УДК 004.832.3

Е.Н. МАЩЕНКО, В.И. ШЕВЧЕНКО

Севастопольский национальный технический университет, Украина

ИССЛЕДОВАНИЕ КРИТИЧЕСКИХ СИТУАЦИЙ В ИТ-ИНФРАСТРУКТУРАХ МЕТОДАМИ КЛАСТЕРНОГО АНАЛИЗА

Рассматривается задача исследования критических ситуаций в ИТ-инфраструктурах методами кластерного анализа. Исследуются информационные системы с одинаковой структурой метаданных и сравнимыми значениями характеристик качества, с целью выделения групп критичности. Предлагается способ нормирования классификационных признаков при предварительном анализе данных. По результатам кластеризации выделено три группы ИС с разными уровнями критичности: ИС с малым количеством ошибок и предупреждений об ошибках; ИС с малым количеством ошибок и большим количеством предупреждений об ошибках; ИС с большим количеством ошибок и большим количеством предупреждений об ошибках.

Ключевые слова: *критическая инфраструктура, информационная система, кластерный анализ, соглашение об уровне качества.*

Введение

В условиях развития информационных технологий (ИТ), проблемы автоматизации критических инфраструктур во многом связаны с резким увеличением объема обрабатываемой информации, необходимой для принятия решений управленческого характера. Скорость принятия и обоснованность этих решений во многом определяют критичность процессов управления и напрямую зависят от качества функционирования ИТ-сервисов.

Особенно актуальной эта проблема является для информационных систем, обеспечивающих функционирование критических систем управления (КСУ), таких как: системы управления опасными производствами и объектами атомной энергетики; системы управления космическими полетами, воздушным или железнодорожным движением; системы управления военного назначения; системы управления органов государственной власти; банковские и экономические информационные системы.

Для организации надежной и эффективной работы критических ИТ-инфраструктур необходим мониторинг, анализ и адаптивное управление уровнем ключевых показателей эффективности (Key Performance Indicators - KPI), зафиксированных в соглашении об уровне обслуживания (Service Level Agreement - SLA).

Обеспечение выполнения жестких соглашений об уровне обслуживания, как правило, гарантируется избыточным количеством ИТ-ресурсов в расчете на наихудшие, пиковые нагрузки, однако применение такого подхода приводит к построению негибких, дорогих в обслуживании ИТ-инфраструктур. Альтернативным направлением развития служб поддержки ИТ-сервисов критических систем является концепция управления ИТ-службами (IT Service Management, ITSM) [1, 3], которая предлагаетподход к организации функционирования ИТ-подразделений, основанный на базе «эталонных» моделей и принципов, изложенных в Библиотеке передового опыта в области управления информационными технологиями (IT Infrastructure Library, ITIL).

В рамках методологии ITIL и стандарта ISO 20000 для обеспечения контроля качества ИТ-сервисов, применяется особая система показателей качества [2]. В данной работе проведено исследование подмножества показателей качества, непосредственно связанных с надежностью и безопасностью критических систем.

1.Постановка задачи

Объектом исследования являются информационные системы с одинаковой структурой метаданных и сравнимыми значениями характеристик качества. Цель исследования – используя методы кластерного анализа, определить группы ИС, близких по критичности (уровню риска нарушения SLA, уровню количества нарушений SLA).

Формальная постановка задачи кластеризации критических ИТ-инфраструктур следующая:

Заданы: G – множество ИТ-инфраструктур (объектов кластеризации); M - множество номеров кластеров в интервале 1..m; d - функция расстояния между объектами (метрика); X - конечная обучающая выборка объектов; S - способ нормирования исходных данных; A – алгоритм кластеризации; W - критерий оптимальности разбиения.

Требуется на основании данных, содержащихся в выборке X, разбить множество объектов G на m (m – целое) непересекающихся подмножеств (кластеров) Q_1, Q_2, …, Q_m, так, чтобы объекты разных кластеров существенно отличались по уровню критичности.

Используемый критерий оптимальности разбиения – внутригрупповая сумма квадратов отклонения

$$W = \sum_{j=1}^{n} (x_j - \overline{x})^2 = \sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2 .$$

2. Предварительный анализ данных

Данные для анализа исследуемых информационных систем выбраны из журналов регистрации событий (логов) за 2010 год и представляют собой следующие критерии качества [2]: количество ошибок (рис. 1) и количество предупреждений об ошибках (рис. 2), которые являются также классификационными признаками для кластерного анализа.

Были построены проекции значений критериев качества ИС по месяцам в двумерном пространстве признаков.

Из рис. 3 видно, что исследуемые ИС четко подразделяются на подмножества-кластера по уровню критичности.

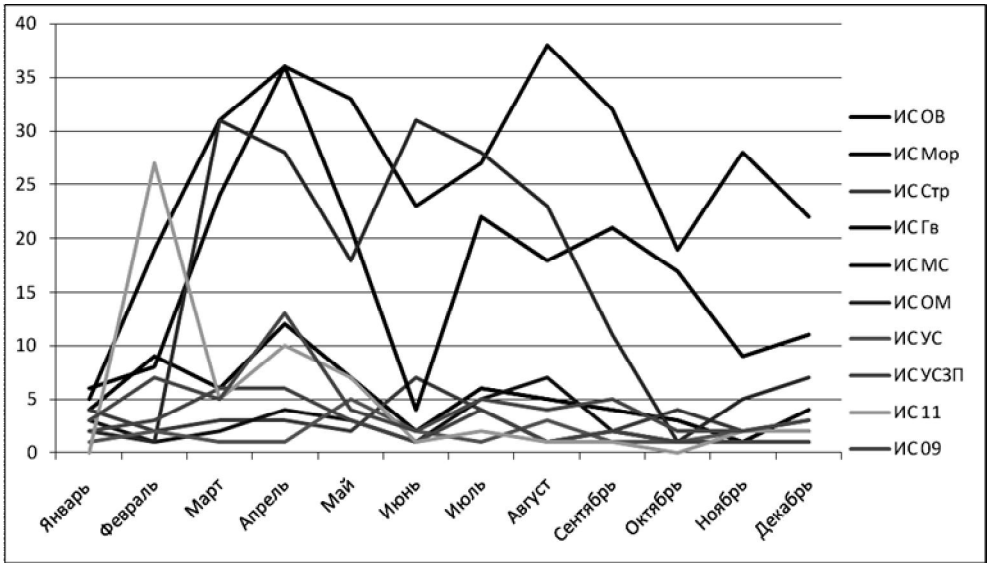


Рис. 1. Количество ошибочных ситуаций в исследуемых ИС за 2010 год

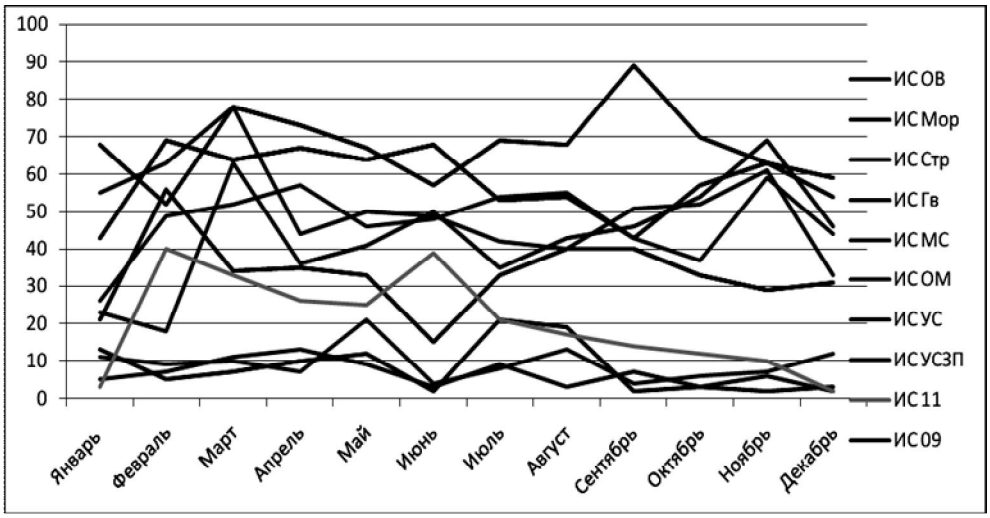


Рис. 2. Количество предупреждений об ошибках за 2010 год

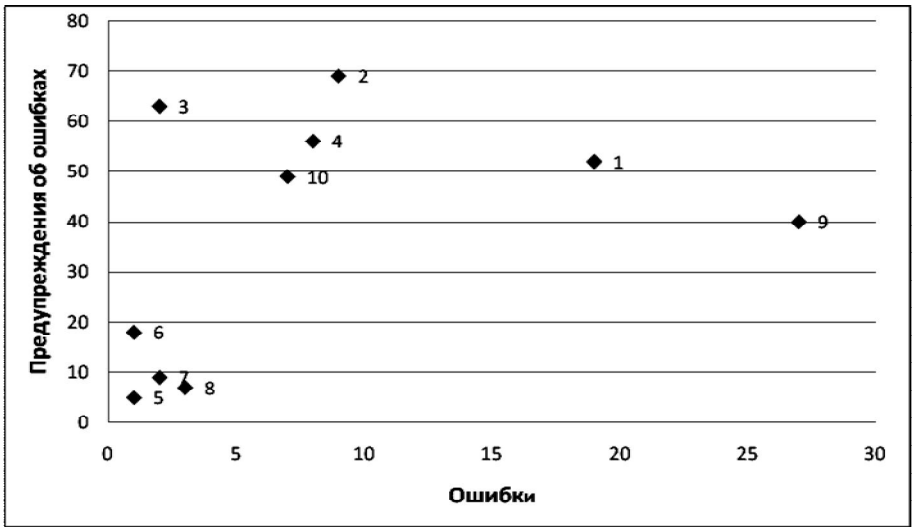


Рис. 3. Значения критериев качества ИС за февраль 2010 года

Соответствие точек в пространстве признаков и реальных ИС приведено в табл. 1

Таблица 1

Соответствие точек
в пространстве признаков и реальных ИС

ИС	ИС ОВ	ИС Стр	ИС Мор	ИС Гв	ИС ОМ
№ точки	1	2	3	4	5
ИС	ИС ОВ	ИС Стр	ИС Мор	ИС Гв	ИС ОМ
№ точки	6	7	8	9	10

Одной из основных проблем предварительного анализа данных при кластеризации является проблема сравнимости шкал классификационных признаков. Эта проблема решается с помощью процесса стандартизации (standardization) или нормирования (normalization).

Существуют следующие способы нормирования [4]:

1) деление исходных данных на некоторые эталонные (нормативные) значения;

2) вычисление Z-вклада или стандартизованного вклада

$$Z_i = \frac{x_i - \bar{x}}{S},$$

где x_i – значение данного наблюдения, x – среднее, S – стандартное отклонение. Результаты нормирования приведены на рис. 4, а, б.

Как видно из рис. 4, исследованные способы нормирования данных не оказывают влияния на характер пространства признаков. Для последующей кластеризации будем использовать способ нормирования делением исходных данных на некоторые эталонные (нормативные) значения, так как эти значения определены предельными значениями критериев качества в соглашениях об уровне обслуживания, а близость точки в нормированном пространстве признаков к единице означает близость критической ситуации в ИС к предельно допустимому уровню.

По результатам предварительного анализа данных можно выделить три кластера ИС с разными уровнями критичности:

1) ИС с малым количеством ошибок и предупреждений об ошибках (точки 5, 6, 7, 8);

2) ИС с малым количеством ошибок и большим количеством предупреждений об ошибках (точки 2, 3, 4, 10); 3) ИС с большим количеством ошибок и большим количеством предупреждений об ошибках (точки 1, 9).

3. Кластерный анализ

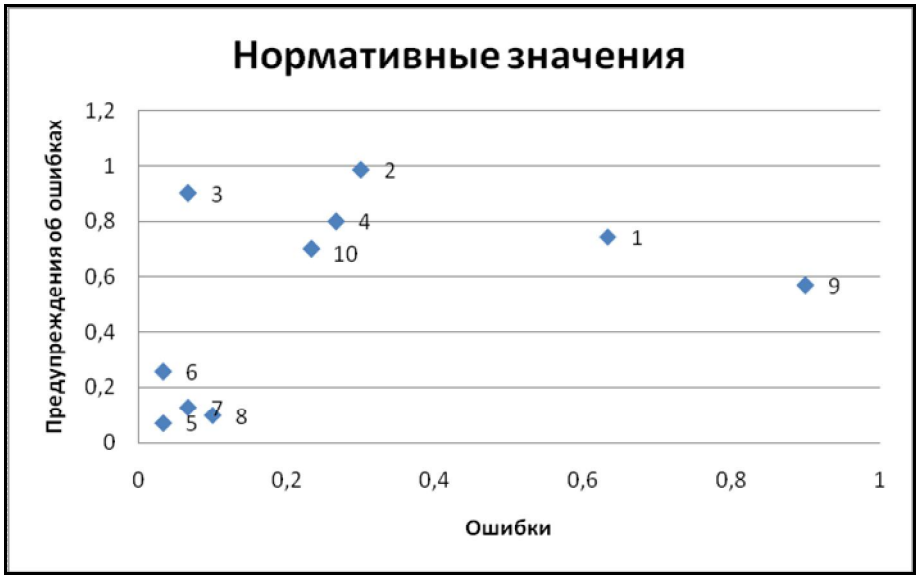
Кластерный анализ проводился в среде STATISTICA с использованием следующих алгоритмов:

1) иерархическая кластеризация с разными типами расстояний – евклидово расстояние, манхеттенское расстояние (расстояние городских кварталов), расстояние Чебышева (рис. 5);

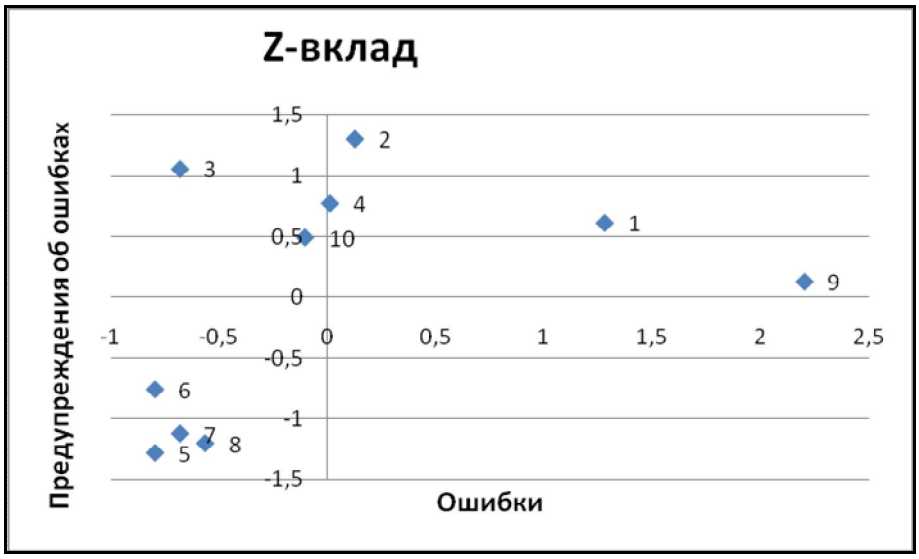
2) кластеризация методом К-средних (табл. 2, рис. 6).

По результатам кластеризации иерархическим методом два из трех способов расчета расстояний дают одинаковый результат деления на кластеры, совпадающий с результатом предварительного анализа данных. Метод К-средних также подтверждает эту классификацию.

Следовательно, по эмпирическому правилу проверки достоверности кластерного решения, а именно если сравниваемые классификации групп, полученные разными методами, имеют долю совпадений более 70 % (более 2/3 совпадений), то кластерное решение принимается.



а



б

Рис. 4. Нормированные данные

Выводы

По результатам кластеризации выделено три группы ИС с разными уровнями критичности:

1) ИС с малым количеством ошибок и предупреждений об ошибках (точки 5, 6, 7, 8);

2) ИС с малым количеством ошибок и большим количеством предупреждений об ошибках (точки 2, 3, 4, 10);

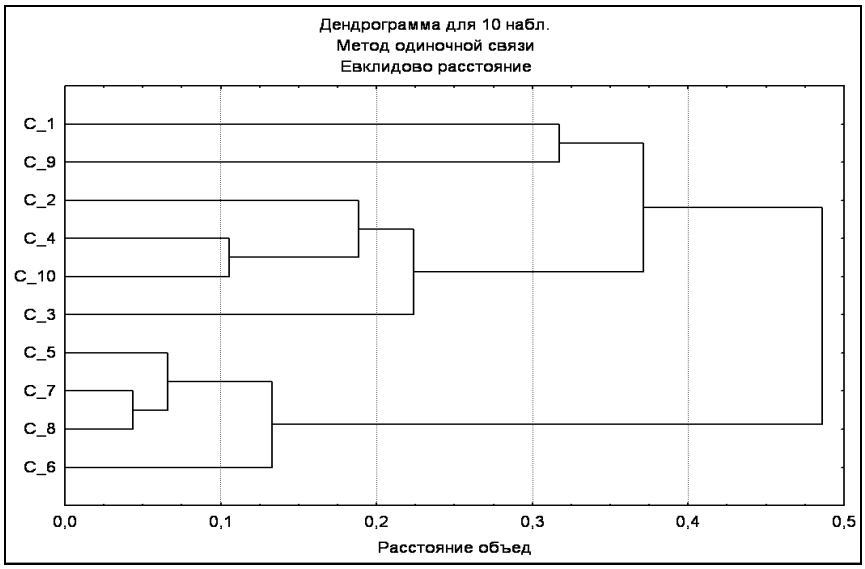
3) ИС с большим количеством ошибок и большим количеством предупреждений об ошибках (точки 1, 9).

Планируется продолжение исследований с целью выявления групп критичности ИС по критериям управления доступностью сервисов и параметрам информационных рисков в рамках методологии ITIL.

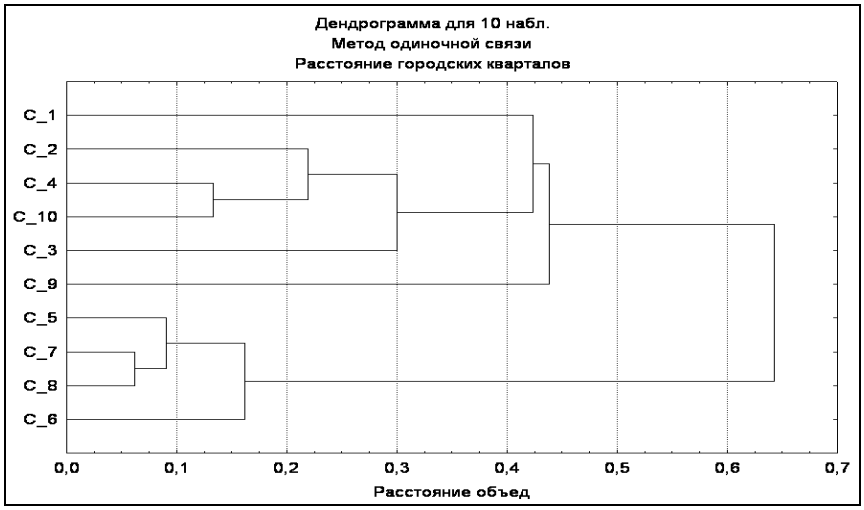
Таблица 2

Результаты кластеризации методом К- средних

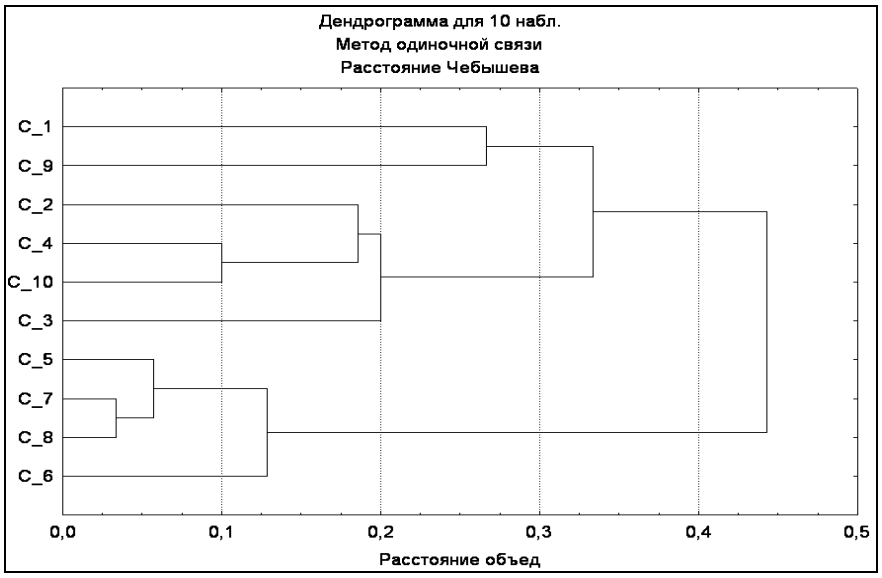
№ кластера	1	2	3
Элементы кластера	5, 6, 7, 8	2, 3, 4, 10	1,9



а

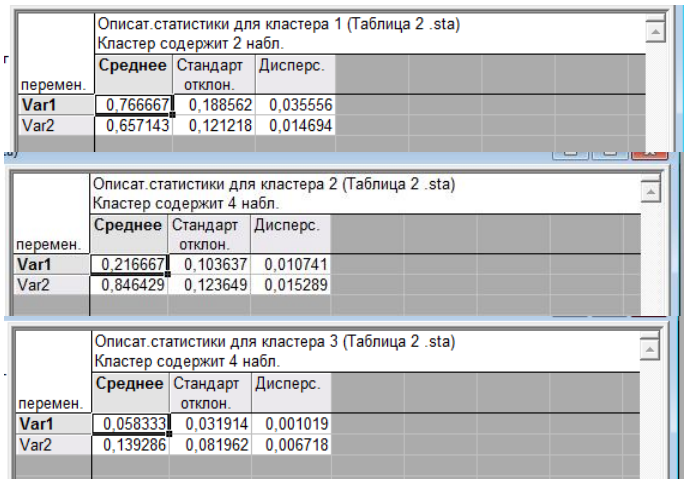


б



в

Рис. 5. Результаты кластеризации иерархическим методом



Описат.статистики для кластера 1 (Таблица 2 .sta)
Кластер содержит 2 набл.

перемен.	Среднее	Стандарт отклон.	Дисперс.
Var1	0,766667	0,188562	0,035556
Var2	0,657143	0,121218	0,014694

Описат.статистики для кластера 2 (Таблица 2 .sta)
Кластер содержит 4 набл.

перемен.	Среднее	Стандарт отклон.	Дисперс.
Var1	0,216667	0,103637	0,010741
Var2	0,846429	0,123649	0,015289

Описат.статистики для кластера 3 (Таблица 2 .sta)
Кластер содержит 4 набл.

перемен.	Среднее	Стандарт отклон.	Дисперс.
Var1	0,058333	0,031914	0,001019
Var2	0,139286	0,081962	0,006718

Рис. 6. Результаты кластеризации методом К-средних в среде STATISTICA.

Литература

1. Баранов, В.В. Процессы принятия управленческих решений, мотивированных интересами [Текст] / В.В. Баранов. – М: ФИЗМАТЛИТ, 2005. – 296 с.

2. Брукс, П. Метрики для управления ИТ-услугами [Текст]: пер. с англ. / Питер Брукс. – М: Альпина Бизнес Букс, 2008. – 283 с.

3. Ян, В.Б. ИТ Сервис-менеджмент, введение [Текст] / Я.В. Бон, Г. Кеммерлинг, Д. Пондаман; под ред. М.Ю. Потоцкого (русская версия). – М: IT Expert, 2003. – 215 с.

4. Мандель, И.Д. Кластерный анализ [Текст] / И.Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с.

Поступила в редакцию 1.03.2012

Рецензент: д-р техн. наук, проф., зав. каф. Информационных управляющих систем О.Е. Федорович, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ» , Харьков, Украина

ДОСЛІДЖЕННЯ КРИТИЧНИХ СИТУАЦІЙ В ІТ-ІНФРАСТРУКТУРАХ МЕТОДАМИ КЛАСТЕРНОГО АНАЛІЗУ

О.М. Мащенко, В.І. Шевченко

Розглядається задача дослідження критичних ситуацій в ІТ-інфраструктурах методами кластерного аналізу. Досліджуються інформаційні системи з однаковою структурою метаданих і порівнянними значеннями характеристик якості, з метою виділення груп критичності. Пропонується спосіб нормування класифікаційних ознак при попередньому аналізі даних. За наслідками кластеризації виділено три групи ІС з різними рівнями критичності: ІС з малою кількістю помилок і попереджень про помилки; ІС з малою кількістю помилок і великою кількістю попереджень про помилки; ІС з великою кількістю помилок і великою кількістю попереджень про помилки.

Ключові слова: критична інфраструктура, інформаційна система, кластерний аналіз, угода про рівень якості.

RESEARCH OF CRITICAL SITUATIONS IN IT- INFRASTRUCTURES OF CLUSTER ANALYSIS METHODS

E.N. Maschenko ,V.I. Shevchenko

Problem of investigating the critical situation in the IT-infrastructures was considered. Information systems with the same structure of metadata and comparable values of quality characteristics was study. Valuation method of classification features a preliminary analysis of data was proposed. On results a clusterization three groups are selected IC with the different levels of criticism: IC with a few of errors and warnings of errors; IC with a few of errors and plenty of warnings of errors; IC with plenty of errors and plenty of warnings of errors.

Key words: critical infrastructure, information systems, cluster analysis, Service Level Agreement.

Мащенко Елена Николаевна – канд. техн. наук, доцент кафедры кибернетики и вычислительной техники Севастопольского национального технического университета, Севастополь, Украина.

Шевченко Виктория Игоревна – ст. преподаватель кафедры кибернетики и вычислительной техники Севастопольского национального технического университета, Севастополь, Украина.