

20. Moritz T. Geo-referencing the natural and cultural world, past and present: Towards building a distributed, peer-reviewed gazetteer system // Digital Gazetteer Information Exchange Workshop. Smithsonian Institute. October 12–14. 1999.
21. Майоров А.А., Цветков В.Я. Геореференция как применение пространственных отношений в геоинформатике // Геодезия и аэрофотосъемка. 2012. № 3. С. 87–89.
22. Paul Getty Trust Getty Thesaurus of Geographic Names Online: Place Type Looking. 2004. URL: http://www.getty.edu/research/conducting_research/vocabularies/tgn/
23. Кужелев П.Д. О применении геостатистики в науках о Земле // Науки о Земле. 2012. № 4. С. 77–81.
24. Rendu J.-M. An introduction to geostatistical methods of mineral evaluation. Technical Report. South African Institute of Mining and Metallurgy. Johannesburg, 1981.
25. Pichler G. Computer-Programme der Geostatistik. Master's Thesis. Institut für Statistik. Technische Universität, Graz, Austria, 1982.
26. Майоров А.А., Матерухин А.В. Геоинформационный подход к задаче разработки инструментальных средств массовой оценки недвижимости // Геодезия и аэрофотосъемка. 2011. № 4. С. 92–97.

Spatial relations in inventories

Vasyutinskaya Stanislava Igorevna, PhD, Associate Professor, Department of Economics and Business. Faculty of Economics and management of territories, Moscow State University of Geodesy and Cartography

The article analyzes the spatial relationship as the spatial and economic factors. The article shows the role of Geoinformatics to identify and describe spatial relationships. This article describes the features of the application of spatial relations in the inventory. This article describes an example of formalization of spatial relations. The article shows that the formalization of spatial relationships helps automate and intelligent processing of inventory information

Keywords: applied geoinformatics, cadastre, spatial relationships, information models, hierarchical relationships

УДК 004.041

БОЛЬШИЕ ДАННЫЕ В ФОТОГРАММЕТРИИ И ГЕОДЕЗИИ

*Андрей Иванович Павлов, канд. техн. наук, доцент, научный сотрудник
экспериментально-технологического отделения № 28,
E-mail: andravlav.51@mail.ru,*

НИИ оснований и подземных сооружений ОАО «НИЦ «Строительство»

Раскрывается проблема «больших данных» в фотограмметрии и геодезии. Статья описывает причины и факторы, которые ведут к появлению больших данных. Дается сравнение больших данных и обычных данных. Показано, что проблема больших данных состоит не только в больших объемах данных, но и в нечеткости информации, сложности информационных моделей и требовании оперативной обработки. Дается анализ инструментария, применяемого при обработке больших данных.

Ключевые слова: данные, большие данные, сложность, методы обработки, фотограмметрия, геодезия, информационные технологии

Введение. Проблема «больших данных» (BigData) [1–4] в явной форме обсуждается в последние 7–8 лет. Ее связывают, в первую очередь, с большими объемами данных. К ней присовокупляют еще ряд факторов, таких как необходимость обработки слабо структурированных и неструктурированных данных больших объемов. Формально, появление термина отмечают 2008 годом, с работами Клиффорда Линча – редактора журнала Nature [5]. Однако это лишь признание данной проблемы в широких слоях общества.

Проблема, обозначаемая как большие данные, отмечена в 1987 году академиком В.М. Глушковым как «информационный барьер» [6]. С этой проблемой впервые столкну-

лись специалисты в области дистанционного зондирования Земли более 50 лет назад [7, 8]. Тогда она была узкоспециальной. Но в явной форме эта проблема открылась лишь в последние годы для бизнес-аналитиков и журналистов, что и привело к их повышенному вниманию к этому явлению и появлению соответствующего термина.



А.И. Павлов

Однако, бросая взгляд в историю человечества, можно констатировать, что данная проблема является условной. Она отражает невозможность на определенном этапе развития обрабатывать большие объемы или сложные данные существующими теоретическими и технологическими средствами [4]. На протяжении развития человечество получает информацию в информационном поле [9]. Глубина поиска в этом поле определяется инструментариями и теорией. По мере проникновения в это поле возникают ситуации, когда теории не хватает для описания и для обработки новых данных.

При этом надо констатировать, что процесс накопления знаний включает накопление явных и неявных знаний. В последующем неявные знания также трансформируются в явные знания. По мере накопления данных и опыта происходит формирование накопление явных и неявных знаний, а также совокупности описаний объектов, явлений и процессов.

Процесс извлечения информации и данных всегда включал два этапа: первичный и вторичный. Первичное описание включает измерения и получения данных как таковых. Оно включает формирование количественных и качественных свойств, характеристик объектов. Вторичная обработка детализирует и уточняет описание и включает формирование моделей, признаков и отношений между ними. Это описание представляет собой информационные ситуации, в которых находятся объекты исследования. Чем сложнее ситуация и объект исследования, тем большего количества формальных параметров и отношений между ними требует его описание. Это влечет рост объемов и сложности информационных конструкций, описывающих такие объекты и ситуации.

Рост объемов собираемой информации и требование ее обработки и хранения делают актуальным исследования в области методов и алгоритмов анализа больших и сверхбольших наборов данных. В работе [10] предполагается, что выявление закономерностей в больших массивах данных становится также инструментом исследования и получения новых знаний. Рост объемов данных характерен как для IT-компаний, так и для научной сферы [11]. Рост объемов данных и появление больших данных имеет место в самых различных областях [12]. Поскольку проблема больших данных первоначально появилась в Науках о Земле, то она не исчезла и теперь. Это определяет актуальность исследования этой проблемы в геодезии и фотограмметрии как характерных направлениях, связанных с появлением и обработкой больших данных.

Описание больших данных. Описания и модели больших данных, применяемых в науке и технологиях требуют проведения исследований и разработок, направленных на разработку масштабируемых аппаратных и программных решений. Пока пределом возможностей приложений, ориентированных на обработку больших объемов данных, являются петабайтные коллекции данных и гигабайтные потоки данных. В аспекте существующей тенденции ожидаются еще большие масштабы и объемы данных.

При работе с большими данными, приходится сталкиваться со следующими модификациями этой проблемы: большие объемы данных [1], интенсифицированные потоки данных, существенное сокращение допустимого времени анализа данных, предел времени для принятия решений при большом количестве данных [4], возрастание морфологической сложности моделей, возрастание структурной сложности [12] моделей и систем, возрастание вычислительной сложности [12], относительный рост слабоструктурированной исходной информации, относительный рост нечеткой

информации, рост потребностей в параллельных вычислениях [5] и т.д. Упрощенно проблемы работы с данными большого объема приведены в таблице 1, в которой дается качественное сравнение обычных данных и больших данных.

Таблица 1

Сравнительные характеристики больших данных

Характеристика	Обычные данные	Большие данные
Формат	Однородный	Неоднородный
Объем	Мегабайты гигабайты	Петабайты
Распределенность данных	нет	есть
Тип задачи	Первого рода	Второго рода
Тип моделей решателей	Алгоритмические	Статистические
Тип моделирования	Имитационное моделирование	Стохастическое
Топологическая сложность	Приемлемая	Высокая
Вычислительные ресурсы	Обычные	Повышенной мощности

Приложения, ориентированные на обработку больших объемов данных, имеют дело с наборами данных, имеющими информационные объемы от нескольких терабайт до петабайта. На практике эти данные поступают в разных форматах и часто распределены между несколькими приложениями. Обработка больших объемов данных обычно происходит в режиме многошагового аналитического конвейера, включающего стадии преобразования, обработки и интеграции обработанных данных.

Требования к скорости вычислений обычно почти линейно возрастают при росте объема данных. Часто такие вычисления осуществляют на основе предварительного распараллеливания данных и последующей параллельной обработке. К основным техническим проблемам больших вычислений относятся управление данными, методы фильтрации и интеграции данных, эффективная поддержка запросов и распределения данных.

Особо следует подчеркнуть, что проблема распределения данных сама по себе создает проблемы даже при не очень большом объеме. Это мотивирует разработку специальных пространственных моделей данных [13], которые включают свойство распределения в пространстве.

Источники больших данных в геодезии и фотограмметрии. *Неявные знания как источник появления больших данных.* Сами по себе неявные знания [14, 15] характеризуются слабой структурированностью, что является признаком «больших данных». Особенность получения информации при аэрофотосъемке и, особенно, при дистанционном зондировании в том, что данные первоначально накапливают и спустя какой-то временной период начинают обрабатывать. При этом на аэрокосмических снимках могут содержаться изображения непонятных явлений и сложных комбинаций разных объектов.

При аэрокосмической съемке изображение формируется как совокупность наложенных друг на друга объектов. Например, облака могут закрывать часть территории, при фотографировании водной поверхности она становится прозрачной на определенную глубину и все объекты в пределах видимости отпечатываются как на одном слое изображения.

Сложность как источник больших данных. Особенность термина «сложность» в том, что он является связанной сущностью (или атрибутом) и с другой сущностью. Это порождает различные виды сложности. Например, различают виды сложности по связи с объектом: сложность организационно-технической системы [16], сложность процесса (действия), сложность явления, условная колмогоровская сложность, простая колмогоровская сложность, префиксная сложность [17], сложность ситуации, сложность теории и т.д.

Следовательно, термин «сложность» требует указания связанного объекта, по отношению к которому сложность оценивается. В противном случае оценка сложности будет неадекватной. Сложности качественно разных сущностей или разных атрибутов могут быть не сопоставимы.

Различают разные сложности одного и того же объекта:

- структурная сложность объекта [12];
- сложность процессов, в которых участвует объект [17];
- сложность получения решения в допустимое время – временная сложность [18, 19];
- сложность, обусловленная ограниченным объемом памяти вычислительной системы при больших объемах перерабатываемой информации – ёмкостная сложность [19];
- сложность определения положения в пространстве – пространственная сложность позиционирования;
- сложность формы объекта – морфологическая сложность;
- сложность ситуации, в которой находится объект – ситуационная сложность;
- сложность позиции, в которой находится объект – позиционная сложность;
- сложность декодирования объекта [12] – криптографическая сложность;
- сложность описания явления, с которым взаимосвязан объект;
- сложность теории, описывающей поведение объекта, и т.д.

Следовательно, для полноты исследования можно говорить об «обобщенной сложности» объекта и об «атрибутивной сложности».

Общие характеристики больших данных. Для характеристики «больших данных» часто применяют критерий «три V»: объём (*volume* – $v1$), скорость (*velocity* – $v2$), многообразие (*variety* – $v3$), К этому необходимо добавить сложность [12, 19] (*complex* – $c1$).

Обработка серий фотоснимков, как и в 60-е годы приводит к получению файлов большого объема. Ситуация усугубляется появлением сканеров высокого разрешения, что на порядок повышает информационную емкость снимков и создает проблему их обработки. В чистом виде проблема больших данных появляется при создании и ведении мультимасштабных карт [20].

Критерий $v1$ проявляется в области наук о Земле при хранении гигабайтных и терабайтных файлов. Это встречается при работе с мультимасштабными картами [20] и банками пространственных данных [21].

Критерий $v2$ проявляется в области наук о Земле при уравнивании больших систем уравнений [22]. Этот критерий проявляется также при оперативном управлении подвижными объектами.

Критерий $v3$ проявляется в области наук о Земле при моделировании сложных систем [23] большого территориального охвата. Он также проявляется при семантическом анализе информационных объектов [24]. Критерий $c1$ проявляется в области наук о Земле при топологическом анализе сложных транспортных и иных сетей [25].

Выводы. Возникновение проблемы больших данных можно рассматривать как отражение процессов глобализации. Анализ данных больших объемов требует привлечения технологий и средств реализации высокопроизводительных вычислений. Основными факторами проблемы являются, в первую очередь, сложность, и, во-вторую – физический объем информационной коллекции. Большие объемы данных порождают проблемы при формировании информационных ресурсов из таких данных. По существу большие данные являются новой формой информационного барьера [4]. Большие данные качественно отличаются от обычных данных тем, что создают семантический разрыв при их обработке и анализе. Большие данные, с одной стороны, обуславливают постановку и решение новых задач [26], с другой стороны – 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 познавательной точки зрения преодоление информационного барьера «большие данные» способствует развитию познания окружающего мира и построению его целостной картины.

Литература.

1. Майер-Шенбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим. М.: 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. Р. 36–44.
4. Tsvetkov V.Ya., Lobanov A.A. Big Data as Information Barrier // European Researcher. 2014. Vol.(78). № 7-1. Р. 1237–1242.
5. Lynch C. Bigdata. How do your data grow? // Nature. 2008. Т. 455. № 7209. Р. 28–29.
6. Глушков В.М. Основы безбумажной информатики. М.: Наука, 1987. 557 с.
7. Космические исследования земных ресурсов. Методы и средства измерений и обработки информации. М.: Наука, 1976. 386 с.
8. Цветков В.Я. Методы и системы обработки и представления видеонформации. М.: ГКНТ, ВНИЦентр, 1991. 113 с.
9. Цветков В.Я. Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. 2014. № 5. Ч. 2. С. 178–180.
10. The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009. URL: <http://research.microsoft.com/enus/collaboration/fourthparadigm>
11. LoekEssers: CERN pushes storage limits as it probes secrets of universe. URL: <http://news.idg.no/cw/art.cfm?id=FF726AD5-1A64-6A71-CE987454D9028BDF>
12. Tsvetkov V.Ya. Complexity Index // European Journal of Technology and Design. 2013. Vol.(1). № 1. Р. 64–69.
13. Tsvetkov V.Ya. Spatial Information Models // EuropeanResearcher. 2013. Vol.(60). № 10-1. Р. 2386–2392.
14. Polanyi M. *The tacit dimension*. London: Routledge and Kegan Paul. 1966.
15. Сизов А.С., Цветков В.Я. Неявное знание: оппозиционный логический анализ и типологизация // Вестник Российской Академии Наук. 2015. Т. 85. № 9. С. 800–804. DOI: 10.7868/S0869587315080319.
16. Тихонов А.Н., Иванников А.Д., Соловьёв И.В., Цветков В.Я. Основы управления сложной организационно-технической системой. Информационный аспект. М.: МаксПресс, 2010. 228 с.
17. Вьюгин В.В. Колмогоровская сложность и алгоритмическая случайность. М.: ИППИ РАН, 2012. 131 с.
18. Tsvetkov V.Ya. Multipurpose Management // European Journal of Economic Studies. 2012. Vol.(2). № 2. Р. 140–143.
19. Адигеев М.Г. Введение в теорию сложности. Ростов н/Д.: РГУ, 2004. 35 с.
20. Цветков В.Я., Железняков В.А. Мультимасштабная электронная карта как основа системы учёта земель // Государственный советник. 2014. № 1. С. 28–37.
21. Железняков В.А. Интеллектуальное обновление информации в банке геоданных // Инженерные изыскания. 2012. № 5. С. 58–61.
22. Михайлович К. Геодезия (уравнительные вычисления) / пер. С.В. Лебедева; под ред. ВД Большакова. М.: Недра, 1984.
23. Бусленко В.Н. Автоматизация имитационного моделирования сложных систем. М.: Наука, 1977.
24. Цветков В.Я. Семиотический подход к построению моделей данных в автоматизированных информационных системах // Геодезия и аэрофотосъемка. 2000. № 5. С. 142–145.
25. Свами М., Тхуласираман К. Графы, сети и алгоритмы. М.: Мир, 1984. Т. 198.
26. Herodotou H. et al. Starfish: A Self-tuning System for Big Data Analytics // CIDR. 2011. Т. 11. Р. 261–272.

Big data photogrammetry and geodesy

Andrey Ivanovich Pavlov, Ph.D., Associate Professor

Researcher experimental technology department number 28, Institute of Foundations and underground structures of «SIC» Construction, E-mail: andpavlov.51@mail.ru

This article describes the problem of «big data» in the photogrammetry and geodesy. This article describes the causes and factors that lead to the emergence of big data. The article compares the big data and traditional data. The article shows that the problem of large data depends on large amounts of data. The article shows that the problem of large data depends also on fuzzy information, information model complexity and demands high processing speed. The analysis tools used in the processing of big data.

Keywords: Data, big data, complexity, methods of processing, photogrammetry, surveying, information technology