

# Интеллектуальный термостат



**Беседовал Отабек Арзикулов**



«Первопроходцами «Больших данных» должны быть стартапы. Именно они могут позволить себе эксперименты с технологиями, которые по определению связаны с огромными рисками. Из ста выживает один — это основной принцип стартапов и венчурного капитализма в целом. Ни государство, ни устоявшийся бизнес на такие эксперименты не способны», — рассказывает Дмитро Карамшук, специалист в области обработки данных Королевского колледжа Лондона.

— **Сегодня термин «Большие данные» уже порядочно намозолил глаза. Так ли прекрасна технология, как ее преподносят?**

— «Большие данные» не совсем технология. Это скорее общее название тенденции в IT, которая подразумевает монетизацию аналитики, основанной на данных. Интеллектуальный термостат, база данных финансовых транзакций, социальная сеть, интернет-магазин — каждый по-своему определяет, что для него «Большие данные» и каким образом их можно монетизировать. Современные возможности анализа данных, в том числе тех, которые люди оставляют добровольно в Интернете, — настоящая кладезь информации. На основе их можно делать целые маркетинговые исследования, изучать поведение потребителей, прогнозировать будущий спрос, тенденции. Например, проанализировав количество «чекинов» в Foursquare\*, можно определить наиболее выгодные места для открытия ресторана или магазина. Или получить ценную информацию о предпочтениях зрителей, изучив историю просмотров телепередач.

На самом деле большинство данных доступны и открыты. Вопрос в создании алгоритмов, которые позволят проанализировать и получить ценные результаты. Как говорил Уинстон Черчилль, истинная гениальность — это умение оценивать непонятную, рискованную и противоречащую информацию.

Сама идея собирать данные и извлекать финансовую выгоду из них стара как мир. Уникальность ситуации, которая сложилась в последнее время, заключается в том, что

появились доступные инструменты хранения и обработки сколь угодно больших объемов данных. Речь идет о Cloud Computing («облачные» вычисления). К примеру, загрузить несколько терабайт данных и обработать сложный алгоритм в течение пары часов на сотни компьютеров Amazon Web Services будет стоить порядка \$10. Еще лет 10 назад это казалось фантастикой.

**Идея собирать данные и извлекать из них финансовую выгоду стара как мир. Уникальность сегодняшней ситуации — в наличии доступных инструментов хранения и обработки сколь угодно больших объемов данных**

На мой взгляд, именно доступный Cloud Computing стал толчком к развитию целого семейства технологий, которые сейчас подразумевают, когда говорят о «Больших данных». Это и алгоритмы параллельной обработки данных, и средства визуализации, и множество тематических моделей поиска закономерностей и прогнозирования — то, что называется predictive analytics.

— **В каких сферах «Большие данные» наиболее актуальны?**

— В современной экономике правила диктует рынок. Результаты ждут там, где его проще всего добиться и где это наиболее выгодно. Яркий пример массивной монетизации данных — это, конечно же, Интернет. Google в самой своей сути Big Data — компания, которая построила бизнес вокруг сбора и обработки данных со всех сайтов и трендов

в Интернете. Большинство интернет-гигантов, включая Facebook, Yahoo, Twitter, Google, Pinterest и т.д. монетизируются с помощью сложных алгоритмов анализа громадных данных о поведении людей в Интернете и предсказаниях по количеству кликов на баннеры. Это колоссальный рынок, на который работают сотни и тысячи технологий и, к слову, научных умов мира.

Индустрия финансовых технологий — еще одна область, где базы данных существовали со времен самых первых компьютеров. Этот рынок более консервативен и менее открыт, нежели интернет-компания, но он также в центре революции «Больших данных».

Особенно интересно проникновение «Больших данных» в сферы, которым компьютерная обработка данных не так свойственна. К примеру, энергетика, городское планирование или здравоохранение. Компания Apple недавно анонсировала платформу для разработки приложений, сбора и обмена данными в области здравоохранения, фитнеса и здорового образа жизни. Представьте точность медицинского диагноза, поставленного с учетом симптомов у миллионов пользователей приложениями для астматиков, «сердечников», людей, страдающих раком.

## Открытия, сделанные благодаря «Большим данным»

Большое количество чипов в iPhone делают невозможным вложение денег в Интернет по сравнению с мобильными устройствами, но получают выгоду от них: с помощью других устройств формируются новые устройства. Это касается сигналов и интерпретации данных, и стимулирует их разработку компьютерными интерфейсами. В конечном итоге, чем больше информации, тем больше возможностей для компаний.

1

3

Многие исследования в области биомедицины и генетики позволяют выявить новые связи между генами, которые связаны с определенными заболеваниями. Исследования показывают, что гены могут влиять на развитие болезни, а также на реакцию организма на лечение. Это позволяет врачам лучше понимать механизмы болезни и разрабатывать новые методы лечения.

2

Полный анализ генома человека в 2003 году на Гемплтонской конференции по геномике и геномике, чтобы быстрее выявить, где находятся гены, связанные с болезнями. Американские и европейские ученые начали предсказывать, по какой причине они заболевают. Так, Twitter использует свои данные для выявления тенденций в работе, чтобы лучше понять поведение пользователей и улучшить сервис.

— *Big Data* — технология не из дешевых. И если для развитых стран вопрос ее внедрения может оказаться делом копейным, то как быть странам с

Первопроходцами должны быть стартапы. Стартапы могут позволить себе эксперименты с технологиями, которые по определению связаны с огромными рисками. Из ста

**Если банк собирает информацию о финансовых транзакциях клиентов, то она должна быть защищена вне зависимости от того, занимается этот банк аналитикой данных или нет**

**развивающейся экономикой? Кто должен быть первопроходцем в технологиях «Больших данных»: государственные структуры или же частные компании?**

выживает один — это основной принцип стартапов и венчурного капитализма в целом. Ни государство, ни устоявшийся бизнес на такие эксперименты не способны.

Рынок огромен, равно как и спектр предлагаемых услуг и продуктов. Хорошей новостью является то, что многие компании, включая таких гигантов, как Facebook, Google, Microsoft, Oracle, практикуют Open Source в качестве модели разработки и поддержки программных систем. Поэтому софтверная «экосистема» развивается открыто и доступно. Существует множество решений, которые совершенно бесплатны и поддерживаются огромным сообществом программистов.

Есть широкий спектр открытых систем для параллельной обработки данных, их визуализации, математического моделирования и т.п. Есть также и коммерческие продукты, которые собирают весь этот «салат» в красивую и удобную упаковку.

**— Но даже стартапы, идущие на риск, задаются вполне закономерным вопросом: каков уровень информационной безопасности при использовании «Больших данных»?**

— На мой взгляд, этот вопрос связан больше со сбором и хранением информации в целом. Если банк собирает информацию о финансовых транзакциях своих клиентов, то она должна быть защищена вне зависимости от того, занимается ли этот банк какой-либо аналитикой данных или нет. Что касается безопасности алгоритмов, не думаю, что такая проблема стоит остро. Если бы кто-либо и пытался взломать алгоритм, то с той лишь целью, чтобы понять, как он работает, и использовать его в своих целях, поскольку алгоритмы

обычно являются интеллектуальной собственностью и коммерческим преимуществом компаний.

Да, существует проблема приватности. Чтобы сделать полезный алгоритм, необходимо понять, каким образом пользователи ведут себя в той или иной ситуации. Соответственно, чтобы получить хороший анализ, необходимо

**Такие концепции, как «интернет вещей», «умные» города уже долгое время развиваются за кулисами и могут в любой момент стать прорывом**

рассказать системе многое о себе. К примеру, в магазине смогут порекомендовать хороший продукт, только если будут знать, что вы покупаете, а что нет. Или приложение на вашем телефоне по запросу сможет предложить хороший ресторан, только если будет знать информацию о вашем местоположении и ресторанах, которые вы посещали до этого.

Думаю, ключевой вопрос в том, насколько мы готовы делиться своими приватными данными взамен на сервис, получаемый от системы.

**— С Big Data IT-технологии уже, можно сказать, почти достигли своего апофеоза. Ждут ли нас новые IT-революции после эры «Больших данных»?**

**WESTMINSTER**  
INTERNATIONAL UNIVERSITY IN TASHKENT  
An Accredited Institution of the University of Westminster (UK)

Международный Вестминстерский Университет в Ташкенте объявляет о наборе на магистерские программы:

**МЕЖДУНАРОДНЫЙ БИЗНЕС И МЕНЕДЖМЕНТ (MAIBM)**  
**МЕЖДУНАРОДНОЕ КОММЕРЧЕСКОЕ ПРАВО (LLM)**

Программы магистратуры имеют следующие преимущества:

- программы позволяют обучение без отрыва от производства (занятия проходят в вечернее время с 18:30)
- набор студентов проводится ежегодно осенью и зимой
- успешно закончившие студенты получают диплом Вестминстерского Университета (Великобритания)

Последний срок подачи заявления на осенний набор:  
4 сентября 2015 г.  
[www.wiut.uz/en/courses/postgraduate](http://www.wiut.uz/en/courses/postgraduate)

(+986 71) 238 74 00/44; [www.westminster.uz](http://www.westminster.uz); [info@wiut.uz](mailto:info@wiut.uz)  
Услуги лицензированы

— Это не первая революция IT. На моей памяти dot-com bubble в начале 2000-х; web 2.0 в середине 2000-х; smatrphone в 2008-м. Теперь вот Big Data. Думаю, еще пару лет «Большие данные» будут в тренде, но неизбежно интегрируются в общий «компот» и уступят место новой мании.

Наиболее перспективными тенденциями сейчас являются концепции internet of things («интернет вещей»), smart cities («умные» города) и технологии

в области энергетики. Хотя все из этих областей, как, впрочем, и Big Data в свое время, уже долгое время развиваются за кулисами и могут в любой момент стать прорывом.

#### Примечание:

\*Foursquare (сокращенно: 4sq) — социальная сеть с функцией геопозиционирования, предназначенная в основном для работы с мобильными устройствами. Пользователи отмечают (check-in) в различных заведениях с помощью мобильной версии веб-сайта, sms-сообщения или же специального приложения.