

Современные тенденции поиска и анализа информации



О.Н. Шорин,
зам. ген. директора
РНБ

В течение последних нескольких десятилетий было модно делить молодых людей на поколения: поколение X, затем Y, и, наконец, Z. Изначально деление на поколения происходило на основе различных экономических, демографических, политических, маркетинговых признаков. Однако исследователи сходятся во мнении, что основными признаками принадлежности к поколению Y и Z является период жизни, в который произошел технологический бум. Считается, что детство людей поколения Y прошло в те годы, когда цифровое будущее еще не наступило. Люди же из поколения Z с самого детства окружены различными гаджетами и не представляют себе, как можно жить несколько дней без Интернета.

Но технологические прорывы случаются гораздо чаще, чем раз в 20–30 лет, поэтому есть необходимость исследовать более детально, какие же события произошли в сфере информационных технологий за последние 20 лет и как они отразились на культуре чтения и письма в частности.

Google

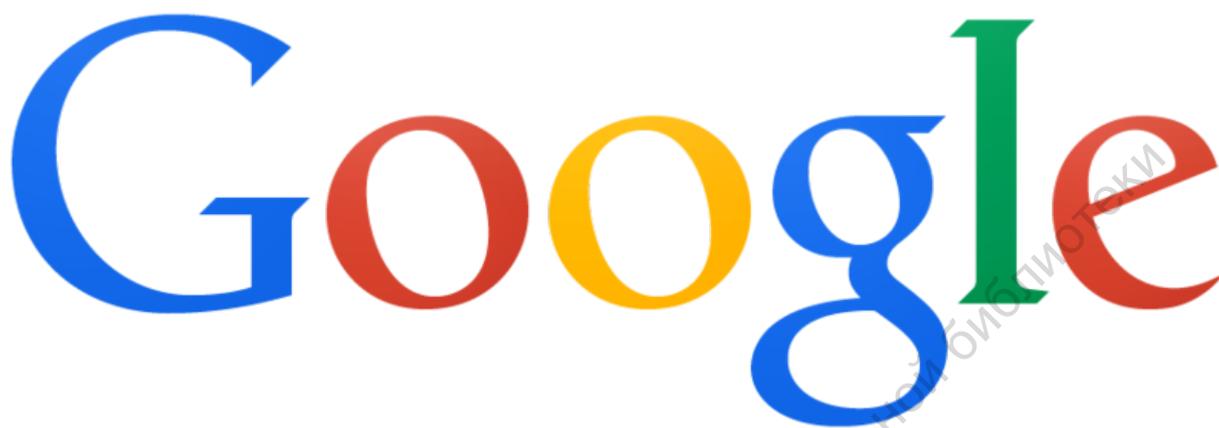
16 лет назад появился поисковый сервис Google. Нельзя не вспомнить, что и до Google существовали поисковые движки, такие как Yahoo!, Altavista, Rambler, Yandex, но именно Google совершил прорыв, который очень сильно повлиял на поведение пользователей в Глобальной сети. Существовавший до этого времени поиск был рассчитан на людей, которые разбирались в компьютерных технологиях, могли составить сложный поисковый запрос с использованием специализированных символов, проанализировать несколько страниц результатов и при необходимости переформулировать поисковое требование.

Google, конечно же, обладал всеми этими возможностями, но основной упор делался на простой поиск, когда пользователь мог вводить фразу на естественном языке, не особо задумываясь о том, как же машина интерпретирует тот или иной падеж, сочетание и порядок слов. Но еще более сногшибательным оказалось то, что алгоритм ранжирования результатов Google был устроен таким образом, что на 2-й, 3-й и последующих страницах не было ничего существенно важного. Отпала необходимость перелистывать страницы: если на первой странице не было интересующих результатов, то можно было смело переформулировать запрос.

У поколения Google, как окрестили молодых пользователей Интернета библиотечные эксперты, можно выделить несколько особенностей:

- 1) сбор информации начинается с поиска в Интернете, а не в библиотеке или на библиотечном сайте;
- 2) простой поиск превалирует над расширенным;
- 3) поиск задается с орфографическими ошибками;
- 4) по умолчанию считается, что если искомая информация отсутствует на первой странице результатов поиска, то ее не существует вообще.

В Российской национальной библиотеке библиотекари столкнулись с подобным поведением пользователей. За последний год (с 18 ноября 2013 г. по 18 ноября 2014 г.) в электронном каталоге РНБ был совершено 1485971 простой поиск против 540318 расширенных поисков. Почти трехкратный отрыв. А если учесть, что каталогом РНБ пользуются профессиональные библиотекари из региональных библиотек, которые,



скорее всего, в силу своей специфики используют расширенный поиск, то разрыв между количеством осуществленных простых и расширенных поисков рядовыми читателями значительно усиливается.

Также нередки случаи, когда поисковый запрос делается с орфографическими ошибками или опечатками, например «харошо» или «А зори зднсь тихие». Примечательно, что для большей части таких ошибок выдается результат. Происходит это по двум причинам:

1. Программное обеспечение электронного каталога РНБ обладает неким интеллектом и понимает, что вводя слово «харошо» пользователь, скорее всего, имел в виду «хорошо». В таком случае электронный каталог выдает подсказку типа «возможно вы искали «хорошо»?» и предлагает в один клик осуществить поисковый запрос без ошибок.

2. Если в электронной библиотеке есть отсканированная копия книги, то поиск осуществляется и по полному тексту тоже. Однако распознавание полного текста происходило полностью в автоматическом режиме, что приводило к появлению ошибок. И существует вероятность того, что слово «хорошо» было напечатано не очень четко и из-за этого оно было распознано с ошибкой – «харошо».

Очевидно, что если пользователь вводит запрос на так называемом в Интернете

«олбанским» языке, например «пазитифф», то список результатов с большой долей вероятности будет пустым. Возникает вопрос, что делать в данной ситуации: должны ли библиотеки идти на поводу у пользователей и придумывать различные алгоритмы, выявляющие подобные случаи, или же не стоит реагировать на подобные запросы, вынуждая читателей повышать уровень грамотности? Подобный вопрос может расколоть сообщества библиотекарей и читателей на два лагеря. Однако, как будет видно далее, эволюция интерфейса взаимодействия с пользователем позволит избежать необходимости отвечать на этот непростой вопрос.

Apple

Уверен, что никто не будет спорить с тем, что в 2007 г. компания Apple произвела настоящую революцию в сфере высоких технологий, выпустив свой смартфон – iPhone. Стив Джобс и сотрудники его компании избегали и избегают применять слово «смартфон» по отношению к своему устройству, так как до этого на рынке уже существовали смартфоны, но для их использования необходимо было иметь специальность, как минимум, системного программиста. iPhone, а через три года и iPad, позволили обыкновенным пользователям получить доступ к доселе невиданным для них возможностям.

Нетрудно догадаться, что самым значи-

мым достижением, которое получили пользователи с появлением iPhone/iPad, является мобильность. Если раньше мобильным считался сотрудник, который мог ответить на электронное письмо не только со своего рабочего места в офисе, но и со своего домашнего компьютера, то теперь невозможность проверить электронную почту, отредактировать документ, поучаствовать в онлайн-конференции, находясь в поезде, аэропорте или даже на отдыхе, расценивается как недостаток. Ведущие аналитические агентства однозначно утверждают, что продажи ноутбуков и уж тем более десктопных решений сокращаются несколько кварталов подряд.

Еще одним немаловажным аспектом, оказавшим влияние на мышление людей, стало появление в 2009 г. магазина приложений App Store для мобильных устройств Apple. Компании удалось создать удивительную «экосистему» из своих устройств и приложений сторонних разработчиков, которая по аналогии со снежным комом привлекала новых пользователей. А столь огромный рынок потенциальных клиентов, в свою очередь, привлекал новых разра-

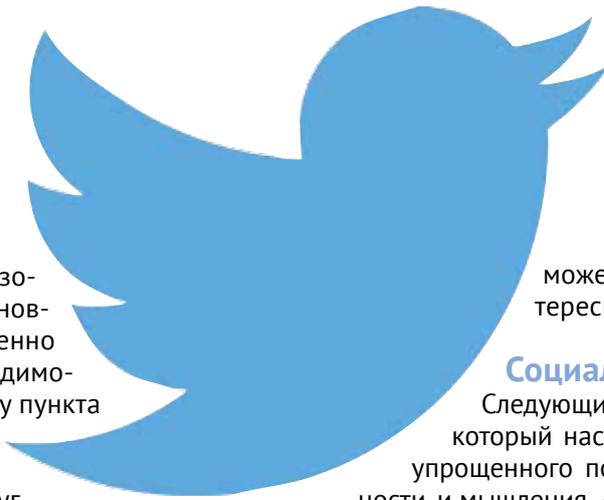
ботчиков, стимулируя их создавать все более и более качественные приложения.

Данная модель развития была скопирована различными компаниями, выпускающими как смартфоны, так и телевизоры, автомобили и даже домашнюю технику. Доля Smart TV среди телевизоров неуклонно растет. Появились магазины приложений для автомобилей, например AT&T Drive, в которых самым востребованным является приложение, помогающее осуществлять параллельную парковку. Компания LG анонсировала серию холодильников с возможностью установки приложений. Предполагается, что для таких холодильников большим спросом будут пользоваться приложения, предлагающие рецепты приготовления блюд из продуктов, имеющихся в холодильнике.

Мышление среднестатистического пользователя сильно изменилось. Считается, что для любого аспекта в жизни должно существовать отдельное приложение. Хочет ли человек заняться бегом, подтянуть знания правил дорожного движения, начать коллекционировать монеты – для всего этого должно быть отдельное приложение. В английском языке даже есть устойчивое выражение «There's an app for that», что в вольном переводе означает «И для этого тоже есть приложение». Как мы понимаем, библиотеки – не исключение.

Наличие множества приложений, которые собирают, генерируют, каким-то образом обрабатывают только ограниченную часть информации о человеке, его увлечениях и интересах, логичным образом вылилось в создание «облаков данных», куда стекается информация из разных приложений. В этом «облаке» все данные собираются, анализируются, и на их основе в приложения поступает персонализированный результат. Например, информация из календаря о том, что у человека запланирована встреча, проанализированная в совокупности с данными о текущем





местонахождении пользователя и дорожной обстановкой, позволяет своевременно напомнить ему о необходимости выдвигаться в сторону пункта назначения. Подобный анализ больших объемов, не связанных друг с другом данных, поступающих из разнообразных источников, получил название Big Data – «Большие данные».

Библиотеки, как бы странно это ни звучало, тоже сталкиваются с проблемой «больших данных». В информационные системы библиотеки поступают данные о том, какие поисковые запросы читатель использовал, на какие онлайн-сервисы подписался, какие электронные книги просматривал и как долго, какие книги взял в читальный зал, какие выставки посетил и т.д., и т.п. Проанализировав всю эту информацию, вполне возможно создать сервисы библиотечной направленности, которые по своему уровню не уступали бы каждодневно используемым сервисам. Например, можно оповещать читателя о проводимой в библиотеке выставке, которая заведомо

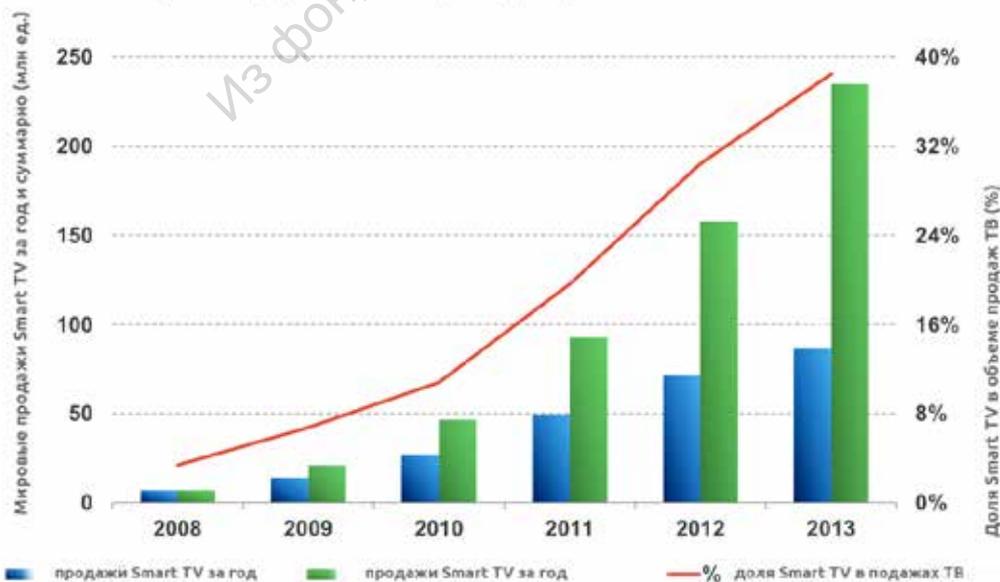
может быть ему интересна.

Социальные сети

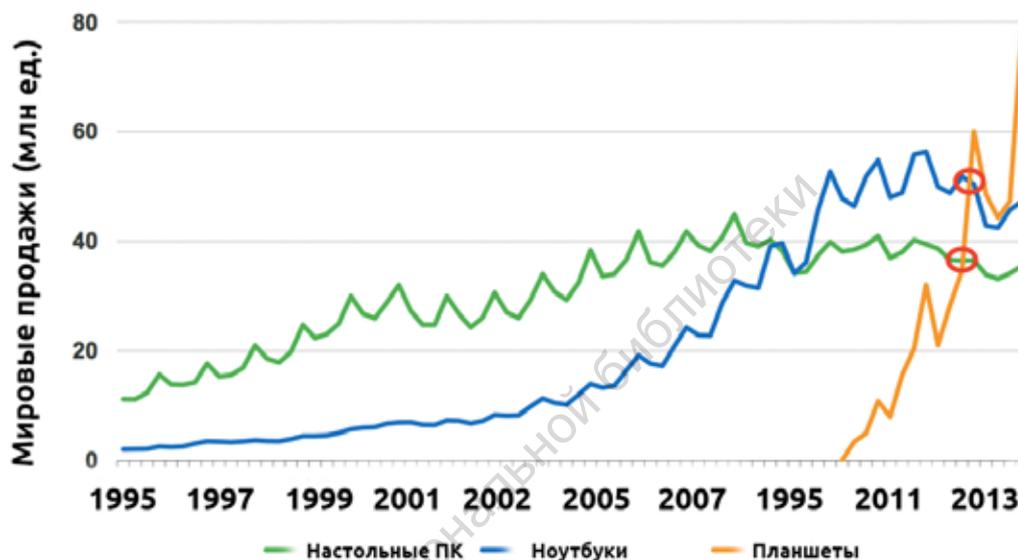
Следующим фактором, который наложилась поверх упрощенного поиска, мобильности и мышления, ориентированного на приложения, и который оказал самое непосредственное влияние на читателя, на наш взгляд, является появление Twitter. Сами социальные сети появились относительно давно, на волне развития так называемого Web 2.0. Концепция Web 2.0 заключалась в том, что контент в Сети формировали сами пользователи с использованием удобных, понятных простому обывателю инструментов. В то время как до появления этих инструментов создание, наполнение и развитие сайтов в Интернете было делом избранных профессионалов.

Создатель протокола HTTP, языка разметки HTML – Тим Бернерс-Ли, которого называют изобретателем Интернета, не считает Web 2.0 чем-то выдающимся, поскольку подход к созданию контента пользователями полностью основывался

Мировые продажи телевизоров с функцией Smart TV в 2008-2013 гг.



Продажа настольных ПК, ноутбуков и планшетов



©КРСВ Источник: Morgan Stanley Research. Примечание: К ноутбукам относятся и нетбуки

на возможностях созданных им механизмов, существовавших с самого начала. Он считает, что Web 2.0 нужен был для маркетинговых целей, так как лопнувший в начале 2000-х гг. мыльный пузырь доткомов существенно подорвал доверие к интернет-отрасли.

Тем не менее, следуя идеологии Web 2.0, появились такие понятия, как блоги, возможность оставлять комментарии, предоставлять рейтинги и тэги. Консолидируя в себе все эти идеи, возникли социальные сети. Одной из первых появилась сеть LiveJournal, которая представляла собой платформу для ведения дневников-блогов с возможностью последующего обсуждения. Таким образом, получалось, что эта социальная сеть мотивировала своих участников к созданию нового контента. В отличие от LiveJournal, самая популярная в мире социальная сеть Facebook ориентирована не на создание контента, а на публикацию ссылок на уже существующий контент. Вероятно, именно разницей трудозатрат по ведению своей страницы можно объяснить популярность Facebook и стагнацию LiveJournal.

Но, как уже говорилось ранее, наиболее глубокий след в плане потребления информации оставил сервис микроблогинга Twitter. Этот сервис позволяет вести дневник, создавая сообщения, длина которых ограничивается 140 символами. Объясняется такое ограничение необходимостью интеграции с SMS-сообщениями. Такое объяснение выглядит немного странным, так как, во-первых, многие современные люди не пишут SMS, а используют альтернативные сервисы – iMessage, WhatsApp, Viber, в которых нет ограничений на количество символов, а во-вторых, интеграция Twitter с SMS работает всего лишь в двух странах – США и Великобритании. Тем не менее ограничение в 140 символов существует в Twitter и по сей день, и отказываться от этого ограничения сервис не собирается.

Twitter сильно изменил способ подачи информации: на смену малому количеству хорошо продуманных, выверенных статей пришло огромное количество маленьких высказываний, из которых складывается целостная картина. При таком подходе вес отдельно взятой публикации ничтожно

мал, что влечет за собой изменение способа восприятия информации. Пользователь пролистывает бесконечный поток этих коротеньких сообщений, особо сильно не вчитываясь и не вдаваясь в подробности каждого из них. Находясь в потоке постоянного информационного шума, современный человек практически не способен отделить действительно важное сообщение от ничего не значащей чепухи.

Такой метод восприятия информации практически отучил людей вдумчиво читать и мыслить. Появился даже специальный термин – «оторванное чтение». Это такой способ чтения, при котором человек, прочитав текст, не может его понять, осмыслить, сформулировать основную мысль текста. Во главу угла ставится не ценность полученной информации, а сам процесс потребления этой информации. Именно «потребления», как fast food – быстро, урывками, на ходу, 140 символов, не тратя на это ни времени, ни усилий.

Поскольку чтение текстов неотрывно связано с их созданием, необходимо отметить, что молодежь перестает грамотно писать. Пытаясь уместиться в 140 символов, она приносит в жертву предлоги, слова искусственно сокращаются, вводятся в оборот специальные словосочетания, например IMHO, ROTFL, ASAP, придумываются различные смайлики, не сочетаются падежи, склонения, слова пишутся с ошибками.

К повальной безграмотности и неумению составлять из слов осмысленные тексты людей подталкивают современные технологии. В последние пару лет активно развивается голосовой интерфейс взаимодействия с устройствами. Зачем что-то писать, если можно нажать на кнопку и произнести то, что тебе нужно? На всех современных платформах существуют голосовые помощники: Apple Siri, Google Now, Microsoft Cortana. Очевидно, что в ближайшем будущем человечество все дальше и дальше будет отходить от использования текстовой информации. Она будет вытесняться мультимедийной, голо-

совой, визуальной, так как эти виды представления информации более удобные, более наглядные, быстрее производящие необходимый эффект.

Выводы

В этом году в нашей стране реализуется крупнейший проект в библиотечной области – создание Национальной электронной библиотеки. Наличие слова «электронная» в названии этого проекта означает, что существовать этой библиотеке суждено наряду с современными системами в интернет-пространстве. И для того чтобы суметь ужиться в этой среде, быть востребованным, придется играть по правилам современного электронного мира, оказывать тот уровень сервиса, к которому привыкло современное поколение.

Национальная электронная библиотека должна обладать следующими свойствами:

- 1) необходимо наличие приложений для всех современных платформ;
- 2) информация с различных устройств пользователя должна автоматически синхронизироваться с использованием «облачных технологий»;
- 3) для каждого конкретного пользователя библиотеки информация, рекомендации, подборки, списки литературы должны персонализироваться в зависимости от его интересов, прав доступа, местоположения, устройства, с которого осуществляется работа;
- 4) должен быть предусмотрен голосовой метод ввода информации;
- 5) для быстрого ознакомления с основной сутью того или иного документа следует создать сервисы автоматического реферирования и аннотирования текстов.