

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное учреждение высшего образования

**«Уральский федеральный университет имени  
первого Президента России Б. Н. Ельцина»**

**Уральский гуманитарный институт  
Департамент «Факультет журналистики»**

**Кафедра периодической печати и сетевых изданий**

## **АЛГОРИТМИЗАЦИЯ ПРИ РАБОТЕ НАД КОНТЕНТОМ**

Допустить к защите:  
заведующий кафедрой,  
доктор философских наук,  
профессор Олешко В.Ф.



Выпускная квалификационная работа  
студентки IV курса  
Балезиной Надежды Владимировны



Нормоконтролер:  
кандидат филологических наук,  
доцент Глебович Т.А.



Научный руководитель:  
доктор филологических наук,  
доцент Олешко Е.В.



Екатеринбург  
2021

## ОГЛАВЛЕНИЕ

|  |    |
|--|----|
| <b>ВВЕДЕНИЕ</b> .....  | 4  |
| <b>ГЛАВА 1.</b> ....   | 7  |
| §1.1 Рекомендательные системы: понятие, история, развитие..... | 7  |
| §1.2 Классификация рекомендательных систем.....                | 10 |
| §1.3 Bubble filter.....  | 14 |
| §1.4 Алгоритмизация поисковых систем.....                      | 15 |
| 1.4.1 Работа поисковой системы «Яндекс».....                   | 15 |
| 1.4.2 Работа поисковой системы Google.....                     | 20 |
| § 1.5 Сбор данных в сети интернет.....                         | 23 |
| <b>ГЛАВА 2.</b> ....   | 30 |
| § 2.1 Работа сервиса «Яндекс.Дзен».....                        | 30 |
| § 2.2 Работа социальной сети «ВКонтакте».....                  | 34 |
| <b>ГЛАВА 3.</b> ....   | 38 |
| § 3.1. Эмпирическая база исследования.....                     | 38 |
| § 3.2 Общие данные.....  | 38 |
| § 3.3. Работа СМИ с платформой «Яндекс.Дзен».....              | 40 |
| § 3.4. Работа СМИ с платформой «ВКонтакте».....                | 42 |
| <b>ЗАКЛЮЧЕНИЕ</b> .....  | 46 |
| <b>БИБЛИОГРАФИЧЕСКИЙ СПИСОК</b> .....                          | 49 |

|                         |    |
|-------------------------|----|
| <b>ПРИЛОЖЕНИЕ</b> ..... | 55 |
|-------------------------|----|

## **ВВЕДЕНИЕ**

Каждый день мы заглядываем в ленту социальных сетей и мессенджеров, чтобы посмотреть новости мира и своих друзей. Где-то между фотографиями знакомых попадают новости разного уровня. Как к пользователю в ленту попадает СМИ, в какой момент и почему именно определенная новость? Может, вы на него подписаны, а может, умная лента решила, что вам будет интересно, потому что вашему другу «понравилась» эта новость.

Так что же делать медиа, чтобы чаще оказываться на глазах пользователя? Как адаптировать свои сайты и материалы, чтобы пользователи дольше на них оставались?

Актуальность работы заключается в том, что каждый из нас использует социальные сети и встречается с рекомендательными сервисами, нейросетями, сидит в своем комфортном информационном пузыре. Понимание работы таких сервисов полезно при создании и распространении контента на различных платформах.

Цель: рекомендовать пути работы «Яндекс.Дзен» и «ВКонтакте» для СМИ, определить, как попасть в ленту потенциальной аудитории.

Для достижения этой цели в дипломной работе поставлены следующие задачи:

- изучить историю рекомендательных систем;
- изучить виды рекомендательных систем и алгоритмов;
- рассмотреть работу поисковых систем;
- изучить законодательство в сфере персональных данных и его применение в сети интернет;
- исследовать работу алгоритмов рекомендаций и нейросетей на платформах «Яндекс.Дзен» и «ВКонтакте»;
- исследовать, как СМИ с этими платформами работают;

Во время изучения будем рассматривать тактики работы с текстовыми и графическими материалами СМИ на указанных платформах.

Объектом исследования являются рекомендательные сервисы и работа их алгоритмов с контентом пользователей.

Предметом исследования является анализ работы СМИ на платформах с рекомендательными системами «ВКонтакте» и «Яндекс.Дзен».

В дипломной работе широко используются такие методы исследования, как анализ литературы и нормативно-правовой документации по теме, изучение и обобщение отечественной и зарубежной практики, интервьюирование.

Практическая значимость исследования заключается в разработке рекомендаций по работе на платформах «Яндекс.Дзен» и «ВКонтакте».

В процессе написания дипломной работы нами проведен анализ статей, учебных материалов о рекомендательных сервисах и нормативно-правовых актов РФ, в частности федерального закона «О персональных данных».

Теоретическую основу дипломной работы составляют труды таких исследователей, как Pariser. E., Francesco R., Prem M., и др.

Работа состоит из введения, трех глав, заключения и списка использованной литературы.

В первой главе раскрываются теоретические основы рекомендательных систем, дается характеристика видам и особенностям каждой из них, рассматривается история их появления. Обозначаются аспекты работы поисковых алгоритмов и сбора персональных данных.

Во второй главе исследуются особенности работы сервисов «Яндекс.Дзен» и «ВКонтакте» с кликбейтом, оригинальным и дублирующимся контентом, хейтспичем и отбором материалов.

В третьей главе проведен анализ ответов сотрудников разных СМИ об их работе на вышеобозначенных платформах.

## ГЛАВА 1.

### §1.1 Рекомендательные системы: понятие, история, развитие

Определение рекомендательных систем различается:

**Рекомендательные системы** — помощь в преодолении информационной перегрузки, предоставляя персонализированные предложения, основанные на истории лайков и дизлайков пользователя.<sup>1</sup>

**Рекомендательные системы** — программные средства и методы, предоставляющие рекомендации по предметам, которые могут быть полезны пользователю.<sup>2</sup>

Главная цель рекомендательной системы состоит в том, чтобы генерировать значимые рекомендации для пользователя среди вещей, которые могут быть ему интересны. То есть рекомендательные системы можно назвать предсказателями желаний потребителей.

Использование рекомендательных систем началось примерно 30 лет назад с советов о том, что посмотреть пользователю, и долгое время развивались в видеосфере. Одним из первых устройств, которое пыталось предугадать предпочтение пользователя, оказалась кабельная приставка компании TiVo. Она использовала коллаборативную систему.

---

<sup>1</sup> Prem Melville, Raymond J. Mooney, Ramadass Nagarajan Content-boosted collaborative filtering for improved recommendations. // Eighteenth national conference on Artificial intelligence. [Электронный ресурс] URL: <https://www.cs.utexas.edu/~ml/papers/cbcf-aaai-02.pdf> (дата обращения 20.05.2020).

<sup>2</sup> Ricci L. Rokach B. Shapira P. B. Kantor. Recommender Systems Handbook. // Springer Science+Business Media. [Электронный ресурс] URL: <https://link.springer.com/book/10.1007/978-0-387-85820-3> (дата обращения 15.05.2020).

Все устройства TiVo ночью отправляли данные о последовательностях кликов (какие именно кнопки нажимал пользователь, какие программы смотрел и записывал) на центральный сервер, на основании этих данных рекомендательный алгоритм на сервере вычислял персональные рекомендации, и они затем загружались на устройства. Сервер сравнивал пользователей, которые смотрят одно и то же шоу, и, если потребитель N начинал смотреть что-то новое, это же и предлагалось потребителю M. Система работала не идеально, например, она могла подумать, что вы гей. А если переучить ее и начать показывать ей фильмы про войну, она может подумать, что вы фанат Третьего Рейха.<sup>3</sup>

Система выполняла достаточно простой подбор и быстро достигала предела анализируемых данных.<sup>4</sup>

В это же время для развития рекомендательных систем компания Netflix организует Netflix Prize. Лучшие команды разработчиков в течение трех лет добились улучшения точности рекомендаций еще на 10%, но предложенные решения требовали совершенно неадекватных операционных затрат. Коллаборативная фильтрация, как основной метод рекомендаций, встала в тупик.

Другие компании поняли, что качественная информация о рекомендуемом контенте могла бы служить серьезной основой для тонкой подстройки под интересы потребителя. Эта новая волна исследований в основном использовала **байесовскую классификацию**, которая позволяет понять и

---

<sup>3</sup> ZaslowStaff, J. If TiVo Thinks You Are Gay, Here's How to Set It Straight. // The Wall Street Journal. [Электронный ресурс]. URL: <http://web.archive.org/web/20190210040707/https://www.wsj.com/articles/SB1038261936872356908>. (дата обращения: 20.05.2020).

<sup>4</sup> Пройдл А.. ИСТОРИЯ СИСТЕМ РЕКОМЕНДАЦИЙ. // «Теле-Спутник / [Электронный ресурс] URL: <https://telesputnik.ru/materials/video-glavnoe/article/istoriya-sistem-rekomendatsiy/> (дата обращения: 20.05.2020).

оценить, почему пользователь предпочитает тот или иной контент. У этой классификации были свои пределы: чем больше данных и категорий в каталоге, тем незначительнее становятся объекты для системы.

С появлением облачных вычислений задача усложнилась. Пользователям нужна комбинация редакторских рекомендаций, фильтрации по контенту на основе персональных предпочтений, рекомендаций сообщества людей с близкими вкусами и удачные находки в этих рекомендациях. А подбор контента соответствовал сиюминутному настроению и типу используемого устройства. При этом пользователь должен понимать, что это не какой-то там Большой брат следит за ним, а он сам может повлиять на результат.

Рекомендательный движок должен давать результат сразу, как только будут получены данные о контексте, учитывать все варианты алгоритмов, пояснять рекомендации пользователю и давать ему постоянный контроль над параметрами.

Сейчас рекомендательные системы стоят на этой ступени разработки. Системы в основном подстраиваются под задачи бизнеса, относительно запроса они могут выдавать:

- Рекомендации по новостям или актуальным данным — предложение нового контента, повышающего заинтересованность пользователей. Цель состоит в том, чтобы познакомить пользователей с новым контентом, который может заинтересовать их и побудить потреблять больше контента на нашей платформе.
- Рекомендации по акциям — предложение акций, которые являются наиболее прибыльными для клиентов. Рекомендации могут быть акциями, которыми они пользовались в прошлом. Новизна здесь не имеет значения, только рентабельность акций.

- Рекомендации по продукту — предложение сочетания старых и новых продуктов. Старые продукты, известные из прошлых данных пользователей, служат напоминанием. Также важно предложить новые продукты, которые пользователи могут захотеть попробовать.

## **§1.2 Классификация рекомендательных систем**

За все время развития программисты выработали достаточное количество рекомендательных систем, но можно выделить их два основных типа, с которых все начиналось и которые продолжают модифицироваться.

### **Фильтрация по контенту/ контент-ориентированные/ фильтрация на основе содержания (Content-based)<sup>5</sup>:**

- Пользователю рекомендуются объекты, похожие на те, которые он уже просмотрел или/и использовал.
- Сходства оцениваются по признакам содержимого объектов.
- Сильная зависимость от предметной области, полезность рекомендаций ограничена.

### **Коллаборативная фильтрация (Collaborative Filtering)<sup>6</sup>:**

- Для рекомендации используется история оценок как самого пользователя, так и других пользователей.
- Более универсальный подход, часто дает лучший результат.

Делится на два основных подвида.

#### **На основе пользователя:**

---

<sup>5</sup> Content-based filtering. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/content-based-filtering/> (дата обращения: 20.05.2020).

<sup>6</sup> Collaborative filtering. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/collaborative-filtering/> (дата обращения: 20.05.2020).

Основная идея таких систем состоит в том, что если пользователи имели одинаковые интересы в прошлом, то в будущем их предпочтения также будут совпадать.<sup>7</sup>

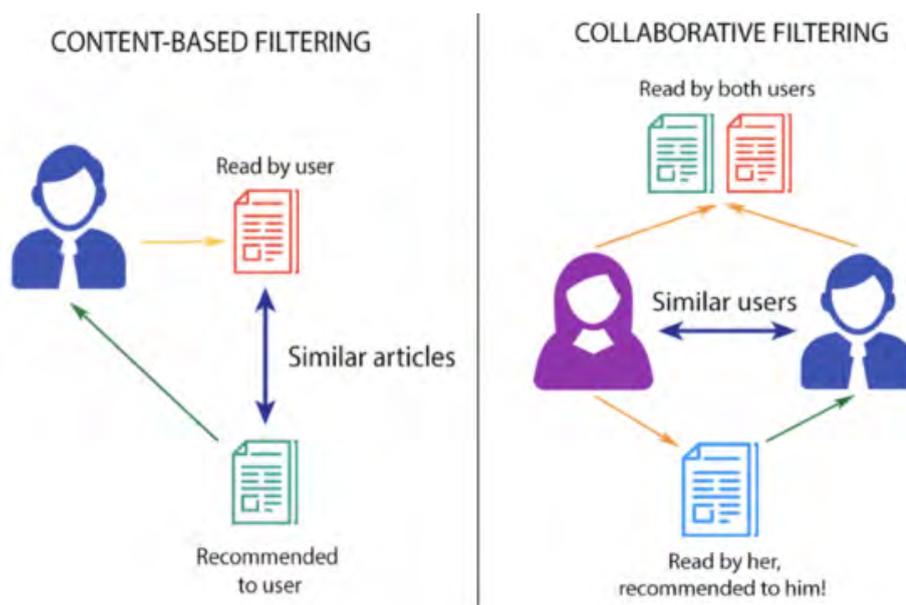
### На основе объекта:

Основная идея состоит в том, что если пользователям, которые оценили два продукта, понравились оба, то пользователям, которые попробовали только один, можно предлагать второй, который, вероятнее всего, им понравится.

Системы похожи так, как их основная цель — схожесть выбора.

Эти две системы можно объяснить в простой схеме.

Рисунок 1. Как работают рекомендательные системы<sup>8</sup>



<sup>7</sup> Кутянин, А. Р. Рекомендательные системы: обзор основных постановок и результатов. // Интеллектуальные системы / [Электронный ресурс] URL: <http://intsysjournal.org/pdfs/21-4/18-30-Kutanin.pdf> (дата обращения: 20.05.2020).

<sup>8</sup> Как работают рекомендательные системы. // Neurohive. [Электронный ресурс]. URL: <https://neurohive.io/ru/osnovy-data-science/rekomendatelnye-sistemy-modeli-i-ocenka/> (дата обращения: 26.03.2021).

Проще говоря, цель контент-ориентированных методов — создать «профиль» для каждого пользователя и каждого предмета. В коллаборативном подходе нет признаков, соответствующих пользователям или предметам, все строится на рейтинге.

Системы не идеальны, у каждой есть свои пределы и недочеты.

**Фильтрация по контенту/ контент-ориентированные/ фильтрация на основе содержания:**

- Система не сможет предложить объект, если предыдущий пользователь не представил данных для этого. Необходимо добавить дополнительные методы, чтобы дать системе возможность делать предложения за пределами того, к чему пользователь уже проявил интерес.
- Проблемы в возможности различить субъективную информацию, как мнение и юмор.

**Коллаборативная фильтрация:**

- Эта система не может предложить рекомендации новых объектов, пока нет оценок от пользователя, так как на этой системе строится предположение. Даже если пользователи начали проставлять оценки, нужно время, чтобы получить критическое количество оценок для точной рекомендации. Также неточной будет рекомендация для пользователей, которые мало проставляют оценки.
- В большинстве сфер существующее количество объектов превышает количество человек, которые могли бы это изучить. Это усложняет поиск рейтинговых объектов, на которые может опираться система в своих предложениях/предположениях.
- Группы пользователей с противоположными или пересекающимися (наслаивающимися) запросами. Даже если такая группа существует,

отдельные пользователи, кто не полностью согласен или не согласен с любой из групп, получит неточную рекомендацию.

Но что, если мы объединим эти две системы, — компенсируют ли они тогда минусы друг друга?<sup>9</sup>

Если есть данные об описании продуктов, о профиле пользователя и истории его покупок, то мы можем улучшить рекомендательную систему путем объединения методов коллаборативной фильтрации и алгоритмов фильтрации по содержанию. Таким образом, в случае появления нового пользователя в системе, о котором нет истории покупок, мы сможем использовать рекомендации на основе алгоритмов фильтрации по содержанию, и, в случае большого объема статистических данных, строить более точный прогноз, используя методы коллаборативной фильтрации.

Есть еще одна интересная система, о которой стоит упомянуть.

**Кластеризация** — рекомендательная система, в которой отсутствует большое количество первоначальных данных. То есть, очень мало исторических/размеченных данных, можно кластеризовать наблюдения на основе набора признаков, а затем назначать рекомендации для кластеров на основе меток, которые имеются у объектов в этом кластере.<sup>10</sup>

Этот метод хорошо подходит для начинающей рекомендательной системы на самом старте, чтобы набрать базу данных.

Таким образом, можно сказать, что рекомендательные сервисы пережили большие изменения, разработчики научились компенсировать минусы

---

<sup>9</sup> Hybrid Recommender Systems. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/hybrid-recommender-systems/> (дата обращения: 20.05.2020).

<sup>10</sup> Doshi, N. Recommendation Systems — Models and Evaluation. // Towards Data Science. [Электронный ресурс]. URL: <https://towardsdatascience.com/recommendation-systems-models-and-evaluation-84944a84fb8c>. (дата обращения: 20.05.2020).

одной системы, создавая новую. Смешение различных рекомендательных систем в одну дает сильную систему, у которой почти не остается слабых мест.

### §1.3 Bubble filter

Термин Bubble filter, информационный пузырь, был введен Илаем Парайзером в 2011 году на TED talk<sup>11</sup>, после чего он опубликовал книгу об этом эффекте. Он рассматривал, как часто в его ленте Facebook показываються публикации с иными политическими взглядами<sup>12</sup>. Он вывел интересную тенденцию, что от политического взгляда зависит процент фильтрации алгоритма.<sup>13</sup>

Информационный пузырь — это ваша уникальная онлайн-вселенная из информации в ленте, которая зависит от ваших интересов и действий. Алгоритмы создают вокруг вас информационный пузырь, основанный на ваших персональных данных, который закрывает от вас другие точки зрения.

Таким образом, вместо разнообразного и порой неприятного инфополя, алгоритмы загоняют пользователя в его комфортную зону, не пуская туда непривычный контент.

---

<sup>11</sup> Beware online «filter bubbles». // Ted talk [Электронный ресурс]. URL:

[https://www.ted.com/talks/eli\\_pariser\\_beware\\_online\\_filter\\_bubbles](https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles) (дата обращения: 17.02.2021).

<sup>12</sup> Pariser, E. Did Facebook’s Big New Study Kill My Filter Bubble Thesis? // Backchannel. [Электронный ресурс]. URL:

<https://medium.com/backchannel/facebook-published-a-big-new-study-on-the-filter-bubble-here-s-what-it-says-ef31a292da95>. (дата обращения: 20.05.2020).

<sup>13</sup> Pariser, E. Fun facts from the new Facebook filter bubble study. // Medium. [Электронный ресурс]. URL:

<https://medium.com/@elipariser/fun-facts-from-the-new-facebook-filter-bubble-study-444d4a658f5d>. (дата обращения 20.05.2020).

То есть, даже если пользователь подписан на новостное агентство, ему может выдаваться не весь информационный поток, а только определенные новости с либеральным взглядом.

Также такая односторонность материалов может притуплять критичность мысли и делать пользователя более податливым на манипуляции.<sup>14</sup>

## **§1.4 Алгоритмизация поисковых систем**

### **1.4.1 Работа поисковой системы «Яндекс»**

Поисковая система «Яндекс» была разработана в 1997 году. С 2005 года «Яндекс» 24 раза менял алгоритмы поисковой системы.

Поиск в интернете состоит из двух частей. Первая — специальная компьютерная система — поисковый робот — регулярно обходит интернет, выкачивает документы и обрабатывает их. Создается своего рода слепок интернета, который хранится на серверах поисковика и обновляется при каждом новом обходе. У «Яндекса» два поисковых робота — основной и быстрый (Orange).

Основной робот индексирует интернет в целом, а Orange отвечает за то, чтобы в поиске можно было найти самые свежие документы, которые появились минуты или даже секунды назад. У каждого робота есть список адресов документов, которые нужно проиндексировать.<sup>15</sup>

Вторая — пользователь задает запрос и получает ответ с серверов поисковика.

---

<sup>14</sup> Weisberg J. Is Web personalization turning us into solipsistic twits? // Slate. [Электронный ресурс]. URL: <https://slate.com/news-and-politics/2011/06/eli-pariser-s-the-filter-bubble-is-web-personalization-turning-us-into-solipsistic-twits.html>. (дата обращения: 20.05.2020).

<sup>15</sup> Индексирование интернета. // Яндекс. [интернет-портал] URL: <https://yandex.ru/company/technologies/searchindex/> (дата обращения 27.05.2021).

«Яндекс» ищет по поисковому индексу — базе данных, где для всех слов, которые есть на известных поиску сайтах, указано их местонахождение — адрес страницы и место на ней.

Каждый день «Яндекс» отвечает на десятки миллионов запросов. Около четверти из них — неповторяющиеся. Поэтому невозможно написать для поисковой системы такую программу, в которой предусмотрен каждый запрос и для каждого запроса известен лучший ответ. Поисковая система должна уметь принимать решения самостоятельно, то есть, сама выбирать из миллионов документов тот, который лучше всего отвечает пользователю. Для этого нужно научить ее обучаться.

Поисковая система должна научиться строить правило, которое определяет для каждого запроса, какая страница является хорошим ответом на него, а какая — нет. Для этого поисковая машина анализирует свойства веб-страниц и поисковых запросов.

Одна из проблем, которая возникает при машинном обучении, — переобучение. Переобучившаяся машина похожа на студента, который перезанимался, — например, прочитал очень много книжек перед экзаменом по психологии. Он мало общается с живыми людьми и пытается объяснить простые поступки слишком сложными моделями поведения.

Для построения формулы ранжирования «Яндекс» использует собственный метод машинного обучения — «Матрикснет». Он устойчив к переобучению.<sup>16</sup>

Наверху поисковой выдачи «Яндекс» показывает среди первых результатов те документы, которые содержат наиболее подходящие пользователю ответы, — наиболее релевантные заданному запросу.

---

<sup>16</sup> Ранжирование и машинное обучение. // Яндекс. [интернет-портал] URL: <https://yandex.ru/company/technologies/learning/> (дата обращения 27.05.2021).

С принципом релевантности связаны правила, которые «Яндекс» применяет при ранжировании сайтов<sup>17</sup>.

1. Страницы, которые ухудшают качество поиска, система автоматически исключает из поиска или понижает в ранжировании. Они специально созданы с целью обмануть поисковую систему. Для этого, например, на странице размещают невидимый или бессмысленный текст. Или промежуточные страницы, которые перенаправляют посетителей на сторонние сайты. Некоторые сайты умеют замещать страницу, с которой перешел пользователь, на какую-нибудь другую. То есть, когда пользователь переходит на такой сайт по ссылке из результатов поиска, а потом хочет снова вернуться к поиску и посмотреть другие результаты, он видит какой-то другой ресурс.

2. Сайты с рекламой, которая затрудняет восприятие информации, система располагает в поисковой выдаче ниже, чем сайты, на которых пользователь может найти ответ на свой вопрос без лишних проблем. Например, есть сайты, которые содержат popunder-баннеры (они перемещаются по экрану вслед за прокруткой страницы и закрывают ее содержание, а при попытке закрытия такого баннера открывается новое окно) и clickunder-рекламу (она неожиданно для пользователя открывает рекламную страницу при любом клике по сайту, в том числе — по ссылкам).

3. Понижение в поиске или исключение результатов со взрослым контентом. Дело в том, что порноресурсы — это бизнес. Необязательно в виде прямой продажи доступа к контенту. Сайты «для взрослых» часто распространяют вирусы, которые потом используются для спама и мошенничества. Получить легальную рекламу порноресурсы по понятным причинам не могут и для продвижения используют в основном серые и

---

<sup>17</sup> Принципы ранжирования поиска Яндекса. // Миссия Яндекса. [Электронный ресурс]. URL: <https://yandex.ru/company/rules/ranking/> (дата обращения 27.05.2021).

черные методы. Поэтому, в частности, такие сайты могут появляться в результатах поиска по самым разнообразным запросам.<sup>18</sup>

4. Страницы с вирусом получают специальную отметку. Если обнаружилось, что сайт заражен, в результатах поиска рядом с ним появляется предупреждающая пометка. При этом зараженные сайты не исключаются из поиска и не понижаются в результатах поиска — может быть, на таком ресурсе находится нужный пользователю ответ, и он все равно захочет туда перейти.

Основная метрика качества поиска «Яндекса» — это то, насколько пользователю пригодились найденные результаты. Иногда по запросам пользователей невозможно определить, какой ответ ему подойдет. Например, человек, задавший запрос «Пушкин», возможно, ищет информацию о поэте, а возможно — о городе. Среди первых результатов поиска, по мнению «Яндекса», должны найтись ответы для всех случаев. Поисковая система умеет определять многозначные запросы и показывает разнообразные ответы. Для этого используется технология «Спектр».

Технология «Спектр» умеет учитывать множество неявных целей пользователей и показывать соответствующие ответы. В основе работы «Спектра» лежит статистика поисковых запросов. Система исследует запросы всех пользователей «Яндекса» и выделяет в них различные объекты — это могут быть имена людей, названия фильмов и книг, модели автомобилей и тому подобное. Каждый объект относится к одной или нескольким категориям.

Кроме того, «Спектр» при поиске умеет учитывать различные потребности пользователей. У каждой категории есть список возможных

---

<sup>18</sup> Поиск Яндекса и контент для взрослых. // Миссия Яндекса. [Электронный ресурс]. URL: <https://yandex.ru/company/rules/filtration/> (дата обращения 27.05.2021).

потребностей — тех намерений, с которыми пользователи ищут тот или иной объект.

Алгоритм оценивает процент людей, которые ищут этот объект с каждой из возможных целей. Эти данные используются при ранжировании результатов поиска по многозначным запросам.<sup>19</sup>

#### **1.4.2 Работа поисковой системы Google**

Инструменты ранжирования Google упорядочивают сотни миллиардов страниц в поисковом индексе, чтобы увеличить скорость поисковой выдачи.

Эти системы ранжирования включают целый ряд алгоритмов. Чтобы дать пользователю наиболее полезную информацию, они учитывают множество факторов, включая поисковый запрос, релевантность и удобство найденных страниц, их надежность, а также местоположение и настройки. Вес каждого фактора варьируется в зависимости от характера запроса. Например, дате публикации контента придается большее значение, когда пользователь ищет актуальные новости, и меньшее, если интересуется определением из словаря.<sup>20</sup>

Обычно, когда пользователь вводит запрос, алгоритмы ищут в индексе подходящие страницы, а также определяют, как часто ключевые слова встречаются на странице и в каких ее разделах (например, в заголовке или основном тексте).

---

<sup>19</sup> Спектр. // Яндекс. [Электронный ресурс]. URL: <https://yandex.ru/company/technologies/spectrum/> (дата обращения 27.05.2021).

<sup>20</sup> Принципы работы алгоритмов Google Поиска. // Google Поиск. [Электронный ресурс]. URL: <https://www.google.com/intl/ru/search/howsearchworks/algorithms/> (дата обращения 19.05.2021).

В первую очередь, о релевантности информации говорит наличие на странице тех же ключевых слов, что и в поисковом запросе. Если они действительно есть, например, в заголовке или основном тексте, значит, информация с большей вероятностью соответствует введенному запросу. В дополнение к этому, релевантность оценивается на основе обобщенных и анонимизированных данных о том, как пользователи взаимодействуют с результатами поиска. Эти данные преобразуются в сигналы, которые позволяют поисковым алгоритмам, использующим машинное обучение, лучше справляться с такой оценкой.

### **Принципы работы алгоритмов Google-поиска**

При том огромном объеме информации, который есть в интернете, находить нужные сведения практически нереально без какого-то решения, которое помогало бы упорядочить их. Чтобы пользователи за доли секунды получали актуальные и полезные результаты, инструменты ранжирования Google упорядочивают сотни миллиардов страниц в поисковом индексе.

Эти системы ранжирования включают целый ряд алгоритмов. Чтобы выдавать наиболее полезную информацию, они учитывают множество факторов, включая поисковый запрос, релевантность и удобство найденных страниц, их надежность, а также местоположение и настройки пользователя. Вес каждого фактора варьируется в зависимости от характера вашего запроса. Например, дате публикации контента придается большее значение, когда вы ищете актуальные новости, и меньшее, если вас интересует определение из словаря.

Чтобы алгоритмы поиска отвечали самым высоким стандартам, используется процедура онлайн-тестирования, к которой привлекаются

тысячи сторонних специалистов по оценке качества результатов поиска, представляющие разные страны и прошедшие дополнительную подготовку.

### **Анализ слов и выражений**

Для этого выполняется ряд действий – от интерпретации орфографических ошибок до определения типа введенного запроса на основе результатов последних исследований в области понимания естественного языка. Google использует систему синонимов, чтобы подобрать правильное значение слову. Система разрабатывалась в течение пяти лет и увеличила качество поисковых запросов на 30%. Также, если выражение относится к популярному сейчас запросу, то будут выдавать актуальные запросы и ниже другие варианты.

### **Подбор подходящих страниц**

Обычно, когда пользователь вводит запрос, алгоритмы ищут в индексе подходящие страницы, а также определяют, как часто ключевые слова встречаются на странице и в каких ее разделах, например, в заголовке или основном тексте.

Алгоритмы не только сопоставляют ключевые слова, но и определяют, насколько полная информация содержится в предполагаемых результатах поиска. Например, пользователя, указавшего запрос «собаки», вряд ли интересует страница, в которой сотни раз повторяется это слово. Так, алгоритмы поиска определяют, представлен ли на страницах нужный контент, например, изображения собак, видео с их участием или список

пород. Пользователю в приоритете будут показаны страницы на языке запроса.

### **Ранжирование страниц**

Чтобы предоставлять наиболее актуальную информацию, эти алгоритмы оценивают сотни самых разных факторов – от новизны контента и количества повторов запроса до удобства просмотра страницы. Для выявления источников достоверных и актуальных сведений по той или иной теме отслеживаются наиболее популярные ресурсы среди пользователей, указывающих похожие запросы. Если на страницу ссылаются авторитетные в своей области сайты, это означает, что на ней, вероятно, представлен качественный контент.

### **Данные пользователя**

Результаты поиска показываются исходя из сведений о стране и местоположении пользователей. Например, если житель Чикаго вводит запрос «футбол», ему, в первую очередь, будут показаны сведения об американском футболе и клубе «Чикаго Беарз». Если же запрос «футбол» указывает пользователь из Москвы, на более высоких позициях окажутся результаты, связанные с обычным футболом и Российской футбольной премьер-лигой. Кроме того, при подборе наиболее подходящих результатов учитываются настройки поиска.

В некоторых случаях учитывается история поиска пользователей, чтобы предоставлять им персонализированные результаты. Например, если пользователь ранее вводил запрос «Ростов против Спартак», а затем указал «Ростов», вероятно, он ищет сведения не о городе, а о футбольном клубе.

Некоторые функции поиска также дают возможность персонализировать результаты на основе данных о действиях пользователя в аккаунте Google. Так, по запросу «мероприятия рядом» могут быть показаны те мероприятия, которые относятся к потенциально интересным пользователю категориям.

Поисковые системы, что были рассмотрены выше, работают по-разному, но акцент делают на ранжировании и актуальной выдаче информации. Обе системы отталкиваются от географии пользователя, его интересов и запросов, которые высказывались ранее и которые сохраняются в базе браузера (поисковой системы). Главная задача разработчиков — создать поисковый алгоритм, который будет быстро выдавать информацию, которая при этом будет персонализирована и актуальна для определенного пользователя.

## **§ 1.5 Сбор данных в сети интернет**

Во время поискового запроса, использования сайта или социальных сетей пользователь передает свои данные. Эти данные помогают алгоритмам и нейросетям точнее подбирать информацию под запрос.

Часть информации пользователь добровольно передает в руки операторов, например, в социальной сети «ВКонтакте» можно указать свои интересы, любимые книги, фильмы, шоу и так далее — по пользовательскому соглашению сайт вправе использовать эту информацию, в том числе и для таргетированной рекламы.<sup>21</sup>

---

<sup>21</sup> Правила пользования Сайтом ВКонтакте. // ВКонтакте. [интернет-портал] URL: <https://vk.com/terms> (дата обращения 21.05.2021).

Самый распространенный способ для сбора данных пользователя — cookie-файлы.

**Файл cookie** — это небольшой фрагмент текста, передаваемый в браузер с сайта, который вы посещаете. С его помощью сайт запоминает информацию о ваших посещениях и с каждым разом становится удобнее и полезнее для вас.

Некоторые файлы cookie используются для сохранения настроек пользователя. Например, у большинства пользователей в браузере хранится файл cookie NID. В нем содержится уникальный идентификатор, который позволяет получать данные о настройках и предпочтениях пользователя таких, как язык, параметры безопасного поиска Google и желаемое количество результатов на странице. Срок действия всех файлов NID истекает через шесть месяцев после последнего использования.<sup>22</sup>

Файлы cookie можно поделить на несколько типов в зависимости от цели<sup>23</sup>:

#### **Длительность хранения**

- Сеансовые — эти файлы cookie являются временными и истекают после закрытия браузера.
- Постоянные — это файлы cookie, которые остаются на жестком диске до тех пор, пока пользователь или браузер не удалит их. Все постоянные файлы cookie имеют дату истечения срока действия, записанную в их коде, но их продолжительность может варьироваться. Согласно директиве ePrivacy, они не должны храниться более 12 месяцев,

---

<sup>22</sup> Как Google использует файлы cookie. // Google Политика конфиденциальности и условия пользования. [Электронный ресурс]. URL: <https://policies.google.com/technologies/cookies> (дата обращения 25.04.2021).

<sup>23</sup> Cookies, the GDPR, and the ePrivacy Directive. // GDPR.eu. [интернет-портал]. URL: <https://gdpr.eu/cookies/> (дата обращения 26.05.2021).

но на практике они могут оставаться на устройстве гораздо дольше, если вы не предпримете никаких действий.

### **Происхождение**

- Основные — размещаются на устройстве непосредственно сайтом, который вы посещаете.
- Сторонние — размещаются на устройстве третьей стороной такой, как рекламодатель или аналитические системы.

### **Цель**

- Необходимые — необходимы для просмотра сайта и использования его функций таких, как доступ к защищенным областям сайта. Эти файлы cookie, как правило, являются основными.
- Предпочтения или «функциональные файлы cookie» — позволяют сайту запоминать выбор, который пользователь сделал в прошлом, например, какой язык предпочитает, для какого региона хочет получать сводки погоды, имя пользователя и пароль для автоматического входа в систему.
- Статистические или «файлы cookie производительности» — собирают информацию о том, как используется сайт, например, какие страницы посещал пользователь и по каким ссылкам перешел. Ни одно из этих сведений не может быть использовано для идентификации. Все это агрегировано и, следовательно, анонимизированно. Их единственная цель — улучшить функции сайта.
- Маркетинговые файлы cookie — эти файлы cookie отслеживают онлайн-активность, чтобы помочь рекламодателям предоставлять более релевантную рекламу или ограничивать количество просмотров рекламы. Данные файлы cookie могут передавать эту информацию другим

организациям или рекламодателям. Это постоянные файлы cookie и почти всегда стороннего происхождения.

Браузеры и сайты предоставляют пользователям определенный контроль над файлами cookie: разрешают блокировать все файлы cookie или только те файлы, которые были переданы третьим лицам; выборочно включать или отключать файлы cookie для каждого сайта или разрешать использование для конкретных сайтов.<sup>24</sup>

Многие из крупнейших и наиболее посещаемых сайтов используют файлы cookie по умолчанию, так, файлы cookie практически неизбежны. Отключение файлов cookie приведет к блокировке доступа пользователя, например, на таких сайтах, как Youtube, Gmail, Yahoo mail.<sup>25</sup>

Европейский союз использует GDPR, который устанавливает, что пользователи могут быть опознаны с помощью онлайн-идентификаторов своих устройств, приложений и протоколов, например, через IP-адрес или идентификационные файлы, хранящиеся в системе посетителя. Эти данные могут оставлять цифровые следы пользователя, которые в сочетании с уникальными идентификаторами и другой информацией могут использоваться для создания профилей людей и их идентификации. Таким образом, с точки зрения GDPR, информация, полученная с помощью файлов cookie, может быть признана персональной, но только в

---

<sup>24</sup> Плотников А. В. Конфиденциальность персональных данных о поведении онлайн-пользователей. // Образование и право. – 2019. – №1. [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/konfidentsialnost-personalnyh-dannyh-o-povedenii-onlayn-polzovateley> (дата обращения 21.05.2021).

<sup>25</sup> Key tips for safe and responsible cookie-based Web browsing. // All about Cookies. [интернет-портал]. URL: <https://www.allaboutcookies.org/> (дата обращения 26.05.2021).

совокупности с уникальным ключом, которая позволяет идентифицировать конкретного человека.<sup>26</sup>

Европейские домены eu должны проходить специальную комиссию Cookie Consent Kit, для прохождения нужно, чтобы сайт предоставлял:<sup>27</sup>

- JavaScript для автоматического отображения заголовка баннера на 24 языках;
- Мастер объявления файлов cookie и ссылку на страницу уведомления о файлах cookie;
- API JavaScript с методами и функциями, которые помогают предотвратить предварительное хранение файлов cookie;
- Корпоративное согласие для запоминания выбора пользователя на сайтах;
- Шаблон для страницы уведомления о файлах cookie.

Российское законодательство дает широкое определение личным данным.

**Персональные данные** — любая информация, относящаяся к прямо или косвенно определенному или определяемому физическому лицу (субъекту персональных данных).<sup>28</sup>

Получается, что при наличии базы cookie-файлов и идентификационных ключей, можно опознать определенного человека. Так, если на каком-то сервере пользователь оставлял свой телефон и дал разрешение на сбор cookie-файлов, имея эти данные, методом сопоставления их базы с номерами телефонов и базой cookie другого сайта можно найти номер

---

<sup>26</sup> General Data Protection Regulation (GDPR). // GDPR.eu. [интернет-портал]. URL: <https://gdpr.eu/tag/gdpr/> (дата обращения 26.05.2021).

<sup>27</sup> The Europa Web Guide. // European Commission. [интернет-портал] URL: <https://wikis.ec.europa.eu/display/WEBGUIDE/04.+Cookies> (дата обращения 26.05.2021).

<sup>28</sup> Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 30.12.2020) «О персональных данных» / [Электронный ресурс] URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61801/4f41fe599ce341751e4e34dc50a4b676674c1416/](http://www.consultant.ru/document/cons_doc_LAW_61801/4f41fe599ce341751e4e34dc50a4b676674c1416/) (дата обращения 20.04.2021)

пользователя, который сейчас посещает сайт, позвонить и продать что-либо.

Сначала мы обозначим, как законодательство определяет:

**Обработка персональных данных** — любое действие (операция) или совокупность действий (операций), совершаемых с использованием средств автоматизации или без использования таких средств с персональными данными, включая сбор, запись, систематизацию, накопление, хранение, уточнение (обновление, изменение), извлечение, использование, передачу (распространение, предоставление, доступ), обезличивание, блокирование, удаление, уничтожение персональных данных.

Но cookie-файлы собираются не людьми, а автоматически, получается, что они автоматизированы.

**Автоматизированная обработка персональных данных** — обработка персональных данных с помощью средств вычислительной техники;

**Предоставление персональных данных** — действия, направленные на раскрытие персональных данных определенному лицу или определенному кругу лиц;

**Информационная система персональных данных** — совокупность содержащихся в базах данных персональных данных и обеспечивающих их обработку информационных технологий и технических средств;

Также в законе уточняется, что обработка персональных данных должна ограничиваться достижением конкретных целей, не допускается обработка персональных данных, несовместимая с целями сбора персональных данных.

Исследование целей обработки личных данных, указанных в пользовательских соглашениях, позволяет сделать вывод о том, что в большинстве случаев ими являются предоставление прав использования дополнительных программных продуктов компании и проведение маркетинговых кампаний. Таким образом, в целях реализации норм международного законодательства, конституционного принципа, гарантирующего любому право на неприкосновенность частной жизни, личную и семейную тайну, операторы должны получать согласие пользователей на обработку персональных данных, содержащихся в учетной записи, а в случаях, если обработка имеет своей целью продвижение продуктов и услуг с возможностью прямого контакта с субъектом таких данных, получать предварительное согласие пользователя.

Кроме того, в случае необходимости обработки оператором информации, хранящейся на окончательном оборудовании пользователя, последний должен быть уведомлен в доступной форме об объемах такой обработки и видах данных, которые могут быть получены оператором в случае предоставления им согласия. При этом, пользовательское соглашение должно включать информацию о сроках, по истечению которых предоставленные пользователем сведения будут уничтожены.

## ГЛАВА 2.

### § 2.1 Работа сервиса «Яндекс.Дзен»

«Яндекс.Дзен» — это платформа, которая подбирает контент специально для пользователя. В «Дзене» есть статьи и видео на разные темы от блогеров и медиа.<sup>29</sup>

«Дзен» начинался с агрегации сайтов. Сейчас 80% времени, проведенного на платформе, приходится на публикации блогеров. Блогерский контент изначально состоял исключительно из статей, но в этом году видео в «Дзене» уверенно догоняет тексты. Пользователи проводят на платформе по 45 минут в день.<sup>30</sup>

«Яндекс.Дзен» использует алгоритм под названием «Альфа-центавра». Технология, основанная на методе поиска «ближайших соседей» (k-NN), позволяет уже на этапе первичного отбора работать с содержанием каждой публикации. Она умеет очень быстро сравнивать интересы пользователя и содержание материала, находя ближайшие по содержанию публикации среди всего множества материалов «Дзена». А также она может анализировать намного больше публикаций за раз, значительно расширяя «воронку отбора».<sup>31</sup>

- Первичный отбор — более 1 млн.

---

<sup>29</sup> Ваш личный Дзен. // Яндекс Дзен. [Электронный ресурс]. URL: <https://zen.yandex/about> (дата обращения: 15.03.2021).

<sup>30</sup> Дзен в 2020 году: рост видео, запуск стримов и другие события и планы. // Журнал Яндекс.Дзена. [интернет-портал]. URL: <https://zen.yandex.ru/media/zenmag/dzen-v-2020-godu-rost-video-zapusk-strimov-i-drugie-sobytiia-i-plany-5fd37d730b82510af5def547> (дата обращения: 26.03.2021).

<sup>31</sup> Альфа Центавра – новый алгоритм Яндекс.Дзена. // Журнал Яндекс.Дзена. [интернет-портал]. URL: <https://zen.yandex.ru/media/zenmag/alfa-centavra--novyi-algoritm-iandeksdzena-5c8bb4a817bca800b1877ca5> (дата обращения: 26.02.2021).

- Ранжирование — до 10 тысяч.
- Лента пользователя — 30 публикаций.

На этапе ранжирования тех публикаций, которые прошли первичный отбор, алгоритм в качестве одного из параметров учитывает реакцию пользователей. Система рекомендует статьи более точно: она учитывает оценки не только от конкретного пользователя «Дзена», но и от тех, кто на него похож. Используется **алгоритм матричных разложений (matrix factorization)**, который обычно предназначен для рекомендаций «долго живущего» контента — например, музыки или фильмов.

Эта технология похожа на коллаборативную фильтрацию — сбор оценок пользователей и их сравнение.

Уникальность алгоритма «Дзена» заключается в системе «Проксима» — она умеет отличать быстро теряющий актуальность контент, то есть, новости.

«Проксима» дает возможность набрать аудиторию публикациям, независимо от того, когда они были опубликованы. Она снижает влияние новостей на ленту пользователя и быстрее «восстанавливает в правах» статьи, которые попали в «Дзен» во время информационной волны. Технология **continuous exploration**<sup>32</sup> увеличивает время, за которое алгоритм собирает реакции на опубликованный контент. Поэтому любой материал получает шанс снова попасть в ленты пользователей.

Как пользователь влияет на свои рекомендации:

---

<sup>32</sup> Continuous Exploration. // Scaled Agile, Inc. [Электронный ресурс]. URL: <https://www.scaledagileframework.com/continuous-exploration/> (дата обращения: 29.03.2021).

- **Оценка материала.** После прочтения статьи можно пройти специальный опрос и оценить, понравилась ли публикация, соответствует ли она интересам и готов ли пользователь рекомендовать ее своим друзьям.
- **Подписка.** Эксперименты показали, что пользователи оценивают материалы блогеров, на которых они подписаны, в пять раз выше и готовы чаще рекомендовать публикации из подписок своим друзьям и близким.
- **Реакция.** Читатели напрямую влияют на рекомендации: если публикация не нравится пользователям, она меньше рекомендуется широкой аудитории. При этом негативные реакции не влияют на показ статьи для подписчиков, они в любом случае увидят материал в своей ленте.

Реакции — не единственный критерий статей, также они дополняют другие сигналы: переходы из ленты, время, проведенное в статье, и другие показатели.

Алгоритм анализирует статистику канала по пяти показателям<sup>33</sup>:

- **Оригинальность.** Тематическая направленность публикаций. При расчете учитывается не только популярность темы, но и качество ее проработки. Уникальность текста и авторская подача материала.
- **Качество.** Позитивная реакция аудитории. При подсчете этого показателя учитываются подписчики канала и постоянные пользователи.
- **Вовлеченность аудитории.** Процент полного прочтения публикации. Среднее время, которое пользователи тратят на чтение. Количество комментариев, выставленных публикации.
- **Развитие канала.** Рост числа новых подписчиков. Регулярность появления новых публикаций на канале.

---

<sup>33</sup> Карма. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/channel/karma.html> (дата обращения: 17.03.2021).

- **Влиятельность.** Прямые переходы на канал из других источников. Количество материалов, которыми пользователи делятся друг с другом. Популярность блогера за пределами «Дзена».

Алгоритмам нужно время, чтобы определить тематику канала и накопить необходимую статистику. Как правило, на это требуется пара недель.

Эти показатели доступны для автора канала, относительно них можно делать выводы о популярности материалов.

Также алгоритмы «Дзена» отслеживают повторный контент и запрещают его. Дублирующий контент — это:<sup>34</sup>

- Повторно опубликованные на одном канале.
- Удаленные из «Дзена» и опубликованные заново.
- По содержанию полностью или почти полностью совпадающие с публикациями, уже размещенными в «Дзене».
- Опубликованные на нескольких каналах, даже если все эти каналы принадлежат одному блогеру.

Под надзором алгоритма при создании контента находятся:

- **Материалы с заимствованным контентом** — представляют для пользователей меньшую ценность, чем оригинальные авторские публикации.
- **Кликбейт в «Дзене»** — это оформление карточки, которое побуждает открыть публикацию, но обманывает ожидания пользователя.<sup>35</sup>

---

<sup>34</sup> Требования к контенту. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/requirements/rules.html> (дата обращения: 26.03.2021).

<sup>35</sup> Кликбейт. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/requirements/clickbait.html> (дата обращения: 17.03.2021).

- **Ложная информация** — попытки выдать себя за третье лицо, информационные мистификации, некоторые виды кликбейтных заголовков, публикации с недостоверными фактами и т.д.
- **Фейки** — это разновидность ложной информации. Фейком в «Дзене» называется публикация, которая основана на недостоверных фактах.
- **Частичный контент** — это публикация с неполным содержанием или ограниченным доступом. Чтобы дочитать ее, пользователь должен перейти по ссылке на сторонний сайт, другую публикацию в «Дзене», видеоролик и т. д. Каждая публикация должна быть полноценной и самостоятельной. Допустимо ссылаться на первоисточники, дополнительные справочные материалы, публикации других каналов. Главное условие — пользователь не должен покидать материал, чтобы дочитать его.

Материалы, в которых замечены эти нарушения показываются только подписчикам каналов. То есть, в ленту рекомендаций неподписанного пользователя контент с таким содержанием не выйдет.

Подводя итог, можно сказать: алгоритм «Альфа-центавра» построен на анализе публикуемого контента и выявления в нем уникальности, а также на постройке профиля пользователя. Чтобы попадать в интересы читателя, нужно публиковать оригинальный и нишевый контент. Нишевый контент попадает в конкретные интересы аудитории, потребителей такого контента может быть много.

## § 2.2 Работа социальной сети «ВКонтакте»

«ВКонтакте» перешли с хронологической ленты на умную. Тогда это вызвало большой ажиотаж среди пользователей. Со временем такие ленты доказали свою эффективность.

### **Алгоритм тематических лент**

В тематические ленты — разработку 19 года — отбирают материалы алгоритмы и люди. Участники открытой редакции, эксперты сферы голосуют за или против публикации. Чем больше положительных откликов, тем выше публикация поднимется в рекомендациях.<sup>36</sup>

Благодаря алгоритму «ВКонтакте» увеличил просматриваемость записей на 30% только за первый год его работы.<sup>37</sup>

### **Алгоритм «Прометей»**

Он создан для поиска и поддержки интересных авторов и сообществ.

Искусственный интеллект постоянно находит создателей уникального контента и следит за их достижениями. За публикации, отмеченные особым вниманием аудитории, автор или его сообщество получают метку «огня» и повышенные охваты. В мобильных приложениях сервис рекомендаций покажет записи талантливого автора людям, которые еще не знакомы с его творчеством, но оно может быть им интересно.<sup>38</sup>

### **Алгоритм «Немезида»**

Специальный алгоритм защиты уникального контента.<sup>39</sup>

---

<sup>36</sup> Тематические ленты: лучшее про Ваши увлечения. // Блог ВКонтакте. [Социальная сеть]. URL: <https://vk.com/blog/themefeeds> (дата обращения: 20.03.2021).

<sup>37</sup> Новая эпоха авторов. // Блог ВКонтакте. [Социальная сеть]. URL: <https://vk.com/blog/creators> (дата обращения: 20.03.2021).

<sup>38</sup> Что такое Прометей? // ВКонтакте с авторами. [Социальная сеть]. URL: <https://vk.com/@authors-prometheus> (дата обращения: 20.03.2021).

<sup>39</sup> Под защитой Немезиды. // Блог ВКонтакте. [Социальная сеть]. URL: <https://vk.com/blog/nemesis> (дата обращения: 26.03.2021).

Искусственный интеллект ищет повторяющийся контент и отправляет на сверку модераторам. Сообщества, замеченные за нарушением правил «Немезиды», получают предупреждение, в серьезных случаях — отключение монетизации.

Какой вывод сделали сотрудники «ВКонтакте» за год работы алгоритма:<sup>40</sup>

- **Оригинальность привлекает внимание.** Благодаря тому, что снизился охват скопированных публикаций, уникальный контент получил заметный приоритет в новостной ленте. Просмотров уникальных записей стало больше на 40%, а отметок «Нравится» — на 35%. Количество копий удалось снизить почти на 49%.

«ВКонтакте» в 2020 году ввел ограничение на хейтспитч. Модерация проходит через алгоритмы автоматического поиска и нейронные сети. Также пользователи могут пожаловаться на контент по причине «Враждебные высказывания».

Хейтспитч — пропаганда идеи о том, что представители какой-то социальной группы заслуживают, чтобы над ними издевались, их изолировали, избивали, уничтожали или сравнивали с грязью.

За два квартала 2020 года заблокировано 1320 личных страниц, 2470 сообществ и 520 500 постов.

«Экспериментальная функция для борьбы с оскорбительными комментариями позволила сократить их количество на 36% и дополнительно исследовать употребление языка вражды. На базе этих знаний и анализа жалоб пользователей мы обучаем новую нейросеть для автоматической разметки материалов и комментариев, которые содержат

---

<sup>40</sup> Новости Немезиды. // ВКонтакте для сообществ. [Социальная сеть]. URL: <https://vk.com/@adminsclub-nemesis-results> (дата обращения: 20.03.2021).

риторику вражды и ненависти, а также материалов, которые провоцируют комментаторов использовать такую риторику». <sup>41</sup>

Мы рассмотрели две платформы, которые используют алгоритмы для фильтрации контента.

Что схожего у алгоритмов:

- Фильтрация по дублирующему контенту;
- Фильтрация по «языку вражды».
- Внимание на полное прочтение и охват;
- Внимание на вовлеченность аудитории;
- Учитывается оригинальность и уникальность контента.

Также есть ряд отличий. У «ВКонтакте» есть алгоритм поощрения и тематическая разбивка с полуавтоматической системой отбора материалов.

Главное в алгоритмах — сформировать достаточно интересную ленту, чтобы задержать читателя на платформе. Значит, контент должен быть интересным для читателя, тогда алгоритмы будут повышать контент, чаще показывать его в ленте и выдавать его потенциальной аудитории, отталкиваясь от профиля, и той аудитории, которая уже наработана.

Под «оригинальностью» чаще всего понимается неповторяющийся контент или тот, который имеет первенство публикации на определенной платформе.

---

<sup>41</sup> Нормы платформы. // Центр безопасности ВКонтакте. [Социальная сеть]. URL: <https://vk.com/safety?section=standards> (дата обращения: 26/03/2020).

## **ГЛАВА 3.**

### **§ 3.1. Эмпирическая база исследования**

Для изучения работы алгоритмов рекомендательных систем мы провели опрос среди сотрудников нескольких СМИ по поводу их работы над контентом. Опрос включал в себя три блока вопросов:

- Общие вопросы;
- Вопросы о работе с «Яндекс.Дзен» — охваты, популярные темы, целевая аудитория;
- Вопросы о работе на платформе «ВКонтакте» — охваты, популярные темы, целевая аудитория.

Опрос прошли сотрудники 11 СМИ разного уровня: федерального, регионального и местного. Направления и тематика работы у всех разная, что позволило изучить то, как алгоритмы работают с различным контентом.

### **§ 3.2 Общие данные**

Большинство СМИ, участвовавших в опросе, используют для распространения контента «Инстаграм», «ВКонтакте» и «Фейсбук».

Большинство, 81% респондентов, обозначает главной целью использования социальных платформ — вовлечение аудитории и привлечение новой аудитории. Затем, 63%, — коммуникация с аудиторией. Самой неактуальной целью является направление трафика на сайт, можно сказать, что социальные сети становятся важным каналом распространения контента, который создает СМИ.

То есть, главная цель использования разных каналов распространения информации — это различного рода взаимодействие с аудиторией платформы.

Все, кто прошел опрос, корректируют контент для разных платформ.

- «Фейсбук» используется почти только для экономических текстов. Во «ВКонтакте» может появиться расширенная версия текста с комментариями, которые не вошли в печать. Для «Одноклассников» и «ВКонтакте» иногда вставляется картинка в баннер со ссылкой. Если бы на «Фейсбуке» можно было бы вставить картинку в баннер, это делалось бы и там.
- «Для каждой соцсети — свой подход. Начиная с вординга и подбора иллюстраций и заканчивая подбором и форматом контента.»
- «Мы пишем отдельно пост для «Инстаграма», иногда делаем картинки с цитатами, чтобы привлечь внимание аудитории.»
- «Для «Телеграма» пишем полный текст, а не только подводку, чтобы было удобнее читать.»
- «Меняем текст и картинки для «Инстаграма». Стараемся уложить всю новость в пост для «Фейсбука», т. к. пользователи не любят переходить по ссылке.»

Как мы видим, индивидуальная работа с платформой связана с особенностями потребления контента на ней и аудиторией. Также важно понимать, насколько это действенно.

- «Показатели растут только в случае, если контент подобран и упакован в соответствии с предпочтениями аудитории на конкретной платформе и в максимально органичном (комфортном) для пользователя формате.»

- «Все зависит от темы. Что-то интересное или важное будут лайкать и без картинки. Если в начале поста указываем, что текст есть только в соцсетях (или какой-то комментарий в нем), то интерес тоже к нему больше.»

Разные платформы заточены под разный вид контента. Так, для СМИ с большими текстами не очень подходят платформы, где ленты с информацией листаются («Инстаграм», «ВКонтакте», Tik-Tok), больше подходят те, где пользователь целенаправленно садится почитать текст («Телеграм», «Яндекс.Дзен»).

«Все-таки соцсети рассчитаны на то, чтобы их листать, а мы пишем такое, что нужно читать подолгу».

Поэтому важно подстраивать контент под способ потребления на платформе, также у разных платформ разные запросы для контента, например, количество знаков, формат медиаматериалов.

«Адаптируем и сокращаем текст для «Инстаграма» и «Телеграма», отдельно работаем над картинками для этих же соцсетей».

Когда формат СМИ не совпадает с форматом социальной сети, может и не совпасть целевая аудитория, что отражается на показателях: охватах, просмотрах, взаимодействиях.

### **§ 3.3. Работа СМИ с платформой «Яндекс.Дзен»**

По данным на февраль 2021<sup>42</sup> года, месячная аудитория «Яндекс.Дзен» — 65 млн, среднее время посещения — 45 минут в день и 21 млн

---

<sup>42</sup> Новый медиакит Яндекс.Дзен. // Яндекс.Дзен для маркетологов. [Электронный ресурс]. URL: <https://zen.yandex.ru/media/zenexperts/novyi-mediakit-iandeksdzena-5e4289de24a0e07212071b81> (дата обращения 4.06.2021).

пользователей, которые открывают «Яндекс.Дзен» ежедневно. Также 27% пользователей читают контент, 24% — просматривают ленту, в 23% — смотрят видеоконтент. То есть, ровно половина людей точно просматривает и взаимодействует с контентом на платформе. Половина ежедневной аудитории — это более 1,5 млн пользователей. Так, месячный охват «Делового квартала» составляет от 3 до 20 тысяч, у «Lenta.ru» многомиллионные показатели.

Сотрудники СМИ выделяют, что большой охват на этой платформе получают нишевые или узко тематические материалы, однако большим спросом пользуется и тема лайфхаков, которую сложно назвать узкой или нишевой, так как материалы на эту тему находятся в каждой сфере.

Как мы узнали выше, алгоритм выбирает материалы в воронку отбора по тематике, что позволяет более нишевым материалам чаще попадать в поле зрения своей точечной аудитории.

Стоит отметить, что большой отклик получают материалы с эмоциональной составляющей: «Материалы, которые напрямую связаны с жизнью читателей, и те, что вызывают максимальный эмоциональный отклик».

Однако за счет того, что «Яндекс.Дзен» построен на рекомендациях, охватывается и новая аудитория. Так, у некоторых СМИ аудитория с целевой совпадает на 30%. Такое неточное совпадение аудитории позволяет расширить границы и повысить узнаваемость бренда.

«Яндекс.Дзен» — удобная платформа для СМИ. Так, материалы «Lenta.ru», «Медуза» (издание признано иностранным агентом в Российской Федерации) и «Деловой квартал» размещены в виде ссылок. У

платформы есть формат, когда они могут выкладывать материалы, которые при клике на него ведут на сайт. Со стороны пользователя такой материал СМИ оформлен так же, как и материалы, размещенные непосредственно на «Дзене»: заголовок, лид и фотография или видео. Что позволяет СМИ получать просмотры наравне с блогерами. Таким образом, СМИ не перекладывает материалы с сайта, не затрачивает ресурсы на его оформление, что делает платформу привлекательной для медиа.

Вышеперечисленные факты позволяют нам сделать вывод о том, что «Яндекс.Дзен» создает удобные условия для работы СМИ на платформе и позволяет выходить на новую аудиторию благодаря рекомендательному алгоритму, на котором строится лента пользователя. Это позволяет получать охваты до 21 млн ежедневно.

### **§ 3.4. Работа СМИ с платформой «ВКонтакте»**

Меньше всего данных мы получили о том, как СМИ выстраивают свою работу во «ВКонтакте». По данным Mediascope<sup>43</sup>, среднее время просмотра ленты с телефонов и компьютеров — 31 минута в день, и 22 879 тыс. человек заходят на платформу хотя бы раз в день.

Среднемесячный охват аудитории у опрашиваемых СМИ больше 5-10 тыс. «ВКонтакте» в охвате считает просмотры поста в новостной ленте и на стене сообщества, не засчитывается за охват быстрое проматывание в ленте (пользователь должен задержать внимание на тексте или изображении).

---

<sup>43</sup> Топ ресурсов: Vk.com. \ Mediascope. [Электронные ресурс]. URL: <https://webindex.mediascope.net/report?id=16571> (дата обращения 8.06.2021).

По данным опроса, у разных СМИ разные темы набирают популярность. У «Медузы» — политика, у «Российской газеты» — фоторепортажи, у других — конкурсы и опросы.

Мы дополнительно изучили паблики СМИ, которые прошли опрос, во «ВКонтакте», чтобы понять, какие темы у них набирают большую популярность.

«Уралнаш» — независимое издание о Екатеринбурге. Это уникальные и интересные статьи, факты и истории. Это необычные фотографии и авторские наблюдения о городе. Статьи полезны и жителям, и гостям города.

Большой охват получают материалы с локальной идентичностью, видео- и фотоматериалы про город и область, а также посты, которые активно вовлекают аудиторию, — опросы. Половину от среднего охвата или меньше получают публикации со ссылками, так как «ВКонтакте» занижает в ленте публикации, которые ведут на сторонние сайты. Все это связано с тем, чтобы удержать аудиторию на своем ресурсе.

У филиала «Российской газеты» в Свердловской области паблик с маленькой аудиторией, большинств постов — это ссылки на новости с заголовком, картинка автоматически подтягивается из ссылки. Среди всех постов охват получают оригинальные (не просто ссылки) посты с текстом и качественными фоторепортажами. Также внимание привлекают технологии, происшествия и их последствия и День Победы.

Если рассматривать главный паблик «Российской газеты», то большую часть просмотров получают материалы по темам «Происшествия», «Законодательство», «Спортивные достижения» и «Ковид». Посты похожи

на региональное отделение: заголовок, ссылка. Однако материалы с развернутым лидом получают охват выше.

Можно заметить, что региональные интересы и интересы по России отличаются.

С другой стороны, оппозиционная «Медуза» собирает больше просмотров на тематику запретов, протестов, происшествий. Отдельно стоит выделить нелепые или смешные новости, а также картинки. Пост строится по типу: заголовок, обязательная надпись для иноагентов, лид и ссылка.

«Lenta.ru» собирает просмотры на новостях об Украине и непосредственно из нее (как с положительной, так и отрицательной окраской), на странных и комичных новостях («Итальянский художник продал невидимую скульптуру за 1,3 миллиона рублей»), на смешной подаче («Эту новость не видят жители Камчатки. А все потому, что из-за обрыва кабеля «Ростелекома» полуостров остался без интернета») и на новостях о детях. Все посты уникальные, то есть, это заголовок, лид и оригинальная картинка или мем, никаких ссылок на сайт.

Мы посчитали соотношение просмотров к количеству подписчиков. У «Lenta.ru» и «Медузы» просматривают контент 3-4% пользователей, в то время, как у «Российской газеты» — всего 1%. Связываем мы это с оформлением и подходом к материалам, так как частота и количество публикаций у этих пабликов близка друг к другу.

Площадка заинтересована в том, чтобы пользователь потреблял контент на платформе, поэтому больше просмотров набирают материалы,

сделанные по правилам площадки без ссылок на сайт медиа и с интерактивом (вопросом, который обсуждают в комментариях, опросом).

Но зачастую у маленьких или региональных СМИ нет ресурсов на отдельного человека, который адаптировал бы материалы под платформу.

## ЗАКЛЮЧЕНИЕ

В дипломной работе было рассмотрено функционирование рекомендательных систем, поисковых алгоритмов, платформ «Яндекс.Дзен» и «ВКонтакте», а также работа СМИ на этих площадках.

За свою историю рекомендательные сервисы пережили большие изменения, разработчики научились компенсировать минусы одной системы, создавая новую. Смешение различных рекомендательных систем в одну дает сильную систему, у которой почти не остается слабых мест.

Но идеально подобранный алгоритм создает bubble filter. Так, вместо разнообразного и порой неприятного инфополя, алгоритмы загоняют пользователя в его комфортную зону, не пропуская туда непривычный, неприятный контент, что ограничивает аудиторию для СМИ — сложно выйти за зону комфортной аудитории.

Поисковые системы, что были рассмотрены, работают по-разному, но акцент делают на ранжировании и актуальной выдаче информации. Обе системы отталкиваются от географии пользователя, его интересов и запросов, которые высказывались ранее и которые сохраняются в базе браузера (поисковой системы). Главная задача разработчиков — создать поисковый алгоритм, который будет быстро выдавать информацию, которая при этом будет персонализирована и актуальна для данного пользователя.

Однако для грамотной работы всех рекомендательных и поисковых систем, оператору нужно собирать данные пользователя, которые считаются персональными. Исследование целей обработки личных данных

позволяет сделать вывод о том, что в большинстве случаев ими являются предоставление прав использования дополнительных программных продуктов компании и проведение маркетинговых кампаний. Таким образом, операторы должны получать согласие пользователей на обработку персональных данных, содержащихся в учетной записи, а в случаях, если обработка имеет своей целью продвижение продуктов и услуг с возможностью прямого контакта с субъектом таких данных, получать предварительное согласие пользователя.

Важной задачей был анализ работ двух платформ — «Яндекс.Дзен» и «ВКонтакте», которые используют алгоритмы для фильтрации контента.

У алгоритмов есть сходства:

- Фильтрация по дублирующему контенту;
- Фильтрация по «языку вражды».
- Внимание на полное прочитывание текста и охват;
- Внимание на вовлеченность аудитории;
- Учитывается оригинальность и уникальность контента.

Также есть ряд отличий. У «ВКонтакте» есть алгоритм поощрения и тематическая разбивка с полуавтоматической системой отбора материалов.

Главное в алгоритмах — сформировать достаточно интересную ленту, чтобы задержать читателя на платформе. Значит, контент должен быть интересным для читателя, тогда алгоритмы будут повышать контент, чаще показывать его в ленте и выдавать его потенциальной аудитории, отталкиваясь от профиля той аудитории, которая уже наработана.

Из проведенного исследования можно подчеркнуть, что главная цель использования разных каналов распространения информации — это различного рода взаимодействие с аудиторией платформы.

«Яндекс.Дзен» создает удобные условия для работы СМИ на платформе: перенаправляет поток на сайт и позволяет не создавать дублирующийся контент; позволяет выходить на новую аудиторию благодаря алгоритму, на котором строится лента пользователя, и получать охваты до 21 млн ежедневно.

В то время, как «ВКонтакте» заинтересован в том, чтобы удерживать аудиторию на своей платформе. Тем самым снижает охваты за ссылки на сайт, где пользователь может прочитать весь материал. Однако платформа лояльно относится, когда СМИ адаптирует материалы под нее.

Подводя итоги, можно дать несколько рекомендаций:

- Лучше подстраивать материал под условия платформы — это задерживает пользователя на материале и поднимает его в ленте рекомендаций.
- Локальные материалы получают больше охвата от региональной аудитории.
- Аудитория по России больше заинтересована в больших происшествиях и столкновениях политических интересов.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Альфа Центавра – новый алгоритм Яндекс.Дзена. // Журнал Яндекс.Дзена. [Электронный ресурс]. URL: <https://zen.yandex.ru/media/zenmag/alfa-centavra--novyi-algoritm-iandeksdz-ena-5c8bb4a817bca800b1877ca5> (дата обращения: 26.02.2021).
2. Ваш личный Дзен. // Яндекс Дзен. [Электронный ресурс]. URL: <https://zen.yandex/about> (дата обращения: 15.03.2021).
3. Дзен в 2020 году: рост видео, запуск стримов и другие события и планы. // Журнал Яндекс.Дзена. [Интернет-портал]. URL: <https://zen.yandex.ru/media/zenmag/dzen-v-2020-godu-rost-video-zapusk-stri-mov-i-drugie-sobytiia-i-plany-5fd37d730b82510af5def547> (дата обращения: 26.03.2021).
4. Индексирование интернета. // Яндекс. [Электронный ресурс]. URL: <https://yandex.ru/company/technologies/searchindex/> (дата обращения 27.05.2021).
5. Как работают рекомендательные системы. // Neurohive. [Электронный ресурс]. URL: <https://neurohive.io/ru/osnovy-data-science/rekomendatelnye-sistemy-modeli-i-ocenka/> (дата обращения: 26.03.2021).
6. Как Google использует файлы cookie. // Google Политика конфиденциальности и условия пользования. [Электронный ресурс]. URL: <https://policies.google.com/technologies/cookies> (дата обращения 25.04.2021).
7. Карма. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/channel/karma.html> (дата обращения: 17.03.2021).
8. Кликбейт. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/requirements/clickbait.html> (дата обращения: 17.03.2021).

9. Кутянин, А. Р. Рекомендательные системы: обзор основных постановок и результатов. // Интеллектуальные системы. [Электронный ресурс]. URL: <http://intsysjournal.org/pdfs/21-4/18-30-Kutanin.pdf> (дата обращения: 20.05.2020).
10. Новая эпоха авторов. // Блог Вконтакте. [Социальная сеть]. URL: <https://vk.com/blog/creators> (дата обращения: 20.03.2021).
11. Новости Немезиды. // Вконтакте для сообществ. [Социальная сеть]. URL: <https://vk.com/@adminsclub-nemesis-results> (дата обращения: 20.03.2021).
12. Новый медиакиТ Яндекс.Дзен. // Яндекс.Дзен для маркетологов. [Электронный ресурс]. URL: <https://zen.yandex.ru/media/zenexperts/novyi-mediakit-iandeksdzena-5e4289de24a0e07212071b81> (дата обращения 4.06.2021).
13. Нормы платформы. // Центр безопасности Вконтакте. [Социальная сеть]. URL: <https://vk.com/safety?section=standards> (дата обращения: 26/03/2020).
14. Плотников А. В. Конфиденциальность персональных данных о поведении онлайн-пользователей. // Образование и право. – 2019. – №1. [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/konfidentsialnost-personalnyh-dannyh-o-povedenii-onlayn-polzovateley> (дата обращения 21.05.2021).
15. Под защитой Немезиды. // Блог Вконтакте. [Социальная сеть]. URL: <https://vk.com/blog/nemesis> (дата обращения: 26.03.2021).
16. Поиск Яндекса и контент для взрослых. // Миссия Яндекса. [Электронный ресурс]. URL: <https://yandex.ru/company/rules/filtration/> (дата обращения 27.05.2021).
17. Правила пользования Сайтом ВКонтакте. // Вконтакте. [Социальная сеть]. URL: <https://vk.com/terms> (дата обращения 21.05.2021)

18. Принципы работы алгоритмов Google Поиска. // Google Поиск. [Электронный ресурс]. URL: <https://www.google.com/intl/ru/search/howsearchworks/algorithms/> (дата обращения 19.05.2021).
19. Принципы ранжирования поиска Яндекса. // Миссия Яндекса. [Электронный ресурс]. URL: <https://yandex.ru/company/rules/ranking/> (дата обращения 27.05.2021).
20. Пройдл А.. ИСТОРИЯ СИСТЕМ РЕКОМЕНДАЦИЙ. // «Теле-Спутник / [Электронный ресурс]. URL: <https://telesputnik.ru/materials/video-glavnoe/article/istoriya-sistem-rekomentatsiy/> (дата обращения: 20.05.2020).
21. Ранжирование и машинное обучение. // Яндекс. [Электронный ресурс]. URL: <https://yandex.ru/company/technologies/learning/> (дата обращения 27.05.2021).
22. Ройзнер, М. Как работают рекомендательные системы. Лекция в Яндексе. // Habr. [Электронный ресурс]. URL: <https://habr.com/ru/company/yandex/blog/241455> (дата обращения: 20.05.2020).
23. Спектр. // Яндекс. [Электронный ресурс]. URL: <https://yandex.ru/company/technologies/spectrum/> (дата обращения 27.05.2021).
24. Тематические ленты: лучшее про Ваши увлечения. // Блог Вконтакте. [Социальная сеть]. URL: <https://vk.com/blog/themefeeds> (дата обращения: 20.03.2021).
25. Требования к контенту. // Яндекс Справка. [Электронный ресурс]. URL: <https://yandex.ru/support/zen/requirements/rules.html> (дата обращения: 26.03.2021).

26. Топ ресурсов: Vk.com. \ Mediascope. [Электронные ресурс]. URL: <https://webindex.mediascope.net/report?id=16571> (дата обращения 8.06.2021).
27. Федеральный закон от 27.07.2006 N 152-ФЗ (ред. от 30.12.2020) «О персональных данных» / [Электронный ресурс]. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61801/4f41fe599ce341751e4e34dc50a4b676674c1416/](http://www.consultant.ru/document/cons_doc_LAW_61801/4f41fe599ce341751e4e34dc50a4b676674c1416/) (дата обращения 20.04.2021).
28. Что такое Прометей? // Вконтакте с авторами. [Социальная сеть]. URL: <https://vk.com/@authors-prometheus> (дата обращения: 20.03.2021).
29. Beware online «filter bubbles». // Ted talk. [Электронный ресурс]. URL: [https://www.ted.com/talks/eli\\_pariser\\_beware\\_online\\_filter\\_bubbles](https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles) (дата обращения: 17.02.2021).
30. Collaborative filtering. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/collaborative-filtering/> (дата обращения: 20.05.2020).
31. Content-based filtering. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/content-based-filtering/> (дата обращения: 20.05.2020).
32. Continuous Exploration. // Scaled Agile, Inc. [Электронный ресурс]. URL: <https://www.scaledagileframework.com/continuous-exploration/> (дата обращения: 29.03.2021).
33. Cookies, the GDPR, and the ePrivacy Directive. // GDPR.eu. [Интернет-портал]. URL: <https://gdpr.eu/cookies/> (дата обращения 26.05.2021).
34. Doshi, N. Recommendation Systems — Models and Evaluation. // Towards Data Science. [Электронный ресурс]. URL: <https://towardsdatascience.com/recommendation-systems-models-and-evaluation-84944a84fb8e> (дата обращения: 20.05.2020).

35. General Data Protection Regulation (GDPR). // GDPR.eu. [Интернет-портал]. URL: <https://gdpr.eu/tag/gdpr/> (дата обращения 26.05.2021).
36. Hybrid Recommender Systems. // Recommender system. [Электронный ресурс]. URL: <http://recommender-systems.org/hybrid-recommender-systems/> (дата обращения: 20.05.2020).
37. Key tips for safe and responsible cookie-based Web browsing. // All about Cookies. [Интернет-портал]. URL: <https://www.allaboutcookies.org/> (дата обращения 26.05.2021).
38. Pariser, E. Did Facebook's Big New Study Kill My Filter Bubble Thesis? // Backchannel. [Электронный ресурс]. URL: <https://medium.com/backchannel/facebook-published-a-big-new-study-on-the-filter-bubble-here-s-what-it-says-ef31a292da95>. (дата обращения: 20.05.2020).
39. Pariser, E. Fun facts from the new Facebook filter bubble study. // Medium. [Электронный ресурс]. URL: <https://medium.com/@elipariser/fun-facts-from-the-new-facebook-filter-bubble-study-444d4a658f5d>. (дата обращения 20.05.2020).
40. Prem Melville, Raymond J. Mooney, Ramadass Nagarajan Content-boosted collaborative filtering for improved recommendations. // Eighteenth national conference on Artificial intelligence. [Электронный ресурс]. URL: <https://www.cs.utexas.edu/~ml/papers/cbcf-aaai-02.pdf> (дата обращения 20.05.2020).
41. Ricci L. Rokach B. Shapira P. B. Kantor. Recommender Systems Handbook. // Springer Science+Business Media. [Электронный ресурс]. URL: <https://link.springer.com/book/10.1007/978-0-387-85820-3> (дата обращения 15.05.2020).

42. The Europa Web Guide. // European Commission.  
[Интернет-портал]. URL:  
<https://wikis.ec.europa.eu/display/WEBGUIDE/04.+Cookies> (дата обращения 26.05.2021).
43. Weisberg J. Is Web personalization turning us into solipsistic twits? // Salte. [Электронный ресурс]. URL:  
<https://slate.com/news-and-politics/2011/06/eli-pariser-s-the-filter-bubble-is-web-personalization-turning-us-into-solipsistic-twits.html>. (дата обращения: 20.05.2020).
44. Zaslowsky, J. If TiVo Thinks You Are Gay, Here's How to Set It Straight. // The Wall Street Journal. [Электронный ресурс]. URL:  
<http://web.archive.org/web/20190210040707/https://www.wsj.com/articles/SB1038261936872356908>. (дата обращения: 20.05.2020).

## ПРИЛОЖЕНИЕ

Таблица 1. Результаты опроса СМИ

| Название СМИ                           | Какие социальные сети для распространения информации вы используете                           | С какой целью вы используете каналы распространения информации, отмеченные выше? | Какие платформы приносят больше результата?  | Корректируете ли вы материалы для определенных соцсетей? Укажите, что меняете и по какой причине                       | Если в прошлом вопросе, вы ответили ДА. Насколько изменения повышают охват/просмотры/взаимодействие с публикацией   |
|--|---|--|--|--|---|
| Журнал «Стольник»                      | Инстаграм   | Вовлечение аудитории, Коммуникация с аудиторией                                  | Инстаграм  | Мы пишем отдельно пост для Инстаграма, иногда делаем картинки с цитатами, чтобы привлечь внимание аудитории            |   |
| <a href="http://Lenta.ru">Lenta.ru</a> | Вконтакте, Одноклассники, Яндекс.Дзен, Телеграм, Инстаграм, Фейсбук, Твиттер, YouTube, TikTok | Вовлечение аудитории, Коммуникация с аудиторией                                  | Все вышеперечисленные, на каждой платформе своя аудитория со своими предпочтениями | Для каждой соцсети -- свой подход. Начиная с вординга и подбора иллюстраций и заканчивая подбором и форматом контента. | Показатели растут только в случае, если контент подобран и упакован в соответствии с предпочтениями аудитории на конкретной платформе и в максимально органичном (комфортном) для пользователя формате. |

|                                  |  |  |   |  |  |
|----------------------------------|--|--|---|--|--|
| <a href="http://DK.RU">DK.RU</a> | Вконтакте, Одноклассники, Яндекс.Дзен, Телеграм, Инстаграм, Фейсбук, Твиттер   | Привлечение новой аудитории, Вовлечение аудитории, Перенаправление потока на сайт, Коммуникация с аудиторией | Фейсбук, Инстаграм, Телеграм  | Создается уникальный лид   | В структуре трафика соцсети занимают 15-30% в среднем  |
| Издательский Дом «Банзай»        | Вконтакте, Инстаграм, Фейсбук, YouTube   | Привлечение новой аудитории, Вовлечение аудитории, Коммуникация с аудиторией                                 | Фейсбук, Инстаграм  | Заголовок, иллюстрации, меняем отметки   | —  |
| Медуза                           | Вконтакте, Яндекс.Дзен, Телеграм, Инстаграм, Фейсбук, Твиттер, YouTube, TikTok | Привлечение новой аудитории, Вовлечение аудитории  | Инстаграм и твиттер   | Для телеграма пишем полный текст, а не только подводку, чтобы было удобнее читать  | К сожалению, не слежу  |
| Российская газета - Урал         | Вконтакте, Одноклассники, Фейсбук, YouTube                                     | Привлечение новой аудитории, Перенаправление потока на сайт  | У уральского филиала РГ - Вконтакте. Отметим, что у федеральной РГ есть аккаунты в инстаграме, телеграме и на яндекс.дзене, там тоже публикуются региональные тексты, но не так часто. Если анализировать федеральные | Фейсбук используется почти только для экономически х текстов. В ВК может появиться расширенная версия текста с комментариями, которые не вошли в печать. Для ОК и ВК иногда вставляется картинка в баннер со ссылкой. Если | Все зависит от темы. Что-то интересное или важное будут лайкать и без картинки. Если в начале поста указываем, что текст есть только в соцсетях (или какой-то комментарий в нем), то интерес тоже к нему больше. |

|  |  |  |   |   |  |
|--|--|--|---|---|--|
|  |  |  | аккаунты, то самый эффективный - дзен.  | бы на ФБ можно было бы вставить картинку в баннер, это делалось бы и там.   |  |
| <a href="http://JustMedia.ru">JustMedia.ru</a> | Вконтакте, Телеграм, Инстаграм, Фейсбук                                  | Привлечение новой аудитории, Вовлечение аудитории, Коммуникация с аудиторией | Фейсбук, Инстаграм  | Да. Меняем текст и картинки для Инстаграма, Стараемся уложить всю новость в пост для Фейсбука, тк пользователи не любят переходить по ссылке.   | В несколько раз.   |
| Е-катарсис                                     | Яндекс.Дзен, Телеграм, Инстаграм   | Привлечение новой аудитории, Вовлечение аудитории, Коммуникация с аудиторией | Инстаграм — более живой инструмент, там больше интерактива, но в телеграме я начинала, там более проверенная аудитория, я его использую больше для анонсов и коротких заметок с хештегом #сталоизвестно | Для инстаграма больше важен визуал, для телеграма - текст. Соответственно в инстаграме больше работаю с картинкой, иногда добавляю текст на неё | Трудно сказать   |
| Регионы России                                 | Вконтакте, Одноклассники, Телеграм, Инстаграм, Фейсбук, Твиттер, YouTube | Привлечение новой аудитории, Вовлечение аудитории                            | YOUTUBE   | Корректирую. это связано с возрастом аудитории соц сетей  | Я сам это не отслеживаю, это замеряют другие специалисты |

|                        |  |  |           |   |  |
|------------------------|--|--|-----------|---|--|
| Уралнаш                | Вконтакте, Телеграм, Инстаграм, Фейсбук, Твиттер | Привлечение новой аудитории, Перенаправление потока на сайт  | вконтакте | В вк стараемся выкладывать в посте со ссылкой по несколько фотографий, в фб раньше тоже так делали, но потом он стал их выкладывать в виде дурацкой плитки, поэтому теперь там просто размещаем ссылки с той обложкой, какую он сам предложит | в вк если в посте есть несколько фото, а не просто обложка ссылки, то охват больше |
| Онлайн-журнал «Нетабу» | Телеграм, Инстаграм, TikTok                      | Привлечение новой аудитории, Вовлечение аудитории, Перенаправление потока на сайт, Коммуникация с аудиторией | Инстаграм | Адаптируем и сокращаем текст для Инстаграма и Телеграма, отдельно работаем над картинками для этих же соц сетей   | Если не делать эту адаптацию, то охваты и все остальное значительно меньше         |

|              |                                     |  |   |  |   |
|--------------|-------------------------------------|--|---|--|---|
| Название СМИ | Использует ли ваше СМИ Яндекс.Дзен? | Какой у вас охват аудитории в Яндекс Дзен в месяц? | Какие материалы набирают большую популярность в Яндекс. Дзен? | Получает ли материалы с конкретной тематикой больше охватов, чем другие? | Совпадает ли аудитория в Яндекс Дзена с вашей ЦА? |
|--------------|-------------------------------------|--|---|--|---|

|  |    |   |   |                               |                           |
|--|----|---|---|-------------------------------|---------------------------|
| <a href="http://Lenta.ru">Lenta.ru</a> | Да | точные цифры не раскрываем, но это многомиллионные показатели | Материалы, которые напрямую связаны с жизнью читателей, и те, что вызывают максимальный эмоциональный отклик. | Да                            | Да                        |
| <a href="http://DK.RU">DK.RU</a>       | Да | От 3 тыс. до 20 тыс., в зависимости от месяца                 | Лайфхаки: психология, мотивация, менеджмент   | Да                            | Совпадает примерно на 30% |
| Медуза                                 | Да | Не знаю((   | Мы не следим за этим. По крайней мере, наш отдел  | Да                            | Да                        |
| Е-катарсис                             | Да | Не занимаюсь им   | Самый популярный материал - «7 лайфхаков, как ходить в филармонию и получать удовольствие»                    | Не знаю, у меня одна тематика | Не знаю                   |

| Название СМИ                           | Какой у вас охват аудитории во ВКонтакте в месяц? | Какие материалы набирают большую популярность во ВКонтакте? | Совпадает ли аудитория в Вк с вашей ЦА?                                   |
|--|---|---|---|
| Журнал «Стольник»                      | -   | -   | Нет, не совпадает. Наша аудитория просто не сидит в этой социальной сети. |
| <a href="http://Lenta.ru">Lenta.ru</a> | точные цифры не раскрываем                        | Касающиеся непосредственно жизни пользователя               | да  |
| <a href="http://DK.RU">DK.RU</a>       | Более 5 тыс.                                      | Новости, городская повестка                                 | До 30%  |
| Медуза                                 | Около миллиона                                    | Про политику  | Да  |

|  |   |                            |  |
|--|---|----------------------------|--|
| Российская газета - Урал                       | если считать средний показатель за последний год, то 35 тысяч | фоторепортажи              | аудитория в нашем паблике совпадает с ЦА издания   |
| <a href="http://JustMedia.ru">JustMedia.ru</a> | 10000   | Конкурсы                   | Да   |
| Е-катарсис                                     | Не веду вконтакте   | Не веду вконтакте          | Не веду вконтакте  |
| Регионы России                                 | не владею данной информацией                                  | не знаю                    | не знаю  |
| Уралнаш  | 10000   | опросы                     | все-тки соцсети рассчитаны на то, чтобы их листать, а мы пишем такое, что нужно читать подолгу, так что не вполне совпадает, но большей частью |
| Онлайн-журнал «Нетабу»                         | Не используем эту площадку                                    | Не используем эту площадку | Не используем эту площадку   |