

ФОРМИРОВАНИЕ И ИСПОЛЬЗОВАНИЕ БАЗЫ ОБЩИХ ДАННЫХ ДЛЯ ВЫБОРА КОМПОНЕНТОВ ВЫЧИСЛИТЕЛЬНОЙ ИНФРАСТРУКТУРЫ С ЖЕЛАЕМЫМИ ХАРАКТЕРИСТИКАМИ ИЗ БАЗ ДАННЫХ РАЗЛИЧНЫХ ПОСТАВЩИКОВ

С. В. Павлов, Е. С. Брекоткина, В. Е. Брекоткин, А. С. Павлов

Аннотация. Качество информационных систем во многом определяется компоновкой (подбором компонентов с заданными характеристиками) их вычислительной инфраструктуры, которая, с одной стороны, улучшает производственные и управленческие процессы, а, с другой – изменяет уязвимость предприятия за счет появления новых угроз. Эти угрозы определяются включением в структуру организации новых более сложных компонентов (технических и организационных), нарушение нормального функционирования которых может привести к нарушению или неправильному функционированию предприятия. Вследствие этого представляется возможной и актуальной задача подбора компонентов вычислительной инфраструктуры предприятия на основе интеллектуального анализа возможных характеристик этих компонентов, представленных в запросах потребителей и базах данных поставщиков. В данной работе предложен метод формирования и использования базы общих данных (БОД) для выбора компонентов с желаемыми характеристиками, который за счет введенных взаимно однозначных соответствий между значениями желаемых характеристик компонентов, указанных в запросе потребителя, и фактическими значениями характеристик компонентов, имеющихся у поставщиков и информация о которых находится в БОД, позволяет осуществить поиск и последующий выбор требуемых компонентов, а также наиболее подходящих поставщиков этих компонентов. В дальнейшем совокупность предложенных критериев и методов позволит автоматизировать и тем самым ускорить процесс поиска информации о необходимых для компоновки вычислительной инфраструктуры компонентах. Предложенный метод формирования базы общих данных основан на построении взаимно однозначных соответствий между значениями желаемых характеристик компонентов, указанных в запросе потребителя, и фактическими значениями характеристик компонентов, имеющихся у поставщиков, и информация о которых находится в БОД. Использование этих соответствий в виде специальных отношений позволяет осуществить поиск и последующий выбор требуемых компонентов, а также наиболее подходящих поставщиков компонентов.

Ключевые слова: выбор компонентов сложных технических систем; распределенная база данных; база данных общих данных; синтез распределенной модели данных; отображение набора атрибутов; трансформация отношений.

ВВЕДЕНИЕ

В связи с ростом интенсивности и масштабов внедрения информационных систем в управление крупными предприятиями и организациями они сами становятся одной из важнейших составляющих этих предприятий, а разработка, внедрение и эксплуатация таких информационных систем становятся самостоятельной проблемой, имеющей организационные, технические, научные и финансово-экономические аспекты. Качество информационных систем во многом определяется компоновкой (подбором компонентов с заданными характеристиками) их вычислительной инфраструктуры, которая, с одной стороны, улучшает производственные и управленческие процессы, а, с другой – изменяет уязвимость предприятия за счет появления новых угроз. Эти угрозы определяются включением в структуру организации новых более сложных компонентов (технических и организационных), нарушение нормального функционирования которых может привести к нарушению или неправильному функционированию предприятия [1–4]. Вследствие этого представляется возможной и актуальной задача

подбора компонентов вычислительной инфраструктуры предприятия на основе интеллектуального анализа возможных характеристик этих компонентов, представленных в запросах потребителей и базах данных поставщиков.

В общем виде постановка этой задачи и некоторые этапы ее решения представлены в предыдущих работах авторов настоящей статьи [1, 2]. В этих работах показано, что для организации эффективного поиска информации о необходимых потребителю компонентах с заданными характеристиками необходимо проанализировать формы и способы структурирования, хранения и использования этой информации у всех возможных поставщиков и разработать механизмы (методы) их формализованного описания и взаимодействия между потребителями и поставщиками компонентов на основе создания базы общих данных и ее интеграции с базами данных всех поставщиков. Необходимо отметить, что потребители и поставщики компонентов создают и используют базы данных об этих компонентах и их характеристиках для своих производственных целей, вследствие чего их корпоративные базы данных об одних и тех же компонентах содержат, вообще говоря, различную информацию, и содержание баз данных отличается как по составу хранимой информации, так и по структуре. По результатам анализа информации из рассматриваемой предметной области предлагаются критерии классификации компонентов вычислительной инфраструктуры по возможным значениям их характеристик и разработка метода формального анализа возможных характеристик компонентов из распределенных баз данных с последующим синтезом единой базы данных желаемых и фактических характеристик компонентов.

В данной работе предложен метод формирования и использования базы общих данных для выбора компонентов с желаемыми характеристиками, который за счет введенных взаимно однозначных соответствий между значениями желаемых характеристик компонентов, указанных в запросе потребителя, и фактическими значениями характеристик компонентов, имеющихся у поставщиков и информация о которых находится в БОД, позволяет осуществить поиск и последующий выбор требуемых компонентов, а также наиболее подходящих поставщиков этих компонентов.

ФОРМАЛИЗАЦИЯ ЗАДАЧИ ВЫБОРА КОМПОНЕНТОВ С ЗАДАНЫМИ ХАРАКТЕРИСТИКАМИ ИЗ РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ

Для формализованного описания всей совокупности необходимой для осуществления выбора компонентов информации, отдельных ее частей и связей между ними воспользуемся введенными в [1] обозначениями: базу данных об имеющихся компонентах и их характеристиках у i -го поставщика обозначим через $БДП_i$, $i = \overline{1, n}$, где n – общее число поставщиков; БОД – база общих данных для всех потребителей и поставщиков данных, а всю распределенную совокупность необходимой для выбора компонентов информации обозначим через РБД.

Необходимо учитывать, что $БДП_i$, для всех $i = \overline{1, n}$ имеют, вообще говоря, различную структуру и информационное наполнение, так как отражает взгляды конкретного поставщика на всю необходимую в его деятельности информацию, в том числе – описание соответствующих компонентов и их характеристик, поставляемых данным поставщиком на рынок. БОД включает в себя информацию обо всех востребованных потребителями и имеющимися у всех, точнее – хотя бы у одного из всех, поставщиков компонентов.

Для описания предлагаемого метода выбора воспользуемся введенными в [1] обозначениями: будем считать, что для формирования вычислительной инфраструктуры используется множество компонентов

$$A = \{a_p\}, \quad p = \overline{1, N}, \quad (1)$$

где a_p – наименование компонента; N – количество различных видов компонентов. В свою очередь, все множество компонентов разбивается на NK взаимно непересекающихся классов

$$K = \{K_t\}, \quad t = \overline{1, NK}. \quad (2)$$

Каждый компонент попадает в некоторый класс и может быть отнесен только к этому классу, то есть для $\forall p_0: a_{p_0} \in A, \exists t_0 \in K_{t_0}$ и

$$\bigcup_{t=1}^{NK} K_t = A.$$

В БОД предлагается выделить независимые части БОД_t, содержащие общую информацию о компонентах и их характеристиках, принадлежащих отдельному t -му классу, а также сформировать некоторую связующую часть, назовем ее СБОД. В силу введенных обозначений

$$\text{РБД} = \text{БОД} \bigcup_{i=1}^n \text{БДП}_i = \text{СБОД} \bigcup \left(\bigcup_{t=1}^{NK} \text{БОД}_t \right) \bigcup \left(\bigcup_{i=1}^n \text{БДП}_i \right). \quad (3)$$

Каждая из БОД_t содержит в себе информацию обо всех характеристиках (идентификационных, технических и коммерческих) всех компонентов (независимо от поставщиков), отнесенных к t -му классу, а связующая часть – СБОД содержит, в соответствии с введенными в [1] критериями классификации, краткую описательную часть каждого класса, позволяющую определить отнесение любого компонента к тому или иному классу, и ссылку на соответствующую БОД_t. Обобщенная схема взаимосвязи введенных в рассмотрение частей РБД представлена на рис. 1.

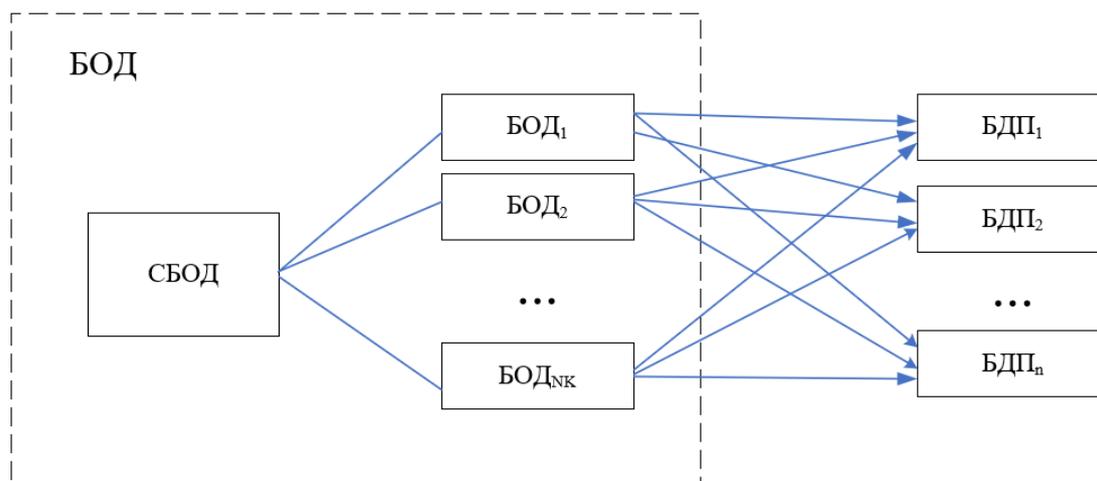


Рис. 1 Схема взаимосвязи частей РБД

В [2] показана возможность представления структуры каждой из БОД_t в следующем виде:

$$\text{БОД}_t = \{R_{t1}^0, R_{t2}^0, R_{t3}^0, C_{t1}^0\}, \quad (4)$$

где R_{t1}^0 – отношение, содержащее значения идентификационных характеристик; R_{t2}^0 – отношение, содержащее значения технических характеристик; R_{t3}^0 – отношение, содержащее информацию об условиях поставки и поставщиках компонентов; C_{t1}^0 – отношение, содержащее связи между данными, находящимися в отношениях $R_{t1}^0, R_{t2}^0, R_{t3}^0$, для каждого t -го класса, $t = \overline{1, NK}$.

Для того чтобы при организации выбора компонентов с заданными характеристиками можно было обратиться к соответствующей БОД_t, в СБОД создается некоторое связующее отношение RC_1 , которое содержит информацию обо всех классах компонентов и ссылки на соответствующие этим классам БОД_t. Предлагается следующая структура этого отношения

$$RC_1 = RC_1(KLC, K_t, t, \text{БОД}_t), \quad (5)$$

оно содержит в себе информацию о наименовании класса компонентов, его номере во всем множестве классов, наименовании и ссылке для перехода к БОД_t для t -го класса. В наиболее общем виде схема взаимосвязи этих частей БОД представлена на рис. 2.

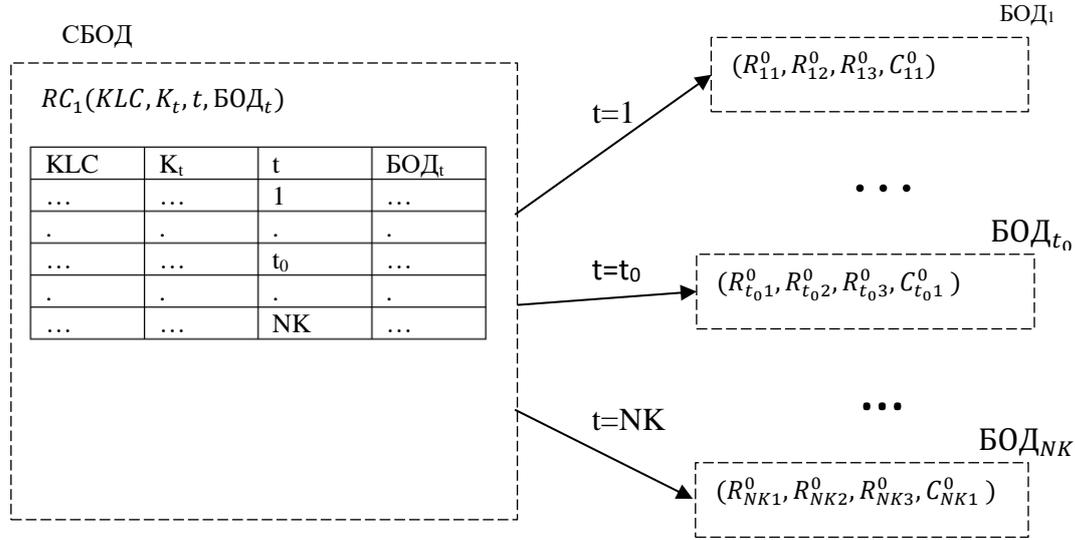


Рис. 2 Укрупненная схема взаимосвязи частей БОД

В дальнейшем для выбора компонентов с желаемыми характеристиками используются особенности предложенной структуры БОД и ее связей со множеством БДП_i, $i = \overline{1, n}$, более того эта структура как раз и разрабатывалась для обеспечения эффективности процедуры выбора необходимых компонентов и, по сути, является частью метода выбора.

Каждый компонент $a_p \in A$ описывается в [2] множеством характеристик

$$X_p = \{x_{pg}\}, \quad p = \overline{1, N}; \quad g = \overline{1, KN_p}, \quad (6)$$

где KN_p – количество характеристик p -го компонента, которые разделяются на три подмножества XI_p, XT_p, XK_p , так что

$$X_p = XI_p \cup XT_p \cup XK_p. \quad (7)$$

Во множество XI_p входят идентификационные характеристики из X_p во множество XT_p – технические, а во множество XK_p – коммерческие характеристики из X_p . Введем следующие обозначения: через GI_p – обозначим множество номеров характеристик $x_{pg} \in X_p$, отнесенных к множеству XI_p

$$GI_p = \{g_1, g_2, \dots, g_{KI_p}\}, \quad (8)$$

где $x_{pg} \in XI_p$, для $g \in GI_p$, KI_p – количество идентификационных характеристик компонента a_p ; GT_p – множество номеров характеристик $x_{pg} \in X_p$, отнесенных к множеству XT_p

$$GT_p = \{g_{KI_p+1}, g_{KI_p+2}, \dots, g_{KI_p+KT_p}\}, \quad (9)$$

где $x_{pg} \in XT_p$, для $g \in GT_p$, KT_p – количество технических характеристик компонента a_p ; GK_p – множество номеров характеристик $x_{pg} \in X_p$, отнесенных к множеству XK_p

$$GK_p = \{g_{K_p}, g_{K_p-1}, g_{K_p-2}, \dots, g_{K_p-KQ_p+1}\}, \quad (10)$$

где $x_{pg} \in XK_p$, для $g \in GK_p$, KQ_p – количество коммерческих характеристик компонента a_p . Из определения этих множеств следует, что

$$KI_p + KT_p + KQ_p = KN_p,$$

$$GI_p \bigcup GT_p \bigcup GK_p = \{1, 2, \dots, KN_p\} \quad (11)$$

для всех $p = \overline{1, N}$.

При выборе компонентов потребитель в своем запросе может указать (и чаще всего указывает) не все возможные характеристики и их желаемые значения, а только их часть, поэтому введем в рассмотрение множество используемых в запросе характеристик

$$X_p^* = \{x_{pg_1}, x_{pg_2}, \dots, x_{pg_{kk_p}}\}, \quad (12)$$

где $KK_p \leq KN_p$ и $X_p^* \subset X_p$. По аналогии с (7) для каждого $a_p \in A$ разобьём все множество используемых в запросе характеристик X_p^* на три части

$$X_p^* = XI_p^* \bigcup XT_p^* \bigcup XK_p^*, \quad (13)$$

где $XI_p^* \subset XI_p$ – множество идентификационных характеристик; $XT_p^* \subset XT_p$ – множество технических характеристик; $XK_p^* \subset XK_p$ – множество коммерческих характеристик, участвующих в запросе пользователя.

Для установления взаимосвязи между этими множествами по аналогии с (8)–(11) введем в рассмотрение множества номеров этих характеристик GI_p^* , GT_p^* и GK_p^* , каждое из которых содержит KI_p^* , KT_p^* и KQ_p^* элементов, так что

$$GI_p^* \subset GI_p, KI_p^* \leq KI_p, \quad (14)$$

$$GT_p^* \subset GT_p, KT_p^* \leq KT_p, \quad (15)$$

$$GK_p^* \subset GK_p, KQ_p^* \leq KQ_p, \quad (16)$$

$$KI_p^* + KT_p^* + KQ_p^* = KK_p \leq KN_p. \quad (17)$$

Учитывая тот факт, что x_{pg} – наименование g -й характеристики p -го компонента a_p , а компонент с таким наименованием (например $a_p = \text{«компьютер»}$) и с таким перечнем характеристик (но различными значениями этих характеристик) много, то для каждой характеристики каждого компонента существует множество ее возможных значений

$$Y_{pg} = \{y_{pg}^1, y_{pg}^2, \dots, y_{pg}^{KK_{pg}}\}, \quad (18)$$

где KK_{pg} – количество элементов этого множества для g -й характеристики p -го компонента, и соответствие между наименованием характеристики и множеством ее возможных значений будем обозначать

$$x_{pg} \rightarrow Y_{pg}. \quad (19)$$

Так как это соответствие справедливо для всех $g = \overline{1, K_p}$ характеристик для всех $p = \overline{1, N}$ компонентов, то в дальнейшем будет полезной еще одна группа множеств

$$Y_p = \{Y_{pg}\}_{g=\overline{1, KN_p}}, \quad (20)$$

для всех $p = \overline{1, N}$, которое вводится в рассмотрение для каждого компонента a_p и включает в себя набор множеств возможных значений всех характеристик X_p этого компонента. Соответствие множества характеристик и множеств их возможных значений будем обозначать тем же символом, что и в (19)

$$X_p \rightarrow Y_p. \quad (21)$$

Напомним, что x_{pg} – наименование характеристики, которое обозначается словом (например, «диагональ», «разрешение», «напряжение» др.), а значения этих характеристик могут

быть как символами, так и числами (например, «19», «1700×2600», и др.), поэтому с математической точки зрения и в соответствии с определением множеств X_{pg}, X_p^*, Y_{pg} , тот факт, что некоторая характеристика $x_{p_0g_0}$ принимает конкретное значение $y_{p_0g}^{s_0} \in Y_{p_0g_0}$, где $s_0 \in [1, KKK_{p_0g_0}]$, будем обозначать специально вводимым символом, отличающимся от (19)

$$x_{p_0g_0} \Rightarrow y_{p_0g}^{s_0} \in Y_{p_0g_0}. \quad (22)$$

И с учетом введенных обозначений требования (запрос) на выбор необходимого компонента можно представить в следующей форме: в РБД найти характеристики поставщиков, у которых для компонента с наименованием a_{p_0} значения множества характеристик $X_{p_0}^*$ определяются множеством

$$Y_{pg}^* = \{y_{p_0g_1}^{s_1}, y_{p_0g_2}^{s_2}, \dots, y_{p_0g_{KK_p}}^{s_{KK_p}}\}, \quad (23)$$

где

$$y_{p_0g_e}^{s_e} \in Y_{p_0g_e}, e = \overline{1, KK_p}, \quad (24)$$

то есть каждая характеристика $x_{p_0g_e}, e = \overline{1, KK_p}$ принимает некоторое конкретное значение $y_{p_0g_e}^{s_e}$ из своего множества возможных значений этой характеристики:

$$x_{p_0g_e} \Rightarrow y_{p_0g_e}^{s_e}. \quad (25)$$

или

$$X_{p_0}^* \rightarrow Y_{p_0g}^*. \quad (26)$$

Заметим, что за счет введения в рассмотрение множеств номеров (индексов) характеристик (8)–(10) и (14)–(16) отнесение конкретного множества Y_{pg} или элемента Y_{pg}^* к идентификационным, техническим или коммерческим характеристикам определяется принадлежностью значения индекса g к одному из множеств из (8)–(10), (14)–(16).

ОПРЕДЕЛЕНИЕ КЛАССА ТРЕБУЕМОГО КОМПОНЕНТА

Для определения принадлежности компонента a_{p_0} некоторому классу K_{t_0} в БОД предлагается включить справочник соответствия наименований компонентов и классов, который в рамках выбранной реляционной модели может быть представлен в виде специального отношения с достаточно простой структурой

$$RC_2 = RC_2(A, P, K, t). \quad (27)$$

Это отношение, наряду с RC_1 , входит в состав СБОД, $RC_2 \in$ СБОД, и в одном кортеже RC_2 находится наименование каждого возможного компонента a_p и наименование класса K_t , которому принадлежит этот компонент, $a_p \in K_t$.

Пример заполнения этого отношения приведен в табл. 1. Это справочное отношение позволяет по наименованию компонента, например, $a_{35} =$ «монитор», устанавливать номер и наименование класса, в данном случае $t_0 = 4$, $K_4 =$ «дисплей», которому принадлежит компонент с требуемым наименованием. Фактически отношение RC_2 является функцией, задаваемой в табличной форме и устанавливающей соответствие между множествами A и K :

$$A \xrightarrow{RC_2} K, \quad (28)$$

а именно – по заданному наименованию компонента a_{p_0} определяются наименование и номер класса K_{t_0} , к которому он относится, и с использованием RC_1 осуществляется переход к соответствующей БОД $_{t_0}$ для поиска конкретных экземпляров компонентов этого класса, удовлетворяющих требованиям запроса потребителя, то есть соотношений (25)–(26).

Таблица 1

Пример заполнения отношения RC_2 в виде таблицы

A	p	K	t
компьютерная мышь	1	мышь	1
мышь	2	мышь	1
мышь манипулятор	3	мышь	1
манипулятор	4	мышь	1
компьютерный манипулятор	5	мышь	1
мышка	6	мышь	1
...
видеодисплей	34	дисплей	4
монитор	35	дисплей	4
экран	36	дисплей	4
дисплей	37	дисплей	4
видеомонитор	38	дисплей	4
компьютерный дисплей	39	дисплей	4
компьютерный монитор	40	дисплей	4
устройство отображения	41	дисплей	4

С учетом введения справочных отношений RC_1 и RC_2 , а также справочного отношения PST , введенного в [2] и содержащего справочную информацию по всем известным поставщикам, СБОД будет иметь следующий вид:

$$\text{СБОД} = \{PST, RC_1, RC_2\}. \quad (29)$$

ОПРЕДЕЛЕНИЕ СООТВЕТСТВИЯ ЖЕЛАЕМЫХ ХАРАКТЕРИСТИК КОМПОНЕНТОВ И МЕСТА ХРАНЕНИЯ ИНФОРМАЦИИ ОБ ИХ ФАКТИЧЕСКИХ ЗНАЧЕНИЯХ

С технической точки зрения, реализация процедуры поиска осуществляется средствами выбранного программного обеспечения, например реляционной СУБД, а именно – осуществляется выполнение операции селекции по условиям совпадения элементов данных карточек с записями в запросе пользователя, а для этого необходимо определиться с тем, какие характеристики к какому отношению $БОД_{t_0}$ относятся и как они в нем называются. Другими словами, необходимо установить соответствие между множеством характеристик всех компонентов класса t_0 и множеством имен доменов отношений $R_{t_0,1}^0, R_{t_0,2}^0, R_{t_0,3}^0$ из (4).

Это соответствие устанавливается с использованием метода формирования структуры БОД, который изначально предполагал организацию хранения именно этих видов характеристик, поэтому отношение $R_{t_0,1}^0$ предназначено для хранения значений идентификационных характеристик, отношение $R_{t_0,2}^0$ – технических характеристик и отношение $R_{t_0,3}^0$ – коммерческих характеристик компонентов a_p , отнесенных к этому классу, для $p_0: a_{p_0} \in K_{t_0}$, множество значений характеристик XI_{p_0} размещается в отношении $R_{t_0,1}^0$, множество значений характеристик XT_{p_0} – в отношении $R_{t_0,2}^0$, множество значений характеристик XK_{p_0} – в отношении $R_{t_0,3}^0$. При этом в каждом домене соответствующего отношения размещаются фактические (то есть присутствующие в описании данного компонента характеристики хотя бы у одного поставщика) значения одной характеристики, то есть в некотором домене D_{t_0,j_0}^0 осуществляется хранение значений некоторой характеристики $x_{p_0,g_0} \in XI_{p_0}$ и необходимо задать это соответствие, то есть каждому номеру домена j поставить в соответствие номер характеристики g .

В дальнейшем нам понадобится структура всех отношений, входящих в (4), обоснование и построение которой подробно приводится в [2] и имеет следующий вид:

$$R_{tj} = R_{tj}(D_{tj}^0, D_{tj}^1, D_{tj}^2, \dots, D_{tj}^{m_{tj}}), \quad (30)$$

где D_{tj}^l – l -й домен, j -го отношения t -го класса; $j = \overline{1,3}$; $t = \overline{1, NK}$; $l = \overline{1, m_{tj}}$, предназначенный для хранения значений характеристик, D_{tj}^0 – ключевой домен отношения; m_{tj} – количество доменов каждого из трех отношений каждого класса. Так как компоненты различных классов описываются значениями, вообще говоря, различного количества компонентов, то $m_{t_1 j_1} \neq m_{t_2 j_2}$, для $t_1 \neq t_2$ или $j_1 \neq j_2$. Тогда хранение значений характеристик некоторой компоненты $a_{p_0} \in A$, осуществляется в одном из доменов (30), первый индекс t определяется номером класса, в который попал компонент:

$$t = t_0: a_{p_0} \in K_{t_0}. \quad (31)$$

Второй индекс определяется номером характеристики в соответствующей группе (то есть в одном из множеств GI , GT и GK), к которой относится характеристика $x_{p_0 g_0}$ этого компонента, обозначим этот номер (индекс) i_0 :

$$\begin{aligned} &\text{если } x_{p_0 g_0} \in XI_{p_0}, \text{ тогда } i_0: g_{i_0} \in GI_{p_0} \text{ и } j = 1; \\ &\text{если } x_{p_0 g_0} \in XT_{p_0}, \text{ тогда } i_0: g_{i_0} \in GT_{p_0} \text{ и } j = 2; \\ &\text{если } x_{p_0 g_0} \in XK_{p_0}, \text{ тогда } i_0: g_{i_0} \in GK_{p_0} \text{ и } j = 3. \end{aligned} \quad (32)$$

Так как в соответствии с одним из основных свойств отношений порядок размещения доменов в отношении не влияет на его информационное содержание, для определенности будем размещать значения характеристики $x_{p_0 g_0}$, номер которой g_0 – во множестве X_{p_0} , во множестве GI_{p_0} , GT_{p_0} или GK_{p_0} (в зависимости от того, в которое из них попал этот номер в соотношениях (32)) этот номер стоит на i_0 месте в j -м домене соответствующего отношения. Для обозначения этого соответствия наименования характеристики и домена, в котором хранятся ее фактические значения у различных поставщиков, вполне подходит введенный в (19) символ, а само соответствие запишется для трех случаев:

$$\begin{aligned} &\text{если } x_{p_0 g_0} \in XI_{p_0} \text{ и } g_{i_0} \in GI_{p_0}, \text{ то } x_{p_0 g_0} \rightarrow D_{t_0 1}^{i_0}; \\ &\text{если } x_{p_0 g_0} \in XT_{p_0} \text{ и } g_{i_0} \in GT_{p_0}, \text{ то } x_{p_0 g_0} \rightarrow D_{t_0 2}^{i_0}; \\ &\text{если } x_{p_0 g_0} \in XK_{p_0} \text{ и } g_{i_0} \in GK_{p_0}, \text{ то } x_{p_0 g_0} \rightarrow D_{t_0 3}^{i_0}. \end{aligned} \quad (33)$$

Значения каждой характеристики $x_{p_0 g_0}$ хранятся в домене $D_{t_0 1}^{i_0}$ отношения $R_{t_0 1}$, а номер домена i_0 определяется исходя из места номера g_0 во множестве GI_{p_0} , то есть число g_0 находится во множестве GI_{p_0} на i_0 месте, так что

$$g_0 = g_{i_0}. \quad (34)$$

Аналогичным образом устанавливается соответствие между техническими и коммерческими характеристиками и местом их хранения (номером домена) в отношениях R_{t_2} и R_{t_3} соответственно.

На основе множеств GI_p , GT_p , GK_p вводятся в рассмотрение функции, отображающие множество наименований характеристик во множество имен доменов соответствующих отношений, в которых хранятся значения этих характеристик:

$$X_p \xrightarrow{FGI_p} \{D_{t_1}^i\}_{i=\overline{1, m_{t_1}}}; \quad X_p \xrightarrow{FGT_p} \{D_{t_2}^i\}_{i=\overline{1, m_{t_2}}}; \quad X_p \xrightarrow{FGK_p} \{D_{t_3}^i\}_{i=\overline{1, m_{t_3}}}. \quad (35)$$

Эти функции FGI_p , FGT_p , FGK_p по аналогии с RC_1 и RC_2 (см. (5) и (27)) задаются в табличной форме, структура и пример заполнения этой функции для идентификационных характеристик компонента «мышь» приведен в табл. 2.

Таблица 2

Пример функции FGI_p для компонента «мышь»

Наименование характеристики	Номер характеристики в X_p	Номер характеристики в GI_p	Номер места домена в отношении R_{t1}	Домен
x_{p0g}	g	i	i	D_{t01}^i
Код компонента	2	1	1	D_{11}^1
Наименование	3	2	2	D_{11}^2
Наименование группы	4	3	3	D_{11}^3
Бренд	8	4	4	D_{11}^4
КТРУ	16	5	5	D_{11}^5
Модель	17	6	6	D_{11}^6
Артикул	24	7	7	D_{11}^7
Код товарной позиции	27	8	8	D_{11}^8
Серия	34	9	9	D_{11}^9

Так как каждая компонента принадлежит одному конкретному классу, то значения ее характеристик хранятся в одной конкретной БОД_t, а функции FGI_p , FGT_p , FGK_p имеют фиксированный вид (одну и ту же таблицу) для всех компонентов этого класса, поэтому хранятся эти функции в виде специализированных справочных отношений в БОД_t. Наименование отношений совпадает с наименованием функций, а структура каждой БОД_t изменится по отношению к (4) с учетом (30)

$$\text{БОД}_t = \{R_{t1}^0, R_{t2}^0, R_{t3}^0, C_{t1}^0, FGI_t, FGT_t, FGK_t\}. \quad (36)$$

После создания новых справочных отношений структура данных БОД будет иметь следующий вид

$$\text{БОД} = \{PST, RC_1, RC_2\} \cup \left(\bigcup_{t=1}^{NK} \{R_{t1}, R_{t2}, R_{t3}, C_{t1}, FGI_t, FGT_t, FGK_t\} \right). \quad (37)$$

ОРГАНИЗАЦИЯ ПОИСКА ХАРАКТЕРИСТИК С ЖЕЛАЕМЫМИ ЗНАЧЕНИЯМИ

Все возможные значения характеристики x_{pg} в соответствии с (18)–(21) описываются множеством Y_{pg} , и, следовательно, в соответствующем домене БОД_t хранятся элементы этого множества, но так как множества желаемых и фактических характеристик, вообще говоря, не совпадают, то есть не всегда компоненты с желаемыми характеристиками могут быть поставлены хотя бы одним поставщиком, то

$$Y_{pg} \neq \bigcup_{j=1}^3 \bigcup_{l=1}^{m_{tj}} D_{tj}^l. \quad (38)$$

Физический смысл элементов множеств Y_{pg} и D_{tj}^l в соответствии с их определениями одинаков: они содержат значения соответствующих характеристик. Но с точки зрения описания метода выбора компонентов с желаемыми характеристиками они несут разную смысловую нагрузку: множество Y_{pg} содержит желаемые для потребителей значения характеристик,

а множество D_{tj}^l – фактические значения характеристик компонентов, имеющих у поставщиков. И задача выбора заключается в том, чтобы в базе данных БОД_t найти все кортежи всех отношений, в которых фактические значения каждой характеристики совпадают (или близки) с заданными в запросе, то есть с Y_{pq}^* (см. (23), (24)). Введем в рассмотрение множество элементов каждого домена, фактически присутствующих в заполненном на данный момент времени варианте отношения R_{tj}

$$DD_{tj}^i = \{d_{tj}^{il}\}_{l=1, kd_{tj}}, \quad (39)$$

где $d_{tj}^{il} \in D_{tj}^i$, kd_{tj} – количество кортежей j -го отношения в БОД_t, $j = \overline{1,3}$; $t = \overline{1, NK}$, $i = \overline{1, m_{tj}}$. По сути, DD_{tj}^i – это столбец, содержащий значения домена D_{tj}^i в табличной интерпретации отношения R_{tj} . Множества D_{tj}^i и DD_{tj}^i состоят из одних и тех же элементов – значений некоторой характеристики, но формально они не совпадают

$$DD_{tj}^i \neq D_{tj}^i. \quad (40)$$

Во множестве D_{tj}^i каждый элемент – возможное значение характеристики, встречается только один раз, а во множестве DD_{tj}^i каждый элемент соответствует заполненному полю столбца для определенного кортежа, поэтому одно и то же фактическое значение характеристики может встречаться несколько раз. Тогда поисковый критерий формулируется следующим образом: в БОД_{t₀} в каждом из отношений найти множество кортежей, для которых выполняется условие

$$d_{tj}^{il} = y_{pg}^i, \quad (41)$$

для всех $y_{pg}^i \in Y_{pg}^*$ и $d_{tj}^{il} \in DD_{tj}^i$; $j = \overline{1,3}$; $l = \overline{1, kd_{tj}}$. Множество ключей этих кортежей обозначим

$$DK_{tj}^0 = \{d_{tj}^{0ls}\}_{s=1, kkd_{tj}}, \quad (42)$$

где kkd_{tj} – общее количество найденных кортежей (и следовательно – их ключей); $kkd_{tj} \leq kd_{tj}$, а множество номеров элементов этого множества

$$LD_{tj} = \{l_1, l_2, \dots, l_{kkd_{tj}}\} \quad (43)$$

содержит порядковые номера кортежей в табличной интерпретации отношения R_{tj} , то есть номера строк таблицы, для которых выполняется условие (41) и

$$DK_{tj}^0 \subset D_{tj}^0, \quad (44)$$

а каждый элемент этого множества определяет именно тот кортеж, все элементы которого, участвующие в запросе потребителя (то есть с номерами характеристик из (32)), удовлетворяют условию (41). Поясним физический смысл множеств DK_{tj}^0 : DK_{t1}^0 – множество ключей отношения R_{t1} (чаще всего это множество номеров строк таблицы, интерпретирующей отношение R_{t1}), определяющих кортежи с желаемыми значениями идентификационных характеристик; DK_{t2}^0 – множество ключей отношения R_{t2} , определяющих кортежи с желаемыми значениями технических характеристик; DK_{t3}^0 – множество ключей отношения R_{t3} , определяющих кортежи с желаемыми значениями характеристик условий поставки.

Далее, используя отношение связи C_{t1}^0 и выполняя операцию селекции, находим в нем все кортежи, содержащие значения ключей, входящих во множества DK_{tj}^0 , $j = \overline{1,3}$, то есть удовлетворяющих условию (41). Обозначим результирующее отношение CK_t , оно по определению операции селекции имеет ту же структуру, что и C_{t1}^0 , то есть

$$CK_t = CK_t(D_{t1}^0, D_{t2}^0, D_{t3}^0, IDP), \quad (45)$$

и содержит необходимую информацию о множестве различных экземпляров требуемого компонента a_{p_0} с набором желаемых характеристик $Y_{p_0g}^*$, поставляемых всеми известными поставщиками. Очевидно, что в отношении CK_t количество кортежей может быть существенно больше одного, так как, во-первых, этот компонент может поставляться более чем одним (а зачастую, почти всеми) поставщиком, а во-вторых, так как KK_{p_0} (напомним, что это количество указываемых в запросе характеристик, введенное в (11)) меньше или равно K_{p_0} (это общее количество известных характеристик компонента a_{p_0} , введенное в (6)), то при равных значениях KK_{p_0} характеристик, остальные $K_{p_0} - KK_{p_0}$ характеристики могут иметь различные значения. Второе замечание указывает на то, что и один поставщик может поставлять несколько компонентов с одинаковыми значениями заданных в запросе характеристик, но с различными значениями других (не указанных в запросе, но существующих в БОД_{*t*}) характеристик.

ЗАКЛЮЧЕНИЕ

Предложенный метод формирования базы общих данных основан на построении взаимно однозначных соответствий между значениями желаемых характеристик компонентов, указанных в запросе потребителя, и фактическими значениями характеристик компонентов, имеющихся у поставщиков, и информация о которых находится в БОД. Использование этих соответствий в виде специальных отношений, входящих в СБОД и БОД_{*t*}, $t = \overline{1, NK}$, позволяет осуществить поиск и последующий выбор требуемых компонентов, а также наиболее подходящих поставщиков компонентов, информация о которых находится в БДП_{*i*}, $i = \overline{1, n}$. В дальнейшем совокупность предложенных критериев и методов позволит автоматизировать и тем самым ускорить процесс поиска информации о необходимых для компоновки вычислительной инфраструктуры компонентах.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Брекоткин В. Е., Брекоткина Е. С., Павлов А. С., Павлов С. В. Классификация компонентов компьютерной инфраструктуры распределенной организации на основе интеллектуального анализа и структурирования их характеристик // International Journal of Open Information Technologies. 2022. Т. 10. № 10. С. 101-110. EDN RVACMO. [[Brekotkin V. E., Brekotkina E. S., Pavlov A. S., Pavlov S. V. Classification of components of the computer infrastructure of a distributed organization based on intellectual analysis and structuring of their characteristics // International Journal of Open Information Technologies. 2022. Vol. 10, No. 10, pp. 101-110. EDN RVACMO. (In Russian).]]
2. Брекоткин В. Е., Брекоткина Е. С., Павлов А. С., Павлов С. В. Информационная технология подбора компонентов сложных технических систем на основе интеллектуального анализа их желаемых и фактических характеристик из распределенных баз данных // Системная инженерия и информационные технологии. 2022. Т. 4. № 2 (9). С. 13-23. EDN XGTHJD. [[Brekotkin V. E., Brekotkina E. S., Pavlov A. S., Pavlov S. V. Information technology for selecting components of complex technical systems based on intelligent analysis of their desired and actual characteristics from distributed databases // Systems Engineering and Information Technologies. 2022. Vol. 4, No. 2 (9), pp. 13-23. EDN XGTHJD. (In Russian).]]
3. Брекоткина Е. С., Павлов С. В., Трубин В. Д., Христуло О. И. Управление проектом создания сложной геоинформационной системы на основе специализированного справочника // Системная инженерия и информационные технологии. 2021. Т. 3. № 2(6). С. 66-73. DOI 10.54708/26585014_2021_32666. EDN HQSJSJ. [[Brekotkina E. S., Pavlov S. V., Trubin V. D., Christodoulo O. I. Project management for creating a complex geoinformation system based on a specialized reference book // System Engineering and Information Technologies. 2021. Vol. 3, No. 2(6). pp. 66-73. DOI 10.54708/26585014_2021_32666. EDN HQSJSJ. (In Russian).]]
4. Брекоткина Е. С., Павлов А. С., Павлов С. В., Христуло О. И. Анализ и синтез структуры пространственных данных для управления сложными распределенными системами // Информационные технологии и системы: Тр. Восьмой Всероссийской научной конференции с международным участием, Ханты-Мансийск, 17–21 марта 2020 года. Ханты-Мансийск, 2020. С. 120-125. EDN IMQENU. [[Brekotkina E. S., Pavlov A. S., Pavlov S. V., Christodoulo O. I. Analysis and synthesis of spatial data structure for managing complex distributed systems // Information technologies and systems: proceedings of the Eighth All-Russian scientific conference with international participation, Khanty-Mansiysk, March 17–21, 2020. Khanty-Mansiysk, 2020. pp. 120-125. EDN IMQENU. (In Russian).]]

Поступила в редакцию 1 марта 2024 г.

МЕТАДАННЫЕ / METADATA

Title: Formation and use of a database of common data for the selection of components of the computing infrastructure with the desired characteristics from databases of various suppliers.

Abstract: The quality of information systems is largely determined by the layout (selection of components with specified characteristics) of their computing infrastructure, which, on the one hand, improves production and management processes, and, on the other hand, changes the vulnerability of the enterprise due to the emergence of new threats. These threats are determined by the inclusion of new, more complex components (technical and organizational) in the structure of the organization, the disruption of the normal functioning of which can lead to disruption or improper functioning of the enterprise. As a result, it seems possible and urgent to select the components of the enterprise's computing infrastructure based on an intelligent analysis of the possible characteristics of these components presented in consumer requests and vendor databases. In this paper, a method is proposed for the formation and use of a database of general data for the selection of components with desired characteristics, which, due to the introduced one-to-one correspondences between the values of the desired characteristics of the components specified in the consumer's request and the actual values of the characteristics of the components available from suppliers and information about which is in the database, allows to search for and then select the required components, as well as the most suitable suppliers of these components. In the future, the set of proposed criteria and methods will allow automating and thereby speeding up the process of searching for information about the components necessary for the layout of the computing infrastructure. The proposed method of forming a database of general data is based on the construction of one-to-one correspondences between the values of the desired characteristics of the components specified in the consumer's request and the actual values of the characteristics of the components available from suppliers and information about which is in the database. The use of these correspondences in the form of special relationships included in the SBOD and BDT allows you to search for and then select the required components, as well as the most suitable component suppliers, information about which is available in the BDI.

Key words: selection of components of complex technical systems; distributed database; database of general data; synthesis of a distributed data model; mapping of a set of attributes; transformation of relations.

Язык статьи / Language: русский / Russian.

Об авторах / About the authors:

ПАВЛОВ Сергей Владимирович

ФГБОУ ВО «Уфимский университет науки и технологий», Россия.
Проф. каф. геоинформационных технологий.
E-mail: psvgis@mail.ru
URL: https://elibrary.ru/author_items.asp?authorid=108929

БРЕКОТКИНА Елена Сергеевна

ООО «Атриум», Россия.
Бизнес-аналитик.
E-mail: brekotkina@mail.ru
URL: https://elibrary.ru/author_items.asp?authorid=781961

БРЕКОТКИН Валерий Евгеньевич

ООО «Медиалюкс», Россия.
Директор.
E-mail: veb@ya.ru
URL: https://elibrary.ru/author_items.asp?authorid=781961

ПАВЛОВ Александр Сергеевич

ООО «ГК ИННОТЕХ», Россия.
Вед. сист. аналитик, дивизион технол. развития розн. бизнеса.
E-mail: asp.gis@gmail.com
URL: https://elibrary.ru/author_items.asp?authorid=743354

PAVLOV Sergey Vladimirovich

Ufa University of Science and Technologies, Russia.
Professor of the Department of Geoinformation Technologies.
E-mail: psvgis@mail.ru
URL: https://elibrary.ru/author_items.asp?authorid=108929

BREKOTKINA Elena Sergeevna

Atrium LLC, Russia.
Business analyst.
E-mail: brekotkina@mail.ru
URL: https://elibrary.ru/author_items.asp?authorid=781961

BREKOTKIN Valeriy Evgenyevich

Medialux LLC, Russia.
Director.
E-mail: veb@ya.ru
URL: https://elibrary.ru/author_items.asp?authorid=781961

PAVLOV Aleksandr Sergeevich

LLC "GK INNOTECH", Russia.
Leading systems analyst, Retail Business Development Division.
E-mail: asp.gis@gmail.com
URL: https://elibrary.ru/author_items.asp?authorid=743354