

DOI:

МОДЕЛИ И МЕТОДЫ СИНТЕЗА РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННО-УПРАВЛЯЮЩЕЙ СТРУКТУРЫ ПАТЕНТНЫХ ИНФОРМАЦИОННЫХ ФОНДОВ

Кульба В.В., Сиротюк В.О.

Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва

ул. Профсоюзная д.65

kulba@ipu.ru, vsirotyuk@ipu.ru

Аннотация: Рассмотрены особенности формирования, сопровождения и развития патентного информационного фонда (ПИФ), играющего важную роль при проведении хозяйствующими субъектами патентных исследований. Предложены формальные методы построения объектной модели распределенной информационно-управляющей структуры (РИУС) ПИФ, предоставляющей доступ к локальным и внешним удаленным патентным базам данных, поиска информации и ее использования.

Ключевые слова: патентный информационный фонд (ПИФ); предметная область ПИФ; патентная база данных; распределенная информационно-управляющая структура (РИУС) ПИФ; объектная модель требований пользователей ПИФ; обобщенная модель РИУС ПИФ; нормализация обобщенной модели РИУС ПИФ.

Введение

Патентные информационные фонды (ПИФ), информационные технологии сбора, хранения, обработки и передачи данных, информационно-телекоммуникационные и сетевые инфраструктуры образуют информационную среду (патентно-информационное пространство), которая формируется для эффективного выполнения патентными ведомствами и хозяйствующими субъектами поставленных перед ними задач [1].

Формирование и сопровождение ПИФ осуществляется патентными ведомствами в ходе рассмотрения заявок на изобретения, выдачи патентов, публикации патентной документации и предоставления к ней доступа общественности. При этом одной из важнейших функций патентных ведомств является проведение патентной экспертизы, в ходе которой заявка на изобретение оценивается на новизну, изобретательский уровень и промышленную применимость. Очевидно, что эффективность и качество экспертизы определяются, в первую очередь, полнотой и достоверностью данных ПИФ, наличием развитых инструментальных средств доступа экспертов к локальным и внешним удаленным источникам патентной и связанной с ней непатентной информации.

Высокие требования к качеству ПИФ предъявляются также его внешними пользователями - хозяйствующими субъектами (патентными поверенными, заявителями и изобретателями, предприятиями и организациями, частными лицами), проводящими патентные исследования технического уровня и тенденций развития объектов хозяйственной деятельности, их патентоспособности, патентной чистоты, конкурентоспособности на основе патентной информации. Такие исследования являются неотъемлемой составной частью обоснования принимаемых хозяйствующими субъектами решений, связанных с созданием, производством, реализацией, совершенствованием и использованием объектов хозяйственной деятельности.

Центральное место в ПИФ занимают патентные базы данных (ПБД), на которые возложены функции хранения, интеграции и консолидации патентно-информационных ресурсов. Создание ПБД является мощным средством активизации национальных информационных ресурсов за счет перевода значительной их части на электронные носители и последующего многоаспектного индивидуального и коллективного использования. Среди ПБД ПИФ различают локальные ПБД, создаваемые владельцем ПИФ (национальным или региональным патентным ведомством), и внешние ПБД, формируемые другими патентными ведомствами и организациями (службами) – провайдерами патентной информации и непатентной литературы, к которым организуется доступ с целью повышения полноты проводимого патентного поиска [2,3].

Ввиду этого инфраструктура современного ПИФ должна иметь распределенную информационно-управляющую структуру (РИУС), предоставляющую доступ к локальным и внешним ПБД, поиск информации и ее использование. При этом обращение к внешним ПБД должно осуществляться на основе обоснованной стратегии патентного поиска через единый пользовательский интерфейс, предоставляющий средства доступа к данным, поиска и обработки данных, а также сервисные средства, включая машинный перевод. Это обуславливает необходимость логической интеграции данных локальных и внешних ПБД.

В работе предложены формализованные модели и методы анализа информационных и функциональных требований пользователей ПИФ и формирования объектно-ориентированных

моделей спецификаций требований пользователей, проектирования обобщенной объектной модели РИУС ПИФ и ее нормализации, построения оптимальной объектной модели РИУС ПИФ. Исходными данными для проектирования являются формализованные описания и характеристики предметной области ПИФ, топологии и инфраструктуры ПИФ. Предложенные модели, методы и алгоритмы использовались при разработке структуры ПИФ и ПБД евразийского патентного информационного пространства [1,4].

Методы построения моделей спецификаций требований пользователей

Рассмотрим процедуры построения моделей спецификаций информационных и функциональных требований пользователей ПБД.

Введем ряд необходимых определений.

Под объектом данных ПИФ будем понимать некоторую совокупность информационных элементов, описывающих патентную и непатентную документацию, методов (процедур) их обработки и поиска, а также отношений между ними, составляющих единое целое с точки зрения семантических и процедурных аспектов предметной области. Объект данных обладает четко выраженным поведением (набором процедур) и набором свойств (информационных элементов). Поведение объекта характеризуется методами, которые объект предоставляет другим объектам (интерфейс объекта) или использует для реализации своего поведения (реализационная часть). В качестве объектов данных патентной документации выступают такие элементы, как сведения о заявителях и патентообладателях, информация из полных описаний изобретений к заявкам и к патентам, формула, реферат изобретения и их части и т.д. Для непатентной документации объектами данных являются сведения об авторах публикации, информация о публикации (название статьи, монографии; номер издания; место публикации и имя издателя; год публикации и др.); реферат статьи, теги текста монографии (статьи, нормативно-правового документа) и др. элементы.

Класс объектов – один или несколько объектов, объединенных в один кластер по критерию максимального интерфейса между ними (информационной и технологической связности объектов).

Объектная модель требований пользователей - информационно-функциональная структура, формируемая в результате выполнения операций наложения на информационные структуры пользователей технологических составляющих, описывающих процедуры поиска и обработки данных.

Под *предметной областью ПИФ* понимается информация о совокупности пользователей, объектов хранения и процедурах обработки данных ПИФ, их характеристиках и отношениях между ними, которая представляется в виде специальным образом организованных структурах данных, хранится в ПБД и используется при проведении патентных поисков и патентных исследований.

Описание предметной области включает следующие основные компоненты:

- пользователи ПБД, к которым относятся эксперты патентных ведомств, изобретатели, заявители, патентообладатели, патентные поверенные и третьи лица, а также сотрудники службы администратора ПБД, отвечающие за создание, функционирование и сопровождение ПБД;
- процедуры поиска и обработки данных, последовательность их выполнения и характеристики;
- объекты данных, информационные элементы и их характеристики;
- отношения между объектами данных, информационными элементами и процедурами поиска и обработки данных.

Данная информация формируется на этапе изучения источников, свойств и характеристик патентной информации, а также информационных и функциональных требований пользователей и фиксируется в стандартных формах материалов обследования [5,6].

Формализовано предметная область ПИФ описывается с помощью множеств $\{H, U, O, E\}$ и булевых матриц смежности:

$$HU = \|hu_{jk}\|, HH = \|hh_{jj}\|, HO = \|ho_{je}\|, HE = \|he_{jp}\|, OE = \|oe_{ep}\|,$$

характеризующих наличие отношений (взаимосвязей) между компонентами предметной области. Здесь:

$H = \{h_j / j = \overline{1, J}\}$ - множество процедур, состоящее из подмножества процедур поиска данных $H_n \subseteq H$ и подмножества процедур обработки данных $H_{обп} \subseteq H$;

$U = \{u_k / k = \overline{1, K_0}\}$ - множество пользователей;

$O = \{o_\varepsilon / \varepsilon = \overline{1, \varepsilon_0}\}$ - множество объектов данных ПИФ;

$E = \{e_p / p = \overline{1, P}\}$ - множество информационных элементов ПИФ, к которым относятся библиографические данные патентных документов и другие элементы, описывающие объекты (части) патентных документов;

$R = \{r_y / y = \overline{1, Y}\}$ - множество отношений (взаимосвязей) между компонентами $\{H, U, O, E\}$.

Выделяются следующие виды отношений:

$r_1(H, U)$ - отношения «процедуры – пользователи». Каждый кортеж отношения r_1 определяет использование процедур поиска и/или обработки данных пользователями;

$r_2(H_n, H_n)$ - отношение «процедура поиска – процедура поиска». Каждый кортеж отношения r_2 определяет последовательность выполнения процедур поиска, которая формирует структуру запроса пользователя;

$r_3(H, O)$ - отношение «процедуры – объекты данных». Каждый кортеж отношения r_3 определяет перечень объектов ПИФ, используемых процедурами поиска и/или обработки данных;

$r_4(H, E)$ - отношение «процедуры – данные». Каждый кортеж отношения r_4 определяет использование информационных элементов при выполнении определенных поисковых процедур и/или процедур обработки данных;

$r_5(O, E)$ - отношение «объекты – данные». Каждый кортеж отношения r_5 характеризует информационное содержание (описание) определенного объекта данных ПИФ.

Формализованное описание предметной области ПИФ используется в дальнейшем при формировании моделей спецификаций информационных требований и требований обработки данных пользователей (функциональных требований). Рассмотрим методы их построения.

Модель спецификации информационного требования k -го пользователя представим в виде $M_{снец}^k = \langle \alpha R \beta \rangle$, где α и β - структурные элементы предметной области (объекты данных и информационные элементы), R - отношение между элементами. Для патентной информации, имеющей строго формализованный, стандартизированный и унифицированный вид, характерно наличие в основном бинарных отношений, например, отношений типа «номер заявки» - «ЗАЯВКА», «имя заявителя» - «ЗАЯВКА», «ФИО патентообладателя» - «ПАТЕНТ» и т.д. [1,2].

Структурными элементами моделей спецификаций информационных требований пользователей являются элементы множеств $O = \{o_\varepsilon / \varepsilon = \overline{1, \varepsilon_0}\}$ и $E = \{e_p / p = \overline{1, P}\}$. Обозначим полное множество структурных элементов как $D = \{d_l / l = \overline{1, L}\}$, $L = L_{ob} \cup L_{el}$, где L_{ob} - мощность множества индексов объектов данных, L_{el} - мощность множества индексов информационных элементов.

Алгоритм формирования моделей спецификаций информационных требований пользователей состоит из следующих шагов:

1. На основании анализа матриц HU и HH определяется перечень процедур поиска и процедур обработки данных по каждому пользователю и последовательность их выполнения.
2. На основании информации, полученной в п. 1., с использованием матрицы HO формируются пары структурных элементов $\langle OR_{об} O \rangle$, где $R_{об}$ - отношения между объектами патентного фонда.
3. На основании анализа матрицы OE формируются пары $\langle OR_a E \rangle$, где R_a - отношение принадлежности (входимости) информационных элементов объектам данных.
4. Проводится совместный анализ матриц HE , HH и OE , в результате которого устанавливается противоречивость и несогласованность описаний объектов.
5. На основании полученных в пп.1÷4 результатов по каждому требованию пользователя формируются бинарные модели спецификаций $M_{снец}^k = \langle \alpha R \beta \rangle$, представляемые в виде списка парных отношений между структурными элементами $d_l \in D$:

$S_k = \{(d_l R d_{l'})\}$, где $d_l, d_{l'} \in D_k, D_k \subseteq D$.

Формализовано модель спецификации информационного требования k -го пользователя на основании списка S_k представляется в виде матрицы смежности структурных элементов предметной области $B_k = \|b_{ll'}^k\|$ и графа $G_k^{in}(D_k, R_k)$. Элемент матрицы B_k равен 1, если $(d_l R d_{l'})$ и равен 0 в противном случае.

Модель спецификации функциональных требований k -го пользователя (модель спецификации инцидентов) формализовано представим в виде предиката $M_{np}^k = h^k(d_l | d_l \in D_k \subseteq D)$, множество элементов d_l которого включает структурные элементы предметной области k -го пользователя ПИФ (объекты данных и информационные элементы). Модель спецификации инцидентов формируется на основании матриц НУ, НЕ и НО.

Алгоритм формирования модели спецификации функционального требования k -го пользователя включает следующие шаги.

1. На основании матрицы НУ определяется множество процедур обработки данных, входящих в состав k -го требования пользователя $H_k = \{h_j^k / j \in J_k \subseteq J\}$.

2. С использованием матриц НЕ и НО для каждого функционального требования пользователя формируется матрица технологии его обработки $W_k = \|w_{jl}^k\|$, $j = \overline{1, J_k}$, $l = \overline{1, L_k}$, индексируемая по строкам множеством процедур обработки данных, а по столбцам - структурными элементами. Элемент w_{jl}^k матрицы W_k принимает следующие значения: $w_{jl}^k = 1$, если элемент d_l используется процедурой h_j^k и $w_{jl}^k = 0$, в противном случае.

3. Проводится совместный анализ матриц НЕ, НО, НУ и W_k на предмет принадлежности процедур обработки данных пользователям, согласованности использования ими объектов данных и информационных элементов, а также исключения не востребуемых информационных элементов и выявления конфликтов по данным (deadlock).

4. На основании полученных в пп.1÷3 результатов по каждому требованию пользователя формируются модели спецификаций функциональных требований $M_{np}^k = h^k(d_l | d_l \in D_k \subseteq D)$, представляемые строками матрицы $W_k = \|w_{jl}^k\|$.

Матрице W_k ставится в соответствие граф технологической структуры k -го требования пользователя G_k^{pr} с двумя типами вершин, соответствующих структурным элементам множества D_k и процедурам обработки данных H_k , и одним типом дуг (d_l, h_j) , соответствующих записям $w_{jl}^k = 1$ в матрице W_k . Таким образом, дуги графа G_k^{pr} отражают технологию обработки структурных элементов.

Разработанные процедуры и методы формирования и анализа моделей спецификаций информационных и функциональных требований пользователей последовательно применяются для каждого требования пользователей, в результате выполнения которых формируются множества структурированных матриц $\{B_k\}$, $\{W_k\}$ и соответствующих им графов $\{G_k^{in}\}$, $\{G_k^{pr}\}$.

Далее осуществляется построение объектных моделей требований пользователей ПБД.

Формирование объектных моделей требований пользователей

Построение объектных моделей требований пользователей осуществляется путем последовательного наложения на модели спецификаций информационных требований пользователей моделей спецификаций функциональных требований.

Исходными данными для реализации этого этапа являются:

- формализованные описания информационных требований пользователей в виде матриц смежности $\{B_k\}$ и графов $\{G_k^{in}\}$;
- формализованные описания функциональных требований пользователей в виде матриц технологии $\{W_k\}$, графов $\{G_k^{pr}\}$ и структур запросов на поиск данных.

Структура отдельного запроса представляется в виде дерева поиска данных на графе информационной структуры G_k^{in} . В качестве корневой вершины дерева поиска запроса указывается объект, начиная с которого осуществляется формирование выходных структур. К корневой вершине

должны быть заданы альтернативные варианты доступа, начинающиеся в одной из точек входа графа G_k^{in} . Варианты доступа к корневой вершине связаны с прохождением некоторого множества связей и объектов информационной структуры пользователя, которые определяют дерево поиска. Каждая ветвь дерева поиска задает путь (или фрагмент пути) к объектам, из которых выбираются элементы, необходимые при формировании выходных структур. Задание условий поиска определяет направление и стратегию обхода вершин дерева поиска. При этом стратегия обхода в значительной степени определяется рекурсивностью деревьев. Алгоритм построения дерева поиска запроса приведен в [6].

Отображение требований обработки данных и запросов пользователей на графах информационных структур пользователей G_k^{in} производится следующим образом. Использование некоторой j -ой процедурой l -го элемента формально представляется на графе G_k^{in} петлей на объекте O_m , в который он входит ($d_l \in O_m$), что свидетельствует об обработке данного объекта (целиком или частично по отдельным входящим в него элементам).

Деревья поиска, требуемые для обработки данных, отображаются дополнительными дугами на графе G_k^{in} .

С учетом введенных элементов объектная модель требований k -го пользователя формализовано представляется в виде мультиграфа $G_k^{ob}(D_k, U_k)$ с одним типом вершин и двумя типами дуг, где $D_k = \{d_l^k / l = \overline{1, L_k}, L_k \subseteq L\}$ - множество структурных элементов k -го пользователя (объектов данных и информационных элементов), $U_k = U_k^{in} \cup U_k^{pr}$ - множество дуг между элементами, где U_k^{in} - множество дуг, описывающих информационные взаимосвязи между элементами, а U_k^{pr} - множество дуг, характеризующих технологию обработки данных в виде совокупности процедур поиска и непосредственной обработки данных, включая петли и дуги.

Основными характеристиками объектной модели требований G_k^{ob} являются:

1. Вектор технологических весов вершин $Z_\mu^k = \{z_{lk}^\mu\}$, где z_{lk}^μ - технологический вес d_l -й вершины графа G_k^{ob} .

2. Вектор технологических толщин дуг $Z_\eta^k = \{z_{ll',k}^\eta\}$, где $z_{ll',k}^\eta$ - технологическая толщина дуги $(d_l, d_{l'})$ на графе G_k^{ob} .

Технологический вес $z_{lk}^\mu = \text{const}$, где $\text{const} \in \{0, 1, 2, \dots, N\}$ означает степень использования информационного элемента d_l множеством процедур. Чем больше значение z_{lk}^μ для некоторого d_l , тем более важным является элемент d_l в процессах обработки данных (выборки, считывания, арифметико-логических действиях и т.п.).

Технологическая толщина дуги $(d_l, d_{l'})$ $z_{ll',k}^\eta = \text{const}$, где $\text{const} \in \{0, 1, 2, \dots, N\}$ означает степень использования этой дуги в процессах поиска требуемых данных. Чем больше значение $z_{ll',k}^\eta$ для некоторой дуги $(d_l, d_{l'})$, тем более часто используется данная взаимосвязь в путях доступа к требуемым для обработки данным.

Определение технологических весов вершин осуществляется путем суммирования значений матрицы $W_k = \|w_{jl}^k\|$ по столбцам (структурным элементам):

$$z_{lk}^\mu = \sum_{\forall j \in J_k} w_{jl}^k.$$

Определение технологических толщин дуг осуществляется путем суммирования дуг, входящих в множество U_k^{pr} :

$$z_{ll',k}^\eta = \sum_{l,l' \in U_k^{pr}} (d_l, d_{l'})$$

Вычислив основные характеристики графа G_k^{ob} , можно перейти к другому формализованному представлению объектной модели требований пользователя - взвешенному графу $G_k^{636}(D_k^{636}, U_k^{636})$, где каждой вершине $d_l \in D_k^{636}$ и дуге $(d_l, d_{l'}) \in U_k^{636}$ приписаны соответствующие им веса.

Взвешенный граф $G_k^{636}(D_k^{636}, U_k^{636})$ характеризует количественные аспекты поведения объектов данных объектно-ориентированных моделей информационных и функциональных требований пользователей: интерфейсной части объектов (показатель $z_{ll',k}^{\eta}$) и реализационной части (показатель z_{lk}^{μ}). Последующий анализ графа $G_k^{636}(D_k^{636}, U_k^{636})$ позволяет выделить классы объектов, обладающих одинаковым поведением и набором информационных элементов [6,7].

В дальнейшем в соответствии с предлагаемой методологией осуществляется построение обобщенной объектной модели РИУС ПИФ и ее нормализация. Методы и алгоритмы ее построения приводятся в следующем разделе.

Модели и методы синтеза и нормализации РИУС ПИФ

Под обобщенной объектной моделью РИУС ПИФ будем понимать интегрированное множество объектных моделей требований пользователей и формализованных характеристик предметной области ПИФ, проектируемое с учетом топологии и характеристик вычислительной среды ПИФ, характеристик внешних источников патентной информации и непатентной литературы, средств доступа к ним и поиска информации.

Построение обобщенной объектно-ориентированной (объектной) модели РИУС ПИФ осуществляется путем анализа объектных моделей требований пользователей и синтеза интегрированной структуры данных путем последовательного объединения мультиграфов объектных моделей требований пользователей $G_k^{ob}(D_k, U_k)$, анализа и нормализации сформированной обобщенной объектно-ориентированной структуры данных.

Построенная в результате разработанных методов и процедур обобщенная объектная модель РИУС ПИФ описывается множествами объектов данных, информационных элементов, процедур поиска и обработки патентной информации и отношений между ними. Характеристиками обобщенной модели РИУС ПИФ являются свойства и параметры объектов данных, информационных элементов, взаимосвязей и отношений между ними, поисковых запросов пользователей и транзакций обработки патентной информации, параметры топологии информационно-вычислительной среды функционирования ПИФ. В дальнейшем данная информация используется при проектировании оптимальных структур ПБД и базы метаданных репозитория ПИФ, определения стратегии взаимодополняющего доступа к локальным и внешним удаленным ПБД и поиска в них.

Разработанная процедура построения обобщенной объектной модели РИУС ПИФ основана на совмещении идентичных структурных элементов объектных моделей требований пользователей ПИФ независимо от уровня их размещения на графах $G_k^{ob}(D_k, U_k)$.

Исходными данными для ее построения являются объектные модели требований пользователей, формализовано представляемые в виде мультиграфов $G_k^{ob}(D_k, U_k)$ и матриц смежности $W_k = \|w_{ij}^k\|$, и их характеристики. Характеристиками информационных элементов объектов данных мультиграфов $G_k^{ob}(D_k, U_k)$ являются длина (в байтах) $g = \{g_{lk}\}$, тип (целый, вещественный, символьный) $\alpha_k = \{\alpha_{lk}\}$, количество экземпляров данных $\pi_k = \{\pi_{lk}\}$ (в виде среднего или максимального значения). Среди информационных элементов для каждого объекта данных с использованием методов и процедур, предложенных в [5,6], выделяются первичные ключи, образующие множество ключей $D_k^{key} = \{d_{ki}^{key}\}, i \in L_{key}^0$, и атрибуты данных, образующие множество атрибутов $D_k^{atr} = \{d_{ki}^{atr}\}, i \in L_{atr}^0$.

Оперативность получения i -го информационного элемента по запросу k -го пользователя определяется как время, в течение которого пользователю должны быть представлены требуемые результаты. Эта характеристика задается в виде вектора $\tau_k^3 = \{\tau_{ik}^3\}, k = 1, K_0$.

Тип информационного отношения (1:1, 1:M, M:1, M:M) между объектами данных d_{ki} и d_{kj} на графах $G_k^{ob}(D_k, U_k)$ задается матрицей $B_k = \|b_{kij}\|$, элементы которой принимают значения: $b_{kij} = 11$,

если $w_{kij}=1$ и тип отношения между объектами «один к одному»; $b_{kij}=1M$, если $w_{kij}=1$ и тип отношения между объектами «один ко многим»; $b_{kij}=M1$, если $w_{kij}=1$ и тип отношения между объектами «многие к одному»; $b_{kij}=MM$, если $w_{kij}=1$ и тип отношения между объектами «многие ко многим»; $b_{kij}=0$, если $w_{kij}=0$.

Важной компонентой объектных моделей требований пользователей является описание методов (процедур) обработки данных, задаваемое в виде двух множеств: множества запросов пользователей $Z_k = \{z_{pk}\}, p = \overline{1, P_k^0}, k = \overline{1, K_0}$ и множества транзакций (требований по пополнению, обновлению и корректировке информации ПИФ), инициируемых службой администратора ПБД ПИФ $K_k = \{k_{sk}\}, s = \overline{1, S_k^0}, k = \overline{1, K_0}$. Основной характеристикой методов обработки данных является частота (интенсивность) их выполнения на заданном интервале времени, которая задается в виде вектора частоты выполнения запросов $Q_k^3 = \{q_{pk}\}, p = \overline{1, P_k^0}$ и вектора частоты выполнения транзакций $Q_k^k = \{q_{sk}\}, s = \overline{1, S_k^0}$.

Вид использования информации задается в объектной модели в виде векторов, определяемых на множествах пользователей и информационных элементов: вектора $\zeta_k^3 = \{\zeta_{ki}^3\}$, где ζ_{ki}^3 принимает единичное значение, если пользователь u_k использует i -й информационный элемент в запросах (в противном случае $\zeta_{ki}^3=0$), и вектора $\zeta_k^k = \{\zeta_{ki}^k\}, i = \overline{1, L_k^0}$, где ζ_{ki}^k принимает единичное значение, если k -й пользователь использует i -й информационный элемент в корректировках, в противном случае $\zeta_{ki}^k=0$.

С учетом введенных характеристик формируется расширенная объектная модель требований k -го пользователя, которая представляется в виде графа $G_k^{ob}(D_k, U_k)$ и формально описывается с помощью матрицы смежности $W_k = \|w_{ij}^k\|, i, j = \overline{1, L_k^0}$, матрицы типов отношений между объектами данных $B_k = \|b_{kij}\|$, множеств ключей и атрибутов объектов данных $D_k^{key} = \{d_{ki}^{key}\}, i \in L_{key}^0$, $D_k^{atr} = \{d_{ki}^{atr}\}, i \in L_{atr}^0$, $D_k^{key} \cup D_k^{atr} = D_k$; множеств запросов пользователей ПИФ $Z_k = \{z_{pk}\}$ и транзакций $K_k = \{k_{sk}\}$; векторов, описывающих характеристики информационных элементов: $g_k = \{g_{ik}\}, \alpha_k = \{\alpha_{ik}\}, \pi_k = \{\pi_{ik}\}, \tau_k^3 = \{\tau_{ik}^3\}$; векторов характеристик методов обработки данных ПИФ: $Q_k^3 = \{q_{pk}^k\}, Q_k^k = \{q_{sk}^k\}$; векторов технологии поиска данных $\zeta_k^3 = \{\zeta_{ki}^3\}$ и обработки данных $\zeta_k^k = \{\zeta_{ki}^k\}$.

Построение обобщенной объектной модели РИУС ПИФ на основе расширенных объектных моделей требований пользователей осуществляется в несколько этапов.

На первом этапе определяется полное множество структурных элементов (объектов данных и информационных элементов) ПИФ $D = \{d_i\}, i = \overline{1, n_0}$. Оно образуется путем объединения элементов множеств D_k , в результате которого одинаковые структурные элементы, принадлежащие различным множествам, заменяется одним, т.е. формируется безызбыточное множество структурных элементов.

На втором этапе для элементов множества D задаются интегральные характеристики образующих его элементов. Вектор длин $g^0 = \{g_i^0\}, i = \overline{1, n_0}$ получается путем объединения векторов $g^0 = U_k g_k$. Значение длины информационного элемента в обобщенной модели равно максимальному из всех возможных значений. Аналогично определяется значение вектора максимальных (средних) значений количества экземпляров $\pi^0 = \{\pi_i^0\}, i = \overline{1, n_0}, \pi^0 = U_k \pi_k$. Вектор типов данных $\alpha^0 = \{\alpha_i^0\}, i = \overline{1, n_0}$ есть объединение векторов $\{\alpha_k\}$ объектных моделей пользователей $\alpha^0 = U_k \alpha_k$. Характеристики оперативности получения i -го информационного элемента по запросу k -го пользователя представляются в обобщенной внешней модели в виде матрицы $T^0 = \|\tau_{ki}^0\|$, проиндексированной по

осям множеством информационных элементов $D = \{d_i\}, i = \overline{1, n_0}$ и множеством пользователей $U = \{u_k\}, k = \overline{1, K_0}$. Исходная информация для матрицы T^0 содержится в векторах $\tau_k^3 = \{\tau_{ik}^3\}$.

На третьем этапе осуществляется построение обобщенной модели РИУС ПИФ. Исходной информацией для ее формирования являются данные матриц семантической смежности объектных моделей требований пользователей. Структура обобщенной объектной модели РИУС ПИФ задается в виде матрицы смежности $W = \|w_{ij}\|$, проиндексированной по осям полным множеством объектов данных и информационных элементов. Элемент матрицы w_{ij} принимает значение, равное единице, если хотя бы в одной из матриц смежности $W_k = \|w_{ij}^k\|$ объектных моделей требований пользователей между элементами d_i и d_j имеется информационная или функциональная связь, в противном случае $w_{ij} = 0$. Полученной таким образом матрице смежности, ставится в соответствие оргграф обобщенной объектной модели РИУС ПИФ $G(D, R)$, множеством вершин которого является множество объектов данных и информационных элементов ПИФ, а множеством дуг – информационные и функциональные связи между элементами.

Сформированная структура обобщенной объектной модели РИУС является инвариантной по отношению к структурам данных, поддерживаемым конкретными системами управления локальными и внешними удаленными ПБД, а также информационной и обеспечивающей инфраструктуре ПИФ.

На четвертом этапе определяются интегральные характеристики РИУС ПИФ: типы отношений между данными; множества запросов и транзакций и их характеристики; виды использования информации и характеристики вычислительной среды. Данные характеристики вычисляются с помощью методов и процедур, описанных в [1].

Матрица типов отношений между элементами $B = \|b_{ij}\|$ формируется на основании матриц $W = \|w_{ij}\|$ и совокупности матриц $B_k = \|b_{kij}\|$.

Безызыбочные множества запросов пользователей и транзакций определяются из выражений:

$$Z = Z_1 U Z_2 - Z_{1,2} U Z_3 - (Z_{2,3} U Z_{1,3}) U \dots U Z_{K_0} - \prod_{i=1}^{k_0-1} Z_{i, k_0},$$

$$K = K_1 U K_2 - K_{1,2} U K_3 - (K_{2,3} U K_{1,3}) U \dots U K_{K_0} - \prod_{i=1}^{k_0-1} K_{i, k_0}$$

Для множества иницируемых запросов и транзакций задаются соответствующие матрицы их частот: матрица $\Lambda^3 = \|\xi_{pk}^3\|, p = \overline{1, p_0}$ и матрица $\Lambda^K = \|\xi_{sk}^k\|$, отражающие частоту использования p -го запроса и s -го задания на корректировку соответственно k -м пользователем.

На множестве запросов, корректировок и множестве пользователей задаются технологические матрицы смежности $\Phi^3 = \|\phi_{kp}^3\|, k = \overline{1, K_0}, p = \overline{1, p_0}$ и $\Phi^K = \|\phi_{ks}^k\|, k = \overline{1, K_0}, s = \overline{1, s_0}$, где ϕ_{kp}^3 и ϕ_{ks}^k принимают единичные значения, если k -й пользователь использует соответственно p -й запрос и s -ю корректировку, в противном случае ϕ_{kp}^3 и ϕ_{ks}^k принимают нулевые значения.

Вид использования информационных элементов множества пользователей ПБД задается в обобщенной внешней модели в виде модифицированных матриц Хоффера $Z^3 = \|\zeta_{ki}^3\|$ и $Z^k = \|\zeta_{ks}^k\|, k = \overline{1, K_0}$, где $\zeta_{ki}^3 = 1$, если k -й пользователь использует i -й информационный элемент в запросах, $\zeta_{ki}^3 = 0$, в противном случае; $\zeta_{ks}^k = 1$, если k -й пользователь использует i -й информационный элемент в транзакциях, $\zeta_{ks}^k = 0$, в противном случае.

Технические параметры вычислительной среды (ВС) РИУС ПИФ в обобщенной объектной модели задаются в виде матрицы пропускной способности каналов связи $C^{\Pi} = \|c_{\tau_1, \tau_2}^n\|, \tau_1, \tau_2 = \overline{1, \tau_0}$, где c_{τ_1, τ_2}^n - пропускная способность каналов связи между узлами τ_1 и τ_2 , и вектора доступной пользователям внешней памяти в узлах ВС $\eta^{B3Y} = \{\eta_{\tau}^{B3Y}\}$.

На пятом этапе осуществляется нормализация обобщенной структуры РИУС ПИФ, в результате которой производится упорядочение информации по уровням иерархии, выявление и устранение дублируемых объектов данных, информационных элементов и избыточных взаимосвязей (транзитивных зависимостей). Нормализация осуществляется с использованием методов, предложенных в [5-7]. Этап нормализации обобщенной структуры заканчивается формированием оптимальной объектной модели РИУС ПИФ, формализовано представляемой в виде орграфа $G(D, R)$

Объектная модель РИУС ПИФ описывается множествами (U, D, Z, K) , векторами $(g^0, \alpha^0, \pi^0, \eta^{BZY})$ и матрицами $\{W, B, T^0, C^I, Z^3, Z^k, \Phi^3, \Phi^k, \Lambda^3, \Lambda^k\}$. Она является инвариантной используемым СУБД (ОСУБД).

Сформированная модель РИУС ПИФ используется в дальнейшем при проектировании эффективных структур ПБД и базы метаданных (БМД) репозитория ПИФ, определения оптимальной стратегии и тактики проведения патентных поисков, выбора методов и средств доступа к внешним удаленным БД патентной информации и непатентной литературы и поиска в них, сервисных средств обработки патентной информации, включая средства метапоиска, машинного перевода, конвертирования данных и других.

Заключение

В работе предложены модели, методы и алгоритмы анализа и синтеза распределенной информационно-управляющей структуры патентного информационного фонда. Рассмотрены процедуры построения моделей спецификаций информационных и функциональных требований пользователей ПИФ, формирования объектных моделей требований пользователей, построения обобщенной объектной модели РИУС ПИФ и ее нормализации. Сформированная в результате реализации предложенных процедур объектная модель РИУС ПИФ является оптимальной и инвариантной используемым программным и техническим средствам ПИФ. В последующем она используется при проектировании оптимальных структур локальных ПБД, БМД репозитория ПИФ, определения оптимальной стратегии доступа к внешним ПБД и поиска в них, выбора сервисных средств обработки патентной информации и непатентной литературы.

Предложенные методология, модели и методы использовались при формировании, совершенствовании и развитии РИУС ПИФ евразийского патентного информационного пространства, создаваемого в рамках официальной деятельности международной патентной организации - Евразийского патентного ведомства (ЕАПВ) [4,8]. Разработаны структура управления ПБД ПИФ, структуры локальных ПБД евразийского региона, ПБД международных патентных организаций и национальных патентных ведомств, средства доступа к внешним ПБД и БД научно-технической литературы, а также сервисные средства обработки патентной информации [1].

Поисковые средства информационной системы ЕАПВ ЕАПАТИС дополнились в соответствии со структурой ПИФ (РИУС ПИФ) средствами виртуального доступа и метапоиска во внешних ПБД и БД непатентной литературы, а также средствами машинного перевода. В настоящее время в ЕАПАТИС поддерживается более 25 постоянно пополняемых локальных ПБД ПИФ ЕАПВ. Общий объем содержащейся в ПБД информации превышает 85 млн. документов [8].

Использование предложенных методов, процедур и информационных технологий позволило повысить полноту и достоверность данных ПИФ, эффективность доступа к источникам патентной и непатентной информации, что предоставило возможность проведения в ЕАПВ полноценных патентных поисков международного типа по евразийским заявкам по ряду тематик.

В свою очередь, хозяйствующие субъекты евразийского региона получили в свое распоряжение эффективные и легкие в использовании инструментальные средства доступа к ПБД, навигации и поиска в них, формирования тематических баз данных и использования полученной информации при проведении патентных исследований.

Литература

1. Кульба В.В., Сиротюк В.О. Формализованная методология повышения эффективности и качества патентных информационных фондов и опыт ее использования при формировании и развитии евразийского патентно-информационного пространства. - М.:ИПУ РАН. Монография, 2019. - 236с.
2. Кульба В.В., Сиротюк В.О. Формализованная методология обеспечения полноты патентных информационных фондов. - Управление развитием крупномасштабных систем MLSD-2018. Труды 11-й международной конференции. Том 3, с.127-138.

3. *Сиротюк В.О.* Формализованные модели и методы анализа и оценки полноты патентных информационных фондов (на примере международной патентной организации) - ж. Проблемы управления, М.:ИПУ РАН, №6/2018, с.35-43.
4. *Фаязов Х.Ф., Сиротюк В.О., Овчинников А.В., Бурцев А.Б.* Формирование и развитие евразийского патентно-информационного пространства. М.: ИНИЦ «Патент», 2010.-124 с.
5. *Мамиконов А.Г., Ашимов А.А., Кульба В.В., Косяченко С.А., Сиротюк В.О.* Оптимизация структур данных в АСУ. М.: Наука, 1988. - 256 с.
6. *Кульба В.В., Ковалевский С.С., Косяченко С.А., Сиротюк В.О.* Теоретические основы проектирования оптимальных структур распределенных баз данных. Серия «Информатизации России на пороге XXI века». М.: СИНТЕГ, 1999. - 660 с.
7. *Кульба В.В., Микрин Е.А., Сиротюк В.О., Сиротюк О.В.* Модели и методы проектирования оптимальных структур объектно-ориентированных баз данных в автоматизированных информационно-управляющих системах. Научное издание. М.: ИПУ РАН, 2005. -103 с.
8. Материалы веб-портала Евразийской патентной организации: www.eapo.org