Шароян Сирануш Вардановна, аспирантка кафедры теории и экономики СМИ факультета журналистики МГУ имени М. В. Ломоносова; e-mail: siranush.sharoyan@gmail.com

ФАКТОРЫ ЧИТАТЕЛЬСКОГО ПРЕДПОЧТЕНИЯ АУДИТОРИИ ДЕЛОВЫХ ОНЛАЙН-СМИ (НА ПРИМЕРЕ САЙТА РБК)

Возможность точно и быстро измерять предпочтения аудитории, появившаяся у средств массовой информации благодаря развитию инструментов веб-аналитики, в корне меняет работу редакций онлайн-изданий и их отношения с аудиторией. Журналисты и редакторы, изучая веб-метрики, все больше узнают о вкусах своих потребителей и пытаются адаптироваться к ним. А умение грамотно анализировать и интерпретировать эти данные становится ключевым для выживания СМИ в условиях растущего экономического давления. При этом научных исследований о предпочтениях аудитории СМИ, основанных на анализе данных, полученных с помощью систем веб-аналитики, в России до сегодняшнего дня не было.

Данная статья представляет результаты анализа самых популярных материалов одного из крупнейших российских онлайн-СМИ – сайта РБК, ежемесячная аудитория которого превышает 25 миллионов читателей. Статья основана на анализе более тысячи материалов, которые в течение года с июля 2015 г. набрали наибольшее количество просмотров. Автор впервые в российской научной практике делает регрессионный анализ, позволяющий оценить влияние различных факторов на потенциальную популярность статьи среди читателей сайта. Результаты анализа показывают, что на потенциальное количество просмотров сильнее всего влияют попадание материала в определенные рубрики, количество слов в заголовке и наличие в нем слов-маркеров, жанровая принадлежность и количество перепостов в социальных сетях.

Ключевые слова: *веб-метрики, веб-аналитика, аудитория СМИ, измерение аудитории, онлайн-журналистика, новости, статьи.*

Siranush V. Sharoyan, PhD student at the Chair of Media Theory and Economics, Faculty of Journalism, Lomonosov Moscow State University; e-mail: siranush.sharoyan@gmail.com

FACTORS OF READING PREFERENCES OF THE ONLINE BUSINESS MEDIA AUDIENCE (A CASE STUDY OF THE RBC WEBSITE)

The media's ability to accurately and quickly measure audience preferences has emerged through the development of web analytics tools and dramatically changed the operation of online media outlets and their relations with readers. Studying web metrics, editors and reporters are now becoming more aware of and adaptable to

consumers' tastes. The skill of analyzing and interpreting these data correctly is a factor of a media outlet's survival in the face of growing economic pressure. In Russia, however, there is still no scientific quantitative research into media audience preferences.

This paper presents the results of a study of the most popular articles on the RBC website, one of the largest Russian online media with a monthly audience of more than 25 million readers. The paper is based on an analysis of more than a thousand articles that within one year from July 2015 scored the highest number of views. For the first time in Russian scientific practice, the author carries out a regression analysis allowing to evaluate the influence of various factors on the potential popularity of the article among the readers of the website. The results show that the number of views largely depends on the rubric the article falls within, the number of words in the headline and the presence of key words in it as well as the article's genre identity and the number of reposts on social networks.

Key words: *web metrics, web analytics, media audience, audience measurement, online journalism, news, articles.*

Веб-аналитика и СМИ

Исследователи средств массовой информации на протяжении десятилетий изучают процесс создания новостей, акцентируя особое внимание на том, *как* журналисты и редакторы фильтруют и трансформируют множество событий, о которых они узнают в течение дня, в ограниченное количество материалов, выбирая только ту информацию, которая, на их взгляд, достойна освещения. Сам процесс отбора обозначается английским термином «gatekeeping», который можно перевести как «контроль доступа» (Shoemaker, 1991), и его значение сложно переоценить: результат отбора тем формирует повестку дня и оказывает сильное влияние на взгляды аудитории (Shoemaker, Vos, 2009).

Единого рецепта для выбора тем и создания хорошего материала, конечно, нет. Но существует, например, пять критериев, которые воспроизводятся в разных журналистских кругах, преподаются в школах журналистики и обсуждаются в профессиональных журналах: это своевременность, релевантность, близость к аудитории, сенсационность и наличие конфликта (Schultz, 2007). При этом редакторы и журналисты, вырабатывая эти критерии и оценивая любую информацию по ним, полагались, в первую очередь, на «внутреннее чувство», в то время как интересы и предпочтения аудитории часто оставались без внимания (Boczkowski, Peer, 2011).

Ситуацию изменило внедрение в СМИ веб-аналитики – систем сбора, измерения, анализа и интерпретации информации о посетителях веб-сайтов, которую используют для улучшения и

оптимизации самих сайтов и их контента. История веб-аналитики началась в середине 1990-х, когда владельцы сайтов смогли устанавливать простые счетчики, показывающие количество просмотров страницы: каждый раз, когда на сайт заходил новый посетитель, число на счетчике увеличивалось на единицу.

С развитием технологий возможности веб-аналитики стали почти безграничными. Сейчас СМИ могут получить практически любую информацию о своей аудитории: от данных о том, сколько людей просмотрели материал, до того, когда и откуда читатели приходят на каждую страницу, как долго ее изучают, куда уходят потом и т.д. (Kaushik, 2009). Интересно, что все это не требует никаких усилий со стороны аудитории, активность пользователей регистрируется независимо от того, осознает это читатель или нет. Статистика, основанная на записи онлайн-активности посетителей, используется для выделения паттернов поведения аудитории и выявления тенденций (McGregor, 2007).

Теперь руководство большинства редакций онлайн-СМИ начинает утренние совещания с краткой сводки о самых популярных материалах предыдущего дня (Peters, 2010), а редакторы сайтов в режиме реального времени корректируют положение новостей на веб-страницах, ориентируясь на количество просмотров той или иной истории (Lee at al, 2014). Эти изменения приводят к появлению «повестки аудитории» (Anderson, 2011), меняя устоявшиеся представления об односторонности массовой коммуникации, «контроле доступа», самостоятельном выборе повестки дня редакцией и степени влияния аудитории на медиаконтент (Singer, 2011).

Таким образом, появление систем веб-аналитики в средствах массовой информации побудило их сотрудников «увидеть свою аудиторию, осознать ее размер и желания, позволив вместо смутных впечатлений об их предпочтениях получить точные количественные данные» (Napoli, 2010). Тем не менее, научных исследований предпочтений аудитории и факторов, влияющих на эти предпочтения, на сегодняшний день в России нет. Научная новизна работы заключается в том, что автор, проанализировав материалы, которые набирали наибольшее количество просмотров в одном из крупнейших онлайн-изданий России – сайте РБК (www.rbc.ru), выделяет факторы, которые влияют на интерес аудитории к материалам, измеряет степень этого влияния и выделяет среди многочисленных количественных данных тенденции и тренды.

Автор впервые в российской научной практике проводит регрессионный анализ популярных материалов, который позволяет объяснить, какие факторы влияют на читаемость статей в целом, анализируя каждую статью не как нечто целостное, но как набор отдельных элементов (заголовка, лида, основного текста), пытаясь измерить влияние на популярность материалов каждого из них. Для этого нужно ответить на два вопроса:

Вопрос 1. Какие материалы оказываются самыми читаемыми на сайте РБК?

Вопрос 2. Какие факторы оказывают влияние на популярность материалов?

Данные и методология

Веб-аналитика становится неотъемлемой частью жизни любой редакции. При этом пока главные вопросы, на которые она может дать ответ, – чего хочет аудитория от СМИ, какие материалы кажутся ей наиболее интересными, что влияет на просматриваемость – пока остаются без ответа. Проанализировав наиболее популярные материалы сайта РБК, который входит в тройку самых популярных онлайн-СМИ России с ежемесячной аудиторией превышающей 25 млн читателей, можно приблизиться к получению ответов на эти вопросы.

Работа автора основана на анализе более тысячи материалов с сайта РБК (www.rbc.ru), которые за год с 1 июля 2015 г. по 31 июня 2016 г. еженедельно набирали наибольшее количество просмотров. Для составления выборки определялись двадцать самых популярных статей за каждую неделю выбранного периода (т.е. статей, которые за неделю с момента публикации набирали наибольшее количество просмотров). Анализ именно популярных материалов позволил понять, какие статьи имеют наибольшее количество просмотров.

Данные собирались с помощью системы веб-аналитики AT Internet. Несмотря на то, что руководство РБК установило и другие системы веб-аналитики, такие как Google Analytics и Яндекс.Метрика, именно показатели счетчика AT Internet департамент маркетинга и руководство сайта в рассматриваемый период использовали как основной для того, чтобы отследить поведение аудитории сайта. Каждое утро встречи в редакции начинались с обсуждения данных по трафику сайта в целом и читаемости отдельных статей.

Руководство редакции (главный редактор, его заместители и редакторы отделов) каждое утро получали отчет о двадцати самых читаемых и пяти самых нечитаемых статьях за предыдущий день с указанием нескольких ключевых характеристик для каждой из них: количестве просмотров, среднем времени чтения и источниках трафика. Доступ к этим отчетам могли получить и другие сотрудники редакции, например журналисты, которым было интересно, как аудитория взаимодействует с их материалами.

Среди данных, которые позволяет получить счетчик AT Internet, есть, например, информация о посещаемости каждого материала, среднем времени, которое читатели проводили за его чтением, поведении после – покинули они сайт или перешли по одной из гиперссылок на сайте, предлагающих ознакомиться с другими материалами (об источниках трафика, количестве репостов и т.д.). Помимо этих данных, автор также выявил для анализа факторы, которые теоретически могли оказывать влияние на интерес аудитории к отдельным статьям.

Так, одним из наиболее значимых факторов, определяющих интерес к материалу, является его тематика. В первую очередь, она отражается в **рубрике**, в которую попадает материал. На сайте РБК существует 6 тематических рубрик: «политика», «экономика», «финансы», «технологии и медиа», «бизнес» и «свое дело».

Кроме того, тематика материала отражается в его **заголовке**, поэтому автор подробно анализировал заголовки выбранных статей. Для начала все заголовки были разделены на четыре типа, согласно делению, которое существует в редакции РБК. Первый тип заголовка – «*новостной*» – коротко и максимально подробно обозначает, какая информация содержится в статье. Это классический пример заголовка новости, построенной по принципу перевернутой пирамиды – все самое важное читатель узнает уже из первых строк, авторы статей с такими заголовками надеются, что читатель заинтересуется обозначенной проблемой и откроет материал, чтобы прочитать подробности. Приведем пример такого заголовка: «Сеть Kira Plastinina получила десятки исков на общую сумму в 45 млн руб»[1].

Второй тип заголовка – «*открытый*». Его цель – заинтриговать читателя, поэтому автор пытается сохранить в таком заголовке интригу. Например: «ЦБ отозвал лицензию у двух банков»[2]. Большинству читателей, у которых есть депозиты в банках, скорее всего, будет интересно, не об их ли банке идет речь. Другой пример:

«“Когалымавиа” назвала "единственную разумную причину разрушения" A321»[3]. Что это за проблемы – из заголовка непонятно, но, теперь мы знаем, что они есть, и нам интересно, о чем пойдет речь в статье. Открытый заголовок – это почти всегда выбор автора, ведь для таких заметок можно было бы легко подобрать соответствующие новостные заголовки, которые звучали бы, например, так: «ЦБ отозвал лицензии у Эл Банка и Мострансбанка»; «“Когалымавиа” назвала внешнее воздействие единственной возможной причиной крушения A321».

Третий тип заголовка – «*двухчастный*». Он обычно используется в неновостных жанрах, материалах, где автор пытается проанализировать и объяснить какие-то события, тренды и т.д. Первая часть в таком заголовке – более образная, метафоричная, вторая – более содержательная, дающая читателю понять, о чем пойдет речь в статье. Нередко такие заголовки используются в авторских колонках, и тогда первая часть может отражать отношение автора к тому, о чем написан материал. Приведем несколько примеров таких заголовков: «За Сирию, за Асада: когда и зачем Москва решила воевать», «Удар на $44 млрд: во сколько может обойтись конфликт России и Турции»[4]; «Новая реальность: что ожидает россиян и бизнес после обвала рубля»[5].

Наконец, четвертый тип заголовка – это «*заголовок-цитата*». Чаще всего используется для обозначения интервью, например: «Вагит Алекперов — РБК: «У нефтяников сейчас не остается никаких денег»[6]. Но бывают и исключения, когда таким заголовком обозначают, например, общее настроение или самое яркое высказывание из нескольких выступлений, например: «Мы не нуждаемся в лекциях»: как страны ЕС бунтуют против правил союза»[7].

Заголовки выбранных материалов были также проанализированы по их **тональности**, чтобы можно было определить, есть ли какая-то зависимость от того, хорошая эта новость или плохая, и ее читаемостью. По этому принципу выделено три типа заголовка: негативные («Центральный банк отозвал лицензию у двух банков»), нейтральные («Греф предсказал скорый отказ клиентов Сбербанка от пластиковых карт») и позитивные («Порошенко предложил объявить в Донбассе перемирие на Пасху»).

Наконец, была проанализирована и зависимость читаемости статьи от **длины заголовков** и наличия в них **слов-маркеров**, привлекающих внимание читателей. Для того чтобы выделить такие слова, автор проанализировал тексты заголовков и выявил наиболее

часто встречающиеся слова (приложение 1). Как оказалось, это слова «Путин» (178 раз), «Сирия» (106 раз), «Украина» (94 раза), «Россия» (88 раз), «США» (69 раз). Автор также проанализировал вероятное влияние этих слов на популярность статьи среди читателей сайта РБК.

Еще один «тематический» фактор, влияние которого анализировал автор, – это попадание материалов в **сюжеты**, которые редакция РБК создавала, когда происходили какие-то важные события, требующие долгого освещения с помощью большого количества материалов. К таким сюжетам в анализируемый период относились, например, «Война в Сирии», «Украинский кризис», «Валютный кризис» и т.д. Названия сюжетов отражаются в верхней части сайта РБК, и, открывая любой из них, читатель попадает на страницу со всеми материалами по интересующей его теме – от коротких новостей до развернутых аналитических статей и авторских колонок.

Анализировалась и **общественная значимость** материала с точки зрения формирования общественной повестки. Исследователи Reuters Institute так описывают разделение на «жесткие» и «мягкие» новости (hard/soft news): первые обычно относятся к важнейшим для общественной жизни темам, таким как политика, экономика, международные отношения и т.д. В то время как «мягкие» новости менее значимы с общественной точки зрения, к ним относят, например, материалы о стиле жизни (lifestyle), развлечениях и т.д.

Кроме того, автор рассматривал в качестве одного из факторов «**эксклюзивность**» материалов. Речь идет об имеющей общественную значимость информации, которую журналисту удалось добыть раньше всех.

Еще один важный фактор, влияющий на читаемость материала, – его **жанровая принадлежность**. Под журналистскими жанрами подразумеваются «устойчивые типы публикаций, объединенных сходными содержательно-формальными признаками» (Тертычный, 2000). Издание чаще всего работает с небольшим числом форматов (Амзин, 2013). Поэтому при анализе выбранных статей автор взял за основу жанровое разделение, которое принято в РБК. В результате, среди анализируемых материалов оказались статьи девяти жанров:

- *новость* – короткое сообщение, построенное по принципу перевернутой пирамиды и содержащее ответы на шесть ос-

новных вопросов (кто? что? где? когда? почему? каким образом?);

- *заметка* – сообщение, содержащее, помимо описания сути события и ответов на основные вопросы, подробный рассказ о нем с деталями и бэкграундом;
- *фичер* – история, написанная от третьего лица, но дающая возможность пережить случившееся (Колесниченко, 2008);
- *интервью* – жанр, который представляет собой диалог репортера с одним человеком или несколькими, представляющими интерес для зрителей или читателей;
- *расследование* – «материал, основанный, как правило, на собственной работе и инициативе, на важную тему, которую отдельные лица и организации хотели бы оставить в тайне» (Уллмен, 1998);
- *репортаж* – освещение события/мероприятия непосредственно с места событий;
- *онлайн-трансляция* – относительно новый жанр интернет-журналистики, который выходит не единым материалом, а небольшими отрывками, при этом предлагая «максимально документальную передачу и максимально жесткое совпадение вещания по времени с тем, что происходит» (Лосева, 2016);
- *фотогалерея* – жанр, представляющий собой объединенные в одной статье фотографии с места события, которое имело яркую визуальную составляющую;
- *мнение* – жанр, представляющий собой высказывание приглашенного эксперта на злободневную тему.

Следующий фактор, который анализировался автором, всегда остается под пристальным вниманием руководства РБК – это **источники трафика**. Анализировались два основных источника, которые приносят сайту РБК большую часть просмотров – прямые заходы, заходы с новостных агрегаторов. При *прямых заходах* пользователи переходят на сайт из «закладок», сделав ее стартовой страницей или набрав прямой адрес сайта в адресной или поисковой строке. Такой тип поведения считается наиболее ценным и интерпретируется как относящийся к сверхлояльной аудитории, ядру медиаресурса (Лосева, 2016). Второй источник – *новостные агрегаторы*, в основном Яндекс Новости и Новости Google. При благополучном стечении обстоятельств трафик с агрегаторов спо-

собен удвоить суточную аудиторию даже такого крупного интернет-СМИ, как РБК.

Анализировались также показатели лояльности пользователей: внутренние переходы, время, проведенное за чтением статьи, и количество перепостов статей в социальных сетях. **Внутренние переходы** отражают долю посетителей сайта, которые не покидают его, прочитав один материал (показатель отказов), а переходят на другие статьи по расположенным на странице ссылкам. В свою очередь, **глубина прочтения** говорит о высоком уровне вовлеченности аудитории и активном взаимодействии с ресурсом. Чтобы рассчитать этот показатель, автор вычислял время, которое бы потребовалось на прочтение статьи среднестатистическому читателю. Для этого он делил количество слов в каждом материале на среднее количество слов, которое человек может прочесть за минуту (120), получив предполагаемое время чтения. Наконец, для определения вовлеченности читателей среднее время чтения, которое все читатели, открывшие статью, провели за ее просмотром, делилось на предполагаемое время чтения. Чем выше показатель, тем интереснее материал. О том, что статья показалась читателям интересной и полезной, говорят и **перепосты в социальных сетях** – с их помощью пользователи пытаются привлечь внимание друзей к заинтересовавшему их материалу.

Наконец, автор попытался найти зависимость количества просмотров материалов от дня недели, в который они были опубликованы, и времени публикации.

Для анализа влияния различных факторов на читаемость статьи было решено провести **регрессионный анализ**. Используя данный метод и удалив 10% аномальных наблюдений (чтобы избавиться от проблемы аномальных наблюдений), удалось получить много интересных и статистически значимых результатов. Мы предлагаем линейную модель с константой:

$$\begin{aligned} \mathrm{No}_{\mathrm{read}} = {} & \beta_1 * Words + \beta_2 * Big\ topic + \beta_3 * Reposts + \beta_4 * Soft + \beta_5 * Exclusive \\ & + \sum \beta_i * Key\ Words_i + \sum \beta_i * \mathrm{G}_i + \sum \beta_i * \mathrm{G}_i \\ & + \sum \beta_i * MD_i + \sum \beta_i * R_i + \sum CV + \mathrm{Const} \end{aligned}$$

Где:

- Words – количество слов в заголовке;
- Big topic – индикатор того, что тема попала в большой сюжет;

- Reposts – количество репостов статьи в соцсетях;
- Key words – наличие ключевых слов в заголовке (Путин, Сирия, Россия, Украина, США);
- Soft – общественная значимость;
- G – жанр статьи;
- MD – тон заголовка;
- R – рубрика;
- CV – включает ряд стандартных для подобных исследований контрольных переменных для исключения проблемы эндогенности:
    - месяц публикации статьи (m_{1-12});
    - день публикации статьи (d_{1-7});
    - время публикации статьи, 12 двухчасовых интервалов (t_{1-12});
    - предположительное время чтения статьи (est_read);
    - показатель внутренних переходов (ent_rate);
    - количество прочтений от пользователей, пришедших через прямой трафик (drct_traff);
    - количество прочтений от пользователей, пришедших по ссылкам с новостных аггрегаторов (indrct_traff).

В данной модели выполняются все необходимые условия для выполнения регрессии:

1. Топ статей определяется *случайным образом,* исключительно исходя из предпочтений читателей сайта РБК.
2. Были учтены все возможные измеряемые количественные факторы, которые могли повлиять на читаемость материалов.
3. Входящие в модель переменные независимы друг от друга.
4. Условие нулевого математического ожидания ошибок выполняется автоматически, так как модель имеет константу.

В заключение стоит отметить, что из-за включения в выборку только статей, попавших в список самых читаемых, некорректно обобщать оценки, полученные в данной работе, без поправки на смещение. С высокой долей вероятности полученные в данной работе оценки смещены вверх, из чего можно сделать вывод о более низком влиянии исследуемых факторов на читаемость статей. Чтобы определить смещение, необходимы дополнительные данные, достать которые на данном этапе не пред-

ставляется возможным. Однако полученные оценки хорошо отражают влияние факторов на читаемость популярных статей и являются состоятельными, несмещенными и лучшими оценками в классе линейных.

Результаты анализа

Как отмечалось выше, веб-аналитика нужна, в первую очередь, для того, чтобы анализировать существующие в поведении аудитории веб-сайтов тренды. Созданная автором модель помогает понять, какие факторы влияют на читаемость статей на одном из основных онлайн-СМИ России – сайте РБК.

В анализируемую выборку попало больше тысячи статей. Из них больше половины – 53% относились к рубрике «политика», еще 14% – к «обществу», по 10% – к «бизнесу» и «финансам», 9% – к «экономике». Меньше всего популярных статей оказалось в рубрике «технологии и медиа» – всего 4%, причем среднее количество прочтений материалов этой рубрики также оказалось минимальным – 346 тысяч просмотров (против 385 тысяч просмотров в среднем по выборке) (см. табл. 1).

Таблица 1

Распределение выборки по рубрикам

Рубрика	Количество материалов, шт.	Общее количество просмотров, тыс. просмотров	Среднее количество просмотров, тыс. просмотров
Политика	547	228 463	418
Общество	143	56 487	395
Бизнес	109	40 719	374
Финансы	108	43 275	401
Экономика	92	34 932	380
Технологии и медиа	22	7 618	346

Источник: результаты исследования, расчеты автора.

Среднее количество слов в заголовке анализируемых материалов составило 8,5 слов. Среди них оказалось 53% материалов с «новостными» заголовками, 32% материалов с «открытыми» заголовками, 11% материалов с «двухчастными» заголовками и только в 4% статей использовались «заголовки-цитаты». При этом большинство статей имеют или негативный (51%), или нейтральный (40%) заголовок, на материалы, которые имеют позитивные заголовки приходится только 9% анализируемых статей (см. табл. 2).

Таблица 2

Распределение заголовков

Тип заголовка	**Количество материалов, шт.**	**Общее количество просмотров, тыс. просмотров**	**Среднее количество просмотров, тыс. просмотров**
Новостной	557	219792	395
Открытый	331	136177	411
Двухчастный	119	49524	416
Заголовок-цитата	14	6001	429

Источник: результаты исследования, расчеты автора.

В выборку попали материалы девяти жанров, причем подавляющее большинство – 67% относилось к жанру «новости». Также в выборке много заметок (24%). Все это неудивительно, учитывая, что именно эти жанры являются базовыми для сайта РБК и требуют меньше всего времени на создание. Остальные жанры либо требуют долгой работы для написания материала (например, расследования или фичер), либо используются только тогда, когда происходят какие-то выходящие из ряда вон события (например, онлайн-трансляции и фоторепортажи), поэтому их в выборке меньше (см. табл. 3).

Таблица 3

Распределение выборки по жанрам

Жанр	Количество материалов, шт.	Общее количество просмотров, тысяча просмотров	Среднее количество просмотров, тысяча просмотров
Новость	686	272288	397
Заметка	260	105154	404
Расследование	14	5317	380
Фотогалерея	13	7137	549
Фичер	13	4514	347
Мнение	12	4889	407
Интервью	11	4438	403
Репортаж	6	2530	422
Онлайн-трансляция	6	2210	368

Источник: результаты исследования, расчеты автора.

Стоит отметить, что статьи публикуются достаточно равномерно в течение рабочего дня. Большая часть материалов, получивших максимальное количество просмотров, была опубликована в промежуток с 10 утра до 18 часов вечера (54%), меньше всего – с полуночи до 8 утра (6,7%). Публикации по рабочим дням также довольно равномерно распределены. При этом в выходные вышло только 10% материалов, которые попали в выборку (см. табл. 4).

Таблица 4

Распределение выборки по времени публикации

Время публикации	Количество материалов, шт.	Общее количество просмотров, тысяча просмотров	Среднее количество просмотров, тысяча просмотров
12:00 – 14:00	167	69452	416
14:00 – 16:00	149	60616	407
16:00 – 18:00	119	48155	405
10:00 – 12:00	117	47074	402

Окончание таблицы 4

22:00 – 00:00	88	34710	394
20:00 – 22:00	83	36007	434
18:00 – 20:00	74	29760	402
8:00 – 10:00	71	25155	354
0:00 – 02:00	60	24436	407
06:00 – 08:00	35	13334	381
02:00 – 04:00	32	13029	407
04:00 – 06:00	26	9765	376

Источник: результаты исследования, расчеты автора.

Средняя глубина прочтения анализируемых материалов составила 6 минут, средний процент внутренних переходов с этих материалов – 63% (т.е. только 37% людей, которые просмотрели материал из выборки, уходят с сайта РБК, остальные продолжают чтение, перейдя по одной из предложенных ссылок).

Для статей, попавших в выборку, средний показатель прямых переходов составляет 12%, переходов из новостных агрегаторов – 20%. Только 14% попавших в выборку статей оказались эксклюзивными, при этом 70% материалов оказались общественно значимыми. Среднее количество перепостов анализируемых статей – 628.

Таким образом, отвечая на поставленный выше ***Вопрос 1*** о том, какие материалы оказываются самыми читаемыми на сайте РБК, можно отметить, что это в основном новости и заметки, связанные с общественной и политической тематикой. Чаще всего это материалы с негативными или нейтральными заголовками, большинство из них затрагивают общественно значимые вопросы.

Дальнейшие результаты регрессионного анализа, проведенного автором, можно увидеть в таблице 5. Стоит отметить, что в своих выводах автор фокусируется только на тех факторах, которые оказались значимыми по итогам используемой модели (иначе говоря, выводы, приведенные далее, верны в 90–99% случаев). Остальные факторы получились статистически незначимыми, то есть полученной оценке можно доверять менее чем в 90% случаев. Однако это не означает, что они не влияют на количество прочтений. При получении более полных данных (например, данных по всем статьям РБК, а не только по попавшим в список самых читаемых) эти факторы также могут стать значимыми.

Таблица 5

Влияние факторов на читаемость материалов

Фактор	Значение	Влияние	Значимость
Рубрика	Экономика	18 657	99%
	Бизнес	23 637	99%
	Политика	31 955	не значимо
	Общество	25 942	не значимо
	Финансы	10 586	не значимо
	Технологии и медиа	23 637	не значимо
Тип заголовка	Новостной	-52 778	не значимо
	Открытый	-62 038	не значимо
	Двухчастный	-75 490	не значимо
	Заголовок-цитата	75 490	не значимо
Тональность заголовка	Позитивная	-631	99%
	Нейтральная	-621	не значимо
	Негативная	-630	не значимо
Количество слов в заголовке	Каждое новое слово	-869	99%
Слова-маркеры	Путин	11 448	90%
	Украина	-3 136	99%
	Сирия	25	не значимо
	Россия	-4 779	не значимо
	США	7 881	не значимо
Попадание в сюжет	Да	-3 404	не значимо
Общественная значимость	Нет	-1 055	99%
	Да	1 344	не значимо
Эксклюзивность	Да	5 184	не значимо
Жанр	Новость	4 599	не значимо
	Заметка	17 150	не значимо
	Фичер	13 672	не значимо
	Расследование	-25 499	не значимо
	Интервью	-11 880	не значимо
	Репортаж	-16 757	не значимо
	Онлайн-трансляция	88 735	не значимо
	Фотогалерея	33 744	не значимо
	Мнение	96 634	99%
Перепосты в социальных сетях	Каждый репост	2,5	99%

Источник: результаты исследования, расчеты автора.

Как мы видим, попадание статьи в рубрики «экономика» и «бизнес» при прочих равных увеличивает количество просмотров. Если материал относится к рубрике «бизнес», его откроют на 6,1% (или на 23,6 тысячи) больше читателей сайта, если к экономике – на 4,8% (или на 18,7 тысяч). Попадание в остальные рубрики – политика, общество, технологии и медиа, бизнес и свое дело – не дает статистически значимых результатов. При этом, как показал анализ, попадание материалов в крупные тематические сюжеты на популярность материала на сайте РБК не влияет. Влияние общественной значимости на просматриваемость материалов минимальна: анализ показывает, что при прочих равных условиях статьи, которые относятся к незначимым, просматривают на 0,2% меньше.

Как показал регрессионный анализ, тип заголовка не оказывает особого влияния на показатель просмотров. Незначимым оказался и фактор «тональности» заголовка. Согласно модели, статистически значимое влияние оказывает только негативный тон, но и оно оказывается почти неощутимым – всего 0,2%. Такое низкое влияние на читаемость подтверждает, что люди больше ориентируются на контент, чем на настроение. Стоит отметить, что каждое новое слово в заголовке незначительно уменьшает количество просмотров. По прогнозам модели, добавление каждого нового слова к заголовку статьи на сайте РБК уменьшает количество просмотров на 0,2% (или на ~900 прочтений).

Слова-маркеры в заголовке влияют на количество просмотров. Так, слово «Путин» в заголовке увеличивает, а «Украина» незначительно уменьшает читаемость статьи. Наличие слова «Путин» в заголовке при прочих равных факторах увеличивает количество прочтений на 3% (или 11,4 тысячи прочтений). В то же время «Украина» незначительно (~1%, или 3 тысячи прочтений) уменьшает количество прочтений при прочих равных. Такие результаты, можно объяснить тем, что люди не хотят пропустить новости о событиях, связанных с президентом страны Владимиром Путиным, так как на подсознательном уровне воспринимают их как важные, и автоматически открывают их, когда видят фамилию президента в заголовке. В то же время тематика Украины значительно утомила читателя, и некоторые новости по этой теме могут намерено пропускаться. Остальные выделенные автором слова-маркеры – «Россия», «США» и «Сирия» – не оказывали статистически значимого эффекта на количество просмотров.

Жанр «мнение» значительно увеличивает читаемость статьи. Анализ показал, что в аудитория РБК предпочитает статьи в жанре «мнение». При прочих равных их прочитает на ~25% больше читателей (или ~96,6 тыс. просмотров), чем другие материалы на сайте. Эта находка ещё раз подтверждает устоявшееся мнение, что сегодня аудитория веб-сайтов отдает предпочтение не «журналистике факта», а «журналистике мнений».

Анализ показывает, что каждый репост статьи, при прочих равных факторах, увеличивает количество прочтений в 2,5 раза. Учитывая, что среднее количество репостов у анализируемых статей равно 628, можно сделать вывод, что в среднем благодаря репостам статьи читают на полторы тысячи человек больше. Этот вывод совпадает с интуицией: каждый репост статьи увеличивает число людей, которые узнают о существовании статьи, следовательно, увеличивается количество потенциальных читателей. Поэтому активное продвижение статьи в социальных сетях приносит прирост в количестве прочтений, хотя и не очень значимый.

Таким образом, отвечая на ***Вопрос 2*** о том, какие факторы оказывают влияние на популярность материалов, можно отметить, что положительное влияние на количество просмотров при прочих равных оказывает попадание в рубрики «экономика» и «бизнес», слово «Путин» в заголовке, принадлежность материала к жанру «мнение» и большое количество репостов. Отрицательное влияние оказывает слово «Украина» в заголовке, добавление новых слов, которые делают заголовки чересчур длинными, и низкая общественная значимость материалов.

Заключение

Появившиеся два десятилетия назад системы веб-аналитики сегодня могут рассказать почти все об аудитории онлайн-изданий. Открывая любую страницу, посетитель оставляет «цифровой след», который позволяет владельцам сайта, имеющим доступ к веб-метрикам, понять, как он ее нашел (зайдя на сам сайт, увидев ссылку в одном из новостных агрегаторов или в социальных сетях и т.д.), как долго просматривал, покинул ли сайт после прочтения материала или решил продолжить изучение других статей на сайте, поделился ли прочитанным с друзьями и т.д. Благодаря таким «следам», редакции и руководство средств массовой информации

получили возможность очень точно и быстро измерять предпочтения своей аудитории.

Такая возможность имеет важное значение для любого СМИ, ведь от количества просмотров зависит экономическая состоятельность подавляющего большинства изданий. Поэтому в выигрыше оказываются редакции, которые лучше понимают свою аудиторию и ее предпочтения, выраженные уже не в смутных интуитивных ощущениях, а в четких количественных показателях.

Приблизиться к пониманию аудитории помогает анализ самых популярных материалов одного из крупнейших веб-сайтов страны. С его помощью мы можем понять, какие же статьи пользуются наибольшей популярностью. Работа автора основана на анализе более тысячи материалов с сайта РБК (www.rbc.ru), которые за год с 1 июля 2015 г. по 31 июня 2016 г. еженедельно набирали наибольшее количество просмотров. Для осуществления анализа автор также выявил факторы, которые теоретически могли оказывать влияние на интерес аудитории к отдельным статьям – от рубрики и типа заголовка до попадания статьи в один из крупных тематических сюжетов и ее общественной значимости. Полученные данные обрабатывались с помощью регрессионного анализа, а сделанные выводы оказались верны в 90-99% случаев. Перечислим основные итоги:

- несмотря на то что большинство статей, попавших в выборку, были посвящены политическим и общественным событиям, регрессионный анализ показал, что попадание в рубрики «бизнес» и «экономика» при прочих равных увеличивает потенциальное количество просмотров;
- попадание материалов в крупные тематические сюжеты, тип заголовка, его тон, а также жанровая принадлежность практически не влияют на популярность статей на сайте РБК;
- при этом добавление каждого нового слова к заголовку статьи уменьшает количество просмотров, но незначительно: при переходе большинство читателей ориентируется на контент заголовка, а не на количество слов в нем;
- значимое влияние оказывает наличие слов-маркеров. Например, слово «Путин» в заголовке при прочих равных факторах увеличивает количество прочтений на 3%, а слово «Украина» незначительно (~1%,) уменьшает количество прочтений.

Примечания

[1] http://www.rbc.ru/business/22/07/2015/55af9b1e9a79470b7922870d
[2] http://www.rbc.ru/finances/05/05/2016/572ae14e9a79475fee178fbc
[3] http://www.rbc.ru/society/02/11/2015/563730199a7947cf10134658
[4] http://www.rbc.ru/politics/30/09/2015/560bffdd9a794744eb92da3b
[5] http://www.rbc.ru/economics/20/01/2016/569f8f9c9a79474677892137
[6] http://www.rbc.ru/interview/business/20/01/2016/569fa24f9a79475cf2473e51
[7] http://www.rbc.ru/photoreport/06/07/2016/577bbf789a79475798cd46eb

Библиография

Амзин А. Новостная интернет-журналистика. М.: Аспект-пресс, 2011.

Колесниченко А.В. Практическая журналистика: учеб. пособие. М.: Изд-во Моск. ун-та, 2008.

Кройчик Л.Е. Система журналистских жанров // Основы творческой деятельности журналиста / ред.-сост. *С.Г. Корконосенко*. СПб: Знание, СПбИВЭСЭП, 2000.

Лосева Н. Аудитория новых медиа // Как новые медиа изменили журналистику. Екатеринбург: Гуманитарный университет, 2016.

Мелихов Д., Сарматов И. Веб-аналитика: шаг к совершенству. К.: Аналитик Интеллект Сервис, 2010.

Тертычный А.А. Жанры периодической печати. Учебное пособие. М.: Аспект Пресс, 2000.

Уллмен Дж. Журналистские расследования: современные методы и техника. М., 1998.

Яковлев А., Довжиков А. Веб-аналитика: основы, секреты, трюки. СПб.: БХВ-Петербург, 2010.

Anderson C. W. (2011) Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism* 12(5).

Boczkowski P.J. (2004) *Digitizing the news: Innovation in online newspapers*. Cambridge, MA: MIT Press.

Boczkowski, P.J., Peer, L. (2011) The choice gap: The divergent online news preferences of jour-nalists and consumers. *Journal of Communication* 61(5).

Gans H.J. (2004) *Deciding what's news: A study of CBS Evening News, NBC Nightly News, Newsweek, and TIME.* New York, NY: Pantheon.

Garber M. (2011) *Newsbeat, Chartbeat's news-focused analytics tool, places its bets on the entrepreneurial side of news orgs.* Nieman Journalism Lab.

Hamilton J. (2004) *All the News That's Fit to Sell: How the Market Transforms Information into News.* Princeton University Press.

Kaushik A. (2009) *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity.* Sybex.

Lee A. M., Lewis S.C., Powers M. (2014) Audience clicks and news placement: A study of time-lagged influence in online journalism. *Communication Research* 41(4): 505–530.

Lowrey W. (2009) *Institutional roadblocks: Assessing journalism's response to changing audiences.* Journalism and citizenship: New agendas.

McGregor P. (2007) Tracking the online audience. Metric data start a subtle revolution. *Journalism studies* 8(2).

Napoli P.M. (2010) *Audience evolution: New technologies and the transformation of media audiences.* New York, NY: Columbia University Press.

Schultz I. (2007) The journalistic gut feeling: Journalistic doxa, news habitus and orthodox news values. *Journalism Practice* 1(2).

Shoemaker P.J. (1991) *Gatekeeping.* Newbury Park, CA: SAGE. Shoemaker P. J., Vos, T. P. (2009) *Gatekeeping theory.* New York, NY: Routledge.

Singer J.B. (2010) Quality control: Perceived effects of user-generated content on newsroom norms, values and routines. *Journalism Practice* 4(2).

Notes

http://www.rbc.ru

References

Amzin A. (2011) *Novostnaya internet-zhurnalistika.* [Online News Journalism]. Moscow: Aspekt Press Publ.

Anderson C.W. (2011) Between Creative and Quantified Audiences: Web Metrics and Changing Patterns of Newswork in Local US Newsrooms. *Journalism* 12(5).

Boczkowski P. J. (2004) *Digitizing the news: Innovation in online newspapers.* Cambridge, MA: MIT Press.

Boczkowski, P.J., Peer, L. (2011) The Choice Gap: the Divergent Online News Preferences of Journalists and Consumers. *Journal of Communication* 61(5).

Gans H. J. (2004) *Deciding What's News: a Study of CBS Evening News, NBC Nightly News, Newsweek, and TIME.* New York, NY: Pantheon.

Garber M. (2011) *Newsbeat, Chartbeat's News-Focused Analytics Tool, Places Its Bets on the Entrepreneurial Side of News Orgs.* Nieman Journalism Lab.

Hamilton J. (2004) *All the News That's Fit to Sell: How the Market Transforms Information into News.* Princeton University Press.

Kaushik A. (2009) *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity.* Sybex.

Kolesnichenko A.V. (2008) *Prakticheskaya zhurnalistika. Uchebnoe posobie.* [Practical Journalism. Study guide]. Moscow, MSU Publ.

Kroychik L.E. (2000) Sistema zhurnalistskikh zhanrov. [The System of Journalistic Genres] In S.G. Korkonosenko (ed.) *Osnovy tvorcheskoy deyatel'nosti zhurnalista* [Fundamentals of Journalists' Creative Work]. SPb.: Znanie, SPbIVESEP Publ.

Lee A. M., Lewis S. C., Powers M. (2014) Audience Clicks and News Placement: a Study of Time-Lagged Influence in Online Journalism. *Communication Research* 41(4): 505– 530.

Loseva N. (2016) Auditoriya novykh media. [New Media Audience]. In *Kak novye media izmenili zhurnalistiku*. [How New Media Have Changed Journalism] Ekaterinburg: Gumanitarnyy universitet Publ.

Lowrey W. (2009) *Institutional Roadblocks: Assessing Journalism's Response to Changing Audiences*. Journalism and Citizenship: New Agendas.

McGregor P. (2007) Tracking the Online Audience. Metric Data Start a Subtle Revolution. *Journalism studies* 8(2).

Melikhov D., Sarmatov I. (2010) *Veb-analitika: shag k sovershenstvu*. [Web Analytics: a Step towards Excellence]. K.: Analitik Intellekt Servis Publ.

Napoli P.M. (2010) *Audience Evolution: New Technologies and the Transformation of Media Audiences*. New York, NY: Columbia University Press.

Schultz I. (2007) The Journalistic Gut Feeling: Journalistic Doxa, News Habitus and Orthodox News Values. *Journalism Practice* 1(2).

Shoemaker P.J. (1991) *Gatekeeping*. Newbury Park, CA: SAGE. Shoemaker P.J., Vos, T. P. (2009) *Gatekeeping Theory*. New York, NY: Routledge.

Singer J.B. (2010) Quality Control: Perceived Effects of User-Generated Content on Newsroom Norms, Values and Routines. *Journalism Practice* 4(2).

Tertychnyy A.A. (2000) *Zhanry periodicheskoy pechati. Uchebnoe posobie*. [Genres of Print Media. Study guide]. Moscow: Aspekt Press Publ.

Ullmen Dzh. (1998) *Zhurnalistskie rassledovaniya: sovremennye metody i tekhnika*. [Journalistic Investigations: Modern Methods and Technique]. Moscow.

Yakovlev A., Dovzhikov A. (2010) *Veb-analitika: osnovy, sekrety, tryuki*. [Web Analytics: Fundamentals, Secrets, Tricks]. SPb.: BKhV-Peterburg Publ.

Поступила в редакцию
21.04.2017