DOI 10.37882/2223–2966.2022.06.35

О ВИЗУАЛИЗАЦИИ БОЛЬШИХ ДАННЫХ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Ушкова Надежда Николаевна
Ассистент, РТУ МИРЭА, г. Москва
ushkova.n.n@mail.ru
Чесалин Александр Николаевич
К.т.н., доцент, РТУ МИРЭА, г. Москва
chesalin_an@mail.ru
Болотин Кирилл Викторович
Ассистент, РТУ МИРЭА, г. Москва
bolotin@mirea.ru

Аннотация. В работе исследуются методы визуализации больших данных в информационных системах, на примере SIEM систем в информационной безопасности. Рассматриваются современные инструменты визуализации информации, их достоинства, недостатки возможные варианты использования. Приводятся рекомендации и методы построения эргономичных аналитических панелей. Предлагается обобщенный алгоритм построения эргономичных информационных панелей SIEM систем.

Ключевые слова: большие данные, визуализация информации, SIEM системы, человеко-машинный интерфейс.

VISUALIZATION OF BIG DATA IN INFORMATION SYSTEMS

N. Ushkova
A. Chesalin
K. Bolotin

Summary. The paper explores methods of big data visualization in information systems, on the example of SIEM systems in information security. Modern information visualization tools, their advantages, disadvantages and possible use cases are considered. Recommendations and methods of ergonomic analytical panels construction are given. A generalized algorithm of building ergonomic information panels of SIEM systems is proposed.

Keywords: big data, information visualization, SIEM systems, human-machine interface.

Состояние проблемы

Современные информационные системы информационной безопасности обрабатывают огромные массивы разнородной информации, которые принято называть большими данными. При этом в задаче управления событиями информационной безопасности анализ информации и принятие решений отводится человеку. Для принятия решения применяются разнообразные вербальные методы анализа информации, к примеру, известны семь новых инструментов управления качеством, направленные на анализ вербальной информации (помимо данных семи методов, конечно же, нельзя не отметить важность семи базовых простых методов, семи новейших и семи современных интеллектуальных методов [1, 2].

Основной системой информационной безопасности, в которой производится анализ информации является SIEM (Security information and event management) системы. Современные SIEM системы имеют в своем составе разнообразные инструменты визуализации и анализа данных. На рисунке 1 представлена аналитическая панель Open-source SIEM Wazuh, которая может быть гибко настроена для различных задач (мониторинга, анализа, прогнозирования и пр.). Разнообразие и обилие средств визуализации позволяет представить информацию «с разных ракурсов», при этом иногда этом может быть

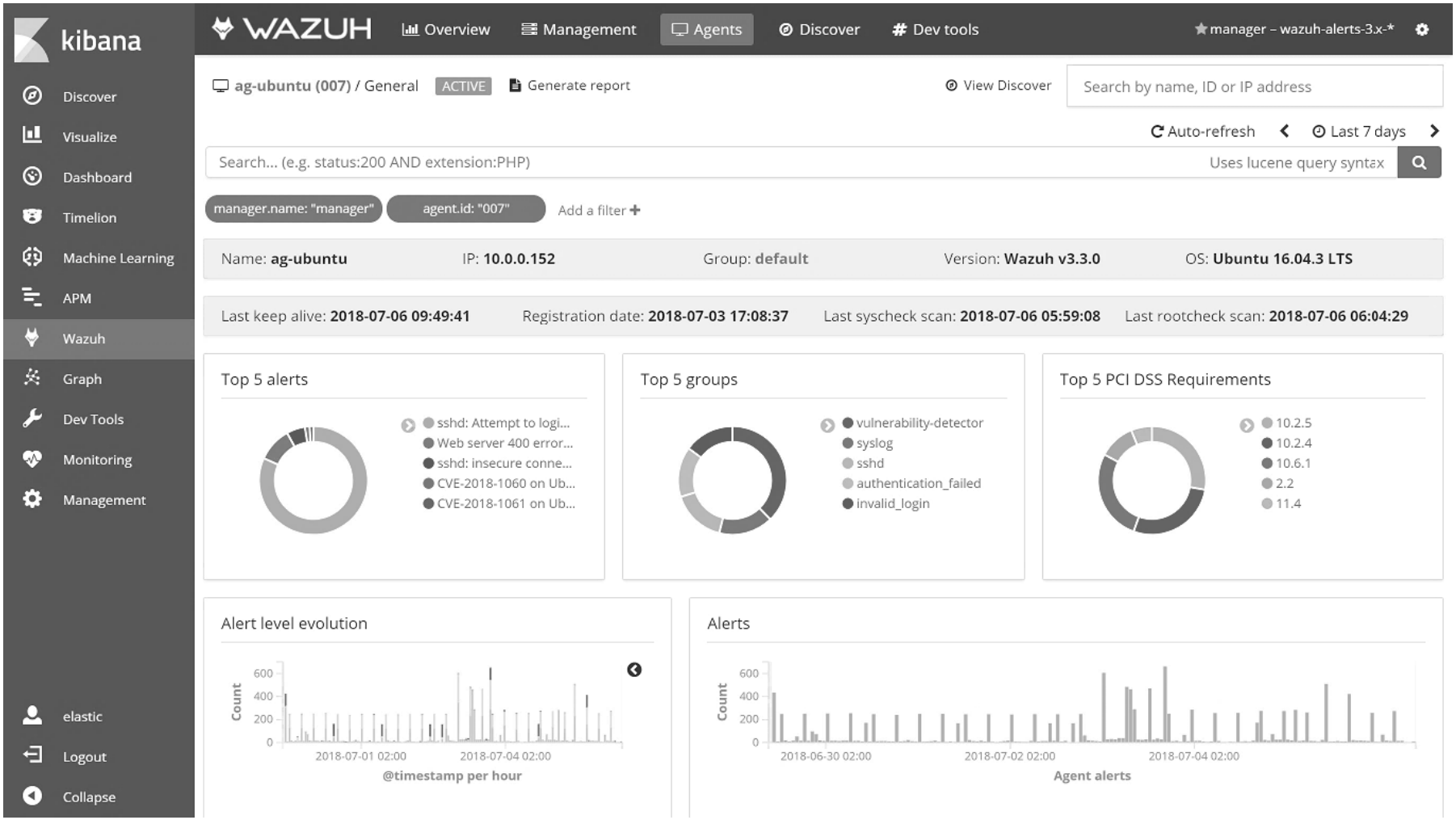

Рис. 1. Информационно-аналитическая панель в SIEM Wazuh

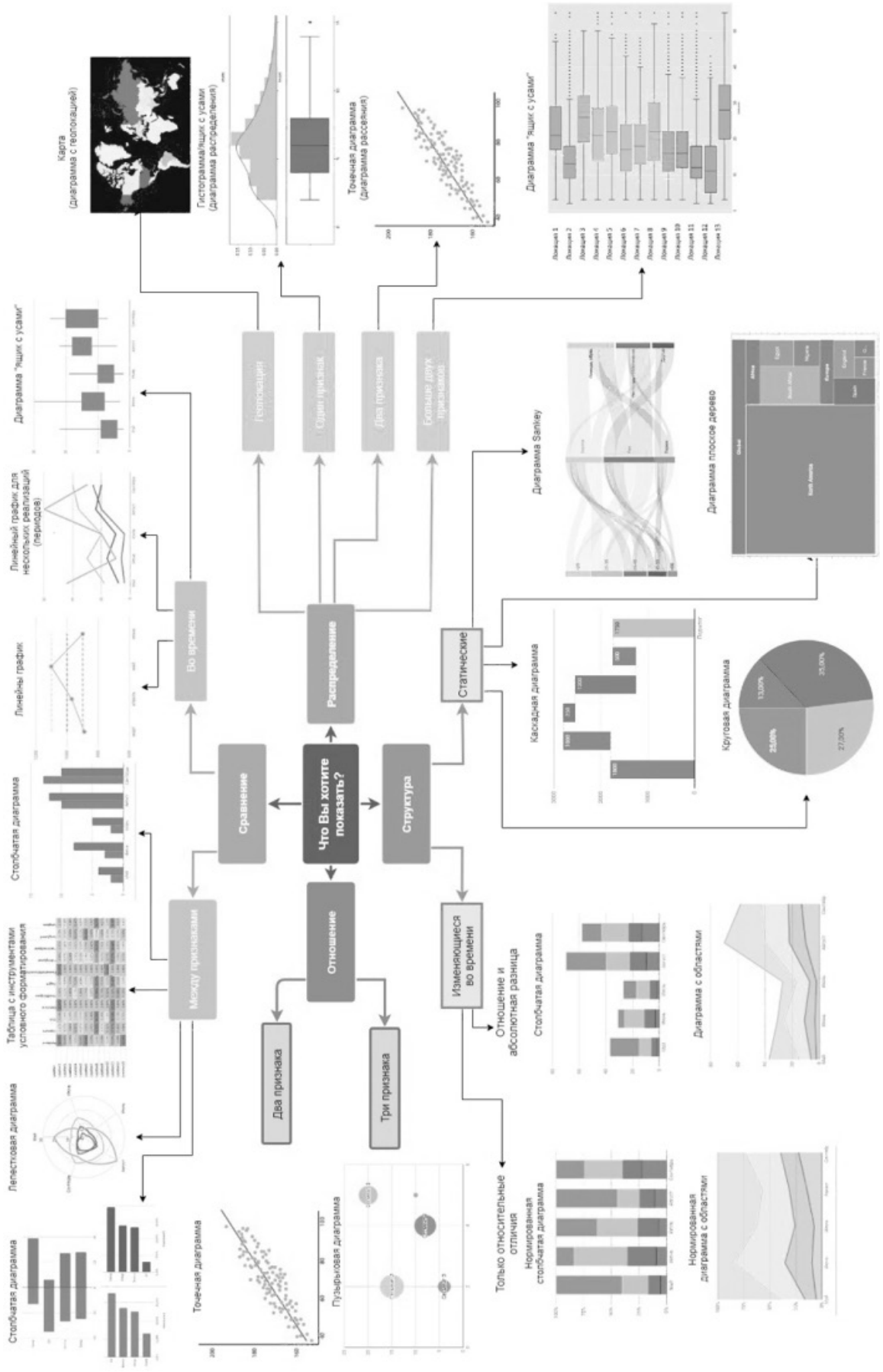

Рис. 2 Диаграмма выбора способа визуализации данных [https://i2.wp.com/s.dou.ua/storage-files/im-1_a9hNNcM.png]

Рис. 3. Диаграммы, используемые в аналитических панелях

и минусом — когда панель перегружена информацией, или элемент сложен для анализа, быстро меняется, расположен далеко от других связанных элементов и пр.

Поэтому важно проектировать аналитическую панель таким образом, что информация на ней была достаточной, при этом минимизирована, с учетом особенностей ее использования в конкретных задачах, конкретным лицом и заданных условий использования.

Постановка задачи

Рассматриваются современные инструменты визуализации больших данных, проводится их анализ и предлагаются варианты их эффективного использования и разработки эргономичных аналитических панелей (dashboards).

Инструменты визуализации больших данных

При проектировании аналитических панелей рекомендуется использование набор различных диаграмм, обладающих различной информативностью и сложностью восприятия, позволяющих системно анализировать информацию и своевременно принимать решения.

Выбор той или иной диаграммы можно осуществлять с помощью матрицы выбора Эндрю Абела [3], расширенный вариант которой представленной на рисунке 2.

Анализ диаграмм, используемых в аналитических панелях.

Рассмотрим некоторые диаграммы, как наиболее используемые, так и рекомендуемые авторами для включения в аналитические панели, с указанием их достоинств и недостатков, а также анализ их применимости (Рис. 3)

График изменения данных (Рис. 3. а)

Линейные графики используют для отображения количественных показателей. К преимуществам данного графика относится возможность демонстрации общей картины полученных данных. К недостаткам — сложность применение графиков к большим данным.

Столбчатая диаграмма (гистограмма и полигон) (Рис. 3. б)

Применимо для опытных специалистов. Удобно использовать для сравнения показателей за разный период, для проведения анализа данных. Недостатком является оформление диаграммы при большом количестве данных.

Пузырьковая диаграмма (Рис. 3. в).

Пузырьковая диаграмма пригодится для визуализации анализа взаимодействий и распределения. Удобно использовать для сравнения и отображения взаимосвязей, для анализа корреляций. К недостатку данной диаграммы относится ограничение по количеству данных, так как большое количество пузырьков может сделать график не читаемым

Диаграмма «бублик» (Рис. 3. г)

Наиболее простая и информативная диаграмма, более оптимально использовать динамическую диаграмму данного вида. Преимущество данной диаграммы в том, что можно легко сравнивать несколько диаграмм данного типа и проводить анализ параметром. К недостаткам относятся: ограничение по количеству отражаемых значений, занимает много места, неудобны для точных сравнений.

Диаграмма усы-ящики (ящик с усами, диаграмма японские свечи) (Рис. 3. д)

Диаграмма «ящик с усами» показывает распределение данных по квартилям, выделяя их выбросы. «Усы» (линии) указывают на отклонение за пределами верхнего и нижнего квартилей, и если имеется точка за пределами «усов», то такая точка считается выбросом. Преимущество графика в экономии пространства, что удобно при анализе больших данных.

Радарная диаграмма (лепестковая) (Рис. 3. е)

Радарная диаграмма позволяет выполнять сравнение разных вариантов по нескольким параметрам. Данный график удобен для выявления переменных с одинаковыми значениями, для определения значений этих выбросов, для сравнения максимальных и минимальных значений переменных. Недостатком является ограничение по набору данных для углов многоугольника, чем больше углов, тем менее читаемой становится диаграмма.

Мнемосхема (Рис. 3. ж)

Мнемосхема отображает графическую структуру системы. Удобно использовать для понимания архитектуры системы и управления.

Граф (сеть) (Рис. 3. з)

Данный тип визуализации демонстрирует взаимосвязь объектов в виде узлов/вершин и связующих их линий и помогает высветить тип взаимоотношений в рамках группы объектов. Недостатком является ограничение по объему данных, так как чем больше количество узлов, тем сеть становится нечитаемой. Преимуществом метода является его гибкость и возможность визуализации разнородной информации

Отображение на карте (картограмма) (Рис. 3. и)

Картограмма показывает зависимость данных от географии. Используя картограмму, можно отобразить разного рода информацию, например атаки на сайты разных регионов. Пузырьковая карта — на этой карте данных, окружности изображаются поверх определенных географических регионов и их размер пропорционален значению данных. Карты удобны для сравнения пропорциональных значений в разных географических регионах. Недостатком таких карт является наложение больших по размеру пузырьков друг на друга, а также выход за границу маленького региона.

Тепловые карты (Рис. 3. к)

Карта рисков в ИБ для многомерного анализа данных и выявления корреляций. Недостаток: данная схема больше подходит для демонстрации более обобщенных числовых данных, поскольку точно различить цветовые оттенки и извлечь из них конкретные данные может быть сложно.

Древовидная диаграмма (Рис. 3. л)

Древовидная диаграмма — метод визуального представления иерархии в древовидной структуре. Преимущество метода — легкость считывания информации. Недостатком диаграммы является невозможность визуализации достаточно больших деревьев

Лучевая диаграмма (Sunburts) (Рис. 3. м)

Такой тип диаграмм используется для представления иерархических данных в компактной форме, что является преимуществом.

Облако тегов (Рис. 3. н)

Каждому элементу в облаке тега присваивается определенный весовой коэффициент, который коррелирует с размером шрифта. Преимущества: простота и легкость понимания. К недостаткам относится наличие длинных слов, которые обращают на себя большее внимание, чем короткие слова, неудобно использовать для анализа.

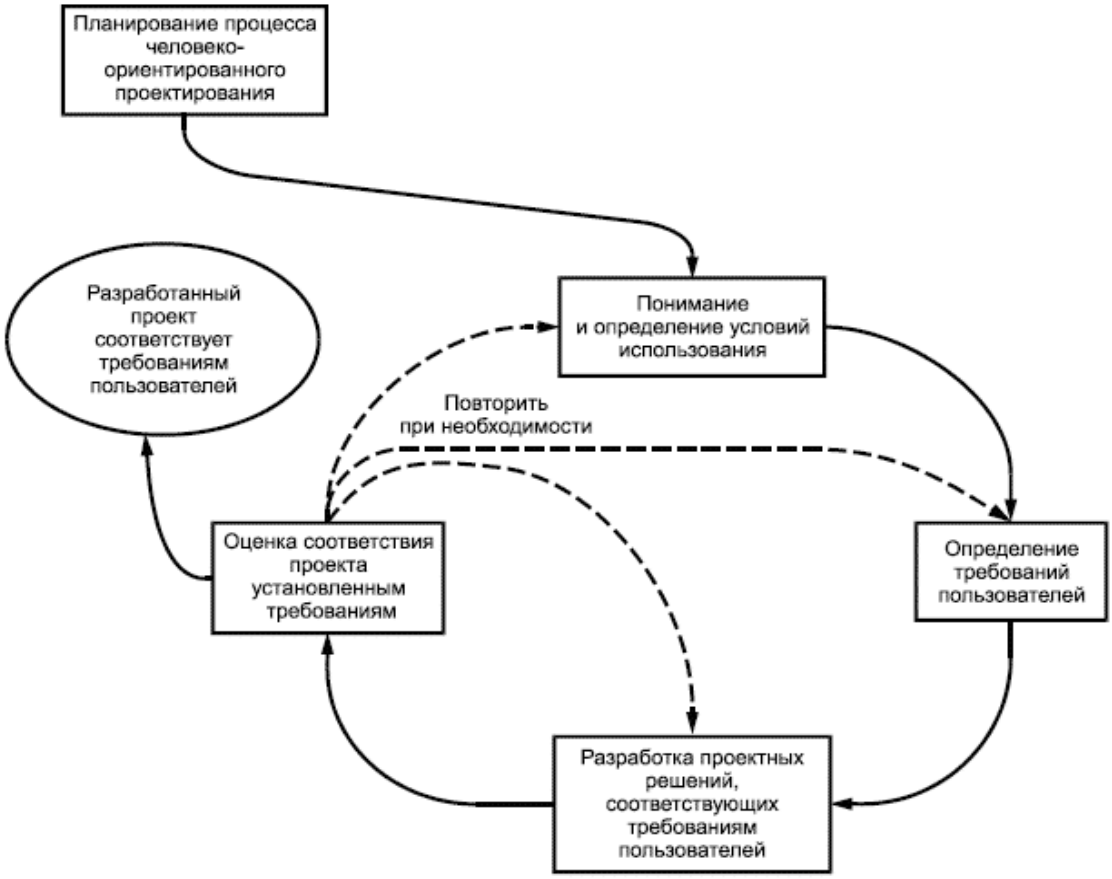

Рис. 4. Взаимосвязь этапов человеко-ориентированного проектирования

Рекомендации по проектированию пользовательских интерфейсов. Рассмотрим некоторые стандарты и рекомендации по проектированию пользовательских интерфейсов SIEM систем.

- **ГОСТ Р ИСО 9241–210–2016** «Человеко-ориентированное проектирование интерактивных систем руководство по человеко-ориентированному проектированию компьютерных интерактивных систем». В стандарте представлены:
- способы улучшения взаимодействия человек-система за счет аппаратных и программных компонентов интерактивных систем.
- принципы человеко-ориентированного проектирования:
- точное определение пользователей, задач и среды
- вовлечение пользователей в проектирование и разработку
- улучшение проекта за счет его оценки пользователями
- итеративное совершенствование проекта
- учет восприятия пользователем системы
- включение в группу специалистов с навыками и знаниями в различных областях
- Аспекты планирования работ по человеко-ориентированного проектирования:
- контрольный перечень требований к пользовательскому интерфейсу

ГОСТ Р ИСО 14915–1–2016 «Эргономика мультимедийных пользовательских интерфейсов. Часть 1 — Принципы проектирования и структура»

В стандарте представлены:

- требования и рекомендации для эргономического проектирования мультимедийных приложений
- принципы организации диалога**:**

о пригодность интерфейса для выполнения производственного задания.

- о информативность
- о управляемость
- о соответствие ожиданиям пользователя
- о устойчивость к ошибкам
- о пригодность к индивидуализации
- о пригодность для изучения
- о пригодность для целей коммуникации

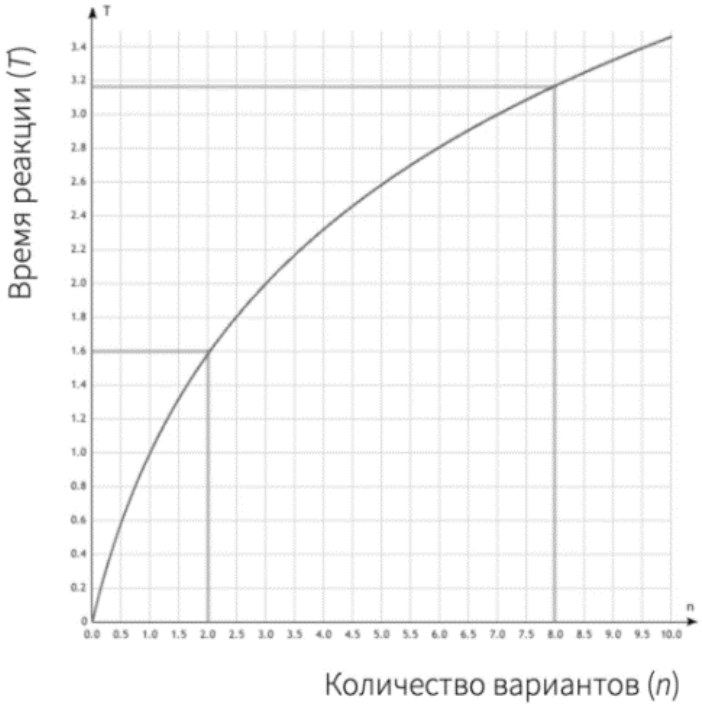

Рис. 5. зависимость времени реакции от количества вариантов

- о пригодность для восприятия и понимания
- о пригодность для изучения
- о привлекательность

«ГОСТ Р ИСО 14915–2–2016 «Эргономика мультимедийных пользовательских интерфейсов. Часть 2 — Навигация и управление мультимедийными средствами». В стандарте представлены:

- руководство и требования к разработке мультимедийных пользовательских интерфейсов с учетом организации информационного наполнения, навигации и управления формами представления информации.
- рекомендации по анализу структуры информационного наполнения и разработке навигации и управления в мультимедийных приложениях.

«Магическое число семь плюс-минус два» (закон Миллера)[4]. Кратковременная память человека способна запоминать в среднем девять двоичных цифр, восемь десятичных цифр, семь букв алфавита и пять односложных слов — то есть человек способен одновременно помнить 7 ± 2 элементов.

Данное правило рекомендуется использовать при проектировании пользовательского интерфейса и ограничивать количество элементов, которые могут быть одновременно представлены на аналитической панели — 7 ± 2 диаграммами

Закон Фиттса [5]:

$$T = a + b\,log_2\left(\frac{D}{W} + 1\right)$$

где:

T — среднее время, затрачиваемое на совершение действия

a — среднее время запуска/остановки движения

b — величина, зависящая от типичной скорости движения

D — дистанция от точки старта до центра цели

W — ширина цели, измеренная вдоль оси движения

Чем дальше или точнее выполняется движение, тем больше коррекции (времени) необходимо для его выполнения

Рекомендуется применять закон Фиттса при проектировании пользовательского интерфейса и проектировать элементы интерфейса, их размеры и расположение для удобного (простого) использования пользователем

Закон Хика [6]:

$$T = b\log_2\left(n + 1\right)$$

Где:

T — среднее время реакции

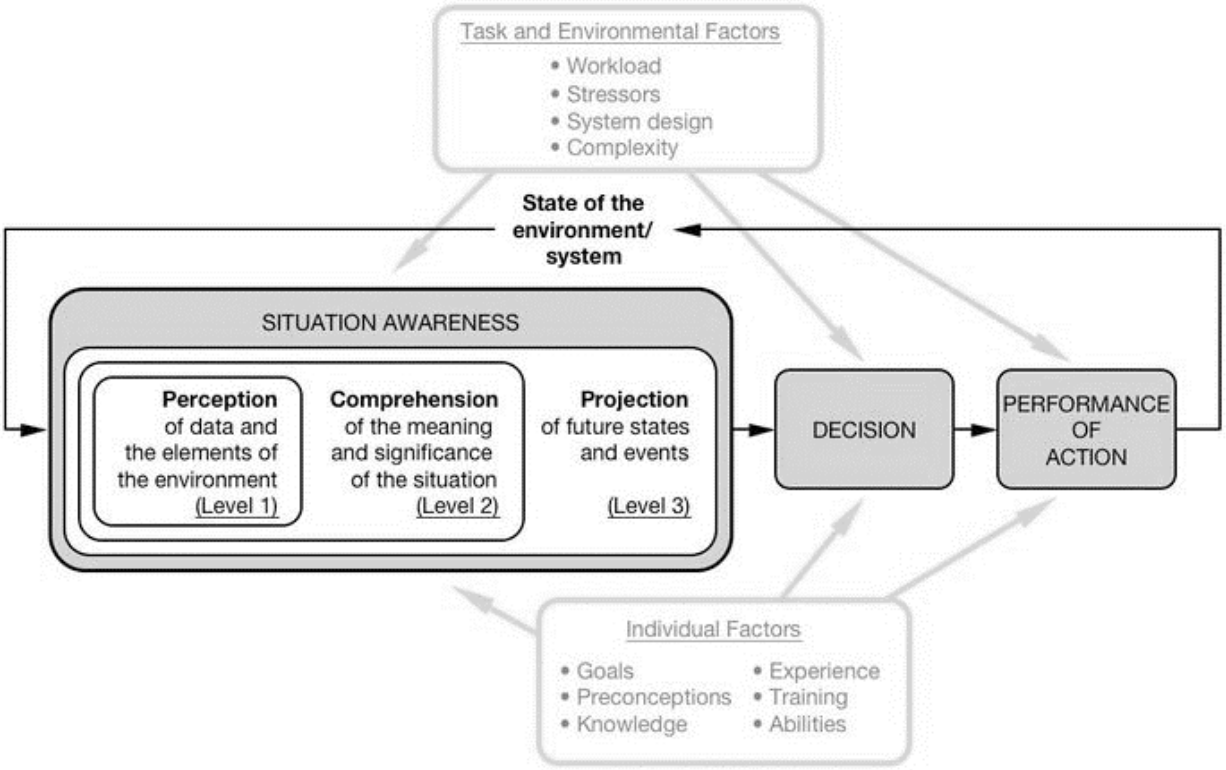

Рис. 6. Место ситуационной осведомленности в процессе принятия решения

b — константа, которая описывает индивидуальные особенности восприятия, такие как задержка перед выполнением задания и индивидуальный коэффициент скорости принятия решения

n — количество вариантов действий.

- Время реакции, которое, необходимо пользователю для выполнения задачи, увеличивается (логарифмическая зависимость) с количеством доступных вариантов

Определение необходимого уровня ситуационной осведомленности [7, 8]

При проектировании аналитической панели необходимо определить необходимый уровень ситуационной осведомленности, и в соответствии с ним определить необходимые диаграммы и данные, выводимые на интерфейс. Для каждой роли свой функционал: 1-й уровень (1 SA) — восприятие необходимых данных (мониторинг инцидентов) 2-й уровень (2 SA) — понимание текущей ситуации (интерпретация и оценку состояния) 3-й уровень (3 SA) — проекция будущего состояния (определение вероятного развития ситуации)

Десять эвристик Нильсена [9] — 10 принципов проектирования пользовательских интерфейсов, разработанных Я. Нильсеном:

- Осведомленность о состоянии системы (Обратная связь).
- Схожесть системы с реальным миром (Следование общепринятым соглашениям в обычном мире)
- Свобода действий и контроль. (Возможность отменить и восстановить действия пользователя)
- Целостность и следование стандартам. (Использование общепринятых правил)
- Предотвращение ошибок. (Защита пользователя от совершения ошибок)
- Минимизация необходимости запоминания информации пользователем. (Перемещаясь по системе, пользователь не должен запоминать информацию — все инструкции к системе должны быть видны или легко извлекаться при необходимости)
- Гибкость и эффективность. (Возможность подстроить интерфейс «под себя»)
- Эстетика и минимализм в дизайне. (Не должно быть лишней (неважной) информации)
- Понимание проблем и их решение. (Сообщения об ошибках должны быть написаны простым языком)
- Справочная информация. (Наличие справки: краткой, актуальной)

8 золотых правил дизайна интерфейса Б. Шнейдермана[10]. Данные правила очень схожи с эвристиками и направлены на повышение юзабилити пользователя:

Рис. 7. Методы оценки качества пользовательских интерфейсов

- Стремление к последовательности
- Использование постоянными пользователями клавиш быстрого доступа
- Обязательная обратная связь
- Диалог с логичным завершением
- Простой процесс обработки ошибок
- Легкий механизм отмены действий
- Внутренний фокус контроля
- Уменьшение кратковременной нагрузки на память

Оценка качества интерфейса. На сегодняшний день применяется довольно большое количество разнообразных методов оценки качества пользовательских интерфейсов [11]:

Показатели качества программных средств рекомендуется выбирать, основываясь на ГОСТ 28195–89 «Оценка качества программных средств» с учетом специфики интерфейса. В дополнение к показателям, указанным в стандарте, для оценки качества инструментов визуализации больших данных предлагается использовать показатели, основанные на свойства больших данных (8V):

- **Volume — объем** — Инструмент визуализации позволяет анализировать большие объемы информации;
- **Velocity– скорость** — Инструмент визуализации позволяет анализировать высокоскоростную обработку данных;
- **Variety — разнообразность** — Инструмент визуализации позволяет анализировать многообразие больших данных;
- **veracity — достоверность** — Инструмент визуализации позволяет оценивать достоверность больших данных;
- **viability — жизнеспособность** — Инструмент визуализации позволяет определять актуальные данные;
- **value — ценность** — Инструмент визуализации позволяет определять ценность больших данных;
- **variability — переменчивость** — Инструмент визуализации позволяет определять динамику больших данных;
- **visibility** — визуализация — Инструмент позволяет визуализировать данные в подходящей форме (наглядно);

А также общие показатели, присущие инструментам визуализации

- информативность (насыщенность элемента информацией);
- достаточность (инструмент позволяет получить всю необходимую информацию о данных);
- простота (понятность) (инструмент не требует больших умственных затрат для анализа);
- удобство (не требует больших зрительных затрат на анализ);
- отзывчивость (гибкость) (инструмент позволяет получать дополнительную информацию при изменении параметров);
- точность (инструмент достоверно описывает данные);
- актуальность (элемент содержит только нужную информацию);
- динамичность (изменение диаграммы в реальном времени).

Оценку качества пользовательского интерфейса рекомендуется проводить с использованием достаточного количества экспертных оценок и автоматизации для по-

лучения наиболее точных результатов. При этом, при отсутствии средств автоматизации и малых ресурсов для проведения оценки целесообразно использовать эффективные методы экспертного оценивания, к примеру, метод попарных сравнений, с расчетом коэффициента конкордации для оценки согласованности экспертов W, определением весомости показателей и коэффициента вариации для каждого критерия.

Коэффициент Конкордации рассчитывается следующим образом:

$$W = {12\ \sum_{j=1}^{n}\left(S_j - \bar{S}\right)^2}\Big/{n^2(m^3 - m)},$$

где $S_j = \sum_{i=1}^{n} R_{ij}$ — сумма ранговых оценок экспертов по каждому показателю,

$$\bar{S} = \frac{1}{2}n(m+1) —$$

средняя сумма рангов для всех показателей, m — количество показателей, n — количество экспертов. Если $W = 0$, то имеет место абсолютная несогласованность экспертов, если $W \geq 0.6$, то считают согласованность приемлемой.

Коэффициент весомости показателя g_i рассчитывается, как

$$g_j = \frac{nm - S_j}{\frac{1}{2}nm(m-1)}.$$

Существенно значимыми (весомыми) считаются показатели, для которых: $g_j > 1/m$. Коэффициент вариации C_{Rj} рассчитывается, как

$$C_{R_j} = \frac{\sigma_{R_j}}{\overline{R_j}}100\%.$$

Согласованность экспертов высокая, если $C_{Rj} < 40\%$.

При работе с противоречивыми, неточными и субъективными данными рекомендуется дополнительно применять аппарат нечеткой логики, позволяющий получить достоверные решения при нечетких входных данных. Аппарат нечеткой логики возможно применять в различных ситуациях [12] как при наличии только экспертной информации, так и при наличии статистических данных (к примеру, используя нечёткие нейронные сети (ANFIS)).

Далее предлагается обобщённый алгоритм построения пользовательского интерфейса для анализа больших данных.

Алгоритм построения интерфейса

В результате проведенного анализа современных методов проектирования пользовательских интерфейсов, предлагается обобщенный алгоритм построения пользовательского интерфейса для визуализации больших данных.

Рис. 8. Обобщенный алгоритм построения пользовательского интерфейса для визуализации больших данных

Заключение

В работе проведено исследование методов визуализации больших данных в информационных системах, и в частности, в SIEM системах. Рассмотрены современные инструменты визуализации информации, отмечены их достоинства, недостатки возможные варианты использования. Проанализированы некоторые аспекты построения и оценки эргономичных аналитических панелей, предложен обобщенный алгоритм построения эргономичных информационных панелей для SIEM. Для оценки качества инструментов визуализации больших данных предложено использовать показатели, основанные на свойства больших данных.

Результаты работы могут быть использованы как разработчиками интерфейсов SIEM, так и других систем анализа больших данных.

ЛИТЕРАТУРА

1. Чесалин А.Н., Гродзенский С.Я., Нилов М.Ю., Фам Ван Ты Интеллектуальные инструменты управления качеством цифрового производства // Стандарты и качество, 2020, № 3, с. 68–72
2. Chesalin A.N., Grodzenskiy S. Ya., Nilov M. Yu., Pham Van Tu Intelligent quality management tools for digital production and knowledge management system for their application электронная IOP Conference Series: Materials Science and Engineering, 862 (2020), 042032, 9 p. doi:10.1088/1757–899X/862/4/042032
3. C. Knaflic. Storytelling with Data: A Data Visualization Guide for Business Professionals, 2015.
4. George A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information // The Psychological Review. — 1956. — Vol. 63. — Pp. 81–97.
5. Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement // Journal of Experimental Psychology. — 1954. — Vol. 47. — No. 6. — Pp. 381–391.
6. Hick, W.E. (1 March 1952). "On the rate of gain of information" (PDF). Quarterly Journal of Experimental Psychology. 4 (1): 11–26

7. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32–64, 1995.
8. Кузьмина Н. Человеко-ориентированный подход при проектировании систем визуализации автоматизированных объектов // СТА, 1, 2015.
9. Nielsen, Jakob. (1996). Usability Metrics: Tracking Interface Improvements. IEEE Software. 13. 12–13. 10.1109/MS.1996.8740869.
10. Шнейдерман, Б. Психология программирования: человеческие факторы в вычислительных и информационных системах / Б. Шнейдерман; пер. с англ. А.И. Горлина, Ю.Б. Котова.— Москва: Радио и связь, 1984.—304 с
11. Зенг Валерия Андреевна Оценка качества проектирования пользовательских интерфейсов нового поколения // Известия ТулГУ. Технические науки. 2019. № 12.
12. Чесалин А.Н. Гродзенский С.Я., Фам Ван Ты, Нилов М.Ю., Агафонов А.Н. Технология оценки рисков на этапах жизненного цикла продукции с использованием нечеткой логики // Российский технологический журнал, 2020, № 8(6), с. 167–183. doi:10.32362/2500–316X-2020–8–6–167–183.
13. Саати Т.Л. Принятие решений. Метод анализа иерархий.— М.: Радио и связь, 1989.— 316 с.

© Ушкова Надежда Николаевна (ushkova.n.n@mail.ru),
Чесалин Александр Николаевич (chesalin_an@mail.ru), Болотин Кирилл Викторович (bolotin@mirea.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»

МИРЭА