

УДК 004.921

DOI 10.17513/snt.39902

ДИСПЕТЧЕР НАУЧНЫХ РАБОЧИХ ПРОЦЕССОВ: МЕТОДЫ И СРЕДСТВА ВИЗУАЛИЗАЦИИ ДАННЫХ

Воскобойников М.Л.

*ФГБУН Институт динамики систем и теории управления имени В.М. Матросова
Сибирского отделения Российской академии наук, Иркутск, e-mail: mikev1988@mail.ru*

Статья посвящена актуальным вопросам визуализации научных рабочих процессов в распределенных пакетах прикладных программ (научных приложениях) и расчетных данных, получаемых при выполнении таких процессов. В рамках исследования разработаны новые инструментальные средства визуализации научных рабочих процессов и расчетных данных на разных этапах их обработки и анализа. Данные инструментальные средства обеспечивают конфигурирование и настройку используемых графиков и диаграмм. Предложена спецификация параметров процесса визуализации данных на языке JSON, расширяющая описание предметных областей приложений. Применение данной спецификации позволяет существенно повысить качество и гибкость процесса визуализации в сравнении с известными системами управления научными рабочими процессами. В качестве базового программного обеспечения для визуализации научных рабочих процессов приложений и расчетных данных предложено использовать соответственно библиотеки Graphviz и Highcharts. Рассмотрены преимущества этих библиотек в сравнении с другими библиотеками подобного назначения. Результаты исследования использованы при создании и применении приложения для исследования живучести энергетических систем. Приведены иллюстративные примеры визуализации одного из научных рабочих процессов приложения и расчетных данных, отражающих степень живучести исследуемой системы при массовых отказах ее элементов.

Ключевые слова: научный рабочий процесс, результаты расчетов, визуализация, спецификация, шаблоны

Исследование проведено при поддержке Министерства науки и высшего образования Российской Федерации, проект № FWEW-2021-0005 «Технологии разработки и анализа предметно-ориентированных интеллектуальных систем группового управления в недетерминированных распределенных средах».

SCHEDULER OF SCIENTIFIC WORKFLOW: METHODS AND TOOLS FOR DATA VISUALIZATION

Voskoboynikov M.L.

*Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch
of the Russian Academy of Sciences, Irkutsk, e-mail: mikev1988@mail.ru*

The paper is devoted to relevant issues of visualization of scientific workflows in distributed applied software packages (scientific applications) and computational data obtained during the workflow execution. Within this research, new tools for visualization of scientific workflows and computational data at different stages of their processing and analysis have been developed. These tools allow configuration and customization of the graphs and diagrams used. A specification of parameters for the data visualization process in the JSON language is proposed. This specification extends the description of application domains. The use of this specification can significantly improve the quality and flexibility of the visualization process compared to known workflow management systems. We propose to use the Graphviz and Highcharts libraries, respectively, as basic software for the visualization of scientific workflows of applications and computational data. The advantages of these libraries in comparison with other libraries for similar purposes are considered. The research results were used to create and apply an application for studying the resilience of energy systems. Illustrative examples of visualization of one of the scientific workflows of the application and computational data reflecting the degree of resilience of the system under study in case of mass failures of its elements are given.

Keywords: scientific workflow, computation results, visualization, specification, templates

The research was carried out with the support of the Ministry of Science and Higher Education of the Russian Federation, project No. FWEW-2021-0005 “Technologies for the development and analysis of subject-oriented intelligent group control systems in non-deterministic distributed environments”.

Визуализация информации является актуальной задачей в широком спектре современных направлений исследований, таких как картография и климатология, геоинформатика, социально-экономическое развитие регионов и природных территорий, обработка данных дистанционного зондирования Земли, навигация в мобильной ро-

бототехнике, прогнозирование, интеллектуальный анализ результатов моделирования природных и технических систем и др. [1]. При этом, как правило, основное внимание уделяется методам и средствам визуализации многомерных временных рядов данных, являющихся результатом наблюдений или выполнения натурных и/или вычисли-

тельных экспериментов и позволяющих настраивать (конфигурировать) отображение данных.

Визуализация призвана облегчить восприятие информации с целью ее дальнейшего анализа, выявления закономерностей в зависимости от влияния параметров исследуемых объектов и систем, а также поддержки принятия решений экспертами [2]. Поэтому важно иметь в наличии развитые методы и средства визуализации как первичных данных, так и вторичных данных, прошедших предварительную обработку и агрегирование. Это в равной степени относится как к визуализации текущих данных, собираемых различными системами измерения и контроля, так и к накопленной ретроспективной информации.

Зачастую визуализация данных реализуется простейшую задачу отображения зависимости функции одной переменной (параметра). Сложность визуализации возрастает при отображении зависимости функции от нескольких переменных (параметров). Разнообразие данных и необходимость обеспечения их четкого восприятия обуславливает применение набора атрибутов процесса визуализации, к основным из которых относятся размер, форма, ориентация, цвет, текстура и значение цвета [3]. Эти атрибуты могут быть заданы с помощью различного рода шкал и уровней их значений. Поэтому любой метод визуализации должен сопровождаться инструментом, позволяющим настроить (конфигурировать) изображение данных.

В рамках решения задач экологического мониторинга [4] возникает необходимость визуализации результатов расчетов, получаемых в результате выполнения ресурсоемких научных приложений (распределенных пакетов прикладных программ) в высокопроизводительной вычислительной среде. В этой связи целью исследования является

разработка инструментальных средств поддержки процессов визуализации расчетных схем и данных в приложениях.

Материалы и методы исследования

Распределенный пакет прикладных программ [5] представляет собой научное приложение, которое характеризуется следующими особенностями: решаемые задачи допускают их декомпозицию на более простые взаимосвязанные подзадачи; разработка прикладного ПО приложения производится на основе модульного подхода к созданию больших программных комплексов; предполагается разбиение исходных данных на блоки и независимая параллельная обработка этих блоков экземплярами модулей; расчеты осуществляются на основе выполнения научного рабочего процесса (НПП, англ., Scientific Workflow), представляющего собой схему решения задачи; управление вычислениями осуществляет диспетчер НПП. Инструментарии для разработки и применения подобных пакетов относятся к классу систем управления научными рабочими процессами. Примером такого инструментария является фреймворк Orlando Tools [6].

В рамках диспетчера НПП Orlando Tools разработаны технологические решения и инструментальные средства для визуализации параметров НПП и самих процессов [7]. НПП и их параметры задаются в описании предметной области на пользовательском входном языке CMDL [8]. Причем данное описание расширено новыми разделами, определяющими способы и параметры визуализации данных. Разработан конвертер описания предметной области на CMDL во внутреннее представление вычислительной модели диспетчера и планировщик НПП [9]. Конвертер производит преобразование описания предметной области на CMDL в ее формализованное описание вида

$$Inputs(; Z_{Inputs}^{out}), o_1(Z_1^{in}; Z_1^{out}), o_2(Z_2^{in}; Z_2^{out}), \dots, o_m(Z_m^{in}; Z_m^{out}), Outputs(Z_{Outputs}^{in}),$$

где $Z = \{z_1, z_2, \dots, z_n\}$ – множество параметров модели, o_i – i -я операция модели, $Inputs$ – операция, определяющая входные параметры НПП, $Outputs$ – операция, определяющая выходные параметры НПП, $Z_i^{in}, Z_{Outputs}^{in} \subset Z, Z_i^{out}, Z_{Inputs}^{out} \subset Z, i = \overline{1, m}$.

Планировщик строит план решения задачи в виде направленного ациклического графа (англ., Directed Acyclic Graph – DAG) по непроцедурной постановке задачи вида «дано – вычислить» на вычислительной модели. Для построения плана решения задачи в планировщике применяются алгоритмы прямой и обратной волны [10].

Предложена спецификация параметров визуализации данных на языке JSON (<https://www.json.org/>). Данная спецификация включается в расширенное описание предметной области на CMDL. Спецификация определяет параметры визуализации данных, такие как название, тип, цвет диаграммы, максимальные и минимальные

значения данных по осям X и Y, название и местоположение легенды, размеры и начертания шрифтов и т.д. Для разных видов диаграмм (диаграмма, график, круговая диаграмма, график с накоплением и др.) созданы типовые спецификации.

В листинге 1 приведены фрагменты типовой спецификации для построения графика. Параметр min (max) – это минимальное (максимальное) значение данных на координатной оси. Если параметры min и max явно

не заданы, то подсистема визуализации автоматически определит и установит значения этих параметров. Параметр type – тип графика/диаграммы (line, spline, area, areaspine, column, bar, pie, scatter, gauge, arearange, areasplinerange и columnrange). Если в параметре series не заданы цвета рядов данных, то они определяются автоматически библиотекой Highcharts. При этом описание предметной области на CMDL расширено новым шаблоном, приведенным в листинге 2.

```
chart: { type: 'Тип графика/диаграммы',
  title { text: 'Название графика/диаграммы',
  verticalAlign: 'top | bottom | middle',
  align: 'left | right | center' },
  xAxis: {
    title: { text: 'Название оси X',
      style: { fontSize: 'Размер шрифта подписи по оси X' } },
    label: { style: { fontSize: 'Размер шрифта', color: 'Цвет шрифта' } }
    min: 0, max: 100 },
  yAxis: {
    title: { text: 'Название оси Y',
      style: { fontSize: 'Размер шрифта подписи по оси Y' } },
    min: 0, max: 100 },
  legend: {
    itemStyle: { "color": "цвет", "fontSize": "размер",
      "fontWeight": "начертание" },
    verticalAlign: 'top | bottom | middle', align: 'left | right | center' },
  marker: { symbol: 'Тип маркера', radius: 'Радиус',
    lineColor: 'Цвет контура маркера',
    fillColor: 'Цвет заливки маркера', },
  series: [{
    name: 'Название легенды ряда данных',
    marker: { symbol: 'Название вида маркера' },
    color: 'Цвет линии данных',
    lineWidth: 'Толщина линии данных' }],
}
```

Листинг 1. Фрагменты спецификации

Data Visualization <файл спецификации параметров визуализации расчетных данных на JSON> > <список сокращенных имен параметров предметной области приложения>

Листинг 2. Шаблон спецификации параметров визуализации расчетных данных

В качестве библиотеки для визуализации данных выбрана библиотека Highcharts 11.2.0 (<https://www.highcharts.com>) – одна из самых многофункциональных и популярных библиотек, таких как AnyChart, Chart.js, Chartist.js и др. (<https://habr.com/ru/articles/457946/>), написанных на языке JavaScript, для построения графиков и диаграмм в формате HTML с рендерингом в формате SVG (VML), в том числе в интерактивном режиме. Данная библиотека является достаточно легковесной, поддерживающей широкий спектр разнообразных типов визуализаций (графиков, диаграмм и др.) и обеспечивающей высокую производительность. Кроме того, функции библио-

теки поддерживают автоматический выбор некоторых стилей графических объектов, если параметры этих стилей не заданы в спецификациях. Визуализация с использованием формата SVG осуществляется в стандартных браузерах, таких как Chrome, Firefox, Internet Explorer и др. Библиотека Highcharts имеет открытый исходный код и может бесплатно использоваться в некоммерческих целях. Разработано инструментальное средство на языке PHP 7.4.3, которое позволяет вывести результаты визуализации на экран и/или сохранить в виде файла в формате PNG. Примеры визуализации графика и графика с накоплением представлены на рис. 1 и 2.

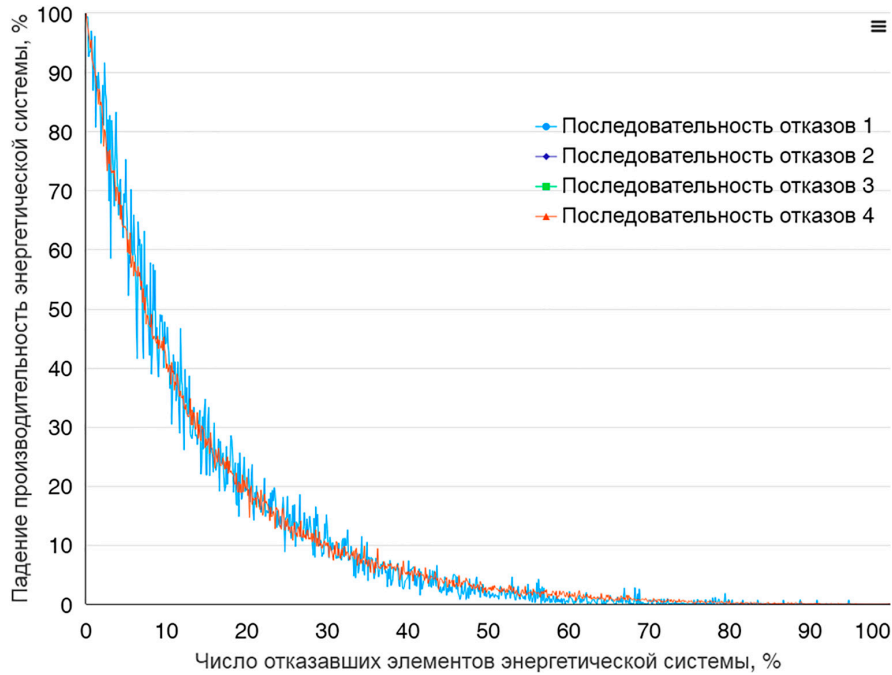


Рис. 1. Визуализация параметра на графике



Рис. 2. Визуализация параметра на графике с накоплением

Разработано инструментальное средство на языке Python 3.8.10, которое выполняет визуализацию НРП в виде двудольного ориентированного графа. В качестве библиотеки для визуализации НРП используется библиотека Graphviz 0.20.1 (<https://graphviz.org/>) – одна из самых популярных библиотек, таких как NetworkX, iGraph,

Graph-tool и др. (<https://tochmah.ru/graphviz-python-razberem-sya-v-bazovyh-funktsiyah-i-vozmozhnostyah-biblioteki/>), разработанных для работы с графами на языке Python.

Библиотека Graphviz представляет собой набор программ с открытым исходным кодом для верстки и визуализации графов. Данная библиотека поддерживает

веб-ориентированный интерактивный пользовательский интерфейс, включает набор вспомогательных инструментов и программных библиотек, а также обеспечивает привязку к различным форматам и языкам представления данных. В частности, Graphviz включает функции для изображения графов в форматах SVG для веб-страниц и Postscript для PDF-документов. Graphviz также поддерживает формат GXL, диалект языка XML.

Граф описывается на простом и интуитивно понятном языке DOT (Graph Description Language). Кроме того, Graphviz поддерживает автоматическую раскладку

графа с целью достижения оптимального расположения вершин и ребер графа на его изображении. Он также предоставляет подпрограммы для задания параметров конкретных графов, таких как цвета, шрифты, макеты узлов, стили линий, гиперссылки и пользовательские фигуры.

На основе описания графа на языке DOT разработан набор базовых спецификаций визуализации НРП. Описание предметной области на CMDL расширено дополнительным шаблоном, представленным в листинге 3. Пример визуализации НРП представлен на рис. 3.

Workflow Visualization <файл спецификации параметров визуализации НРП на JSON> > <список сокращенных имен НРП>

Листинг 3. Шаблон спецификации параметров визуализации НРП

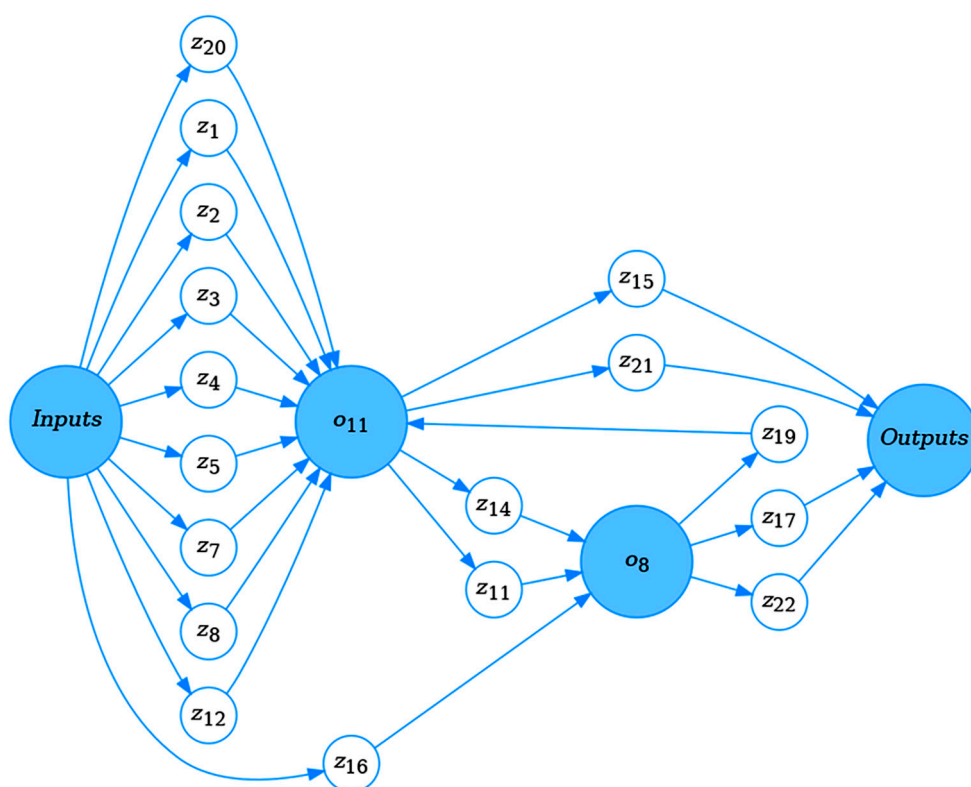


Рис. 3. Визуализация НРП

Результаты исследования и их обсуждение

В отличие от бизнес-рабочих процессов, которые обычно ориентированы на потоки управления, НРП, как правило, ориентированы на потоки данных, что определяет дополнительные требования по визуализации процессов и данных. К сожалению, в известных системах управления НРП

(например, Galaxy, BPEL Designer Project, GeoJModelBuilder и др. [4]) зачастую используется ограниченный набор встроенных типов диаграмм и графиков. Поэтому расширение функциональных возможностей Orlando Tools разработанными инструментальными средствами конфигурируемой и настраиваемой визуализации предопределяет его дополнительные преимущества

в сравнении с вышеупомянутыми системами. Разработанные спецификации параметров визуализации НРП и расчетных данных позволяют разработчику приложения выбрать и детализировать используемые графики и диаграммы для акцентирования необходимых аспектов схемы решения задачи и результатов ее выполнения для конечных пользователей приложения.

Данные средства допускают их автономное использование. В частности, они будут использованы в цифровых двойниках компонентов различных информационно-вычислительных систем. Их применение позволит наглядно отобразить текущее состояние и изменение характеристик вычислительных ресурсов и выполняемых в них процессов.

Заключение

В рамках исследования разработаны новые инструментальные средства визуализации научных рабочих процессов и расчетных данных, а также спецификация параметров визуализации. Результаты исследования обеспечивают повышение качественных характеристик процесса визуализации в сравнении с известными системами управления научными рабочими процессами.

Список литературы

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика. От данных к знаниям. СПб.: Питер, 2013. 622 с.
2. Романова И.К. Современные методы визуализации многомерных данных: анализ, классификация, реализация,

приложения в технических системах // *Машиностроение и компьютерные технологии*. 2016. № 3. С. 133–167. DOI: 10.7463/0316.0834876.

3. Bertin J. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, 1983. 415 p.

4. Feoktistov A., Edelev A., Tchernykh F., Gorsky S., Basharina O., Fereferov E. An Approach to Implementing High-Performance Computing for Problem Solving in Workflow-based Energy Infrastructure Resilience Studies // *Computation*. 2023. Vol. 11. № 12. P. 243. DOI: 10.3390/computation11120243.

5. Feoktistov A., Kostromin R., Sidorov I.A., Gorsky S.A. Development of Distributed Subject-Oriented Applications for Cloud Computing through the Integration of Conceptual and Modular Programming // *Proceedings of the 41st International Convention on information and communication technology, electronics and microelectronics (MIPRO-2018)*. Riejk: IEEE, 2018. P. 256–261. DOI: 10.23919/MIPRO.2018.8400044.

6. Feoktistov A., Gorsky S., Sidorov I., Bychkov I., Tchernykh A., Edelev A. Collaborative Development and Use of Scientific Applications in Orlando Tools: Integration, Delivery, and Deployment // *Communications in Computer and Information Science*. 2020. Vol. 1087. P. 18–32. DOI: 10.1007/978-3-030-41005-6_2.

7. Феоктистов А.Г., Костромин Р.О., Воскобойников М.Л., Ли-Дэ Д.И. Организация вычислительной среды разработки и применения научных рабочих процессов на основе контейнеризации // *Вычислительные технологии*. 2023. Т. 28, № 6. С. 151–164. DOI: 10.25743/ICT.2023.28.6.013.

8. Edelev A., Beresneva N., Gorsky S., Sidorov I., Feoktistov A. Representation of Subject Knowledge from the Field of Vulnerability Analysis of Energy Systems in Distributed Applied Software Packages // *Advances in Intelligent Systems Research*. 2019. Vol. 169. P. 184–188. DOI: 10.2991/iwci-19.2019.32.

9. Воскобойников М.Л. Планировщик схем решения задач для распределенных пакетов прикладных программ Свидетельство о государственной регистрации программы для ЭВМ № 2023685789. М.: Федеральная служба по интеллектуальной собственности (РОСПАТЕНТ), 2023.

10. Горбунов-Посадов М.М., Корягин Д.А., Мартынюк В.В. Системное обеспечение пакетов прикладных программ. М.: Наука, 1990. 208 с.