

## Интеграция инструментария научной визуализации SciVi с информационной системой Семограф

К.В. Рябинин<sup>1</sup>, Д.А. Баранов<sup>2</sup>, К.И. Белоусов<sup>1</sup>

belousovki@gmail.com|kostya.ryabinin@gmail.com|baranov@semograph.com

<sup>1</sup>Пермский государственный национальный исследовательский университет, Пермь, Россия;

<sup>2</sup>Оренбургский государственный университет, Оренбург, Россия

*Данная работа посвящена вопросу выделения из адаптивной мультиплатформенной клиент-серверной системы научной визуализации SciVi настраиваемого модуля визуализации, пригодного для включения в состав сторонних по отношению к этой системе программных продуктов, в частности, в состав информационной системы графосемантического моделирования Семограф. Выделенный модуль, являясь, по сути, самостоятельным программным средством, тем не менее остаётся частью системы SciVi, наследуя от неё управляемость онтологической базой знаний о доступных графических объектах, типах сцен и семантических фильтрах. Как следствие, он имеет возможность расширения функциональности и гибкой настройки на различные задачи визуализации. В этот модуль добавлены средства для отображения графов, успешно протестированные на реальных данных, полученных из информационной системы Семограф. Детально проработана концепция встраивания созданного модуля в состав системы Семограф. Следующим этапом работы в рамках описанного проекта является интеграция указанных программных продуктов, а также расширение системы SciVi новыми графическими возможностями, в частности, для поддержки отображения геоинформационных данных, включающих как пространственные, так и временные координаты.*

**Ключевые слова:** научная визуализация, информационная система, графосемантическое моделирование, Web-приложение.

## Integration of Scientific Visualization Toolset SciVi with Information System Semograph

K.V. Ryabinin<sup>1</sup>, D.A. Baranov<sup>2</sup>, K.I. Belousov<sup>1</sup>

belousovki@gmail.com|kostya.ryabinin@gmail.com|baranov@semograph.com

<sup>1</sup>Perm State University, Perm, Russia

<sup>2</sup>Orenburg State University, Orenburg, Russia

*This paper describes issues of extracting data rendering module from adaptive multiplatform client-server scientific visualization system SciVi to integrate it into third-party software. The concept of integration with graphosemantic modelling information system Semograph is elaborated. While the extracted rendering module is actually stand-alone application, it still remains to be a part of SciVi and therefore is driven by ontology knowledge base, which contains knowledge about graphical objects, scenes' types and semantic filters available. This module can be easily extended and set up for arbitrary visualization tasks. Currently it was tuned to visualize graphs and was tested on the real data obtained from Semograph. The next step of this project's work is integration of this module into Semograph and extending its rendering capabilities with the tools ready to visualize geoinformation data including spatial and temporal coordinates.*

**Keywords:** scientific visualization, information system, graphosemantic modelling, Web-application.

### 1. Введение

При разработке информационных систем (ИС), предназначенных для работы со сложноорганизованными данными в различных предметных областях, часто возникает задача поддержки в составе функциональности этих систем средств визуализации. Для максимально полного покрытия задач отображения данных зачастую требуется реализация модулей визуализации с нуля, либо серьёзная модификация существующих библиотек рендеринга. В обоих случаях трудоёмкость требуемых доработок оказывается достаточно высокой, особенно если структуры данных, подлежащих наглядному изображению, нетривиальны.

Возможным вариантом экономии трудовых затрат является использование адаптивных систем научной визуализации [3] – специализированного программного обеспечения (ПО), взаимодействующего с разрабатываемой ИС по какому-либо стандартизированному протоколу (ТСР/IP стек, общая база данных, текстовые файлы, в которые производится экспорт подлежащих визуализации данных и т. п.). Такой подход оказывается очень удобным в контексте узкоспециализированных решателей (англ.

*Solvers*) и хранилищ данных, то есть средств, играющих роль исключительно источников данных.

Однако в контексте построения сложных ИС, обладающих собственным многофункциональным интерфейсом обработки данных, такой подход оказывается недостаточно эргономичным: пользователю ИС придётся прерывать работу с ней всякий раз, когда ему необходима будет визуализация. В лучшем случае прерывание будет состоять в простом переходе к другому приложению на уровне смены активного окна, а в худшем – в совершении каких-то подготовительных действий, например, экспорта данных.

С точки зрения эргономики значительно лучше оказывается включить визуализатор в состав самой ИС так, чтобы пользователь имел возможность совершать все интересующие его действия как по обработке, так и по визуализации данных в рамках единого интерфейса.

Для обеспечения такой возможности авторами было принято решение доработать предложенную ранее концепцию построения адаптивных мультиплатформенных систем научной визуализации [3] с учётом поддержки встраивания этих систем как отдельных модулей в состав другого ПО.

В результате предыдущих исследований была создана адаптивная система научной визуализации SciVi [8], основанная на принципах онтологического инжиниринга [1, 9]. Её поведение гибко конфигурируется посредством пополнения и/или изменения онтологической базы знаний о языках программирования совместимых с ней решателей, поддерживаемых семантических фильтрах входных данных, визуальных объектах и типах графических сцен. Благодаря этому данная система может быть быстро настроена на решение задач визуализации из любой предметной области и на взаимодействие с широким спектром различных источников данных.

Параллельно с разработкой системы SciVi развивалась ИС Семограф [5]. ИС Семограф предназначена для решения широкого спектра научных, образовательных и прикладных задач, связанных с обработкой и анализом языковых и текстовых данных. Типовыми задачами, решаемыми с помощью инструментария системы, являются: создание классификаторов и тезаурусов предметных областей, проведение психолингвистических, социолингвистических и пр. экспериментов (а также анализ полученных данных), создание и разметка выборок и корпусов текстов, построение информационных моделей предметных областей, и другие задачи, возникающие в ходе анализа текстового контента.

Информационная система реализована как Web-приложение и в настоящее время состоит из модулей, обеспечивающих следующую функциональность:

1. Поиск (с помощью поискового робота и парсера, созданного на основе фреймворк Scrapy на языке Python).
2. Импорт данных (с помощью поискового сервера на основе Apache Solr; кроме того поддерживается импорт файлов формата CSV).
3. Экспорт результатов во внешние приложения, в частности, R (статистическую среду анализа данных), Gephi (средство построения и анализа графов), а также в табличные форматы.
4. Администрирование и организационная деятельность (добавление приглашённых пользователей системы к проектам, управление правами доступа, создание открытых проектов, коммуникация).
5. Исследование (набор инструментов проведения экспериментов, анализа языкового контента, разметки корпусов, классификационной деятельности и др.).

ИС Семограф имеет API, который может использоваться для работы с внешними приложениями или разработки приложений на основе этой ИС с расширенной в каком-либо аспекте функциональностью.

Целью данной работы является интеграция систем SciVi и Семограф для создания отдельного модуля научной визуализации в ИС Семограф. Внедрение инструментария, позволяющего пользователям работать с языковыми и текстовыми данными, обогащёнными наглядной визуализацией, позволит реализовать возможность осуществления полного цикла исследования: от постановки проблемы до генерации отчёта по научному, образовательному или прикладному проекту.

## 2. Встраиваемая версия системы SciVi

Система научной визуализации SciVi состоит из серверной и клиентской частей. Сервер служит в первую очередь для связи с источником данных и их тяжеловесной предобработки (если таковая требуется при решении конкретной задачи визуализации). Клиент отвечает за связь с пользователем, предоставляя в первую очередь графический интерфейс системы визуализации. На данный момент активно развиваются толстый и тонкий клиенты.

Рендеринг данных адаптивно распределяется между клиентом и сервером с учётом производительности клиентской машины, загрузки сервера и скорости сетевого соединения.

Для включения системы SciVi в виде отдельного компонента в состав стороннего по отношению к ней ПО требуется возможность объединения в рамках этого компонента необходимой функциональности сервера и клиента. Это позволит избавиться от внешних зависимостей и минимизировать накладные расходы на пересылку данных внутри итоговой ИС.

Сервер и клиент SciVi имеют модульную архитектуру. Их функциональные блоки полностью изолированы друг от друга и взаимодействуют между собой посредством вызова интерфейсных методов, использования функций обратного вызова (англ. *Callback*) и передачи декларативных описаний (онтологических профилей [4]) обрабатываемых сущностей. Благодаря этому сборка на основе системы SciVi встраиваемых компонентов визуализации оказывается возможной и не требует глубинных изменений реализованной и отлаженной ранее базовой функциональности, такой, например, как механизмы взаимодействия с источником данных, рендеринг графических сцен и т.п.

## 3. Концепция интеграции с системой Семограф

ИС Семограф реализована с использованием современных практик разработки Web-приложений. В основе лежит принцип SPA (англ. *Single Page Application* – одностраничное приложение), что означает отсутствие лишних переходов между страницами, характерных для большинства обычных сайтов. Вместо этого приложение подгружает необходимые данные по мере надобности с помощью AJAX (англ. *Asynchronous Javascript and XML*). Благодаря этому достигается высокая отзывчивость пользовательского интерфейса при минимальном использовании сетевого трафика.

В качестве ключевых решений для организации клиентской части используются библиотеки React и Redux, для серверной – фреймворк Flask и СУБД PostgreSQL. Клиентская часть использует стандарт ECMAScript 6 и при сборке преобразуется в ECMAScript 5 с помощью пакета Babel. Для сборки клиентской части используется Webpack. Такой подход повышает скорость разработки, но накладывает ряд ограничений на используемые библиотеки, например, они должны поддерживать экспорт методов в формате CommonJS или совместимом с ним.

Для интеграции системы SciVi с ИС Семограф предполагается разработка компонента (в терминологии React), адаптирующего программный интерфейс тонкого клиента SciVi к стандартам приложений, разработанных с использованием React и Redux. При таком подходе данные будут следовать стандартному циклу ИС Семограф – загружаться по требованию с сервера через API системы Семограф, помещаться в состояние Redux, при необходимости обрабатываться и передаваться в новый компонент. Большим преимуществом в данном случае является отсутствие необходимости модифицировать сервер или его API, а также возможность использовать существующую систему контроля прав доступа.

Жизненный цикл компонента-адаптера должен включать конструирование, передачу данных и управление модулем визуализации на основе тонкого клиента SciVi. Также данный компонент должен обеспечить передачу управляющих команд от модуля визуализации в основное приложение посредством механизма функций обратного вызова. Таким образом, компонент должен реализовать как

функции представления, так и контроллера (согласно паттерну проектирования Model-View-Controller).

Полученный компонент может быть обособлен от ИС Семограф, унифицирован и собран в виде отдельного модуля, что позволит применять его в других приложениях, базирующихся на React и Redux.

#### 4. Пополнение набора средств визуализации системы SciVi

Многие аналитические задачи, решаемые в ИС Семограф, предполагают работу с данными о связности различных сущностей. Такие данные нагляднее всего представимы в виде графов [6, 7].

Для обработки и отображения графов в систему SciVi были добавлены новые семантические фильтры (позволяющие по различным правилам вычислять положения вершин и характеристики связей графа) и визуальные объекты (позволяющие отображать именованные вершины и связи между ними). Добавление осуществлено путём пополнения соответствующих онтологий, входящих в базу знаний системы, а также путём подключения дополнительной библиотеки визуализации.

Каждая вершина визуализируемого графа характеризуется названием, идентификатором кластера, в который входит связанная с вершиной сущность, и весом. Кластеризация сущностей, отображаемых в виде графа, происходит на стороне ИС Семограф; на данный момент эта функциональность находится в стадии разработки.

Названия вершин отображаются в виде текстовых меток, повернутых под определённым углом, который для каждой вершины определяется вычисляющим её положение фильтром.

Принадлежность вершины к конкретному кластеру отображается цветом: каждому кластеру назначается свой определённый цвет из заданной по умолчанию таблицы; далее у пользователя есть возможность изменять эти цвета. Вершины располагаются радиально и группируются по принципу принадлежности к одному кластеру.

Вес вершин отображается в виде гистограммы, каждый столбец которой рисуется под текстовой меткой соответствующей ему вершины.

Каждая связь в графе характеризуется весом и отображается в виде квадратичной кривой Безье, толщина которой пропорциональна её весу. Цвет связи определяется цветами соединяемых ей вершин (если вершины имеют разные цвета, связь окрашивается градиентом).

В дальнейшем набор атрибутов вершин и связей может быть расширен.

Визуализатор графов поддерживает выделение вершины (с автоматическим подсвечиванием более ярким цветом всех инцидентных ей связей и всех смежных с ней вершин), а также произвольное перемещение вершин курсором мыши (либо пальцем в случае сенсорного экрана мобильного устройства). Таким образом пользователь может осуществлять перегруппировку вершин графа и исследовать их связность.

Настройка конкретного способа визуализации осуществляется пользователем путём редактирования диаграммы потока данных средствами высокоуровневого графического Web-интерфейса [8]. Эта диаграмма определяет последовательность шагов предобработки данных (семантической фильтрации), а также связь данных с теми или иными атрибутами поддерживаемых системой визуализации графических объектов. Набор доступных для построения диаграммы вершин, олицетворяющих семантические фильтры, графические объекты и типы сцен, описан в соответствующих онтологиях системы.

На рис. 1 представлен пример диаграммы потока данных, задающей настройку визуализации графа с радиальным расположением вершин.

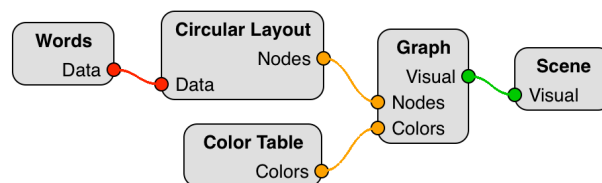


Рис. 1. Диаграмма потока данных, определяющая настройки визуализации графа понятий в системе научной визуализации SciVi.

Результат визуализации графа, состоящего из 166 вершин и 4917 связей, показан на рис. 2. На текущий момент это единственный реализованный способ визуализации графов, однако в дальнейшем, при необходимости, функциональность SciVi по отображению связанных и иерархических сущностей может быть расширена.

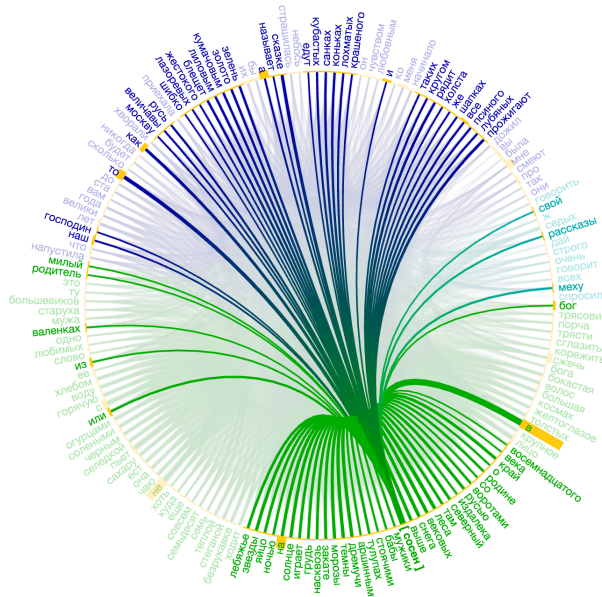


Рис. 2. Визуализация графа понятий в системе научной визуализации SciVi, построенного по данным, извлечённым из ИС Семограф.

Входные данные представляют собой результаты психолингвистического эксперимента по тематическому картированию текста. Информантам предъявляется текст и ставится задача определить его тему, выделить микротемы и к каждой микротеме отнести слова, представляющие её в тексте. Эксперимент проводится в ИС Семограф; одним из результатов исследования становится генерация семантической карты текста, представляющей собой матрицу  $n \times n$ , где  $n$  – число слов в тексте. В ячейках отображаются значения совместной встречаемости двух слов во всех выделенных микротемах. Особенностью матрицы является фиксированный порядок столбцов и строк, отражающий порядок слов в анализируемом тексте.

В результате эксперимента с визуализацией было выяснено, что приемлемой производительности (отклика быстрее, чем за 125 мс [2]) на желаемых объёмах данных удаётся достичь только используя аппаратное ускорение рендеринга. Так, например, использование для визуализации графа стандарта SVG (посредством библиотеки D3.js) в браузерах Mozilla Firefox (версия 51) и

Google Chrome (версия 56) на компьютере с процессором Intel Core i7 (4 ядра тактовой частотой 2 ГГц, без поддержки гипертрединга), объёмом оперативной памяти 8 Гб и видеокартой Intel Iris Pro с объёмом видеопамати 1536 Мб перерисовка графа при перемещении курсора над текстовыми метками длится в среднем 1000 мс.

В связи с этим для рендеринга графа было решено использовать библиотеку PixiJS, предоставляющую удобную в использовании обёртку над низкоуровневым графическим API WebGL. Данная библиотека позволила добиться средней скорости отклика в 20 мс, что вполне удовлетворяет требованиям, предъявляемым к производительности.

## 5. Заключение

В результате проделанной работы сформулирована концепция встраивания системы научной визуализации SciVi в сторонние по отношению к ней приложения, а также произведены подготовительные действия по интеграции этой системы с ИС Семограф. Детально проработана последовательность шагов, необходимая для такой интеграции. В систему SciVi добавлены новые средства визуализации, служащие для наглядного представления данных, обрабатываемых в ИС Семограф, в частности средства для отображения графов. Из тонкого клиента системы SciVi выделен автономный модуль, готовый к встраиванию в сторонние Web-приложения. В целях обеспечения высокой эффективности рендеринга в нём использованы механизмы аппаратного ускорения графики, основанные на API WebGL. Работоспособность данного модуля успешно протестирована на реальных данных, извлечённых из ИС Семограф.

Полученный модуль визуализации может быть использован как автономное программное средство, но при этом остаётся составной частью системы SciVi. Это позволяет развивать его функциональность одновременно с развитием системы SciVi, тестировать новые средства визуализации, используя адаптационные механизмы этой системы, пополнять репозиторий доступных визуальных объектов, типов графических сцен и семантических фильтров путём расширения онтологий, составляющих базу знаний SciVi.

В дальнейшем планируется встроить созданный модуль визуализации в ИС Семограф, а также по мере необходимости пополнять его новыми графическими возможностями. В частности, приоритетным является реализация средств отображения геоинформационных данных с поддержкой временных срезов.

## 6. Благодарности

Работа выполнена в рамках государственного задания Минобрнауки России (проект 34.1505.2017/4.6).

## 7. Литература

- [1] Chuprina S., Nasraoui O. Using Ontology-based Adaptable Scientific Visualization and Cognitive Graphics Tools to Transform Traditional Information Systems into Intelligent Systems // *Scientific Visualization*. – М.: National Research Nuclear University "MEPhI", 2016. – Q. 1, V. 8, No1. – С. 23–44.
- [2] Keval H., Sasse M.A. To catch a thief – you need at least 8 frames per second: the impact of frame rates on user performance in a CCTV detection task // *Proceedings of the 16th ACM international conference on Multimedia*. – ACM, 2008. – P. 941–944.

- [3] Ryabinin K., Chuprina S. Development of Ontology-Based Multiplatform Adaptive Scientific Visualization System // *Journal of Computational Science*. – Elsevier, 2015. – Vol. 10. – P. 370–381.
- [4] Ryabinin K., Chuprina S. High-Level Toolset For Comprehensive Visual Data Analysis and Model Validation // *Procedia Computer Science*. – Elsevier, 2017. – Vol. 108. – P. 2090–2099.
- [5] Баранов Д.А., Белоусов К.И., Влацкая И.В., Зелянская Н.Л. Система графосемантического моделирования [программа для ЭВМ] – М.: Свидетельство о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011.
- [6] Касьянов В.Н., Золотухин Т.А. Visual Graph — система для визуализации сложно структурированной информации большого объема на основе графовых моделей // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2015. – К. 4, Т. 7, №4. – С. 44–59.
- [7] Касьянов В., Касьянова Е. Визуализация информации на основе графовых моделей // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2014. – К. 1, Т. 6, №1. – С. 31–50.
- [8] Рябинин К.В., Чуприна С.И., Бортников А.Ю. Автоматизация настройки систем научной визуализации на специфичу разнообразных источников данных // *Научная визуализация*. – М.: Национальный исследовательский ядерный университет МИФИ, 2016. – К. 4, Т. 8, №4. – С. 1–14.
- [9] Чуприна С.И., Зиненко Д.В. ОНТОЛИС: адаптируемый визуальный редактор онтологий // *Вестник Перм. нац. исследоват. поли-техн. ун-та*. – 2013. – №3(22). – С. 106–110.

## Об авторах

Константин Валентинович Рябинин, к.ф.-м.н., доцент кафедры математического обеспечения вычислительных систем Пермского государственного национального исследовательского университета. Его электронный адрес [kostya.ryabinin@gmail.com](mailto:kostya.ryabinin@gmail.com).

Дмитрий Александрович Баранов, преподаватель кафедры компьютерной безопасности и математического обеспечения информационных систем Оренбургского государственного университета. Его электронный адрес [baranov@semograph.com](mailto:baranov@semograph.com).

Константин Игоревич Белоусов, докт. филол. н., профессор кафедры теоретического и прикладного языкознания Пермского государственного национального исследовательского университета. Его электронный адрес [belousovki@gmail.com](mailto:belousovki@gmail.com).