

ЭКСПЕРТИЗА И АНАЛИТИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ

О ПРИМЕНЕНИИ ТЕХНОЛОГИЙ «БОЛЬШИХ ДАННЫХ» В ИНФОРМАЦИОННОЙ СИСТЕМЕ ФЕДЕРАЛЬНОГО РЕЕСТРА ЭКСПЕРТОВ НАУЧНО-ТЕХНИЧЕСКОЙ СФЕРЫ МИНОБРНАУКИ РОССИИ

Е.А. Марышев, зам. дир. центра ФГБНУ НИИ РИНКЦЭ канд. техн. наук, emarysh@extech.ru

М.В. Сергеев, гл. научн. сотр. ФГБНУ НИИ РИНКЦЭ, sergeev@extech.ru

Статья посвящена оценке возможного использования технологий больших данных для решения задач государственной научной и научно-технической экспертизы. Проведен обзор задач экспертизы в научной научно-технической сфере, для которых целесообразно применение данных технологий. Сделан вывод о том, что технологии больших данных могут стать реальным инструментом повышения эффективности принятия решений в сфере управления научно-технологическим комплексом России, обеспечивая руководящие органы более качественной и достоверной научно-аналитической информацией, полученной с помощью экспертного сообщества.

Ключевые слова: научная и научно-техническая экспертиза, большие данные, Big Data, неструктурированные данные.

ON THE APPLICATION OF «BIG DATA» TECHNOLOGIES IN THE INFORMATION SYSTEM OF FEDERAL ROSTER EXPERTS OF SCIENTIFIC AND TECHNICAL SPHERE OF MINISTRY OF EDUCATION AND SCIENCE OF RUSSIA

E.A. Marishev, Deputy Head of Centre, SRI FRCEC, Doctor of Engineering, emarysh@extech.ru

M.V. Sergeev, Chief Researcher, SRI FRCEC, sergeev@extech.ru

The article evaluates the possible use of big data technologies to solve problems and the state of scientific and technological expert examination. The article presents a review of expert examination tasks in scientific and scientific-technological sphere, where the appropriate application of these technologies is necessary. The article concludes that big data technologies can become a real tool for improving the effectiveness of decision making in the management of scientific-technological complex of Russia, providing the governing bodies of higher quality and accurate scientific-analytical information obtained by the expert community.

Keywords: the scientific and technological expert examination, Big Data, unstructured data.

Технологии «больших данных», несмотря на то, что находятся на начальном этапе своего развития, уверенно завоевывают популярность в мире. Как одно из ключевых направлений компьютерной науки, они, несомненно, оказывают огромное влияние на все направления развития высокотехнологического сектора и быстро становятся частью нашей повседневной жизни.

Существует множество определений термина «большие данные», тем не менее, все они подразумевают, что это подходы, инструменты, методы для обработки огромных и непрерывно растущих объемов данных, разнообразных по структуре и источникам, позволяющие существенно повысить эффективность хранения, управления потоками, анализа информации. Другими словами, технология «больших данных» имеет три отличительных признака, часто обо-

значаемых как «три V»: volume (анализируются массивы данных объемом в десятки терабайт), velocity (накопление и обработка данных идут с высокой скоростью), variety (вариативность – обрабатываются данные самых разных типов из одного или нескольких источников) [1].

Одной из ключевых технологий, относящихся к обработке «больших данных», является платформа Hadoop с открытым исходным кодом, позволяющая обрабатывать огромные массивы данных в распределенной среде [2]. Hadoop позволяет не только сократить время на обработку и подготовку данных для аналитических систем, но и существенно расширяет возможности по анализу, позволяя оперировать слабоструктурированными или неструктурированными данными. Платформа активно развивается, многие изначально связанные с ней проекты и технологии впоследствии стали самостоятельными.

Многие представители научного сообщества и бизнеса отмечают значительный потенциал «больших данных» как стимула инноваций, двигателя прогресса и торговли [3, 4, 5], полагая, что такие технологии могут изменить методы научных исследований и организации бизнеса, обеспечив более оперативные и точные аналитические действия для принятия более продуманных решений.

Перспективы развития «больших данных» в ближайшие годы позитивно оцениваются в исследовании компании IDC «Глобальный рынок технологий и сервисов, связанных с “Большими данными”, прогноз на 2013–2017 годы». Авторы уверены, что «большие данные» и связанный с ними рынок покажет рост, в шесть раз превышающий развитие рынка традиционных информационных и телекоммуникационных технологий. По прогнозам команды аналитиков IDC, к 2017 г. объем рынка «больших данных» достигнет 32,4 млрд долл. США.

В августе 2014 г. компания Gartner выпустила ежегодный «Цикл зрелости новых технологий» [6]. По мнению аналитиков компании, технологии «больших данных» уже прошли «пик завышенных ожиданий» (Peak of Inflated Expectation) и в настоящее время перемещаются в сторону «впадины разочарования» (Trough of Disillusionment). Этот процесс происходит довольно быстро, так как согласованный подход к этой технологии уже сложился, и большинство новых достижений носят характер «добавок», а не революционных перемен. Тем не менее, многие эксперты полагают, что ничего страшного с «большими данными» не происходит – сейчас имеет место определенное разочарование среди тех, кто поддался первоначальной шумихе, но в перспективе технологию ждет выход на «плато продуктивности» (Plateau of Productivity).

С точки зрения реализации научно-технической и инновационной политики современной России технологии «больших данных» могут стать реальным инструментом повышения эффективности принятия решений в сфере управления научно-технологическим комплексом государства путем обеспечения руководящих органов более качественной и достоверной научно-аналитической информацией, полученной с помощью экспертного сообщества.

Рассмотрим вопрос целесообразности использования технологий «больших данных» в процедурах научной и научно-технической экспертизы.

Прежде всего следует отметить, что единой законодательной и нормативной правовой базы экспертизы и экспертной деятельности в научно-технической сфере в настоящее время не существует [7]. Участники экспертного процесса руководствуются положениями Федерального закона «О науке и государственной научно-технической политике» (от 23 августа 1996 г. № 127-ФЗ). Кроме того, отношения, возникающие в связи с назначением и проведением научной и научно-технической экспертизы, регулируются модельным законом «О научной и научно-технической экспертизе», принятым Межпарламентской ассамблеей государств-участников СНГ в ноябре 2003 г.

В соответствии с требованиями Федерального закона «О техническом регулировании» (от 27 декабря 2002 г. № 184-ФЗ), в ФГБНУ НИИ РИНКЦЭ создан типовой технологический процесс экспертизы в виде стандарта организации (СТО), отвечающий современным требованиям к организации и проведению государственной экспертизы в сфере науки [8]. Типовая модель механизма государственной научной и научно-технической экспертизы представлена на рис 2.

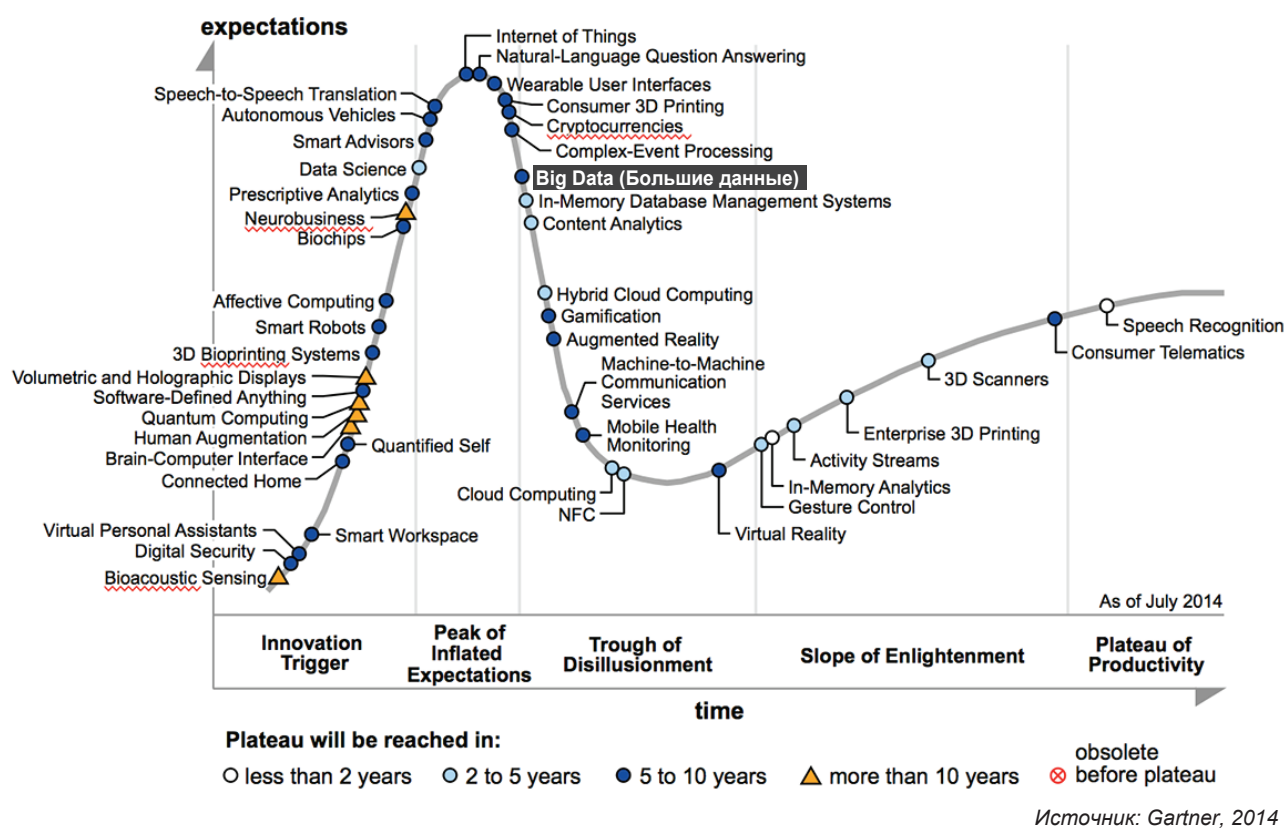


Рис. 1. Цикл зрелости новых технологий (Hype Cycle) по состоянию на август 2014 г.

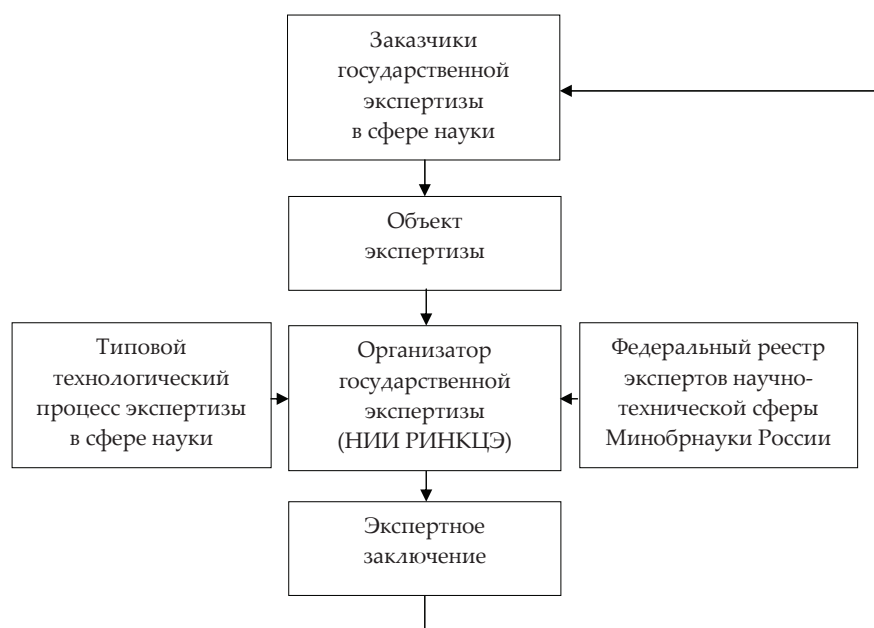


Рис. 2. Модель механизма государственной научной и научно-технической экспертизы

Как следует из рис. 2, объект экспертизы, сформированный заказчиком экспертизы, поступает организатору государственной экспертизы – ФГБНУ НИИ РИНКЦЭ. Государственная научная и научно-техническая экспертиза представленного объекта экспертизы организуется после подбора экспертов, в частности, из Федерального реестра экспертов научно-технической сферы Минобрнауки России (далее – ФРЭ), с использованием типового технологического процесса экспертизы в сфере науки. Сформированное экспертами экспертное заключение после рассмотрения его руководством НИИ РИНКЦЭ направляется заказчику.

Типовой технологический процесс экспертизы предусматривает следующие процедуры:

- прием и регистрация документов на объект, подлежащий экспертизе;
- резолюция (решение) руководства о проведении экспертизы;
- формирование пакета документов для экспертизы, в том числе тиражирование материалов;
- изучение, классификация и анализ объекта экспертизы;
- формирование технического задания на экспертизу;
- проведение информационного поиска;
- подбор экспертов (из числа аккредитованных в ФРЭ) и формирование экспертных пулов;
- проведение экспертизы в информационной системе ФРЭ (reestr.extech.ru) и формирование экспертного заключения;
- обработка экспертных заключений;
- подготовка и согласование заключения государственной экспертизы.
- архивация материалов;
- отправка материалов экспертизы заказчику.

С целью определения целесообразности применения технологий «больших данных», более подробно рассмотрим отдельные процедуры технологического процесса экспертизы.

В ходе изучения, классификации и анализа объекта экспертизы необходимо, в частности, установить состав и значения его классификационных признаков, таких как принадлежность к научно-технической продукции, к предметной области знаний; выявить возможные факты уже существующего финансирования данного объекта и повторности поступления его на экспертизу.

Проведение предварительного информационного поиска по тематике объекта экспертизы осуществляется в базах данных общего доступа, служебных базах данных, локальном архиве, удаленных источниках.

Процедура подбора экспертов неформализована, весьма ответственна и состоит в выявлении независимых компетентных экспертов по узкой тематике объекта экспертизы. Отметим, что существующие в настоящее время подходы к данной проблеме часто основаны на субъективных суждениях или результатах психологических и социальных исследований. Другие методы, использующие, например, оценки непротиворечивости суждений эксперта [9], сложны в реализации традиционными средствами, так как предполагают обработку данных обширных архивов, содержащих заключения оцениваемых экспертов.

Особенность проведения экспертизы состоит в том, что заключение формируется на основе субъективных оценок эксперта. Поэтому для принятия объективного решения ему необходимо проанализировать и переработать большой объем информации, учитывая влияние различных факторов и оценив вероятные последствия того или иного решения.

Следует также отметить, что эксперты ФРЭ привлекаются не только для проведения государственной научной и научно-технической экспертизы. Они активно участвуют в подготовке информационно-аналитических материалов о состоянии и перспективах развития российской и зарубежной сфер исследований, разработок и инновационной деятельности.

Все вышеперечисленные процедуры научной и научно-технической экспертизы имеют ряд особенностей, применимых к термину «большие данные»:

- значительные объемы данных, расположенные на множестве разных компьютеров;
- данные слишком большие для резервного копирования;
- данные могут быть структурированными, слабо-структурированными и неструктурированными;
- использование традиционных решений для аналитической обработки данных в реальном времени или для организации хранилищ данных решений не подходит для анализа данных.

Смысл концепции «больших данных» – в получении совершенно новых знаний из результатов выявления ранее незаметных взаимосвязей данных или поиска неочевидных фактов.

В качестве таких данных могут выступать:

- публикации (статьи, монографии, аналитические обзоры и т.п.);
- результаты научно-технической деятельности (НТД) (отчеты о НИР, патенты, ноу-хау, и т.п.);
- информация специализированных баз данных;
- комментарии на веб-сайтах, форумах, в социальных сетях;
- другие источники.

Основное отличие традиционных средств бизнес-аналитики (аналитическая обработка данных в реальном времени, технологии Data Mining) от технологий «больших данных» заключается в том, что они используют преимущественно структурированные данные. Проблема заключается в том, что в настоящее время огромные объемы информации существуют в неструктурированном виде. К таким данным можно отнести файлы различных форматов (фото, аудио и видео, электронная почта), сообщения (службы мгновенных сообщений, социальные сети, форумы и блоги) и т.д. Неструктурированные данные, в отличие от структурированных, не имеют ни известных типов атрибутов (например, Integer, Character), ни назначения (например, Salary, ZipCode), они неоднозначны и могут содержать различный смысл в зависимости от контекста. Например, в фразе «Tom Brown has brown eyes» встречающееся два раза слово «brown» имеет разный контекст, и компьютерные программы должны быть способны обнаруживать такие различия. Кроме того, такие данные часто носят субъективный характер.

Все это затрудняет обработку неструктурированной информации традиционными средствами. Для того, чтобы использовать в них неструктурированные данные, необходимо сначала преобразовать их в структурированные.

Следует добавить, что несмотря на внешнюю схожесть задач бизнес-аналитики и аналитических решений «больших данных», между ними существуют серьезные различия. Аналитики компании O'Reilly Radar выделяют три отличия [10]: «большие данные» предназначены для обработки более значительных массивов данных; «большие данные» предназначены для обработки более быстро получаемых и меняющихся сведений и требуют интерактивности; «большие данные» изначально неструктурированы и требуют интерпретации и очистки.

В заключение отметим, что ажиотаж вокруг модной в 2012 г. темы «больших данных» понемногу идет на убыль, что означает приближение периода зрелости данной технологии. Уже сегодня всемирно известные и нишевые вендоры предлагают решения в данной области. Эти решения существенно сокращают долю «ручного труда» при обработке массивов информации в задачах анализа, прогноза и экспертизы в научно-технической сфере, позволяют улучшить качество и достоверность получаемых результатов.

Представляется целесообразным в ближайшее время начать работы по реализации пилотного проекта по применению технологий обработки «больших данных» в интересах проведения экспертно-аналитических исследований.

В статье приведены результаты, полученные при выполнении работ в рамках Государственного задания 2015/Н7 Минобрнауки России.

Список литературы

1. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим // Манн, Иванов и Фербер, 2013.
2. Уайт Т. Hadoop. Подробное руководство // Питер, 2013.
3. Потенциал Больших Данных. Available at: <http://polit.ru/article/2013/03/11/lobzovsky>.
4. Фрэнкс Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики // Манн, Иванов и Фербер, 2014.
5. Шилина М.Г. Data-коммуникация как новый формат взаимодействия в публичном пространстве // Бизнес. Общество. Власть, 2014. № 19, с. 91–98.
6. Hype Cycle for Emerging Technologies. Available at: <http://www.gartner.com/newsroom/id/2819918>.
7. Белоусов В.Л., Дегтярев Ю.И., Сергеев М.В. Концептуальные основы формирования многоотраслевой системы государственной экспертизы // Автоматизация и современные технологии, 2013. № 6, с. 30–38.
8. Викулов О.В., Бухарин С.Н., Дивуева Н.А. Типовой технологический процесс проведения научно-технической экспертизы, реализованный в ФГБНУ НИИ РИНКЦЭ // Инноватика и экспертиза, 2014, № 2 (13), с. 101–114.
9. Литвак Б.Г. Экспертная информация. Методы получения и анализа. М. 2009.
10. Slocum M. Big data goes to work // O'Reilly Rada. Available at: http://radar.oreilly.com/2011/11/big-data-business-enterprise.html#disqus_thread.

References

1. Meyer-Shenberger V., Kukier K. (2013) *Bol'shie dannye. Revolyutsiya, kotoraya izmenit to, kak my zhivem, rabotaem i myslim* [Big Data. A revolution that will change the way we live, work and think]. *Mann, Ivanov i Ferber* [Mann, Ivanov and Ferber].
2. White T. (2013) *Hadoop. Podrobnoe rukovodstvo* [Hadoop. Detailed manual]. *Piter* [Peter].
- 3 *Potentsial Bol'shikh Danykh* [The Potential Of Big Data]. Available at: <http://polit.ru/article/2013/03/11/lobzovsky>.
4. Franks B. (2014) *Ukroshchenie bol'shikh danykh. Kak izvlekat' znaniya iz massivov informatsii s pomoshch'yu glubokoy analitiki* [Taming big data. How to extract knowledge from data arrays using deep analytics]. *Mann, Ivanov i Ferber* [Mann, Ivanov and Ferber].
5. Shilina M.G. (2014) *Data-kommunikatsiya kak novyy format vzaimodeystviya v publichnom prostranstve* [Data-communication as a new format of interaction in the public space]. *Biznes. Obshchestvo. Vlast'* [Business. Society. Power], no. 19, pp. 91–98.
6. Hype Cycle for Emerging Technologies. Available at: <http://www.gartner.com/newsroom/id/2819918>.
7. Belousov V.L., Degtyarev Y.I., Sergeev M.V. (2013) *Kontseptual'nye osnovy formirovaniya mnogo-otraslevoy sistemy gosudarstvennoy ekspertizy* [Conceptual bases of formation of a diversified system of state expert examination]. *Avtomatizatsiya i sovremennye tekhnologii* [Automation and modern technologies], no. 6, pp. 30–38.
8. Vikulov O.V., Bukharin S.N., Divuyeva N.A. (2014) *Tipovoy tekhnologicheskij protsess provedeniya nauchno-tekhnicheskoy ekspertizy, realizovanny v FGBNU NII RINKTsE* [Typical technological process of conducting scientific and technological expert examination, implemented SRI FRCEC]. *Innovatika i ekspertiza* [Innovation and assessment], no. 2 (13), pp. 101–114.
9. Litvak B.G. (2009) *Ekspertnaya informatsiya. Metody polucheniya i analiza* [Expert information. Methods of preparation and analysis]. Moscow.
10. Slocum M. Big data goes to wor. O'Reilly Radar. Available at: http://radar.oreilly.com/2011/11/big-data-business-enterprise.html#disqus_thread.