

## Решение задач анализа данных с помощью аналитических визуальных моделей

А.А. Захарова, Е.В. Вехтер, А.В. Шкляр  
zaa@tpu.ru|vehter@tpu.ru|shklyarav@tpu.ru

Институт кибернетики Томского политехнического университета, Томск, Россия

*В работе показаны возможности и преимущества использования авторского метода комплексного визуального анализа, возникающие при решении практических задач исследования данных. Показан результативный подход к построению визуальных моделей, являющихся инструментами визуальной аналитики и обладающих предсказуемой логичностью и контролируемой результативностью. Успешность решения задач анализа обеспечена обоснованным использованием свойств визуальных моделей. Предложена базовая методика проведения аналитического исследования произвольных данных, позволяющая ввести понятие формальной визуальной модели.*

**Ключевые слова:** визуальная аналитика, визуальная модель, анализ данных, интерпретация, визуальное восприятие.

## Data analysis problems' solving with a use of analytical visual models

A.A. Zakharova, E.V. Vekhter, A.V. Shklyar  
zaa@tpu.ru|vehter@tpu.ru|shklyarav@tpu.ru

Institute of Cybernetics, Tomsk Polytechnic University, Tomsk, Russian Federation

*The work shows the features and benefits of using copyright method integrated Visual analysis that arise when solving practical tasks of research data. Shows a productive approach to building Visual models are tools for Visual Analytics and with predictable consistency and controlled performance. The success of decision analysis problems is provided justified using the properties of Visual models. Proposed basic methodology of analytical study of arbitrary data, allowing to introduce the notion of a formal Visual model.*

**Keywords:** visual analytics, visual model, data analysis, visual interpretation, visual perception.

### 1. Введение

Визуализация является одним из способов моделирования и выполняет функции инструмента познания [1]. Полученный наблюдателем зрительный образ является искусственной формой, находящейся между чувственным восприятием и мышлением. Это позволяет объединить зрительную обработку информации и мысленный анализ данных, хранящихся в памяти исследователя.

Результатом визуального анализа является ответ на вопрос исследования, полученный с помощью визуальной модели в пределах поставленных ограничений [3]. Результативность визуального анализа возрастает в случае реализации комплексного подхода в функции визуального моделирования. Комплексный подход к визуальному анализу данных подразумевает сбалансированное и целенаправленное участие в функции визуального моделирования всех значимых для результата анализа составляющих.

В рамках комплексного подхода, в визуальном анализе выделяют три основные составляющие функции моделирования – функциональную, семиотическую и психоэмоциональную. Построение визуальной модели, обладающей предварительно заданными возможностями, соответствующими особенностям решаемой задачи, на основании комплексного подхода к визуализации данных эквивалентно определению способа и степени активности участия каждой из указанных составляющих.

Структура визуальной модели, предназначенной для проведения анализа данных, в т.ч. для задач с низким уровнем формализации, определяет последовательность и логику рассуждений исследователя. Таким образом, свойства визуальной модели оказывают прямое влияние на получаемое решение [2]. С другой стороны, визуальная модель, позволяющая получить решение таких задач,

становится полноценным аналитическим инструментом, реализующим логику рассуждений исследователя и его понимание изучаемой предметной области. Кроме того, визуальная модель становится воспринимаемой репрезентацией нового знания, возникающего в ходе проводимого анализа.

### 2. Структура визуального анализа

Анализ, в общем случае, является целенаправленным действием, которое может рассматриваться как последовательность, начинающаяся с постановки вопроса и продолжающаяся поиском гипотез, содержащих ответ на этот вопрос. В таком случае, правильность результата анализа зависит от двух факторов: сути вопроса, которая определяется степенью предварительного понимания анализируемой информации, а также от формы этого вопроса. Форма определяется языком, на котором поставлен вопрос, внутренней структурой, логикой и необходимостью взаимодействия с новыми источниками информации [5, 8].

Анализ, осуществляемый с помощью визуального образа исследуемых данных, можно представить аналогичным образом. Сам образ в момент начала анализа является моделью исходных данных. Вместе с моделью или на основании ее изучения у пользователя формируется вопрос дальнейшего исследования. Возникает циклически повторяемая ситуация, для которой характерно наличие поисковой и верифицирующей процедур. Все следующие действия наблюдателя направлены на построение и оценку гипотезы промежуточного ответа на основании информации, представленной в визуальной модели и, вероятно, с использованием знаний пользователя [14]. Следовательно, многошаговый последовательный процесс визуального анализа состоит в переходе от визуального образа-вопроса, являющегося формулировкой решаемой задачи, к мысленному образу-ответу, интерпретируемому

пользователем как обнаруженная закономерность в исследуемых данных (Рис. 1).

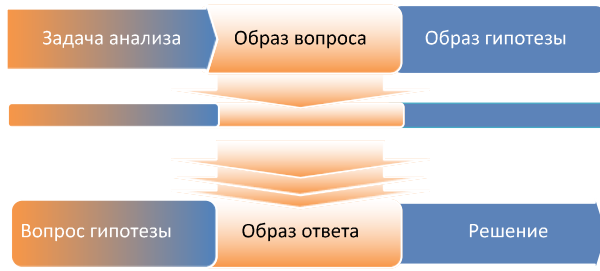


Рис. 1. Структура визуального анализа.

### 3. Структурная единица визуального анализа

Введем определение структурной единицы визуального анализа – элемента визуальной модели. Структурной единицей визуализации (визуальным элементом) будем считать состояние визуальной модели, интерпретация которого предоставляет наблюдателю объем информации, необходимый для последующего получения общего результата проводимого анализа. Согласно приведенному описанию структуры визуального анализа, такое состояние является визуально воспринимаемым образом, интерпретируемым как однозначный ответ на промежуточный вопрос. Сложность такого вопроса определяется ограниченностью времени, предоставляемого процедурой анализа, а также ресурсоемкостью процессов создания структурной единицы и верификации сведений, полученных пользователем, благодаря ее изучению.

Структурная единица визуального анализа представляет собой управляемую систему  $S$  с обратной связью  $R(t)$ . Объем исследуемых данных поступает на неуправляемый вход  $V(t)$  структурной единицы (Рис. 2). Особенностью такой системы является возможность изменения состояния управляемого входа системы  $U(t)$  в зависимости от полученных результатов  $Y(t)$ . Пользователь  $P$  является, с точки зрения комплексного подхода к визуальному анализу, обязательным участником структурной единицы.

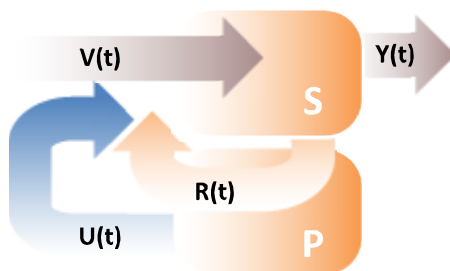


Рис. 2. Структурная единица визуального анализа.

### 4. Комбинированная модель

В ряде случаев, например, при исследовании сложных и объемных данных [7], также как при изучении особенностей изменяющихся состояний, может оказаться недостаточным использование одной визуальной модели. Это происходит в тех случаях, когда визуальный образ не может, по любым причинам, содержать в себе всего объема необходимых сведений, либо, если технически это возможно, но адекватная интерпретация вызывает у пользователя затруднения. Решением, позволяющим найти выход из этого затруднения, может быть построение набора визуальных моделей, каждая из которых соответствует некоторому фрагменту исходного объема данных. В результате, происходит построение сложной визуальной модели, имеющей более широкие возможности.

Визуальные модели различаются по своим возможностям, назначению и применимости для решения различных видов задач анализа. На основании предложенного описания структурной единицы визуального анализа, тип визуальной модели характеризуется параметрами активности использования информационных входов и выходов единиц, составляющих модель. В одном из простейших случаев, визуальный образ используется как индикатор состояния наблюдаемой системы. Целью подобной модели данных является оповещение наблюдателя об изменении состояния или, точнее, о наступлении ожидаемого события.

Визуальные модели данных получили широкое применение в системах поддержки принятия решений, обладающих различным уровнем сложности. В этом случае, задача анализа требует быстрого осмысления поступающих данных, формирования одной или нескольких гипотез ответа на вопрос исследования и выбор наилучшей из них. Для этого типа визуального анализа характерно активное использование дополнительных сведений и знаний пользователя при формулировании промежуточных вопросов и получении выводов, оказывающих влияние на дальнейшее существование исследуемой системы.

Объединение структурных единиц в функционирующую систему, приводит к построению визуальной модели, когнитивное значение которой превосходит результативность исследования отдельных образов. Необходимо наличие правила, объединяющего отдельные элементы в единую систему и обеспечивающего получение ожидаемого результата анализа. Для изменяющихся данных, таким правилом может быть построение хронологической последовательности, позволяющей наблюдать и сравнивать динамические характеристики системы. Однако, использование в качестве объединяющего правила естественных закономерностей, связанных с происхождением данных, не является обязательным.

Эмерджентность визуальной модели обеспечивается связями составляющих ее элементов, поэтому особое значение приобретает выделение и описание именно этих свойств структурных единиц. Условно, они могут быть разделены на две функциональные группы: информативные и управляющие связи.

Управляющие связи структурных единиц, ответственные за верификацию появляющихся гипотез в визуальных моделях, содержащих совокупность анализируемых образов, могут, а иногда и обязаны учитывать при выполнении своих функций всю информацию, полученную пользователем на уже пройденных этапах анализа. С точки зрения повышения достоверности результатов проводимого анализа, это обстоятельство является положительным достижением. При этом, более сложная процедура проверки может быть связана с дополнительными затратами времени, являющегося одним из основных критериев результативности визуальной модели, и это потребует уточнения разработанной системы образов для достижения сбалансированного решения.

В описанной таким образом логической структуре визуальной модели наряду с последовательностью наблюдаемых пользователем образов данных формируются еще два функциональных объекта, участвующих в визуальном анализе. Первым является цепочка промежуточных вопросов, руководящих процессом исследования данных и ответственных за его логичность и обоснованность. Вторым объектом становится управляющая последовательность, состоящая из процедур верификации, активно использующих возможности

визуального восприятия и эмоциональной оценки наблюдаемых исследователем образов. Успешность применения модели для решения задачи анализа достигается сбалансированным и полноценным использованием всех логических элементов, образующих структуру функционирования визуальной модели.

## 5. Аналитическая модель

Структура визуальной модели, предназначенной для проведения анализа данных в задачах с низким уровнем формализации, определяет последовательность и логику рассуждений исследователя (Рис. 3). Таким образом, свойства визуальной модели оказывают прямое влияние на получаемое решение. Следует обратить внимание на то обстоятельство, что результативность составляющих визуальной модели (функциональной, информативной и управляющей) зависит от использования определенных возможностей пользователя [4].

В случае сложной модели, последовательное приближение к решению задачи приводит к непрерывному и обязательному присутствию пользователя в процессе анализа. Таким образом, полнота логической схемы функционирования визуальной модели не может быть достигнута без включения в нее четвертого обязательного элемента – возможностей конкретного пользователя, влияющих на результат анализа на протяжении прохождения всей последовательности образов.

Наличие интерактивного управления визуальным образом обеспечивает прямое участие пользователя в манипулировании образом и является основанием для глубокого анализа данных [9]. Обоснованная система управления создает условия для постановки новых вопросов исследования и быстрое получение ответов, ускоряя достижение цели анализа. Таким образом, дополнение визуальной модели управляющим интерфейсом, соответствует предложенной структуре визуального анализа и позволяет сделать его инструментом аналитического решения любой поставленной задачи.

Основной идеей, обеспечивающей увеличение когнитивной результативности визуального анализа, становится направленное использование возможностей пользователя, реализуемое благодаря свойствам визуальной модели. Это создает возможность привлечения дополнительного ресурса для решения поставленной задачи анализа [10]. Перспективными направлениями в этом случае становятся привлечение не только информированности пользователя, но и особенностей его мышления, как для формирования гипотезы решения, так и для принятия решения о ее достоверности.

В случае исследования данных, относящихся к вопросам, не имеющим полноценного формального описания структуры соответствующих знаний, использование визуальных моделей становится не только способом обнаружения новых закономерностей, но и формирует понимание их смысла. Следовательно, приобретение визуальной моделью статуса аналитического инструмента возможно лишь в том случае, когда созданы

условия для интерактивного управления способом представления данных с ее помощью. Визуальная модель, предлагающая наблюдателю на первоначальном этапе исследования образ не только исходных данных, но и обнаруженных закономерностей, может использоваться в качестве *формальной визуальной модели* области знаний.

Анализ логической структуры визуальной модели, полученной в результате приведенных выше рассуждений, позволяет дать ответы на вопросы, относящиеся к увеличению результативности использования визуального анализа [13]. Основными критериями эффективности визуальной модели выступают два параметра: достоверность полученного результата и время, затраченное на его достижение. Наличие в полной схеме, как минимум, четырех функциональных элементов и их взаимное влияние приводит к необходимости определения правил, соблюдение которых необходимо для получения наиболее эффективной комбинации. С учетом динамичного характера проведения процедуры анализа, соблюдение и адаптация этих правил к конкретным условиям становятся еще одной целью интерактивного управления визуальной моделью.

Система интерактивного управления моделью является одним из способов верификации полученных решений и потому обеспечивает ускоренное достижение цели анализа в случае, когда этот способ удобен для пользователя. Интерактивность визуальной модели становится условием ее высокой результативности [6, 11].

Комбинирование единичных элементов позволяет получать информативные, визуально воспринимаемые объекты, которые могут быть использованы для решения самых разных задач. В их число входят визуальное представление информации, целенаправленное использование когнитивного потенциала исследователя, активизируемое воздействием образного восприятия, исследование данных, не имеющих формального описания или объяснения. Особый интерес представляет задача построения визуальных моделей, содержащих данные целых предметных областей. Это позволяет проводить визуальный анализ данных, имеющих различное происхождение, достоверность, формат и смысл, с целью их обобщения, верификации и поиска объяснения.

Таким образом, аналитический потенциал визуальной модели определяется выбором способов визуального моделирования изучаемых данных, максимально возможной скоростью построения зрительного образа при сохранении всех необходимых, с точки зрения функциональности, нюансов визуального представления [12]. Кроме этого, возможно, наиболее значимым элементом модели становится ее интерфейс, необходимый для использования информативного и когнитивного потенциала самого пользователя.

## 6. Исследование предметной области

В качестве практической задачи, для исследования результативности использования визуальных моделей при изучении многомерных данных, была решена задача

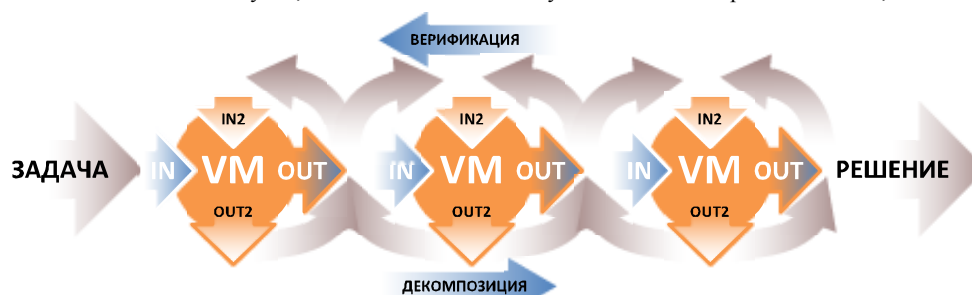


Рис. 3. Структура аналитической модели.



представления общей совокупности знаний в предметной области, содержащей экспериментальные сведения. Осложняющими обстоятельствами для подобной задачи стали большое число разнородных источников информации (публикации) и различающиеся структуры описаний сравниваемых данных. Источником исходной информации в базе исследуемых данных служили статьи, опубликованные в мировых рецензируемых изданиях.

В результате, построена *общая визуальная модель* эмпирических данных о современном состоянии изучения процессов получения карбида кремния электродуговым методом (Рис. 4). Модель представляет собой набор горизонтальных плоскостей, каждая из которых соответствует одной из характеристик исследуемых объектов. Линия, проведенная через точки, соответствующие одному объекту становится визуальным образом, характеризующим этот объект. Множество таких визуальных объектов (трек-образов) составляет визуальную модель, позволяющую проводить сравнение объектов между собой и анализировать весь объем исследуемых данных. Одним из значимых результатов проведенного анализа стало определение стратегии развития предметной области. Кроме того, проведена оценка достоверности опубликованных данных и изменений областей концентрации внимания исследователей во времени.

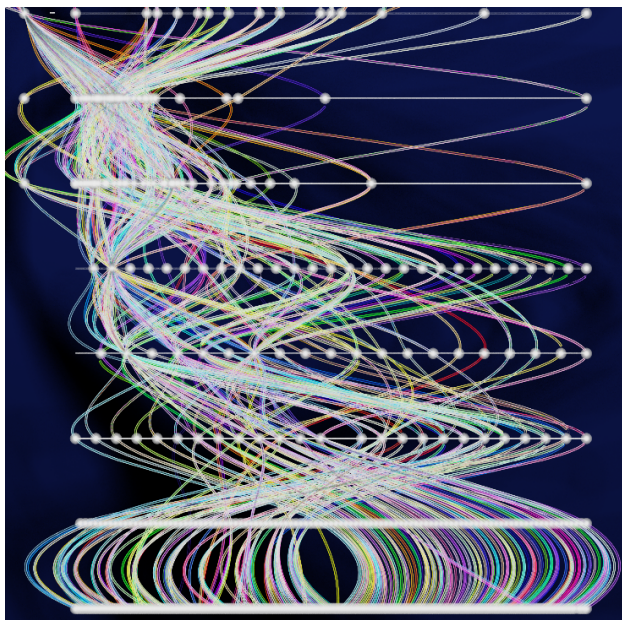


Рис. 4. Визуальная модель эмпирических данных.

С помощью построенной визуальной модели проанализированы сведения о 260 экспериментах, посвященных получению углеродных ультрадисперсных материалов в плазме электрической дуги постоянного тока. Получены выводы о росте объема экспериментальных данных преимущественно в отношении получения четырех наиболее популярных продуктов. При сравнении групп отдельных параметров (показатели эксперимента – получаемый продукт) обнаружено, что производительность рассматриваемых экспериментальных систем практически не оценивается за исключением обрывочных несистематизированных данных, что говорит о неготовности исследуемого метода к внедрению в промышленное производство [14].

Значительный интерес представляет использование фрагментированных моделей, представляющих образы комбинаций отдельных параметров изучаемых данных [13]. Создание таких моделей соответствует введенному

понятию декомпозиции исходной задачи анализа. Кроме того, построение подобных «частичных» моделей обеспечивает пользователя ответами на самостоятельно сформулированные промежуточные вопросы. Полученная исследователем новая информация позволяет формировать вероятные гипотезы ответов на основной вопрос исследования, принимать решение о переходе к следующему шагу анализа или проводить поиск внутренних закономерностей, неявно присутствующих в исходных данных.

Модель соответствия между годами исследований и напряжением, характеризующим отдельный эксперимент (Рис. 5), позволяет провести анализ изменений мировых тенденций и перейти к формулировке вопроса об их причинах. Преимуществом подобной модели становится быстрое определение особенностей данных (в приведенном примере, выделение значения напряжений, при которых проведено наибольшее число исследований на протяжении длительного времени). Одним из следствий подобного алгоритма исследований становится обоснованное исключение из анализа части данных, не представляющих интерес на основании промежуточных выводов. Этот шаг увеличивает результативность анализа при проведении исследований больших объемов данных.

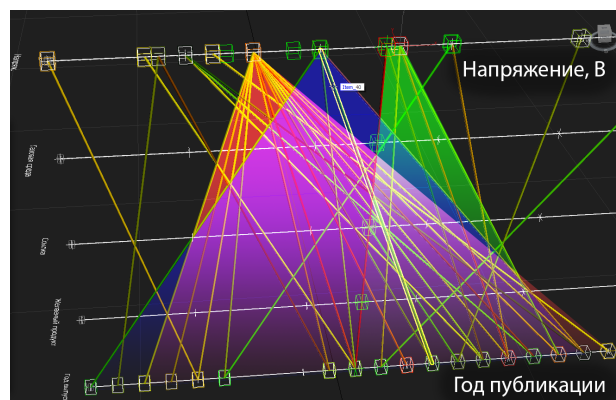


Рис. 5. Фрагментированная визуальная модель.

В результате проведенного исследования были подтверждены основные предположения, описывающие структуру визуального анализа. Благодаря использованию интерфейса управления визуальной моделью, достигнута высокая результативность анализа. Это связано с сокращением времени анализа, по сравнению с традиционными способами изучения подобных данных, для которых время исследования измеряется несколькими днями, даже с учетом опыта и квалификации специалистов.

По совокупности представленных сведений в отношении анализа диаграммы энергетических характеристик получен вывод о том, что полем для дальнейших исследований метода синтеза углеродных материалов в плазме дуги постоянного тока должна стать взаимосвязь уровня напряжения и электрической мощности с другими параметрами экспериментальных установок для получения материала с заданным фазовым составом и свойствами в контексте требований к электрической мощности системы и объема потребления электроэнергии. Основаниями для подобных выводов являются точки концентрации данных на определенных величинах выбранных параметров (Рис. 6), а также наблюдаемые в модели явные области недостаточной изученности.

## 7. Заключение

Предложенный в работе подход к решению задач анализа с помощью визуальных моделей данных позволяет

оперативно получать ответы на вопросы исследования, в том числе, высокой степени сложности. Построение модели, предоставляющей возможность визуализировать исходные данные и формулировать промежуточные вопросы с целью поиска неизвестных ранее внутренних закономерностей, дает пользователю аналитический инструмент для исследования поставленной перед ним задачи. Помимо возможности достичь цели анализа с помощью этого инструмента, пользователь получает возможность существенно сократить время, затрачиваемое им на поиск ответа или принятие решения в ситуации выбора наилучшего варианта. Сделанные утверждения проверены при анализе опубликованных данных, посвященных получению углеродных ультрадисперсных материалов.

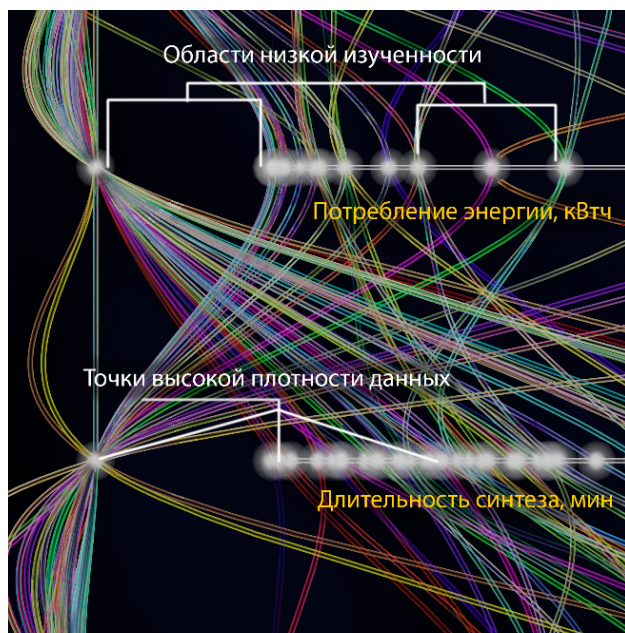


Рис. 6 Визуальный поиск закономерностей.

Работа выполнена в рамках госзадания № 2.1642.2017/ПЧ на выполнение проекта по теме «Когнитивные методы визуализации и анализа многомерных данных при моделировании нелинейных динамических систем».

## 8. Литература

- [1] Bondarev A.E., Galaktionov V.A.: Multidimensional data analysis and visualization for time-dependent CFD problems. *Programming and Computer Software* 41(5), 247–252 (2015). DOI: 10.1134/S0361768815050023.
- [2] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, Hanspeter Pfister, What Makes a Visualization Memorable?, *IEEE Transactions on Visualization and Computer Graphics*, v.19 n.12, p.2306-2315, December 2013 [DOI: 10.1109/TVCG.2013.234]
- [3] Chen, C.: *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. 2nd ed. London: Springer, (2013).
- [4] Chen C.: Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications* 25(4), 12–16 (2005).
- [5] Eppler, M., Burkhard, R.A.: Visual representations in knowledge management: Framework and cases. *Journal of Knowledge Management*, vol. 11(4), pp. 112-122.

QEmerald Group Publishing Limited (2007). ISSN 1367-3270 DOI 10.1108/1367327071076275.

- [6] Huang, D., Tory, M., Adriel Aseniero, B., Bartram, L., Bateman, S., Carpendale, S., Tang, A., Woodbury, R.: Personal visualization and personal visual analytics. *IEEE Trans. Vis. Comput. Graph.* 21, 420–433 (2015).
- [7] Klimentov, A.; Buncic, P.; De, K.; Jha, S.; Maeno, T.; Mount, R.; Nilsson, P.; Oleynik, D.; Panitkin, S.; Petrosyan, A.; Porter, R. J.; Read, K. F.; Vaniachine, A.; Wells, J. C.; Wenaus, T.; Iop, Next Generation Workload Management System For Big Data on Heterogeneous Distributed Computing. *Proceedings from the 16th International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, Prague, Czech Republic, September 1–5, 2014; Iop Publishing Ltd: Bristol, U.K., 2015; Vol. 608.
- [8] Korobkin D., Fomenkov S., Kravets A., Kolesnikov S., Dykov M.: Three-Steps Methodology for Patents Prior-Art Retrieval and Structured Physical Knowledge Extracting. In: *Creativity in Intelligent Technologies and Data Science First Conference, CIT&DS 2015*, pp.124-138. Volgograd, Russia (2015).
- [9] Ozpeynirci, O., Ozpeynirci, S., Kaya, A.: An interactive approach for multiple criteria selection problem. *Computers and Operations Research* 78, 154-162 (2016).
- [10] Sedig Kamran, Parsons Paul, Hai-Ning Liang, Jim Morey: Supporting Sensemaking of Complex Objects with Visualizations. *Visibility and Complementarity of Interactions*, *Informatics* 3(4), 20 (2016). doi:10.3390/informatics3040020.
- [11] Shneiderman, B., Plaisant, C., Cohen, M.S., Jacobs, S.M., Elmqvist, N., Diakopoulos, N.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 6th ed. Pearson: Upper Saddle River, NJ, USA (2016).
- [12] Zakharova, A.A., Shklyar, A.V.: Visual presentation of different types of data by dynamic sign structures. *Scientific Visualization* 8(4), 28-37 (2016).
- [13] Zaripova V.M., Petrova I. Yu., Kravets A., Evdoshenko O.: Knowledge Bases of Physical Effects and Phenomena for Method of Energy-Informational Models by Means of Ontologies. In: *Creativity in Intelligent Technologies and Data Science First Conference, CIT&DS 2015*, pp. 224-237. Volgograd, Russia (2015).
- [14] Zavyalov, D.A., Zakharova, A.A., Shklyar, A.V., Bagutdinov, R.A.: An integrated approach to modeling by an example of a landfill of disposal of liquid oil waste. *Software systems and computational methods [Programmnye sistemy i vychislitel'nye metody]* 1, 22-30 (2017).

## Об авторах

Захарова Алена Александровна, Институт кибернетики, Томский политехнический университет, Томск, Россия, zaa@tpu.ru

Вехтер Евгения Викторовна, Томский политехнический университет, Томск, Россия, vechter@tpu.ru

Шкляр Алексей Викторович, Институт кибернетики, Томский политехнический университет, Томск, Россия, shklyarav@mail.ru