

История понятия «большие данные» (Big Data): словари, научная и деловая периодика

В статье кратко рассмотрена история понятия «большие данные» (Big Data) и уточняются некоторые аспекты появления и развития понятия в словарях, научной и деловой периодике. Что важно – выявлена фактологическая неточность относительно зарождения термина, которая реплицируется во многих отечественных публикациях (в том числе научных). В связи с чем дается уточнение по поводу появления и бытования термина в российских научных, научно-популярных и деловых публикациях. Статья также предлагает посмотреть немного шире на эволюцию представлений о «больших данных» и возможности для гуманитарных исследователей.

Ключевые слова: большие данные, Big Data, digitalhumanities, media-studies, цифровая гуманитаристика, терминология, цифровая среда, интернет-исследования.

Проблема определения, понимания и истории термина «большие данные» напрямую влияет на возможности использования в гуманитарных исследованиях методик и инструментов, предлагаемых данным обширным направлением. Сегодня знать принципы работы с большими массивами данных – ключевой навык практика-журналиста: существуют специализированные направления по исследованиям данных (datascience) и журналистике данных (datajournalism). Развитие компетенций в этих направлениях сложно представить без ориентирования в базовых терминах и понятиях.

Парадокс: несмотря на принадлежность к точным наукам, цифровой среде и IT-сфере, понятие «большие данные» («Big Data») не имеет четкого определения. Множество авторов, организаций и сообществ пытаются по-разному интерпретировать понятие. Более того, от изначального непонимания сути явления и суженного

представления о возможностях сферы больших данных в дальнейшем развивается методическая слепота и инструментальная беспомощность исследователя.

Поскольку термин англоязычный, логично начать обзор с авторитетных англоязычных словарей и ресурсов.

В июне 2013 г. Оксфордский английский словарь (The Oxford English dictionary, OED) добавил определение термина «Big Data»¹, которое можно перевести так: «Данные очень большого размера, как правило, в том смысле, что представляют серьезные трудности в материально-техническом обеспечении по манипуляциям и управлению ими; (также) направление вычислений с использованием такого типа данных»². Здесь доминирующая характеристика в определении – это операционные трудности (важнее «как?», а не «зачем?»).

Что интересно, сайт oxforddictionaries.com, который фокусируется на текущих, актуальных значениях слов и практическом употреблении, дает следующее определение «Big Data»: «Очень большие массивы данных, которые могут быть проанализированы с помощью компьютеров, чтобы выявить закономерности, тенденции и взаимосвязи, в особенности в отношении поведения людей и их взаимодействий»³. Технологический термин в данной интерпретации тесно связан с социально-гуманитарной компонентой, т. е. по отношению к первоначальному определению наблюдается эволюция понятия в сторону его гуманизации, а содержание понятия меняется от «серьезных трудностей» к практическим возможностям анализа человеческой деятельности.

Сайт Кембриджского словаря дает несколько иное определение, также отсылающее к центральной роли человека, но в неразрывной связке с технологиями: «Очень большие массивы данных, производимые людьми в процессе пользования интернетом, которые могут быть сохранены, поняты и использованы только с помощью специальных инструментов и методов»⁴.

Поиск по электронным версиям российских академических и энциклопедических словарей, а также сервисов (gramota.ru, например) результатов по запросам «Big Data» и «большие данные» не выдает.

Предыстория широкого употребления понятия еще более интересна. В Рунете популярна версия, что термин «Big Data» ввел в научный оборот редактор научного журнала «Nature» Клиффорд Линч в специальном номере от 3 сентября 2008 г., посвященном драматическому росту объемов информации⁵. Эта версия появилась и воспроизводится в статьях (в том числе научных)⁶ и бизнес-презентациях с подачи издания «Открытые системы. СУБД»

с 2011 г.⁷ Распространению информации также могло послужить включение этой версии в русскоязычную статью Википедии о «больших данных»⁸.

Однако даже самый поверхностный анализ отсылает к более ранним референциям. Так, раз уж упомянута русская версия Википедии, ее англоязычная версия отсылает к исследователю Джону Машей, специалисту в области информатики (John Mashey, computerscientist), который популяризовал термин в 1990-х гг.⁹

Исследователь, предприниматель и автор блога whatsthebigdata.com Гил Пресс (Gil Press) в публикации «Очень короткая история больших данных» на сайте Forbes.com ведет отсчет проблематики больших данных с появления термина «информационный взрыв» (information explosion) в Оксфордском английском словаре в 1941 г.¹⁰ Тогда появились первые попытки количественной оценки темпов роста объемов данных. Гил Пресс также фиксирует первое осмысленное использование термина «Big Data» в октябре 1999 г. в цифровой библиотеке Ассоциации компьютерных вычислений (ACM Digital Library) – в статье исследователей NASA, посвященной проблемам визуализации информации¹¹.

Кроме того, в 2001 г. вышло основополагающее исследование Дуга Ланей (Doug Laney) для Meta Group, которое определило три ключевых параметра больших данных: объем, скорость, разнообразие (так называемые три «V»: Volume, Velocity, Variety)¹². Большие данные характеризуются гигантскими размерами (точных определений нет), большими скоростями их новой генерации и притока, неоднородностью и неупорядоченностью. И эти три «VVV» в том или ином виде признают все эксперты, что свидетельствует о более раннем появлении понятия и его эволюции.

Поэтому говорить о «дате рождения термина»¹³ в 2008 г., как минимум, некорректно. Если обратиться к сервису Google Trends, который как раз использует большие данные для анализа (массивы поисковых запросов и анализ документов на наличие ключевых слов), то рост популярности поискового запроса и темы «большие данные» (Big Data) в мире начинается с середины 2011 г.¹⁴

Можно предположить, что рост популярности в более узких специализированных научных и деловых кругах начался благодаря публикации «Nature» в том числе, что среди прочих авторов стимулировало популяризацию темы и привело к массовому интересу. Но прямую корреляцию установить невозможно.

Анализ эволюции понятия приводит к тому важнейшему выводу, что большинство исследователей и практиков признает сложность выработки единого определения «больших данных». Инструментальность понятия позволяет сфокусироваться на таком

определении «больших данных» и его понимании, которое лучше всего подходит для исследовательских задач. Поэтому необходимо не замыкаться на ограниченном круге определений, а использовать различные характеристики понятия, обращаться к различным исследованиям и подходам в определении «больших данных». При этом важен социогуманитарный аспект, когда применение технологий больших данных согласуется с интересами людей, их ценностями, этикой и моралью¹⁵.

Примечания

- ¹ Tweet geekery and epic crowdsourcing: an Oxford English Dictionary update. URL: <http://blog.oxforddictionaries.com/2013/06/oed-june-2013-update/> (дата обращения: 05.07.2017).
- ² «Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges; (also) the branch of computing involving such data». URL: <http://www.oed.com/view/Entry/18833#eid301162177> (дата обращения: 05.07.2017).
- ³ «Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions». URL: https://en.oxforddictionaries.com/definition/big_data (дата обращения: 05.07.2017).
- ⁴ «Very large sets of data that are produced by people using the internet, and that can only be stored, understood, and used with the help of special tools and methods». URL: <http://dictionary.cambridge.org/dictionary/english/big-data> (дата обращения: 05.07.2017).
- ⁵ Lynch C. Big data: how do your data grow? // Nature. 2008. Vol. 455. № 7209. P. 28–29.
- ⁶ См., напр.: Чехарин Е.Е. Большие данные: большие проблемы [Электронный ресурс] // Психологическая наука и образование. 2016. № 3 (21). URL: <http://cyberleninka.ru/article/n/bolshie-dannye-bolshie-problemy> (дата обращения: 06.07.2017); Толстова Ю.Н. Социология и компьютерные технологии // Социологические исследования. 2015. № 8. С. 3–14.
- ⁷ Черняк Л. Большие Данные – новая теория и практика // Открытые системы. СУБД. 2011. № 10. С. 18–25; цит. по: URL: <https://www.osp.ru/os/2011/10/13010990/> (дата обращения: 05.07.2017).
- ⁸ Большие данные. URL: https://ru.wikipedia.org/wiki/Большие_данные
- ⁹ Big data. URL: https://en.wikipedia.org/wiki/Big_data#cite_note-13
- ¹⁰ Press G. A Very Short History of Big Data [Электронный ресурс] // Forbes.com. 09.05.2013. URL: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#92208b965a18> (дата обращения: 5.07.2017).

- ¹¹ *Cox M., Ellsworth D.* Application-controlled demand paging for out-of-core visualization [Электронный ресурс] // Proceedings of the 8th conference on Visualization '97 (VIS '97). Phoenix, Arizona, USA, 1997. URL: https://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf (дата обращения: 5.07.2017).
- ¹² *Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety. 06.02.2001 [Электронный ресурс] // URL: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (дата обращения: 5.07.2017).
- ¹³ «BigData относится к числу немногих названий, имеющих вполне достоверную дату своего рождения – 3 сентября 2008 г., когда вышел специальный номер старейшего британского научного журнала “Nature”...» (*Черняк Л.* Указ. соч.).
- ¹⁴ Сервисы Google Trends. URL: [tps://trends.google.ru/trends/explore?date=all&q=big%20data,%2Fm%2F0bs2j8q](https://trends.google.ru/trends/explore?date=all&q=big%20data,%2Fm%2F0bs2j8q)
- ¹⁵ См., напр.: *Карчагин Е.В.* Справедливость в информационной области: этический аспект BigData // Вестник ВолГУ. Серия 7: Философия. Социология и социальные технологии. 2016. № 1 (31) [Электронный ресурс] URL: <http://cyberleninka.ru/article/n/spravedlivost-v-informatsionnoy-oblasti-eticheskiy-aspekt-big-data> (дата обращения: 09.07.2017).