

13. Топ главных событий рынка упаковки [Электронный ресурс]. – Основа. – 2016. – Режим доступа: <http://article.unipack.ru/60863>
14. Ecolan – легкая упаковка [Электронный ресурс]. – Режим доступа: <https://www.ecolean.com/ru/package/#>
15. Лучко, А. 13 примеров «умного» дизайна упаковок [Электронный ресурс] / А. Лучко // Look at me. – 2014. – Режим доступа: <http://www.lookatme.ru/mag/live/inspiration-lists/202077-smart-package>
16. ДСТУ 4518:2008. Продукти харчові. Маркування для споживачів [Текст]. – К.: Держспоживстандарт України, 2008. – 39 с.
17. Про основні принципи та вимоги до безпечності та якості харчових продуктів [Текст]. – Верховна Рада України, 1997. – № 771/97-ВР. – Режим доступа: <http://zakon3.rada.gov.ua/laws/show/771/97-%D0%B2%D1%80>
18. Гигиенические нормативы ГН 2.3.972-00 [Текст]. – М.: Федеральный центр госсанэпиднадзора Минздрава России, 2000. – 55 с. – Режим доступа: http://docs.nevacert.ru/files/sanpin/gn_2.3.972-00.pdf
19. Про вимоги до предметів та матеріалів, що контактують з харчовими продуктами [Текст]. – Міністерство аграрної політики та продовольства України, 2012. – № 9450. – Режим доступа: <http://minagro.gov.ua/node/18243>

Дата надходження рукопису 04.01.2017

Калініна Олена Сергіївна, інженер I категорії, кафедра метрології, стандартизації та сертифікації, Національний університет «Львівська політехніка», вул. С. Бандери, 12, м. Львів, Україна, 79013
E-mail: anel105@rambler.ru

Байцар Роман Іванович, доктор технічних наук, професор, кафедра метрології, стандартизації та сертифікації, Національний університет «Львівська політехніка», вул. С. Бандери, 12, м. Львів, Україна, 79013
E-mail: baitzar@ukr.net

УДК 658.51/52

DOI: 10.15587/2313-8416.2017.93457

АНАЛИЗ ТЕХНОЛОГИЙ «BIG DATA» НА БАЗЕ ДИСКРЕТНО-ВЕРОЯТНОСТНОЙ МЕРЫ ИНФОРМАЦИИ

© **И. В. Руженцев, С. В. Луцкий**

В статье рассматриваются вопросы системно-информационного подхода к анализу видов технологий обработки «Big Data» с позиции определения меры дискретно-вероятностной информации объектов при исследовании реального физического мира. Изложены основные недостатки современных технологий обработки «Big Data». Представлена структура самоорганизационной информационной системы «Big Data». Предложена структура видов технологий обработки «Big Data» с позиции дискретно-вероятностной информационной меры

Ключевые слова: мера, технологии «Big Data», системно-информационный подход, дискретно-вероятностная информация

1. Введение

Информационная система «Big Data» является компьютерным отражением реального физического мира, в тоже время является его частью. Она должна подчиняться общим информационным законам Вселенной. Отсутствие понимания общих информационных законов и закономерностей реального физического мира не позволяет в полной мере понять принципы самоорганизации информационной системы «Big Data». Свойство самоорганизации реального физического мира является фундаментальным атрибутом Мироздания, которое находится в основе принципов его существования и развития.

Многозначное определение понятия «информации» в научной среде сдерживает развитие информационных технологий «Big Data» и требует новых принципов и подходов при разработке «современной теории информации», которая бы объединила новые и старые подходы.

Разработка научных основ, методов и подходов по разработке технологий обработки самоорганизационной информационной системы «Big Data» на базе «системно-информационного (СИ) подхода» является требуемой необходимостью для развития перспективных информационных технологий «Big Data».

2. Литературный обзор

Гипотеза о том, что Вселенная является цифровым компьютером, впервые была выдвинута Конрадом Цузе в книге Rechner Raum («Вычислительное пространство»), переведена на английский язык как Calculating Space). Термин «цифровая физика» использовался Эдвардом Фрэдкином, который потом предпочел термин «цифровая философия». Среди тех, кто рассматривал Вселенную как гигантский компьютер находятся Стивен Вольфрам, Юрген Шмидхубер и нобелевский лауреат Герард 'т Хоофт.

Эти авторы считали, что несомненно вероятностная природа квантовой физики не обязательно является несовместимой с идеей вычислимости. Квантовая версия цифровой физики недавно была предложена Сетом Ллойдом, Дэвидом Дойчем и Паоло Цицци.

Похожими идеями являются теория протоальтернатив Карла Фридриха фон Вайцзеккера, панкомпьютериализм, вычислительная теория Вселенной, теория «вещества из информации» Джона Уилера и гипотеза математической Вселенной («Конечный ансамбль») Макса Тегмарка.

Взгляд на Вселенную как на большую вычислительную машину, или как сеть вычислительных процессов, которая вычисляет следующее состояние фундаментальных физических законов из текущего состояния предполагает, что существует, по крайней мере, в принципе – программа, которая вычисляет в реальном времени эволюцию Вселенной. Этот компьютер может быть, например, гигантским клеточным автоматом (Цузе, 1967), или универсальной машиной Тьюринга, как предположил Шмидхубер (1997). Они обращали внимание на то, что существует очень короткая программа, которая может вычислить все возможные вычислимые Вселенные асимптотически оптимальным путем.

Представление о том, что информация является фундаментальной величиной в ядре физики, принадлежит Фредерику Кантору, физику из Колумбийского университета «Информационная механика» (1977). 1990 году Джон Арчибалд Уиллер высказал предположение, что информация является фундаментальной концепцией физики. Согласно его доктрине «it from bit» все физические сущности являются информационно-теоретическими в своей основе. Дэвид Чалмерс из Австралийского национального университета высказал мнение, что информация истинно фундаментальна, и по которой она обладает двумя базовыми аспектами, соответствующими физической и воспринимаемой сторонами действительности. Кристофер Ланган также усилил взгляды Уиллера в своей эпистемологической метатеории.

Развитие информационных технологий обработки «Big Data» по принципу Вселенского цифрового компьютера в настоящее время является актуальной темой в мире. Приведение к единому информационному знаменателю базы данных «Big Data» послужит локомотивом для решения накопившихся задач.

Системно-информационный подход [1] является новым научным направлением в области информационного исследования процессов и систем реального физического Мира.

Системно-информационный подход к исследованию процессов и систем как научное направление базируется на:

а) разработанной концепции определения понятия меры дискретно-вероятностной информации [2, 3];
б) методологии численного определения количества, качества и ценности дискретно-вероятностной информации [4];

в) научных принципах дискретно-вероятностной информационной алгебры [5];

г) сформулированных законах и закономерностях дискретно-вероятностной информации [6];

д) методологии системно-информационный подхода к моделированию процессов и систем [7];

е) дискретно-вероятностных моделях процессов и систем [7];

ж) методологии системно-информационный подхода к анализу и синтезу процессов и систем [7, 8];

з) методологии разработки критических технологий на базе дискретно-вероятностных моделей процессов и систем [9].

Концептуальной основой методологии системно-информационный подхода являются системно-информационные принципы определения числового значения количества, качества и ценности дискретно-вероятностной информации, которыми характеризуется объект [10].

К категории «Big Data» относится информация, которую уже невозможно обрабатывать традиционными способами, в том числе структурированные данные, медиа и случайные объекты. Термин «Big Data» относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации.

«Big Data» рассматриваются сразу в трех плоскостях (Gartner):

1) рост объемов информации;

2) рост скорости обмена данными;

3) увеличение информационного разнообразия.

Проблема «Big Data» заключается в том, что большая их часть представлена в формате, плохо соответствующем традиционному структурированному формату баз данных, – это веб-журналы, видеозаписи, текстовые документы, машинный код и т. д. Данные самых разных типов, как структурированные, так и неструктурированные хранятся 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.

Необходимо учесть, что данные обновляются все чаще, и возникает ситуация, в которой традиционные методы анализа информации не могут угнаться за огромными объемами постоянно обновляемых данных, что в итоге и открывает дорогу технологиям «Big Data».

Основная особенность используемых в рамках концепции подходов «Big Data», является возможность обработки информационного массива целиком для получения более достоверных результатов анализа.

Основные методики анализа массивов данных, базируются на основе инструментария заимствованного из статистики и информатики.

В ожидании новых возможностей, которые принесут с собой технологии обработки «Big Data»,

уже сейчас многие организации организуют процесс сбора и хранения различного рода информации. По мнению *Gartner*, в гонке данных победят именно те, кто научится обращаться с самыми разными *источниками информации*.

Задача по «освоению» «Big Data» является уникальной. Главным ответом компаний на вызовы «Big Data» является модернизация корпоративных центров обработки данных (ЦОД).

Разнообразие задач и вендорские решения в сфере «Big Data» пока не приобрели ярко выраженной отраслевой направленности. Рынок находится на стадии активного формирования, в самом начале этой стадии.

Недостатки технологий «Big Data» на современном этапе заключаются в следующем:

1. Объективная потребность в новых технологиях «Big Data» – определена.
2. Цель и задачи «Big Data» – определены.
3. Проблемные аспекты «Big Data» – определены частично.
4. Области применения технологий «Big Data» – определены частично.
5. Открытые системы анализа «Big Data» используемые в настоящее время – определены частично.
6. Основные направления развития открытых систем «Big Data» – определены частично.
7. Источники формирования «Big Data» – определены частично.
8. Методики анализа массивов данных – определены частично.
9. Спектр развития технологий «Big Data» – не определен.
10. Компании, заинтересованные в технологиях «Big Data» – определены частично.
11. Проблема ценообразования технологий «Big Data» – определена частично.

Резюме. Главный недостаток современного подхода к методологии разработок технологий «Big Data» заключается: первое – отсутствие единого подхода к определению *количественной, качественной и ценностной меры информации* «Big Data», второе – отсутствие единой *информационной инфраструктуры* корпоративных центров обработки данных как структурированной, так и неструктурированной информации, третье – отсутствие единого подхода к программно-аппаратным решениям и сопутствующим сервисам, направленных на *выявление скрытых информационных закономерностей* в «Big Data».

3. Цель и задачи исследования

Цель исследования – анализ технологий формирования, обработки и использования «Big Data» с позиции определения меры дискретно-вероятностной информации объектов реального физического Мира.

Для достижения цели были поставлены следующие задачи:

1. Определить основные научные положения методологии системно-информационного подхода к

технологиям формирования, обработки и использования «Big Data».

2. Определить особенности алгоритмов синтеза технологий «Big Data».

3. Определить задачи при разработке технологий «Big Data» на базе теоретических положений системно-информационного подхода.

4. Сформулировать основополагающие принципы формирования, обработки и использования «Big Data».

5. Разработать структурную схему самоорганизационной информационной системы «Big Data».

4. Анализ технологий формирования, обработки и использования «Big Data» на базе дискретно-вероятностной меры информации.

Основные научные положения методологии системно-информационного подхода к технологиям обработки «Big Data» заключаются в следующем.

1. Методология системно-информационного подхода позволяет рассчитывать количество, качество и ценность дискретно-вероятностной информации как структурированных, так и неструктурированных данных любой природы, на базе которой создается единая *информационная инфраструктура* для корпоративных центров обработки данных.

2. *«Информационная структура»* созданная на базе дискретно-вероятностной информации подчиняется *общим закономерностям*, выявленных методологией системно-информационного подхода.

3. Структурированные и неструктурированные данные подчиняются одним и тем же разработанным по системно-информационному подходу *дискретно-вероятностным информационным законам*, на основе которых *выявляются скрытые, конкретные закономерности* в «Big Data»

4. Алгоритм функционирования «Big Data» должен основываться на принципах *самоорганизации структур данных*, теоретические основы которых обеспечивает методология системно-информационного подхода.

5. Основной информационной платформой для решения задач на базе системно-информационного подхода можно использовать разработанные нами принципы программного обеспечения «DISLUT».

6. Проблема ценообразования решается расчетом количества дискретно-вероятностной информации, которым обладает продукт – (скрытая закономерность) цена продукта эквивалентна дискретно-вероятностной информации, которая была передана продукту при его производстве (это наиболее объективный принцип формирования стоимости продукта и измеряется она в битах).

Особенности алгоритма синтеза технологий формирования и обработки «Big Data» заключаются в следующем.

1. Особенность алгоритма самоорганизационной информационной системы «Big Data» на базе системно-информационного подхода заключается в структуризации данных по информационным признакам.

2. На основе теоретических положений системно-информационного подхода разрабатывается база данных с перечнем информационных признаков для свойств объектов любой природы.

3. Данные корпоративных центров постоянно добавляются с присвоением признаков из базы данных и на этой же основе самоструктурируются с присвоением коэффициентов информационной связи по отношению к данным с другими разрешенными признаками.

4. Заказчик информации «Big Data» вводит «заявку», при этом алгоритм выделяет перечень признаков свойств объектов, которые соответствуют «заявке» клиента.

5. По «заявке» производится выборка данных посредством коэффициентов информационной связи по признакам из базы данных, а выбранная информация подвергается требуемому анализу и выводится результат.

Задачи по разработке технологий формирования и обработки «Big Data» на базе теоретических положений системно-информационного подхода.

1. Разработать на базе системно-информационного подхода основы технологий (метод, способ, средство, процесс) «Big Data».

2. Разработать алгоритмы приема и преобразования дискретно-вероятностной информации для системы «Big Data».

3. Разработать алгоритмы хранения дискретно-вероятностной информации для системы «Big Data».

4. Разработать алгоритмы переработки дискретно-вероятностной информации для системы «Big Data».

5. Разработать алгоритмы передачи дискретно-вероятностной информации для системы «Big Data».

6. Разработать генетические основы дискретно-вероятностной информации для системы «Big Data».

7. Разработать классификацию внутренних факторов воздействия на систему «Big Data».

8. Разработать классификацию внешних факторов воздействия на систему «Big Data».

9. Разработать функциональный анализ синтеза структур коэффициентов информационной связи элементов системы «Big Data».

10. Разработать принципы самоорганизации системы «Big Data» под воздействием внутренних факторов воздействия.

11. Разработать принципы самоорганизации системы «Big Data» под воздействием внешних факторов воздействия.

12. Разработать методологию математического обеспечения системы «Big Data».

13. Разработать методологию программного обеспечения системы «Big Data».

14. Разработать программные продукты по управлению системой «Big Data» на основе системно-информационного подхода.

15. Разработать алгоритмы, которые позволяют строить гибкие запросы при обработке массивов данных «Big Data».

Основополагающие принципы формирования системы «Big Data».

1. Объекты окружающей среды объединены в системы: физические, биологические, технические, информационные и социальные, которые подчиняются одним и тем же информационным законам.

2. Окружающая среда состоит из экологических ниш, в которой существуют виды объектов систем.

3. Экологическая ниша есть объем в n -мерном пространстве с минимальными и максимальными значениями факторов среды.

4. Каждый объект экологической ниши с одной стороны представляет собой определенный вид, к которому он относится, с другой стороны в нем заложена генетическая информация (дискретно-вероятностная), которая строго индивидуальна.

5. Виды объектов в экологической нише характеризуются определенным количеством признаков, которые отличают объекты друг от друга.

6. Место разнообразия видов объектов в экологической нише имеет строгое место и подчиняется степенному закону, который был открыт на основе большого количества статистических данных, (предложен в 1928 г. С. Гартсайденом, развивался в работах Р. Фишера, А. Коберта, Престона, Мак-Артура, что подтверждает вывод Ч. Дарвина о вероятностном распределении популяции биологических видов в экосистеме).

7. Объекты в экологической нише, в зависимости от заложенной в них генетической информации, ведут борьбу за место на базе закона информационного отбора (*естественный отбор в биол.*), т. е. их место в n -мерном пространстве постоянно меняется во времени самоорганизационно.

8. Пространство «Big Data» состоит из множества экологических ниш, в которых в определенном месте находятся данные видов объектов, при пополнении данных структура «Big Data» постоянно меняется по определенному закону самоорганизационно.

9. Алгоритм, который позволяет строить гибкие запросы, состоит из:

– определения вида запроса, в соответствии с приведенной классификацией;

– определения места объекта в n -мерном пространстве «Big Data» на основе известного закона распределения данных видов объектов.

10. Приведенный принцип заложен природой и является определяющим в издержках на обслуживание структуры «Big data».

Структура самоорганизационной информационной системы «Big data» на представлена на рис. 1.

Структура видов технологий обработки «Big Data» на дискретно-вероятностной информационной основе для внутренних и внешних факторов указана в табл. 1, 2.



Рис. 1. Структура самоорганизационной информационной системы Big data

Таблица 1

Внутренние факторы

| | Форма информации Источник информации | Буква (текстовая) | Число (цифровая) | Физические величины (свойства объектов) |
|---|---|----------------------|---------------------|--|
| 1 | Текущие данные | 1 | 2 | 3 |
| 2 | Исторические данные | 4 | 5 | 6 |
| 3 | Данные из мониторов и датчиков | 7 | 8 | 9 |
| 4 | Данные в реальном времени | 10 | 11 | 12 |
| 5 | Неструктурированные данные | 13 | 14 | 15 |

Таблица 2

Внешние факторы

| | Вид информации Потребители информации | Априорная (Прогнозирование) | Информационный процесс (Управление) | Постаприорная (новые технологии жизнедеятельности общества) |
|----|--|--------------------------------|--|--|
| 1 | Ритейл | 1 | 2 | 3 |
| 2 | Финансы | 4 | 5 | 6 |
| 3 | Логистика | 7 | 8 | 9 |
| 4 | Телеком | 10 | 11 | 12 |
| 5 | Энергетика | 13 | 14 | 15 |
| 6 | ЖКХ | 16 | 17 | 18 |
| 7 | Нефтегаз | 19 | 20 | 21 |
| 8 | Аэрокосмическая отрасль | 22 | 23 | 24 |
| 9 | Медицина | 25 | 26 | 27 |
| 10 | Машиностроение | 28 | 29 | 30 |

Примечание: в таблице № 1,2 – порядковые значения 1,2,3...n – номер вида технологии.

7. Результаты исследований.

Проведенные в статье исследования показали, что технологии «Big Data» развивались эволюционно по принципу удовлетворения возникших социальных потребностей. Такой принцип развития технологий «Big Data» привел к возникновению целого ряда недостатков и прежде всего к разобщению методологий разработки технологий по формированию, обработке и использованию накопленной на носителях информации. Отсутствие единого системного подхода к определению меры информации объектов реального

Мира в «Big Data» сдерживает развитие технологий, по этому, как следствие потери этой информации для потребителей. В статье представлены основные принципы системно-информационного подхода к решению накопившихся проблем в «Big Data».

6. Выводы

Создание «Big Data» на базе единой меры информации является ключевым аспектом, который позволяет формировать, обрабатывать и использовать информацию «Big Data» на основе сформулиро-

ванных научных информационных закономерностей. Определение основных научных положений системно-информационного подхода к технологиям формирования, обработки и использования «Big Data» с позиции дискретно-вероятностной меры информации призвано выполнить эту задачу. Практически это обеспечивается особенностями разработки алгоритмов анализа и синтеза в задачах по разработке технологий «Big Data» на базе теоретических положений системно-информационного подхода. Сформулированные основополагающие принципы формирования системы «Big Data» с позиции дискретно-вероятностной меры информации позволяют со-

здавать «Big Data» со свойствами самоорганизации заложенных в нее функций. Разработанная схема самоорганизационной информационной системы «Big Data» отражает структурные основы формирования, обработки и использования информации «Big Data».

Предложенный подход к разработке технологий формирования, обработки и использования «Big Data» на базе дискретно-вероятностной меры информации существенно уменьшает затраты ресурсов при решении задач анализа и синтеза, за счет использования информационных закономерностей коррелированных свойств данных.

Литература

1. Луцкий, С. В. Теоретические основы системно-информационного подхода к технологическим процессам и системам [Текст]: монография / С. В. Луцкий. – Х.: ХНАДУ, 2008. – 328 с.
2. Луцкий, С. В. Стратегия формирования концепции информационного подхода и ее развития [Текст]: сб. тр. междунар. науч.-техн. конф. / С. В. Луцкий // Машиностроение и техносфера XXI века. – 2007. – Т. 2. – С. 267–273.
3. Луцкий, С. В. Системная методология информационного подхода [Текст]: зб. наук. пр. / С. В. Луцкий // Вісник Національного технічного університету “Харківський політехнічний інститут”. Тематичний випуск: Технології в машинобудуванні. – 2004. – № 28. – С. 60–64.
4. Руженцев, І. В. Міра дискретно-імовірної інформації при дослідженні процесів та систем [Текст]: наук.-вироб. журн. / І. В. Руженцев, С. В. Луцький, В. П. Фетьків // Метрологія та прилади. – 2016. – № 3. – С. 43–45.
5. Луцкий, С. В. Числовые характеристики информации технологических систем [Текст]: сб. тр. междунар. науч.-техн. конф. / С. В. Луцкий // Машиностроение и техносфера XXI века. – 2002. – Т. 2. – С. 61–66.
6. Руженцев, І. В. Дискретно-вероятностные информационные закономерности фактор повышения эффективности производства [Текст]: сб. тр. XX междунар. науч.-техн. конф. / І. В. Руженцев, С. В. Луцький, В. П. Фетьків, О. І. Подзигун // Метрологія и измерительная техника. – Х., 2016. – С. 21.
7. Луцкий, С. В. Моделивання, аналіз і синтез механообробних комп'ютерно-інтегрованих технологічних систем з елементами самоорганізації [Текст]: мат. наук.-пр. конф. / С. В. Луцький, Н. Е. Тернюк // Сучасні технології промислового комплексу. – Херсон: ХНТУ, 2015. – С. 104–108
8. Луцкий, С. В. Последовательность структурно-информационного синтеза систем [Текст]: сб. ст. мат. 10-ой междунар. науч.-метод. конф. / С. В. Луцький // Технологии XXI века. – 2003. – Т. 2. – С. 105–109.
9. Луцкий, С. В. Структуризация информационных процессов в технологических системах [Текст]: наук.-техн. журн. / С. В. Луцький // Механіка та машинобудування. – 2007. – № 2. – С. 117–125.
10. Луцкий, С. В. Отличительные особенности сущности информационного подхода к описанию систем [Текст]: сб. тр. междунар. науч.-техн. конф. / С. В. Луцький // Машиностроение и техносфера XXI века. – 2002. – Т. 3. – С. 61–66.

Дата надходження рукопису 16.01.2017

Руженцев Игорь Викторович, доктор технических наук, профессор, заведующий кафедрой, кафедра «Метрологии и технической экспертизы», Харьковский национальный университет радиоэлектроники, пр. Науки, 14, г. Харьков, Украина, 61166
E-mail: igor.ruzhentsev@nure.ua

Луцкий Сергей Владимирович, кандидат технических наук, старший преподаватель, кафедра «Метрологии и технической экспертизы», Харьковский национальный университет радиоэлектроники пр. Науки, 14, г. Харьков, Украина, 61166
E-mail: lutsk.sv6@gmail.com