

- активное участие в разработке системы заинтересованных сторон: Министерства образования, уполномоченных представителей УВО и научных организаций, потенциальной организации-администратора Системы и разработчиков;
- широкое использование новых технических средств и современных высокоэффективных информационных технологий;
- управление информацией, представляемой в Системе, в соответствии с согласованными регламентами ее получения, изменения, обмена и распространения (разграничение прав доступа);
- масштабируемость Системы – возможность ее дальнейшего развития без необходимости внесения значительных изменений в уже существующую структуру;
- возможность отслеживать не только текущие (годовые) значения показателей, но и динамику их изменения по временным интервалам;
- издержки на последующее сопровождение и обслуживание Системы должны быть минимальны (не более чем 1 специалист (администратор) при штатном функционировании Системы).

Создание и внедрение Системы позволит обеспечить:

- формирование единого информационно-аналитического пространства в области научно-исследовательской и инновационной деятельности в рамках общей системы информационной поддержки принятия управленческих решений на уровне Министерства образования;
- улучшение качества и сокращение сроков подготовки и принятия управленческих решений по различным направлениям научно-исследовательской и инновационной деятельности в системе Министерства образования;
- повышение эффективности использования информации за счет оптимизации ее процесса прохождения от первоисточника к основному потребителю – уполномоченным представителям Министерства образования;
- повышение оперативности и качества представляемой Министерству образования подведомственными УВО и научными организациями аналитической информации о ходе выполнения государственных научно-исследовательских и научно-технических программ (ГПНИ, ГНТП);
- снижение организационных издержек и документооборота Министерства образования по взаимодействию с подчиненными организациями в части сбора аналитической информации по различным аспектам научно-исследовательской и инновационной деятельности.

Важно отметить, что функционирование Системы мониторинга научно-исследовательской и инновационной деятельности в сфере высшего образования позволит обеспечить оперативной и достоверной информацией как непосредственно Министерство образования, так и управленческо-административный персонал на уровне подведомственных УВО и научных организаций.

УДК 004.514

ВИЗУАЛИЗАЦИЯ ДЕНДРОГРАММ ПРИ АНАЛИЗЕ БОЛЬШОГО ОБЪЕМА ДАННЫХ

Иванова Л. В.

ВГУ имени П. М. Машерова

e-mail: portnenko_lv@mail.ru

***Summary.** Studied the methods of hierarchical cluster analysis and the features of their application for processing a big data for its visualization in the most convenient way as results of analysis (clustering) of these data. The purpose of this work is to develop a JavaScript programming language library for visualizing dendrograms as a result of hierarchical cluster analysis.*

В рамках данного исследования изучались методы иерархического кластерного анализа и особенности их применения для обработки большого объема данных для последующей визуализации наиболее удобным способом результатов анализа (кластеризации) этих данных.

Целью данной работы является разработка библиотеки языка программирования JavaScript для визуализации дендрограмм, получаемых как результат иерархического кластерного анализа.

Так в большинстве практических задач обработки данных применяются методы статистического и кластерного анализа. Визуализация, применяемая в исследованиях по математической статистике, широко представлена в различных специализированных программах, таких как Statistica, MathCAD или Maple; в программах общего назначения, таких как MS Word или MS Excel; в библиотеках для разных языков программирования, с помощью которых визуализацию данных можно осуществлять в разрабатываемых системах. Но визуализация, применяемая в кластерном анализе, особенно в иерархическом кластерном анализе, представлена не так широко, а в тех системах, где она присутствует, например, в программе Statistica или в библиотеках таких языков программирования, как Python и R, удобство представления информации на так называемых дендрограммах невысокое и не позволяет гибко настраивать это представление.

Так при планировании нагрузки преподавателей кафедры с помощью генетического алгоритма распределяются отдельные элементы нагрузки между преподавателями кафедры. Количество вариантов распределения нагрузки оказывается огромным, поэтому на первом этапе элементы нагрузки кластеризуются с целью выявления отдельных групп элементов нагрузки и формирования правил селекции популяций для генетических алгоритмов.

Таким образом, основной задачей является разработка библиотеки визуализации дендрограмм с удобной настройкой получаемых изображений без необходимости повторного анализа существующих данных.

Визуализация будет выполняться в браузерах средствами языка программирования JavaScript. Поэтому в качестве формата входных данных для этой библиотеки используется формат JSON. Но требуется также предусмотреть в будущем возможность гибкого добавления других форматов данных, прежде всего XML.

Так как иерархическая кластеризация может выполняться для очень большого количества исходных объектов, а для анализа результатов этой кластеризации может быть достаточно нескольких верхних уровней иерархии, то в библиотеке визуализации реализована возможность получения данных частично с возможностью последующей подгрузки остальных данных по запросу пользователя. Построенная дендрограмма позволяет сворачивать и разворачивать содержимое каждого кластера. При разворачивании кластера, если данные о подкластерах отсутствуют, происходит подгрузка данных с сервера.

Важными также являются требования по визуальному оформлению дендрограммы. Помимо цветов и стилей фигур и текста на дендрограмме реализована возможность менять ориентацию (вертикальную и горизонтальную) и направление диаграммы.

Отметим, что язык программирования JavaScript, являясь интерпретируемым языком программирования, имеет проблемы с быстродействием. Поэтому оценивалась производительность скриптов при прорисовке изображений на стандартном HTML-компоненте Canvas.

В качестве параметров для оценки эффективности выступало время выполнения скрипта в зависимости от количества отображаемых объектов, а также объем используемой памяти. По последнему параметру важно не только оценить объем выделяемой при прорисовке графических примитивов памяти, но и исследовать скорость освобождения памяти при удалении прорисовываемого объекта. Ввиду специфики решаемой задачи по визуализации дендрограмм для большого количества исходных объектов, как отмечалось выше, прорисовывать планируется лишь несколько верхних уровней иерархии с реализацией динамического развертывания и свертывания поддеревьев нижележащих уровней. При таком подходе как раз и может возникнуть вопрос освобождения памяти при сворачивании поддерева.

Проведенные исследования показали, что объемы памяти для визуализации графических объектов сопоставимы с количеством логических объектов, визуализируемых на дендрограмме, и линейно зависят от этого количества.

Скорость прорисовки объектов, как и скорость выделения памяти позволяют реализовать анимацию при визуализации дерева со скоростью не ниже 25 кадров в секунду.

Скорость удаления памяти, наоборот, не очень высока. Связано это со спецификой реализации механизма сборки мусора в большинстве интерпретаторов JavaScript. При реализации библиотеки в будущем планируется в связи с этим освободить ранее прорисованные, но в настоящий момент скрытые объекты только после отсутствия активных действий пользователя с дендрограммой в течении некоторого интервала времени, который можно будет задавать при использовании разработанной библиотеки.

При этом отсутствие настройки этого времени будет восприниматься как ситуация, в которой уничтожать объекты не требуется и, таким образом, все объекты будут сохраняться в памяти. Использование такого функционала приведет к использованию большего объема памяти при визуализации больших дендрограмм. Однако, по субъективной оценке, вряд ли пользователю придется часто сталкиваться с ситуацией, когда необходимо отобразить дендрограмму с большим количеством объектов и глубокой иерархией кластеров, и при этом понадобится развернуть все ветви построенного дерева, что приведет к постепенной загрузке всех данных. В таком случае скорость отображения дерева будет крайне низкой.

Если же размер дерева небольшой, или пользователь будет выборочно разворачивать лишь некоторые ветви большого дерева, объем используемой памяти будет небольшим, а наличие кэшированной информации позволит быстрее и более плавно выполнять анимацию сворачивания и разворачивания ветвей дерева.

При этом у разработчиков, которые будут применять библиотеку для визуализации данных, остается возможность настраивать таймаут на удаление данных в свернутых ветвях. При этом у каждой свернутой ветви этот таймаут независимый, что позволит через некоторое время удалять объекты в свернутой части дендрограммы, если эта часть не отображается достаточно длительное время.

Таким образом была разработана библиотека и оценена эффективность подхода по визуализации дендрограмм средствами языка программирования JavaScript. Можно сделать вывод о том, что разработанный подход эффективен и может быть использован на практике.

Список использованных источников

1. Жамбю, М. Иерархический кластер-анализ и соответствия / М. Жамбю // Финансы и статистика. – Москва. – 1988. – С. 342.
2. Фрисби, М. JavaScript для профессиональных веб-разработчиков / М. Фрисби. – Санкт-Петербург. – 2022. – С. 1167.

УДК 004.8

ПЕРИОД ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ТРУДНОСТИ И РЕШЕНИЯ ДЛЯ РАЗВИТИЯ ВИРТУАЛЬНЫХ ВЕДУЩИХ

Жэнь И

*Белорусский государственный университет
e-mail:936400630@QQ.COM*

Summary. *The advent of the age of artificial intelligence is quietly changing and influencing all walks of life. As a new form of artificial intelligence technology applied in the broadcasting and hosting industry, virtual hosts have broken the pattern of real hosts dominating the world by virtue of the inherent advantages of all-day and multilingual broadcasting. Virtual hosts generally have realistic disputes such as “inhumanity” and “expensive research and development”, which hinders its long-term development. Helping virtual hosts out of the development predicament has become a key problem to be solved urgently in the industry.*