

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПОДБОРА КОМПОНЕНТОВ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ИХ ЖЕЛАЕМЫХ И ФАКТИЧЕСКИХ ХАРАКТЕРИСТИК ИЗ РАСПРЕДЕЛЕННЫХ БАЗ ДАННЫХ

В. Е. БРЕКОТКИН ¹, Е. С. БРЕКОТКИНА ², А. С. ПАВЛОВ ³, С. В. ПАВЛОВ ⁴

¹ a3um@ a3um.ru, ² brekotkina@mail.ru, ³ asp.gis@gmail.com, ⁴ psvglis@mail.ru.ru

¹ ООО «Атриум»

^{2,4} ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

³ ПАО Банк «ФК Открытие»

Поступила в редакцию 2.09.2022

Аннотация. Одним из важных направлений повышения эффективности создания и использования сложных технических систем является определение соответствия желаемых характеристик этих систем и входящих в них компонентов их характеристикам в настоящем и будущем. Сложность этой проблемы определяется сложностью составляющих компонентов, большим количеством характеристик этих компонентов и систем в целом, нечеткостью требований к перечню и значениям характеристик со стороны различных участников создания и использования сложных систем. При этом возникает двойственность ситуации: с одной стороны используются большие объемы разнородных данных, на сбор и обработку которых требуется много времени и человеческих ресурсов, с другой стороны – имеются существенные ограничения на время принятия решений на основе этих данных. Предлагается решение задачи выбора компонентов сложных систем на различных этапах их создания (проектирование, разработка, реализация) и эксплуатации, основанное на интеллектуальном анализе всего множества возможных характеристик этих компонентов, хранящихся в множестве разнородных (по структуре, ведомственной принадлежности, средствам реализации) баз данных. Разработаны метод формирования структуры распределенной базы данных с описанием компонентов сложных технических систем и их характеристик и процедура построения модели общих данных, позволяющая осуществлять взаимно однозначное соответствие атрибутов, описывающих характеристики компонентов в различных частях распределенной базы данных: баз данных поставщиков и базы общих для всех потребителей и поставщиков данных. Методы предназначены для их последующей реализации в рамках информационной системы, обеспечивающей информационную поддержку деятельности различных категорий специалистов (конструкторов, технологов, снабженцев и др.), осуществляющих подбор необходимых компонентов с заданными (желаемыми) характеристиками.

Ключевые слова: подбор компонентов сложных технических систем; распределенная база данных; база общих данных; синтез модели распределенных данных; отображение множества атрибутов; преобразование отношений.

ВВЕДЕНИЕ

Одним из важных направлений повышения эффективности создания и использования сложных технических систем является определение соответствия желаемых характеристик этих систем и входящих в них компонентов их характеристикам в настоящем и будущем. Сложность этой проблемы определяется сложностью составляющих компонентов, большим количеством характеристик этих компонентов и систем в целом, нечеткостью требований к перечню и значениям характеристик со стороны различных участников создания и использования сложных систем. При этом возникает двойственность ситуации: с одной стороны используются большие объемы разнородных данных, на сбор и обработку которых требуется много времени и человеческих ресурсов, с другой стороны – имеются существенные ограничения на время принятия решений на основе этих данных. Таким образом разработка новых цифровых технологий, обеспечивающих информационную поддержку выбора компонентов сложных технических систем на основе интеллектуального анализа их желаемых и фактических характеристик из распределенных баз данных является важной и актуальной научной и технологической задачей [8, 9]. В данной статье предлагается решение задачи выбора компонентов сложных систем на различных этапах их создания (проектирование, разработка, реализация и эксплуатации), основанное на интеллектуальном анализе всего множества возможных характеристик этих компонентов, хранящихся в множестве разнородных (по структуре, ведомственной принадлежности, средствам реализации) баз данных. Предлагается развитие теоретико-множественного подхода к формальному описанию баз данных желаемых характеристик компонентов, а также баз данных фактических характеристик этих компонентов, в результате чего создается единая распределенная база

данных желаемых и фактических характеристик. На основе методов теории систем и системного анализа формируются критерии и методы выбора компонентов, фактические характеристики которых наиболее соответствуют желаемым для данной сложности системы. При этом необходимо обеспечить интеграцию моделей и методов формального описания больших разнородных данных, включая новые источники и технологии получения и предоставления данных, а также методов анализа и выбора требуемых комбинаций данных, максимально соответствующих желаемым характеристикам компонентов [9,10,11]. Предлагаемые в данной статье методы предназначены для их последующей реализации в рамках информационной системы, обеспечивающей информационную поддержку деятельности различных категорий специалистов (конструкторов, технологов, снабженцев и др.), осуществляющих подбор необходимых компонентов с заданными (желаемыми) характеристиками.

ПОСТАНОВКА ЗАДАЧИ ПОДБОРА КОМПОНЕНТОВ СЛОЖНЫХ ТЕХНИЧЕСКИХ СИСТЕМ

В качестве примера предметной области рассматривается задача подбора компонентов вычислительной инфраструктуры некоторого крупного подразделения большой организации (в нашем примере это кафедра геоинформационных систем «Уфимского государственного авиационного технического университета»). Пусть в результате анализа потребностей пользователей [3, 4] разработана структура вычислительной инфраструктуры этого подразделения, включающая фактическое размещение ее компонентов по территории (помещениям) этого подразделения, приведенная на рис. 1. Технические характеристики каждого компонента определяются (предлагаются) системным интегратором – разработчиком вычислительной инфраструктуры исходя из потребностей пользователей и решаемых ими задач.

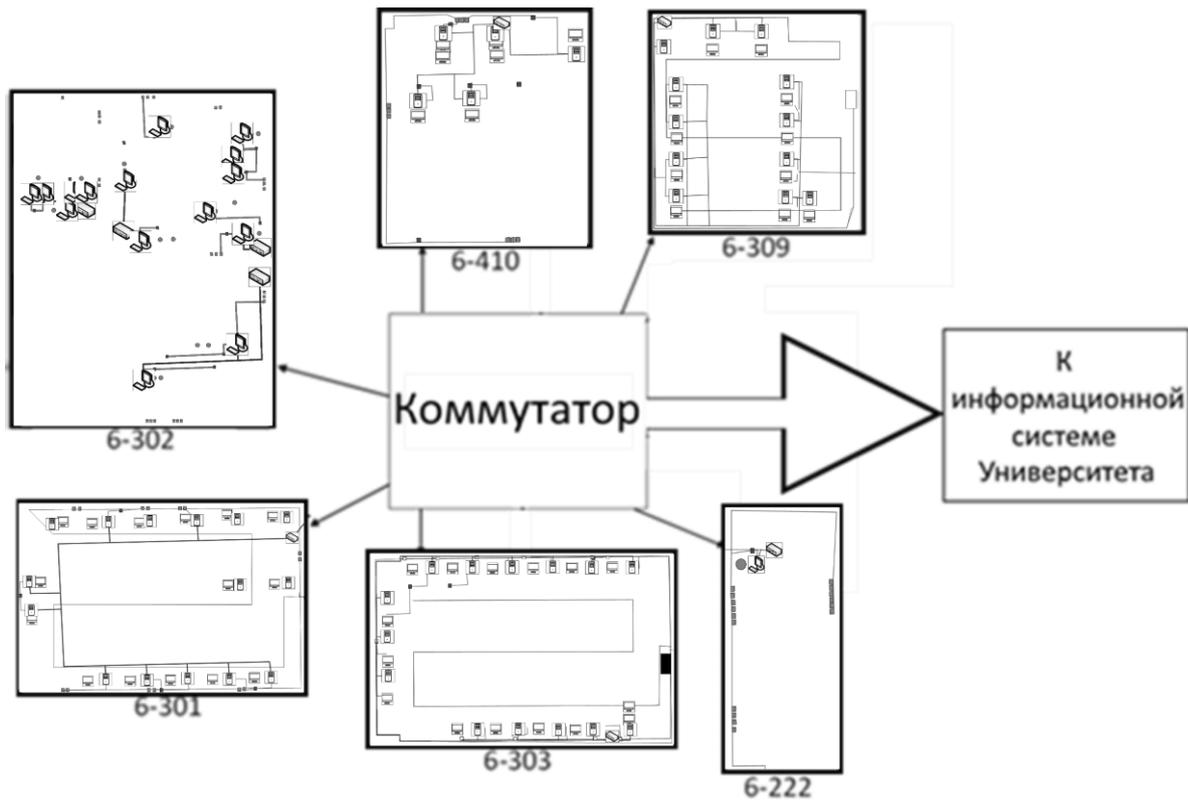


Рис. 1. Размещение компонентов вычислительной инфраструктуры по помещениям подразделения

Для практической реализации вычислительной инфраструктуры необходимо выбрать поставщиков ее компонентов с наибольшим соответствием желаемых и фактических технических характеристик и с учетом некоторых финансовых и организационных ограничений: стоимости компонентов и их доставки, сроков поставки, гарантии, необходимости предоплаты и др. Учитывая, что такая задача подбора компонентов решается даже в одной организации многократно, а также большое количество потенциальных поставщиков необходимых компонентов, автоматизация решения этой задачи представляется актуальной.

При решении этой задачи используется информация о необходимых потребителю компонентах с желаемыми характеристиками и информация о существующих (доступных к приобретению) компонентах с фактическими характеристиками у различных поставщиков. Учитывая, что потребители и поставщики для принятия решений

(для каждого из них свои) используют информацию об одних и тех же объектах – компонентах вычислительной инфраструктуры – необходимо обеспечить технологическую и информационную среду для информационного обмена между потребителями и поставщиками, позволяющую сократить финансовые и временные затраты на поиск необходимой информации для подбора компонентов. Для организации хранения, обновления и доступа к этой информации всех заинтересованных пользователей (в первую очередь потребителей искомых компонентов) предлагается разработать информационную систему (структура которой представлена на рис. 2), обеспечивающую многопользовательскую обработку распределенно-хранящейся информации о потребностях и наличии искомых компонентов (сокращенно данную систему обозначим ИСПК) в заданной предметной области (в нашем случае – вычислительная инфраструктура подразделения).

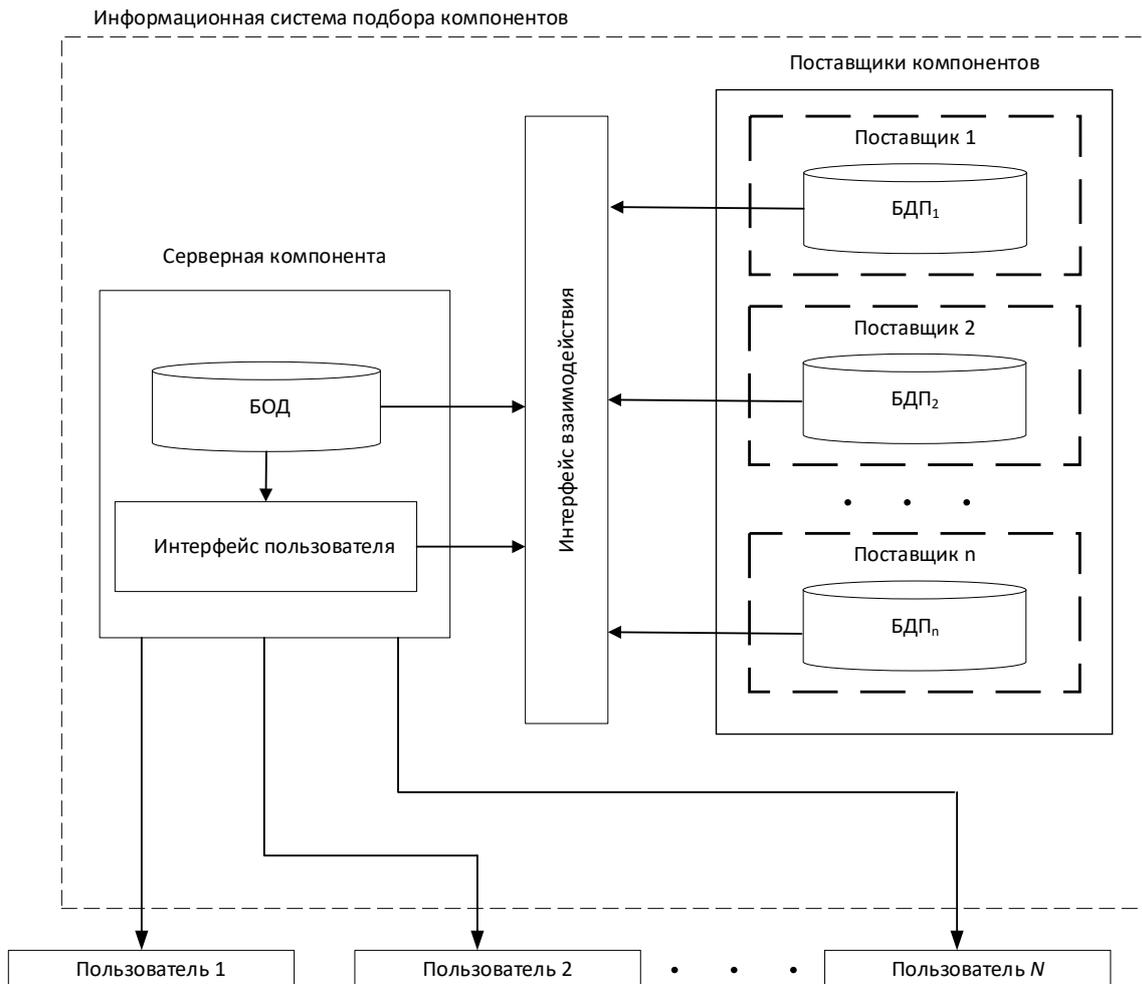


Рис. 2. Обобщенная структура информационной системы подбора компонентов

Одним из ключевых, трудоемких и сложных в научном плане моментов при разработке ИСПК является формирование структуры распределенной базы данных (РБД), для чего необходимо произвести анализ требований к информации о компонентах и их характеристиках со стороны всех участников ИСПК – как поставщиков так и потребителей, выявить общую для всех участников процесса информацию и перенести ее в базу общих данных (БОД), размещенную на серверной части ИСПК, сохраняя при этом ее полноту и непротиворечивость. Для этих целей предлагается теоретико-множественный подход к формальному описанию и анализу структуры множества взаимосвязанных баз данных и синтезу структуры базы общих данных на основе взаимно-однозначного (функционального) преобразования структур всех используемых данных, который ранее использовался авторами данной статьи в других предметных областях [1–3].

СИНТЕЗ СТРУКТУРЫ БАЗЫ ОБЩИХ ДАННЫХ НА ОСНОВЕ АНАЛИЗА СТРУКТУР МНОЖЕСТВА ВЗАИМОСВЯЗАННЫХ БАЗ ДАННЫХ ПОСТАВЩИКОВ

Для формализованного описания всей совокупности необходимой информации, отдельных ее частей и связей между ними введем обозначения: 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 (например, наименование компонентов, наименование характеристик компонентов и др.), а всю распределенную совокупность информации обозначим через РБД, тогда

$$РБД = БОД \bigcup_{i=1}^n БДП_i. \quad (1)$$

Необходимо отметить 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.

Каждая из принятых к рассмотрению баз данных разрабатывается на основе соответствующей модели данных (МД), поэтому обозначим модель каждой из БДП_{*i*} через МДП_{*i*}, $i = \overline{1, n}$ причем МДП_{*i*} ≠ МДП_{*j*} при $i \neq j$, модель данных БОД – через МОД, а модель данных РБД – через МРД. Каждая база данных состоит из множества отношений различной структуры и связей между ними, что может быть отображено соотношением

$$\text{МД}_i = \{(R_1^i, R_2^i, \dots, R_{k_i}^i, C_1^i, C_2^i, \dots, C_{t_i}^i)\}, \quad (2)$$

для всех $i = \overline{1, n}$. Здесь k_i – количество отношений; t_i – количество связей между отношениями i -ой МДП_{*i*}. Аналогичные обозначения вводятся для БОД

$$\text{МОД} = \{(R_1^0, R_2^0, \dots, R_{k_0}^0, C_1^0, C_2^0, \dots, C_{t_0}^0)\}, \quad (3)$$

где k_0 – количество отношений; t_0 – количество связей между отношениями в МОД. У каждого поставщика (например [5–7] в его модели данных МД_{*i*} присутствует информация с описанием всех поставляемых компонентов, их технических характеристик и условий поставки, представленных в соответствующих отношениях, при этом количество и структура этих отношений у различных поставщиков может различаться.

Предположим, что информация о каждой из основных групп характеристик компонентов (идентификационных, технических, условий поставки) компактно размещается в каждой из баз данных БДП_{*i*} в отдельном отношении. Первое из входящих в МД_{*i*} отношений содержит в себе некоторую общую информацию о каждом компоненте (будем называть ее идентификационной), которая

позволяет однозначно идентифицировать его во всем множестве компонентов, поставляемых данным поставщиком (например: наименование, производитель, наименование группы, код производителя и др.). Пусть это отношение входит в множество (2) под номером 1 и содержит в себе $m_1^i + 1$ доменов, тогда это отношение можно обозначить

$$R_1^i = R_1^i (KL_1^i, A_{1,1}^i, A_{1,2}^i, \dots, A_{1,m_1^i}^i), \quad (4)$$

где $A_{1,j}^i, j = \overline{1, m_1^i}$, наименование j -го атрибута (домена) содержащего значение j -ой идентификационной характеристики, а каждая строка этого отношения содержит значения всех идентификационных характеристик одного компонента; KL_1^i – ключевой домен этого отношения, который не имеет физического смысла, но однозначно определяет конкретный экземпляр компонента из всего их множества для данного поставщика и вводится для организации связи с другими видами информации в БДП_{*i*}.

Второе из входящих в МД_{*i*} отношений содержит в себе значения технических характеристик компонентов, необходимых для удовлетворения запроса потребителей, примеры таких характеристик для рассматриваемой в настоящей статье предметной области приведены в [4–6]. Пусть это отношение входит в множество (2) под номером 2 и содержит в себе $m_2^i + 1$ доменов, тогда это отношение можно обозначить

$$R_2^i = R_2^i (KL_2^i, A_{2,1}^i, A_{2,2}^i, \dots, A_{2,m_2^i}^i), \quad (5)$$

где $A_{2,j}^i, j = \overline{1, m_2^i}$, наименование j -го атрибута (домена) содержащего значение j -ой технической характеристики, а каждая строка этого отношения содержит значения всех технических характеристик одного экземпляра компонента; KL_2^i – ключевой домен этого отношения, который по смыслу аналогичен KL_1^i .

Третье из входящих в МД_{*i*} отношений содержит в себе значения характеристик условий поставки компонентов данным поставщиком и по аналогии с (4) и (5) может быть представлено

$$R_3^i = R_3^i (KL_3^i, A_{3,1}^i, A_{3,2}^i, \dots, A_{3,m_3^i}^i), \quad (6)$$

здесь переменная m_3^i и домены $KL_3^i, A_{3,j}^i$, $j = \overline{1, m_3^i}$ по смыслу аналогичны предыдущим, но применительно к условиям поставки компонентов i -ым поставщиком.

После введения обозначений можно установить связи между рассмотренными частями данных в форме отношения связи

$$C_1^i = C_1^i(KL_1^i, KL_2^i, KL_3^i), \quad (7)$$

в каждом кортеже (строке) которого содержится информация о том, какой экземпляр компонента с какими характеристиками удовлетворяет конкретным условиям поставки. Учитывая, что для нашей задачи содержание всех остальных отношений $R_4^i, \dots, R_{k_i}^i$, а также связи между другими частями данных $C_2^i, \dots, C_{t_i}^i$ для всех $i = \overline{1, n}$, не имеют существенного значения, соотношение (2) может быть представлено в следующей форме

$$MD_i = \{(R_1^i, R_2^i, R_3^i, C_1^i)\}, \quad (8)$$

для всех $i = \overline{1, n}$.

Задача формирования из совокупности $\{БДП_i\}$, $i = \overline{1, n}$, единой для всех потребителей и поставщиков БОД заключается в выделении в каждой БДП_{*i*} общей хотя бы для двух поставщиков информации и уникальной, характерной для i -го поставщика информации для последующего включения в БОД. Для того, чтобы любой из N пользователей ИСПК мог выбрать необходимые ему компоненты с заданными характеристиками в БОД в одном месте должна находиться взаимоувязанная информация о всех компонентах с их характеристиками и о поставщиках выбранных компонент, которая распределена (хранится по отдельности) во всем множестве $\{БДП_i\}$, $i = \overline{1, n}$. Для этого в МОД должны быть сформированы отношения с перечнем (описанием) всех компонентов и их возможных характеристик, а также с описанием поставщиков этих компонентов (размещением этой информации в соответствующих БДП). В зависимости от конкретной предметной области использования ИСПК количество и структура этих отношений может различаться, но можно предложить единый подход (метод) по их формированию и использованию.

Первое из входящих в МОД отношений содержит в себе общую информацию о каждом компоненте данной предметной области, которая позволяет однозначно идентифицировать искомую компоненту во всем их множестве, это может быть название компоненты, производитель, внутренний код производителя, наименование группы и подгруппы компонентов. Пусть это отношение входит в множество (3) под номером 1 и содержит в себе $m_1^0 + 1$ доменов, тогда его можно обозначить

$$R_1^0 = (KL_1^0, A_{1,1}^0, A_{1,2}^0, \dots, A_{1,m_1^0}^0), \quad (9)$$

по аналогии с (4) $A_{1,j}^0$, $j = \overline{1, m_1^0}$, наименование j -го атрибута (домена), содержащего значение j -ой идентификационной характеристики, а каждая строка (кортеж) этого отношения содержит значения всех идентификационных характеристик одного компонента в БОД; KL_1^0 – ключевой домен этого отношения, который однозначно определяет конкретный экземпляр компонента из всего их множества в БОД (для всех поставщиков).

Обозначим через A_1^i множество неключевых атрибутов первого отношения во всех рассматриваемых МД_{*i*}

$$A_1^i = \{A_{1,j}^i\}_{j=1, m_1^i}, i = \overline{0, n}, \quad (10)$$

тогда

$$A_1^0 = \bigcup_{i=1}^n A_1^i, \quad (11)$$

то есть, каждая идентификационная характеристика компонента, встречающаяся хотя бы у одного поставщика, обязательно входит в перечень идентификационных характеристик этого компонента, при его описании в БОД. Причем входит только один раз, то есть если одна и та же идентификационная характеристика присутствует в описании этого компонента у нескольких (или всех) поставщиков, то в описание этого компонента в БОД она входит только один раз.

Второе из входящих в МОД отношений содержит в себе значения технических характеристик компонентов, необходимых для удовлетворения запроса потребителей. Пусть это отношение входит в множество (3)

под номером 2 и содержит в себе $m_2^0 + 1$ доменов, тогда по аналогии с (5) его можно обозначить

$$R_2^0 = R_2^0(KL_2^0, A_{2,1}^0, A_{2,2}^0, \dots, A_{2,m_2^0}^0), \quad (12)$$

здесь $A_{2,j}^0$, $j = \overline{1, m_2^0}$ – наименование j -го атрибута (домена), содержащего значение j -ой технической характеристики, а каждая строка этого отношения содержит значения всех технических характеристик одного экземпляра компонента в БОД; KL_2^0 – ключевой домен этого отношения, который по смыслу аналогичен KL_1^0 .

Обозначим через A_2^i – множество неключевых атрибутов второго отношения во всех рассматриваемых МД _{i}

$$A_2^i = \{A_{2,j}^i\}_{j=1, m_2^i}, i = \overline{0, n}. \quad (13)$$

тогда

$$A_2^0 = \bigcup_{i=1}^n A_2^i, \quad (14)$$

то есть, каждая техническая характеристика компонента, встречающаяся хотя бы у одного поставщика, обязательно входит в перечень технических характеристик этого компонента при его описании в БОД. Причем, как и при описании идентификационных характеристик входит только один раз, то есть если одна и та же техническая характеристика присутствует в описании этого компонента в БДП _{i} у нескольких или всех ($2 \leq i \leq n$) поставщиков, то в описании этого компонента в БОД она входит только один раз.

Третье из входящих в МОД отношений содержит в себе информацию о всех поставщиках, включая описание и ссылку на БДП _{i} для всех $i = \overline{1, n}$ и по аналогии с (9) и (12) может быть представлено

$$R_3^0 = R_3^0(KL_3^0, A_{3,1}^0, A_{3,2}^0, \dots, A_{3,m_3^0}^0), \quad (15)$$

здесь $A_{3,j}^0$, $j = \overline{1, m_3^0}$ – содержит основные характеристики поставщиков (включая ссылку на БДП _{i}), необходимые для выбора и приобретения компонентов; KL_3^0 – ключевой домен, значения которого соответствуют номеру или другому идентификатору соответствующего поставщика, то есть в отношении

R_3^0 – ровно n кортежей и в самом простом случае KL_3^0 принимает одно из значений от 1 до n : $KL_3^0 = \overline{1, n}$.

После введенных обозначений в МОД можно установить связи между рассматриваемыми частями данных в форме отношения связи

$$C_1^0 = C_1^0(KL_1^0, KL_2^0, KL_3^0), \quad (16)$$

в каждом кортеже (строке) которого содержится информация о том, какой экземпляр компонента с какими конкретными характеристиками поставляется поставщиком с конкретным номером.

В силу введенных обозначений МОД может быть представлена следующим соотношением

$$\text{МОД} = \{R_1^0, R_2^0, R_3^0, C_1^0\}, \quad (17)$$

и, в свою очередь, с учетом (1)

$$\begin{aligned} \text{МРД} &= \\ &= \text{МОД} \bigcup_{i=1}^n \left(\bigcup_{i=1}^n \text{МД}_i \right) = \\ &= \{R_1^0, R_2^0, R_3^0, C_1^0\} \bigcup_{i=1}^n \{R_1^i, R_2^i, R_3^i, C_1^i\}. \end{aligned} \quad (18)$$

ОБОСНОВАНИЕ КОРРЕКТНОСТИ СИНТЕЗА СТРУКТУРЫ БАЗЫ ОБЩИХ ДАННЫХ

Теперь покажем справедливость, или точнее – реализуемость соотношения (8), то есть возможность описания групп характеристик (идентификационных, технических, условий поставки) соответственно в виде отношений R_1^i, R_2^i, R_3^i , наиболее трудоемкую часть которых составляют множества A_1^i, A_2^i, A_3^i из соотношений (10), (11), (13) и (14). Для этого еще раз обратим внимание на то, что все БДП _{i} создаются каждым поставщиком самостоятельно, независимо друг от друга, исходя из собственных производственных потребностей, поэтому в реальной действительности атрибуты, относящиеся к описанию той или иной части компонентов и их характеристик, могут размещаться в нескольких отношениях.

Несколько упрощенно покажем это на примере формирования множества A_2^i для одного любого поставщика, при этом, для простоты изложения, верхний индекс i в соотношении (14) пока использовать не будем, то есть полагаем

$$A_2 = \{A_{21}, A_{22}, \dots, A_{2m_2}\}, \quad (19)$$

а модель базы данных этого поставщика представим совокупностью отношений

$$\text{МД} = \{\bar{R}_1, \bar{R}_2, \dots, \bar{R}_k\}, \quad (20)$$

каждое j -ое отношение содержит в себе M_j атрибутов, $j = \overline{1, k}$

$$\bar{R}_j = \bar{R}_j(B_{j1}, B_{j2}, \dots, B_{jM_j}). \quad (21)$$

Предположим, что информация о всех технических характеристиках некоторого компонента у этого поставщика размещается в P отношениях $\bar{R}_{l_1}, \bar{R}_{l_2}, \dots, \bar{R}_{l_p}$, где l_1, l_2, \dots, l_p – номера этих отношений в множестве (20) и для них справедливо

$$1 \leq l_1 < l_2 < \dots < l_p \leq k, \quad (22)$$

обозначим это множество номеров

$$L = \{l_i\}_{i=\overline{1, p}}. \quad (23)$$

Учитывая (21), структура каждого из этих отношений выглядит следующим образом

$$\bar{R}_{l_i} = \bar{R}_{l_i}(B_{l_i1}, B_{l_i2}, \dots, B_{l_iM_{l_i}}), \quad i = \overline{1, p}, \quad (24)$$

причем не все атрибуты этих отношений содержат значения технических характеристик, а только некоторые из них. Пусть в каждом l_i -ом отношении содержится MO_{l_i} атрибутов ($MO_{l_i} \leq M_{l_i}$), с описанием технических характеристик

$$B_{l_i T_1^{l_i}}, B_{l_i T_2^{l_i}}, \dots, B_{l_i T_{MO_{l_i}}^{l_i}}, \quad (25)$$

где $T_1^{l_i}, T_2^{l_i}, \dots, T_{MO_{l_i}}^{l_i}$, номера этих атрибутов в множестве (24) и для них справедливо

$$1 \leq T_1^{l_i} < T_2^{l_i} < \dots < T_{MO_{l_i}}^{l_i} \leq M_{l_i}, \quad (26)$$

обозначим это множество номеров

$$T = \{T_q^{l_i}\}_{q=\overline{1, MO_{l_i}}, i=\overline{1, p}}. \quad (27)$$

После введенных обозначений информация о всех технических характеристиках описывается атрибутами из различных отношений, количество этих атрибутов в соответствии с (19) должно равняться m_2

$$m_2 = \sum_{i=1}^p MO_{l_i}, \quad (28)$$

а все множество атрибутов безотносительно их принадлежности к конкретному отношению можно определить следующим образом

$$B = \left\{ B_{l_i T_q^{l_i}} \right\}_{i=\overline{1, p}, q=\overline{1, MO_{l_i}}}. \quad (29)$$

Для реализации предложенного выше в соотношениях (2)–(18) метода формирования МОД, необходимо показать возможность однозначного формирования множества A_2 из атрибутов исходных баз данных, а для этого необходимо установить взаимно однозначные соответствия между элементами множества B из соотношения (29) и множества A_2 из соотношения (19).

Установим следующий порядок перемещения нужных нам атрибутов (в нашем примере тех, которые описывают технические характеристики) из соотношений \bar{R}_i в множество A_2^i : в начало (с первого значения индекса j) этого множества размещаются необходимые атрибуты (их MO_{l_1} штук) l_1 -го отношения; потом атрибуты (их MO_{l_2}) l_2 -го отношения и так далее; в конце множества (с последними номерами элементов) будут размещаться атрибуты (их MO_{l_p}) l_p -го отношения. В общем случае взаимосвязи номеров атрибутов из множества (19) и (29) вычисляются следующим образом. Пусть атрибут, содержащий информацию о некоторой технической характеристике, входит под номером F в некоторое отношение под номером G в МД (20), то есть

$$B_{G,F} \in \bar{R}_G, \quad G \in L, \quad F \in T, \quad (30)$$

тогда, в силу определения множества L , \exists число

$$MM: G = l_{MM} \in L, 1 \leq MM \leq P. \quad (31)$$

В соответствии с (22) отношению \bar{R}_G в (20) предшествует $(MM-1)$ отношений; в каждом из которых MO_{l_i} атрибутов, $i = \overline{1, (MM-1)}$.

Аналогично, в силу определения множества T , \exists число

$$NN: F = T_{NN}^G, 1 \leq NN \leq MO_{l_{MM}} = MO_G, \quad (32)$$

и атрибуту B_{G, T_{NN}^G} в соответствии с (26) в \bar{R}_G предшествует $(NN-1)$ атрибутов.

Тогда атрибут $B_{G, F} \in \bar{R}_G$ будет входить в A_2 под номером j_0 , который определяется следующим образом

$$j_0 = \begin{cases} NN, & \text{для } MM = 1 \\ \sum_{i=1}^{MM} MO_{l_{i-1}} + NN, & \text{для } MM \geq 2. \end{cases} \quad (33)$$

Для сокращения записи определения номера j_0 введем в рассмотрение фиктивный элемент $MO_{l_0} = 0$, который не имеет физического смысла и не участвует в описании модели данных, тогда

$$j_0 = \sum_{i=1}^{MM} MO_{l_{i-1}} + NN, \quad (34)$$

и атрибут $B_{G, F}$ будет входить в множество (19) под номером j , определяемым (34)

$$A_{2, j_0} = B_{G, F}. \quad (35)$$

Теперь покажем обратное соответствие: в каком атрибуте какого отношения из (20) будет размещаться атрибут $A_{2, j_0} \in A_2$. В силу выбранного порядка размещения атрибутов в множестве A_2 \exists некоторое количество отношений с номерами $l_i \in L$, обозначим это количество через MM , все атрибуты которых с номерами $T_q^{l_i}$, $q = \overline{1, MO_{l_i}}$, $i = \overline{1, MM}$, размещаются перед A_{2, j_0} , то есть

$$\sum_{i=1}^{MM} MO_{l_i} \leq j_0 \leq \sum_{i=1}^{MM+1} MO_{l_i}, \quad (36)$$

обозначим $G = l_{MM+1}$ тогда атрибут A_{2, j_0} размещается в отношении \bar{R}_G , а число

$$NN = j_0 - \sum_{i=1}^{MM} MO_{l_i}, \quad (37)$$

задает номер атрибута T_{NN}^G в этом отношении, следовательно

$$B_{G, T_{NN}^G} = A_{2, j_0}. \quad (38)$$

Так как исходные определения (19)–(21) вводились для произвольного (любого) поставщика, то соотношения (35) и (38) также справедливы для любого поставщика и правомерность использования в соотношениях (4)–(18) отношений R_1^i, R_2^i, R_3^i , множеств A_1^i, A_2^i, A_3^i , и в целом метод формирования МОД из МДП_i вполне корректны.

ЗАКЛЮЧЕНИЕ

Предложен метод формирования структуры распределенной базы данных с описанием компонентов сложных технических систем и их характеристик, необходимый для создания информационной системы подбора компонентов на различных этапах создания и эксплуатации сложных технических систем. Метод основан на использовании и развитии теоретико-множественного подхода к формальному анализу и описанию баз данных желаемых и фактических характеристик компонентов и последующему синтезу модели общих данных, позволяющих осуществить оперативный доступ большого числа пользователей ко всей информации. Предложена процедура формирования модели общих данных, позволяющая осуществлять взаимно однозначное соответствие атрибутов, описывающих характеристики компонентов в различных частях распределенной базы данных: баз данных поставщиков и базы общих для всех потребителей и поставщиков данных, а также обоснована корректность этой процедуры и метода в целом.

БЛАГОДАРНОСТИ

Результаты исследований, представленные в статье, поддержаны грантом РФФИ №20-08-00301\20-А «Методологические и теоретические основы управления уязвимостью на основе парирования угроз в сложных распределенных системах».

СПИСОК ЛИТЕРАТУРЫ

1. **Создание** инфраструктуры пространственных данных Республики Башкортостан на основе геоинформационных технологий / Р. Н. Бахтизин [и др.]. Уфа: Нефтегазовое дело, 2008. 103 с. [R. N. Bakhtizin, *Creation of spatial data infrastruc-*

ture of the Republic of Bashkortostan on the basis of geoinformation technologies, (in Russian). Ufa: Neftgazovoe delo, 2008.

2. **Брекоткина Е. С., Павлов С. В., Павлов А. С.** Метод организации больших массивов данных региона в рамках перехода к цифровой экономике // Материалы XVIII Международной научной конференции «Управление экономикой: методы, модели, технологии». Уфа: УГАТУ, 2018. С. 34–37. [E. S. Brekotkina, S. V. Pavlov, A. S. Pavlov, "Method of organizing large data arrays of the region in the framework of the transition to the digital economy", (in Russian), in Proceedings of the XVIII International Scientific Conference "Economic Management: methods, models, technologies", pp. 34-37, 2018.]

3. **Анализ** и синтез структуры пространственных данных для управления сложными распределенными системами / Е. С. Брекоткина [и др.] // Информационные технологии и системы: труды Восьмой Всероссийской научной конференции с международным участием. Ханты-Мансийск: Югорский НИИ информационных технологий, 2020. С. 120–125. [E. S. Brekotkina, et al., "Analysis and synthesis of spatial data structure for managing complex distributed systems", (in Russian), in Proc. Information Technologies and Systems: proceedings of the Eighth All-Russian Scientific Conference with international participation, pp. 120-125, 2020.]

4. **Информационная** поддержка управления уязвимостью сложных распределенных систем на основе обработки пространственной информации / Е. С. Брекоткина [и др.] // Моделирование, оптимизация и информационные технологии. 2020. Т. 8, № 2 (29). DOI: 10.26102/2310-6018/2020.29.2.023. [E. S. Brekotkina, et al., "Information support for vulnerability management of complex distributed systems based on spatial information processing", (in Russian), in Modelirovanie, optimizaciya i informacionny'e tehnologii, vol. 8, no 2 (29), 2020. DOI: 10.26102/2310-6018/2020.29.2.023.]

5. **Инструкция** по использованию Marvel-B2B API версия 3.10. [Электронный ресурс]. URL: <https://b2b.marvel.ru/> (дата обращения 22.04.2022). [Instructions for using Marvel-B2B API version 3.10 (2022, Apr. 22). [Online]. Available: <https://b2b.marvel.ru/22/04/2022>]

6. **API-коннектор** B2B документация к сервису. [Электронный ресурс]. URL: <https://b2b.ocs.ru/> (дата обращения 22.04.2022). [API connector B2B service documentation (2022, Apr. 22). [Online]. Available: <https://b2b.ocs.ru/22/04/2022>]

7. **Open** Catalog Interface (OCI) a set of standards and file formats used in the Iccat XML repository. It gives the basics of how to use the Iccat repository or how to publish data to Iccat conforming the OCI. [Electronic resource]. URL: <https://www.iccat.biz/>. (Accessed 22.04.2022).

8. **Бэгули Ф.** Управление проектом. М.: ФАИР-ПРЕСС, 2002. 208 с. [F. Baguli, *Project management*, (in Russian). Moscow: FAIR PRESS, 2002.]

9. **Гвоздев В. Е., Колоденкова А. Е.** Системные вопросы проектирования программных продуктов: учебное пособие. Уфа: Гилем, 2010. 188 с. [V. E. Gvozdev, A. E. Kolodenkova, *System issues of software product design: Textbook*, (in Russian). Ufa: Gilem, 2010.]

10. **Багутдинов Р. А., Степанов М. Ф.** Методы интеграции, уменьшение размеров и нормализация обработки разнородных и разномасштабных данных // International Journal of Open Information Technologies. 2021. Т. 9, № 2. С. 39–44. [R. A. Bagutdinov, M. F. Stepanov, "Methods of inte-

gration, size reduction and normalization of processing of heterogeneous and multi-scale data", (in Russian), in *International Journal of Open Information Technologies*, vol. 9, no. 2, pp. 39-44, 2021.]

11. **Беркун С.** Искусство управления IT-проектами. СПб.: Питер, 2007. 400 с. [S. Berkun, *The Art of IT project management*, (in Russian). St. Petersburg: Peter, 2007.]

ОБ АВТОРАХ

БРЕКОТКИН Валерий Евгеньевич, ген. директор ООО «Атриум».

БРЕКОТКИНА Елена Сергеевна, доц. каф. экономики предпринимательства УГАТУ.

ПАВЛОВ Александр Сергеевич, гл. спец. группы развития систем риск-менеджмента в ПАО Банк «ФК Открытие».

ПАВЛОВ Сергей Владимирович, проф. каф. геоинформационных технологий УГАТУ.

METADATA

Title: Information technology for selecting components of complex technical systems based on intelligent analysis of their desired and actual characteristics from distributed databases.

Authors: V. E. Brekotkin ¹, E. S. Brekotkina ², A. S. Pavlov ³, S. V. Pavlov ⁴

Affiliation: ¹ Atrium LLC, Russia.

^{2,4} Ufa State Aviation Technical University (UGATU), Russia.

³ PJSC Bank "FC Otkritie", Russia.

Email: ¹ a3um@ a3um.ru, ² brekotkina@mail.ru, ³ asp.gis@gmail.com, ⁴ psvgis@mail.ru.ru

Language: Russian.

Source: SIIT (scientific journal of Ufa State Aviation Technical University), vol. 4, no. 2 (9), pp. 13-23, 2022. ISSN 2686-7044 (Online), ISSN 2658-5014 (Print).

Abstract: One of the important ways to increase the efficiency of creating and using complex technical systems is to determine whether the desired characteristics of these systems and their components correspond to their characteristics in the present and future. The complexity of this problem is determined by the complexity of the components, the large number of characteristics of these components and systems as a whole, the vagueness of the requirements for the list and values of characteristics from various participants in the creation and use of complex systems. At the same time, a duality of the situation arises: on the one hand, large volumes of heterogeneous data are used, the collection and processing of which requires a lot of time and human resources, on the other hand, there are significant restrictions on the time of decision-making based on this data. It offers a solution to the problem of selecting components of complex systems at various stages of their creation (design, development, implementation) and operation, based on an intelligent analysis of the entire set of possible characteristics of these components stored in a variety of heterogeneous (in structure, departmental affiliation, means of implementation) databases. A method of forming the structure of a distributed database with a description of the components of complex technical systems and their characteristics and a procedure for constructing a general data model that allows for a one-to-one correspondence

of attributes describing the characteristics of components in various parts of a distributed database: supplier databases and databases common to all consumers and suppliers of data. The methods are intended for their subsequent implementation within the framework of an information system that provides information support for the activities of various categories of specialists (designers, technologists, suppliers, etc.) who select the necessary components with the specified (desired) characteristics

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 relationships.

About authors:

BREKOTKIN, Valery Evgenievich, General Director of "Atrium" LLC.

BREKOTKINA, Elena Sergeevna, Assoc. Prof., Dept. of economics of entrepreneurship of Ufa State Aviation Technical University.

PAVLOV, Aleksandr Sergeevich, Chief Specialist of the Risk Management Systems Development Group of PJSC Bank "FC Otkritie".

PAVLOV, Sergey Vladimirovich, Prof., Dept. of geoinformation technologies of Ufa State Aviation Technical University.