Метод численной оценки технической интероперабельности

**Человек:** Предмет исследований является метод оценки качества средств технической интероперабельности открытых информационных систем в соответствии с моделью интероперабельности по ГОСТ Р 55062–2012. В работе предложен метод расчёта оценочных характеристик технической интероперабельности на основе оценки детектирования и преобразования форматов данных узлами открытой информационной системы. В качестве иллюстрации в работе выполнены практические оценки технической интероперабельности, по критериям детектирования и преобразования форматов, некоторых программных пакетов обработки графических изображений. В работе отмечено, что при необходимости наблюдений за эволюцией интероперабельности информационной системы, необходимо сохранение методической основы формирования исходных параметров. Предложенный метод оценки базируется на законах теории информации и определяется формальной моделью позволяющей вычислить частную и общую информационную энтропию, характеризующую техническую интероперабельность как способность исследуемой информационной системы детектировать данные в различных форматах и способность исследуемой информационной системы производить корректные преобразования данных в различных форматах. Основным результатом исследований является разработанный метод оценки технической интероперабельности позволяющий производить численную оценку свойств технической интероперабельности с точки зрения детектирования и преобразования форматов. Практическим результатом работы является получение возможности численной оценки интероперабельности, наблюдение за динамикой интероперабельности различных систем в условиях взаимодействия с непрерывно изменяющимся межсистемным информационным пространством.

**Key words:** интероперабельность, открытые информационные системы, информационная энтропия, Преобразование форматов, модель оценки интероперабельности, численная оценка интероперабельности, уровни интероперабельности, детектирование формата, преобразование данных, MATLAB

=================================

**FastText\_KMeans\_Clean:** 3. ISO/IEC 2382-1:1993 Information technology – Vocabulary. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Для вычисления – матрицы условных вероятностей преобразования форматов, необходимо определить матрицу вероятностей преобразования данных в формат при получении данных в формате : . Общая информационная энтропия . Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации. – мощность множества форматов Paint 5 . "один ко многим" и "многие к одному" для различных УИС.

**Key words part:** 0.8666666666666667

=================================

**FastText\_KMeans\_Raw/:** 3. ISO/IEC 2382-1:1993 Information technology – Vocabulary. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации. – мощность множества форматов GIMP2 ;. Для каждого УИС, сформируем матрицы учитывая, что программные пакеты позволяют производить взаимные преобразования по всем форматам, т.е. обладают полнотой взаимных преобразований, получим, что при равных вероятностях взаимных преобразований и условии полного взаимного преобразования форматов, значения частных энтропий для определённого узла УИС "один ко многим" будут равны значениям "многие к одному": при (рис. 3). Распределение частных энтропий. Полные энтропии УИС имеют следующие значения:. Авторы не претендуют, на то что предлагаемый метод предоставляет возможность произвести оценку технической интероперабельности в полном объёме, в соответствии с эталонной моделью по ГОСТ Р 55062–2012, при этом предлагаемый метод предоставляет исследователю готовый математический аппарат оценки детектирования и преобразования форматов, что относится к технической интероперабельности.

**Key words part:** 0.8666666666666667

=================================

**FastText\_PageRank\_Clean/:** Значительный вклад в формирование модели интероперабельности внесли следующие отечественные исследователи: В. К. Батоврин, Ю. В. Гуляев, А. Я. Олейников. 1. ISO/IEC 24765:2009, Systems and Software Engineering – Vocabulary;. 2. ISO/IEC 19500-2:2003 Information technology – Open Distributed Processing;. 3. ISO/IEC 2382-1:1993 Information technology – Vocabulary. – "один ко многим": ;. – "многие к одному": . Общая информационная энтропия . Распределение частных энтропий.

**Key words part:** 0.5666666666666667

=================================

**FastText\_PageRank\_Raw/:** Значительный вклад в формирование модели интероперабельности внесли следующие отечественные исследователи: В. К. Батоврин, Ю. В. Гуляев, А. Я. Олейников. 1. ISO/IEC 24765:2009, Systems and Software Engineering – Vocabulary;. 2. ISO/IEC 19500-2:2003 Information technology – Open Distributed Processing;. 3. ISO/IEC 2382-1:1993 Information technology – Vocabulary. Определение объектов оценивания. – "один ко многим": ;. – "многие к одному": . Распределение частных энтропий.

**Key words part:** 0.5

=================================

**Mixed\_ML\_TR/:** В работе предлагается модель оценки способности информационной системы детектировать форматы данных и производить преобразования форматов данных, что в большей степени относится к техническому уровню интероперабельности. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Тогда вероятности встречи детектируемого формата в потоке информационного обмена для всех УИС () возможно, рассчитать как произведение матриц: . Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации. – мощность множества форматов Potoshop CS2 ;. Далее произведём оценку технической интероперабельности тех же УИС с учётом фактической вероятности присутствия форматов в информационном обмене. Авторы не претендуют, на то что предлагаемый метод предоставляет возможность произвести оценку технической интероперабельности в полном объёме, в соответствии с эталонной моделью по ГОСТ Р 55062–2012, при этом предлагаемый метод предоставляет исследователю готовый математический аппарат оценки детектирования и преобразования форматов, что относится к технической интероперабельности.

**Key words part:** 0.9

=================================

**MultiLingual\_KMeans/:** Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Тогда вероятности встречи детектируемого формата в потоке информационного обмена для всех УИС () возможно, рассчитать как произведение матриц: . – мощность множества форматов Potoshop CS2 ;. Далее произведём оценку технической интероперабельности тех же УИС с учётом фактической вероятности присутствия форматов в информационном обмене. Авторы не претендуют, на то что предлагаемый метод предоставляет возможность произвести оценку технической интероперабельности в полном объёме, в соответствии с эталонной моделью по ГОСТ Р 55062–2012, при этом предлагаемый метод предоставляет исследователю готовый математический аппарат оценки детектирования и преобразования форматов, что относится к технической интероперабельности.

**Key words part:** 0.8

=================================

**Multilingual\_PageRank/:** В отечественной печати практически отсутствуют публикации ориентированные на количественную оценку технической интероперабельности как уровня эталонной модели интероперабельности по ГОСТ Р 55062–2012. Далее возможно перейти к расчёту частных информационных энтропий:. – мощность множества форматов Paint 5 . В частности, и , т.е. множество форматов не содержит формат применяёмый для создания графических меток-логотипов изображений в настольных и сетевых приложениях. Распределение частных энтропий. Очевидно, что по сравнению с предыдущими расчётами, значение энтропии снизилось, что является следствием использования реальной плотности распределения форматов (а не равномерной предполагаемой) и уменьшения числа форматов до фактически присутствующих в МИП. Распределение по форматам частных энтропий. Выявлено отсутствие практического инструментария количественной оценки технической интероперабельности.

**Key words part:** 0.7333333333333333

=================================

**RuBERT\_KMeans\_Without\_ST/:** Например, в системах САПР ориентированных на картографию и мониторинг окружающего пространства (геоинформационные системы) [6,15,16], при преобразовании форматов данных, может потребоваться достаточно сложный анализ и обработка данных (например, распознавание образов, решение дифференциальных и интегральных уравнений, интеграция данных из нескольких форматов в один, разделение данных одного формата на данные в многих форматах и др.). Для каждого УИС, сформируем матрицы учитывая, что программные пакеты позволяют производить взаимные преобразования по всем форматам, т.е. обладают полнотой взаимных преобразований, получим, что при равных вероятностях взаимных преобразований и условии полного взаимного преобразования форматов, значения частных энтропий для определённого узла УИС "один ко многим" будут равны значениям "многие к одному": при (рис. 3). Далее, учитывая, что детектирование формата производится в автоматизированном режиме, т.е. с участием человека-оператора, рассчитаем информационную энтропию, характеризующую техническую интероперабельность каждого УИС. Авторы не претендуют, на то что предлагаемый метод предоставляет возможность произвести оценку технической интероперабельности в полном объёме, в соответствии с эталонной моделью по ГОСТ Р 55062–2012, при этом предлагаемый метод предоставляет исследователю готовый математический аппарат оценки детектирования и преобразования форматов, что относится к технической интероперабельности.

**Key words part:** 0.8666666666666667

=================================

**RuBERT\_KMeans\_With\_ST/:** Учитывая наличие отечественного ГОСТа определяющего модель интероперабельности и соответствующий терминологический аппарат, в исследованиях будем использовать, в качестве базовой, модель рассмотренную в ГОСТ Р 55062–2012 и модель предшествующая выходу ГОСТ опубликованную в работе [7]. В работе предлагается модель оценки способности информационной системы детектировать форматы данных и производить преобразования форматов данных, что в большей степени относится к техническому уровню интероперабельности. Вероятность получения системой формата из множества F , известна, она задана диагональной матрицей . Распределение по форматам частных энтропий.

**Key words part:** 0.8666666666666667

=================================

**RUBERT\_page\_rank\_Without\_ST/:** Оценка преобразования форматов. – мощность множества форматов Potoshop CS2 ;. Мощность объёдинённого множества форматов . Полные энтропии УИС имеют следующие значения:. Полные энтропии УИС имеют следующие значения:.

**Key words part:** 0.5333333333333333

=================================

**RUBERT\_page\_rank\_With\_ST/:** – сумма вероятностей матрицы будет равна единице, т.е.: ;. Зачастую, в практических задачах, ситуация обстоит следующим образом. – "один ко многим": ;. – "многие к одному": . Распределение частных энтропий.

**Key words part:** 0.3333333333333333

=================================

**RUSBERT\_KMeans\_Without\_ST/:** Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Для оценки преобразования форматов необходимо определить матрицу условных вероятностей , формата необходимого преобразования от формата получения в спросе и предложении информационного обмена. – сумма каждой строки матрицы , равна вероятности преобразования формата в форматы для ("один ко многим");. Мощность объёдинённого множества форматов . Далее есть смысл исключить из рассмотрения форматы, которые отсутствуют в МИП, а значит по условию, что обмен информацией между УИС производится через МИП, отсутствуют и в информационном обмене.

**Key words part:** 0.6333333333333333

=================================

**RUSBERT\_KMeans\_With\_ST/:** Если УИС, из множества форматов F , не детектирует один единственный формат, то эта ситуация эквивалентна детектированию всех форматов, т.к. в данном случае формат распознаваем методом исключения, т.е. детектируем, достаточно лишь знать что не детектируемый формат единственный. – сумма каждого столбца матрицы , равна вероятности преобразования в формат при получении данных в форматах для ("многие к одному"). Мощность объёдинённого множества форматов . В работе выполнен анализ методов оценки интероперабельности в соответствии с эталонной моделью по ГОСТ Р 55062–2012 и методов технической совместимости узлов информационных систем.

**Key words part:** 0.8333333333333334

=================================

**RUSBERT\_page\_rank\_Without\_ST/:** – "один ко многим": ;. – "многие к одному": . Полные энтропии УИС имеют следующие значения:. Полные энтропии УИС имеют следующие значения:. Выявлено отсутствие практического инструментария количественной оценки технической интероперабельности.

**Key words part:** 0.5333333333333333

=================================

**RUSBERT\_page\_rank\_With\_ST/:** Определение объектов оценивания. – "один ко многим": ;. – "многие к одному": . Общая информационная энтропия . Выявлено отсутствие практического инструментария количественной оценки технической интероперабельности.

**Key words part:** 0.6

=================================

**Simple\_PageRank/:** Таким образом, системе преобразования необходимо принимать корректные решения, возможно, на основе информации содержащейся в преобразуемых данных, что относится к способности взаимодействующих в процессе коммуникации информационных систем одинаковым образом понимать смысл информации [8], которой они обмениваются, т.е. семантической интероперабельности [17,18], а с учётом прагматических аспектов преобразования, безусловно, затрагивает и организационный уровень интероперабельности в информационном обмене [19-23]. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Тогда вероятности встречи детектируемого формата в потоке информационного обмена для всех УИС () возможно, рассчитать как произведение матриц: . При решении практических задач значения элементов матрицы может быть вычислить затруднительно, при этом мы не можем гарантировать, но можем предположить, что эти условные вероятности взаимных преобразований хорошо соотносятся с вероятностями вычисленными как: , где – матрица размером N ×N значения всех элементов которой равны единице. Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации. Отметим, что для удобства всем форматам присвоены условные номера, порядок нумерации форматов не влияет на конечный результат – оценку, но определяет вид диаграммы распределения.

**Key words part:** 0.9

=================================

**TextRank/:** Целью исследований является разработка метода численной количественной оценки технической интероперабельности в смысле оценки детектирования и преобразования форматов данных и протоколов информационной поддержки. В работе предлагается модель оценки способности информационной системы детектировать форматы данных и производить преобразования форматов данных, что в большей степени относится к техническому уровню интероперабельности. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации. Для каждого УИС, сформируем матрицы учитывая, что программные пакеты позволяют производить взаимные преобразования по всем форматам, т.е. обладают полнотой взаимных преобразований, получим, что при равных вероятностях взаимных преобразований и условии полного взаимного преобразования форматов, значения частных энтропий для определённого узла УИС "один ко многим" будут равны значениям "многие к одному": при (рис. 3). В соответствии с целью исследований произведена разработка метода численной оценки детектирования и преобразования форматов данных, что в большей степени относится к уровню технической интероперабельности.

**Key words part:** 0.9333333333333332

=================================

**TF-IDF\_KMeans/:** В работе предлагается модель оценки способности информационной системы детектировать форматы данных и производить преобразования форматов данных, что в большей степени относится к техническому уровню интероперабельности. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. . Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Для вычисления – матрицы условных вероятностей преобразования форматов, необходимо определить матрицу вероятностей преобразования данных в формат при получении данных в формате : . Так же как и при оценке детектирования при оценке преобразования форматов в качестве дополнительного показателя оценки преобразования различных форматов, возможно, использовать визуализацию матрицы , с учётом признаков преобразования , используя матричные или трёхмерные диаграммы типа scatter plot. – мощность множества форматов Paint 5 . В результате оценки технической интероперабельности получим равную для всех УИС энтропию, и общую энтропию .

**Key words part:** 0.9

=================================

**Текст:** Введение. В системах агрегирующих большие количества различных и разнородных данных [1-3], в системах с большим числом внешних (межсистемных) взаимодействий [3-5], в распределённых системах [6] использующих разнообразные каналы связи [2,4,6] форматы [3-6] и формы транспорта [4, 5] данных существенной задачей является повышение качества интероперабельности [7]. Устоявшаяся модель определения интероперабельности в отечественной практике исследования открытых систем сформировалась к 2012 году в результате чего был опубликован ГОСТ Р 55062–2012 [8], разработанный специалистами «Института радиотехники и электроники Российской академии наук им. В. А. Котельникова РАН». Значительный вклад в формирование модели интероперабельности внесли следующие отечественные исследователи: В. К. Батоврин, Ю. В. Гуляев, А. Я. Олейников. Следует отметить, что в зарубежной стандартизации ISO определено как минимум три модели интероперабельности [7]:. 1. ISO/IEC 24765:2009, Systems and Software Engineering – Vocabulary;. 2. ISO/IEC 19500-2:2003 Information technology – Open Distributed Processing;. 3. ISO/IEC 2382-1:1993 Information technology – Vocabulary.. Учитывая наличие отечественного ГОСТа определяющего модель интероперабельности и соответствующий терминологический аппарат, в исследованиях будем использовать, в качестве базовой, модель рассмотренную в ГОСТ Р 55062–2012 и модель предшествующая выходу ГОСТ опубликованную в работе [7].. Существуют частные качественные решения численной оценки интероперабельности [9-11], но нет общепринятых, достаточно универсальных количественных оценок. Так авторы работы [9] предлагают комплексную ранговую модель оценки интероперабельности. Авторы работы [10] предлагают разработанные ими варианты шкал для оценивания интероперабельности на основе модели нечёткой логики с учётом весов влияющих факторов. В работе [11] предложены наборы показателей и варианты шкал для оценки приемлемости решений при выборе стандартов открытых систем. Основой формирования показателей оценивания предложенных методик является экспертные заключения.. Существует значительно число работ в области исследования технической совместимости различных устройств и информационных систем. Что более относится к оценке родственных с интероперабельностью направлений [8], такие как совместимость на уровне протоколов передачи данных (Coexistent), способность к соединению (Interconnectable), взаимодействию (Interworkable) и т.д. Так в работе [12], в соответствии с ГОСТ 22315-77 КС определены следующие виды совместимости: конструктивная, технологическая. В руководящем документе РД 50-726-92 [13] предлагается производить комплексную оценку совместимости (технических средств размещённых на морских подвижных объектах) на основе совокупного анализа электромагнитных излучений, условий размещения и эксплуатации, что больше относится к обеспечению снижения искажений в процессе информационного обмена, чем, например, к оценке совместимости информационных каналов. Близкий, по сути, способ оценки технической совместимости предложен в работе [14]. В отечественной печати практически отсутствуют публикации ориентированные на количественную оценку технической интероперабельности как уровня эталонной модели интероперабельности по ГОСТ Р 55062–2012.. В то же время, в работе [7] отмечается что «теория интероперабельности ждёт своего развития с учётом достижений теории информации К. Шенона и В. А. Котельникова…», здесь возможно, добавить выдающегося учёного А. Н. Колмогорова, внёсшего значительный вклад в развитие теории информации. Таким образом, желательно иметь метод, базирующийся на основах теории информации для численной оценки интероперабельности.. Целью исследований является разработка метода численной количественной оценки технической интероперабельности в смысле оценки детектирования и преобразования форматов данных и протоколов информационной поддержки.. Предлагается способ численной оценки интероперабельности базирующийся на методах и положениях теории информации.. . Определение объектов оценивания. Для разработки модели оценивания необходимо определить объекты оценивания в общей структуре эталонной модели интероперабельности [8]. Предлагается построение модели оценки детектирования формата (или протокола) и возможности преобразования данных из одного формата (или протокола) в другой.. Под термином детектирование формата будем понимать возможность определения, что некоторый информационный пакет имеет вполне определённый формат соответствующий заданным стандартам и/или требованиям.. Под термином преобразование данных будем понимать возможность корректного преобразования данных из одного или нескольких детектируемых форматов в другой или другие детектируемые форматы.. Необходимо определить к каким уровням модели интероперабельности относятся параметры: детектирование и преобразование формата.. В соответствии с эталонной моделью межсистемного обмена информацией и интероперабельности принятой по стандарту ГОСТ Р 55062–2012 [8], определено несколько иерархических уровней отвечающих за эволюционные качества интероперабельности. В соответствии с [8], термин интероперабельность определяется как способность двух или более систем, или компонентов к обмену информацией и к использованию информации, полученной в результате обмена.. Из определения трёх уровней (технический, семантический, организационный ) эталонной модели интероперабельности по ГОСТ Р 55062–2012 вполне очевидно, что детектирование формата, относится к техническому уровню интероперабельности.. Преобразование данных выходит за определённый в ГОСТ Р 55062–2012 уровень технической интероперабельности. Преобразование данных может быть достаточно сложной не всегда обратимой процедурой. Например, в системах САПР ориентированных на картографию и мониторинг окружающего пространства (геоинформационные системы) [6,15,16], при преобразовании форматов данных, может потребоваться достаточно сложный анализ и обработка данных (например, распознавание образов, решение дифференциальных и интегральных уравнений, интеграция данных из нескольких форматов в один, разделение данных одного формата на данные в многих форматах и др.).. Таким образом, системе преобразования необходимо принимать корректные решения, возможно, на основе информации содержащейся в преобразуемых данных, что относится к способности взаимодействующих в процессе коммуникации информационных систем одинаковым образом понимать смысл информации [8], которой они обмениваются, т.е. семантической интероперабельности [17,18], а с учётом прагматических аспектов преобразования, безусловно, затрагивает и организационный уровень интероперабельности в информационном обмене [19-23].. В модели интероперабельности, рассмотренной в работе [7] и опирающейся на исследования [18, 25-27], содержится девять видов интероперабельности: техническая, синтаксическая, семантическая, прагматическая, динамическая, организационная, концептуальная, на основе обмена, интеграционная. В соответствии с этой моделью преобразование форматов в большей степени относится к синтаксическому виду интероперабельности, т.е. способности к обмену данными. В модели ГОСТ Р 55062–2012 синтаксический уровень, как отдельный вид интероперабельности, отсутствует и включён в состав технического уровня.. В данной работе будем считать, что детектирование и преобразование форматов, в большей степени относится к виду технической интероперабельности. При решении практических задач, следует учитывать рекомендации ГОСТ Р 55062–2012 «для систем конкретных классов на базе эталонной модели интероперабельности должны создаваться проблемно-ориентированные модели интероперабельности, которые могут иметь большее число уровней».. В работе предлагается модель оценки способности информационной системы детектировать форматы данных и производить преобразования форматов данных, что в большей степени относится к техническому уровню интероперабельности. Детектирование и преобразование данных является неотъемлемыми частями любого информационного обмена [19-24], поэтому с практической стороны их оценка возможна и полезна.. . Оценка детектирования форматов. Предположим, что в межсистемной информационной среде (МИС) допустимо присутствие форматов из упорядоченного множества F = (f 1, f 2, f 3, …, fN ) , тогда мощность множества |F | = N будет равна числу допустимых форматов.. Пусть каждый формат МИС из множества F имеет, вполне определённую (экспериментально или иначе) вероятность присутствия в информационном обмене. Запишем эти вероятности в виде диагональной матрицы [28, 29] вероятностей размером N ×N : , причём вероятности с индексами соответствуют вероятности присутствия в информационном обмене формата. Суммарная вероятность присутствия в информационном обмене, допустимых форматов равна единице, т.е. .. Пусть для оценки детектирования и преобразования форматов определено упорядоченное множество узлов информационных систем (УИС) , могут быть и другие УИС взаимодействующие с МИС но не входящие в подлежащее оценке множество УИС. Пусть каждый УИС, взаимодействующий с МИС имеет ограниченное упорядоченное множество детектируемых форматов , причём и все детектируемые форматы образующие кортеж взяты из упорядоченного множества , т.е. . Признак детектирования формата является бинарным.. Введём множество бинарных признаков детектирования форматов для всех УИС из , с помощью матрицы признаков, строки которой будут соответствовать различным УИС, а столбцы различным форматам:. .. Тогда вероятности встречи детектируемого формата в потоке информационного обмена для всех УИС () возможно, рассчитать как произведение матриц: . УИС могут получать и не детектируемые форматы, т.к. не детектируемые форматы УИС не могут отличить друг от друга, то возможно определить общую вероятность получения, не детектируемого формата для каждого УИС как: . Таким образом, суммарная вероятность получения детектируемых и не детектируемых форматов будет: . Если УИС, из множества форматов F , не детектирует один единственный формат, то эта ситуация эквивалентна детектированию всех форматов, т.к. в данном случае формат распознаваем методом исключения, т.е. детектируем, достаточно лишь знать что не детектируемый формат единственный. Строки матрицы характеризуют вероятностную способность определённого УИС идентифицировать форматы из всего множества F , столбцы матрицы характеризуют вероятностную способность идентифицировать определённый формат из множества F , всем множеством узлов .. Далее можно рассчитать информационную энтропию [30, 31] каждого УИС:. , здесь и далее логарифм берётся по основанию 2. Используя свойство аддитивности энтропии [31], энтропию всего множества УИС, вычислим как: .. Показатели и предлагается рассматривать как относительные оценки детектирования форматов множеством УИС взаимодействующих в общем информационном пространстве. В соответствии со свойствами информационной энтропии [31] максимальные значения энтропии возможны, когда все детектируемые форматы в информационном пространстве равновероятны и вероятность отлична от нуля.. В качестве дополнительного показателя оценки детектирования различных форматов различными УИС возможно использовать визуализацию матрицы на матричной или трёхмерной диаграмме типа scatter plot.. . Оценка преобразования форматов. Для оценки преобразования форматов необходимо определить матрицу условных вероятностей , формата необходимого преобразования от формата получения в спросе и предложении информационного обмена. Определить такую матрицу можно различными способами, например, экспериментальным наблюдением за значительным множеством преобразований с последующим расчётом статистики. Так же возможна теоретическая оценка, различными способами. При необходимости наблюдений за эволюцией технической интероперабельности информационной системы, необходимо сохранение методической основы формирования матрицы взаимных вероятностей, т.к. взаимные вероятности могут быть вычислены различными способами. Это связанно с тем, что получаемые информационные характеристики (энтропия) могут значительно варьироваться при различных распределениях вероятности. При сохранении методики расчёта вероятностей, распределение вероятности будет меньше подвержено вариабельности в результате методической ошибки, т.е. ошибки измерений или моделирования. Поэтому необходимо сохранять методику расчёта вероятностей, если имеется необходимость сравнительных наблюдений.. Предлагается следующая методика расчёта матрицы условных вероятностей.. Вероятность получения системой формата из множества F , известна, она задана диагональной матрицей . Для вычисления – матрицы условных вероятностей преобразования форматов, необходимо определить матрицу вероятностей преобразования данных в формат при получении данных в формате : .. Далее возможно рассчитать матрицу условных вероятностей получения данных в формате и преобразования в формат , как: .. В результате получим матрицу взаимных вероятностей преобразования форматов, обладающую следующими свойствами:. – сумма вероятностей матрицы будет равна единице, т.е.: ;. – сумма каждой строки матрицы , равна вероятности преобразования формата в форматы для («один ко многим»);. – сумма каждого столбца матрицы , равна вероятности преобразования в формат при получении данных в форматах для («многие к одному»).. При решении практических задач значения элементов матрицы может быть вычислить затруднительно, при этом мы не можем гарантировать, но можем предположить, что эти условные вероятности взаимных преобразований хорошо соотносятся с вероятностями вычисленными как: , где – матрица размером N ×N значения всех элементов которой равны единице.. Учитывая, что в реальных системах не все форматы преобразуемы друг в друга, необходимо ввести матрицу признаков преобразования форматов, что возможно, сделать при помощи бинарной матрицы:. .. Каждую строку матрицы , возможно, рассматривать как упорядоченные множества признаков преобразования формата с индексом в форматы с индексами , или в терминологии реляционной алгебры строка матрицы выражает возможность преобразования как «один ко многим». Каждый столбец матрицы выражает признаки возможности преобразования многих форматов с индексами , в один формат с индексом (для каждого столбца), или в терминологии реляционной алгебры столбцы матрицы выражают возможность преобразования как «многие к одному». Понятно, что на главной диагонали матрицы всегда будут единицы.. Зачастую, в практических задачах, ситуация обстоит следующим образом. Система получает множество данных, из определённого класса, в различных форматах. Обработка данных при этом производится в одном (или нескольких) заданном формате, причём часто этот формат не выходит за пределы информационной системы, т.е. является внутренним форматом. Поэтому для получения результатов обработки системе необходимо иметь возможность преобразования различных форматов в тот (или те) в которых производится обработка, что собственно относится к преобразованиям «многие к одному». Если информационная система производит экспорт данных, то скорее всего возникнет необходимость преобразования форматов данных как «многие к одному», так и «один ко многим».. Далее возможно перейти к расчёту частных информационных энтропий:. – «один ко многим»: ;. – «многие к одному»: .. Общая информационная энтропия .. Показатели , и предлагается рассматривать в качестве обобщённых оценок преобразования форматов.. Так же как и при оценке детектирования при оценке преобразования форматов в качестве дополнительного показателя оценки преобразования различных форматов, возможно, использовать визуализацию матрицы , с учётом признаков преобразования , используя матричные или трёхмерные диаграммы типа scatter plot.. Модели оценки технической интероперабельности, предлагаемые в данной работе, построены с применением методов теории информации, в частности для оценки детектирования и преобразования форматов использована информационная энтропия, соответственно сохраняющая все свойства энтропии известные из теории информации.. . Пример оценки технической интероперабельности программных микросистем. В качестве примера рассмотрим сравнительные оценки технической интероперабельности распространённых графических программных систем (Potoshop CS2 9.0, GIMP 2.8.14, Paint 5.1) работающих OS Windows, каждую из которых будем рассматривать как узел информационной системы требующей оценки качества технической интероперабельности (рис. 1). Расчёты производились в программной среде MatLab [32-34].. . Рисунок 1. Структура УИС взаимодействующих в информационной среде. В результате анализа читаемых и конвертируемых форматов, было выявлено, что все рассматриваемые программы практически позволяют выполнить взаимно-обратную конвертацию определённых форматов:. – мощность множества форматов Potoshop CS2 ;. – мощность множества форматов GIMP2 ;. – мощность множества форматов Paint 5 .. Мощность объёдинённого множества форматов . Следует отметить, что каждый из продуктов способен идентифицировать все форматы из множества , это обусловлено тем что данные системы являются системами с интерфейсом пользователя-человека (автоматизированными системами), т.е. идентификация форматов производится пользователем с помощью средств интерфейса программ. В автоматических системах распознавание форматов производится только машинными средствами, без вмешательства пользователя человека. Будем считать, что в данном случае, что уровень квалификации пользователя позволяет произвести детектирование форматов МИП.. Взаимное пересечение множеств форматов, возможно, определить на диаграмме Эйлера-Вена как (рис. 2). В частности, и , т.е. множество форматов не содержит формат применяёмый для создания графических меток-логотипов изображений в настольных и сетевых приложениях.. . Рисунок 2. Диаграмма отношений множеств форматов. Первичную оценку произведём исходя из предположения, что все форматы равновероятны в информационной среде. При мощности множества , вероятности в матрице главной диагонали будут равны , для .. В результате оценки технической интероперабельности получим равную для всех УИС энтропию, и общую энтропию .. Для каждого УИС, сформируем матрицы учитывая, что программные пакеты позволяют производить взаимные преобразования по всем форматам, т.е. обладают полнотой взаимных преобразований, получим, что при равных вероятностях взаимных преобразований и условии полного взаимного преобразования форматов, значения частных энтропий для определённого узла УИС «один ко многим» будут равны значениям «многие к одному»: при (рис. 3). Отметим, что для удобства всем форматам присвоены условные номера, порядок нумерации форматов не влияет на конечный результат – оценку, но определяет вид диаграммы распределения. Поэтому в случае использования характера диаграмм распределения для формирования эволюционной или сравнительной оценки необходимо сохранять выбранный ранее порядок соответствия номер-формат.. . Рисунок 3. Распределение частных энтропий. от номера формата для различных УИС. Полные энтропии УИС имеют следующие значения:. – Potoshop CS2 ;. – GIMP2 ;. – Paint 5 .. Таким образом, уровень технической интероперабельности графической системы GIMP2 почти в три раза превосходит уровень технической интероперабельности системы Potoshop CS2 и почти в 40 раз уровень технической интероперабельности графического редактора Paint 5.. Далее произведём оценку технической интероперабельности тех же УИС с учётом фактической вероятности присутствия форматов в информационном обмене. В качестве общего МИП использовалась файловая база данных, хранящая 175127 файлов с графической и смежной информацией в различных форматах. Из 49, используемых программными пакетами форматов, в МИП выявлено 10 форматов данных, так же выявлено 17 форматов не преобразуемых программными пакетами (рис.4).. . Рисунок 4. Общее распределение форматов в логарифмическом масштабе, присутствующих в МИП, форматы разделены на две группы: преобразуемые УИС и не преобразуемые УИС. . Далее есть смысл исключить из рассмотрения форматы, которые отсутствуют в МИП, а значит по условию, что обмен информацией между УИС производится через МИП, отсутствуют и в информационном обмене. В результате получим 27 рассматриваемых форматов данных. Далее, учитывая, что детектирование формата производится в автоматизированном режиме, т.е. с участием человека-оператора, рассчитаем информационную энтропию, характеризующую техническую интероперабельность каждого УИС. Полученное значение равное для всех УИС, и общую энтропию . Очевидно, что по сравнению с предыдущими расчётами, значение энтропии снизилось, что является следствием использования реальной плотности распределения форматов (а не равномерной предполагаемой) и уменьшения числа форматов до фактически присутствующих в МИП.. Следующим действием произведём оценку преобразования форматов, учитывая, что рассматриваемые УИП обладают свойством полноты преобразования форматов то: , при (рис. 5).. . . Рисунок 5. Распределение по форматам частных энтропий. «один ко многим» и «многие к одному» для различных УИС. . Полные энтропии УИС имеют следующие значения:. – Potoshop CS2 ;. – GIMP2 ;. – Paint 5 .. По сравнению с расчётами без учёта реального распределения форматов в МИП заметно, что существенно изменились абсолютные значения полных энтропий и величины их взаимных отношений, при этом общая тенденция «больше-меньше» сохранилась. Наибольшая величина информационной энтропии, характеризующая преобразование форматов, осталась у программного пакета GIMP2, затем Potoshop CS2 и на последнем месте Paint 5.. . Заключение. В работе выполнен анализ методов оценки интероперабельности в соответствии с эталонной моделью по ГОСТ Р 55062–2012 и методов технической совместимости узлов информационных систем. Выявлено отсутствие практического инструментария количественной оценки технической интероперабельности. В соответствии с целью исследований произведена разработка метода численной оценки детектирования и преобразования форматов данных, что в большей степени относится к уровню технической интероперабельности. Предложенный в работе метод базируется на основах теории информации.. Авторы не претендуют, на то что предлагаемый метод предоставляет возможность произвести оценку технической интероперабельности в полном объёме, в соответствии с эталонной моделью по ГОСТ Р 55062–2012, при этом предлагаемый метод предоставляет исследователю готовый математический аппарат оценки детектирования и преобразования форматов, что относится к технической интероперабельности..