lexicography, lexicographic research, encyclopedia, lexical entry, selfie

Википедия (Wikipedia) – некоммерческий проект Ларри Сэнгера и Джимми Уэйлса, основанный в 2001 году и представляющий собой публичную интернет-энциклопедию, создаваемую и редактируемую на добровольных началах самими пользователями Всемирной Сети. В настоящее время функционируют 276 разделов Википедии на национальных языках, 12 из них содержат свыше миллиона статей, 39 – свыше ста тысяч [2].

В состав Википедии входят несколько самостоятельных проектов, в частности мультифункциональный Викисловарь, в настоящее время включающий в себя 3 850 899 статей на 1510 языках и диалектах (в том числе искусственных), из них 560 321 в англоязычной и 542 640 в русскоязычной версии [5, 8]. Статья Викисловаря

содержит информацию о грамматических характеристиках слова, его произношении, основных значениях, синонимах, антонимах, гиперонимах, гипонимах, этимологии, фразеологизмах, в состав которых оно входит; для каждого из значений приводятся примеры словоупотребления.

В 2003 году англоязычный, а в 2006 году и русскоязычный разделы Википедии обогнали знаменитую Британскую Энциклопедию по количеству статей [1]. Сегодня Википедия – крупнейший и наиболее доступный из энциклопедических источников. Её главное достоинство, позволившее «свободной энциклопедии» в кратчайшие сроки достигнуть фантастических, по меркам традиционных изданий, масштабов информационного охвата, одновременно является и её «ахиллесовой пятой»: авторами

и редакторами подавляющего большинства статей становятся энтузиасты-непрофессионалы, что отражается на качестве материалов: содержащиеся в них сведения нередко оказываются неполными, неточными, а порой и недостоверными; многократно зафиксированы случаи намеренного искажения текста статей, представляющие собой акты вандализма, информационной агрессии или PR-акции.

Впрочем, пресса и научное сообщество склонны преувеличивать отрицательные эффекты «волонтёрского» подхода в лексикографии. В 2005 году в журнале «Nature» были опубликованы результаты сравнительного исследования 42 статей из Википедии и Британской энциклопедии, показавшего, что процент ошибок (как грубых, предполагающих существенные концептуальные искажения, так и незначительных) в этих источниках примерно одинаков [6]. Хотя акты вандализма происходят регулярно, среднее время «жизни» деструктивных модификаций (до восстановления исходного текста) не превышает нескольких дней, а в статьях, посвящённых популярной тематике, – нескольких часов.

Социологические опросы показывают, что «кредит доверия» Википедии у населения развитых стран сравнительно высок. По данным ВЦИОМ, «информации в печатных справочниках респонденты доверяют меньше, чем той, что размещена в интернет-энциклопедиях (54% против 57%)», причем сведения, приведённые в Википедии воспринимаются как более достоверные в сравнении с теми, что размещены на официальных сайтах различных организаций [3].

Крупнейшие мировые агрегаторы информации сегодня позиционируют Википедию как авторитетный и заслуживающий доверия источник. Как показало исследование, проведённое в 2012 году компанией «Intelligent Positioning» [7], при вводе в поисковую строку Google 1000 произвольно отобранных слов, в 99% случаев ссылка на соответствующую статью Википедии выдаётся в числе первых 10 результатов поиска, причём в 56% случаев она идёт первой строкой в поисковой выдаче, и только в 24 и 9% случаев занимает соответственно второе и третье места. Подобной стратегии придерживаются и другие крупнейшие поисковые системы, в частности Яндекс. Таким образом, первое, что увидит пользователь, решивший уточнить в Интернете значение незнакомого слова или получить информацию по интересующему его предмету, – это, в большинстве случаев, ссылка на статью Википедии (или Викисловаря, если соответствующая статья отсутствует в основном корпусе Википедии).

Безоговорочное доминирование Википедии в результатах поисковых запросов обусловлено радикальным пересмотром требований к информационно-справочным ресурсам, который был осуществлён разработчиками интеллектуальных поисковых систем. Традиционно в качестве главного достоинства справочных источников рассматривалась степень достоверности и полноты предоставляемой ими информации. Современные поисковые системы признают решающим критерий релеванности полученной информации исходному пользовательскому запросу. Иными словами, идеальным признаётся не идеально точный и содержательный ответ, а ответ, идеально соответствующий потребностям пользователя. Многократно отмечалось, что в большинстве случаев обращение к энциклопедическим материалам обусловлено необходимостью составить общее представление о предмете (для получения детальной и разносторонней информации даже эталонной энциклопедической статьи оказывается недостаточно, требуется изучение специальной научной литературы). Избыточная полнота и точность информации становится в данном случае таким же существенным изъяном, как её неточность и неполнота. Критерий безусловной научной достоверности для современного потребителя информации теряет актуальность и замещается критерием приемлемой степени достоверности.

Немаловажное значение приобретает и тот факт, что «волонтёрская» лексикография нередко опережает профессиональную, более оперативно подмечая происходящие в языке изменения, фиксируя неологизмы. В эпоху интернет-технологий частота обновления информационного ресурса является одним из важнейших преимуществ.

Можно констатировать, что «народная» лексикография сегодня не только вступает в конкурентные отношения с научной, более того, «свободная энциклопедия» и (в меньшей степени) её дочерний проект Викисловарь становятся для значительной части населения безальтернативным источником лексикографических сведений.

В этой связи изучение открытых энциклопедических и лексикографических интернет-источников, а также принципов «народной» лексикографии становится одной из приоритетных научных задач.

Системообразующие принципы функционирования Википедии таковы, что применительно к каждому опубликованному материалу она накапливает большой объём статистических данных, которые могут быть использованы для получения метаин-

формации (знаний о знаниях, которые со-держатся в энциклопедии). Это делает её не только источником материалов для научной аналитики, но одновременно эффективным научно-аналитическим инструментом.

Перечислим важнейшие, с точки зрения исследователя, аналитические средства «свободной энциклопедии»:

1. Википедия хранит полную хронологически упорядоченную (с точностью до минуты) историю правок каждой статьи с момента её создания, что позволяет достоверно установить, какие коррективы были внесены в документ на том или ином этапе, какие из них (и как скоро) были отвергнуты сообществом пользователей; проследить, как постепенно выкристаллизовывалась дефиниция и как она менялась с течением времени; изучить в диахроническом срезе соотношение между дефинитивной и описательной частью энциклопедической статьи, отсмотреть, какие семантические компоненты вытеснялись из дефиниции в описательную часть статьи, и наоборот.

2. Википедия хранит полную историю обсуждений, относящихся к каждой осуществлённой пользователями правке. Поскольку «свободная энциклопедия» является сообществом анонимных энтузиастов, правомерность тех или иных редакторских процедур не может быть узаконена ссылкой на профессиональную компетентность редактора (даже если пользователь выступает под своим реальным именем, сложно доказать, что это имя, а не псевдоним). Как следствие, любые принципиальные изменения принято обосновывать ссылкой на авторитетный источник либо, если это по каким-либо причинам неосуществимо, логической аргументацией. История обсуждений даёт возможность изучать «народную» лексикографию «изнутри», не только рассматривая результаты лексикографической деятельности как объективную данность, но и анализируя её мотивационную составляющую.

3. Википедия позволяет в хронологическом порядке отслеживать все правки, сделанные любым зарегистрированным пользователем во всех статьях, и все обсуждения, в которых он участвовал. Таким образом, становится возможным провести мониторинг лексикографической активности определённого автора / редактора, составить его «лексикографический профиль».

Рассмотрим указанные особенности на примере конкретной энциклопедической статьи – «Селфи», – посвящённой слову (и явлению), получившему в молодёжной среде практически повсеместную известность в течение последних полутора-двух лет. Ещё раз отметим, что мы обращаемся

к материалам энциклопедического (Википедия), а не лексикографического (Викисловарь) источника, поскольку в настоящий момент именно Википедия остаётся для большинства интернет-пользователей приоритетным источником лексикографической информации. Сведения о дате и времени, имени пользователя, а также цитаты из текста статьи по состоянию на тот или иной момент времени приводятся нами по данным метастраницы «Селфи – история изменений» [4], а также, в ряде случаев, связанных с нею метастраниц, доступных по прямым ссылкам.

Статью «Селфи» создаёт в Википедии анонимный автор в 16.57 7 декабря 2013 года. Уже в 23.07 того же дня программный механизм Википедии вносит несколько технических корректив, относящихся к шаблону статьи. В 07:00 8 декабря 2013 г. анонимный пользователь производит незначительную стилистическую правку. В 16:31 того же дня пользователь VAP+VYK заменяет в тексте статьи прямые кавычки «французскими» кавычками («ёлочками»), исправляет ‘е’ на ‘ё’ в слове ‘ещё’. Вслед за этим он открывает обсуждение статьи, в котором ставит под сомнение целесообразность её сохранения в Википедии: ‘Возможно, значимо. Но очень уж похоже на орисс’ (‘орисс’, от словосочетания ‘оригинальное исследование’, либо ‘перисс’, от ‘первоначальное исследование’, – термины, возникшие в среде редакторов Википедии и обозначающие впервые публикуемые материалы, ещё не получившие научного и общественного признания).

В 03:15 9 декабря 2013 г. пользователь Andshel соотносит статью с категориями «Фотография», «Автопортреты», «Интернет-культура», а также добавляет в неё две ссылки на авторитетные источники. Полемизируя с VAP+VYK, он указывает в обсуждении: ‘Точно значимо. Это слово в 2013 году официально признано Оксфордским словарём...’. В 18:01 14 декабря 2013 г. он проставляет ударение в заглавном слове и дорабатывает этимологическую справку: ‘Сѐлфи (англ. Selfie от англ. Self – сам, само)...’. В 23:30 25 декабря 2013 г. пользователь Камарад Че включает в этимологическую справку разговорный русскоязычный синоним заглавного слова: ‘Сѐлфи (самострел, англ. Selfie от англ. Self – сам, само)...’. В 03:55 26 декабря 2013 г. пользователь Andshel снабжает дефиницию гиперссылками на статьи «Автопортрет», «Мобильный телефон», «Фотоаппарат», вносит в текст ряд стилистических правок, заменяет (для англоязычных слов) «французские» кавычки прямыми.

Нет необходимости детально рассматривать здесь полную историю правок (всего за 11 месяцев с момента создания статьи их было сделано 163). Обозначим некоторые выводы, к которым позволяет прийти анализ метаинформации, связанной со статьями Википедии.

Википедическое сообщество оперативно. Динамика правок показывает, что работа по усовершенствованию статей ведётся постоянно и начинается сразу вскоре после их опубликования. Редактура осуществляется по нескольким направлениям: орфографические и пунктуационные, грамматические и стилистические правки, технические коррективы, уточнение структуры статьи, шаблонизация, содержательные уточнения, добавление новых фактов, добавление ссылок на авторитетные источники, устранение признаков авторского субъективизма и т.д. Сообщество своевременно реагирует на акты вандализма. В среднем с момента порчи до момента восстановления статьи проходит от нескольких минут до нескольких дней. Так, первый акт вандализма в отношении анализируемого нами материала был предпринят в 09:17 3 марта 2014 г., но уже спустя несколько секунд другим пользователем был произведён «откат» к стабильной версии. Второй акт вандализма состоялся в 09:51 7 июня 2014 г., в 13.15 его последствия были устранены.

Википедическое сообщество консервативно. Хотя большинство редакторов прекращают работу над статьёй после одной или нескольких правок, нередко выделяются группы «эпизодических» и постоянных редакторов (последние «присматривают» за статьёй регулярно). Статьи «свободной энциклопедии» на протяжении своего жизненного цикла сохраняют известную устойчивость. Как правило, на начальном этапе их существования, в течение некоторого времени после публикации, правки носят в основном технический характер. Симптоматично замечание пользователя Анатолич1, одного из первых редакторов статьи «Селфи»: *‘Хотел сам начать с нуля, но уважаемый аноним опередил меня, и теперь имеем то, что имеем’*. На следующем этапе активным переработкам подвергается описательная часть статьи. Впоследствии длительные периоды «вялотекущих» косметических правок перемежаются редкими всплесками лексикографической активности (обычно проявляемой «случайными» пользователями, не входящими в число «постоянных» редакторов). Наиболее устойчивой частью статьи является

дефиниция. Так, первая попытка кардинально переписать дефинитивную часть статьи «Селфи», в своём первоначальном виде абсолютно не соответствовавшую энциклопедическим стандартам, относится к 29 апреля 2014 г. (проходит 5 месяцев с момента опубликования текста).

Эволюция статей Википедии протекает в непрерывном противоборстве разговорно-бытового и научного дискурсов (с постепенным вытеснением первого вторым). История правок позволяет детально проследить этот процесс. В качестве иллюстрации сравним актуальную энциклопедическую дефиницию понятия (*‘Разновидность автопортрета, заключающаяся в запечатлении самого себя на фотокамеру. Термин приобрёл известность в конце 2000-х, начале 2010-х гг. благодаря развитию встроенных функций фотоаппарата мобильных устройств. Поскольку селфи чаще всего выполняется с расстояния вытянутой руки, держащей аппарат, изображение на фото имеет характерный ракурс и композицию – под углом, чуть выше или ниже головы’*) с исходной: *‘Снимок самого себя, сделанный самолично на мобильный телефон или фотокамеру, что называется, с руки. Как правило, изображения на фото получается под углом, так как снимок производится удерживая камеру чуть выше или ниже головы’* (нижним подчёркиванием выделены разговорные элементы).

Двойственное положение Википедии, которая одновременно выполняет функции и толкового словаря, и энциклопедического источника, провоцирует постоянные конфликты лексикографического и энциклопедического стилей изложения. Так, 6 сентября 2014 года пользователь Тортилиус заменяет в рассматриваемой нами статье энциклопедическое определение словарным: *‘Запечатление самого себя на камеру, чаще всего с вытянутой руки’*. Примечателен комментарий пользователя в обсуждении: *‘Написала всё простым языком, без умных слов) Просто понятно и просто)’*.

Итак, можно сделать вывод, что в ситуации обостряющегося конкурентного противостояния академической и «народной» лексикографии изучение последней является актуальной задачей, решение которой могут существенно облегчить аналитические инструменты Википедии, позволяющие отслеживать полную историю правок любой статьи, составлять «лексикографические профили» отдельных пользователей, на основе редакторских «маргиналий» изучать субъективную мотивацию внесённых в текст изменений.

Список литературы

1. Бабицкий А., Терентьев И. 22 главных тренда десятилетия (Основные тенденции 2000-х по версии Forbes) [Электронный ресурс] // Forbes Online: интернет-журнал. – 24.12.2010. – URL: <http://www.forbes.ru/ekonomika-photogallery/rynki/61509-22-glavnyh-trenda-desyatiletija/photo/16> (дата обращения: 11.11.2014).
2. Википедия: Список Википедий // Википедия: сайт. – URL: https://ru.wikipedia.org/wiki/Википедия:Список_Википедий (дата обращения: 11.11.2014).
3. Виртуальная реальность vs реальная жизнь: выбор «интернетчиков» // Всероссийский центр изучения общественного мнения: сайт. – Пресс-выпуск № 2090 (28.08.2012). – URL: <http://wciom.ru/index.php?id=459&uid=112964> (дата обращения: 11.11.2014).
4. Селфи – история изменений // Википедия: сайт. – URL: <https://ru.wikipedia.org/w/index.php?title=Селфи&action=history> (дата обращения: 11.11.2014).
5. Статистика // Викисловарь: сайт. – URL: <http://ru.wiktionary.org/wiki/Служебная:Статистика> (дата обращения: 11.11.2014).
6. Giles J. Special Report Internet encyclopaedias go head to head // *Nature*. – 2005. – № 438 (7070). – P. 900–901.
7. Silverwood-Cope S. Wikipedia: Page one of Google UK for 99% of searches // Intelligent Positioning: сайт. – 08.02.2012. – URL: <http://www.intelligentpositioning.com/blog/2012/02/wikipedia-page-one-of-google-uk-for-99-of-searches/> (дата обращения: 11.11.2014).
8. Wiktionary: Statistics // Wiktionary. URL: <http://en.wiktionary.org/wiki/Wiktionary:Statistics> (дата обращения: 11.11.14).

References

1. Babickij A., Terent'ev I. 22 glavnyh trenda desjatiletija: Osnovnye tendencii 2000-h po versii Forbes (22 main trend of the decade: Main trends of the 2000s according to Forbes), *Forbes Online*, 24.12.2010, available at: <http://www.forbes.ru/ekonomika-photogallery/rynki/61509-22-glavnyh-trenda-desyatiletija/photo/16> (accessed 11 November 2014).

2. Wikipedija: spisok Wikipedij (Wikipedia: List of Wikipedia), *Wikipedia*, available at: https://ru.wikipedia.org/wiki/Википедия:Список_Википедий (accessed 11 November 2014).
3. Virtual'naja real'nost' vs real'naja zhizn': vybor «internetчиков» (Virtual reality vs real life: the choice of «Internet users»), *Russian Public Opinion Research Center*, press release 2090 (28.08.2012), available at: <http://wciom.ru/index.php?id=459&uid=112964> (accessed 11 November 2014).
4. Selfi – istorija izmenenij (Selfie – Revision history), *Wikipedia*, available at: <https://ru.wikipedia.org/w/index.php?title=Селфи&action=history> (accessed 11 November 2014).
5. Statistika (Statistics), *Wiktionary*, available at: <http://ru.wiktionary.org/wiki/Служебная:Статистика> (accessed 11 November 2014).
6. Giles J. Special Report Internet encyclopaedias go head to head, *Nature*, 438, 900-901 (15 December 2005).
7. Silverwood-Cope S. Wikipedia: Page one of Google UK for 99% of searches, *Intelligent Positioning*, available at: <http://www.intelligentpositioning.com/blog/2012/02/wikipedia-page-one-of-google-uk-for-99-of-searches/> (accessed 11 November 2014).
8. Wiktionary: Statistics, *Wiktionary*, available at: <http://en.wiktionary.org/wiki/Wiktionary:Statistics> (accessed 11 November 2014).

Рецензенты:

Гуц Е.Н., д.фил.н., профессор кафедры русского языка, славянского и классического языкознания, ФГБОУ ВПО «Омский государственный университет им. Ф.М. Достоевского», г. Омск;

Косяков Г.В., д.фил.н., профессор, проректор по учебной работе, ФГБОУ ВПО «Омский государственный педагогический университет», г. Омск.

Работа поступила в редакцию 02.12.2014.