

УДК: 004.822 МЕТОДЫ ВИЗУАЛЬНОГО УПРАВЛЕНИЯ ОНТОЛОГИЧЕСКИМИ ДАННЫМИ В ТРЕХМЕРНОМ ВИРТУАЛЬНОМ ПРОСТРАНСТВЕ ОСНОВАННЫЙ НА МЕТРИКАХ ТРУДОЕМКОСТИ ПОИСКА

Аннотация. Цель исследования – сокращение времени и трудоёмкости построения визуальных репрезентаций онтологических данных в трехмерном виртуальном пространстве для навигации и поиска по графам знаний.

В связи с ростом потребности в визуальной репрезентации онтологий, встаёт задача оценки качества создаваемых представлений. При этом качество – это субъективное понятие. В данной работе мы описываем модель представления онтологических данных и соответствующие ей метрики производительности, которые позволяют производить объективную оценку качества представления. Затем на основе проделанной работы мы разрабатываем метод репрезентации в трехмерном пространстве, учитывающий метрики производительности поиска и управления данными.

Введение. С ростом потребность в визуальном представлении онтологий, растёт и количество инструментов, решающих подобные задачи. Задача представления может быть решена множеством различных способов от представления в сыром текстовом виде **turtle** или **jsonld** формата, до систем представлений включающих таблицы, графики и node-link-диаграммы. В связи с этим встаёт задача оценки качества создаваемых представлений. Качество – это субъективное понятие. В данной работе мы описываем модель представления онтологических данных и соответствующие ей метрики производительности, которые позволяют производить объективную оценку качества представления.

Тема разработки метрик визуализации затрагивалась во многих работах, таких как «Evaluation of clusterings--metrics and visual support», «Project dashboard: Concurrent visual representation method of project metrics on 3D building models» и «Metrics for effective information visualization». В этих работах акцент делался на оценку параметров конечной визуализации, а не на представление в целом. Здесь нужно уточнить, что под понятием «представление» мы понимаем не визуализацию как конечный продукт, а объект, имеющий состояние визуализации, и инструменты позволяющие менять это состояние.

Так, например, в статье «Quality metrics for information visualization» авторы предлагают оценивать качества репрезентации по трём входным параметрам: Данные, Пользователь, Задача. Тем самым в статье озвучивая тот факт, что визуализация, данных, не может быть качественной, если она не позволяет решить изначально поставленную перед ней задачу. Используя данную идею как основу, мы постарались выделить основные сценарий использования представлений онтологических данных. В статье «The Value of Visualization» описывалась суть взаимодействия человека с представлением. Любое взаимодействие производится для получения некоторого прироста знания, которое зависит от пользователя, его изначальных знаний и функции восприятия. Все эти параметры сложны в формализации, однако, дают хороший пример того, в каком направлении нужно вести разработку. Основываясь на эти двух статья «Quality metrics for information visualization» и «The Value of Visualization» мы разработали свою модель, метрики и метод представления учитывающий их.

Основная часть. Наша работа сосредоточена на представлении онтологических данных, которые могут быть представлены в виде онтологического графа. Онтологический граф – это конечный набор триплетов. Данные располагаются в пространстве данных и формируют его. В модели существует понятие измерение данных. Онтологические данные – это онтологический граф.

Данные мультимерны и пространство данных представляет собой набор измерений. Каждое измерение соответствует предикату из онтологии. То есть измерение – это множество возможных объектов триплетов с соответствующим предикатом.

Любое представление данных, согласно описываемой модели, представляет собой кортеж из двух элементов: функция визуализации данных и функция управления.

Функция визуализация, как математический объект, представляет собой функцию отображения данных, находящихся в рамках окна поиска, на двухмерное пространство экрана

или трехмерное пространство виртуальной реальности. Визуализация же — это результат подобного отображения, — проекция пространства на двухмерное пространство экрана или трехмерное пространство виртуальной реальности. В каждый момент времени при работе с представлением состояние пользователя описывается объектом, который называется «окно поиска», где окно поиска — это набор пар координата и значение диапазона по каждому измерению данных, которые определяют положение и форму окна в пространстве данных. Кроме окна поиска положение пользователь в пространстве данных характеризуется точкой поиска. В то время как за перемещение окна поиска отвечают инструменты управления, за перемещение точки поиска отвечает сам пользователь, осуществляя поиск данных с использованием сформированной визуализации.

Функция управления представляется инструментами управления. Данная функция принимает в качестве параметра текущее состояние окна поиска и пользовательский ввод. Результатом функции всегда является новое положение окна.

Любое представление строится для достижения определенных целей. В упомянутой «Quality metrics for information visualization», приведены несколько примеров этого: Определение смежности вершин или достижимости, определение расстояния между двумя вершинами, определения размера и структуры, определение принадлежности к группе, плотность вершин в подграфе, сравнение подграфов. Большинство задач можно свести к задачам поиска одного, двух или n -го количества элементов в пространстве данных. Подсчёт количества элементов — это поиск каждого элемента на экране, определение смежности двух элементов, это поочередный поиск каждого из них в пространстве данных, установление типа элемента — это поиск точного положения элемента в пространстве данных, в частности поиск в измерении данных, которое соответствует типам. Конечным шагом поиска всегда идёт шаг Восприятие — на этом шаге идёт восприятие данных пользователем, то есть увеличение количества знаний. Процесс поиска тесно связан с перемещением в мультимерном пространстве. С помощью инструментов управления пользователь перемещает окно поиска по измерениям соответствующим инструментам поиска, то есть функциям управления, затем, когда достигнуто максимальное приближение, при минимальной ширине окна, пользователь переходит к шагу «Осмотр» и начинает изучать данные, лежащие в окне, перемещаясь от одной позиции в многомерном пространстве данных к другой. Позицию пользователя можно описать точкой, в которую устремлен взгляд пользователя. По результатам осмотра принимается решение об окончании или продолжении поиска. Процесс поиска информации состоит из непустой последовательности четырех типов шагов: Пользовательский ввод, Визуализация данных инструментом, Самостоятельный поиск или осмотр пользователем результата, Восприятие пользователем данных. Количество и порядок шагов могут варьироваться от представления к представлению.

Процесс поиска характеризуется тремя величинами: скорость поиска, дистанция поиска, трудоемкость поиска.

Трудоемкость — это последовательность простейших логических операций, таких как: клик мышью, ввод ключа поиска в текстовое поле, переключение флажка фильтра, распознавание элемента пользователем.

Расстояние — это длина отрезка прямой в мультимерном пространстве данных, соединяющей две точки поиска.

Скорость — это то насколько быстро изменяется координата точки поиска в мультимерном пространстве — является отношением расстояния, на которое переместилась точка поиска к затраченной трудоемкости, то есть расстояние, на которое перемещается точкой поиска за одну простейшую операцию.

Оценка представления складывается из оценки визуализации и оценки элемента управления. Оценка представления — это сумма оценок качества функции визуализации и качества функции управления. Оценки строятся для всех возможных операций поиска в пространстве данных D . Сумма оценок всегда больше или равна нулю и меньше или равна единице.

Качество элемента управления для представления складывается из среднего значения качества каждого инструмента управления, где качество каждого отдельного элемента управления – это скорость выполнения шага поиска соответствующего этому элементу управления, для всех возможных операций поиска, умноженная на квадрат охвата пространства. Где охват пространство — это отношение недоступного пространства поиска к доступному.

Качество визуализации для представления складывается из среднего значения качества каждого аспекта визуализации, для всех возможных операций поиска, где качество каждой отдельной визуализации – это скорость выполнения шага осмотра, следующего за шагом визуализации.

При разработке метода, согласно модели, были выделены следующие измерения данных: три измерения координат и масштаб элементов онтологии (соответствующие четырем самым релевантным атрибутам), размерность типов элементов, размерности объектных свойств и свойств данных. Им соответствуют инструменты навигации в трехмерном пространстве, панель фильтров по типа элементов, панель связей между элементами и панель фильтров по атрибутам, включающая поиск по ключу, соответствующий свойствам данных. За основу взята нотация Ontodia, адаптированная под трехмерное пространство. Ленивая визуализация, согласно методу, должна быть применена для уменьшения начальной ширины окна поиска, а также для увеличения покрытия в условиях ленивой визуализации, применены методы визуализации пустых узлов, описанные в статье «Подход к обработке пустых узлов при порционной визуализации данных на примере инструмента ONTODIA».

Выводы. При использовании разработанных метрики трудоемкости, акцент смещается с трудно формализуемых величин таких как качество и красота визуализации на счётные величины такие как скорость поиска, трудоемкость и время, затраченное на поиск, что позволяет выделить узкие места при разработке средств репрезентации онтологических данных, тем самым облегчив процесс разработки решений.

Раздьяконов Д. С. (автор)

Подпись

Муромцев Д. И. (научный руководитель)

Подпись