

УДК 004.4:378

Мельников А. Ю., Баган С. В.

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ФОРМИРОВАНИЯ НОВЫХ МЕТОДОВ ПРЕДСТАВЛЕНИЯ ДАННЫХ В ЧЕТЫРЕХ И БОЛЕЕ ИЗМЕРЕНИЯХ

Анализ информации – это довольно трудный и трудоемкий процесс. Когда информация избыточна и получена с многих источников, ее довольно тяжело анализировать, сравнивать и строить любые заключения. Поэтому методы представления данных в нескольких измерениях могут существенным образом упростить процесс анализа и систематизации информации, визуализируя ее.

Визуализация – это процесс представления данных в виде изображения с целью максимального удобства их понимания; предоставление зримой формы любому мыслимому объекту, субъекту, процесса и т. д. Существует большое разнообразие способов визуального представления данных. От выбора способа визуализации зависит процесс понимания информации пользователем.

Данные, которые представлены в четырех и более измерениях, необходимо либо преобразовывать к трехмерному пространству, либо использовать специальные методы: например, «лица Чернова», базирующиеся на концепции кодировании значений различных переменных в характеристиках человеческого лица; лепестковые диаграммы в виде круга, отображающего данные с помощью углов; диаграммы с параллельными координатами, где каждая из осей отображает значения по выбранному показателю [1].

Каждый из методов имеет свой ареал применения, разработаны приложения для сравнения методов и выбора лучшего при визуализации конкретных данных [2–5]. В то же время ни одно из существующих программных средств не позволяет пользователю самому создать визуализатор согласно собственным предпочтениям.

Была поставлена задача проектирования системы – приложения, позволяющего пользователю при помощи графических примитивов создать некий рисунок (схему) и определить его параметры (для измерений). При этом рисунок-схема со всеми описаниями должен сохраняться в специальном файле, а потом использоваться для визуализации данных.

Объектом исследования являются методы представления данных в четырех и более измерениях. Предмет исследования – создание новых визуализаторов для отображения таких данных.

Многомерные пиктографики – не очень простой, но мощный исследовательский инструмент разведывательного анализа данных. Главная идея такого метода анализа основана на человеческой способности «автоматически» фиксировать сложные связи между многими сменными, если они проявляются в последовательности элементов. С помощью пиктографиков можно представить элементарные наблюдения отдельных элементов графическими объектами, где значения переменных отвечают определенным чертам или размерами объекта [6–8].

Американский математик Герман Чернов в 1973 году опубликовал работу, в которой изложил концепцию использования способности восприятия лица человека для построения пиктографиков. Их применяют, как правило, в двух случаях:

- когда нужно проявить характерные зависимости или группы наблюдений;
- когда необходимо исследовать сложные взаимосвязи между несколькими переменными.

Способ представления многомерных данных в «лицах Чернова» разрешает показать информацию, например, о любом экономическом объекте, в виде «лица». При этом черты

«лица» демонстрируют разные значения параметров данного объекта. «Лицо» представляет собой массив из 18 элементов, каждый из которых принимает значение от 0 до 1. Значению отвечает внешний вид соответствующей части «лица». Параметры исследуемых объектов приводятся к этим значениям. Экстремумы реальных данных будут приняты как 0 и 1, все другие значения лежат в промежутке от 0 до 1. В 1981 году Бернхард Флури и Ганс Ридвил (Bernhard Flury and Hans Riedwyl) улучшили концепцию и прибавили лицам Чернова асимметрию, которое разрешило увеличить вдвое количество переменных (до 36).

Таким образом, проектируемая система должна предоставить пользователю-исследователю данных такие же возможности визуализации, как любое приложение, использующее метод лиц Чернова, но при этом разрешить добавление собственного метода представления данных – то есть создать новый визуализатор.

Информационная модель системы была создана на языке визуального моделирования UML [9]. Функциональные возможности системы представлены в виде диаграммы вариантов использования (рис. 1). Структура информационной системы представлена на диаграмме классов (рис. 2). Типовая работа пользователя представлена на диаграмме состояний (рис. 3).

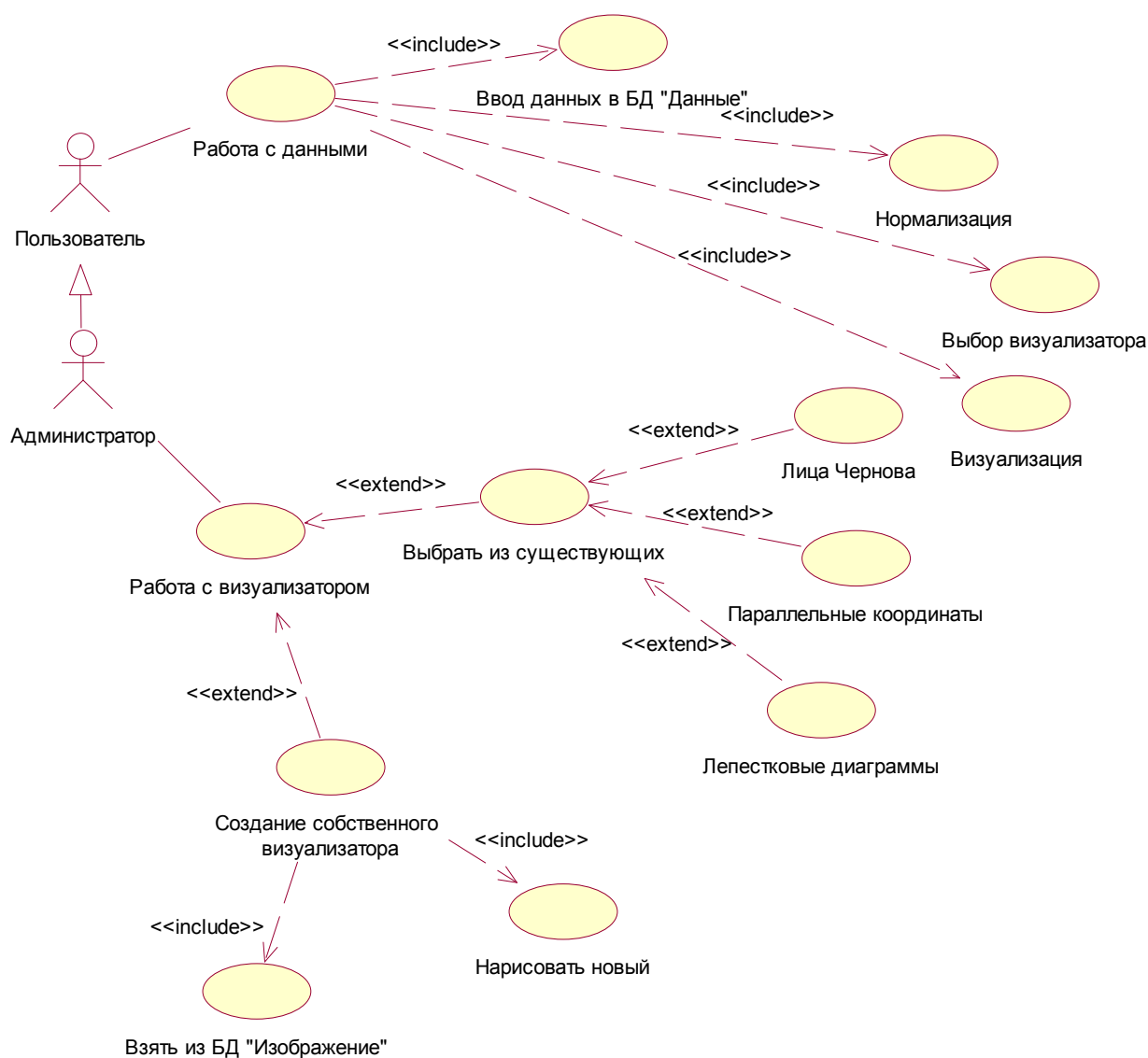


Рис. 1. Диаграмма вариантов использования информационной системы для представления многомерных данных

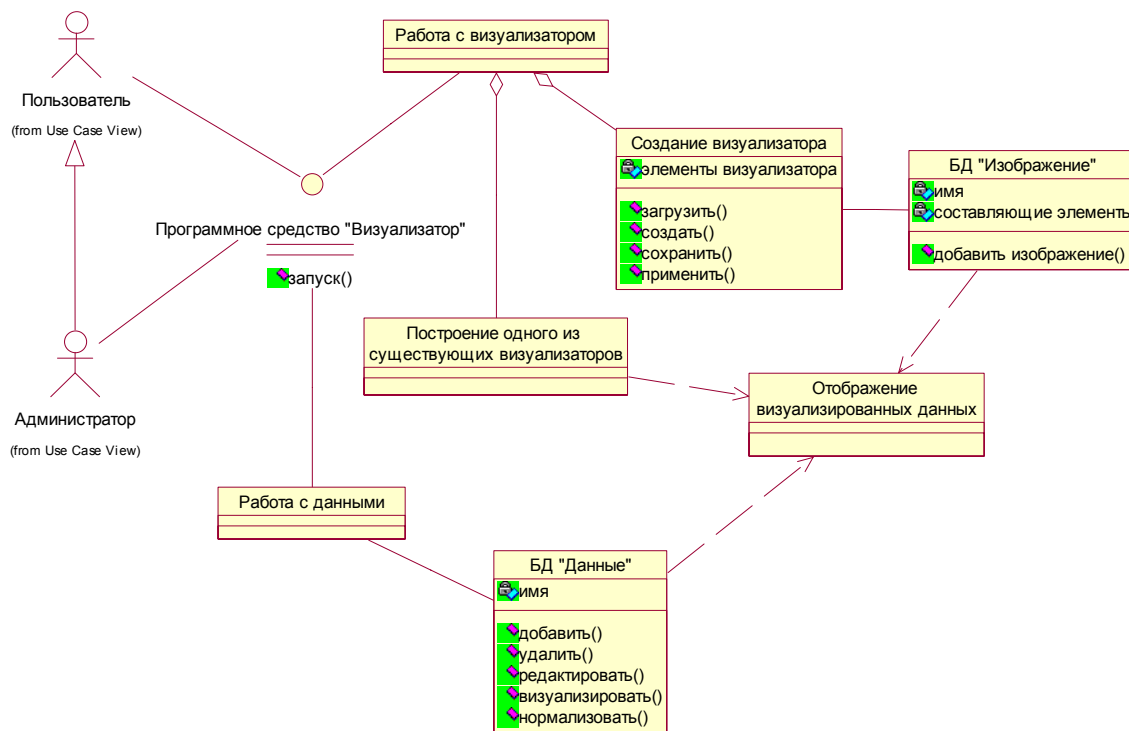


Рис. 3. Диаграмма классов информационной системы для представления многомерных данных

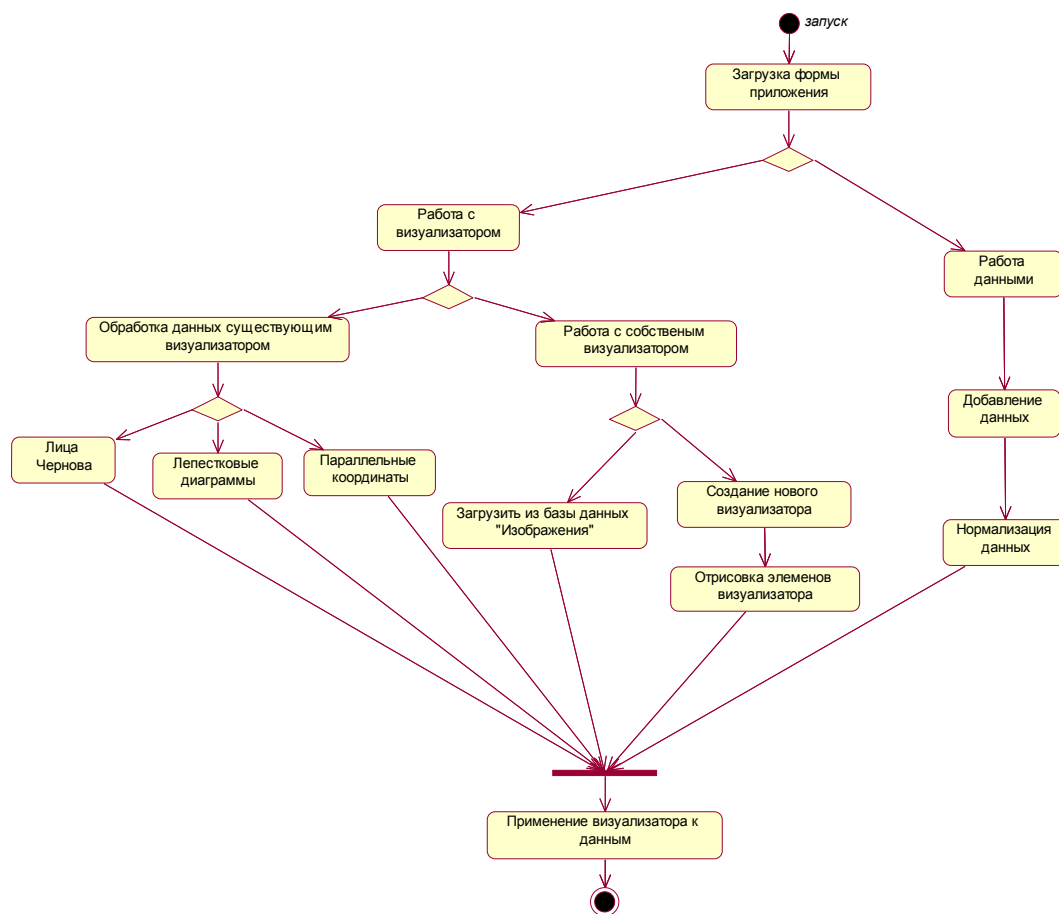


Рис. 4. Диаграмма состояний информационной системы для представления многомерных данных

В настоящее время осуществляется работа по реализации модели в среде визуального программирования. Созданное приложение позволяет:

- работать с данными (рис. 5): загрузить (импортировать из редактора электронных таблиц, при этом автоматически рассчитывается число измерений), нормализовать (рис. 6), сохранить;
- работать с визуализаторами;
- строить многомерные диаграммы.

Работа с визуализаторами, в свою очередь, предполагает или использование стандартных методов (рис. 7–9), или создание собственных визуализаторов (рис. 10).

Наименование	Территория	% водной поверхности	Население	Плотность	ВВП (на душу населения)	ИЧР00	Кол-во тер
Украина	603549	7	42,46546	73,92	8665	0,747	27
Франция	647685	0,26	66,736	116	43000	0,884	27
Великобритания	243809	1,34	63,395574	246	34919	0,892	4
Испания	504782	1,04	47,370542	91,45	33711	0,885	19
Италия	301340	2,4	60,795612	201,1	35811	0,872	20
Россия	17125191	4,22	146,54471	8,56	25411	0,798	85
Швейцария	41284	4,2	7,996026	188	81000	0,917	26
США	9519431	6,76	325,31027	32	49922	0,914	51

Рис. 5. Вкладка «Данные»

Наименование	Территория	% водной поверхности	Население	Плотность	ВВП (на душу населения)	ИЧР00	Кол-во тер
Украина	101	198	109	126	98	100	126
Франция	102	98	117	143	145	179	126
Великобритания	99	114	115	198	134	184	98
Испания	101	110	110	133	133	180	117
Италия	100	130	115	179	136	172	118
Россия	198	157	142	98	121	129	198
Швейцария	98	156	98	174	198	198	125
США	153	194	198	108	155	196	156

Рис. 6. Вкладка «Данные» после нормализации

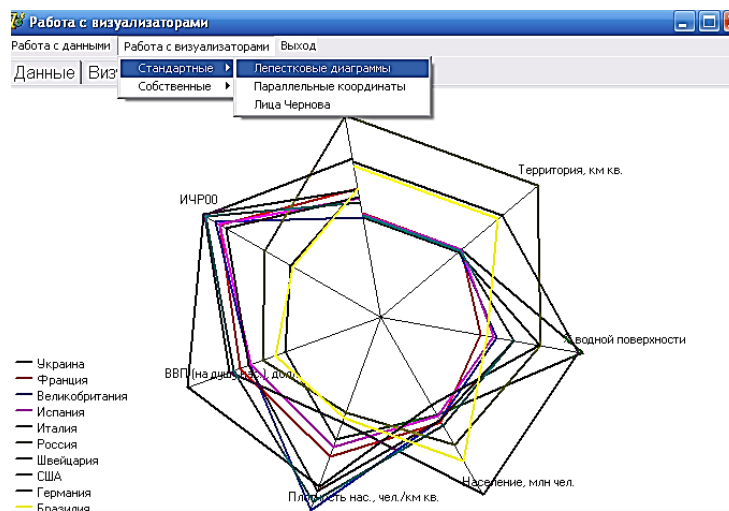


Рис. 7. Лепестковые диаграммы

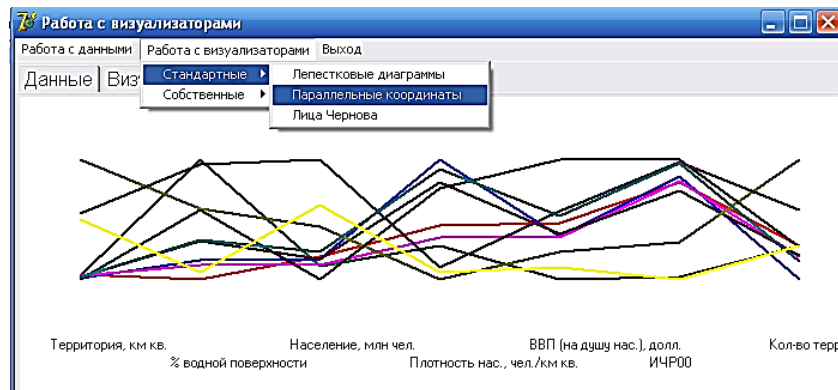


Рис. 8. Параллельные координаты

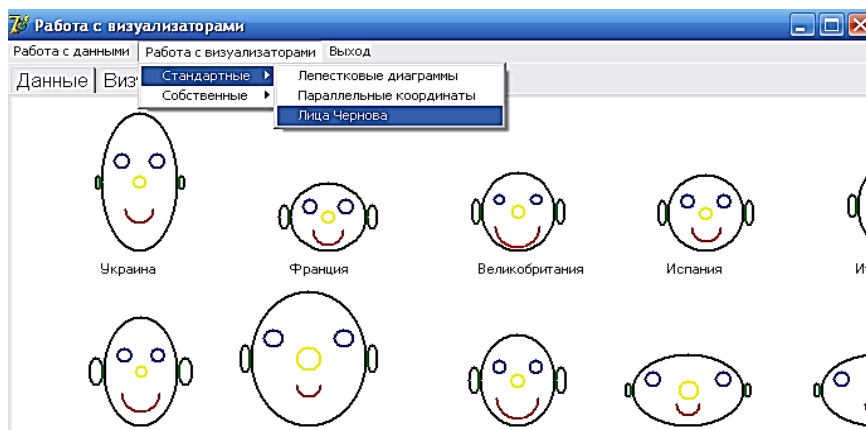


Рис. 9. Лица Чернова

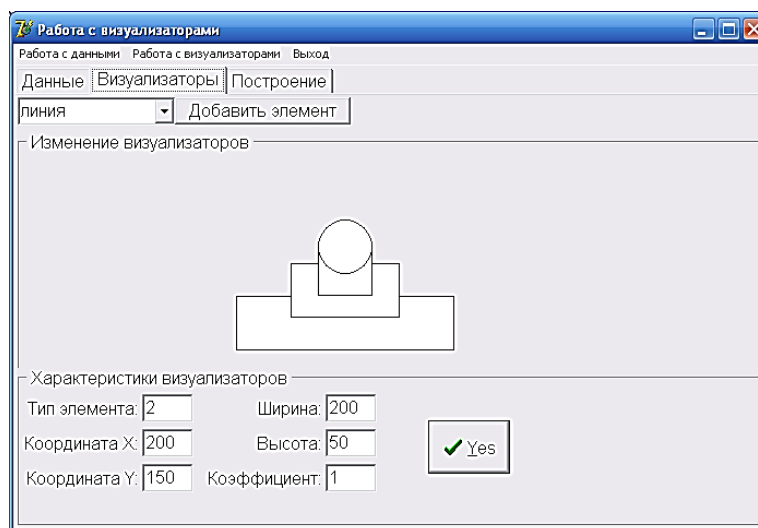


Рис. 10. Вкладка «Визуализаторы»

Пример визуализатора, показанный на рис. 10, содержит 4 элемента: 3 элемента типа «прямоугольник» и 1 элемент типа «круг». Пользователь имеет возможность изменять положение каждого элемента, а также параметр «коэффициент», который показывает, во сколько раз максимальное нормализованное значение будет больше минимального (по умолчанию равен единице). При использовании такого визуализатора для ранее представленных данных будут применены 7 измерений: для каждого прямоугольника – ширина и высота, для круга – радиус.

Далее необходимо сопоставить каждому элементу показатель (рис. 11).

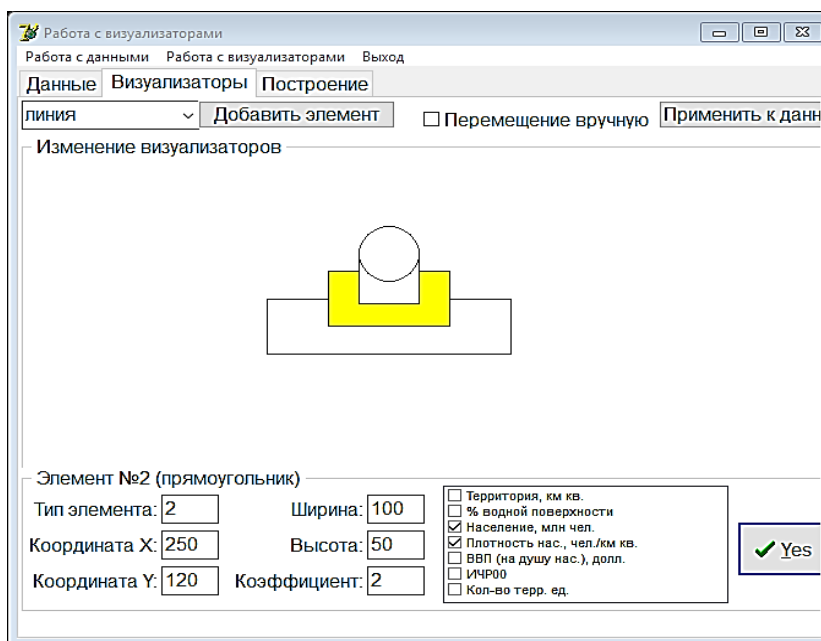


Рис. 11. Применение визуализатора

ВЫВОДЫ

В статье представлено описание информационной системы, которая позволяет работать с многомерными данными как при помощи распространенных методов, так и путем создания собственного визуализатора. Приведен пример работы приложения.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Чубукова И. А. *Data Mining: Учебное пособие* / И. А. Чубукова. – М. : Интернет-Университет Информационных Технологий ; БИНОМ. Лаборатория знаний, 2006. – 382 с.
2. Мельников А. Ю. Разработка приложения для сравнения изображения многомерных данных тремя основными методами визуализации / А. Ю. Мельников, С. В. Баган // *Молодежь в науке: Новые аргументы: Сборник научных работ VI-го Международного молодежного конкурса (Россия, г. Липецк, 30 апреля 2017 г.). Часть I / Отв. ред. А. В. Горбенко.* – Липецк : Научное партнерство «Аргумент», 2017. – С. 115–117.
3. Мельников А. Ю. Использование вспомогательного приложения для выбора лучшего метода визуализации многомерных данных / А. Ю. Мельников, С. В. Баган // *Сучасна освіта та інтеграційні процеси: збірник наукових праць міжнародної науково-методичної конференції, 22–23 листопада 2017 року, м. Краматорськ / Під заг. ред. С. В. Ковалевського, д-ра техн. наук., проф.* – Краматорськ : ДГМА, 2017. – С. 125–126. – ISBN 978-966-379-817-2.
4. Мельников А. Ю. Проектирование приложения для работы с визуализаторами представления многомерных данных / А. Ю. Мельников, С. В. Баган // *Молодежь в науке: Новые аргументы : Сборник научных работ VIII-го Международного молодежного конкурса (Россия, г. Липецк, 30 марта 2018 г.). Часть I / Отв. ред. А. В. Горбенко.* – Липецк : Научное партнерство «Аргумент», 2018. – С. 63–65.
5. Melnykov O. Yu. *About the task of developing an application for creating multidimensional data view visualizers* / O. Yu. Melnykov, S. V. Bagan // *Матеріали міжнародної науково-практичної конференції «Universum View 6».* – Вінниця: ТОВ «Нілан-ЛТД», 2018. – С. 52–54.
6. Лица Чернова [Электронный ресурс]. – Режим доступа: http://neural.forekc.ru/dm2/index_lisa_chernova_2.htm
7. Кабулов Б. Т. *Метод построения лиц Чернова, ориентированный на интервальные оценки параметров* / Б. Т. Кабулов. – Томск : Техническая кибернетика, 1991. – 250 с.
8. Наука и технологии. *Графические методы анализа данных [Электронный ресурс].* – Режим доступа: <http://www.ligis.ru/effects/stat/modules/stgraph.html>
9. Мельников А. Ю. *Объектно-ориентированный анализ и проектирование информационных систем: учебное пособие* / А. Ю. Мельников. – Изд. 2-е, перераб. и доп. – Краматорск : ДГМА, 2013. – 172 с.