

Построение динамической модели предметной области для решения сложно структурированных задач

А. В. КАРКАНИЦА

Введение

В условиях динамично меняющейся внешней среды и в результате глобализации общества появляется множество задач с новыми свойствами. В частности, наблюдается рост количества сложно структурированных задач с недостаточно формализованной предметной областью и большим количеством удаленных источников информации. Эти «новые задачи» можно охарактеризовать как задачи, для решения которых требуется «внешняя» информация, т.е. информация, которая имеется в распределенных источниках глобальной среды. Типичным примером данной группы задач является задача поддержки выполнения заказных проектов (в том числе, IT-проектов), которые инициируются в одной стране, а выполняются распределенными коллективами исполнителей в других странах [1]. Предметная область такого рода задач обладает структурой, процесс моделирования которой не может быть закончен в ограниченный период времени, так как объекты предметной области могут менять свое состояние и структуру на протяжении всего жизненного цикла (от появления до исчезновения).

На практике таких задач значительно больше и, несмотря на различия в постановках, суть их одинакова: имеется задача, предметная область которой формируется из множества источников информации, распределенных в глобальной среде. Решение задачи должно носить оперативный характер в силу того, что в условиях ускорения динамики бизнес-процессов знания быстро устаревают. Основные требования к решению – актуальность, точность и полнота предметной области.

Таким образом, можно сделать вывод о том, что в настоящее время актуальна проблема разработки технологии, которая должна поддерживать выполнение проектов территориально распределенными группами исполнителей, обеспечивать независимость времени реализации проекта от количества исполнителей, автоматизировать процесс отображения материалов проекта в компьютерную систему [2]. В свою очередь, модели предметных областей должны отражать не только состояние и структуру на момент проектирования и реализации программной системы, но и иметь внутренние механизмы динамической адаптации (в том числе и структурной) модели предметной области на протяжении всего ее жизненного цикла.

Анализ нового класса задач

Типичная схема распределенного решения сложных задач несколькими исполнителями включает следующие этапы [3]:

- 1 центральный орган (руководитель) проводит декомпозицию исходной проблемы на отдельные задачи;
- 2 задачи распределяются между исполнителями;
- 3 каждый исполнитель решает свою задачу или снова разделяет ее на подзадачи, обращаясь за помощью к другим исполнителям;
- 4 для получения общего результата производится композиция частных результатов, соответствующих выделенным задачам.

Двумя важнейшими аспектами в этой схеме являются распределение задач между исполнителями и объединение результатов. На этапе декомпозиции центр может провести разбиение задачи на подзадачи, но не способен найти их решение ввиду ограничений (компе-

тенция, ресурсы, время). Возникает необходимость распределения задач. После получения частных результатов возникает проблема их согласования и интеграции.

Таким образом, онтология предметной области строится в процессе совместной деятельности групп экспертов. Автоматизация этого процесса требует создания программной системы, которая будет поддерживать совместную согласованную работу по формированию онтологии предметной области географически распределенной экспертной группой.

Уточним основополагающие понятия, связанные с описанным классом задач и построим онтологическую модель задачи в ее первом приближении. Для этого используем онтологический подход, стандартизованный в IDEF5 и допускающий уточнение каждого компонента первичной модели [4].

Задачу (**Task**) определим как совокупность следующих составляющих: **S** – постановка задачи; **Req** – требования к решению; **Solution** – результат решения задачи, представленный в форме компьютерной системы:

$$\mathbf{Task} = (\mathbf{S}, \mathbf{Req}, \mathbf{Solution}) \quad (1)$$

Рассматриваемый класс задач решается коллективом распределенных исполнителей, который определим как группу (**Group**), включающую центр (**Center**), инициирующий задачу, и распределенных исполнителей (**Expert**), реализующих решение подзадач:

$$\mathbf{Group} = (\mathbf{Center}, \mathbf{Expert}^1, \mathbf{Expert}^2, \dots, \mathbf{Expert}^n) \quad (2)$$

Объединив кортежи (1), (2), уточним модель задачи:

$$\mathbf{Task} = (\mathbf{S}, \mathbf{Req}, \mathbf{Group}, \mathbf{Solution}) \quad (3)$$

Свойство структурированности **Task** подразумевает декомпозицию исходной задачи **S** на подзадачи $\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^n$, а необходимость оперативного получения решения накладывает ограничения по времени (**T**). Формализованная информация, полученная в процессе решения подзадач, в сумме составляет решение **Task**:

$$\mathbf{Solution} = \mathbf{Solution}^1 + \mathbf{Solution}^2 + \dots \mathbf{Solution}^n \quad (4)$$

Соответственно уточним модель задачи:

$$\mathbf{Task} = (\mathbf{S}, \mathbf{Group}, \mathbf{S}^1, \dots, \mathbf{S}^n, \mathbf{Req}^1, \dots, \mathbf{Req}^k, \mathbf{Solution}^1, \dots, \mathbf{Solution}^n, \mathbf{T}^1, \dots, \mathbf{T}^n), \quad (5)$$

где $n \rightarrow \infty$, k – переменная величина.

Описанный класс задач определим как сложно структурированные задачи (ССЗ). Анализируя компоненты модели (5), выделим общие свойства нового класса задач:

C1. Сложная структура задачи определяется наличием в модели элементов $\mathbf{S}^1, \dots, \mathbf{S}^n$, определяющих иерархию подзадач исходной задачи;

C2. Распределенность – онтология предметной области формируется группой экспертов **Group**;

C3. Фрагментарность – решение целевой задачи представлено совокупностью решений подзадач (фрагментов) $\mathbf{Solution}^1, \dots, \mathbf{Solution}^n$;

C4. Динамичность – топология подзадач меняется по мере подключения новых исполнителей или прекращения работы старых, т.е. количество источников формализованной информации велико и заранее неизвестно ($n \rightarrow \infty$);

C5. Масштабность – группа экспертов (**Center**, $\mathbf{Expert}^1, \mathbf{Expert}^2, \dots, \mathbf{Expert}^n$) формируется из многочисленных географически распределенных коллективов, находящихся в глобальной среде и использующих глобальные коммуникации.

Предметную область **Y** сложноструктурированной задачи определим как формализованную информацию, необходимую и достаточную для решения исходной задачи и будем считать ее целевой, если она содержит паттерны знаний, обеспечивающие решение всех подзадач.

$$\mathbf{Y} = \mathbf{Pat}^1 + \mathbf{Pat}^2 + \dots \mathbf{Pat}^n, \\ \mathbf{Pat}^i = \mathbf{Solution}^i = (\mathbf{S}^i, \mathbf{Sol}^i), \quad (6)$$

где \mathbf{S}^i – постановка подзадачи, \mathbf{Sol}^i – ее практическое решение

Из свойств модели задачи логично вытекают свойства предметной области: распределенность, фрагментарность, динамичность.

Постановка задачи исследования

Очевидно, что для оперативного решения ССЗ требуется создание компьютерной системы поддержки деятельности распределенных экспертных групп по разработке онологий предметных областей. В рамках этой задачи требуется автоматизировать процесс формирования предметной области. Соответственно выделяется задача на разработку средства автоматизации формирования предметных областей, обеспечивающего согласованную совместную работу распределенных экспертных коллективов.

Постановку задачи сформулируем следующим образом:

Пусть имеется некоторая сложно структурированная задача **Task**, включающая постановку общей задачи **S** и **n** ее атомарных подзадач:

$$\mathbf{Task} = (\mathbf{S}, \mathbf{S}^1, \dots, \mathbf{S}^n), n \rightarrow \infty$$

В рамках каждой подзадачи известен текст постановки задачи **Text**, спецификация требований **Req** и ограничения на время решения **Time**, описанные кортежем:

$$\mathbf{S}^i = (\mathbf{Text}^i, \mathbf{Req}^i, \mathbf{Time}^i), \mathbf{Time}^i \rightarrow 0$$

В решении задачи участвуют центр **Center** и **k** территориально распределенных исполнителей **Expert¹, Expert², ..., Expert^k**.

Для решения задачи **S** необходимы информационные ресурсы **Zⁱ** (знания), в сумме составляющие предметную область **Y**.

Требуется разработать средство автоматизации формирования ПрО **Y** на основе знаний **Zⁱ**, полученных из **k** удаленных экспертных источников **Expert¹, Expert², ..., Expert^k**, при условии, что **Timeⁱ → 0, n → ∞**. Результатом решения должна быть формализованная предметная область.

В основе решения поставленной задачи лежит предметная область, следовательно, прежде всего, необходимо разработать ее модель, удовлетворяющую описанным выше свойствам. Только после этого можно перейти к этапу разработки средства ее использования распределенными группами экспертов.

Моделирование предметной области ССЗ

Представим процесс формирования предметной области совокупностью следующих типовых подпроцессов:

1 процесс моделирования **P1**: на вход процесса поступает исходная задача **S**, выполняется ее декомпозиция на подзадачи, построение концептуальной и физической модели, на выходе процесса модель ПрО – **modY**;

2 процесс добычи знаний **P2**, релевантных подзадачам: на входе процесса модель предметной области **modY**, на выходе – совокупность знаний **Solution**, представляющих решение подзадач;

3 процесс формирования ПрО **P3**: на входе процесса – **Solution**, на выходе – формализованная ПрО.

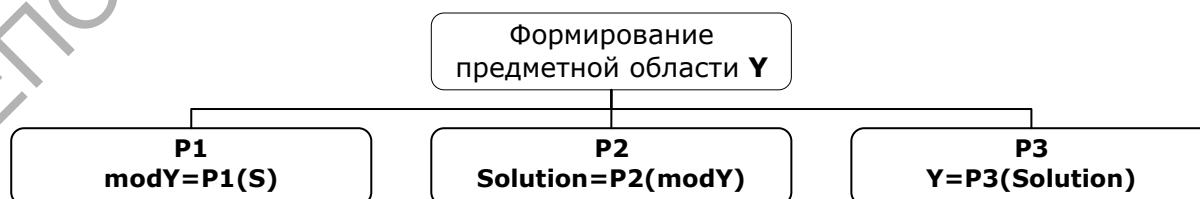


Рисунок 1 – Основные процессы построения ПрО

В настоящее время для решения задачи моделирования ПрО (процесс **P1**) наибольшее распространение получил подход Питера Пин-Чена, известный как структурная методология. Моделирование предметной области базируется на использовании графических диаграмм, включающих небольшое число разнородных компонентов. Смысл подхода заключа-

ется в структурировании важной семантической информации о предметной области путем выделения значимых сущностей (entity) и связей (relationship). Связи между сущностями представляются в форме иерархии, известной как ERD-диаграммы [5]. Онтологическая модель предметной области, согласно подходу Чена, определяется следующим образом:

$$\text{mod}Y=(\text{Entity, Relation, ERD}), \quad (7)$$

где Entity – конечное множество значимых сущностей предметной области, Relation – множество типов связей между сущностями, ERD – диаграммы, представляющие иерархию связей между сущностями. Алгоритм построения модели ПрО следующий:

- 1 проанализировать задачу S;
- 2 выделить сущности $\text{entity}^1, \dots, \text{entity}^n$;
- 3 идентифицировать сущности с помощью ключевых атрибутов;
- 4 определить связи $\text{Relationship}^m=f(\text{entity}^i, \text{entity}^j)$;
- 5 построить иерархию связей;
- 6 построить ERD-модель.

Можно выделить очевидные достоинства методологии «сущность – связь»:

- принцип декомпозиции, заложенный в основу методологии, позволяет получать решение сложных проблем путем их разбиения на простые независимые подзадачи;
- очевидная простота реализации алгоритма построения модели позволяет обеспечить быстроту формирования ПрО;
- жесткость, статичность, неделимость модели, ее ориентация на локальное применение позволяют автоматизировать процесс формирования ПрО.

Рассмотрим вопрос применимости методологии П. Чена для построения модели ПрО сложно структурированных задач. Значимыми сущностями в контексте рассматриваемого класса ССЗ являются подзадачи S^1, \dots, S^n и исполнители $\text{Expert}^1, \text{Expert}^2, \dots, \text{Expert}^n$. Применение алгоритма Чена позволит получить ERD-модель в виде ациклического графа, узлы которого определяют иерархию подзадач и назначение исполнителей, а дуги – уровень их вложенности. Однако это возможно в случае, когда изначально центром строго определена декомпозиция исходной задачи на подзадачи, т.е. известно их количество n и определена группа экспертов. Анализ компонентов модели (7) и свойств задачи (5) позволяет выявить существенные несоответствия свойств С2, С3, С4 нового класса задач со свойствами онтологической модели ПрО, построенной согласно подходу П.Чена:

- статичность модели не позволяет модифицировать ее за счет компетенции удаленных экспертов – не удовлетворяется свойство динамичности;
- неделимость модели противоречит свойству фрагментарности ССЗ;
- ориентация средств построения модели на локальные группы исполнителей осложняет их использование территориально распределенными экспертами – не удовлетворяется свойство распределенности.

В силу выявленных противоречий делаем вывод о том, что прямое использование модели и алгоритма П.Чена для построения модели ПрО задачи (5) невозможно. В случае ССЗ основные процессы формирования ПрО реализуются множеством удаленных экспертов, что требует соответствующей коррекции модели ПрО, алгоритмов реализации процессов P1, P2, P3 и организации специальных средств коммуникации.

Следует все же отметить, что модель П. Чена успешно используется на начальных этапах проектирования программных систем и в случае, если:

- коллектив исполнителей обладает достаточным объемом знаний для компетентной декомпозиции задачи и формирования ПрО;
- исходная задача **Task** хорошо структурированная, количество атомарных подзадач n заранее известно и конечно;
- предметная область **Y** статичная, достаточно формализованная;
- уровень критичности параметра **Time** низкий.

Известные из литературы методологии формирования онтологии предметной области и методы разработки на их основе программных систем не охватывают случай сложно структурированных предметных областей [6]. Однако модель «сущность-связь» считается наиболее общей, из нее могут быть порождены все существующие модели данных (иерархическая, сетевая, реляционная, объектная), и поэтому она может быть взята за основу моделирования ПрО ССЗ. Предлагается модификация онтологической модели и алгоритма П. Чена с целью устранения свойства статичности и наделения модели свойством распределенности и фрагментарности.

Модель предметной области ССЗ должна удовлетворять свойствам С2, С3, С4 задачи (5). Проведем модификацию онтологической модели П. Чена за счет добавления динамического компонента, представленного фреймами **P** и **Z**:

$$\text{modY}=(\text{Entity, Relation, ERD} \mid \mathbf{P, Z}). \tag{8}$$

Структура фрейма **P** определяет постановку задачи эксперту: **Id** – идентификатор задачи; **Text** – текст постановки задачи; **Spec** – требования к решению. Слоты фрейма заполняются центром. Структура фрейма **Z** определяет шаблон решения задачи: **Y** – решение, **Tool** – средство реализации **Y**. Слоты фрейма **Z** пусты. В дальнейшем они уточняются за счет знаний экспертов.

Первые три компонента модели (8) представлены ERD-графом, вершиной которого является задача **S**, узлы (**Entity**) определяют иерархию подзадач, дуги (**Relation**) – уровень их вложенности. Фреймы **P** и **Z** включаются к терминальным узлам дерева. В результате модель представлена ациклическим графом с терминальными узлами, представленными фреймами и, в отличие от модели (7), состоит из двух частей: статической и динамической (рис. 2, 3, 4).

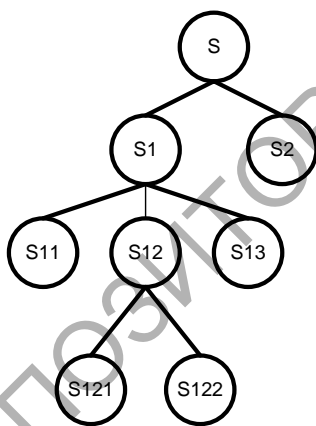


Рисунок 2 – Статическая часть модели ПрО

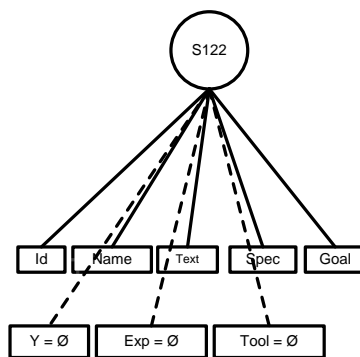


Рисунок 3 – Фрагмент модели ПрО

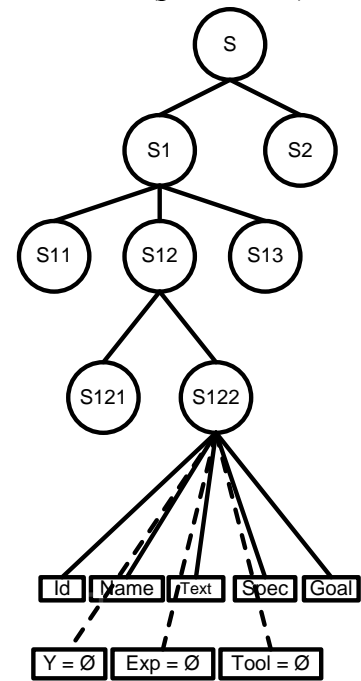


Рисунок 4 – Динамическая модель предметной области

Для формирования предметной области алгоритм П. Чена также необходимо модифицировать, разбив его на два этапа:

- на первом этапе построить статическую часть модели, представляющую собой иерархию подзадач;
- на втором этапе построить динамическую часть модели, включив к терминальным узлам дерева фреймы **P** и **Z**.

В результате формируется динамическая модель предметной области. Изолированность частей модели дает возможность построения второй части удаленными экспертами, что удовлетворяет свойству распределенности. Фрагментарность модели обеспечена тем, что каждый терминальный узел, представленный фреймами P и Z , определяет вполне определенный фрагмент модели предметной области, который в дальнейшем может быть доставлен эксперту. Таким образом, предложенная модификация модели Π . Чена обладает свойствами, выделенными как свойства ССЗ.

Заключение

В работе рассмотрен класс задач, предметная область которых формируется в результате совместной работы распределенной группы экспертов. В первом приближении представлена онтологическая модель задачи, выделены свойства нового класса задач и соответствующие свойства предметной области: динамичность, распределенность, фрагментарность. Для построения модели ПрО предложена модификация известной модели и соответствующего алгоритма за счет добавления динамического компонента.

Резюме. В представленной статье рассматривается новый класс задач, предметная область которых формируется из множества источников информации, распределенных в глобальной среде. Решение задачи – это результат совместной работы распределенной группы экспертов. Вводится понятие сложно структурированной задачи (ССЗ), с использованием онтологического подхода строится ее модель, выделяются свойства нового класса задач. Предлагается динамическая модель предметной области ССЗ, полученная путем модификации известной модели «сущность-связь» за счет добавления динамического компонента.

Abstract. The new class of the problems which subject domain is formed of a set of information sources distributed in the global environment is considered in the article. The problem solving is a result of the expert distributed commission teamwork. The concept of a difficult structured problem (DSP) is entered, its model is constructed with the use of the ontological approach, and new class problem properties are allocated. The dynamic model of DSP subject domain received by updating of the known model “essence-communication” at the expense of a dynamic component addition is offered.

Литература

1. Вальвачев, А.Н. Технология выполнения IT-проектов коллективами распределенных исполнителей / А.Н. Вальвачев, В.В. Краснопрошин, Х.Виссия // Искусственный интеллект. – 2008. – №3 – С.63-69.
2. Вальвачев, А.Н. Интеграция распределенных экспертных знаний: проблемы и решения / В.В. Краснопрошин, Г. Шаках // Информатика. – Минск, 2004. – №1. – С.45-53.
3. Информационные технологии регионального управления / Емельянов С.В. [и др.] – М.: Едиториал УРСС, 2004. – 400 с.
4. Якобсон, А. Унифицированный процесс разработки программного обеспечения / А. Якобсон, Г. Буч, Д. Рамбо. – СПб.: Питер, 2002. – 496 с.
5. Чен, П. Модель “сущность-связь” – шаг к единому представлению данных / П.Чен // СУБД. – 1995. – №3. – С.21-29.
6. Артемьева, И.Л. Сложно структурированные предметные области. Построение многоуровневых онтологий / И.Л. Артемьева // Информационные технологии. – 2009. – №1 – С.16-21.
7. Смирнов, С.В. Онтологии в задачах моделирования сложных систем / С.В. Смирнов // Проблемы управления и моделирования в сложных системах: Труды II международной конф. – Самара: Самарский НЦ РАН, 2000. – С. 66–72.