

Copyright © 2014 by Academic Publishing House *Researcher*

Published in the Russian Federation
European Researcher
Has been issued since 2010.
ISSN 2219-8229
E-ISSN 2224-0136
Vol. 78, No. 7-1, pp. 1237-1242, 2014

DOI: 10.13187/issn.2219-8229
www.erjournal.ru



Engineering sciences

Технические науки

UDC 004.041

Big Data as Information Barrier

¹Victor Ya. Tsvetkov²Alexandr A. Lobanov

¹⁻² Moscow State Technical University of Radio Engineering, Electronics and Automation MSTU MIREA, Russian Federation

¹ Doctor of Technical Sciences, Professor

² PhD, Associate Professor

E-mail: cvj2@mail.ru

Abstract. The article covers analysis of ‘Big Data’ which has been discussed over last 10 years. The reasons and factors for the issue are revealed. It has proved that the factors creating ‘Big Data’ issue has existed for quite a long time, and from time to time, would cause the informational barriers. Such barriers were successfully overcome through the science and technologies. The conducted analysis refers the “Big Data” issue to a form of informative barrier. This issue may be solved correctly and encourages development of scientific and calculating methods.

Keywords: information; data; complexity; informational technologies; analysis; information processing.

Введение. В последние годы достаточно много говорится о проблеме «больших данных» (Big Data) [1, 2, 3]. Чаще всего эту проблему связывают с необходимостью обработки структурированных и неструктурированных данных больших объемов. Достаточно часто для характеристики «больших данных» используют критерий «три V»: объем (*volume*), скорость (*velocity*), многообразие (*variety*). Ради склонности к упрощению в этой совокупности пропущена важная характеристика «сложность». Появление термина соотносят с 2008 годом [4]. Введение термина «большие данные» связывают с Клиффордом Линчем – редактором журнала Nature [4], подготовившему серию работ на эту тему. Это обозначает признание проблемы в некомпьютерных сферах.

Однако еще в 2001 году с этой проблемой столкнулись специалисты в области искусственного интеллекта, когда провалом закончилась программа создания «Интеллектуального изображения». Тогда проблему «больших данных» не обозначили как проблему и расценили эту ситуацию как временные трудности. В целом проблему и факт появления больших данных можно соотнести с появлением компьютеров и технологий аналитической обработки данных.

После появления первых вычислительных машин, уже в 1970–1980-е годы программисты отметили закономерность, которая состоит в том, что «при росте

вычислительной мощности компьютеров появляются новые программные средства, для которых этой мощности всегда недостаточно». Можно констатировать, что проблему больших данных первыми обнаружили специалисты в области дистанционного зондирования Земли более 50 лет назад [5, 6], затем ее отметили программисты 40–50 лет назад. Затем ее зафиксировали аналитики 20–30 лет назад. И только в последние десять лет она открылась для бизнес – аналитиков и журналистов, что и привело к их повышенному вниманию к такому явлению и появлению термина.

Информационные барьеры. Академиком В.М. Глушков [7] было отмечено понятие первого и второго информационных барьеров по критерию сложности (невозможности) управления сложной социальной экономической системой. По существу он связывает эти барьеры только с четвертой информационной революцией и компьютерной обработкой информации. Эти виды информационных барьеров следует считать ресурсными, поскольку они обусловлены недостатком ресурсов для анализа данных.

Анализ эволюции общества показывает, что фактически человечество с древних времен преодолевало различные информационные барьеры. В процессе развития человеческого общества происходит наблюдение человека за объектами, явлениями и процессами окружающего мира. Как результат наблюдения происходит получение информации в информационном поле [8, 9], накопление опыта и формирование описаний объектов, явлений и процессов.

Первичное описание объектов окружающего мира состояло в формировании количественных и качественных свойств, характеристик, признаков и отношений между ними. Это описание представляет собой информационные коллекции. Вторичное описание состояло в формировании моделей и систем. Чем сложнее объект исследования, тем большего количества информации требует его описание и тем объемнее и сложнее информационные коллекции, составляющие такое описание.

Во многих источниках, особенно связанных с информатикой, информационные барьеры связывают с большим объемом и со сложностью информационных связей и информационных отношений. Контекстно информационный барьер связывают с человеком. С этой позиции проблемы освоения, понимания, интерпретации и применения информационных коллекций приводят к информационным барьерам [10, 11].

Факторы, приводящие к информационным барьерам. В практике исследования и анализа окружающего мира человек является завершающим звеном. Поэтому описания или информационные коллекции можно охарактеризовать как человеко анализируемые или человеко не анализируемые [12]. В этом случае речь идет о применении человеческого интеллекта.

Анализируемые человеком информационные коллекции характеризуются следующими свойствами: обозримость, воспринимаемость, временная анализируемость, целевая определенность, ситуационная определенность, полнота, достоверность, актуальность, точность, согласованность, надежность [13].

Обозримость – свойство информационных коллекций, состоящее в том, что человек (в рамках своего человеческого интеллекта) в состоянии *обозреть* совокупность параметров и связей, входящих в информационную коллекцию и *понять* данную информационную коллекцию как целое.

Воспринимаемость – свойство информационных коллекций, состоящее в том, что человек (в рамках своего человеческого интеллекта) в состоянии *воспринять и понять* данную информационную коллекцию как отражение объективной реальности или ее практическое назначение.

Если информационная коллекция необозрима, не воспринимаема или не анализируема за обозримое время, она, как правило, отвергается и не применяется человеком. Это и есть пример информационного барьера.

Типы информационных барьеров. В настоящее время выделяют разные типы информационных барьеров, часть из них описана в [7, 14, 15]. Для связи с проблемой «больших данных» будем использовать характеристики 3V и 1C. Соответственно V1 – объём (volume), V2 – скорость (velocity), V3 – многообразие (variety), C – сложность (complexity).

При этом следует отметить фактор сложности, как наиболее значимый в проблеме больших данных. Так при росте количества данных до величины n , количество связей между

ними растет пропорционально $n!$. Именно количество связей и задает сложность. Поэтому проблема больших данных не сводится к задаче 1 – обработке информационной коллекции объема n , а требует дополнительно решения задачи 2 – анализа связей объема $n!$. Очевидно, что в количественном и в качественном отношении вторая задача является значительно сложнее.

Для описания информационных барьеров используем понятие информационной ситуации. Выделим некоторые виды информационных барьеров.

Временной барьер – информационная ситуация, при которой имеет место разделение источника и приёмника информации во времени [14, 15], что замедляет и затрудняет обработку в допустимое для анализа время (фактор V2). Дополним этот тип случаев, когда обработка информационной коллекции с помощью известных методов или алгоритмов превышает допустимый ситуацией период времени.

Преодоление временного информационного барьера достигается разработкой новых средств коммуникации и новых быстродействующих алгоритмов и методов обработки информации [16].

Ресурсный барьер – информационная ситуация, при которой имеет место нехватка или ресурсов, необходимых для реализации информационных процессов (факторы V1-3, С). Для преодоления такого информационного барьера необходимо получение информационных ресурсов [17].

Пространственный барьер – информационная ситуация, при которой вследствие удаления источника и приёмника информации друг от друга обработка становится затруднительной из-за запаздывания информации или большого уровня помех (фактор V2).

Дополним этот тип барьера случаев, когда информационная коллекция распределена в разных географических точках пространства и обработка или освоение такой информационной коллекции невозможна в допустимый ситуацией период времени. Примером преодоления такого пространственного информационного барьера является дистанционное обучение.

Дополним этот тип барьера случаев, когда информационная коллекция содержит латентные (скрытые) параметры, распределенные в разных географических точках пространства и зависящие от пространственного положения. Примером преодоления такого пространственного информационного барьера является выявление и использование пространственных отношений [18] и применение методов геоинформатики.

Объемный информационный барьер – информационная ситуация, при которой вследствие большого объема информационная коллекция становится не обозримой, не воспринимаемой и не анализируемой в допустимый период времени (фактор V1).

Примером преодоления объемного информационного барьера является применение вычислительной техники и применение алгоритмов ускоренной обработки.

Барьер информационной сложности [19] – информационная ситуация, при которой вследствие большой сложности (сложность описывается графом связей и отношений) информационная коллекция становится не обозримой, не воспринимаемой и не анализируемой – в допустимый период времени (фактор С). Для преодоления барьера информационной сложности применяют интеллектуальных и автоматизированных системы и технологии [11].

Информационный барьер представления – информационная ситуация, при которой вследствие недостатка форм представления информационная коллекция или информационная модель становится не обозримой, не воспринимаемой и не анализируемой с помощью человеческого интеллекта (фактор V3). Примером преодоления информационного барьера представления является применение визуального моделирования, применение новых моделей представления, например, картографической модели.

Взгляд в историю развития человечества показывает, что информационные барьеры и проблемы их преодоления возникали задолго до появления ЭВМ. Эти проблемы или барьеры возникали: при сборе информации, хранении информации, передаче информации, увеличении объема информации, при росте сложности информационных коллекций, при повышении требований к оперативности обработки, при повышении требований к надежности передачи информации и пр.

По мере накопления свойство информационных коллекций, усложнялся процесс их анализа и переработки, возникал информационный барьер, обусловленный низкой пропускной способностью человека как системы обработки и необходимостью обработки больших объемов информации, с которыми эта система не справлялась. Для преодоления этого информационного барьера был разработан метод классификации. Этот метод упростил работу с информацией и позволил выделять нужные классифицированные области для анализа.

Исследования процессов принятия решений и мышления привели к появлению моделей знаний и правил вывода. Это тоже можно рассматривать как преодоление информационного барьера обусловленного необходимостью решения и анализа сложной информации.

С этих позиций интеллектуальные системы [11] можно рассматривать как средство преодоления информационного барьера, обусловленного в первую очередь сложностью, во вторую объемом информации и неспособностью человека как системы обработки и анализа в заданный период времени ее проанализировать и получить адекватное решение.

Свойством интеллектуальных систем является возможность выполнения творческих функций, которые традиционно считаются прерогативой человека. Другими словами, интеллектуальная система, в отличие от информационной системы, способна проявлять активность при отсутствии воздействия или прямых указаний человека.

Интеллектуальные системы не только обрабатывают и упрощают исходную информационную коллекцию, но в ряде случаев решают сложные задачи и в столь короткое время, которые человек принципиально решить не способен или не способен решить их также оперативно. Интеллектуальные системы не только помогают человеку, но и принимают за него решения, включая ту область решений, в которой он не адекватен.

Таким образом, интеллектуальные системы возникли как средство преодоления ряда информационных барьеров и позволяют получать результаты, которые не могут получить информационные системы и многие человеко-машинные системы.

С исследовательских позиций анализ информационных коллекций должен завершаться введением понятий. Определение понятия связано с терминологическими отношениями [20], формированием идеи, определения, суждений, метода, системы понятий. Если система понятий согласована, то уменьшается информационный барьер познания. Если система понятий, которую создает исследователь, не согласована с существующими терминологическими системами, то информационный барьер возрастает. Формирование согласованной системы понятий осуществляется на основе логики и научного подхода.

Выводы. Современная проблема больших данных может быть рассмотрена как очередной этап в развитии человечества, связанный с преодолением очередного информационного барьера. Основными характеристиками проблемы являются в первую очередь сложность и во вторую физический объем информационной коллекции. Большие данные с одной стороны обуславливают постановку и решение новых задач [21]. С другой стороны они обуславливают развитие интегрированных и комплексных систем и технологий.

Преувеличенное внимание к «большим данным» со стороны журналистов и бизнесменов обусловлено отсутствием практики преодоления информационных барьеров и рассмотрением этого явления как совершенно нового, в то время как оно периодически появляется в развитии человечества и «новым» является не само явление, а «новое качество» известного явления. С этим новым качеством специалисты разберутся. С познавательной точки зрения преодоление информационного барьера «большие данные» способствует развитию познания окружающего мира и построению его целостной картины.

Примечания:

1. Майер-Шенбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим. Манн, Иванов и Фербер, 2014. 240 с.

2. Черняк Л. Большие данные – новая теория и практика // Открытые системы. СУБД 2011. №10. с. 18-25.

3. Jacobs A. The pathologies of big data // Communications of the ACM. 2009. Т. 52. № 8. p. 36-44.
4. Lynch C. Big data: How do your data grow? // Nature. 2008. Т. 455. №. 7209. p. 28-29.
5. Космические исследования земных ресурсов. Методы и средства измерений и обработки информации. М.: Наука, 1976. 386 с.
6. Цветков В.Я. Методы и системы обработки и представления видеонформации. М.: ГКНТ, ВНИИЦентр, 1991. 113 с.
7. Глушков В.М. Основы безбумажной информатики. М.: Наука, 1987. 557 с.
8. Цветков В.Я. Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. 2014. №5, ч.2. с. 178-180.
9. Tsvetkov V.Y. Information field. // Life Science Journal. 2014. 11(5). pp. 551-554.
10. Поляков А.А., Цветков В.Я. Прикладная информатика: Учебно-методическое пособие: В 2-х частях: Часть 1 / Под общ. ред. А.Н. Тихонова. М.: МАКС Пресс, 2008. Часть 1. 788 с., часть 2, 860 с.
11. Савиных В.П., Цветков В.Я. Развитие методов искусственного интеллекта в геоинформатике // Транспорт Российской Федерации. 2010. № 5. с. 41-43.
12. Tsvetkov V.Ya. Information interaction // European Researcher, 2013, Vol.(62), № 11-1, p. 2573-2577.
13. Tsvetkov V.Ya. Spatial Information Models // European Researcher, 2013, Vol.(60), № 10-1, p. 2386-2392.
14. Цветков В.Я. Геоинформатика и преодоление информационных барьеров // Геодезия и аэрофотосъемка, №6, 2004. с. 113-118.
15. Цветков В.Я., Маркелов В.М., Романов И.А. Преодоление информационных барьеров // Дистанционное и виртуальное обучение. 2012. № 11. с. 4-7.
16. Казенников А.О., Соловьев И.В. Извлечение структурированного новостного сообщения из веб-страниц при использовании дополнительной информации RSS. // Вестник МГТУ МИРЭА «MSTU MIREA HERALD» 2014, № 2 (3), с. 276-284.
17. Матчин В.Т. Информационные ресурсы как инструмент научного исследования и развития // Вестник МГТУ МИРЭА. 2014, № 2 (3), с. 235-256.
18. Цветков В.Я. Пространственные отношения в геоинформатике // Международный научно-технический и производственный журнал «Науки о Земле». Выпуск 01-2012, с. 59-61.
19. Tsvetkov V.Ya. Complexity Index // European Journal of Technology and Design, 2013, Vol.(1), № 1, p. 64-69.
20. Тихонов А.Н., Иванников А.Д., Цветков В. Я. Терминологические отношения // Фундаментальные исследования, 2009, № 5, с. 146.
21. Herodotou H. et al. Starfish: A Self-tuning System for Big Data Analytics // CIDR. 2011. Т. 11. p. 261-272.

References:

1. Maier-Shenberger V., Kuk'er K. Bol'shie dannye: Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim. Mann, Ivanov i Ferber, 2014. 240 с.
2. Chernyak L. Bol'shie dannye – novaya teoriya i praktika // Otkrytye sistemy. SUBD 2011. №10. s. 18-25.
3. Jacobs A. The pathologies of big data // Communications of the ACM. 2009. Т. 52. № 8. p. 36-44.
4. Lynch C. Big data: How do your data grow? // Nature. 2008. Т. 455. №. 7209. r. 28-29.
5. Kosmicheskie issledovaniya zemnykh resursov. Metody i sredstva izmerenii i obrabotki informatsii. М.: Nauka, 1976. 386 с.
6. Tsvetkov V.Ya. Metody i sistemy obrabotki i predstavleniya videonformatsii. М.: GKNT, VNTITsentr, 1991. 113 s.
7. Glushkov V.M. Osnovy bezbumazhnoi informatiki. М.: Nauka, 1987. 557 s.
8. Tsvetkov V.Ya. Estestvennoe i iskusstvennoe informatsionnoe pole // Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovaniy. 2014. №5, ch.2. s. 178-180.
9. Tsvetkov V.Y. Information field. // Life Science Journal. 2014. 11(5). rr. 551-554.

10. Polyakov A.A., Tsvetkov V.Ya. Prikladnaya informatika: Uchebno-metodicheskoe posobie: V 2-kh chastyakh: Chast' 1 / Pod obshch. red. A.N. Tikhonova. M.: MAKS Press, 2008. Chast' 1. 788 s., chast' 2, 860 s.
11. Savinykh V.P., Tsvetkov V.Ya. Razvitiye metodov iskusstvennogo intellekta v geoinformatike // Transport Rossiiskoi Federatsii. 2010. № 5. s. 41-43.
12. Tsvetkov V.Ya. Information interaction // European Researcher, 2013, Vol.(62), № 11-1, p. 2573-2577.
13. Tsvetkov V.Ya. Spatial Information Models // European Researcher, 2013, Vol.(60), № 10-1, p. 2386-2392.
14. Tsvetkov V.Ya. Geoinformatika i preodolenie informatsionnykh bar'erov // Geodeziya i aerofotos'emka, №6, 2004. s. 113-118.
15. Tsvetkov V.Ya., Markelov V.M., Romanov I.A. Preodolenie informatsionnykh bar'erov // Distantionnoe i virtual'noe obuchenie. 2012. № 11. s. 4-7.
16. Kazennikov A.O., Solov'ev I.V. Izvlechenie strukturirovannogo novostnogo soobshcheniya iz veb-stranits pri ispol'zovanii dopolnitel'noi informatsii RSS. // Vestnik MGTU MIREA «MSTU MIREA HERALD» 2014, № 2 (3), s. 276-284.
17. Matchin V.T. Informatsionnye resursy kak instrument nauchnogo issledovaniya i razvitiya // Vestnik MGTU MIREA. 2014, № 2 (3), s. 235-256.
18. Tsvetkov V.Ya. Prostranstvennye otnosheniya v geoinformatike // Mezhdunarodnyi nauchno-tekhnicheskii i proizvodstvennyi zhurnal «Nauki o Zemle». Vypusk 01-2012, s. 59-61.
19. Tsvetkov V.Ya. Complexity Index // European Journal of Technology and Design, 2013, Vol.(1), № 1, p. 64-69.
20. Tikhonov A.N., Ivannikov A.D., Tsvetkov V. Ya. Terminologicheskie otnosheniya // Fundamental'nye issledovaniya, 2009, № 5, s. 146.
21. Herodotou H. et al. Starfish: A Self-tuning System for Big Data Analytics // CIDR. 2011. T. 11. r. 261-272.

УДК 004.041

Big Data as Informational Barrier

¹ Виктор Яковлевич Цветков

² Александр Анатольевич Лобанов

¹⁻² Московский государственный технический университет радиотехники, электроники и автоматики МГТУ МИРЭА, Российская Федерация

¹ Доктор технических наук, профессор

² Кандидат технических наук, доцент

E-mail: cvj2@mail.ru

Аннотация. Дается анализ проблемы «больших данных» (Big Data), которая интенсивно обсуждается последние десять лет. Раскрываются причины появления проблемы и факторы, которые ведут к ее появлению. Показано, что факторы создающие проблему «больших данных» существовали достаточно давно и периодически приводили к появлению информационных барьеров. Эти барьеры успешно преодолевались за счет развития науки и технологий. Анализ дает основание отнести проблему «больших данных» (Big Data) к новой разновидности информационного барьера. Эта проблема корректно решаема и способствует развитию научных методов и вычислительных методов.

Ключевые слова: информация; данные; сложность; информационные барьеры; информационные технологии; анализ; обработка информации.