

ТЕХНІЧНІ НАУКИ

УДК 004.042 : 004.94

DOI: 10.15587/2313-8416.2017.107048

РАЗРАБОТКА ПОДХОДА К АНАЛИЗУ И ОПТИМИЗАЦИИ ДИАГРАММ ПОТОКОВ ДАННЫХ

© А. М. Копп, Д. Л. Орловский

В исследовании рассмотрены основные особенности диаграмм потоков данных, правила их построения, существующие подходы к анализу диаграмм потоков данных, их преимущества и недостатки. Предложен подход к анализу и оптимизации диаграмм потоков данных, основанный на применении анализа связей. Данный подход позволяет определять недостатки диаграмм потоков данных и формировать рекомендации по их совершенствованию

Ключевые слова: диаграммы потоков данных, структурный анализ, бизнес-процессы, анализ, оптимизация, анализ связей

1. Введение

Одним из основных инструментов структурного анализа и проектирования систем являются диаграммы потоков данных, которые используются для моделирования структуры бизнес-процессов организации и взаимодействия между ними.

Создаваемые модели используются для накопления знаний об организации и могут применяться для анализа и дальнейшего совершенствования бизнес-процессов с помощью систем класса BPM (Business Process Management). Инструментальные средства моделирования осуществляют лишь формальный контроль над разрабатываемыми диаграммами, зачастую только с точки зрения синтаксиса используемой методологии. Поэтому построение моделей бизнес-процессов носит субъективный характер, что приводит к возникновению случайных или системных ошибок, связанных с компетентностью аналитика, недостаточно подробным описанием предметной области и т. д.

Таким образом, обретает актуальность проблема анализа накапливаемых моделей деятельности организации, в том числе и диаграмм потоков данных, на предмет выявления возможных недостатков, требующих устранения.

2. Литературный обзор**2.1. Основные особенности диаграмм потоков данных**

Диаграммы потоков данных DFD (Data Flow Diagram) служат для графического представления потоков данных (data flow), протекающих в информационной системе и, как и большинство структурных моделей, являются иерархическими [1, 2]. Диаграммы потоков данных могут также использоваться

для визуализации обработки данных, связывающих внешние сущности (external entity) и накопители данных (data store) через работы (activity, process) [3, 4].

Для представления работ, потоков данных, накопителей данных и внешних сущностей в DFD используется четыре базовых обозначения. Различают графические нотации Йордана и Гейна-Сарсона, используемые при построении диаграмм потоков данных (табл. 1). Отличия между нотациями заключаются в обозначениях работ и накопителей данных, а также в отсутствии поддержки управляющих потоков данных нотацией Гейна-Сарсона [4].

При построении диаграмм DFD необходимо учитывать следующие правила [4, 5]:

1) внешние сущности не являются частью моделируемой системы, поэтому они могут иметь только входящие, либо только выходящие потоки данных;

2) потоки данных не могут напрямую связывать внешние сущности и накопители данных, они должны проходить через работы;



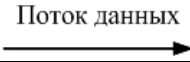
3) работы должны использовать данные – каждый накопитель данных должен иметь хотя бы один выходящий поток данных и, соответственно, каждая работа должна иметь хотя бы один входящий поток данных, направленный от накопителя данных;

4) выполнение работ должно завершаться конкретным результатом – каждая работа должна иметь хотя бы один выходящий поток данных и, соответственно, каждый накопитель данных должен иметь хотя бы один входящий поток данных, направленный от работы;

5) последовательно выполняемые работы должны быть связаны потоками данных и не могут существовать сами по себе.

Таблица 1

Элементы графических нотаций Йордона и Гейна-Сарсона

Обозначение	Нотация Йордона	Нотация Гейна-Сарсона
Работа		
Поток данных		
Накопитель данных		
Внешняя сущность		

Нарушения рассмотренных правил построения диаграмм потоков данных считаются ошибками и являются плохой практикой моделирования [4, 5]. Кроме того, в [4] вводятся следующие понятия, описывающие недостатки диаграмм DFD:

1) «спонтанная генерация (spontaneous generation)» – работа или накопитель данных, который имеет выходящие потоки данных, но не имеет ни одного входящего потока данных;

2) «черная дыра (black hole)» – работа или накопитель данных, который имеет входящие потоки данных, но не имеет ни одного выходящего потока данных;

3) «серая дыра (gray hole)» – работа или накопитель данных, который имеет, по крайней мере, один входящий и один выходящий поток данных, но входных данных очевидно недостаточно для получения выходных данных.

2.2. Существующие подходы к анализу диаграмм потоков данных

Для анализа моделей бизнес-процессов в нотациях ARIS EPC (Event-driven Process Chain) и семейства IDEF (Integrated DEFinition) используют математический аппарат теории графов и сетей Петри [6, 7]. В упрощенном виде диаграммы бизнес-процессов описываются графом [7]:

$$D = (S_0, L_0),$$

где $S_0 = \{s_1, \dots, s_q\}$ – множество элементов модели бизнес-процесса; q – количество элементов; $L_0 = \{l_1, \dots, l_r\}$ – множество связей модели бизнес-процесса; r – количество связей.

Анализ моделей бизнес-процессов в нотации BPMN (Business Process Model and Notation) также осуществляется с использованием взвешенных ориентированных графов [8, 9]. В то время как в литературе, помимо работ [5] и [10], не рассматриваются подходы к анализу диаграмм DFD.

В работе [5] предлагается формализация правил построения диаграмм DFD, а также рассматривается разработка программного обеспечения для построения и проверки диаграмм DFD на соответствие

данным правилам. Предложенный подход позволяет значительно сократить время проверки корректности диаграмм потоков данных [5], но не позволяет вырабатывать рекомендации по их совершенствованию.

Помимо правил построения диаграмм DFD, в работе [5] предлагается следующее формальное описание диаграммы потоков данных:

$$D = \{P, F, S, E\},$$

где $P = \{p_1, p_2, \dots, p_m\}$ – конечное множество работ; $F = \{f_1, f_2, \dots, f_m\}$ – конечное множество потоков данных; $S = \{s_1, s_2, \dots, s_m\}$ – конечное множество накопителей данных; $E = \{e_1, e_2, \dots, e_m\}$ – конечное множество внешних сущностей.

В работе [10] диаграммы DFD используются для оценки трудоемкости и стоимости разработки программного обеспечения на ранних стадиях жизненного цикла разработки. На стадии анализа требований DFD может использоваться для отображения функциональных возможностей программной системы [10].

Для этого предлагается подход к анализу функциональных точек на основе диаграммы DFD [10], включающий подсчет для каждой работы функций следующих типов:

1) внешний вход/выход (External Input/Output) – количество потоков данных, направленных от внешних сущностей или к внешним сущностям соответственно;

2) внутренний логический файл (Internal Logical File) – количество потоков данных внутри моделируемой системы.

Таким образом, количество функциональных точек определяется с помощью следующего выражения [10]:

$$FP = \sum_{i=1}^n \sum_{j=1}^3 N_{ij} \cdot W_{ij},$$

где n – количество работ; N_{ij} – количество функций j -го типа для i -й работы; W_{ij} – весовой коэффициент сложности функции j -го типа для i -й работы.

После этого определяется размер программного обеспечения в строках кода [10]. Трудоемкость реализации системы, моделируемой с помощью диаграммы DFD, рассчитывается с использованием методологии СОСОМО II [10].

Анализ функциональных точек, в том числе и на основе диаграмм DFD, может также использоваться для оценки трудоемкости и стоимости повторного использования и поддержки существующей системы [11].

Предложенный в работе [10] подход предполагает использование диаграмм потоков данных только для определения функциональных точек. Данный подход не направлен на определение недостатков анализируемых диаграмм DFD и возможных путей их устранения.

В русскоязычных источниках рассматривается подход к количественному анализу диаграмм IDEF0 с точки зрения их перегруженности и сложности для восприятия, основанный на расчете коэффициента сбалансированности [12, 13]:

$$K_b = \left| \frac{1}{n} \sum_{i=1}^n A_i - \max_{i \in \{1, n\}} \{A_i\} \right|,$$

где A_i – число дуг, соединенных с i -й работой.

Согласно данному подходу, в рамках одной диаграммы не желательны ситуации, когда для различных блоков значительно отличается число соединенных с ними дуг, что, как правило, является признаком ошибок при проектировании и эксплуатации бизнес-процесса [13]. Зачастую данный подход обобщают и для анализа диаграмм DFD [14], не учитывая при этом наличие на диаграммах потоков данных элементов различной природы и особенности их связей.

Так как рассмотренные подходы обладают определенными недостатками, связанными с ограниченностью их применения, они не могут использоваться для анализа накапливаемых диаграмм потоков данных, что обуславливает необходимость проведения исследований в этом направлении.

3. Цель и задачи исследования

Цель исследования – разработка подхода к анализу и оптимизации диаграмм потоков данных, позволяющего определять недостатки диаграмм DFD и формировать рекомендации по их совершенствованию.

Для достижения цели были поставлены следующие задачи:

- оценка неравномерности распределения потоков данных диаграмм DFD с использованием анализа связей;

- наглядная демонстрация результатов анализа связей диаграмм DFD с использованием диаграмм неравномерности распределения потоков данных;

- разработка математической модели для формирования рекомендаций по совершенствованию диаграмм потоков данных.

4. Анализ и оптимизация диаграмм потоков данных

4.1. Анализ связей диаграмм потоков данных

Поскольку нотация DFD предполагает описание моделируемой системы в виде сети связанных

между собой работ [1, 15, 16], для оценки отношений (связей) между узлами (внешними сущностями, накопителями данных и работами) может быть использован метод анализа данных – анализ связей [17, 18].

Для этого требуется построить следующие матрицы взаимодействия элементов диаграммы DFD:

- 1) $X_{m,r}^2$, $X_{r,m}^4$ – матрицы взаимодействия внешних сущностей и накопителей данных;
- 2) $X_{m,m}^1$ – матрицы взаимодействия внешних сущностей;
- 3) $X_{r,r}^5$ – матрицы взаимодействия накопителей данных;
- 4) $X_{m,n}^3$, $X_{n,m}^7$ – матрицы взаимодействия внешних сущностей и работ;
- 5) $X_{r,n}^6$, $X_{n,r}^8$ – матрицы взаимодействия накопителей данных и работ;
- 6) $X_{n,n}^9$ – матрица взаимодействия работ.

Таким образом, матрица взаимодействия для всех видов элементов диаграммы DFD (внешних сущностей, накопителей данных и работ) примет следующий вид:

$$X = \begin{bmatrix} X_{m,m}^1 & X_{m,r}^2 & X_{m,n}^3 \\ X_{r,m}^4 & X_{r,r}^5 & X_{r,n}^6 \\ X_{n,m}^7 & X_{n,r}^8 & X_{n,n}^9 \end{bmatrix} = (x_{ij}) \in \mathbf{R}^{g \times g},$$

где x_{ij} – количество дуг (потоков данных), направленных из i -го элемента в j -й; m – количество внешних сущностей; r – количество накопителей данных; n – количество работ; g – количество элементов диаграммы DFD, $g = n + m + r$.

Анализ связей нашел свое применение в маркетинговых и медицинских исследованиях, при оптимизации алгоритмов поисковых систем, а также в сферах борьбы с преступностью и информационной безопасности [18, 19].

Одним из основных и наиболее общим показателем структуры сети является сетевая плотность. Плотность сети характеризует количество связей в расчете на один узел сети, тем самым показывает, в какой степени узлы взаимодействуют между собой [18, 20].

Плотность связей в ориентированном графе вычисляется по следующей формуле:

$$\Delta = \frac{L}{g \cdot (g-1)},$$

где L – количество связей в графе.

Роль элементов диаграммы DFD в преобразовании информации можно определить с помощью степени центральности. Данная мера демонстрирует «активность» узла в сети и измеряется количеством его связей с другими узлами [18]. Степень центральности позволяет определить, имеются ли в сети узлы с высоким значением центральности по сравнению с остальными узлами сети, центральность которых значительно ниже [20].

Коэффициент центральности узла v_i сети определяется с помощью следующего выражения:

$$C_i = \deg^+(v_i) = \sum_{j=1}^g x_{ij} = \sum_{j=1}^g x_{ji}.$$

Для того чтобы можно было сравнивать степень центральности узла не только внутри одной сети, но и между сетями различной структуры, вычисляется нормированный коэффициент центральности [20]:

$$C'_i = \frac{\deg^+(v_i)}{g-1} = \frac{\sum_{j=1}^g x_{ij}}{g-1}.$$

Коэффициент центральности для всей сети определяет степень центральности узла с наибольшим значением коэффициента центральности C_{\max} , по отношению ко всем остальным узлам [18, 20]. Он рассчитывается по следующей формуле:

$$C = \frac{\sum_{i=1}^g [C_{\max} - C_i]}{\max_{i \in 1, g} \{C_{\max} - C_i\}}.$$

Для сравнения различных структур и определения, какая из них обеспечивает лучшую централизацию элементов, вводится нормированный коэффициент центральности сети [20]:

$$C' = \frac{\sum_{i=1}^g [C_{\max} - C_i]}{(g-1) \cdot (g-2)}.$$

Нормированный коэффициент центральности сети C' принимает значения, близкие к 0, если узлы сети незначительно отличаются по степени центральности, и значения, близкие к 1, если один узел связан с каждым из остальных узлов сети (табл. 2) [20].

Таблица 2

Значения коэффициентов центральности для различных структур сети

Структура сети	Коэффициент центральности, C'
Древовидная	0,42
Линейная	0,08
Радиальная	0,75
Кольцевая	0
Полная	0

Аналогично коэффициенту сбалансированности для IDEF0, при проектировании диаграмм DFD необходимо стремиться к достижению минимального значения коэффициентов центральности сети C и C' , поскольку данная мера демонстрирует неравномерность в распределении потоков данных между работами, внешними сущностями и накопителями данных [20]. Применительно к особенностям DFD, неравномерность в распределении потоков данных проявляется в виде ошибок, допущенных вследствие нарушения

правил построения диаграмм потоков данных, спонтанных генераций, черных и серых дыр [4, 5].

Согласно результатам исследования архитектуры предприятия [21], неравномерность загрузки элементов бизнес-архитектуры (внешние сущности и работы на диаграмме DFD) и архитектуры приложений и данных (накопители данных на диаграмме DFD) является причиной снижения производительности моделируемой системы. Таким образом, выявление неравномерности в загрузке элементов позволяет определить «узкие места» моделируемой системы [22].

Кроме того, использование диаграмм DFD с неравномерным распределением потоков данных может привести к ошибкам при определении трудоемкости и стоимости разработки, поддержки и повторного использования программного обеспечения [10, 11].

4. 2. Формирование рекомендаций по совершенствованию диаграмм потоков данных

Для наглядной демонстрации неравномерности распределения потоков данных в диаграмме DFD, будем использовать диаграмму, аналогичную функционально-стоимостной диаграмме, которая использовалась для анализа моделей бизнес-процессов в нотации IDEF0 [23]. Она позволит определить диспропорции в количестве связей по каждому элементу диаграммы DFD [24, 25].

В верхней части диаграммы будут располагаться значения сетевой плотности Δ , а в нижней приводиться значения нормированных коэффициентов центральности C'_i элементов диаграммы DFD, что позволит оценить диспропорции в распределении потоков данных [25].

После построения данной диаграммы, также как и при функционально-стоимостном анализе, необходимо выявить и устранить причины несоответствия значений коэффициентов центральности элементов, и плотности диаграммы DFD [23, 26, 27].

Формирование рекомендаций по совершенствованию диаграмм DFD заключается в определении изменений, при которых значение коэффициента центральности сети C' будет минимальным и, соответственно, потоки данных будут распределены равномерно:

1) изменения в количестве потоков данных ΔC_i^1 , направленных от i -го элемента диаграммы к внешним сущностям;

2) изменения в количестве потоков данных ΔC_i^2 , направленных от i -го элемента диаграммы к накопителям данных;

3) изменения в количестве потоков данных ΔC_i^3 , направленных от i -го элемента диаграммы к работам.

Поскольку нотация DFD рассматривает элементы различной природы и назначения, при вычислении коэффициента центральности для диаграммы DFD необходимо учитывать виды элементов (внешние сущности, накопители данных и работы) и особенности связей между ними. Для этого, каждому элементу диаграммы DFD поставим в соответствие весовой коэффициент w_i . Таким образом, коэффициент центральности сети будет определяться по формуле:

$$C'' = \frac{\sum_{i=1}^g [w_i \cdot (C_{\max} - C_i)]}{(g-1) \cdot (g-2)}$$

Для определения значений весовых коэффициентов w_i будем использовать типы связности для данных, определенные в стандарте ISO/IEC/IEEE 24765 (табл. 3).

Таблица 3

Типы связности и соответствующие им значения весовых коэффициентов

Значимость	Тип связности	Описание	Весовой коэффициент, w_i
0	Случайная	Случайная	0
1	Логическая	Данные одного и того же множества или типа	0,17
2	Временная	Данные, используемые в каком-либо временном интервале	0,33
3	Процедурная	Данные, используемые во время одной и той же фазы или итерации	0,5
4	Коммуникационная	Данные, на которые воздействует одна и та же деятельность	0,67
5	Последовательная	Данные, преобразуемые последовательными функциями	0,83
6	Функциональная	Данные, связанные с одной функцией	1

Таким образом, для определения изменений в количестве потоков данных $\Delta C_i^1, \Delta C_i^2, \Delta C_i^3$ необходимо решить задачу, математическая модель которой примет следующий вид:

$$C'' = \frac{\sum_{i=1}^g \left[w_i \cdot \left(\max_{j \in 1, g} \{C_j + \Delta C_j^1 + \Delta C_j^2 + \Delta C_j^3\} - [C_i + \Delta C_j^1 + \Delta C_j^2 + \Delta C_j^3] \right) \right]}{(g-1) \cdot (g-2)} \rightarrow \min,$$

$$C_i + \Delta C_j^1 + \Delta C_j^2 + \Delta C_j^3 \leq C_{\max}, \forall i = 1, \dots, g,$$

$$\Delta C_j^1, \Delta C_j^2, \Delta C_j^3 - \text{целые числа}, \forall i = 1, \dots, g.$$

максимально допустимое количество потоков данных, выходящих из моделируемой системы;

2) потоки данных не могут напрямую связывать внешние сущности и накопители данных, они должны проходить через работы:

$$\sum_{j=1}^m x_{ij} + \Delta C_i^1 = 0, \forall i = 1, \dots, m,$$

$$\sum_{j=m+1}^{m+r} x_{ij} + \Delta C_i^2 = 0, \forall i = 1, \dots, m,$$

$$\sum_{j=1}^m x_{ij} + \Delta C_i^1 = 0, \forall i = m+1, \dots, m+r,$$

В соответствии с правилами построения диаграмм DFD [4, 5], математическая модель данной задачи должна включать следующие ограничения:

1) внешние сущности не являются частью моделируемой системы, поэтому они могут иметь только входящие, либо только выходящие потоки данных:

$$\sum_{j=m+r+1}^g x_{ij} + \Delta C_i^3 \geq 0, \forall i = 1, \dots, m,$$

$$\sum_{j=1}^m x_{ij} + \Delta C_i^1 \geq 0, \forall i = m+r+1, \dots, g,$$

$$I_{\min} \leq \sum_{i=1}^m \left(\sum_{j=m+r+1}^g x_{ij} + \Delta C_i^1 \right) \leq I_{\max},$$

$$O_{\min} \leq \sum_{i=m+r+1}^g \left(\sum_{j=1}^m x_{ij} + \Delta C_i^3 \right) \leq O_{\max},$$

где I_{\min} и I_{\max} – минимально и максимально допустимое количество потоков данных, входящих в моделируемую систему; O_{\min} и O_{\max} – минимально и

$$\sum_{j=m+1}^{m+r} x_{ij} + \Delta C_i^2 = 0, \forall i = m+1, \dots, m+r;$$

3) работы должны использовать данные – каждый накопитель данных должен иметь хотя бы один выходящий поток данных и, соответственно, каждая работа должна иметь хотя бы один входящий поток данных, направленный от накопителя данных:

$$\sum_{j=m+r+1}^g x_{ij} + \Delta C_i^3 \geq 1, \forall i = m+1, \dots, m+r,$$

$$\sum_{i=m+1}^{m+r} \left(\sum_{j=m+r+1}^g x_{ij} + \Delta C_i^3 \right) \geq n;$$

4) выполнение работ должно завершаться конкретным результатом – каждая работа должна иметь хотя бы один выходящий поток данных и, соответственно, каждый накопитель данных должен иметь хотя бы один входящий поток данных, направленный от работы:

$$\sum_{j=m+1}^{m+r} x_{ij} + \Delta C_i^2 \geq 1,$$

$$\forall i = m+r+1, \dots, g,$$

$$\sum_{i=m+r+1}^g \left(\sum_{j=m+1}^{m+r} x_{ij} + \Delta C_i^2 \right) \geq r;$$

5) последовательно выполняемые работы должны быть связаны потоками данных и не могут существовать сами по себе:

$$\sum_{j=m+r+1}^g x_{ij} + \Delta C_i^3 \geq 1,$$

$$\forall i = m+r+1, \dots, g-1,$$

$$\sum_{j=m+r+1}^g x_{ij} + \Delta C_g^3 \geq 0.$$

Кроме того, необходимо ввести ограничения, при которых диспропорции значений Δ и C'_i на диаграмме будут минимальными [9]:

$$\left| \frac{C_i + \Delta C_i^1 + \Delta C_i^2 + \Delta C_i^3}{g-1} - \Delta \right| \leq \delta_i, \forall i = 1, \dots, g,$$

где δ_i – допустимые значения отклонений значений коэффициента центральности C'_i от плотности сети Δ .

Предложенная задача является целочисленной задачей нелинейного программирования и может быть решена с помощью соответствующих математических методов.

5. Результаты исследования и их обсуждение

В качестве примера будем использовать диаграмму потоков данных, описывающую бизнес-процессы поставки продукции (рис. 1). Для создания диаграмм DFD использовалась пробная версия пакета AllFusion Process Modeler.



Рис. 1. Исходная диаграмма DFD

Проанализировав данную диаграмму (рис. 1) на соответствие правилам построения, были определены следующие недостатки, вследствие наличия которых не выполняются неравенства, записанные в ограничениях предложенной задачи оптимизации:

1) потоки данных «Информация о состоянии склада» напрямую связывают внешнюю сущность «Данные о количестве продукции на складе» и накопитель данных «Состояние склада»;

2) поток данных «Информация о предлагаемой продукции» напрямую связывает накопители данных «Поставщики» и «Состояние рынка»;

3) накопитель данных «Состояние рынка» не имеет входящих потоков данных, направленных от работ и является, по сути, «черной дырой»;

4) накопитель данных «Поставщики» является «серой дырой», поскольку входных данных очевидно недостаточно для получения сведений о поставщиках;

5) работы «Согласование с поставщиком сроков и формы оплаты» и «Доставка продукции» не имеют выходящих потоков данных, направленных к накопителям данных.

Наличие выделенных недостатков построения диаграммы DFD, как было отмечено ранее, является

причиной неравномерного распределения потоков данных. Поэтому, для анализируемой диаграммы DFD были определены коэффициенты сбалансированности и центральности (табл. 5, 6), а также была построена диаграмма неравномерности распределения потоков данных (рис. 3, а).

Для определения изменений, при которых потоки данных на диаграмме DFD будут распределены равномерно, и она будет соответствовать правилам построения диаграмм потоков данных, была решена задача оптимизации, рассмотренная в предыдущем разделе.

В качестве минимально и максимально допустимого количества потоков данных, входящих и выходящих из моделируемой системы, были приняты следующие значения:

$$I_{\min} = 2, I_{\max} = 2,$$

$$O_{\min} = 1, O_{\max} = 2.$$

В качестве допустимых значений отклонения C'_i от плотности сети $\Delta = 0,21$, были приняты следующие значения:

$$\delta_i = 0,1, \forall i = 1, \dots, 9.$$

В результате решения данной задачи, были получены следующие изменения в количестве потоков данных, направленных от i -го элемента диаграммы DFD к внешним сущностям, накопителям данных и работам (табл. 4).

Таблица 4

Изменения, при которых потоки данных будут распределены равномерно

i	1	2	3	4	5	6	7	8	9
ΔC_i^1	0	0	0	-1	0	0	0	0	-1
ΔC_i^2	-1	0	0	0	-1	0	1	1	0
ΔC_i^3	0	0	0	0	-1	0	0	0	0

С использованием полученных результатов, были сформированы рекомендации по совершенствованию исходной диаграммы DFD (рис. 1) и построена преобразованная диаграмма потоков данных (рис. 2), соответствующая правилам построения диаграмм DFD, требованиям к сбалансированности и равномерности распределения потоков данных (табл. 5, 6).

Поскольку изменения (табл. 4), полученные в результате решения рассмотренной задачи оптимизации, не предполагают добавление входящих потоков данных

для накопителя данных «Состояние рынка», данный элемент был представлен в виде внешней сущности.

В следующих таблицах приведены значения коэффициентов сбалансированности и центральности, рассчитанные для исходной и преобразованной диаграммы DFD (табл. 5, 6).

Для наглядной демонстрации неравномерности распределения потоков данных в исходной (рис. 3, а) и преобразованной диаграмме DFD (рис. 3, б), были построены следующие диаграммы.



Рис. 2. Преобразованная диаграмма DFD

Таблица 5

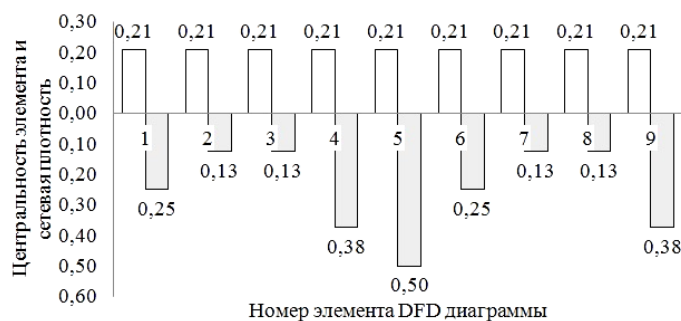
Коэффициенты центральности для элементов диаграмм DFD

№	Элемент диаграммы DFD	Весовой коэффициент, w_i	Исходная диаграмма		Преобразованная диаграмма	
			C_i	C'_i	C_i	C'_i
1	Данные о количестве продукции на складе	0,5	2	0,25	1	0,13
2	Данные о заказах клиентов	0,5	1	0,13	1	0,13
3	Состояние рынка	0,5	1	0,13	1	0,13
4	Состояние склада	0,67	3	0,38	2	0,25
5	Поставщики	0,67	4	0,5	2	0,25
6	Формирование заказа на поставку	0,83	2	0,25	2	0,25
7	Согласование с поставщиком сроков и формы оплаты	0,83	1	0,13	2	0,25
8	Доставка продукции	0,83	1	0,13	2	0,25
9	Прием поставленной продукции	0,83	3	0,38	2	0,25

Таблица 6

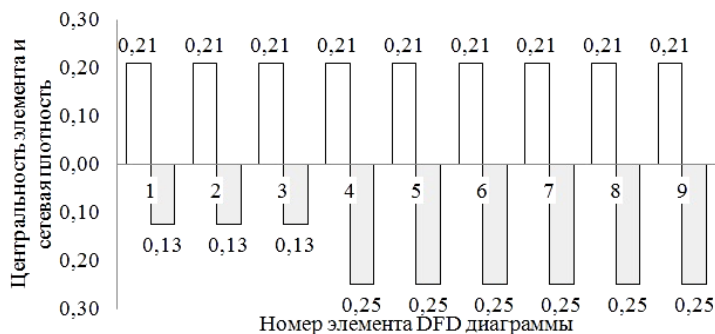
Коэффициенты сбалансированности и центральности для диаграмм DFD

Коэффициент	Исходная диаграмма	Преобразованная диаграмма
K_b	2	0,33
C'	0,32	0,05
C''	0,22	0,03



□ – плотность сети □ – нормированный коэффициент центральности

а



□ – плотность сети □ – нормированный коэффициент центральности

б

Рис. 3. Диаграммы неравномерности распределения потоков данных: а – для исходной диаграммы DFD; б – для преобразованной диаграммы DFD

Очевидно, что диспропорции в распределении потоков данных для преобразованной диаграммы DFD (рис. 3, б) значительно меньше, чем для исходной диаграммы (рис. 3, а). Максимальное отклонение значений коэффициента центральности от плотности сети для преобразованной диаграммы DFD составляет 0,08, в то время как для исходной диаграммы оно составляет 0,5.

6. Выводы

В данном исследовании был разработан подход к анализу и оптимизации диаграмм потоков данных, позволяющий определять недостатки диаграмм DFD и формировать рекомендации по их совершенствованию.

1. Для оценки неравномерности распределения потоков данных диаграмм DFD был использован анализ связей.

2. Для наглядной демонстрации результатов анализа связей диаграмм DFD были предложены диаграммы неравномерности распределения потоков данных.

3. Для формирования рекомендаций по совершенствованию диаграмм потоков данных была разработана соответствующая математическая модель.

В результате применения предложенного подхода, диаграмма потоков данных, описывающая бизнес-процессы поставки продукции, была преобразована в соответствии с правилами построения диаграмм DFD, диспропорции в распределении потоков данных на полученной диаграмме минимальны.

Литература

1. Yourdon, Ed. Just Enough Structured Analysis [Text] / Ed. Yourdon. – 2006. – 643 p. – Available at: <http://zimmer.csufresno.edu/~sasanr/Teaching-Material/SAD/JESA.pdf>
2. Hatley, D. Process for System Architecture and Requirements Engineering [Text] / D. Hatley, P. Hruschka, I. Pirbhai. – Addison-Wesley, 2013. – 456 p.
3. Miller, F. P. Data Flow Diagram: Information System, Visualization, Computer Data Processing, Flowchart, Control Flow Diagram, Data Island, Dataflow, Functional Flow Block Diagram [Text] / F. P. Miller, A. F. Vandome, J. McBrewster. – Alphascript Publishing, 2010. – 80 p.
4. Shelly, G. Systems Analysis and Design [Text] / G. Shelly, H. Rosenblatt. – Cengage Learning, 2009. – 742 p.
5. Ibrahim, R. Formalization of the Data Flow Diagram Rules for Consistency Check [Text] / R. Ibrahim, S. Y. Yen // International Journal of Software Engineering & Applications. – 2010. – Vol. 1, Issue 4. – P. 95–111. doi: 10.5121/ijsea.2010.1406
6. Van der Aalst, W. Modeling Business Processes: A Petri Net-Oriented Approach [Text] / W. Van der Aalst, C. Stahl. – Cambridge: MIT Press, 2011. – 386 p.
7. Дорпер, М. Г. Алгоритм преобразования моделей бизнес-процессов в одноцветные сети Петри [Текст] / М. Г. Дорпер // Модел. и анализ информ. систем. – 2010. – Т. 28, № 12. – С. 5–16.
8. Gomez-Lopez, M. T. Guiding the Creation of Choreographed Processes with Multiple Instances Based on Data Models [Text] / M. T. Gomez-Lopez, J. M. Perez-Alvarez, A. J. Varela-Vaca, R. M. Gasca // Lecture Notes in Business Information Processing. – 2017. – P. 239–251. doi: 10.1007/978-3-319-58457-7_18
9. Roa, J. Ontology-Based Heuristics for Process Behavior: Formalizing False Positive Scenarios [Text] / J. Roa, E. Reynares, M. L. Calusco, P. Villarreal // Lecture Notes in Business Information Processing. – 2017. – P. 106–117. doi: 10.1007/978-3-319-58457-7_8
10. Arnuphaptrairong, T. Early Stage Software Effort Estimation Using Function Point Analysis: An Empirical Validation [Text] / T. Arnuphaptrairong // International Journal of Design, Analysis and Tools for Integrated Circuits and Systems. – 2013. – Vol. 4, Issue 1. – P. 15–21.
11. Singh, K. A. Survey of Various cost & Effort Estimation Models [Text] / K. Singh, U. Dwivedi // International Journal of Advanced Research in Computer Science and Software Engineering. – 2014. – Vol. 4, Issue 8. – P. 1113–1116.
12. Анализ предметной области оборота рабочей конструкторской документации [Электронный ресурс]. – Режим доступа: http://www.kb-ametist.com/rus/18_publish/10.pdf
13. Замятина, О. М. Моделирование систем [Текст]: учеб. пос. / О. М. Замятина. – Томск: Изд-во ТПУ, 2009. – 204 с.
14. Вендров, А. М. Проектирование программного обеспечения экономических информационных систем [Текст]: учеб. / А. М. Вендров. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2005. – 544 с.
15. Helmond, A. The Web as Platform: Data Flows in Social Media [Text]: PhD thesis / A. Helmond. – University of Amsterdam, 2015. – 190 p.
16. Feijs, L. Notations for Software Design [Text] / L. Feijs, H. Jonkers, C. Middelburg. – Springer Science & Business Media, 1994. – 425 p. doi: 10.1007/978-1-4471-2107-7
17. Jawadkar, W. S. Management Information Systems: Text and Cases: a Global Digital Enterprise Perspective [Text] / W. S. Jawadkar. – Tata McGraw-Hill Education, 2013. – 814 p.
18. Fouss, F. Algorithms and Models for Network Data and Link Analysis [Text] / F. Fouss, M. Saerens, M. Shimbo. – Cambridge University Press, 2016. – 543 p. doi: 10.1017/cbo9781316418321
19. Boshmaf, Y. Graph-based Sybil Detection in Social and Information Systems [Text] / Y. Boshmaf, K. Beznosov, M. Ripeanu // Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. – 2013. – P. 466–473. doi: 10.1145/2492517.2492568
20. McCulloh, I. Social Network Analysis with Applications [Text] / I. McCulloh, H. Armstrong, A. Johnson. – John Wiley & Sons, 2013. – 320 p.
21. Iacob, M.-E. Quantitative Analysis of Enterprise Architectures [Text] / M.-E. Iacob, H. Jonkers // Interoperability of Enterprise Software and Applications. – 2006. – P. 239–252. doi: 10.1007/1-84628-152-0_22
22. Nguyen, H. Business Process Performance Mining with Staged Process Flows [Text] / H. Nguyen, M. Dumas, A. Hofstede, M. La Rosa, F. M. Maggi // Advanced Information Systems Engineering. CAiSE 2016. Lecture Notes in Computer Science. – 2016. – P. 167–185. doi: 10.1007/978-3-319-39696-5_11
23. Копп, А. М. Об одном подходе к решению задачи оптимизации структуры бизнес-процессов предприятия [Текст] / А. М. Копп, Д. Л. Орловский // Вестник НТУ «ХПИ». – 2015. – № 58. – С. 102–108.
24. Mead, N. Requirements Prioritization Case Study Using AHP [Text] / N. Mead // Software Engineering Institute. – 2008.

25. Щербаков, В. А. Комплексный экономический анализ хозяйственной деятельности предприятия в рыночной экономике [Текст] / В. А. Щербаков. – Новосибирск: НГАВТ, 2012. – 216 с.
26. Cross, N. Engineering Design Methods: Strategies for Product Design [Text] / N. Cross. – Wiley, 2008. – 230 p.
27. Miles, L. D. Techniques of Value Analysis and Engineering [Text] / L. D. Miles. – Miles Value Foundation, 2015. – 433 p.

*Рекомендовано до публікації д-р техн. наук, професор Годлевський М. Д.
Дата надходження рукопису 22.05.2017*

Копп Андрей Михайлович, аспирант, кафедра программной инженерии и информационных технологий управления, Национальный технический университет «Харьковский политехнический институт», ул. Кирпичева, 2, г. Харьков, Украина, 61002
E-mail: korp93@gmail.com

Орловский Дмитрий Леонидович, кандидат технических наук, доцент, кафедра программной инженерии и информационных технологий управления, Национальный технический университет «Харьковский политехнический институт», ул. Кирпичева, 2, г. Харьков, Украина, 61002
E-mail: ordm@kpi.kharkov.ua