

Лекция 7. Большие данные в мультимедиа

Интернет

Большие данные и развлечение с помощью больших данных пришли в нашу жизнь и появлением Интернета. Что это такое?

Фактически это инфраструктура для связи компьютеров, «межкомпьютерная сеть». Википедия говорит, что это «всемирная система объединенных компьютерных сетей для хранения и передачи информации». Всемирная паутина — это не Интернет, она работает на основе Интернета, и их часто путают.

Концепция связи компьютеров появилась в 1962 году. Но прошло несколько лет, перед тем как компьютеры удалось соединить. Считается, что Интернет вырос из военной сети *APRANet*, которая быстро переместилась в американские университеты. Первый специальный кабель был проложен между Калифорнийским университетом в Лос-Анджелесе, в Санта-Барбаре, Стэнфордским университетом и Университетом штата Юта для того, чтобы исследователи в одном университете могли быстро связаться с компьютером в другом, не тратя времени на дорогу. Первый сеанс связи между университетом в Лос-Анджелесе и Стэнфордским университетом был проведен 29 октября 1969 года. Удалось отправить всего два символа.

Идея Всемирной паутины зародилась в недрах лаборатории ЦЕРН — Европейской организации по ядерным исследованиям — для связывания документов посредством гипертекстовых ссылок и для облегчения обмена информацией между группами исследователей, занимающихся проведением экспериментов на Большом коллайдере. Идея принадлежит британскому ученому Тиму Бернерсу-Ли (р. 1955), который работал в ЦЕРН консультантом по программному обеспечению. Годом рождения Всемирной паутины считается 1989-й. В экспериментах на Большом коллайдере участвовали сотни ученых со всего мира, и им требовался быстрый, часто мгновенный обмен данными. Первоначально проект использовался только во внутренней сети ЦЕРН. В 1991 году были созданы первые в мире веб-сервер и браузер, их «папой» тоже является Тим Бернерс-Ли. Он же разместил первый сайт на первом сервере 6 августа 1991 года. В апреле 1993 года ЦЕРН объявила, что система будет открыта для всех пользователей.

В течение 1990-х годов Интернет объединил большинство существовавших тогда сетей. В 1997 году Сеть включала уже 10 миллионов компьютеров и стала очень популярным средством обмена информацией. В течение пяти лет количество пользователей достигло 50 миллионов человек. Сейчас Интернет состоит из многих тысяч компьютерных сетей — научных, правительственных, корпоративных, домашних. А Всемирная паутина вызвала настоящий взрыв в информационных технологиях и развитии Интернета. В повседневной речи, употребляя слово «Интернет», мы обычно имеем в виду *www* — *World Wide Web*.

Про электронную почту можно говорить начиная с сентября 1973 года. И ее появлению мы обязаны бритве. Леонард Клейнрок (р. 1934), американский инженер и ученый в области информационных технологий и компьютерных сетей, забыл свою бритву в Брайтоне, Великобритания, куда ездил на конференцию по компьютерным сетям, которая проходила в Сассекском университете. Обнаружил он это только дома в Лос-Анджелесе. В дни конференции была установлена временная связь с *APRANet*, сигнал передавался через станцию в Корнуолле и спутник — таким образом обеспечивалась трансатлантическая телефонная связь и передача телепрограмм. К моменту возвращения Клейнрока в Лос-Анджелес связь не была прекращена, так как часть делегатов все еще оставалась в университете. Он нашел коллегу, подключенного

к сети (несмотря на то, что в Брайтоне в то время было 3 часа ночи), и отправил сообщение через программу, разработанную для связи телетайпов. Просьба Клейнрока найти его бритву — это первый *email* в истории.

Первая поисковая система, *AltaVista*, появилась только в 1995 году. В основном там были тексты и немного фотографий, но не самого лучшего качества. Тогда мы даже не могли представить, как Интернет и Всемирная паутина изменят нашу жизнь, жизнь каждого простого человека. Больших данных еще не было, а с их появлением жизнь изменилась еще больше.

Ответ на все

Что было раньше? Представьте, что вы услышали в программе новостей информацию о каком-то человеке и вам захотелось узнать о нем побольше. Вы заглядывали в толстые тома энциклопедий, стоявших у вас дома. У всех (или у большинства) стояли дома какие-то энциклопедии. Если вы не находили нужную информацию дома, то отправлялись в библиотеку. Там было больше справочного материала. Если человек был из мира политики, то можно было посмотреть подшивки газет и журналов. Если это был киноактер, певица, спортсмен, то вы тоже просматривали подшивки газет и журналов — соответствующие издания и разделы. Вы могли провести два дня за поисками и не найти ничего. Вообще ничего. Сегодня такое просто невозможно.

Теперь вы просто достаете из кармана смартфон, набираете имя и фамилию — и информация оказывается у вас перед глазами.

Теперь Интернет иногда называют «оракулом больших данных», и это правильно, потому что он является источником всей информации, которая вам может потребоваться, и предоставляет ее вам тогда, когда она вам нужна, и независимо от места, в котором вы находитесь.

Но Интернет — не энциклопедия. На ваш запрос (например, введенные имя и фамилию) вы можете получить 500 тысяч ссылок, можете миллион, а то и больше. Это одна из самых впечатляющих демонстраций работы больших данных. Вы просто физически не сможете пройти по всем этим ссылкам, даже при легком и быстром доступе. А поскольку информации так много, вся она просто не может быть правильной. Так что нужно критически подходить ко всей всплывающей информации. Никто не гарантирует, что все, что вы прочитаете в Интернете, соответствует действительности. Эту информацию легко опубликовать, легко получить, но ее надо проверять.

Несколько слов о Википедии, которая появилась в 2001 году. Изначально планировалось создать гигантскую мировую энциклопедию, причем бесплатную. И в самом начале она создавалась как обычные энциклопедии — статьи писали эксперты в каждой области. Но задача оказалась непосильной, и уже через несколько лет в проекте мог принять участие любой человек. Ни в одной энциклопедии в истории человечества не было столько информации, сколько в Википедии. По количеству данных они даже к ней не приближались. Эта самая полная энциклопедия за всю историю человечества и по тематическому охвату, и по объему сведений. Сравните с бумажным энциклопедическим словарем, который, возможно, до сих пор хранится у вас дома, или даже многотомной бумажной энциклопедией. К тому же это многоязычная энциклопедия (есть информация на 301 языке), что отличает ее от предыдущих. Википедия — это большие данные, причем и по количеству статей, и по количеству информации в каждой статье. В ней более 40 миллионов статей. Это пятый по посещаемости сайт в мире, при этом — самый популярный и самый крупный справочник. И это быстрый поиск, что тоже является одной из характеристик

технологии *Big Data*. Что и отражено отражено в названии. Английское слово *wiki* из гавайского языка, в котором означает «быстро», то есть дословно Википедия — это «быстрая энциклопедия».

Сейчас ее все больше и больше ругают за неточность, вспоминают, как раньше над энциклопедиями работали ученые и эксперты из разных областей знаний, проверяли и перепроверяли информацию. Статьи в Википедии не проходят экспертной оценки, что является, например, обязательным требованием для *Encyclopaedia Britannica* («Британской энциклопедии»). Но статьи могут редактировать другие участники Википедии, хотя нет гарантии, что они исправят неправильное на правильное.

Не так давно британские ученые провели экспертную оценку ряда статей из Википедии, посвященных научной тематике. Можно ли считать информацию в Википедии надежной? Ученые пришли к выводу, что в статьях на научные темы и статьях, посвященных различным технологиям, ошибок не больше, чем в «Британской энциклопедии», но в Википедии гораздо больше информации, чем в «Британской энциклопедии». Они же указали на предвзятое отношение авторов некоторых разделов, например статей на политические темы. Но ведь если мы говорим о политике, мы всегда пристрастны, не правда ли? И информация в Википедии об одном и том же в политике, опубликованная на английском и русском языках, может отличаться, потому что к нему по-разному относятся в США и России, а статьи на своих родных языках писали американец и русский.

В любом случае мы можем сказать, что Интернет стал причиной информационной революции, или революции данных. Теперь у нас есть легкий и быстрый доступ к огромным массивам самой разной информации. Одной из составляющих информационной революции стали поисковые системы. И это тоже невероятно впечатляющая демонстрация работы *Big Data*. Это невероятно быстрая обработка гигантских массивов данных по запросу. Скорость современных технологий поражает, как и их возможности.

Google — это американская транснациональная публичная корпорация. Компания была зарегистрирована Ларри Пейджем и Сергеем Брином в 1996 году. *Google* управляет более чем миллионом серверов в центрах обработки данных по всему миру. Каждый день они обрабатывают свыше миллиарда поисковых запросов со всего мира и 24 петабайт пользовательских данных (1 петабайт = 1 миллиону гигабайт). Главный продукт компании — это поисковая система, и, слыша или видя слово *Google*, мы в первую очередь думаем не о компании (о компании мы чаще всего вообще не думаем), а об очень удобной и мощной поисковой системе. Слово теперь используется и как глагол, обозначающий поиск с ее помощью. Эта поисковая система охватывает более 60% рынка. Поиск может вестись на 191 языке. Ежедневно регистрируется около 50 миллионов запросов и индексируется более 8 миллиардов веб-страниц. У компании имеются и другие продукты. Например, почтовый сервис (*Gmail*) и социальная сеть (*Google+*). Основной сайт *Google* (*google.com*) является самым посещаемым сайтом Интернета, показателя 1 миллиард уникальных пользователей в месяц компания добилась в мае 2011 года. Международные сайты компании (например, *google.co.uk*) входят в первую сотню по посещаемости. Это относится и к сайтам других сервисов *Google* — *YouTube*, *Blogger*. *Google* называют и самым мощным, и самым дорогим брендом в мире. Это бренд больших данных.

Если говорить о миллиардах страниц, то *Google* покрывает от 47 до 49 миллиардов. Но это не означает, что в Интернете около 50 миллиардов страниц. Их гораздо больше, просто у *Google* нет к ним доступа, так как они являются коммерческой тайной или закрыты из соображений безопасности. Неужели при

каждом запросе система проводит для нас поиск по всем этим миллиардам страниц, находящихся в ее распоряжении? Нет, это не так. У *Google* имеются свои поисковые роботы, которых также называют веб-кроулерами или веб-пауками, которые постоянно рыскают по Всемирной паутине в поисках нового материала. Он заносится в базу данных поисковика и прикрепляется к каким-то указателям в базе данных. Только в одном алфавитном указателе *Google* 100 петабайт данных.

Чтобы найти ответы на вопросы, нам нужно так их формулировать, чтобы слова совпали со словами на веб-страницах. Стоит нам начать печатать вопрос — тут же выходят подсказки и варианты для выбора. Почему одни страницы в ответ на наш вопрос выходят первыми, а другие, на которых мы на самом деле находим ответ, оказываются гораздо дальше? Во-первых, владельцы сайтов могли за это заплатить. Во-вторых, вначале показывается недавняя или самая последняя информация. В-третьих, *Google* вначале показывает те сайты, которые система считает высококачественными, или они связаны с другими важными сайтами.

В памяти *Google* хранятся запросы каждого пользователя, и ответы на его запросы структурируются в соответствии с его прошлыми запросами. Уже немало людей пытались разобраться с рейтинговым алгоритмом *Google*, чтобы поднять свои сайты в их рейтинге. Но инженеры *Google* всегда находят ответ, чтобы противостоять такой «оптимизации» поисковой системы, и усовершенствуют и ее, и рейтинговый алгоритм. Пока победить *Google* никому не удалось.

Конечно, *Google* используют не только и не столько для развлечений. Многие люди используют поисковую систему, чтобы что-то купить, в образовательных, исследовательских целях, для своего бизнеса. Но согласитесь, что это здорово — иметь возможность быстро и легко найти то, что нужно. В любых целях.

Звук и картинка

Вначале цифровой мир оказал серьезное влияние на музыку. Цифровые данные гораздо проще переписывать, чем сохраняемые на физических носителях аналоговые данные. Шоком был переход от пластинок к кассетам, потом появились CD, а цифровые файлы стали частью информационной революции. Такой файл можно переписать и представить миру за несколько секунд, например, с помощью сервиса бесплатного обмена типа *Napster*. Из-за таких сервисов доходы звукозаписывающих компаний упали невероятно.

Что представляет собой цифровая дистрибуция? В настоящее время есть интернет-магазины, которые работают по классической схеме — продают альбомы или песни популярных исполнителей. Таким образом работают *iTunes*, *Google Play*. Вы покупаете альбом или песню, платите за них один раз и закачиваете себе на любые устройства. С другой стороны, есть так называемые стриминговые сервисы. Здесь вы не платите за конкретный альбом или конкретную песню, а оплачиваете подписку. То есть платите определенную сумму (небольшую) один раз в месяц и получаете на этот месяц доступ ко всему каталогу мировой музыки. Если вы подписку не продлеваете, вас отключают от каталога. К стриминговым сервисам относятся, например, *Apple Music* и платформа *Boom*.

Подобное положение дел очень помогает молодым неизвестным исполнителям без связей и денег. Если они сделают актуальную и качественную работу, то могут попасть в топ и стрим-сервисов, и интернет-магазинов, а потом монетизировать успех. Достаточно двух хитов в год. Если они звучат со всех интернет-площадок, включая *YouTube*, *Apple Music* и других, то реально получать порядка 50 тысяч долларов в месяц. А потом уже можно живьем выходить к публике и получать не меньшие гонорары, чем звезды шоу-бизнеса. А можно не тратить на масштабные шоу и турне.

Можно выступать только в Интернете, хотя это зависит от аудитории. Например, популярность стриминговых сервисов постоянно растет, а для поколения 15-20-летних они заменили радио, добавьте к этому реальную работу законов против пиратства. Можно сказать, что сейчас конфликт отцов и детей — это конфликт старых звезд шоу-бизнеса и звезд стриминга. Исполнители новой волны стали новыми героями слушателей. Пока на пике хип-хоп, рэп, поп-музыка, лексика улиц, которые интересны их 15-20-летней аудитории. Что будет дальше, когда они все повзрослеют? Пока мы ответа не знаем. Но эта молодежь, можно сказать, родилась с компьютерами и смартфонами, это другое поколение, которое живет в Интернете и мире больших данных. Вероятно, они и дальше будут жить в этом мире.

Музыкальный бизнес довольно быстро понял, что если сделать законные действия легкими и доступными, это будет гораздо более эффективным, чем трата огромных сумм денег на борьбу с пиратством. Вначале эта борьба активно велась, но как только закрывался один пиратский сайт, в других местах появлялись три новых. Вначале было предоставлено легкое и дешевое скачивание с сервисов типа *iTunes*, потом появился музыкальный стриминг от *Spotify* и их конкурентов, и среднестатистический любитель музыки понял, что нет оснований нарушать закон. Конечно, всегда останутся и пираты, и пользователи пиратских сайтов, но их можно рассматривать как неизбежные потери, которые есть в любом бизнесе. Большинство, по крайней мере в США, скачивает музыку легально. И снова главной причиной эффективного использования больших данных стало удобство. Компании используют силу и гибкость технологии *Big Data*. И это помогает им не нарушать закон и не платить штрафы.

Открывающиеся перед нами сегодня возможности поражают. Теперь мы можем смотреть фильмы не только по телевизору, но и через компьютер и смартфон. Мы можем в любое время посмотреть выпуски любимого ток-шоу, в любое время включить и в любое время остановить просмотр, если возникли неотложные дела. Это невероятно! Раньше так можно было останавливать *DVD*. Но на одном *DVD* хранится только несколько гигабайтов данных. Представьте, сколько данных обрабатывают стриминговые сайты, если обслуживают миллионы клиентов. Это невероятно большие данные. Большие данные в промышленных масштабах.

Пока стриминговые сервисы не получили такого широкого распространения, как традиционное вещание. Данные в США очень сильно отличаются от штата к штату, а в Великобритании примерно четверть домохозяйств подписана на лидера этих услуг *Netflix*. Но людей все больше привлекают сервисы «по требованию», а не традиционная программа передач, которая формируется на неделю вперед. Мне кажется, что через 20 лет останутся только часы прямых включений с мест событий. И крупнейшие телекомпании мира обязательно окажутся в тренде, отказавшись от устаревших способов предоставления доступа к телепрограммам. Каждый из нас, скорее всего, будет их заказывать — те, которые хотим, и тогда, когда хотим. Не любишь спорт — у тебя в программе не будет спорта. Любишь футбол — будет футбол. Любишь хоккей — будет хоккей.

Видеоконтент с самого момента своего появления привлекал пиратов. Пиратство будет всегда, но его можно свести к минимуму, то есть сделать легальный стриминг и скачивание настолько легкими и удобными, насколько возможно. У компаний типа *Netflix* это получилось. Они с самого начала сделали свой продукт доступным через большое количество платформ и добавили много полезных функций, например, остановку просмотра фильма или шоу в любое время и запуск с того же места после перерыва, причем это можно сделать и с другого устройства. В этом и

заключалось отличие от «старой гвардии». А сама возможность появилась с появлением больших данных.

Книги

Но если видеоиндустрия с трудом переходила и все еще продолжает переходить в мир больших данных, что говорить о книгоиздании? Книги в виде цифровых данных появились значительно позже цифровых музыкальных и видеофайлов. У книгоиздателей уже давно сформировалась модель получения максимальной прибыли от продаж каждой книги. Вначале на рынок традиционно выпускается вариант в твердой обложке. Огромного количества экземпляров в твердой обложке не издашь, но даже первый тираж всегда приносит хорошую прибыль. Ведь есть люди, которые с нетерпением ждут книги любимого автора и готовы платить за то, чтобы получить новинку как можно быстрее. Затем, для остальных читателей, не готовых платить лишнее за скорость, издается вариант в мягкой обложке. В США период между изданием в твердой обложке и мягкой обложке может составлять и обычно раньше составлял один год.

Изначально большинство издателей рассматривали электронную книгу как некий вариант книги в мягкой обложке и задерживали ее выход на несколько месяцев, чтобы выжать максимум с твердой обложки. Они усложняли жизнь читателям, вместо того чтобы использовать подход, который требуют *Big Data*, — сделать доступ к легальной версии как можно проще. И в индустрии, ранее не сталкивавшейся с пиратством, оно расцвело пышным цветом. Но проведенные исследования показали, что нелогично задерживать продажу электронных книг, потому что рынок книг в твердой обложке и рынок электронных книг практически не пересекаются. Фактически это показал конфликт *Amazon* с одним из издателей. Какое-то время они одновременно начинали продажу версий книг в твердой обложке и электронной версии, а в 2010 году из-за конфликта предложили посетителям сайта только версию в твердой обложке. Отсутствие электронной версии никак не повлияло на продажу книг в твердой обложке — те, кто покупал их обычно, купил их и в тот раз, те, кто не покупал, не купил.

Проведенное чуть позже исследование показало, что задержка выхода на рынок электронной версии книги, наоборот, существенно снижает продажи. Похоже, что покупатели электронных книг хотят получить их как можно раньше — одновременно с выходом книги в твердой обложке. Так что традиционный подход к книготорговле, существовавший до эры больших данных, себя больше не оправдывает. Он снижает продажи.

Через какое-то время большинство издателей это поняли и стали одновременно выпускать версию в твердой обложке (если она вообще публикуется) и электронную версию. Но даже сейчас остается ряд издателей, преимущественно представителей старшего поколения, которые не понимают современный рынок. У некоторых цена на электронную версию бывает всего лишь чуть-чуть ниже, чем на книгу в твердой обложке, и снижают они ее только после выхода книги в мягкой обложке. Это опасная стратегия, она на руку пиратам, которые прекрасно понимают «цифровых» покупателей.

Неудивительно, что доминирующим игроком на рынке электронных книг является *Amazon*, контролируя 75% рынка США и 95% рынка Великобритании. *Amazon*, как и *Netflix*, собирает данные о своих покупателях для того, чтобы правильно размещать продукцию на своем сайте и облегчить пользователям покупку книг. У них преимущество перед книгоиздателями, потому что они имеют прямой контакт с покупателями из своей базы. А в мире больших данных отсутствие прямого контакта ставит вас в невыгодное, а то и опасное положение.

Но книгоиздатели могут получать информацию от своих читателей непрямым образом, используя большие данные. И они используют эту возможность, пытаясь понять, что сделало бестселлерами некоторые из опубликованных книг и что не сделало другие.

Американский ученый Мэтью Джокере и бывший издатель Джоди Арчер даже написали об этом книгу «Код бестселлера». Издатели притворяются, будто умеют определять потенциальный бестселлер среди многочисленных рукописей, поступающих в издательства. Но никто не ожидал такого успеха «Гарри Поттера» и «50 оттенков серого». Влияет слишком много факторов, которые еще и взаимодействуют друг с другом. Чтобы сделать прогноз, нужно учитывать не только характеристики самой книги, но и общие тенденции, и социальные факторы. Джокере и Арчер решили, что технологии *Big Data* помогут им справиться с трудной задачей прогнозирования. Они разработали программное обеспечение для анализа огромного количества бестселлеров и нахождения общих черт. По их мнению, на основании полученных данных можно анализировать недавно написанные книги для определения потенциальных бестселлеров. То есть большие данные становятся судьей читательских вкусов.

Но... Их алгоритм отмечал использование определенных слов, построение фраз, фактически он находил книги, которые мы охарактеризуем как «хорошо написанные». Но бестселлерами становились и те книги, которые, по мнению разработанного программного обеспечения, относились к «ужасно написанным». Например, этот алгоритм не определил в списке бестселлеров книги Дэна Брауна и «50 оттенков серого». Да, многие книги попали в список бестселлеров *New York Times*. Да, они хорошо написаны. Но литературные критики обычно основывают свои оценки на других факторах.

Нельзя написать книгу, которая понравится всем. Джокере и Арчер дают советы тем, кто хочет написать бестселлер на основании изучения списков бестселлеров и выкладок своего программного обеспечения. Например, следует не писать фантастику, избегать чисто британских тем и описания частей тела и полового акта. Я считаю, что их алгоритм выявляет плохие книги (с описанием полового акта и частей тела отдельно), но не способен выявить настоящие бестселлеры.

Количество проанализированных ими книг впечатляет, но они не сообщают нам, сколько раз алгоритм ошибся. Сколько раз он назвал бестселлерами книги, которые ими не были? Сколько настоящих бестселлеров пропустил? Насколько мне известно, издатели не восприняли исследование Джокерса и Арчера всерьез и не полагаются на него. А вот на технологии *Big Data* полагаются во все большей и большей степени.

Социальные сети

Очень мало кто живет в изоляции от других людей. Конечно, и в наше время есть отшельники, но социальные сети существовали и до появления Интернета. У нас у всех есть круг друзей, есть родственники, коллеги, соседи, есть люди, с которыми мы просто здороваемся, есть те, кому мы просто киваем. С кем-то мы каждый день едем на работу в электричке, но не решаемся первыми заговорить. Но технологии *Big Data* подняли социальные сети и социальные связи на новый уровень. О таком никто не мог и подумать и тем более предположить.

В наше время социальные сети называют по-разному, в частности рассадником фейковых новостей. Об этом стали активно говорить после избрания Дональда Трампа президентом США. Фейковыми новостями или просто фейками называют намеренное распространение дезинформации в социальных медиа и традиционных СМИ для

получения выгоды, в особенности финансовой и политической, увеличения трафика и прибыли. Фейковые новости имеют много общего с желтой прессой и политической пропагандой, которые существуют уже давно и сыграли свою роль в человеческой истории.

В бумажных СМИ тоже, бывало, попадались непроверенные новости, против желтых изданий возбуждались многочисленные иски. Но все-таки раньше печатные издания старались проверять публикуемую информацию, откровенную ложь отсеивали. В социальных медиа такого фильтра нет. А мы имеем склонность верить тому, что нам сообщают друзья по социальным сетям, и скорее им, чем информации из какого-то далекого источника.

Проблема фейковых новостей связана с появлением больших данных и легкостью распространения информации в социальных сетях и Всемирной паутине в целом в сравнении с физическим миром. Если вы получаете какую-то шокирующую, ошеломляющую новость от вашего френда, от вас требуется всего пара кликов, чтобы поделиться ею с огромным количеством людей. Сейчас информация распространяется как вирус — один человек может «заразить» многих, именно так в человеческой истории начинались эпидемии. Виртуальную эпидемию начать еще проще, а остановить невозможно.

Сейчас, к счастью, есть способ как-то ограничить распространение таких новостей — есть специальный сайт СтопФэйк. Как он нем пишут на самом сайте, цитирую:

«Мы находим и проверяем сообщения, которые выглядят неточными или ложными. Выясняем, можно ли их подтвердить, разъяснить или опровергнуть.

Находим первоисточники — публикации в интернете и офлайн-СМИ, высказывания или заявления людей, упомянутых в сообщениях. Выясняем, какой резонанс они вызвали и как восприняли их читатели.

Исследуем спорные доказательства, сравниваем разные теории и изучаем подтверждения.

Анализируем иллюстрации, фотографии и видео, которыми делятся пользователи.

Обращаемся к экспертным оценкам и берем официальные комментарии.

Публикуем факты.».

В настоящее время наибольшее беспокойство вызывает влияние социальных сетей (и соответственно больших данных) на способность людей (в первую очередь молодых людей) сосредоточивать внимание и взаимодействовать с другими. Жизнь современных молодых людей, можно сказать, проходит в социальных сетях и Интернете. Они регулярно заглядывают в смартфоны или вообще не отрывают от них глаз. Это даже приводит к несчастным случаям. В среднем современный молодой человек проверяет телефон 100 раз в день! Многие заглядывают в свои аккаунты или почту каждый раз, когда просыпаются ночью.

Да, социальные связи необходимы, но не в такой же степени! Масса молодых людей сейчас очень тщательно следит за тем, что они выставляют в Сеть, и еще более тщательно за тем, сколько лайков принесли их фотографии. Они постоянно проверяют *Instagram*, *Facebook* и другие сети на предмет количества лайков на свои недавние посты, они проверяют количество полученных ими лайков и количество лайков у других людей, сравнивают количество подписчиков. Это болезненная зависимость! Ученые разных стран, проводившие соответствующие исследования, обнаружили прямую связь между использованием социальных сетей и ощущением себя несчастными. То есть чем больше времени люди проводят в сетях, тем они менее счастливы. Связь здесь не односторонняя, а двухсторонняя. Счастливый человек не

будет проводить много времени в социальных сетях, может ими воспользоваться, чтобы кого-то найти, а может быть, вообще не зарегистрирован нигде.

Но все равно социальные сети — это большой плюс. Это развлечение стало возможным только после появления технологии *Big Data*. Я сама регулярно пользуюсь социальными медиа. Я работаю дома, и социальные сети — это один из способов общения с другими людьми. Я участвую в различных форумах, иногда мне требуется консультация специалиста, хочется что-то обсудить с коллегами, просто посоветоваться.

Big Data все больше и больше вторгаются в нашу жизнь.

Нужно стараться соблюдать разумное равновесие между удовольствием, которое дают те же социальные сети, и формирующейся зависимостью от них. Нужно осознавать, что большие данные могут принести большие проблемы, и сознательно прилагать усилия, чтобы не уйти из мира реального в мир виртуальный. Можно взять лучшее из обоих миров, если вы осознаете возможные проблемы и контролируете себя. Человек должен контролировать большие данные, а не наоборот.

Если же мы действуем правильно, то большие данные помогают нам решать проблемы. Если нет, то они их только добавляют. Так давайте сделаем так, чтобы они приносили нам пользу и удовольствие.

Список использованных источников:

1. Просто Big Data. — СПб.: Страта, 2019. — 148 с.