ТЕРМИНЫ БОЛЬШИХ ДАННЫХ

Е.А. Волнухина

Термин «большие данные» ввел редактор журнала Nature Клиффорд Линч еще в 2008 году в спецвыпуске, посвященном взрывному росту мировых объемов информации. Хотя сами «большие данные» существовали и ранее. По словам специалистов, к категории Big data относится большинство потоков данных свыше 100 Гб в день.

Сегодня под этим простым термином скрывается всего два слова – хранение и обработка данных. В современном мире Big data – социально-экономический феномен, который связан с тем, что появились новые технологические возможности для анализа огромного количества данных.

Огромные объемы данных обрабатываются для того, чтобы человек мог получить конкретные и нужные ему результаты для их дальнейшего эффективного применения. Фактически Big data – это решение проблем и альтернатива традиционным системам управления данными.

«Большие данные» – относительно молодая область информационных технологий. И как в большинстве новых областей терминология еще не полностью устоялась и сформировалась. Однако можно выделить ряд терминов, которые позволят лучше понимать эту новую область.

DATA SCIENCE

Русский аналог – «наука о данных». Это наука о методах анализа данных и извлечения из них ценной информации.

Data science как академическая дисциплина формируется с начала 2010-х. Чтобы стать специалистом в этой области, необходимо прежде всего быть отличным математиком – знать матмоделирование, матстатистику, комбинаторику, теорию графов и многое другое. Ну и, конечно, уметь программировать. Надо заметить, что спрос на дата-сайентистов сильно превышает предложение (особенно в России).

МАШИННОЕ ОБУЧЕНИЕ

Русское словосочетание «машинное обучение» используется так же часто, как английское «machine learning» (что-то вроде «мэшин лернинг»).

«Именно благодаря машинному обучению поисковая машина понимает, какие результаты (и рекламу) показывать в ответ на ваш запрос. Когда вы просматриваете почту, большая часть спама проходит мимо вас, потому что он был отфильтрован с помощью машинного обучения. Если вы решили что-нибудь купить на Amazon.com или заглянули на Netflix в поисках фильма, система машинного обучения услужливо предложит варианты, которые могут прийтись вам по вкусу. С помощью машинного обучения Facebook решает, какие новости вам показывать, а Twitter подбирает подходящие твиты», – с этих слов начинается книга «Верховный алгоритм» исследователя искусственного интеллекта Педро Домингоса.

DATA MINING

Принято говорить «дата/дэйта майнинг», «майнить» – извлекать данные, «намайнить» – извлечь.

Датамайнингом называют как технологии, так и процесс обнаружения в сырых данных неизвестной и полезной информации. Основу data mining составляют всевозможные методы классификации, моделирования и прогнозирования.

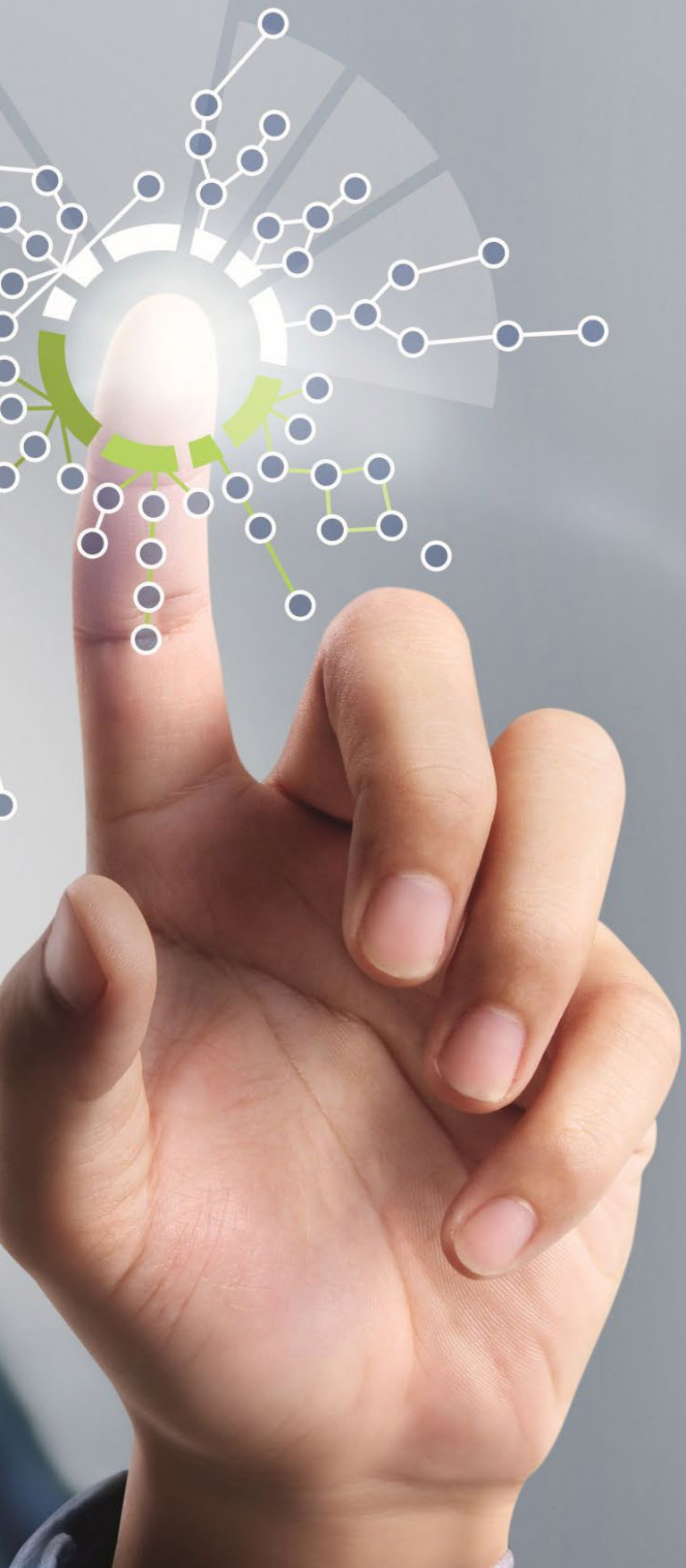
В научный обиход термин ввел израильский математик Григорий Пятецкий-Шапиро еще в 1989 году.

ОБЛАКА

Говорят как «облако/облачный», так и «cloud» (например, «cloud computing» – облачные вычисления).

Держать в голове все задачи на день, месяц, год не очень-то удобно, поэтому мы записываем их в блокнот или заносим на виртуальную доску.

Точно так же наш компьютер не может хранить на своем диске сотни гигабайт видео, фоток и музыки – их приходится закачивать на такие сервисы, как Google Drive или Яндекс.Диск.



Мы имеем постоянный доступ к своим данным через интернет, но физически они находятся на виртуальных серверах соответствующих компаний. При этом пользователь платит лишь за место в хранилище, а это гораздо дешевле аренды целого сервера. Естественно, для работы с большими данными «облака» просто необходимы.

СУПЕРКОМПЬЮТЕР

Это слово начало входить в русский язык еще в конце 1960-х, когда в СССР появился первый суперкомпьютер БЭСМ-6, способный выполнять 1 млн операций в секунду.

Речь идет о вычислительной машине, значительно превосходящей по техническим параметрам и скорости обработки данных обычные персоналки. Как правило, она представляет собой систему высокопроизводительных компьютеров и используется для решения задач в самых разных областях науки и технологий: от разработки атомного оружия до моделирования новых лекарств. Самые мощные российские суперкомпьютеры «Ломоносов» и «Ломоносов-2» находятся в Московском государственном университете им. М.В. Ломоносова.

ИНТЕРНЕТ ВЕЩЕЙ

Популярен и русский вариант, и английский – «internet of things», а также аббревиатура IoT.

Вслед за компьютерами и смартфонами в Сеть вышли фитнес-трекеры, чайники, стиральные машины, телевизоры, датчики и сенсоры.

И все же интернет вещей – это не включение света посредством брюк или удаленный спуск воды в унитазе через смартфон. Есть масса примеров внедрения технологий IoT в медицине, спорте, сельском хозяйстве, промышленности. Например, BigBelly – урна, которая предупреждает сотрудников городской службы по уборке мусора, что ее пора опорожнить. Локальная сеть анализирует данные, полученные от каждой урны, что позволяет планировать частоту вывоза бытовых отходов.

В 1990 году выпускник Массачусетского технологического института, один из отцов протокола TCP/IP Джон Ромки создал первую в мире интернет-вещь. Он подключил к Сети свой тостер. Термин «интернет вещей» появился в 1999 году. Ожидается, что в 2020-м IoT объединит более 30 млрд устройств.

Подготовлено
по материалам сайтов
kot.sh, rb.ru, postnauka.ru ■