

УДК 311.42:658

**Н. Г. Синяк¹, А. Каклаускас², Д. В. Зинькина¹, А. Хабиб¹,
Н. Шариф¹, А. В. Бондаренко¹**

¹Белорусский государственный технологический университет

²Вильнюсский технический университет им. Гедиминаса

ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ В ОЦЕНКЕ ЭКОНОМИЧЕСКОГО ПОТЕНЦИАЛА ОРГАНИЗАЦИЙ

В статье были разъяснены самые широко используемые технологии больших данных в организациях, кратко рассмотрены созданные Э.-К. Завадскас и А. Каклаускасом методы многокритериального анализа: комплексный метод определения значимости критериев; метод многокритериальной комплексной пропорциональной оценки проектов; метод определения степени полезности и рыночной стоимости проектов. Жизненный цикл здания состоит из шести тесно взаимосвязанных этапов: проектирования, строительства, технического обслуживания, управления сооружениями, сносом и утилизацией. Жизненный цикл здания может иметь много альтернативных вариантов. Эти варианты основаны на альтернативном дизайне, проектировании, строительстве, техническом обслуживании, управлении объектами, сносе, процессами утилизации и их составных частей. При создании этих систем интернетной поддержки решений применялись следующие основные принципы и методы: комплексного анализа, функционального анализа, оптимизации соотношения полезности и стоимости, взаимосвязей разных наук, многокритериального вариантного проектирования проектов, многокритериального анализа проектов, зависимости степени эффективности проектов от заинтересованных групп и их целей.

Ключевые слова: большие данные, технологии больших данных, многокритериальный анализ, интернетная поддержка решений, методы анализа.

**N. G. Sinyak¹, A. Kaklauskas², D. V. Zin'kina¹, A. Habib¹,
N. Sharif¹, A. V. Bondarenko¹**

¹Belarusian State Technological University

²Vilnius Gediminas Technical University

USING BIG DATA IN ASSESSMENT OF THE ECONOMIC POTENTIAL OF ORGANIZATIONS

In the article there were explained the most widely used technologies of big data in organizations. Created by E.-K. Zavadskas and A. Kaklauskas methods of multi-criteria analysis such as a comprehensive method for determining the significance of criteria; method of multi-criteria complex proportional evaluation of projects; method for determining the rate of usefulness and the market value of the projects were briefly considered. A building life cycle consists of seven closely interrelated stages: brief, design, construction, maintenance, facilities management, demolition and utilisation. A building life cycle may have a lot of alternative versions. These variants are based on the alternative brief, design, construction, maintenance, facilities management, demolition, utilisation processes and their constituent parts. When creating these systems of internet decision support, the following basic principles and methods have been applied: complex analysis, functional analysis, optimization of the usefulness-value ratio, interaction of the various sciences, multi-criteria variant designing of projects, multi-criteria analysis of projects, dependence on the degree of effectiveness of the projects from the interested groups and their goals.

Key words: big data, big data technology, multi-criteria analysis, interactive decision support, analysis methods.

Введение. Технологии не стоят на месте и сегодня при оценке экономического потенциала организаций ведущую роль играет их готовность использования новых технологий, особенно из области аналитики больших данных.

В гиперконкурентном мире, вынуждающем соперничающие компании постоянно снижать маржинальность, бизнес рассматривает большие данные как возможность получить абсолютное оружие в борьбе за выживание.

Большие данные – серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, сформировавшихся в конце 2000-х годов, альтернативных традиционным системам управления базами дан-

ных и решениям класса Business Intelligence. Под экономическим потенциалом организации понимается способность предприятия финансировать свою деятельность. Оно характеризуется обеспеченностью финансовыми ресурсами, необходимыми для нормального функционирования предприятия, целесообразностью их размещения и эффективностью использования, финансовыми взаимоотношениями с другими юридическими и физическими лицами, платежеспособностью и финансовой устойчивостью [1].

Экономический потенциал организации может быть устойчивым, неустойчивым и кризисным. Способность предприятия своевременно производить платежи, финансировать свою деятельность на расширенной основе свидетельствует о его хорошем финансовом состоянии.

Финансовое состояние предприятия зависит от результатов его производственной, коммерческой и финансовой деятельности. Если производственный и финансовый планы успешно выполняются, то это положительно влияет на финансовое положение предприятия. И наоборот, в результате невыполнения плана по производству и реализации продукции происходит повышение ее себестоимости, уменьшение выручки и суммы прибыли и, как следствие, ухудшение финансового состояния предприятия и его платежеспособности. Устойчивый экономический потенциал в свою очередь оказывает положительное влияние на выполнение производственных планов и обеспечение нужд производства необходимыми ресурсами. Поэтому финансовая деятельность как составная часть хозяйственной деятельности направлена на обеспечение планомерного поступления и расходования денежных ресурсов, выполнение расчетной дисциплины, достижение рациональных пропорций собственного и заемного капитала и наиболее эффективного его использования. Большие данные используются во многих сферах, в том числе в оценке экономического потенциала организаций.

Основная часть. Так же, как и с облачными технологиями, то, что один человек подразумевает при обсуждении больших данных, совсем не обязательно совпадает с тем, что под этим понимает его собеседник.

Из названия можно предположить, что термин «большие данные» относится просто к управлению и анализу больших объемов данных. Согласно отчету McKinsey Institute «Большие данные: новый рубеж для инноваций, конкуренции и производительности» (Big data: The next frontier for innovation, competition and productivity), термин «большие данные»

относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации. И мировые репозитории данных, безусловно, продолжают расти. В представленном в середине 2011 г. отчете аналитической компании IDC «Исследование цифровой вселенной» (Digital Universe Study), подготовку которого спонсировала компания EMC, предсказывалось, что общий мировой объем созданных и реплицированных данных в 2011-м может составить около 1,8 зеттабайта (1,8 трлн. гигабайт) – примерно в 9 раз больше того, что было создано в 2006-м.

В последние годы Большие Данные являются общепризнанным трендом экономического и технологического развития. Им посвящены тысячи публикаций. Они относятся к числу наиболее популярных тем, как в специализированных изданиях, так и в различного рода СМИ, рассчитанных на самую широкую аудиторию. В результате возникло впечатление, что Большие Данные – это нечто само собой разумеющееся, ясное, понятное.

Между тем, дело обстоит совсем не так. Беспристрастный анализ фактических данных показывает, что Беларусь существенно отстает в сфере Больших Данных. Значительная часть компаний только использует этот термин в маркетинговых целях, а по сути, применяет старую, хорошо известную бизнес-аналитику, которая заметно отличается от Больших Данных. В стране практически нет спроса на специалистов по Большим Данным. Достаточно посмотреть наиболее популярные порталы работ, чтобы убедиться, что спрос на специалистов в этой сфере у нас на порядки меньше, чем в США, Европе, Японии, Китае.

Происходящее вызывает особую тревогу в условиях, когда ведущие наднациональные мировые структуры и транснациональные корпорации, правительства ведущих стран мира, бизнес самых различных масштабов, системы управления производственной и социальной инфраструктурой и, конечно же, военноразведывательный комплекс всех основных стран мира уже используют Большие Данные как важнейший стратегический ресурс.

По данным компании Cisco, объем сгенерированных данных в 2012 г. составил 2,8 зеттабайт и увеличится до 40 зеттабайт к 2020 г. Примерно треть передаваемых данных составляют автоматически сгенерированные данные, т. е. управляющие сигналы и информация, характеризующие работу машин, оборудования, устройств, присоединенных к интернету, или к интернету вещей. Причем с каждым годом доля интернета вещей или как его сегодня еще

называют «интернета всего» растет в общем объеме мировых информационных потоков. На 40% ежегодно увеличивается объем корпоративной информации, передаваемой и хранящейся в сети интернет.

Еще в 2011 г. McKinsey Global Institute объявил Большие Данные «следующим рубежом для инноваций, конкуренции и производительности». По данным целого ряда ведущих международных деловых изданий, уже сегодня Большие Данные дают заметный эффект в бизнесе. Например, выяснилось, что в транснациональных компаниях, входящих в список Fortune 500, где, казалось бы, до мелочей отлажены все процедуры и процессы, внедрение технологий Больших Данных на 5–7% увеличило эффективность использования ресурсов – труда, основных производственных фондов, энергии и т. п. и на 7–9% обеспечило рост объемов продаж. Для среднего бизнеса показатели оказались в полтора – два раза выше. Причем, следует отметить, что данные получены в условиях, когда мировая экономика испытывает на себе последствия глубочайшего финансово-экономического кризиса и экономический рост составляет в лучшем случае 1–2%.

На чем же базируется эффективность Больших Данных? Технологии Больших Данных и, прежде всего, методы статистического анализа, компьютерного распознавания образов и т. п., применяемые на огромных, постоянно пополняемых массивах данных позволяют:

- проводить самые различные и сколь угодно подробные классификации той или иной совокупности людей, компаний, иных объектов по самым разнообразным признакам. Такие классификации обеспечивают точное понимание взаимосвязи тех или иных характеристик любого объекта – от человека до компании или организации, с теми или иными его действиями;

- осуществлять многомерный статистический и иной математический анализ. Этот анализ позволяет находить корреляции между самыми различными параметрами, характеристиками, событиями и т. п. Корреляции не отвечают на вопрос – почему. Они показывают вероятность, с которой при изменении одного фактора изменится и другой. В каком-то смысле Большие Данные представляют собой альтернативный традиционной науке метод. Наука на основе теоретических моделей отвечает на вопрос – почему, а затем, получив ответ, делает рекомендации, как действовать. В случае корреляции стадия поиска причины ликвидируется, а действие происходит в тех случаях, когда факторы тесно взаимосвязаны и на один из факторов легко или возможно осуществить целенаправленное воздействие;

- прогнозировать. На основе классификаций и аналитических выкладок осуществляется прогнозирование. Суть прогнозирования состоит в том, чтобы на основе корреляции определить наиболее легкий способ воздействия для того, чтобы один набор факторов, характеризующих тот или иной объект, лицо, компанию, событие и т. п. был преобразован в другой.

По этому поводу руководитель одной из самых перспективных компаний в области прогнозирования Quid, также принадлежащей П. Тиллю, Ш. Горли сказал: «Наибольший эффект Большие Данные дают тогда, когда возможности компьютеров в обработке гигантских массивов информации и выявлении нетривиальных связей соединены с человеческим опытом и профессиональной интуицией. А все, что вам рассказывает Р. Курцвейл про искусственный интеллект, это как минимум на ближайшее будущее просто красивые истории и PR ходы» [2].

В области оценки экономического потенциала организаций Большие Данные вместе с человеческим опытом дают потрясающий эффект.

Тем не менее «большие данные» предполагают нечто большее, чем просто анализ огромных объемов информации. Проблема не в том, что организации создают огромные объемы данных, а в том, что большая их часть представлена в формате, плохо соответствующем традиционному структурированному формату БД, – это веб-журналы, видеозаписи, текстовые документы, машинный код или, например, геопространственные данные. Все это хранится во множестве разнообразных хранилищ, иногда даже за пределами организации. В результате корпорации могут иметь доступ к огромному объему своих данных и не иметь необходимых инструментов, чтобы установить взаимосвязи между этими данными и сделать на их основе значимые выводы.

В сущности понятие больших данных подразумевает работу с информацией огромного объема и разнообразного состава, весьма часто обновляемой и находящейся в разных источниках в целях увеличения эффективности работы, создания новых продуктов и повышения конкурентоспособности. Консалтинговая компания Forrester дает краткую формулировку: «Большие данные объединяют техники и технологии, которые извлекают смысл из данных на экстремальном пределе практичности».

Последний отчет O'Reilly Strata Conference, составленный по результатам опроса ста участников конференций, показал следующие результаты:

- 18% уже имеют решение для работы с большими данными;

– 28% не имеют планов в этой области в настоящее время;

– 22% планируют внедрить решение по работе с большими данными в ближайшие шесть месяцев, 17% – в течение года, 15% – в течение двух лет.

Проведенный фирмой Techaisle в США опрос восьмисот компаний, относящихся к сегменту СМБ (средний и малый бизнес), показал, что независимо от размера бизнеса примерно треть опрошенных заинтересована в перспективе работы с большими данными. Их главной проблемой был недостаток опыта. Из этого можно сделать вывод, что использование больших данных повышает экономический потенциал организаций.

Работа с большими данными не похожа на обычный процесс бизнес-аналитики, где простое сложение известных значений приносит результат: например, итог сложения данных об оплаченных счетах становится объемом продаж за год. При работе с большими данными результат получается в процессе их очистки путем последовательного моделирования: сначала выдвигается гипотеза, строится статистическая, визуальная или семантическая модель, на ее основании проверяется верность выдвинутой гипотезы и затем выдвигается следующая. Этот процесс требует от исследователя либо интерпретации визуальных значений или составления интерактивных запросов на основе знаний, либо разработки адаптивных алгоритмов «машинного обучения», способных получить искомый результат. Причем время жизни такого алгоритма может быть довольно коротким.

До появления больших данных традиционный анализ предусматривал занесение информации в традиционную БД. Этот подход основывался на модели реляционных БД, использующей таблицы для хранения данных и отношений между ними. Данные обрабатывались и запоминались в строках таблиц.

С течением времени БД совершенствовались и сейчас используют массовую параллельную обработку, когда данные разбиваются на небольшие группы и обрабатываются одновременно на многих узлах, что обеспечивает значительное ускорение работы. Вместо хранения информации в строках таблиц БД могут также использовать архитектуры столбцов, которые позволяют обрабатывать только столбцы с данными, необходимыми для формирования результатов запроса, и, кроме того, поддерживают хранение неструктурированной информации.

Технология MapReduce представляет собой комбинацию двух функций, улучшающих обработку данных. Сначала map-функция

разделяет данные на несколько групп, которые затем обрабатываются параллельно. Затем reduce-функция объединяет результаты расчетов в варианты ответов.

Компания Google использовала технологию MapReduce для индексирования сети Интернет и получила патент на свою MapReduce-платформу. Однако постепенно эта методика начинает использоваться все шире и шире, наибольшую известность получила ее реализация в проекте Hadoop на основе открытого кода (об этом чуть ниже).

Подобно технологии MapReduce, MPP обрабатывает данные, распределяя их по множеству узлов, которые выполняют обработку распределенных данных одновременно. На основании полученных данных собирается общий результат.

Однако для обращения к MPP-продуктам используется язык запросов SQL, в то время как MapReduce непосредственно контролируется с помощью Java-кода. MPP также широко используется на специализированных дорогих аппаратных платформах (иногда называемых программно-аппаратными комплексами для работы с большими данными), в то время как для работы технологии MapReduce применяются обычные компьютерные системы.

Методика обработки сложных событий предполагает онлайн-обработку информации из различных источников, зависящей от времени. Например, данных о местоположении мобильных телефонов или информации от систем датчиков для прогнозирования, выделения или классификации интересующих событий. В частности, данные, поступающие от датчиков, могут лечь в основу предсказания сбоя оборудования, даже если они кажутся совершенно не связанными друг с другом. Обрабатывать сложные события в больших объемах данных можно с помощью технологии MapReduce, разделяя информацию на порции, не связанные друг с другом. Например, данные датчиков для каждого участка оборудования могут пересылаться для обработки на отдельный узел.

Платформа Hadoop на основе открытого кода была разработана с помощью технологии MapReduce для параллельной обработки больших объемов информации на множестве узлов стандартного недорогого оборудования.

Информация разделяется на блоки и загружается в файловое хранилище данных, например Hadoop Distributed File System (HDFS), организованное как несколько избыточных узлов на недорогом запоминающем устройстве. Узел name протоколирует размещение данных на конкретных узлах. Данные реплицируются

более чем на одном узле, что обеспечивает их сохранность в случае выхода какого-либо узла из строя.

Затем данные можно анализировать с помощью технологии MapReduce, которая определяет местонахождение необходимых для расчета сведений из узла name. После этого обработка на узлах идет параллельно. Результаты расчетов обобщаются для составления ответа на запрос и затем загружаются на узел, который впоследствии доступен для анализа с помощью других инструментов. В качестве альтернативы возможна загрузка полученных сведений в традиционные хранилища для обработки с помощью транзакций. В качестве наиболее примечательного дистрибутива Hadoop называют Apache.

Системы управления базами данных на основе парадигмы NoSQL не похожи на системы управления реляционными БД, поскольку не используют SQL в качестве языка запросов. Такие системы считаются более подходящими для работы с данными, не укладывающимися в стандартные таблицы реляционных БД. Они обходятся без перезагрузки индексирования, схемы и ACID-свойств транзакций при создании огромных, реплицированных хранилищ данных для выполнения аналитических работ на экономичном оборудовании, подходящем для работы с неструктурированными данными.

Cassandra представляет собой альтернативу для Hadoop HDFS, – это БД, выполненная как NoSQL.

Базы данных, подобные файловому хранилищу Hadoop, затрудняют ad hoc-запросы и аналитику, поскольку требуемое для этого программирование функций map/reduce является весьма сложной задачей. Осознав это обстоятельство в процессе работы с Hadoop, компания Facebook создала собственное хранилище данных Hive, преобразующее SQL-запросы в задания map/reduce, которые выполняются с использованием Hadoop.

Для оценки экспортного потенциала предприятий должны использоваться смешанные (количественные и качественные) методы. Чисто количественные методы основаны на статистических сведениях, которые не способны предоставить качественную информацию. Качественный подход отражает изменчивость поведения в зависимости от контекста. Он сумбурен и неаккуратен.

Стратегия проведения исследования оценки экономического потенциала организаций означает нахождение способа, при помощи которого будут решены проблемы и задачи исследования и достигнуты его цели.

Процесс научного исследования условно разделен на четыре этапа. Первый этап – подготовка к исследованию. Он связан с планированием исследования, т. е. методологическим обоснованием исследования. В нем следует выделить следующее: изучение литературы по вопросу, формулирование темы, обсуждение проблемы исследования (вопроса исследования), характеристику объекта исследования, формулирование гипотезы, установление целей и задач исследования.

Второй этап – организация процесса исследования.

Третий этап – сбор эмпирических данных. На этом этапе собираются данные, способные подтвердить или опровергнуть гипотезу исследования.

Четвертый этап – обработка данных исследования. Применялись следующие методы многокритериального анализа: комплексного определения значимости критериев в зависимости от их качественных и количественных характеристик; многокритериальной комплексной пропорциональной оценки альтернатив; многокритериального вариантного проектирования жизненного процесса альтернатив.

Ввиду многообразия критериев непосредственное сравнение проектов представляет собой сложную задачу. Одной из главнейших целей при решении этой проблемы является определение значимостей критериев. Физический смысл конкретной значимости критерия в том, во сколько раз его полезность для проекта при комплексной оценке альтернатив больше (меньше) другого критерия.

С помощью данных методов были проанализированы: процесс существования здания, средние показатели степеней полезности и предполагаемых стоимостей вариантов проектов, коды альтернатив процесса существования здания и предоставляемой этими кодами концептуальной и количественной информации и др. [3].

Заключение. Большие данные используются относительно недавно. Однако, компании, не использующие их в своей деятельности на данный момент, уже только этим уступают своим конкурентам на 20%. Были описаны различные технологии больших данных, используемые компаниями. А. Каклаускас и Э.-К. Завадскас создали новые многокритериальные методы анализа больших данных, которые могут применяться для оценки экономического потенциала организаций. Они разделили процесс исследования на четыре этапа: подготовка к исследованию, организация процесса исследования, сбор эмпирических данных, обработка данных исследования.

Литература

1. Каклаускас А., Завадскас Э.-К. Биометрическая и интеллектуальная поддержка решений: монография. Вильнюс: Техника, 2012. С. 344.
2. Ардити Д., Гунайдин Х. М. Восприятие качества процесса в строительных проектах // Журнал по вопросам управления в машиностроении. 1999. № 12. С. 18–21.
3. Банайтиене Н., Банайтис А., Каклаускас А., Завадскас Э.-К. Оценка жизненного цикла здания: многовариантные и множественные критерии подхода // Международный журнал наук управления. 2008. № 36. С. 429–441.

References

1. Kaklauskas A., Zavadskas E.-K. *Biometricheskaya i intelektual'naya podderzhka resheniy: monografiya* [Biometric and intelligent decision support: monograph]. Vilnius: Technique Publ., 2012. 344 p.
2. Arditi D., Gunaydin Kh. M. Perceptions of process quality in building projects. *Zhurnal po voprosam upravleniya v mashinostroenii* [Journal of Management in Engineering], 1999, no. 12, pp. 18–21.
3. Banaytiene N., Banaytis A., Kaklauskas A., Zavadskas E.-K. Evaluating the life cycle of a building: A multivariant and multiple criteria approach. *Mezhdunarodniy zhurnal nauk upravleniya* [International Journal of Management Science], 2008, no. 36, pp. 429–441.

Информация об авторах

Синяк Николай Георгиевич – кандидат экономических наук, доцент, заведующий кафедрой организации производства и экономики недвижимости. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: siniakn@mail.ru

Артурас Каклаускас – доктор наук, профессор, заведующий кафедрой экономики строительства и менеджмента недвижимого имущества строительного факультета. Вильнюсский технический университет им. Гедиминаса. E-mail: arturas.kaklauskas@vgtu.lt

Зинькина Диана Васильевна – магистрант кафедры организации производства и экономики недвижимости. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: tzinkina@mail.ru

Хабиб Авада – аспирант кафедры организации производства и экономики недвижимости. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь).

Шариф Нуреддин – аспирант кафедры организации производства и экономики недвижимости. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь).

Бондаренко Анна Владимировна – аспирант кафедры организации производства и экономики недвижимости. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь).

Information about the authors

Sinyak Nikolay Georgievich – Ph. D. Economics, Assistant Professor, Head of the Department Production Organization and Real Estate Economics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: siniakn@mail.ru

Kaklauskas Arturas – D. Sc., Professor, Head of Department of Construction Economics and Property Management. Vilnius Gediminas Technical University (LT-10223, Vilnius, 11, Sauletekio str., Lithuania). E-mail: arturas.kaklauskas@vgtu.lt

Zin'kina Diana Vasil'yevna – undergraduate student, the Department of Production Organization and Real Estate Economics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: tzinkina@mail.ru

Habib Avada – graduate student, the Department of Production Organization and Real Estate Economics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus).

Sharif Nureddin – graduate student, the Department of Production Organization and Real Estate Economics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus).

Bondarenko Anna Vladimirovna – graduate student, the Department of Production Organization and Real Estate Economics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus).

Поступила 10.07.2015