

BIG DATA VISUAL ANALYSIS USING COGNITIVE PATTERNS
Shishkin Yu.
ВИЗУАЛЬНЫЙ АНАЛИЗ БОЛЬШИХ ДАННЫХ С ПРИМЕНЕНИЕМ
ПОЗНАВАТЕЛЬНЫХ ПАТТЕРНОВ
Шишкин Ю. Е.

Шишкин Юрий Евгеньевич / Shishkin Yuriy – аспирант,
кафедра информационных технологий и компьютерных систем,
Федеральное государственное автономное образовательное учреждение высшего образования
Севастопольский государственный университет, г. Севастополь

Аннотация: в статье рассматриваются методы визуализации Больших Данных, использующие базовые познавательные структуры человека с целью увеличения объема воспринимаемой информации и повышения качества принимаемых, на ее основе, управленческих решений. Выделены ключевые системные компоненты средств анализа Больших Данных для решения прикладных управленческих, промышленных и крупномасштабных научных задач. Вводится понятие слабоструктурированных данных как разновидности Больших Данных, с учетом прикладного аспекта разработки системы поддержки принятия решений.

Abstract: the article analyzes the methods of big data visualization using the basic human cognitive structures in order to increase the amount of perceived information and improve the managerial decisions quality based on it. Highlighted big data key system components analysis tools for solving application, industrial and large-scale scientific problems. The concept of semi-structured data was introduced as a type of big data, taking into account the applied aspect of decision support system development.

Ключевые слова: Большие Данные, слабоструктурированные данные, система поддержки принятия решений, агентный подход, имитационное моделирование, визуализация.

Keywords: Big Data, semi-structured data, decision support system, agent-based approach, simulation, visualization.

На сегодняшний день для решения многих прикладных управленческих, промышленных и крупномасштабных научных задач используются автоматизированные экспертные системы принятия решений [1]. Методы визуализации, использующие познавательные паттерны человека, применяются для выполнения анализа данных мониторинга и моделирования, отображения иерархии классов и процессов программного обеспечения, осуществления процессов верификации и валидации, что особенно актуально для систем критического назначения [2].

Целью исследования является произвести анализ существующих методов визуализации больших объемов данных, выделить среди них подмножество методов, использующих врожденные познавательные паттерны человека. Дать определение данным с неявно выраженной структурой в рамках решения прикладной задачи разработки системы поддержки принятия решений.

Большие Данные (БД), вследствие особенностей структуры, эффективно используются при хранении, обработке и исследовании данных мониторинга в плохо формализуемых областях знаний [3, 4]. В широком смысле БД включают в себя структуры данных и методы для работы с ними, делящиеся на классы: структуризации, кластеризации и визуализации. При эффективной организации систем поддержки принятия решений с использованием систем визуализации, многомерной структуры данных и их взаимосвязей лицо, принимающее решение, получает возможность выявления неявных связей в управляемой системе, что, в конечном итоге, на качественном уровне повышает эффективность управленческих решений [5,6]. Результативность описанной системы управления может оцениваться комплексными количественными метриками и критериями, компоненты которых в общем виде могут быть представлены парой векторов: коэффициенты эффективности $\varphi = \langle q_{\text{agil}}, q_{\text{acc}}, q_{\text{deadline}}, q_{\text{fin}} \rangle$ и коэффициенты значимости $\mathcal{J} = \langle i_{\text{agil}}, i_{\text{acc}}, i_{\text{deadline}}, i_{\text{fin}} \rangle$, где индексы компонентов имеют следующий физический смысл: оперативность, прогностическая точность, соблюдение директивных сроков и финансовая эффективность [7]. Таким образом, критерий эффективности, используемый для признания предпочтительности некоторого подмножества из конечного множества стратегий, может быть представлен в виде:

$$\chi(\varphi_i \cdot \mathcal{J}_i) \geq \chi(\varphi_j \cdot \mathcal{J}_j), \forall i, j \in (0, |\chi|), \chi \neq \emptyset,$$

где $|\chi|$ – мощность функционала, характеризующего количественный интегральный показатель эффективности выбранной стратегии, i, j – индексы компонентов критерия эффективности.

Данные мониторинга физических, технических и экономических систем, собираемые с использованием различных технических средств, по своему объему, изменчивости и разнообразию позволяют отнести их к слабоструктурированным данным, таким образом становится возможным применение для их обработки математического аппарата Больших Данных [8]. Существующие определения слабоструктурированных данных не учитывают специфику их применения для разработки систем поддержки принятия решений и не в полной мере удовлетворяют требованиям, предъявляемым к системам анализа и визуализации больших данных [9].

Введем понятие слабоструктурированных данных как компонента системы поддержки принятия решений следующим образом: слабоструктурированными данными называется разновидность структурированных Больших Данных, энтропия которых не возрастает при сколь угодно большом увеличении объема данных, иерархическая структура наборов данных полностью описывается тегами и маркерами записей и полей в самих наборах данных, обладающих необходимым и достаточным многообразием, позволяющим производить полный и объективный интеллектуальный анализ предметной области.

Исследование типовой структуры системы визуализации Больших Данных как инструмента для поддержки принятия решений прикладных управленческих, промышленных и крупномасштабных научных задач позволило выделить следующие ключевые системные компоненты [10, 11].

- подсистема источников данных, системы управления реляционными и многомерными базами данных, системы хранения данных, протоколы обеспечения целостности, безопасности и непрерывности информационных потоков;

- подсистема обработки данных, приведения к общей структуре, фильтрации дубликатов, кластеризации, агрегации и трансформации;

- подсистема моделирования, прогнозирования, оценки и валидации данных;

- подсистема визуализации и анализа, интерпретация результатов, визуализация взаимодействия, отображения внутренних связей, визуализация последствий принимаемых решений.

Существующие методы отображения информации делятся на классы: построенные на статических диаграммах стандартного типа, секторные диаграммы, гистограммы, диаграммы рассеивания, геопространственные и трехмерные карты, средства Text Mining и Data Mining; динамические диаграммы сложного типа с интерактивным управлением, масштабированием обобщенных данных и возможностью осуществления визуальных запросов в данных изменяющихся в реальном времени [12].

Осуществление процесса интерпретации результатов обработки Больших Данных с привлечением экспертов требует дальнейшего развития средств визуализации и отображения информации, таких как трехмерные проекции, голограммы, инструменты дополненной и виртуальной реальности. Качественное повышение объема и степени восприятия данных пользователем, без отсечения существенных деталей, возможно при комплексном использовании математических методов, новых информационных технологий и учетом принципов когнитивной психологии. Ориентирование визуальных средств на, сложившиеся эволюционным путем, базовые познавательные структуры человеческого сознания снижает требуемый уровень минимально необходимых теоретических и специальных знаний пользователя и позволяет значительно увеличить объем одновременно воспринимаемых данных, что приводит к повышению качества принимаемых экспертных решений.

Литература

1. Скатков А. В., Шишкин Ю. Е., Николаева Ю. П. Имитационная модель взаимодействия триады агентов облачной вычислительной среды «Потребитель – Брокер – Провайдер» // Информационные технологии и управление: сб. науч. тр., 2015. Т. 1. № 1. С. 114-119.
2. Шишкин Ю. Е. Анализ моделей взаимодействия пользователей и провайдеров облачных сервисов // Интеллектуальные системы, управление и мехатроника – 2016: материалы всерос. науч.-техн. конфер. молодых ученых, аспирантов и студентов. Севастополь 19-21 мая 2016 г. Севастополь: СевГУ, 2016. С. 289-293.
3. Шишкин Ю. Е., Греков Н. А. Исследование систем управления высокоточными измерениями // Интеллектуальные системы, управление и мехатроника - 2015: материалы междунар. науч.-техн. конфер. молодых ученых, аспирантов и студентов. Севастополь: СевГУ, 2015. С. 221-225.
4. Пасынков М. А. Варианты получения и обработки информации прибора ИСТ-1М // Интеллектуальные системы, управление и мехатроника – 2016: материалы всерос. науч.-техн. конфер. молодых ученых, аспирантов и студентов. Севастополь: СевГУ, 2016. С. 511-512.
5. Скатков А. В., Брюховецкий А. А., Шишкин Ю. Е. Сравнительный анализ методов обнаружения изменений состояний сетевого трафика // Автоматизация и приборостроение: проблемы, решения: материалы междунар. науч.-техн. конфер. Севастополь: СевГУ, 2016. С. 14-15.

6. *Малярова М. В.* Аналитика и визуализация «больших данных»: почему «большие данные» являются большой проблемой? // International Scientific Review, 2016. № 3 (13). С. 66-68.
7. *Shishkin Y. E.* Big Data visualization in decision making // Science in Progress: тез. Всерос. науч.-практ. конфер. магистрантов и аспирантов / Новосибирск: НГТУ, 2016. С. 203-205. ISBN 978-5-7782-3094-1.
8. *Шлюйкова Д. П.* Большие данные: современные подходы к хранению и обработке // Наука, техника и образование, 2016. № 1 (19). С. 75-79.
9. *Шишкин Ю. Е., Скатков А. В.* Решение задачи составления расписаний большой размерности с применением технологии Больших Данных // Информационные технологии и информационная безопасность в науке, технике и образовании «ИНФОТЕХ – 2015»: материалы междунар. науч.-техн. конфер. Севастополь: СевГУ, 2015. С. 103-105.
10. *Греков А. Н., Шишкин Ю. Е.* Моделирование трехкомпонентного акустического измерителя скорости течения // Системы контроля окружающей среды. Севастополь: ИПТС, 2016. № 6 (26). С. 33-40.
11. *Досалиева Н. А.* Роль больших данных в управлении знаниями // European Science. 2016. № 7 (17). С. 25-27.
12. *Шишкин Ю. Е.* Облачные сервисы в системах поддержки принятия решений // Научный журнал, 2017. № 1 (14). С. 19-20.