

УДК 004.6:004.8

## ИНТЕЛЛЕКТУАЛЬНЫЕ ВЫЧИСЛЕНИЯ, АНАЛИЗ, ВИЗУАЛИЗАЦИЯ БОЛЬШИХ ДАННЫХ



**В.Б. Таранчук**

*профессор кафедры компьютерных  
технологий и систем,  
доктор физико-математических  
наук профессор*

*Белорусский государственный университет, Республика Беларусь  
E-mail: taranchuk@bsu.by*

**Аннотация.** Обсуждаются возможности сбора и обработки больших данных средствами системы Wolfram Mathematica. Приведены примеры обработки, визуализации и анализа.

**Ключевые слова:** большие данные, анализ, интеллектуальные вычисления, интерактивная визуализация, Wolfram Mathematica.

**Введение.** В современном обществе вместе со стремительным накоплением информации быстрыми темпами развиваются и технологии анализа данных. Из новых технологий наибольшее внимание уделяется развитию направлений Разреженные данные (Sparse data), Большие данные (Big Data).

Термин Big Data используется в сферах, где актуальна работа со сверхбольшими объемами структурированных и неструктурированных данных, где имеет место постоянное увеличение скорости потока данных. Источниками больших данных являются: интернет-документы, социальные сети, блоги, измерительные устройства, системы аудио- и видеорегистрации. Big Data аккумулируются на разных этапах деятельности человека. Громадные объемы данных накоплены в системах автоматизации производства, управлении, банковской деятельности, маркетинге, телекоммуникациях, логистике и транспорте, веб-аналитике, науке, медицине, образовании. Обычно, говорят о Big Data, когда обсуждают хранилища данных объемом намного больше, чем несколько сотен гигабайт, причем, в условиях непрерывного прироста, распределения по многочисленным узлам компьютерной сети, когда надо анализировать данные, которые невозможно обработать на одном компьютере. Существуют отрасли, где данные собираются и накапливаются очень интенсивно и названные объемы существенно выше. Например, в электроэнергетике по каждой электростанции непрерывный поток данных генерируется для десятков тысяч параметров каждую минуту, а по некоторым параметрам даже каждую секунду.

На текущий момент Большие данные являются одним из самых интенсивно развиваемых направлений информационных технологий. Обычно выделяют следующие типы задач, связанных с Big Data:

–технологии и инструменты сбора, хранения информации (MapReduce, Hadoop,

NoSQL базы данных) в распределенных файловых системах; очень большой объем имеющихся данных и увеличивающаяся скорость их накопления требуют развития нетрадиционных способов обработки данных в реальном времени;

–методы и алгоритмы, способы организации неструктурированной информации, сегментации и кластеризации; одновременной (параллельной) обработки разноформатной структурированной и неструктурированной информации;

–новая парадигма анализа данных, лингвистическая обработка текстов, проверка достоверности и определение ценности данных, углубленная аналитика, методы построения и способы внедрения прогностических моделей.

Основная парадигма распределенной обработки данных – MapReduce [1], [2], модель распределенной обработки больших объёмов данных на компьютерных кластерах. Практическая реализация, разработки MapReduce-программ эффективно выполняются на платформе Hadoop – набор утилит, библиотек и фреймворк; программная платформа, одно из ведущих решений для массово-параллельной обработки больших объемов разнородных данных [3]. Основными составляющими Hadoop являются ее фреймворк обработки данных и распределенная файловая система HDFS, которая только хранит данные, но не обрабатывает их. HDFS принципиально отличается от традиционных БД распределенностью операций. Обработка информации может выполняться разными способами. Один из наиболее популярных, доступный по умолчанию – это написанный на Java движок MapReduce. Он позволяет работать с HDFS как с базой данных, что не просто, так как в традиционных реляционных базах данных необходимая информация находится с использованием запросов на языке SQL. Нереляционные БД могут использовать и другие виды запросов, поэтому их еще называют NoSQL-базами данных.

*Цель данной работы* – привлечь внимание и обозначить возможности интеллектуального анализа, интерактивной визуализации, обсудить, как система компьютерной алгебры Mathematica компании Wolfram [4] может помочь в эффективном использовании Big Data для оптимизации различных процессов, решения задач управления.

Технические вопросы инсталляции в Wolfram Mathematica пакета Hadoop [5] описаны в [6]. Коды и подробные пояснения, как, используя HadoopLink, писать в Mathematica MapReduce-программы и запускать их в кластере Hadoop, можно посмотреть в [7] - [9]. Пример "Hello world!" [7] является для Hadoop аналогом WordCount – подсчитывается количество раз, когда каждое слово появляется в фрагменте текста. В [8] (перевод на русский в [9]) описан пример создания геномной поисковой системы.

Отметим несколько групп инструментов системы Mathematica, которые целесообразно применять при обработке, интерпретации и анализе больших данных.

### **Примеры работы с неструктурированной разноформатной информацией**

Большинство процедур обработки Big Data включают фильтрацию (например, найти все записи с указанного IP-адреса в логах web-сервера), выполнение преобразований данных (например, удалить колонку в csv-логах), сортировки, манипуляции с выборками. Вопросы, которые будут обсуждаться в данном модуле изложения, можно было бы дать иллюстрациями видео рядов из скриншотов, работая в системе Mathematica. Но такой вариант потребовал бы очень большой объем. Поэтому принят другой подход – сформирован представительный, но компактный результат сбора информации и на его примере показаны основные инструменты системы Mathematica.

Приведенный ниже список объектов (создан автоматически, поэтому подобный по представительности наполнения можно сделать сколь угодно большого объема) имитирует возможный комплект собранных данных, сформирован он функциями Mathematica, а именно, использовались: *RandomInteger*, *RandomReal*, *RandomChoice*, *RandomVariate*,

*RandomSample, Table.* Был зафиксирован один из конкретных вариантов, который представлен ниже выводом в формате списка lstA:

```
lstA = {
  {1, A1, B1, C1, D1, E1, F1, G1, H1},
  {2, 22, -NS, acxzs47, 58.5, -3, 105, {88, {48, 81}}, 98},
  {3, 27, -2.011, Текст43, 50.2, -5, аб.с, 83},
  {4, NS, -4.981, b49, аб.с, -3, 112, {83, 81, 48}, 96},
  {5, 24, -5.068, {48, 81, 94}, 53.4, -2, NS, {85, e49}, 92},
  {6, 24, NS, a42, 55.3, -2, 116, {82, Текст81}, 97},
  {7, 24, -1.066, bc47, 52.7, -3, NS, 89, 90},
  {8, NS, -41/6, e49, 54.8, -3, 102, {f81, 81, 97}},
  {9, 24, -1.529, f49, 57.3, -1, 199, 88},
  {10, 22, NS, {46, 94}, 56.1, -3, 104, {87, 97}, 96},
  {11, 23, -45/6, ac44, 55.5, -2, NS, {80, 9, 7, 5}},
  {12, 27, -5.051, {{a43, 81}, 43}, 54.6, -2, 117, "-", 94},
  {13, 29, -3.313, 49, 53.3, -4, 189, "-"}
}
```

и формы:

1	A1	B1	C1	D1	E1	F1	G1	H1
2	22	-NS	acxzs47	58.5	-3	105	{88, {48, 81}}	98
3	27	-2.011	Текст43	50.2	-5	аб.с	83	
4	NS	-4.981	b49	аб.с	-3	112	{83, 81, 48}	96
5	24	-5.068	{48, 81, 94}	53.4	-2	NS	{85, e49}	92
6	24	NS	a42	55.3	-2	116	{82, Текст81}	97
7	24	-1.066	bc47	52.7	-3	NS	89	90
8	NS	$-\frac{41}{6}$	e49	54.8	-3	102	{f81, 81, 97}	
9	24	-1.529	f49	57.3	-1	199	88	
10	22	NS	{46, 94}	56.1	-3	104	{87, 97}	96
11	23	$-\frac{15}{2}$	ac44	55.5	-2	NS	{80, 9, 7, 5}	
12	27	-5.051	{{a43, 81}, 43}	54.6	-2	117	-	94
13	29	-3.313	49	53.3	-4	189	-	

Поясним, что важно для понимания дальнейших действий по обработке данных из lstA. В нотации системы Mathematica всё, что ограничено фигурными скобками, это список. Допустима любая глубина вложенности элементов. Элементами могут быть любые объекты, в частности: цифры, текст, таблицы, графика, математические выражения, файлы мультимедиа, программные модули, гиперссылки. Число элементов не ограничено. По сути, список в Wolfram Mathematica – набор неструктурированной разноформатной информации. Выше показана форма (это не таблица), условно далее будем использовать термины строка и столбец, как общепринятые. Но в списке (фрагмент неструктурированной базы данных) таких понятий нет, а есть номер элемента. В приведенном варианте в списке lstA 13 вложенных списков. В некоторых из них есть другие вложенные списки – например, в форме показан вложенный список в строке 5 столбце C1. Есть более глубокие вложения – например, в строке 2 в столбце G1. При формировании lstA были заданы и поставлены в определенные места в списках последовательные номера 1 – 13, символы A1 – H1. Всё

остальное – результат работы упомянутых выше генераторов, и они дополнялись опциями. На каждой конкретной позиции (в приведенной форме это в столбике) формировались: целые, действительные или рациональные числа, идентификаторы, слова, списки. Несколько типичных ситуаций работы с Big Data воспроизведены записями «NS» (нет значения) и «-» – позиция не заполнена. Например, вторыми элементами списков 1 – 13 указывались случайные целые числа, сформированные *RandomInteger* с ограничением – из интервала 20 – 29. Например, четвертыми элементами (в таблице в колонке C1) присутствуют идентификаторы, списки и обязательно целые из интервала 41 – 49. Следует обратить внимание, что в оригинальной записи *lstA* есть действительные и рациональные числа в форматах строковой записи, идентификаторы могут быть гиперссылками. Приведенная выше форма наглядно демонстрирует данные, она получена в *Mathematica* с использованием *TableForm*, вывод дается в математической нотации (в частности, для рациональных чисел). Сформированный комплект можно трактовать, как представительный пример неструктурированной разноформатной информации. Приведем функции и опции системы, которые обеспечивают простейший анализ содержимого списка *lstA*. Заметим, что мы выполняем анализ с размещенным на конкретном компьютере файлом, но это только с целью, чтобы «все было перед глазами» и легко перепроверить.

Простейший вопрос – сколько чего в списке на разных уровнях? Запрос об архитектуре и числе элементов на разных уровнях обрабатываемой выборки. Для получения ответа достаточно выполнить функцию *Count* (встречаемость). Если в опциях указать уровни 1, 2, 3, 4, {2}, {3}, {4}, {Infinity}, то ответ будет {13, 125, 156, 160, 112, 31, 4, 160}. Как это надо понимать? «1, 2, 3, 4» означают, что надо посчитать, сколько элементов на 1; 1 и 2; 1 – 3; 1 – 4 уровнях. Записи «{2}, {3}, {4}» – запросы о числе элементов на 2, 3, 4 уровнях (только на них). Запись «{Infinity}» – запрос, сколько всего, на всех имеющихся уровнях (полезно использовать для контроля).

Запросы с уточнением типа объектов делаются аналогично. Если перечислить 1, {2}, {3}, {4}, {Infinity} и запросить с опцией *Integer*, ответ будет {0, 54, 21, 3, 78}; для типа *Real*, ответ будет {0, 18, 0, 0, 18}; для типа *Rational*, ответ будет {0, 2, 0, 0, 2}; для типа *Even*, ответ будет {0, 28, 8, 1, 37}.

Можно выполнять запросы, формулируя их с логическими и арифметическими выражениями, с шаблонами. Приведем примеры и конструкции кода, обеспечивающие требуемый запрос. Ниже результат выполнения запроса – выбор элементов (в данном примере списков), в которых только два объекта:

```
{{88, {48, 81}}, {85, e49}, {82, Текст81}, {46, 94}, {87, 97}, {{a43, 81}, 43}}.
```

Для реализации используется функция *Cases* (случаи по образцу), и нужно определять уровень, где искать. Заметим, что элементы {88, {48, 81}} и {a43, 81}, 43} содержат число и список, то есть по 2 объекта.

Следующий результат – выбор элементов, в которых на последнем месте простое число:

```
{{3, 27, -2.011, Текст43, 50.2, -5, аб.с, 83},  
{6, 24, NS, a42, 55.3, -2, 116, {82, Текст81}, 97}}
```

Обратите внимание, что строка 8 не включена в ответ – там на последнем месте не 97, а список. Для реализации используется функция *Select* (выбрать) и опция *Last*. Можно отбирать записи, в которых требуемое на первом месте, тогда надо использовать *First*.

Результат выполнения запроса ниже – выбор и вывод элементов, удовлетворяющих двум условиям: а) список из 2-х объектов, б) в каждом из вбранных на первом месте записано число и оно больше 82:

```
{{88, {48, 81}}, {85, e49}, {87, 97}}.
```

Для реализации используется функция *Cases*, и нужно определять условие  $x_1 > 82$ . Заметим, что запись {83,81,48} не включена в ответ – этот список содержит 3 объекта, а в примере отбираем только те, где их 2.

Следующий результат – пример извлечения списков, в которых присутствует указанный текст «аб.с»:

```
{{3, 27, -2.011, Текст43, 50.2, -5, аб.с, 83},  
{4, NS, -4.981, b49, аб.с, -3, 112, {83, 81, 48}, 96}}.
```

Обратите внимание, что в запросе не имеет значения, на каком месте текст находится. Для реализации используется функция *Select* (выбрать) и опция *MemberQ[#,аб.с]*.

Результат выполнения запроса ниже – пример извлечения списков, в которых ни один из элементов не есть NS:

```
{{9, 24, -1.529, f49, 57.3, -1, 199, 88},  
{12, 27, -5.051, {{a43, 81}, 43}, 54.6, -2, 117, "-", 94},  
{13, 29, -3.313, 49, 53.3, -4, 189, "-"}}.
```

Для реализации используется функция *Cases*, и опция *(#[[\_]]!="NS")*.

Следующий результат – пример извлечения списков (строк), в которых в столбце F1 числа больше или равно 118:

```
{{9, 24, -1.529, f49, 57.3, -1, 199, 88}, {13, 29, -3.313, 49, 53.3, -4, 189, "-"}}.
```

Для реализации используется функция *Cases*, с условием  $x_7 > 118$ , задается номер элемента списка (7 – в столбце F1).

Представляется, что приведенные примеры реализаций запросов – это наиболее часто применяемые в практике. Получаемые выборки можно обрабатывать и выполнять «Разделение списка на части», «Изменение порядка элементов в списке», «Комбинирование списков и работа с множествами» – подробно описано в [10], [11].

### Примеры извлечения и визуализации данных из удаленных серверов

Графика системы Mathematica всегда считалась для многих пакетов и систем эталонной и во многом способствовала ее высокой репутации мирового лидера среди подобных программных приложений. Графические возможности достигаются, как обилием встроенных функций, так и средствами их модификации с помощью директив, опций, примитивов, языка программирования. Mathematica позволяет строить любые виды графиков, причем, обычного пользователя в большинстве случаев удовлетворяют графики, параметры которых устанавливаются по умолчанию. В системе поддерживается работа практически со всеми форматами графики, их импорт и экспорт.

Концептуально графики в Mathematica являются объектами, которые создаются (возвращаются) соответствующими графическими функциями. Они охватывают построение всех типов математических графиков, иллюстраций деловой графики – гистограмм, двумерных и трехмерных секторных и столбчатых, пузырьковых диаграмм, специфических

графиков для таких областей, как финансы и статистика, теория графов, управляющие системы [12]. Достигается многообразие за счет применения функций, опций и директив. Графическим объектам можно присваивать имена, а затем оперировать как любыми объектами. Приведем несколько примеров визуализации Big Data.

Ниже приведены несколько диаграмм:

- TradingChart – "живой" график изменения цен с возможностью указания дополнительных индикаторов анализа; трейдинговая интерактивная диаграмма, показывающая в выбираемом временном интервале цены (минимальную, максимальную, при открытии и закрытии) и объемы продаж на каждую дату диапазона;

- CandlestickChart – диаграмма японские свечи, которая чаще всего используется для отображения результатов биржевых торгов, а именно, значений максимальной цены, минимальной цены, цены открытия и цены закрытия в день торгов.

Диаграммы иллюстрируют актуальные данные за январь 2019 г., отобранные с удаленных серверов с использованием упомянутых выше инструментов. Примеры даются по итогам торгов на Нью-Йоркской Фондовой Бирже, несколько вариантов визуализации данных и сопоставления приведены по показателям компаний EPAM Systems и IBM.

На рисунке 1 представлен "живой" график TradingChart, он включает 2 диаграммы: верхняя – диаграмма японские свечи, нижняя иллюстрирует объемы продаж, также выводятся числовые значения на выбранную дату. Пользователь такой интерактивной диаграммы указателем может выбрать любую из дат диапазона, активизированная дата подсвечивается полосой, обозначается точками на столбиках, а также в верхних частях диаграмм слева выводятся конкретные числовые показатели. Японские свечи можно оформлять разными способами, отдельно назначая атрибуты случаев роста и падения цены (устанавливается по умолчанию, но пользователь может назначить свои атрибуты). В примере рисунка 1 зеленым цветом показаны столбики, когда был рост цены акции, красным – снижение.

В примере рисунка 1 фактически нет никакой математической обработки, только визуализация. Важно, что Wolfram Mathematica предоставляет пользователю не только большое число вариантов оформления, но и возможности подключения, выполнения расчетов и вывода дополнительных графиков встроенных финансовых индикаторов-показателей. Только в ядре системы их больше 50, а также есть специфические в пакетах-приложениях (поставляются с дистрибутивом).

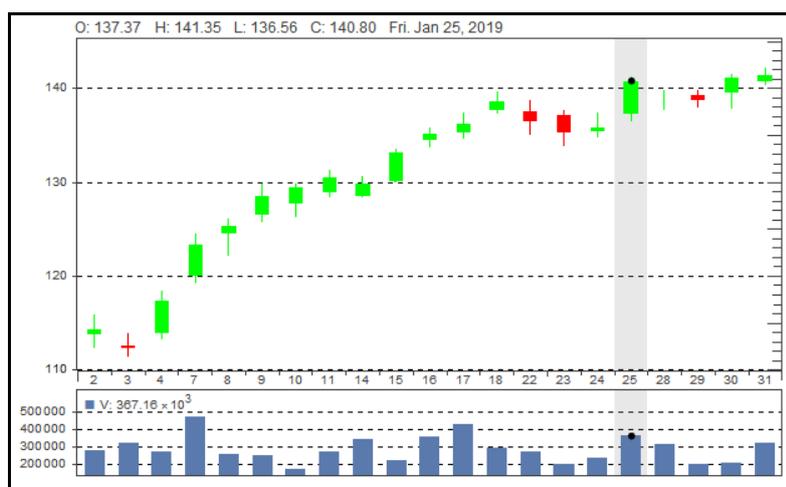


Рисунок 1. Визуализация результатов торгов акций компании EPAM Systems на Нью-Йоркской Фондовой бирже в январе 2019 г.

Например, можно отметить: AbsolutePriceOscillator, AccumulativeSwingIndex, AverageDirectionalMovementIndex, BollingerBands, ChaikinMoneyFlow, CommoditySelectionIndex, DynamicMomentumIndex, ExponentialMovingAverage, FastStochastic, LinearRegressionTrendlines, StochasticMomentumIndex, TimeSeriesForecast, WeightedMovingAverage, SimpleMovingAverage. Перечислены опции системы функций вывода и анализа, они идентифицируются согласно принятой в Wolfram Mathematica концепции названия функций и опций – классический математический термин, записываемый без пробелов и каждое следующее слово с прописной буквы. Другими словами, все приведенные обозначения, их смысл можно найти в математической энциклопедии или справочнике. Все перечисленные выше индикаторы содержат параметры настройки. Пример на рисунке 2 – вариант вывода кроме данных, как на рисунке 1, кривых, иллюстрирующих расчетный индикатор WeightedMovingAverage для случаев числа точек осреднения 5 и 15.



Рисунок 2. Визуализация результатов торгов и кривые индикатора WeightedMovingAverage

Конечно же, система дает возможности пользователю сопоставлять результаты разных запросов, выборок, методов обработки. На рисунке 3 показаны результаты применения графической функции CandlestickChart (диаграмма японские свечи), дается возможность сопоставлять данные торгов по компаниям EPAM Systems и IBM. В подобных иллюстрациях важно использовать разные варианты оформления выводимых диаграмм.

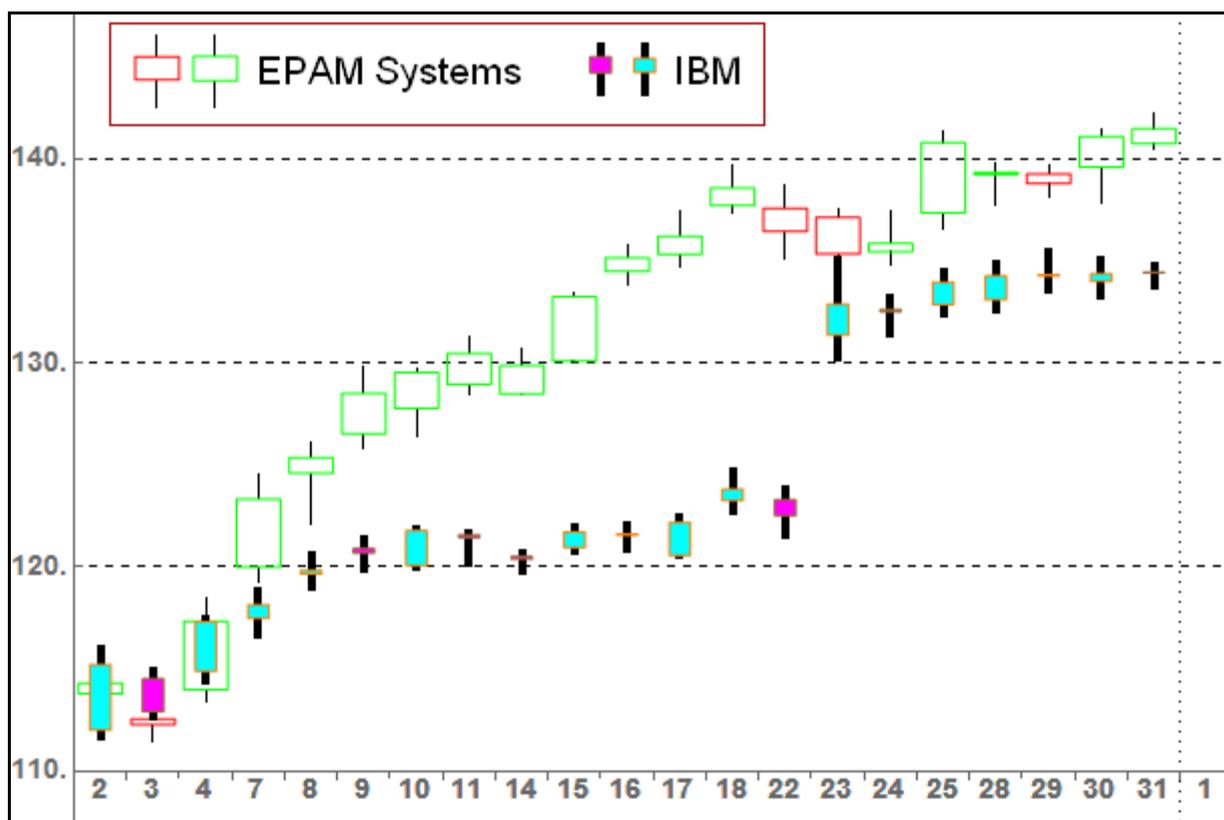


Рисунок 3. Визуализация результатов торгов по компаниям EPAM Systems и IBM

### Возможности анализа Big Data

Для оптимизации разных сфер деятельности, от государственного управления до производства и телекоммуникаций, нужны новые методы анализа накапливаемой информации. Компьютерный анализ данных ориентирован на то, чтобы обнаружить определённые и незаметные человеку закономерности. Быстро возрастающие объёмы неоднородной разноформатной цифровой информации обработать традиционными инструментами невозможно. Техники и методы анализа, применимые к Big Data по McKinsey [13]: Data Mining, краудсорсинг, смешение и интеграция данных, машинное обучение, искусственные нейронные сети, распознавание образов, прогнозная аналитика, имитационное моделирование, пространственный анализ, статистический анализ, визуализация аналитических данных. Перечисленное реализуется разными технологиями и на разных программных платформах. Основой соответствующих алгоритмов являются методы прикладной статистики, достижения в разработке компьютерных программ для анализа статистических данных. Распространяемые программы, пакеты, системы постоянно совершенствуются в направлениях ускорения обработки данных, увеличения числа встроенных статистических процедур и функций, улучшения графической визуализации результатов анализа, повышения удобства интерфейса, совершенствования справочной системы. В ракурсе использования систем компьютерной алгебры отметим, что в обсуждаемом секторе ИТ эксперты констатируют кардинальное изменение ситуации на рынке предложений, которое произошло в 2010 г. Тогда была выпущена система компьютерной алгебры Wolfram Mathematica 8 с существенно расширенными функциональными возможностями реализации вероятностно-статистических методов. По ряду позиций она признана лидером. Например, возможности

Mathematica в сравнении с другими разработчиками иллюстрирует диаграмма на рисунке 4.



Рисунок 4. Wolfram Mathematica – сопоставление возможностей основных лидеров

Спектр задач прикладной статистики, которые решаются в версиях 8 - 10 Mathematica, приведен в [14]. Отметим, что появилось новое в ракурсе конкретных дополнений применительно к анализу Big Data в версии 11.

В первую очередь следует упомянуть аппарат и большое число функций в области машинного обучения, в частности нейронные сети, кластерный анализ. Версия 11 Mathematica дает пользователю возможности создания и тренировки нейронных сетей, включающих в себя полный набор ориентированных на зрительное восприятие слоёв, задание топологий сетей множественными входными и выходными данными и связывающей структурой произвольно ориентированного ациклического графа. В подсистеме помощи даны подробные описания и представительные примеры. В частности, можно изучать особенности настройки и использовать решения: цифровой классификации (обучения свёрточной сети по определению цифр в изображении), обучения классификации точек из различных кластеров, измерения результативности классификации (точность цифрового распознавателя, обученного на базе данных рукописных цифр MNIST), обучения на находящихся во внешней памяти наборах данных изображений. Функции машинного обучения Mathematica 11 дают возможности извлекать признаки моделируемых объектов, оптимизировать их размер, группировать данные, организовывать поиск иерархической группировки для данных, классифицировать гиперпараметры, строить интерпретируемые модели, обучать модели на данных смешанных видов. Функциональные возможности извлечения признаков можно использовать для создания семантических расстояний для поисковых систем, создания встроенного распознавания объектов на изображениях и классификаторов для неклассифицированных данных. Существенно расширен аппарат машинного зрения, в частности, ImageIdentify может определить более 10000 объектов; повышена эффективность работы Classify с изображениями. В системе предоставляются инструменты автоматического поиска формулы, подходящей для кривой, использования метода прогнозирования гауссова процесса.

### *Литература*

- [1]. MapReduce: Simplified Data Processing on Large Clusters [Электронный ресурс] / Режим доступа: <https://ai.google/research/pubs/pub62>. Дата доступа: 30.01.2019.
- [2]. Big Data – Buzz Words: What is MapReduce – Day 7 of 21 [Электронный ресурс] / Режим доступа: <https://blog.sqlauthori-ty.com/2013/10/09/big-data-buzz-words-what-is-mapreduce-day-7-of-21>. Дата доступа: 30.01.2019.
- [3]. Hadoop – слон, на котором стоит Big Data [Электронный ресурс] / Режим доступа: [http://cloudzone.ru/community/blogs/big\\_data/221.html](http://cloudzone.ru/community/blogs/big_data/221.html). Дата доступа: 30.01.2019.
- [4]. WOLFRAM MATHEMATICA. Наиболее полная система для современных технических вычислений в мире [Электронный ресурс] / Режим доступа: <http://www.wolfram.com/mathematica>. Дата доступа: 30.01.2019.
- [5]. Join GitHub today [Электронный ресурс] / Режим доступа: <https://github.com/shadanan/HadoopLink>. Дата доступа: 30.01.2019.
- [6]. Mathematica gets Big Data with HadoopLink [Электронный ресурс] / Режим доступа: <https://blog.wolfram.com/data/uploads/2013/07/HadoopLink.cdf>. Дата доступа: 30.01.2019.
- [7]. Paul-Jean Letourneau. Mathematica Gets Big Data with HadoopLink [Электронный ресурс] / Режим доступа: <https://blog.wolfram.com/2013/07/31/mathematica-gets-bigdata-with-hadooplink>. Дата доступа: 30.01.2019.
- [8]. Paul-Jean Letourneau. Searching Genomes with Mathematica and HadoopLink [Электронный ресурс] / Режим доступа: <https://blog.wolfram.com/2014/02/21/searching-genomes-with-mathematica-and-hadooplink>. Дата доступа: 30.01.2019.
- [9]. Гузенко, К. Поиск по геному с помощью Wolfram Language (Mathematica) и HadoopLink. = Перевод поста Пола-Жана Летуэрно (Paul-Jean Letourneau) "Searching Genomes with Mathematica and HadoopLink". [Электронный ресурс] / Режим доступа: <https://habr.com/ru/company/wolfram/blog/263307>. Дата доступа: 30.01.2019.
- [10]. Таранчук, В.Б. Введение в язык Wolfram : учеб. материалы для студентов фак. прикладной математики и информатики спец. 1-31 03 04 «Информатика» / В.Б. Таранчук. – Минск : БГУ, 2015. – 51 с.
- [11]. Таранчук, В.Б. Основы программирования на языке Wolfram : учеб. материалы для студентов фак. прикладной математики и информатики спец. 1-31 03 04 «Информатика» / В.Б. Таранчук. – Минск : БГУ, 2015. – 49 с.
- [12]. Таранчук, В.Б. Введение в графику системы Mathematica : учеб. материалы для студентов фак. прикладной математики и информатики / В. Б. Таранчук. – Минск : БГУ, 2017. – 53 с.
- [13]. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. Big data: The next frontier for innovation, competition, and productivity. [Электронный ресурс] / Режим доступа: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>. Дата доступа: 30.01.2019.
- [14]. Таранчук, В.Б. Об инструментах системы Mathematica для реализации вероятностно-статистических методов / В.Б. Таранчук // Информационные технологии и системы 2015 (ИТС 2015) : материалы международной научной конференции (БГУИР, Минск, Беларусь, 29 октября 2014) / Минск : БГУИР, 2015 – С. 192-195.

## **INTELLIGENT COMPUTING, ANALYSIS, VISUALIZATION OF BIG DATA**

***V.B. Taranchuk***

*Professor of Department of Computer  
Applications and Systems,  
Doctor Physical-Mathematical  
Sciences, Professor*

*Belarusian State University, Republic of Belarus  
E-mail: taranchuk@bsu.by*

**Abstract.** Possibilities of collecting and processing Big Data by tools of the Wolfram Mathematica are discussed.

**Keywords:** analysis, intelligent computing, interactive visualization, Big Data, Wolfram Mathematica.