ПРИКЛАДНЫЕ ОНТОЛОГИИ ПРОЕКТИРОВАНИЯ

УДК 004.822

Научная статья

DOI: 10.18287/2223-9537-2022-12-1-41-56

Построение хранилища обобщённых вычислительных экспериментов на основе онтологического подхода

© 2022, А.Г. Подвесовский 🖂, Д.А. Коростелёв, Е.А. Лупачёв, Н.В. Беляков

Брянский государственный технический университет, Брянск, Россия Институт прикладной математики им. М.В. Келдыша РАН, Москва, Россия

Аннотация

Рассмотрена задача организации хранения и доступа к информации об обобщённых вычислительных экспериментах, связанных с математическим моделированием реальных физических процессов. Обобщённый вычислительный эксперимент предполагает многократное решение задачи численного моделирования при различных наборах значений определяющих параметров модели, что позволяет получить решение для некоторого класса задач математического моделирования, заданного в многомерном пространстве параметров. Предложен подход к построению хранилища данных, описывающих обобщённые вычислительные эксперименты, включая историю их планирования и корректировки сценариев проведения. Обсуждаются особенности хранения обобщённых вычислительных экспериментов, положенные в основу построения логической и физической моделей хранилища. Для организации эффективного взаимодействия пользователя с хранилищем предложена онтологическая модель обобщённого вычислительного эксперимента и связанная с ней структура метаданных, на основе которой реализуется механизм формирования запросов к хранилищу в терминах предметной области моделирования. Обсуждается применение разработанного подхода для хранения и анализа данных об обобщённом вычислительном эксперименте, связанном с оценкой точности численных моделей программной платформы *OpenFOAM* для трёхмерной задачи сверхзвукового обтекания клина.

Ключевые слова: обобщённый вычислительный эксперимент, многомерные данные, представление данных, хранение данных, онтология, тезаурус, верификация численных методов.

Цитирование: Подвесовский А.Г., Коростелёв Д.А., Лупачёв Е.А., Беляков Н.В. Построение хранилища обобщённых вычислительных экспериментов на основе онтологического подхода // Онтология проектирования. 2022. Т.12, №1(43). С.41-56. DOI: 10.18287/2223-9537-2022-12-1-41-56.

Финансирование: исследование выполнено при поддержке Российского научного фонда (проект 18-11-00215).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение

Математическое моделирование реальных физических процессов в большинстве случаев предполагает проведение серий вычислительных экспериментов (ВЭ), в рамках которых изучается состояние модели и особенности её поведения в различных диапазонах изменения входных переменных и параметров модели. С развитием средств вычислительной техники и технологий параллельных вычислений появилась возможность построения обобщённого ВЭ (ОВЭ), основная идея которого состоит в многократном решении прямой или обратной задачи численного моделирования для различных наборов значений параметров моделирования [1]. Подобный подход позволяет получить решение не для одной, а для класса задач математического моделирования, заданного в многомерном пространстве параметров модели-

рования. В результате появляется возможность одновременно исследовать влияние нескольких параметров на интересующие характеристики модели, в том числе исследовать их совместное влияние в различных сочетаниях диапазонов изменения.

Целью проведения ОВЭ, как правило, является определение либо уточнение характеристик и свойств моделей: поиск функциональных зависимостей или закономерностей для прогнозирования исследуемых характеристик, определяемых результатами моделирования; сравнение точности моделей, а также установление границ областей, в которых определённая модель даёт более точные результаты по сравнению с другими. Известны примеры успешного применения ОВЭ при решении задач вычислительной гидродинамики [1, 2], газовой динамики [3, 4], автоматизации проектирования энергетических установок [5].

При переходе к ОВЭ значительно возрастают объём и размерность обрабатываемых массивов экспериментальных данных. Моделирование реальных физических процессов на основе проведения серии экспериментов со всеми допустимыми комбинациями моделей и параметров моделирования, как правило, не представляется возможным. Поэтому актуальной является задача адаптивного управления ОВЭ.

В работе [6] предложен подход к планированию ОВЭ с возможностью динамической корректировки плана на основе оценки результативности промежуточных состояний ОВЭ. Под планом ОВЭ понимается последовательность экспериментов, подлежащих проведению, для заданного многомерного массива значений параметров моделирования с выбранными методами анализа и интерпретации результатов эксперимента.

Преимуществом динамического планирования ОВЭ является возможность сокращения объёма недостаточно результативных экспериментов и, напротив, обеспечение более детального экспериментального исследования тех диапазонов значений входных данных и параметров моделирования, где требуется подтверждение и уточнение найденных закономерностей. Процедуру планирования целесообразно делать интерактивной, привлекая к ней исследователя и обеспечивая возможность совместного использования традиционных методов обработки экспериментальных данных и методов их интеллектуального анализа, например, визуально-когнитивной аналитики [2, 7, 8].

Для обеспечения автоматизации и программной поддержки методов адаптивного планирования и управления ОВЭ требуется организация хранения результатов ОВЭ и сценариев его проведения. Актуальной задачей является разработка методов и технологий построения хранилища данных ОВЭ. Его создание обеспечит возможность повторного использования экспериментальных данных и результатов их обработки, а также организацию библиотек и репозиториев моделей анализа данных по различным направлениям и классам проблем. Необходимо обеспечить простой и интуитивно понятный механизм взаимодействия пользователя с хранилищем, возможность формирования запросов в терминах предметной области (ПрО). Данная возможность может быть обеспечена за счёт использования онтологического подхода.

В статье рассматриваются особенности хранения данных, связанных с ОВЭ, предлагаются модели данных для построения хранилища ОВЭ, способ построения онтологии ОВЭ, и рассматриваются пути её применения для организации доступа к хранилищу и эффективного взаимодействия с ним пользователя.

1 Формализованное представление ОВЭ

Для формирования требований к структуре хранилища ОВЭ и выделения сущностей, которые должны быть положены в основу моделей представления данных, необходимо выпол-

нить формализацию понятия OBЭ, а также связанных с ним понятий состояния и плана OBЭ. С учётом результатов работы [6] предлагается следующая схема формализации.

1.1 Понятие и структура данных ОВЭ

Пусть математическая модель исследуемой системы задана набором:

$$\langle X, Y, A, F \rangle$$
, (1)

где $X = (x_1, x_2, ..., x_L)$ — вектор входных переменных, L — количество входных переменных, $Y = (y_1, y_2, ..., y_M)$ — вектор выходных переменных, M — количество выходных переменных, $A = \{a_1, a_2, ..., a_N\}$ — множество параметров моделирования, также называемых определяющими параметрами [1], N — количество определяющих параметров, F — реализуемое с помощью модели отображение пространства входных переменных в пространство выходных переменных. Вид данного отображения определяется структурой модели с учётом её определяющих параметров из множества A.

В качестве определяющих параметров моделирования могут выступать как числовые величины (например, некоторые физические характеристики исследуемого процесса), так и качественные показатели, такие как используемые численные методы, характеристики программно-аппаратного обеспечения и др. [2, 3].

Каждый параметр a_k характеризуется областью допустимых значений V_k (k=1,...,N). Для числового параметра это может быть диапазон изменения, а для нечислового — множество возможных значений. На значения параметров могут быть наложены совместные ограничения вида

$$g_t(a_1, a_2, ..., a_N), t = 1, ..., T,$$
 (2)

которые могут задаваться в форме уравнений, неравенств или логических условий. Совокупность областей значений параметров вместе с набором ограничений вида (2) задаёт N-мерное пространство определяющих параметров V, которое в общем случае является подмножеством декартова произведения областей значений каждого параметра:

$$\mathbf{V} \subseteq V_1 \times V_2 \times \dots V_N. \tag{3}$$

Поскольку ОВЭ предполагает решение задачи моделирования при вариации определяющих параметров на некотором дискретном наборе точек пространства этих параметров, то на области V_k значений каждого параметра a_k выбирается множество значений $W_k = \{a_{k1}, a_{k2}, ..., a_{k,pk}\}$, для каждого из которых выполняется моделирование. Здесь a_{kj} – конкретные выбранные значения параметра a_k , $j = 1, ..., p_k$, где p_k – число таких значений для параметра a_k . В простейшем случае для числового параметра W_k может представлять собой набор равноотстоящих точек внутри диапазона его значений.

Разбиение пространства определяющих параметров представляет собой множество всех возможных кортежей выбранных значений каждого параметра, принадлежащих пространству параметров V, т.е. удовлетворяющих ограничениям (2):

$$\mathbf{W} = (W_1 \times W_2 \times \dots W_N) \cap \mathbf{V}. \tag{4}$$

Любая точка, принадлежащая данному разбиению, задаёт определённый набор значений определяющих параметров, участвующий в проведении ОВЭ:

$$w_i = (a_{1,i1}, a_{2,i2}, ..., a_{N,iN}), a_{k,ik} \in W_k.$$
(5)

В качестве примера можно рассмотреть ОВЭ для трёхмерной задачи невязкого обтекания конуса [4]. Для данного ОВЭ определяющими параметрами являлись: число Маха (a_1) , угол полураствора конуса в градусах (a_2) , угол атаки в градусах (a_3) . В качестве W_k были выбраны следующие множества значений определяющих параметров: $W_1 = \{3, 5, 7\}$; $W_2 = \{10, 15, 20\}$;

 $W_3 = \{0, 5, 10\}$. Соответственно, w_i в данном случае представляют собой упорядоченные тройки элементов этих множеств: $w_1 = (3, 10, 0)$; $w_2 = (3, 10, 5)$; ...; $w_{24} = (3, 20, 10)$ (для комбинации углов полураствора и атаки, равных 10° , ВЭ не проводились).

С учётом введённых обозначений *одиночный* ВЭ характеризуется дискретным векторным полем $\mathbf{E}^{\mathbf{S}}$, которое задаёт распределение \mathbf{Y}^* значений вектора выходных переменных Y для заданного множества \mathbf{X} значений вектора входных переменных X при некотором фиксированном наборе w^* значений определяющих параметров, принадлежащем разбиению \mathbf{W} :

$$\mathbf{E}^{\mathbf{S}} = \langle \mathbf{X}, \mathbf{Y}^* \rangle, \tag{6}$$

где

$$Y^* = \{ Y = F(X, w^*), X \in X \}. \tag{7}$$

OBЭ (\mathbf{E}^{Gen}) характеризуется N-мерным массивом, элементы которого задают дискретные векторные поля, соответствующие одиночным ВЭ для всех элементов разбиения \mathbf{W} :

$$\mathbf{E}^{\mathsf{Gen}} = \langle \mathbf{X}, \mathbf{Y}, \mathbf{W} \rangle, \tag{8}$$

где

$$\mathbf{Y} = \{ Y = F(X, w), X \in \mathbf{X}, w \in \mathbf{W} \}. \tag{9}$$

Полученный массив многомерных данных содержит первичные результаты ОВЭ и не может рассматриваться как решение задачи моделирования. Поэтому далее выполняется его обработка с целью получения полезной для исследователя информации об объекте моделирования и выявления закономерностей и взаимосвязей, присущих определяющим его параметрам. Для этого могут использоваться традиционные методы обработки экспериментальных данных и методы интеллектуального анализа данных, а также совместное использование различных подходов и методов.

Таким путём, от первичного многомерного массива экспериментальных данных, задаваемого набором (8), осуществляется переход к обобщённым показателям, которые представляют собой результаты обработки первичных данных. Эти показатели используются далее для интерпретации, поиска закономерностей, формирования и проверки гипотез. Примерами обобщённых показателей могут служить главные компоненты в задачах понижения размерности, нормы векторов ошибок L1 и L2 в задачах оценивания точности различных численных методов при варьировании определяющих параметров моделирования [3, 4] и др.

1.2 План и состояние ОВЭ

План ОВЭ должен задавать:

- 1) серию одиночных ВЭ, которые требуется провести для заданного множества \mathbf{X} значений входных переменных и выбранного разбиения пространства определяющих параметров \mathbf{W} ;
- 2) множество обобщённых показателей для представления результатов обработки и анализа первичных экспериментальных данных, а также набор используемых для этого методов. План ОВЭ можно представить в виде:

$$\langle \mathbf{X}, \mathbf{W}, C, H \rangle$$
, (10)

где $C = \{C_1, C_2, ..., C_S\}$ — множество обобщённых показателей, H — отображение пространства первичных экспериментальных данных в пространство значений обобщённых показателей. Вид данного отображения определяется применяемыми методами обработки и анализа первичных данных.

Состояние ОВЭ определяется совокупностью ВЭ, проведённых в соответствии с выбранным планом, и задаётся многомерным массивом экспериментальных данных, получен-

ных при текущем разбиении пространства определяющих параметров, набором значений обобщённых показателей, полученных в результате обработки данного массива, а также множеством закономерностей, выявленных на основе анализа и интерпретации этих значений. Состояние ОВЭ можно представить в виде:

$$\langle X, Y, W, C(W), R \rangle$$
, (11)

где C(W) – пространство значений множества обобщённых показателей C, полученное при разбиении пространства определяющих параметров W; R – множество выявленных закономерностей.

2 Проектирование структуры хранилища ОВЭ

В контексте задачи построения хранилища, ОВЭ можно рассматривать как набор проведённых одиночных ВЭ и историю изменения планов их проведения. Для хранения набора одиночных ВЭ может быть использован традиционный подход к представлению многомерных экспериментальных данных, основанный на двумерных таблицах. В этом случае часть столбцов таблицы описывает значения определяющих параметров, а другая часть – значения выходных переменных для разных моделей. Это обусловливает первую особенность хранения ОВЭ – возможность использования традиционных подходов к хранению многомерных данных. Хранить значения входных переменных в подобной таблице нецелесообразно, поскольку эти значения могут быть недостаточно структурированными и создавать избыточность данных. Вместе с тем, эти значения должны содержаться в хранилище. Это обусловливает вторую особенность хранения ОВЭ - нецелесообразность табличного хранения значений входных переменных в единой таблице со значениями определяющих параметров и выходных переменных. Значения выходных переменных образуют многомерные массивы большого объёма, но при этом они не используются для анализа результатов напрямую вместо них используются обобщённые показатели. Данный аспект обусловливает третью особенность хранения ОВЭ – нецелесообразность хранения значений выходных переменных ВЭ в единой таблице со значениями определяющих параметров и обобщённых показателей. Целесообразно хранить их отдельно при условии сохранения связи с соответствующим одиночным ВЭ.

При хранении ОВЭ необходимо хранить и историю изменения планов проведения ВЭ. Это связано с тем, что после проведения части запланированных ВЭ возможна корректировка плана ОВЭ, т.е. создание новой последовательности проведения ВЭ, которая может иметь общие части с прошлыми планами и не иметь таковой. В общем случае имеется множество пересекающихся последовательностей одиночных ВЭ, и для их хранения целесообразно использовать соответствующие структуры данных, например, графовые. Это обусловливает четвёртую особенность хранения ОВЭ – целесообразность хранения истории изменений планов проведения ОВЭ в виде графовых структур.

После проведения очередной серии ВЭ оценивается состояние ОВЭ, анализ которого используется при формировании решения о дальнейшем плане проведения ОВЭ. В этой связи состояние ОВЭ можно рассматривать как характеристику проведённой серии последовательных одиночных ВЭ и сопоставлять её с последним экспериментом в данной серии. Это обусловливает *пятую особенность хранения ОВЭ* — необходимость хранения состояния ОВЭ после каждой проведённой серии экспериментов.

Для организации эффективного хранения данных, связанных с ОВЭ, важно учитывать специфику операций, которые будут осуществляться с ними. К таким операциям относятся добавление, редактирование, удаление и поиск (выборка). Операции поиска (выборки) в первую очередь используются для анализа состояния ОВЭ, что служит основой управления

ОВЭ. Хранение и обработка многомерных данных зачастую строится на базе реляционных систем управления базами данных (СУБД) и $OLAP^I$ -хранилищ.

Традиционные хранилища показывают высокую эффективность сбора и выборки данных, но обладают рядом недостатков при представлении данных [9], среди которых ключевыми для рассматриваемой задачи являются следующие.

- Хранение только детализированных данных. Для обеспечения большей нормализации и, как следствие, более быстрого доступа к данным, информация максимально детализируется, однако для последующего их анализа такой подход неудобен ввиду большого объёма дополнительной информации.
- Хранилище может содержать данные в разных форматах. Данное обстоятельство существенно затрудняет анализ данных без их предварительной обработки.
- Доступ к данным пользователей осуществляется на основе заранее составленных запросов. Такой подход не может обеспечить гибкость в выборе данных для анализа.

При хранении данных ОВЭ использование технологий *OLAP* приведёт к дополнительной избыточности, создаваемой структурой базы данных. При построении хранилища следует использовать современные реляционные СУБД, позволяющие проводить необходимую нормализацию отношений между сущностями с целью снижения избыточности и обеспечения целостности данных [10]. В модели данных для хранения ОВЭ целесообразно предусмотреть отдельные таблицы для *описания* входных переменных, определяющих параметров, выходных переменных и обобщённых показателей, а также отдельные таблицы *для хранения полученных в ходе проведения экспериментов значений* входных переменных, определяющих параметров, выходных переменных и обобщённых показателей. Для хранения значений выходных переменных нет необходимости проводить нормализацию, поэтому эти данные могут храниться в сериализованной форме в виде текста с использованием форматов *JSON* (или *XML*) или же в виде ссылки на объект (файл) во внешнем хранилище (например, в *S3*-хранилище [11]). Для этого может быть создана отдельная таблица, связанная с таблицей, описывающей одиночные ВЭ.

Для хранения больших объёмов экспериментальных данных можно использовать специальные иерархические форматы, например, HDF5 [12]. Особенность данного формата состоит в том, что каждый файл имеет свою внутреннюю структуру, и это может быть более эффективным (по сравнению с форматами JSON или XML) в случае, когда требуется обмениваться данными с другими системами (например, системами визуализации экспериментальных данных). Формат HDF5 позволяет получать данные в виде массива, при этом встроенные средства анализа отсутствуют. С использованием данного формата можно построить автономное хранилище, части которого можно легко отделить, что делает его удобным инструментом для транспортировки данных.

Учёт четвёртой особенности хранения ОВЭ приводит к наличию сущности или сущностей для хранения истории изменения планов проведения ОВЭ. Вводится понятие «родительского» плана ОВЭ, под которым понимается начальный план проведения серии ВЭ. Если в ходе его выполнения или после завершения возникает потребность в проведении другой серии ВЭ, то формируется новый план, который должен ссылаться на «родительский» план и на последний проведённый ВЭ «родительского» плана. Новый план может также послужить «родительским» для другого плана. Для того, чтобы хранить такие структуры в базе данных, следует организовать иерархию путём добавления в таблицу с описанием планов проведения ВЭ ссылки на другую запись этой же таблицы с ограничением в виде запрета ссылки на себя. Если в таблице отсутствует ссылка на «родительский» план, то данная запись связана с вер-

¹ OLAP (англ. online analytical processing) — технология обработки данных, заключающаяся в подготовке суммарной информации на основе больших массивов данных, структурированных по многомерному принципу.

шиной иерархии, т.е. с первичным планом проведения серии ВЭ. Описание такой таблицы планов должно включать номер начального ВЭ «родительского» плана. С помощью подобной иерархии планов и структуры таблиц представляется возможным восстановить полную историю проведения ОВЭ с учётом всех изменений.

Поскольку данные, описывающие состояния разных ОВЭ, могут иметь различные типы и размерность, то хранить состояния ОВЭ целесообразно также в сериализованном виде с использованием форматов *JSON* или *XML*. Для обращения к данным, хранящимся в подобных объектах, можно использовать встроенные в СУБД языковые механизмы запросов. Для хранения состояния ОВЭ можно добавить соответствующее поле в таблицу одиночных ВЭ, что позволит хранить состояние ОВЭ после каждого проведённого ВЭ.

Модель базы данных, которая может быть положена в основу организации хранилища ОВЭ, с учётом рассмотренных выше особенностей, представлена на рисунке 1.

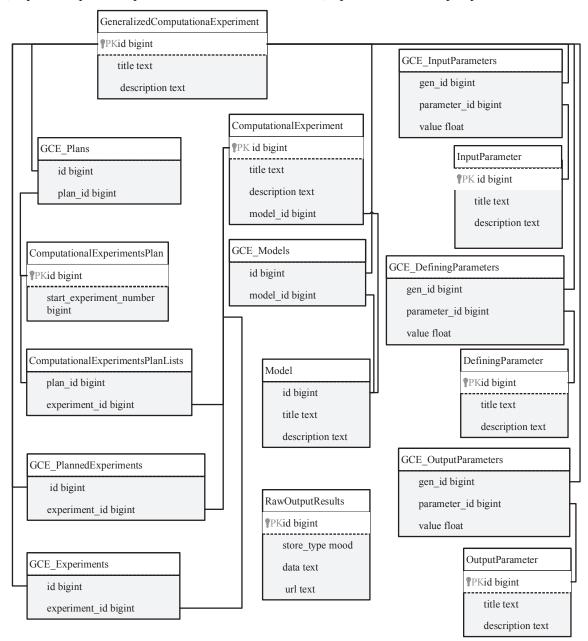


Рисунок 1 – Модель базы данных для организации хранилища ОВЭ

Механизм доступа пользователя к информации по ОВЭ с целью её анализа должен учитывать специфичный для ПрО механизм поиска, группировки и структурирования данных, а также принятую в соответствующей ПрО терминологию. В качестве основы для такого механизма предлагается использовать методологию онтологического инжиниринга. Онтология в рассматриваемой задаче позволяет:

- *добавлять в хранилище дополнительную мета-информацию об ОВЭ* (это обеспечивает возможность дополнять описание хранимых данных, что может быть учтено при построении более эффективных методов обработки соответствующих данных, например, указывать типы данных, используемых для конкретных параметров).
- формировать запросы и структуры данных при выборке (например, запросы пользователя, составленные через некоторый упрощённый человеко-машинный интерфейс, можно преобразовывать в физические запросы к хранилищу, либо уточнять уже существующие запросы [13], либо после получения данных из хранилища приводить к виду, удобному для понимания и последующего анализа).

Поиск и структуризацию данных исследователю удобнее выполнять в терминах той ПрО, к которой относится ОВЭ, а для этого в онтологии предусмотрен соответствующий тезаурус с синонимами, учитывающий в том числе и языковые особенности терминов [14].

3 Онтологическая модель ОВЭ

Онтологическая модель ОВЭ, инвариантная к ПрО, с поддержкой тезауруса ключевых понятий позволяет формировать мета-описание ОВЭ, включая сопоставление терминов, набор сущностей и их типов данных. Мета-описание может использоваться для формирования запросов к хранилищу в упрощённой форме, скрывая от исследователя детали физической реализации хранилища. Для разработки онтологической модели ОВЭ были выделены следующие ключевые сущности (понятия).

- 1) ОВЭ. Является базовым понятием создаваемой онтологии. Связано с остальными понятиями.
- 2) *Модель*. С учётом (1), под моделью поднимается отображение, с помощью которого на основе значений входных переменных и с учётом значений определяющих параметров формируются значения выходных переменных и/или обобщённых показателей.
- 3) Одиночный ВЭ. С учётом (6) и (7), это понятие объединяет данные, включающие в себя значения выходных переменных, полученные для заданного набора значений входных переменных и определяющих параметров с помощью выбранной модели.
- 4) План ОВЭ. В рамках онтологической модели под планом понимается серия одиночных ВЭ в определённом порядке их следования, а также набор правил обработки первичных экспериментальных данных и представления результатов этой обработки в виде значений обобщённых показателей.

Предполагается, что исследователь при проведении ОВЭ будет оперировать в первую очередь перечисленными ключевыми понятиями. Смысл одних и тех же понятий для исследователей из разных ПрО может отличаться, поэтому в онтологическую модель были включены понятия синоним, язык и ПрО. Словарь синонимов для используемых в онтологии понятий составляется для конкретной ПрО. При этом остаётся возможным использовать один и тот же синоним для различных ПрО (например, смежных). Разработанная обобщённая онтологическая схема ОВЭ представлена на рисунке 2.

Помимо неопределённости, связанной с трактовкой понятий ПрО, неопределённость может также возникать при описании самих компонентов онтологической модели. Описания на разных языках отличаются, например, русскоязычному термину «Обобщённый вычисли-

тельный эксперимент» соответствует англоязычный термин «Generalized computational experiment». Разные термины могут использоваться для описания одного и того же понятия внутри одного языка. Например, понятие «модель» в зависимости от ПрО может иметь целый ряд синонимов: функция, алгоритм, мутация, версия и т.д. Фрагмент разработанного тезауруса для некоторых ключевых понятий, входящих в онтологическую модель ОВЭ, представлен в таблицах 1 и 2. Использование исследователем привычных для него терминов повышает эффективность хранения, проведения и управления ОВЭ.

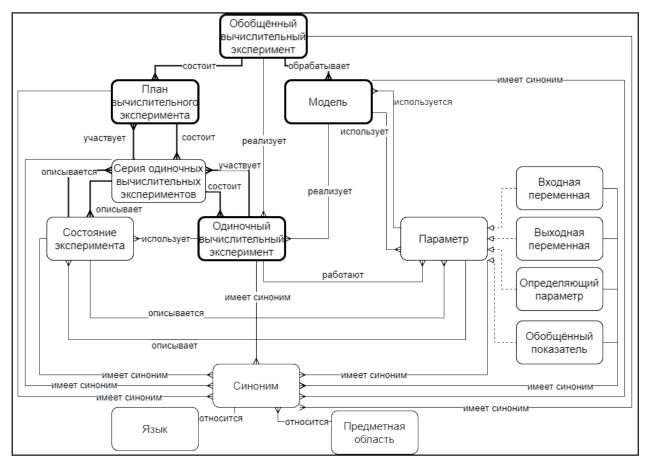


Рисунок 2 – Обобщённая онтологическая схема ОВЭ

Таблица 1 - Примеры синонимов для понятия "Обобщённый вычислительный эксперимент"

Синоним	Предметные области	Язык
Описание эксперимента	Общенаучная	Русский
Generalized computational experiment	Компьютерное моделирование	Английский
Серия экспериментов	Общенаучная	Русский
Экспериментальное исследование	Общенаучная	Русский

Предложенная обобщённая онтологическая модель ОВЭ позволяет создать дополнительный (верхний) уровень абстракции хранилища ОВЭ, с которым исследователь может взаимодействовать более эффективно по сравнению с доступом к хранилищу напрямую. Для обеспечения поддержки функционирования и взаимосвязи этого уровня с нижележащими уровнями хранилища ОВЭ разработана структура метаданных ОВЭ с учётом его онтологических особенностей (рисунок 3). Эта структура позволяет сформировать описание ПрО в привычных исследователю терминах, однозначно сопоставляемых с понятиями хранилища

OBЭ, а также указать спецификацию параметров, показателей и переменных, используемых в OBЭ. Дополнительно исследователь с помощью этого описания может задать метод оценки состояния OBЭ.

Таблица 2 - Примеры	синонимов для	понятия	"Модель"
---------------------	---------------	---------	----------

Синоним	Предметные области	Язык
Солвер	Компьютерное моделирование	Русский
Model	Общенаучная	Английский
Функция	Математика	Русский
Алгоритм	Математика	Русский
Метод	Общенаучная	Русский
Черный ящик	Информационные технологии	Русский
Программа	Общенаучная	Русский
Мутация	Генетика	Русский
Численный метод	Математика	Русский
Вариант	Общенаучная	Русский
Версия	Общенаучная	Русский
Модификация	Общенаучная	Русский



Рисунок 3 – Структура метаданных для работы с хранилищем ОВЭ

4 Апробация и обсуждение результатов

Для апробации предложенных подходов и методов использовались данные ОВЭ, проводившегося с целью верификации и оценки точности солверов (от англ. solver - решатель) платформы *OpenFOAM* для трёхмерной задачи сверхзвукового обтекания клина [15] (в терминологии *OpenFOAM* солверы представляет собой программные модули, в которых реализованы различные численные модели механики сплошных сред [16]).

Фрагмент мета-описания для данного ОВЭ представлен в следующем виде.

1) Словарь терминов < «понятие из онтологии», «используемый синоним»>:

««обобщённый вычислительный эксперимент», «обобщённый вычислительный эксперимент»>; «модель», «солвер»>; «план вычислительного эксперимента», «серия вычислительных экспериментов»>; «одиночный вычислительный эксперимент», «вычислительный эксперимент»>; «входная переменная», «входная переменная»>; «определяющий параметр»>; «выходная переменная», «выходная переменная»>; «выходная переменная»>; «кобобщённый показатель», «норма ошибки»>.

- 2) ПрО: верификация численных методов.
- 3) **Модели:** risoCentralFOAM, pisoCentralFOAM, sonicFOAM, QGDFoam.
- 4) **Входные переменные:** площадь и положение ячеек дискретизации. Тип данных вещественный.
- 5) Выходные переменные: давление в ячейках. Тип данных вещественный.
- 6) Определяющие параметры: угол набегающего потока, число Маха набегающего потока. Тип данных вещественный.
- 7) **Обобщённые показатели:** норма L1, норма L2. Тип данных вещественный. На основе словаря терминов получена онтологическая схема хранилища ОВЭ л

На основе словаря терминов получена онтологическая схема хранилища ОВЭ для ПрО математического моделирования и численных методов (см. рисунок 4).

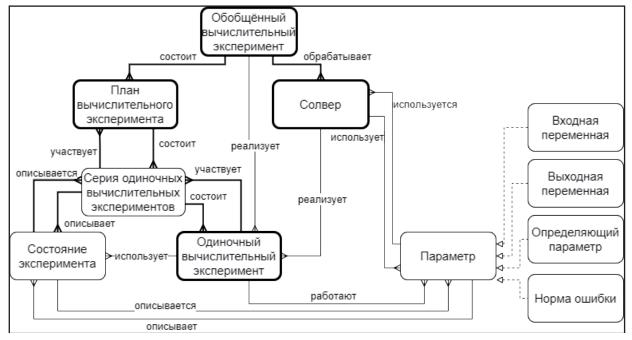


Рисунок 4 – Конкретизированная онтологическая схема ОВЭ для ПрО математического моделирования и численных методов

В хранилище были помещены сведения о 80 проведённых ВЭ. Состояние ОВЭ определялось с помощью метода, основанного на оценке точности двумерной аппроксимации при фиксированных комбинациях подмножества определяющих параметров [6]. Для этих целей с использованием имеющегося словаря терминов и мета-описания были сформированы следующие запросы к хранилищу на получение данных:

- 1) выбрать все возможные значения определяющего параметра «угол набегающего потока».
- 2) для всех комбинаций значений определяющего параметра «угол набегающего потока», модели и обобщённого показателя выбрать множество пар (значение «число Маха набегающего потока», значение обобщённого показателя).
- 3) выбрать все возможные значения определяющего параметра «число Маха набегающего потока».
- 4) для всех комбинаций значений определяющего параметра «число Маха набегающего потока», модели и обобщённого показателя выбрать множество пар (значение «угол набегающего потока», значение обобщённого показателя).

Полученные результаты выборок использовались для построения аппроксимирующих кривых, которые в дальнейшем подвергались визуальному анализу.

В ходе визуального анализа удалось обнаружить возможную неточность в данных для солвера *sonicFOAM* при значении числа Маха набегающего потока, равном 4, и значении угла набегающего потока, равному 20 градусам. Это привело к необходимости повторного проведения соответствующего одиночного ВЭ, в ходе которого данное предположение подтвердилось, а в хранилище ОВЭ были помещены уточнённые значения указанных определяющих параметров для данного одиночного ВЭ.

Отмеченный факт является подтверждением того, что систематическое и структурированное накопление и хранение сведений о проведённых ВЭ позволяет в дальнейшем проводить анализ всего ОВЭ, в ходе которого становится возможным как определение некорректно проведённых ВЭ, так и корректировка плана проведения ОВЭ. Осуществить подобную систематизацию без применения представленного онтологического подхода к формализации представления и обработки данных об ОВЭ было бы затруднительно. Применение тезауруса к онтологической схеме хранилища позволяет получить адаптированную версию онтологической схемы в терминах соответствующей ПрО и упростить исследователю взаимодействие с ней: представляется возможным формирование запросов на выборку данных в более привычных исследователю терминах.

Заключение

Предложен подход к построению хранилища ОВЭ, основанный на применении методов онтологического инжиниринга. Определены основные особенности хранения ОВЭ, разработана и обоснована схема построения хранилища, включающая дополнительный верхний уровень абстракции. На верхнем уровне исследователь взаимодействует с привычными ему терминами, а на нижних уровнях взаимодействие с данными осуществляется посредством традиционных подходов, реализуемых в современных СУБД.

Исследователю предоставляется возможность оперировать привычными ему понятиями, что положительно сказывается на качестве проводимых исследований с использованием разработанного хранилища ОВЭ.

Проведённые экспериментальные исследования на примере оценки точности решения задач механики сплошных сред с применением различных численных методов подтвердили работоспособность предложенного подхода, а также позволили выделить направления дальнейших исследований:

- 1) расширение числа ПрО и соответствующего тезауруса;
- 2) расширение онтологической модели за счёт выделения в ней типовых видов запросов данных, адаптированных к различным ПрО;
- 3) программная реализация предложенных подходов по формированию и преобразованию запросов от исследователя на понятном ему языке к их выполнению в соответствующей СУБД;
- 4) развитие методов и технологий визуализации данных, получаемых из разработанного хранилища ОВЭ.

Список источников

- [1] **Bondarev A.E.** On the Construction of the Generalized Numerical Experiment in Fluid Dynamics // Mathematica Montisnigri. 2018. Vol.XLII. P.52-64.
- [2] *Bondarev A.E., Galaktionov V.A.* Generalized Computational Experiment and Visual Analysis of Multidimensional Data // Scientific Visualization. 2019. Vol.11(4). P.102-114. DOI:10.26583/sv.11.4.09.
- [3] Alekseev A., Bondarev A., Galaktionov V., Kuvshinnikov A., Shapiro L. On Applying of Generalized Computational Experiment to Numerical Methods Verification // In: CEUR Workshop Proc. of 30th International Confer-

- ence on Computer Graphics and Machine Vision GraphiCon 2020. Vol.2744. http://ceur-ws.org/Vol-2744/paper19.pdf. DOI:10.51130/graphicon-2020-2-3-19.
- [4] *Bondarev A.E., Kuvshinnikov A.E.* Analysis of the Accuracy of OpenFOAM Solvers for the Problem of Supersonic Flow Around a Cone // In: Y. Shi et al. (ed.): Computational Science ICCS 2018. Lecture Notes in Computer Science, Vol.10862. Springer, Cham, 2018. P.221-230. DOI:10.1007/978-3-319-93713-7_18.
- [5] *Андреев С.В., Бондарев А.Е., Бондаренко А.В. и др.* Вычислительная технология построения оптимальной формы узла лопастей энергоустановки с учётом конструкционных ограничений // Программирование. 2017. №6. С.65-74.
- [6] Zakharova A., Korostelyov D., Podvesovskii A. Evaluating State Effectiveness in Control Model of a Generalized Computational Experiment // In: A.G. Kravets et. al. (ed.): Creativity in Intelligent Technologies and Data Science. CIT&DS 2021. Communications in Computer and Information Science, Vol.1448. Springer, Cham, 2021. P.207-222. DOI:10.1007/978-3-030-87034-8 16.
- [7] Zakharova A.A., Korostelyov D.A., Podvesovskii A.G., Galaktionov V.A. Methods of Constructing a Visual Map of Generalized Computational Experiment // Scientific Visualization. 2021. Vol.13(4). P.76-92. DOI:10.26583/sv.13.4.07.
- [8] **Zakharova A.A., Korostelyov D.A.** Application of Visual Analytics Methods to Reduce the Dimensionality of Decision-making Problems // Scientific Visualization. 2020. Vol.12(4). P.23-32. DOI:10.26583/sv.12.4.03.
- [9] Liu L., Özsu M.T. OLAP. Springer. 2009. DOI:10.1007/978-0-387-39940-9 3191.
- [10] *Wang M., Chan R.K.* Database Development Process // Encyclopedia of Information Systems. 2003. Vol.1. P.389-402. DOI:10.1016/B0-12-227240-4/00026-5.
- [11] *Li Y., Guo L., Guo Y.* An Efficient and Performance-Aware Big Data Storage System // Cloud Computing and Services Science. 2013. Vol.367. P.102–116. DOI:10.1007/978-3-319-04519-1_7.
- [12] *De Carlo F., Gürsoy D., Marone F., et.al.* Scientific data exchange: a schema for HDF5-based storage of raw and analyzed data // Journal of Synchrotron Radiation. 2014. Vol.21(6). P.1224–1230. DOI:10.1107/s160057751401604x.
- [13] *Leshcheva I., Begler A.* A method of semi-automated ontology population from multiple semi-structured data sources // Journal of Information Science. 2020. DOI:10.1177/0165551520950243.
- [14] *Микони С.В.* О качестве онтологических моделей // Онтология проектирования. 2017. Т.7, №3(25). С.347-360. DOI:10.18287/2223-9537-2017-7-3-347-360.
- [15] *Bondarev A.E., Kuvshinnikov A.E.* Analysis and Visualization of the Computational Experiments Results on the Comparative Assessment of OpenFOAM Solvers Accuracy for a Rarefaction Wave Problem // Scientific Visualization. 2021. Vol.13(3). P.34-46. DOI:10.26583/sv.13.3.04.
- [16] OpenFOAM | Free CFD Software | The OpenFOAM Foundation https://openfoam.org.

Сведения об авторах



Подвесовский Александр Георгиевич, 1972 г. рождения. Окончил Национальный технический университет Украины «Киевский политехнический институт» в 1996 г., к.т.н. (2001). Доцент кафедры «Информатика и программное обеспечение» Брянского государственного технического университета, старший научный сотрудник временной научно-исследовательской лаборатории «Автоматизированные системы обработки и визуализации информации» Института прикладной математики им. М.В. Келдыша РАН. Председатель Брянского регионального отделения Российской ассоциации искусственного интеллекта. В списке научных трудов более 120 работ в области методов и систем поддержки принятия

решений, когнитивного моделирования, анализа данных. Author ID (RSCI): 195986; Author ID (Scopus): 56367146100; Researcher ID (WoS): L-5247-2015. apodv@tu-bryansk.ru. ⊠.

Коростелёв Дмитрий Александрович, 1985 г. рождения. Окончил Брянский государственный технический университет в 2008 г., к.т.н. (2011). Доцент кафедры «Информатика и программное обеспечение» Брянского государственного технического университета, научный сотрудник временной научно-исследовательской лаборатории «Автоматизированные системы обработки и визуализации информации» Института прикладной математики им. М.В. Келдыша РАН. Член Российской ассоциации искусственного интеллекта. В списке научных трудов более 80 работ в области методов и систем поддержки принятия решений, компьютерного и математического моделирования, разработки программного обеспечения, когнитивного моделирования, искусственного интеллекта. Author ID (RSCI): 556496; Author ID (Scopus): 56368046200; Researcher ID (WoS): G-2682-2016. nigm85@mail.ru





Лупачёв Евгений Андреевич, 1995 г. рождения. Окончил Брянский государственный технический университет в 2019 г. Аспирант Брянского государственного технического университета, программист временной научно-исследовательской лаборатории «Автоматизирован-

ные системы обработки и визуализации информации» Института прикладной математики им. М.В. Келдыша РАН. Author ID (RSCI): 1139682. lupacheveugene@gmail.com

Беляков Никита Владимирович, 1999 г. рождения. Магистрант кафедры «Информатика и программное обеспечение» Брянского государственного

технического университета, программист временной научно-исследовательской лаборатории «Автоматизированные системы обработки и визуализации информации» Института прикладной математики им. М.В. Келдыша РАН. Author ID (RSCI): 1142163. nikregul@gmail.com



Поступила в редакцию 22.03.2021, после рецензирования 28.03.2022. Принята к публикации 31.03.2022.

Building a repository of generalized computational experiments based on the ontological approach

© 2022, A.G. Podvesovskii , D.A. Korostelyov, E.A.. Lupachev, N.V. Belyakov

Bryansk State Technical University, Bryansk, Russia Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Moscow, Russia

Abstract

The paper deals with the problem of organizing storage and access to information about generalized computational experiments related to the mathematical modeling of real physical processes. A generalized computational experiment implies a multiple solution of the numerical simulation problem for different sets of values of defining model parameters, and such an approach produces a solution at once for a certain class of problems of mathematical modeling given in a multidimensional space of defining parameters. An approach to building a data repository, describing generalized computational experiments, including the history of their planning and scenario corrections, is proposed. The features of storage of generalized computational experiments, which are the basis for the construction of logical and physical models of the repository, are discussed. To organize effective user interaction with the repository, an ontological model of a generalized computational experiment and the associated metadata structure are proposed, on the basis of which the mechanism for generating queries to the repository in terms of the subject area of modeling is implemented. The application of the developed approach for storing and analyzing data on a generalized computational experiment related to the estimation of the accuracy of numerical models of the OpenFOAM software platform for the three-dimensional problem of supersonic flow around a wedge is discussed.

Key words: generalized computational experiment, multidimensional data, data representation, data storage, ontology, thesaurus, verification of numerical methods.

Citation: Podvesovskii AG, Korostelyov DA, Lupachev EA, Belyakov NV. Building a repository of generalized computational experiments based on the ontological approach [In Russian]. Ontology of designing. 2022; 12(1): 41-56. DOI: 10.18287/2223-9537-2022-12-1-41-56.

Financial Support: The study was supported by the Russian Science Foundation (project 18-11-00215).

Conflict of interest: The author declares no conflict of interest.

List of figures and tables

Figure 1 – A physical database model for organizing a repository of generalized computational experiments

Figure 2 – A generalized ontological scheme of generalized computational experiments

- Figure 3 Metadata structure for working with the repository of generalized computational experiments
- Figure 4 Concretized ontology scheme for the subject area of mathematical modeling and numerical methods
- Table 1 Examples of synonyms for the "Generalized computational experiment" term
- Table 2 Examples of synonyms for the "Model" term

References

- [1] **Bondarev AE.** On the Construction of the Generalized Numerical Experiment in Fluid Dynamics. *Mathematica Montisnigri*. 2018; XLII: 52-64.
- [2] **Bondarev AE, Galaktionov VA.** Generalized Computational Experiment and Visual Analysis of Multidimensional Data. *Scientific Visualization*. 2019; 11(4): 102-114. DOI:10.26583/sv.11.4.09.
- [3] *Alekseev A, Bondarev A, Galaktionov V, Kuvshinnikov A, Shapiro L.* On Applying of Generalized Computational Experiment to Numerical Methods Verification. In: CEUR Workshop Proc. of 30th International Conference on Computer Graphics and Machine Vision GraphiCon 2020; 2744. http://ceur-ws.org/Vol-2744/paper19.pdf. DOI:10.51130/graphicon-2020-2-3-19.
- [4] *Bondarev AE, Kuvshinnikov AE.* Analysis of the Accuracy of OpenFOAM Solvers for the Problem of Supersonic Flow Around a Cone. In: Y. Shi et al. (ed.): Computational Science ICCS 2018. Lecture Notes in Computer Science, Vol.10862. Springer, Cham, 2018. P.221-230. DOI:10.1007/978-3-319-93713-7_18.
- [5] Andreev SV, Bondarev AE, Bondarenko AV. et. al. A computational technology for constructing the optimal shape of a power plant blade assembly taking into account structural constraints [In Russian]. Programming and Computer Software. 2017; 6: 65-74. DOI:10.1134/S0361768817060020.
- [6] Zakharova A, Korostelyov D, Podvesovskii A. Evaluating State Effectiveness in Control Model of a Generalized Computational Experiment. In: A.G. Kravets et. al. (ed.): Creativity in Intelligent Technologies and Data Science. CIT&DS 2021. Communications in Computer and Information Science, Vol.1448. Springer, Cham, 2021. P.207-222. DOI:10.1007/978-3-030-87034-8 16.
- [7] Zakharova AA, Korostelyov DA, Podvesovskii AG, Galaktionov VA. Methods of Constructing a Visual Map of Generalized Computational Experiment. Scientific Visualization. 2021; 13(4): 76-92. DOI:10.26583/sv.13.4.07.
- [8] Zakharova AA, Korostelyov DA. Application of Visual Analytics Methods to Reduce the Dimensionality of Decision-making Problems. Scientific Visualization. 2020; 12(4): 23-32. DOI: 10.26583/sv.12.4.03.
- [9] Liu L, Özsu MT. OLAP. Springer. 2009. DOI:10.1007/978-0-387-39940-9_3191.
- [10] *Wang M, Chan RK.* Database Development Process. *Encyclopedia of Information Systems*. 2003; 1: 389-402. DOI: 10.1016/B0-12-227240-4/00026-5.
- [11] Li Y, Guo L, Guo Y. An Efficient and Performance-Aware Big Data Storage System. Cloud Computing and Services Science. 2013; 367: 102–116. DOI:10.1007/978-3-319-04519-1 7.
- [12] *De Carlo F, Gürsoy D, Marone F. et.al.* Scientific data exchange: a schema for HDF5-based storage of raw and analyzed data. *Journal of Synchrotron Radiation*. 2014; 21(6): 1224–1230. DOI:10.1107/s160057751401604x.
- [13] *Leshcheva I, Begler A.* A method of semi-automated ontology population from multiple semi-structured data sources. *Journal of Information Science*. 2020. DOI:10.1177/0165551520950243.
- [14] *Mikoni SV.* On the quality of ontological models [In Russian]. *Ontology of Designing*. 2017; 7(3): 347-360. DOI:10.18287/2223-9537-2017-7-3-347-360.
- [15] *Bondarev AE, Kuvshinnikov AE.* Analysis and Visualization of the Computational Experiments Results on the Comparative Assessment of OpenFOAM Solvers Accuracy for a Rarefaction Wave Problem. *Scientific Visualization*. 2021; 13(3): 34-46. DOI:10.26583/sv.13.3.04.
- [16] OpenFOAM | Free CFD Software | The OpenFOAM Foundation. https://openfoam.org.

About the authors

Aleksandr Georgievich Podvesovskii (b. 1972) graduated from National Technical University of Ukraine ("Kyiv Polytechnical Institute") in 1996, PhD (2001). He is an Associate Professor at Bryansk State Technical University (Department of Informatics and Software Engineering) and a Senior Researcher at Keldysh Institute of Applied Mathematics of Russian Academy of Sciences (Temporary Research Laboratory of Automated Systems for Information Processing and Visualization). He is the Chair of Bryansk regional branch of Russian Association of Artificial Intelligence. He is a coauthor of more than 120 scientific articles and abstracts in the field of decision support methods and systems, cognitive modeling, and data analysis. Author ID (RSCI): 195986; Author ID (Scopus): 56367146100; Researcher ID (WoS): L-5247-2015. apodv@tu-bryansk.ru.

Dmitriy Aleksandrovich Korostelyov (b. 1985) graduated from Bryansk State Technical University in 2008, PhD (2011). He is an Associate Professor at Bryansk State Technical University (Department of Informatics and Software Engineering) and a Researcher at Keldysh Institute of Applied Mathematics of Russian Academy of Sciences (Temporary Research Laboratory of Automated Systems for Information Processing and Visualization). He is a member of Russian Association of Artificial Intelligence. He is a co-author of more than 80 scientific articles and abstracts in the field of decision support methods and systems, computer and mathematical simulation, software development, cognitive modeling, and artificial intelligence. Author ID (RSCI): 556496; Author ID (Scopus): 56368046200; Researcher ID (WoS): G-2682-2016. nigm85@mail.ru.

Evgenii Andreevich Lupachev (b. 1995) graduated from Bryansk Technical State University in 2019. He is a post-graduate student at Bryansk State Technical University and a Programmer at Keldysh Institute of Applied Mathematics of Russian Academy of Sciences (Temporary Research Laboratory of Automated Systems for Information Processing and Visualization). Author ID (RSCI): 1139682. *lupacheveugene@gmail.com*.

Nikita Vladimirovich Belyakov (b. 1999) is a graduate student at Bryansk State Technical University and a Programmer at Keldysh Institute of Applied Mathematics of Russian Academy of Sciences (Temporary Research Laboratory of Automated Systems for Information Processing and Visualization). Author ID (RSCI): 1142163. *nikregul@gmail.com*.

Received March 22, .2022. Revised March 28, 2022. Accepted March 31, .2022.