

Поисковое поведение пользователя в новостных агрегаторах

Научный руководитель – Садохин Александр Петрович

Чижик Анна Владимировна

Аспирант

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Институт государственной службы и управления, Факультет государственного и муниципального управления, Москва, Россия

E-mail: afrancuzova@mail.ru

Изучение человеческого общества как социального явления возможно только через призму культуры, потому что оно существует только в ее контексте, при этом довольно сложно определить, что включает в себя это понятие. Эмпирически в той или иной форме ученые, представляющие разные научные сферы (от философии до лингвистики), говорят о том, что культура есть система образцов поведения и ценностей, которая формирует социальную среду. Иными словами утверждается, что культура является сложным механизмом, вырабатывающим информацию. Она, опираясь на ресурсы окружающего мира, превращает не информацию в информацию, а является, по сути, антиэнтропийным механизмом человечества. Таким образом, определяется основная роль культуры - быть негенетической памятью человеческого общества [1], тем самым задавая индивидам и социальным группам модель поведения и понимания действительности, что, в свою очередь, активизирует тенденции развития человечества. Культура, являясь внутренне сложно организованной системой, воспроизводит общественное сознание, «объективное» и «субъективное» в историческом процессе.

Большая часть социальной реальности находится за пределами повседневной жизни человека, индивид не может оценить сведения на предмет достоверности, поэтому он находится в режиме доверия или недоверия к СМИ [2] (реальный анализ данных не происходит). Таким образом, возникает тезис о том, что у СМИ есть возможности для создания псевдокартины мира в сознании индивидов. Так вокруг сложных социальных явлений складывается комплекс устойчивых ассоциаций, «картинки в голове», стереотипы.

Современный статистический ландшафт позволяет наблюдать поведение аудитории, ее реакцию, запросы и усталость, собирая базу знаний о возрасте, поле, географии, типичном времени потребления информации, интересующих темах и так далее. Все эти данные дают возможность реалистичного социального прогнозирования. При уникальной ситуации, когда временем потребления новостей стало практически все время бодрствования пользователя, существенно усложнились наиболее важные показатели реагирования аудитории на новостной поток. Например, в данный момент происходит перераспределение долей носителей информации (устройств и платформ) в информационном суточном круге потребителя, а также существенно изменяются показатели длины и глубины сессий потребления. Одним из важнейших показателей настроений и взглядов социума в этой ситуации становится попадание новостного сообщения в топы агрегаторов (чаще всего речь идет о Яндексe и Google), которые формируются на основе алгоритмов, реализованных за счет анализа количественных показателей (количество кликов и репостов). Так формируется социально-ориентированный подход. В этом случае аудитория СМИ рассматривается как субъект процесса массовой коммуникации.

Логическая структура процесса поиска

Введем понятие логической структуры поиска как когнитивного объекта, практически независимого пространства его физической реализации, определяемого доступной по-

исковой средой. Логическая структура поиска является графом зависимостей, вершинами которого являются цели поиска, запросы и результаты поиска. Логическая структура представляет два противоположных типа зависимостей: зависимости «быть результатом» (изменение целей поиска как результат того, что нашла система, запрос как результат цели поиска, найденные системой документы как результат запроса) и «композиционные» зависимости, описывающие составные цели и составные запросы. Зависимости первого типа - линейные и ветвящиеся (исходящих дуг может быть несколько), второго - «сходящиеся» (входящих дуг больше одной). Логическая сессия есть логическая структура с относительно неизменной целью поиска.

Ограниченность памяти человека и вытекающие из этого ограничения на физическую реализацию поиска

На рис.1 показано, как реализация одной и той же логической структуры поиска зависит от ресурсов памяти человека.

Рис. 1. Пределы реализаций логической структуры поиска, налагаемые памятью человека в одно- и многооконной поисковых средах

На первом рисунке (А) - потенциально возможное ветвление поиска, которая на практике (в многооконной среде) в связи с ограничениями кратковременной памяти реализуется в схему (В) и (С). Дело в том, что пользователь не способен поддерживать параллельную работу со многими ветвями, поэтому чаще всего использует линейную модель или слабо разветвленную модели поиска.

Сценарии поведения пользователей на компьютерах, планшетах и телефонах

Из всего потока запросов, который обрабатывают поисковые системы, на момент конца 2016-го года около 60% запросов приходится на мобильные устройства. При этом поиск со смартфонов происходит в два раза чаще, чем с планшетов. Однако если обратиться к статистике запросов по ключевым словам непосредственно к поисковым агрегаторам, окажется, что доля активных пользователей поиска перемещается в сторону пользователей ПК (70% от общего числа поисковых запросов). Такие данные легко объясняются многими характеристиками мобильных устройств (наличие приложений, которые выводят ленту новостей автоматически, удобность просмотра новостей через социальные сети и т.д.). Но самое главное это диктует реагирования механизма вывода новостей в ТОПы поисковых агрегаторов, так как первейшее условие передвижение новостного сообщения на более высокие позиции - клики пользователей.

С компьютеров люди задают примерно одинаковое количество запросов в течение рабочего дня, а на мобильных устройствах заметно меньше ищут утром и больше во временном промежутке с пяти до десяти часов вечера. Один и тот же человек может в течение одного и того же дня использовать разные устройства: например, на работе - компьютер, в обеденный перерыв - телефон, а вечером дома - планшет.

Чаще всего для поиска на всех типах устройств люди используют существительные (56% всех слов в запросах к новостным агрегаторам), при этом обычно пользователи используют всего одно слово-тег.

Насколько тема характерна для того или иного типа устройств, можно оценить с помощью аффинити-индекса, который показывает, на сколько процентов популярность темы на устройстве отличается от средней. В зависимости от используемого устройства тематика запросов распределяется на три группы: образованием люди интересуются, заходя в агрегаторы с телефонов, вопросы, связанные с работой и развлечениями, поступают с планшетов, а экономическое обозрение, проблемы социальной и политической жизни пользователи чаще всего ищут с компьютеров.

Топы новостных агрегаторов как зеркало медиасреды

Агрегаторы новостей рассчитывают «вес» сообщения на основе трёх критериев: цитируемость (сколько ссылок на сообщение в других сообщениях сюжета, без учёта ссылок в аффилированных СМИ и самоцитирования), свежесть (время публикации сообщения по сравнению с другими источниками) и информативность (наполненность сообщения ключевыми фактами сюжета) [3].

Из фрагментов трёх лучших по этим критериям сообщений формируется краткое описание сюжета (аннотация). По тому же принципу выбирается заголовок сюжета, прочитав который пользователь должен сразу понять, о чём идёт речь. Под аннотацией и названием сюжета новостные агрегаторы формируют блок из пяти похожих новостей. Из всех сообщений робот выбирает наиболее цитируемые и ранжирует их по свежести. Остальные новости в виде хронологической ленты можно увидеть, развернув конкретный сюжет.

Изложенный алгоритм, по которому работает робот, в сочетании с вышеизложенными тезисами о предпочтениях читательской аудитории и характеристиками человеческой памяти, задающие дополнительные критерии поисковых запросов, формирует несколько предсказуемых предпосылок формирования основных тем ТОПов новостных агрегаторов. Например, ранним утром в будние дни (6-9 утра), а также в дневное время в выходные первые десять новостных сюжетов в агрегаторах почти всегда связаны с криминальными сводками и хроникой дорожно-транспортных происшествий. С другой стороны в течение рабочего дня пять первых позиций в ТОПах обычно занимают шутливые заголовки с непредсказуемой тематикой. Вечером же обычно некоторый хаос в формировании повестки дня уменьшается, уступая место структурированным спискам тем по экономике и социальной политике.

Благодаря статистическим исследованиям можно определить первичное облако тегов, интересующее аудиторию (а также разбить запросы по возрастным и гендерным признакам на подгруппы). Так, например, наиболее частотные запросы к агрегаторам новостей в Санкт-Петербурге за январь - это следующая пятерка слов: «Петербург», «Полтавченко», «Исаакий», «ЖКХ», «ограничения движения». Исследуя эти слова-теги, можно создавать группы ТОПов, благодаря которым выстраиваются опорные конструкции для социального прогнозирования процессов, происходящих в обществе.

В докладе будут изложены результаты статистических (машинный анализ данных) и социологических (метод опроса) исследований ТОПов новостных агрегаторов, которые проводились в декабре 2016-го и январе-феврале 2017-го годов на базе СМИ, работающих на территории Санкт-Петербурга и Ленинградской области.

Источники и литература

- 1) Ю. М. Лотман. История и типология русской культуры. – СПб, 2002.
- 2) Кастельс М. Информационная эпоха: экономика, общество и культура / Пер. с англ. под науч. ред. О. И. Шкаратана. – М., 2000.
- 3) Документация - Технологии Яндекса: <https://tech.yandex.ru/money/doc/dg/concepts/About-docpage/>

Иллюстрации

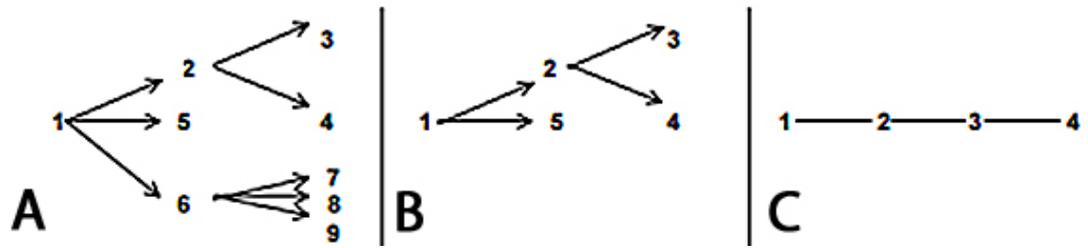


Рис. 1. рис. 1. Пределы реализаций логической структуры поиска, налагаемые памятью человека в одно- и многооконной поисковых средах