

Евсюков А.А.

Федеральный исследовательский центр "Красноярский научный центр СО РАН", Красноярск,
Россия

ИНФОГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ МНОГОМЕРНЫХ ДАННЫХ

АННОТАЦИЯ

Статья посвящена применению инфографических средств для анализа многомерных данных. Выполнен обзор существующих подходов и инструментов построения инфографики, определены виды инфограмм, проанализированы предпосылки интеграции методов инфографического моделирования и технологии оперативной аналитической обработки данных.

КЛЮЧЕВЫЕ СЛОВА

Инфографика; инфограмма; OLAP; представление многомерных данных.

Evsyukov A.A.

Federal Research Centre "Krasnoyarsk Science Centre SB RAS", Krasnoyarsk, Russia

INFOGRAPHIC REPRESENTATION OF MULTIDIMENSIONAL DATA

ABSTRACT

The article is devoted to infographic application for multidimensional data representation. The survey of existing approaches and tools for infographics and the types of infogram is described. The integrating methods of infographic simulation and online analytical processing technology are observed.

KEYWORDS

Infographics; infogram; OLAP; multidimensional data representation.

В задачах поддержки принятия управленческих решений широкое распространение получила технология оперативной аналитической обработки данных OLAP (On-Line Analytical Processing). Технология обеспечивает наглядное представление многомерных данных при высокой скорости выполнения аналитических операций. Основной идеей технологии OLAP является представление анализируемой информации в виде многомерного куба. Специалист-аналитик представляет анализируемые факты в разрезе множества шкал – измерений. Например, показатели экономического состояния населения естественно рассматривать в разрезе возрастного состава, территориальной привязки, профессиональной занятости и т.д. Многомерное представление позволяет наиболее точно оценить состояние анализируемой области, помогает принять обоснованное решение.

Преимуществом OLAP-подхода является то, что аналитик получает не жестко регламентированный отчет, а инструментарий для творческого исследования задачи. Возможность свободного управления данными упрощает получение необходимых наборов данных. Результаты OLAP-запроса представляются в виде двумерной перекрестной таблицы (кросс-таблицы) со сложными иерархическими шапками, для которой доступны операции группировки, поворота, фильтрации и агрегации данных. Важным преимуществом OLAP является использование специалистом не сложных SQL-запросов, а знакомых терминов из предметной области.

С момента возникновения в 1993 году OLAP-технология проделала огромный путь развития. В настоящее время можно выделить два наиболее значимых направления развития технологии. В первую очередь, это увеличение объемов и повышение скорости обработки анализируемой информации. Например, реализация OLAP-моделей на больших данных (big data). Другое важное направление при использовании OLAP, в котором могут быть достигнуты значительные результаты, – увеличение наглядности представления результатов аналитической обработки для стимуляции когнитивных (познавательных) возможностей лиц, принимающих решения [1; 2; 3].

Традиционно технология OLAP использовала интерактивные инструменты представления результатов, такие как кросс-таблицы, диаграммы, а в случае пространственного распределения

данных – картограммы. Если не рассматривать табличное представление данных (кросс-таблицы), то существующие средства графического отображения многомерных данных (диаграммы и картограммы) отображают срезы, где изменяются значения не более двух измерений при фиксации значений остальных измерений. Способ представления результатов аналитической обработки во многом определяет когнитивную эффективность оперативного анализа [4; 5; 6]. Применение принципов инфографического моделирования [7] для стимулирования творческой активности лица принимающего решения позволит максимально полно использовать аналитический потенциал OLAP-средств. Актуальна задача разработки аналитических инфографических моделей для отображения многомерных результатов оперативного анализа с учетом состава и характера данных.

Использование инфографики для представления информации

Инфографика – графический способ представления информации, упрощающий ее восприятие. Помогает интуитивно выявлять закономерности в данных, увеличивает количество усваиваемой информации. Инфограмма – это конкретный образец инфографики на любом из носителей информации. В данной статье под инфограммой будем понимать программное представление инфографики в виде графического контейнера, который может содержать видео, изображения, диаграммы, карты и текст, и соответствует назначению инфографики.

Существуют два противоположных подхода к созданию инфографики, расходящиеся в вопросах значимости эстетической составляющей. Один из них – исследовательский (explorative) подход, берущий начало в традициях оформления научных работ. Основателем данного направления является Эдвард Тафти, автор множества работ по информационному дизайну [8]. Подход характеризуется минимализмом, при котором всё несущественное для передачи информации должно быть опущено, а сама информация должна быть передана максимально точно. Основной целью этого подхода является стремление к донесению сути информации до целевой аудитории. Такой подход оправдан в технологиях анализа данных, научных исследованиях, бизнес-аналитике.

Другой подход называется сюжетным или повествовательным (narrative). Основателем подхода Найгел Холмс наиболее известен по иллюстрациям в журнале Time с 1978 по 1994 год, которые он назвал «поясняющие иллюстрации» (explanation graphics). Для этого подхода характерно стремление к созданию выразительных образов, эффектного дизайна и иллюстративности. Способ подачи информации выбирается интуитивно, исходя из опыта и художественного вкуса дизайнера. В результате знакомства с сюжетной инфографикой читатель получает не просто информацию, но и развлечение. Сферой применения этого подхода можно считать журналистику, блоги, маркетинговые и рекламные материалы.

Сравнение двух подходов выявляет следующую важную деталь: исследовательский подход подразумевает извлечение нужной информации самим читателем, тогда как повествовательный уже содержит заключение, к которому читатель должен прийти. А также лучше подходит для формализации информационного представления. Информационно-аналитические системы (в том числе использующие технологию OLAP), как правило, не предлагают готовых заключений, а является лишь аналитическим инструментом для их поиска. Соответственно исследовательский подход больше чем сюжетный подходит для создания инфографики в таких системах.

Для построения инфограмм необходимо определить размерность отображаемых данных в терминах технологии OLAP, проанализировать способы отображения многомерных данных, выявить основные когнитивные элементы инфограмм. Систематизация разновидностей инфограмм создаст фундамент для разработки аналитических инфографических моделей.

По способу отображения инфограммы разделяются на два вида: статические и динамические. Динамические инфограммы наиболее удобны для отображения многомерных данных. С помощью анимации или встроенных в инфограмму инструментов (списков, фильтров) можно отобразить различные срезы гиперкуба, изменяя фиксированные значения измерений, или выполнить его вращение.

Обзор существующих решений

Существует ряд готовых решений автоматизированных инструментов построения инфографики, в том числе онлайн-сервисы. Среди них можно выделить сервисы, использующие как исследовательский подход, так и сюжетный.

К первой группе относятся такие сервисы как Piktochart.com или Infogr.am, предлагающие пользователю шаблоны страниц и диаграмм. Экспорт данных для диаграмм возможен из таблиц (xls, xlsx, csv), либо после установки динамической связи с данными таблиц Google Spreadsheet [9]. В

зависимости от количества анализируемых показателей изменяется выбор доступных видов диаграмм, подходящих для текущей таблицы. Отметим, что для одного показателя набор способов отображения шире, но в то же время при отображении двух и более показателей появляются новые виды отображений.

К явному преимуществу инфограмм, построенных с помощью перечисленных двух сервисов, относится их динамичность. В построенных инфограммах предусмотрена возможность выбора показателя для отображения, выделение показателя (при наведении курсора мыши) для конкретного значения измерения. На рис. 1 представлен пример анализа накопленной статистики о проведенных научных мероприятиях. Инфограмма состоит из двух диаграмм, отображающих два разных показателя одного многомерного куба данных. Измерениями построенной OLAP-модели являются: год, наименование мероприятия, организация; показателями – количество статей, их авторов и сделанных докладов. Выполнен срез куба, где фиксирован год (2015) и мероприятие (XIV Всероссийская конференция «Проблемы информатизации региона»). Отображены показатели: количество авторов статей, количество прочитанных докладов сотрудниками различных организаций.



Рис.1. Динамическая инфограмма для отображения статистики конференции

Ко второй группе сервисов, использующих сюжетный подход, относятся сервисы Easel.ly и Creately.com, предлагающие использовать существующие макеты для создания инфографики. Easel.ly фактически является графическим редактором построения инфограмм. При этом предусмотрено всего 5 типов диаграмм с минимумом настроек, возможен экспорт из Excel [11].

Сервис Creately.com обладает широким функционалом, специализируется как инструмент для создания схем [12]. При этом динамической привязки внешних данных не выявлено. В отличие от Easel.ly частично соответствует принципам исследовательского подхода (схематичность инфограмм).

Набор диаграмм для отображения многомерных данных

Одним из самых востребованных инструментов для графического представления многомерных данных является диаграмма. Причины этого заключаются в простоте построения диаграмм и наглядности представления информации для пользователей. Инфограммы могут содержать сразу несколько диаграмм, или состоять из единственной диаграммы. Рассмотрим распространенные виды диаграмм и их возможность использования для отображения многомерных данных с различным числом измерений и показателей.

Гистограмма. Прямой перевод с греческого языка – «писать столбиком». Это своеобразный столбчатый график. Диаграммы такого вида могут быть объемные, плоские, вложенные и т.п. На одной гистограмме одновременно могут быть отображены сразу несколько показателей гиперкуба, при условии, что их значения – числа одного порядка. Также есть возможность не фиксировать одно из измерений. В этом случае гистограмма будет разбита на группы, по количеству значений нефиксированного измерения.

Точечная диаграмма. Показывает взаимную связь между числовыми данными (показателями) и представляет собой набор чисел в виде единственного ряда точек в координатах. Виды диаграмм такого типа необходимы для поиска кластеров данных. Лучше всего использовать для анализа взаимозависимости двух показателей, но можно использовать и для большего числа показателей. Значения измерений для точечной диаграммы фиксированы.

Пузырьковая диаграмма. Одна из разновидностей точечной диаграммы. Подходит для

отображения трех показателей. Величина «пузырька», положение которого на координатной плоскости определяется двумя показателями, зависит от величины третьего показателя. Значения измерений для пузырьковой диаграммы фиксированы.

График. Отображает последовательность изменений числовых значений нескольких показателей для значений нефиксированного измерения, например, временного. Для лучшего восприятия значения всех показателей должны быть числами одного порядка.

Диаграммы с областями. Основной целью такой диаграммы является акцент на суммарной величине изменения нескольких показателей для значений выбранного (как правило, временного) измерения. Возможно также отображение доли отдельно взятых значений в общей сумме.

Кольцевая и круговая диаграммы. Данные виды диаграмм весьма схожи по целям. Обе они отображают роль каждого элемента в общей сумме. Их отличие заключается лишь в том, что кольцевая диаграмма может содержать несколько рядов показателей. Каждое вложенное кольцо представляет собой отдельный показатель. Частный случай кольцевой диаграммы – круговая диаграмма строится для одного показателя. Значения измерений для обеих диаграмм фиксированы.

Лепестковая диаграмма. Для данной диаграммы выбирается одно нефиксированное измерение. Каждое его значение представляет индивидуальную координатную ось, исходящую от нулевой точки координат. Затем на каждой такой оси откладываются значения показателей. Данный вид диаграмм позволяет сравнивать значения нескольких показателей.

Мишень. Диаграмма является средством когнитивной графики в виде лепестковой диаграммы (мишени) разбитой на сектора с отображением дополнительных зависимостей. Диаграмма-мишень служит для отображения значений основного показателя в виде секторов мишени и динамики изменения остальных показателей в виде траекторий, отмеченных на секторах мишени. Для каждого объекта может существовать несколько зависимостей: одна для основного показателя и несколько дополнительных, представляющих интерес при анализе того или иного процесса. Каждый сектор мишени соответствует значению выбранного измерения.

Диаграмма на основе 3-симплекс метода. На основе 3-симплекс метода [13] возможно построение когнитивной графики для представления многомерных данных для 4 показателей – по числу граней тетраэдра. Для каждого из них выбирается своя ось, перпендикулярная одной из граней тетраэдра. Значения показателей откладываются на выбранных осях, в результате ячейке многомерного куба ставится в соответствие точка пространства. Преимущество метода заключается в свободе выбора фиксированных и нефиксированных измерений. С помощью 3-симплекс метода визуально решается задача поиска кластеров на множестве анализируемых данных.

Таб.1. Зависимость допустимого числа измерений и показателей от вида диаграммы

	Гистограмма	Точечная диаграмма	Пузырьковая диаграмма	График	Диаграммы с областями	Кольцевая и круговая диаграммы	Лепестковая диаграмма	«Мишень»	Диаграмма на основе 3-симплекс
Число измерений	2	2	1	2	2	1	2	3	4
Число показателей	от 1	1	3	1	1	1	1	3	от 1

Зависимость допустимого числа измерений и показателей от вида диаграммы приведена в таб. 1. Следует понимать, что описание видов диаграмм приведено для создания статических инфограмм для отображения результатов OLAP-анализа. При создании динамических инфограмм можно добавить функцию выбора показателей, или дополнительных измерений для анализа данных.

Создание шаблонов и использование карт для построения инфограмм

Кроме диаграмм инфограмма может содержать видео, изображения, карты и текст. Как правило, данный контент создается не автоматически (за исключением картограмм), а входит в специально подготовленный шаблон с заданной структурой для создания инфограммы. Для графического представления результатов аналитик выбирает подходящий шаблон, куда будут встроены ключевые данные OLAP-анализа.

Рассмотрим механизм отображения данных OLAP-анализа на карте [14]. В основе механизма

динамической связи карты и гиперкуба данных OLAP-системы лежит картографическая привязка данных, устанавливающая соответствие между территориальными объектами карты и одним из измерений гиперкуба – географическим измерением. Привязка географического измерения осуществляется к одному или нескольким картографическим слоям. Чтобы значения показателя гиперкуба отобразить в виде тематической карты, необходимо построить легенду тех слоев, к объектам которых выполнена картографическая привязка. При построении легенды используются методы тематического картографирования, позволяющие объекты слоя разбить на классы на основе соответствующих значений анализируемого показателя. Одновременно на карте могут отображаться сразу несколько показателей, значения которых передаются с помощью цвета, обозначений или состояния объекта (видимый, невидимый, выделенный, мигающий). Кроме привязки данных к существующим объектам возможно динамическое формирование новых картографических слоев, отображающих многомерные данные [15]. Такой вид слоев соответствует представлению на карте объектов, существование, местоположение и форма которых определяется в зависимости от данных, принятых из OLAP-системы. Для формирования новых слоев могут быть использованы таблицы агрегатов, содержащие агрегированные данные, прошедшие предварительную обработку. Это позволяет сформировать новый слой не только на основе собранных статистических данных, но и на основе аналитических результатов.

Для получения информации о сформированных картографических объектах предусмотрены информационные карточки. В карточках отображаются значения показателей, соответствующие данному объекту, и зафиксированные на данный момент измерения. В качестве картографические подложки могут быть использованы современные геосервисы, такие как Яндекс.Карты или Google Maps.

Уровни воды на реках

Дата: вторник, 12 апреля, 2016 - 09:29

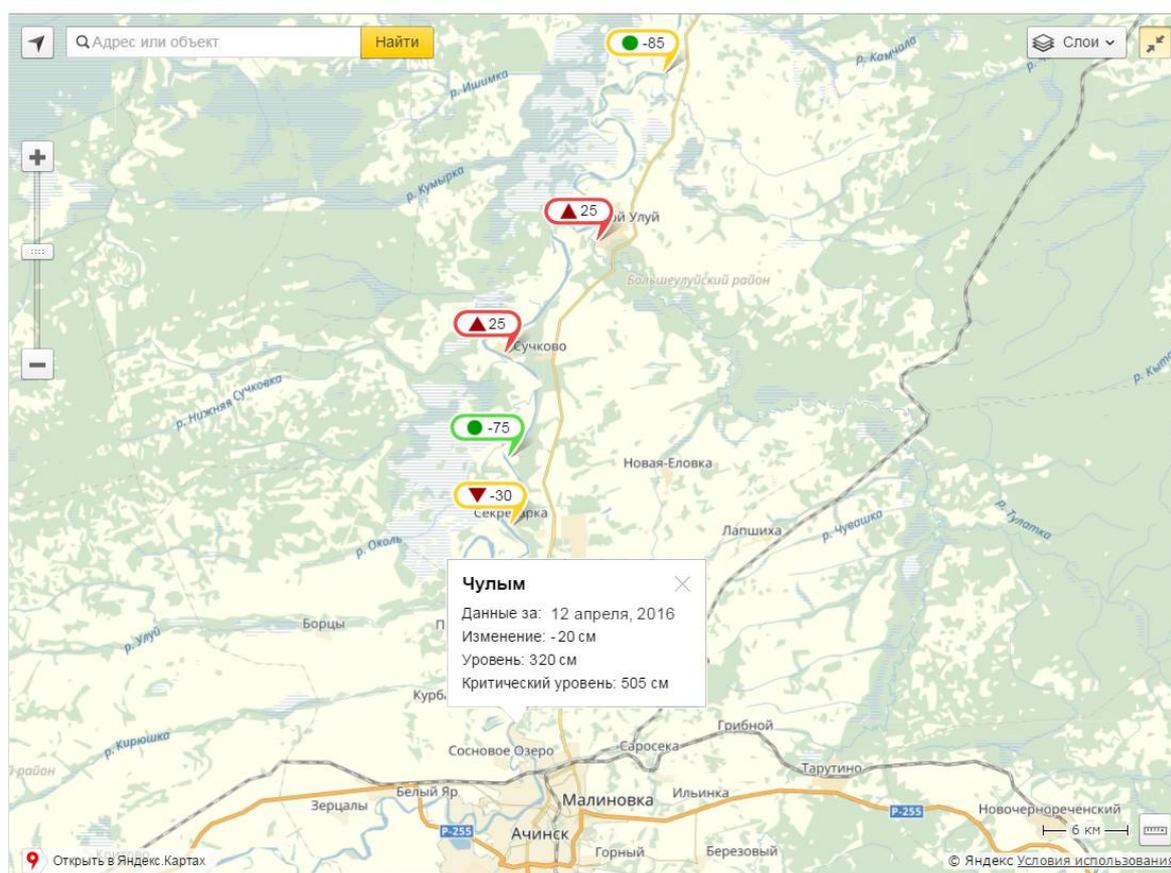


Рис.2. Пример картограммы с динамически сформированными слоями

На рис. 2 приведен пример построения картограммы на основе сервиса Яндекс.Карты для отображения ситуации на реках во время весеннего половодья на реках Красноярского края. Динамически сформированный слой отображает информацию о суточном изменении уровня воды на гидропостах. Красная каемка у меток на карте информирует о превышении критического уровня, желтая – о приближении к критической отметке, зеленый – о нормальном уровне воды. Измерениями построенной OLAP-модели являются: дата и место наблюдения; показателями –

текущий и критический уровни воды, суточное изменение воды. Последний показатель является агрегированным, его значение рассчитывается из разницы уровней воды за два дня. Выполнен срез куба, где фиксирован конкретная дата (12 апреля 2016). Из представленной картограммы видно, что на 12 апреля 2016 года происходило подтопление населенных пунктов Сучково и Большой Улуй.

Заключение

В работе выполнен обзор существующих подходов и инструментов построения инфограмм, определены виды инфограмм, проанализированы предпосылки интеграции методов инфографического моделирования и технологии оперативной аналитической обработки данных.

Актуальной задачей дальнейшего исследования является создание методических основ многомерного когнитивного моделирования и разработка аналитических инфографических моделей. Необходимо решить проблему взаимодействия с OLAP-системами, увеличить скорость построения инфограмм при выполнении операций над гиперкубом. Отдельное внимание стоит уделить вопросам восприятия – использованию разных способов графического представления информации при визуализации результатов многомерного анализа данных.

Литература

1. Stacey M., Salvatore J., Jorgensen A. Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data. – John Wiley & Sons, 2013.
2. Airinei D., Homocianu D. Data visualization in business intelligence //2010 Proceeding of WSEAS MCBEC2010-Recent Advances In Mathematics And Computers In Business, Economics, Biology & Chemistry. – 2010.
3. Pur A. et al. Data Presentation Methods for Monitoring a Public Health-Care System //11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007. – Springer Berlin Heidelberg, 2007. – С. 708-711.
4. Dance D. R. Mind mapping your way to the right software //Strategic Finance. – 1997. – Т. 78. – №. 10. – С. 36.
5. Dos Santos F. A. P., das Neves R. C., Filipe J. B. On the Generation of Dynamic Business Indicators //KDIR. – 2012. – С. 390-394.
6. Brackett K. E., Pal S. Creating high-impact visualisations //Applied Marketing Analytics. – 2015. – Т. 1. – №. 4. – С. 331-341.
7. Павлов А. А. Компьютеризация инфографического моделирования //Интернет-журнал Науковедение. – 2012. – №. 2 (11).
8. Tufte, Edward R (2001) [1983], The Visual Display of Quantitative Information (2nd ed.), Cheshire, CT: GraphicsPress, ISBN 0-9613921-4-2.
9. Create Easy Infographics, Reports, Presentations | Piktochart. [Электронный ресурс]. URL: <http://piktochart.com> (дата обращения: 21.02.2016).
10. Create and share visual ideas online | Easel.ly. [Электронный ресурс]. URL: <http://www.easel.ly> (дата обращения: 21.02.2016).
11. Create online charts & infographics | Infogr.am. [Электронный ресурс]. URL: <https://infogr.am> (дата обращения: 21.02.2016).
12. Online Diagram Software to draw Flowcharts, UML & more | Creately[Электронный ресурс]. URL: <http://creately.com> (дата обращения: 21.02.2016).
13. Yankovskaya A., Krivdyuk N. Cognitive Graphics Tool Based on 3-Simplex for Decision-Making and Substantiation of Decisions in Intelligent System // Proceedings of the IA STED International Conference Technology for Education and Learning TEL '2013. Marina del Rey, USA, November 11-13, 2013. P. 463-469.
14. Евсюков, А.А. Применение технологии OLAP для построения тематических карт / А.А. Евсюков // Сборник научных трудов VIII Международной научно-практической конференции «Современные информационные технологии и ИТ-образование». – Москва: МГУ, 2013. – С. 377-384.
15. Евсюков, А.А. Динамическое формирование картографических слоев в информационно-аналитических системах / А.А. Евсюков // Вестник Сибирского государственного аэрокосмического университета имени академика М. Ф. Решетнева. №1. – Красноярск: 2011. – С. 15-20.

References

1. Stacey M., Salvatore J., Jorgensen A. Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data. – John Wiley & Sons, 2013.
2. Airinei D., Homocianu D. Data visualization in business intelligence //2010 Proceeding of WSEAS MCBEC2010-Recent Advances In Mathematics And Computers In Business, Economics, Biology & Chemistry. – 2010.
3. Pur A. et al. Data Presentation Methods for Monitoring a Public Health-Care System //11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007. – Springer Berlin Heidelberg, 2007. – P. 708-711.
4. Dance D. R. Mind mapping your way to the right software //Strategic Finance. – 1997. – Т. 78. – №. 10. – P. 36.
5. Dos Santos F. A. P., das Neves R. C., Filipe J. B. On the Generation of Dynamic Business Indicators //KDIR. – 2012. – P. 390-394.
6. Brackett K. E., Pal S. Creating high-impact visualisations //Applied Marketing Analytics. – 2015. – Т. 1. – №. 4. – P. 331-341.
7. Pavlov A. A. Computerization of info-graphic simulation // Internet-journal Naukovedenie. – 2012. – №. 2 (11).
8. Tufte, Edward R (2001) [1983], The Visual Display of Quantitative Information (2nd ed.), Cheshire, CT: GraphicsPress, ISBN 0-9613921-4-2.
9. Create Easy Infographics, Reports, Presentations | Piktochart. [Электронный ресурс]. URL: <http://piktochart.com> (дата обращения: 21.02.2016).
10. Create and share visual ideas online | Easel.ly. URL: <http://www.easel.ly> (дата обращения: 21.02.2016).
11. Create online charts & infographics | Infogr.am. URL: <https://infogr.am> (дата обращения: 21.02.2016).

12. Online Diagram Software to draw Flowcharts, UML & more | Creately. URL: <http://creately.com> (дата обращения: 21.02.2016).
13. Yankovskaya A., Krivdyuk N. Cognitive Graphics Tool Based on 3-Simplex for Decision-Making and Substantiation of Decisions in Intelligent System // Proceedings of the IA STED International Conference Technology for Education and Learning TEL '2013. Marina del Rey, USA, November 11-13, 2013. P. 463-469.
14. Evsyukov A.A. OLAP technology usage to build thematic maps / A.A.Evsyukov // Collection of scientific works of the VIII International scientific-practical conference "Modern information technology and IT-education." - Moscow: Moscow State University, 2013. - P. 377-384.
15. Evsyukov A.A. Dynamic generation of map layers in the information-analytical systems / A.A.Evsyukov // Bulletin of the Siberian State Aerospace University named after Academician M.F.Reshetnev. №1. - Krasnoyarsk: 2011. - P. 15-20.

Поступила 20.10.2016

Об авторе:

Евсюков Александр Анатольевич, кандидат технических наук, старший научный сотрудник Отдела прикладной информатики Института вычислительного моделирования Федерального исследовательского центра "Красноярский научный центр СО РАН", alev@icm.krasn.ru.