



ISSN 2072-8441

ЭКОНОМИКА И УПРАВЛЕНИЕ

<http://elibrary.miu.by/journals!/item.eiup.html>

Кузьминов, В.В. Подходы к обработке больших потоков информации при проектировании информационных систем / В.В. Кузьминов, Т.А. Ермакова // Экономика и управление. – 2014. – № 2 (38). – С. 25–30.

ПОДХОДЫ К ОБРАБОТКЕ БОЛЬШИХ ПОТОКОВ ИНФОРМАЦИИ ПРИ ПРОЕКТИРОВАНИИ ИНФОРМАЦИОННЫХ СИСТЕМ

В.В. Кузьминов^а, Т.А. Ермакова^б

КЛЮЧЕВЫЕ СЛОВА

большие данные (Big Data), предпроектное обследование, моделирование информационных потоков, методика исследования информационных потоков

АННОТАЦИЯ

В статье анализируются методы и алгоритмы построения модели информационных потоков большой размерности, охватывающих процесс проектирования корпоративной информационной системы от проведения предпроектного обследования до построения модели предприятия, являющейся основой для проектирования информационной сети предприятия. Авторы приводят методику и разработанный ими алгоритм исследования и определения характеристик потоков информации, циркулирующей в проектируемой системе на примере процесса технической подготовки производства в ОАО «Белшина».

СТАТЬЯ ПОСТУПИЛА В РЕДАКЦИЮ

16 декабря 2013

WEB

<http://elibrary.miu.by/journals!/item.eiup/issue.38/article.5.html>

APPROACHES TO PROCESSING LARGE DATA FLOWS IN DESIGN OF INFORMATION SYSTEMS

V.V. Kuzminov^a, T.A. Yermakova^b

KEYWORDS

Big Data, pre-project analysis, modelling of information flows, methodology for a study of information flows

ABSTRACT

The article examines the methods and algorithms of designing a model of data flows of large dimension, covering the design process of a corporate information system carried out from a pre-project analysis to building a business model which is the basis for the design of the information network of an enterprise. The authors provide the methodology and algorithm that they developed for the study and determination of characteristics of data flows circulating in the projected system on the example of the process of technical preparation of production at JSC «Belshina».

RECEIVED

December 16, 2013

WEB

<http://elibrary.miu.by/journals!/item.eiup/issue.38/article.5.html>

Введение

Во многих крупных компаниях аккумулируется огромное количество данных. Методы анализа данных существенно отличаются друг от друга по производительности, качеству результатов, удобству применения, требованиям к данным.

^а Кузьминов Вячеслав Васильевич, кандидат технических наук, доцент кафедры информатики, статистики и высшей математики Бобруйского филиала Белорусского государственного экономического университета
Kuzminou Viachaslau Vasilevich, PhD in Engineering sciences, associate professor of Department of Information Science, Statistics and Higher Mathematics of Bobruisk Branch of Belarus State Economic University
kuzminovvv@gmail.com

^б Ермакова Татьяна Александровна, кандидат экономических наук, доцент, заведующий кафедрой информатики, статистики и высшей математики Бобруйского филиала Белорусского государственного экономического университета
Yermakova Tatsiana Aleksandrovna, PhD in Economics, Associate Professor, head of Department of Information Science, Statistics and Higher Mathematics of Bobruisk Branch of Belarus State Economic University
ermakova@bfgu.by

Большие данные (Big Data) в информационных технологиях — серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объемов и значительного многообразия для получения воспринимаемых человеком результатов и характеризуются тремя V: volume, variety, velocity (объем, многообразие, скорость) [1].

Производительность при обработке больших объемов данных можно повысить различными способами:

— с помощью оборудования: многопроцессорных систем, оперативных запоминающих устройств большой емкости, RAID-массивы;

— с использованием баз данных: «тяжелые» СУБД (системы управления базами данных реального времени, оптимизированные на быструю запись и чтение больших объемов информации 24 часа в сутки. Примерами могут служить СУБД PowerBuilder, SIAD/SQL 6), разбиение на разделы, оптимальное индексирование;

— через аналитическую платформу: параллельная обработка, кэширование данных, комбинирование простых и сложных моделей;

— с помощью исходной информации: репрезентативные выборки, сегментирование данных, группировка;

— применяя алгоритмы: масштабируемые алгоритмы, иерархические модели.

Анализ большого объема данных требует особого подхода, т.к. технически сложно их переработать при помощи использования только лишь мощного оборудования. Конечно, можно увеличить скорость обработки данных за счет более производительного оборудования, тем более, что современные серверы и рабочие станции используют многоядерные процессоры, оперативную память значительных размеров и мощные дисковые массивы. Однако есть множество других способов обработки больших объемов данных, которые позволяют повысить масштабируемость и не требуют бесконечного обновления оборудования.

Используемые методики обработки больших объемов информации

С точки зрения системного анализа [2] разработка корпоративных информационных систем (КИС) является довольно сложным процессом и некоторые алгоритмы информационного поиска могут требовать значительных информационных ресурсов или объемов занимаемой памяти [3].

Можно выделить семь основных этапов разработки информационных систем:

1. Исследование бизнес-процессов заказчика. Постановка задач.
2. Разработка технического задания на информационную систему.
3. Создание программного кода. Организация рабочей среды для отработки тестовых задач.
4. Тестирование и внедрение информационной системы.
5. Разработка документации. Устранение недостатков.
6. Передача информационной системы в эксплуатацию.
7. Сопровождение. Информационная поддержка.

Применение таких инструментальных CASE-средств как BPwin/ERwin, Rational Rose и др. позволяет значительно повысить качество и эффективность проектирования на данных этапах, а также ускорить процесс разработки [4].

Однако один из наиболее сложных и важных этапов разработки КИС — этап формирования концепции и исследование бизнес-процессов — остается во многом неформализованным. На данном этапе определяется круг задач, которые предполагается решать с использованием разрабатываемой информационной системы, определяются требования к самой системе. Детально исследуется каждая цепочка операций автоматизируемых бизнес-процессов. Ведется активная работа с конечными пользователями на предмет выявления наиболее трудоемких операций. Определяется перечень входной и выходной информации, формы действующей и требуемой отчетности. Исследуются возможности интеграции с другими информационными системами, используемыми заказчиком.

Начальные фазы проекта имеют решающее влияние на достигаемый результат, так как в них закладываются

основные решения, определяющие качество информационной системы. Доля вклада в конечный результат концептуальной фазы достигает 30 % [5].

Самую большую сложность 1-го этапа представляет сбор исходных данных и анализ существующих бизнес-процессов предприятия. Это вызвано огромным объемом информации, который необходимо собрать, упорядочить и обработать, особенно на стадии предпроектного обследования. В настоящее время существует несколько общепотребимых методик обследования [6], однако их общим недостатком является недостаточная формализованность, большинство операций производится вручную. Следствием такого подхода являются значительные издержки как по времени, так и по ресурсам, а результат обследования представляет собой массу несистематизированной информации, которую для дальнейшего анализа еще предстоит предварительное обработать [4].

С другой стороны, правильно и полно собранные материалы обследования содержат в себе информацию о предприятии, достаточную для проектирования всех компонентов КИС, как программных, так и технических. Так, например, проектирование информационной сети предприятия в настоящее время опирается на вполне четко формализованные методики расчета [7, 8], однако ни одна из этих методик не затрагивает способ получения исходных данных для проведения расчетов, т.е. способ получения полной информации (в том числе количественной) о документообороте на предприятии.

В современной практике использования информационных технологий нашли применение ряд моделей, позволяющие представить данные большой размерности [8]. К ним следует отнести булевы модели и векторные модели поиска.

Булева модель является классической и широко используется с целью представления информации [9]. Популярность этой модели обусловлена простотой ее реализации, позволяющей в массивах документов большого объема индексировать и выполнять поиск. В рамках булевой модели документы и запросы представляются в виде множества ключевых слов — термов.

Большинство известных информационно-поисковых систем и систем классификации информации в той или иной мере основываются на использовании векторной модели описания данных. Векторная модель является классической алгебраической моделью [7]. В рамках этой модели документ описывается вектором в некотором евклидовом пространстве. Каждому используемому в документе терму ставится в соответствие его весовой коэффициент. Актуальной задачей в процессе использования векторной модели в информационно-поисковых системах является задача снижения размерности представленных данных. Этот процесс повышает скорость обработки и выполнения поиска по заданному вектору неструктурированных документов.

По существующим оценкам, неструктурированные данные — главным образом, текст — составляют не менее 90 % информации, с которой имеют дело фирмы и организации. И лишь 10 % приходится на структурированные данные, загружаемые в реляционные СУБД [10].

Главная проблема современных коммуникаций — это извлечение действительно ценных сведений из информационных потоков, или, другими словами, получение знаний из информации [11]. Для решения такой проблемы подходят технологии, получившее сегодня название Data Mining и Text Mining [6].

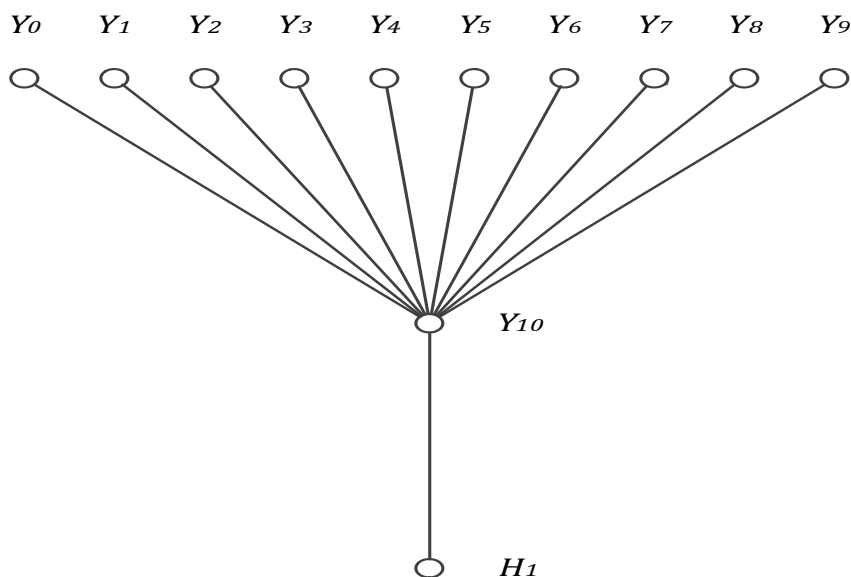


Рисунок 1 — Расширенный структурный граф задачи «Расчёт норм расхода полуфабрикатов на 1000 штук комплектующих»

Технологии Data Mining и Text Mining являются новой тенденцией в развитии средств и методов обработки данных. Они помогают найти скрытые закономерности и отношения в исследуемых данных для того, чтобы можно было бы принять более обоснованные решения. Опыт многих предприятий показывает, что отдача от использования технологий Data Mining и Text Mining может достигать 1000 % [1].

Таким образом, вышеизложенные материалы отражают всю совокупность сложных технологий поиска и обработки данных в информационных потоках. Существующие технологии построены на моделировании характера и сути процессов, происходящих внутри информационного потока.

Применение методики исследования информационных потоков

Информационные потоки наиболее удобно представлять в виде информационной модели, которая разрабатывается в соответствии с принципиальной схемой матричной модели. Она содержит сведения о документах, маршрутах их движения, формируемых показателях потоков, точках возникновения возмущающих факторов. Этот метод основан на использовании теории графов и теории матриц и позволяет автоматизировать процесс анализа потоков информации, расчета ее объемов, определить вычислительную сложность алгоритмов задач и их синтез [5]. В соответствии с этим методикой исследования потоков информации можно разделить на два этапа: получение характеристик по информационным связям задач и получение основных характеристик потока информации [12]. На первом этапе определяется количество независимо решаемых и взаимосвязанных задач, последовательность решения задач. На втором этапе определяется число тактов движения информации, порядок каждой компоненты потока, длина путей, длительность хранения компонента и др.

На примере проектирования подсистемы технической подготовки шинного производства (ТПШП) ОАО «Белшина» рассмотрим применение этой методики для обработки больших объемов информации.

Исходной информацией для анализа потоков сообщений является граф взаимосвязи документов по задачам, решаемым внутри подсистемы ТПШП.

Традиционными способами машинного представления графов являются матрицы, благодаря развитому аппарату анализа которых могут быть получены достаточно хорошие решения структурных задач [13]. Однако ограничением их применения является информационная избыточность. Для моделей больших размерностей и сильносвязных структур это ограничение возрастает [14]. Приведем один из способов, позволяющий избежать этого [12].

Согласно функциям подсистемы ТПШП, технологическому процессу и выбранному комплексу задач, для каждой задачи составляется информационный граф, на основании которого строится матрица смежности A .

Рассмотрим одну из задач этой подсистемы: «Расчёт норм расхода полуфабрикатов на 1000 штук комплектующих». Для ее решения необходимы следующие информационные файлы:

- Y_0 — конструкторская спецификация;
- Y_1 — типы деталей;
- Y_2 — готовая продукция;
- Y_3 — техническая база нормирования;
- Y_4 — сырье и материалы;
- Y_5 — резиновые смеси;
- Y_6 — плановая рецептура;
- Y_7 — комплектность автошин;
- Y_8 — паспорт конструкторской спецификации;
- Y_9 — технологические потери;
- Y_{10} — нормы расхода полуфабрикатов на 1000 штук изделий.

Если результатом решения задачи является документ или файл Y_i , то можно получить расширенный структурный граф задачи, дополняя его еще одной вершиной H_k и соединяя дугой вершину Y_i с вершиной H_k . В полученном графе (рисунок 1) из вершины H_i не выходит ни одна дуга. Далее строится матрица смежности для данной задачи (таблица 1).

Для большого количества задач и сложной взаимосвязи между ними рекомендуется матрицы триангулировать. Матрица A триангулирована тогда и только тогда, когда соответствующий ей граф $G(U, V)$ не имеет контуров. U и V — соответственно вершины и ребра графа. Триангуляцию матриц можно выполнить различными

Таблица 1 — Матрица смежности для задачи «Расчёт норм расхода полуфабрикатов на 1000 штук комплектующих»

$I \setminus j$	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
γ_0	0	0	0	0	0	0	0	0	0	0	1
γ_1	0	0	0	0	0	0	0	0	0	0	1
γ_2	0	0	0	0	0	0	0	0	0	0	1
γ_3	0	0	0	0	0	0	0	0	0	0	1
γ_4	0	0	0	0	0	0	0	0	0	0	1
γ_5	0	0	0	0	0	0	0	0	0	0	1
γ_6	0	0	0	0	0	0	0	0	0	0	1
γ_7	0	0	0	0	0	0	0	0	0	0	1
γ_8	0	0	0	0	0	0	0	0	0	0	1
γ_9	0	0	0	0	0	0	0	0	0	0	1
γ_{10}	0	0	0	0	0	0	0	0	0	0	0

ми алгоритмами, но в основе их лежит идея перенумерации вершин графа.

Если матрицу A представить в виде:

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}, \tag{1}$$

где B, C, D — подматрицы матрицы A , получаемые разбиением ее на четыре равные части, то можно получить алгоритм для вычисления матрицы [8]:

$$A^n = \begin{pmatrix} B^n & B^{n-1} \cdot C \cdot D^0 + B^{n-2} \cdot C \cdot D^1 + \dots + B^0 \cdot C \cdot D^{n-1} \\ 0 & D^n \end{pmatrix}, \tag{2}$$

Следовательно, для триангулированной матрицы при четырехблочном разбиении нет необходимости обрабатывать четвертую часть информации, т.к. она заведомо равна нулю, что экономит 25 процентов занимаемой памяти. При разбиении матрицы на 16 блоков объем экономии составляет уже 37 процентов. Блочное разбиение матрицы уменьшает также число операций суммирования и умножения при обработке матриц (таблица 2) [12].

Зададимся вероятностями появления ошибок на всех стадиях обработки информации: подготовка исходных данных, ввод, расчет и выдача результатов соответственно через P_1, P_2, P_3, P_4 . Если рассматривать обработку информации как независимые события, то число ошибок без разбиения матрицы на блоки γ и с разбиением на блоки γ_1 будут соответственно равны:

$$\gamma = C_1 \cdot P_1 + C_2 \cdot P_2 + C_3 \cdot P_3 + C_4 \cdot P_4; \tag{3}$$

$$\gamma_1 = C_1^1 \cdot P_1 + C_2^1 \cdot P_2 + C_3^1 \cdot P_3 + C_4^1 \cdot P_4, \tag{4}$$

где C_i и C_i^1 — число обрабатываемых символов без разбиения и с разбиением матрицы на блоки.

Аналогичным путем строятся матрицы смежности и для остальных задач. Выполняя суммирование этих матриц, получим матрицу смежности A , отражающую взаимосвязь исходной информации по всем задачам подсистемы:

$$A = A_1 + A_2 + \dots + A_m, \tag{5}$$

которой соответствует граф $G(U, V)$, являющийся результатом объединения графов каждой задачи $G_1(U_1, V_1) \dots G_m(U_m, V_m)$:

$$G(U, V) = G_1(U_1, V_1) \dots G_m(U_m, V_m).$$

Пользуясь известными свойствами графов [5], можно выявить ряд важных характеристик потоков информации, получая последовательность матриц $A_1, A_2, A_3, \dots, A^n$,

$$A\Sigma = \sum_{i=1}^m A_i$$

Если $G(U, V)$ — информационный граф, а A — его матрица смежности, то элемент a_{ij} матрицы, полученный возведением матрицы A в степень n , равен числу различных путей длины n , идущих от γ_i к γ_j , т.е. возводя матрицу в степень n по тех пор, когда $A^n = 0$, можно выявить:

— количество и перечень задач, решаемых независимо;

Таблица 2 — Сравнительная таблица обработки триангулированной матрицы без разбиения на блоки и при четырехблочном разбиении

Степень матрицы	Размерность матрицы	Число операций умножения	Число операций суммирования
С разбиением матрицы на блоки			
2	$N/2$	$4(N/2)^2$	$4(N/2 - 1)(N/2)^2 + 1$
3	$N/2$	$8(N/2)^2$	$5 * 2(N/2 - 1)(N/2)^2 + 1$
4	$N/2$	$12(N/2)^2$	$6 * 3(N/2 - 1)(N/2)^2 + 1$
...
n	$N/2$	$4(n - 1)(N/2)^2$	$(n^2/3 + 2n^2 + 5n/2)(N/2)^2(N/2 - 1)$
Без разбиения матрицы на блоки			
2	N	N^2	$(N - 1)N^2$
3	N	$2N^2$	$2(N - 1)N^2$
4	N	$3N^2$	$3(N - 1)N^2$
...
n	N	$(n - 1)N^2$	$(n - 1)(N - 1)N^2$

- схему информационных связей между выделенными в модели элементами;
- число разновидностей исходной, промежуточной и результирующей информации; частоту использования различных видов информации;
- перечень задач, решаемых с использованием промежуточных результатов;
- установить последовательность решения задач и их связь с различными данными;
- определить объем информации, циркулирующей в системе;
- установить последовательность подготовки, ввода и использования в системе различных данных для подготовки выходных документов или решения определенных задач.

Равенство нулю суммы элементов j -го столбца матрицы смежности A служит признаком для формального выделения исходных данных, а значение $\sum a_{ij} > 0$ равно числу компонент, входящих в Y_j . Если просуммировать все матрицы, то отличные от нуля компоненты j -го столбца матрицы указывают все компоненты, участвующие в формировании Y_j , а ненулевые элементы i -той строки матрицы $A\Sigma$ указывают все результаты, при формировании которых используется компонента Y_i .

Состав и структуру документов в управляющей системе можно представить в виде графа, называемым структурным графом документа. Если реквизитам документа x_1, x_2, \dots, x_n сопоставить вершины графа X_1, X_2, \dots, X_n и каждую пару вершин X_i и X_j соединить дугой, идущей от X_i к X_j в том и только в том случае, когда X_i является составной частью X_j , то полученный граф будет отражать структуру документа как взаимосвязь его реквизитов. При построении графа документа «Конструкторская спецификация» примем следующие обозначения:

- X_0 — номер конструкторской спецификации;
- X_1 — признак детали узла;
- X_2 — шифр резиновой смеси;
- X_3 — процентное содержание синтетического каучука;

- X_4 — марка корда;
- X_5 — марка пропитки;
- X_6 — калибр детали;
- X_7 — ширина детали;
- X_8 — длина детали;
- X_9 — наименование резиновой смеси;
- X_{10} — геометрические размеры детали;
- X_{11} — каркас изделия;
- X_{12} — резиновые прокладки и ленточки.

Структурный граф документа представляет собой граф типа дерева, висячими вершинами которого являются реквизиты.

Для определения объема информации изменим ориентацию всех дуг на противоположную (потоки информации на графе идут снизу вверх, т.е. нам необходимо собрать всю информацию на номер конструкторской спецификации). Каждой дуге графа $U_i = (X_i, X_j)$ присваивается некоторый параметр m_i , представляющий собой число вхождений информационной совокупности X_i в информационную совокупность X_j . Каждой висячей вершине графа-дерева присваивается некоторый параметр l_i , который представляет значность данного реквизита X_i по количеству алфавитно-цифровых символов. Общий объем информации V определяется как сумма объемов информации по каждому пути от висячей вершины до корня дерева X_0 :

$$V = \sum_{i=1}^k v_i \tag{6}$$

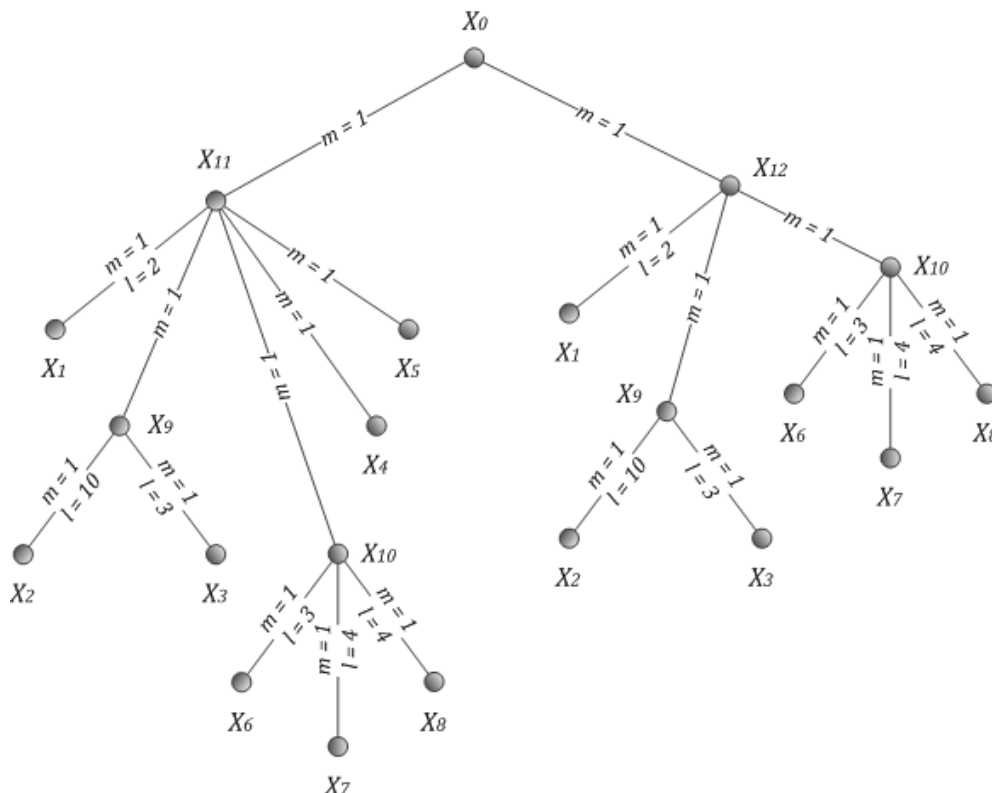


Рисунок 2 — Фрагмент графа взаимосвязи реквизитов документа «Конструкторская спецификация» (каркас автопокрышки)

где v_j — объем информации по пути от j -той вершины;
 k — число висячих вершин.

Объем информации по конкретному пути v_j определяется как произведение соответствующих параметров m_j и l_j , стоящих на этом пути. Пути при расчете могут определяться либо непосредственно по структурным графам документа, либо по соответствующей матрице смежности. Процесс определения путей по графу состоит в определении последовательности смежных дуг, идущих от висячих вершин к корню дерева.

Для определения объема информации используем фрагмент графа взаимосвязи реквизитов документа «Конструкторская спецификация» (рисунок 2).

Рассмотрим процесс определения объемов информации с использованием матрицы смежности. Пусть задана матрица A графа $G(X, Y)$, представляющего структурный граф документа с измененным направлением дуг. Прежде всего, определяются входы графа, соответствующие висячим вершинам графа $G(X, Y)$. Этим вершинам в матрице A будут соответствовать такие столбцы, сумма элементов строк в которых равна нулю. Выходу графа будет соответствовать строка, сумма элементов столбцов которой равна нулю. Таким образом, для определения входов и выходов графа необходимо подсчитать сумму элементов в каждой строке и каждом столбце матрицы смежности этого графа.

Переменная значения зафиксированных при поиске элементов матрицы, отличных от нуля, и значения параметра l_j , соответствующие данной начальной вершине, получаем объем информации по конкретному пути. Суммируя объем информации по всем путям, получим объем информации по документу.

Структурные графы документов и соответствующие им матрицы смежности могут использоваться при определении объема информации по задачам, подсистемам и системам в целом. В этом случае граф должен отражать структуры всех документов и содержать соответствующие вершины.

Заключение

Таким образом, авторами проанализированы методики обработки больших объемов информации. Предложено проводить анализ информационных потоков на стадии проектирования информационных систем с помощью информационного графа и его информационной матрицы, являющихся моделью этих потоков в системе.

Использование такого подхода дает возможность повысить наглядность функционирования системы управления и движения потоков информации, которая ускользает при пользовании другими методами, а также выявить ряд важных характеристик: количество и перечень задач, решаемых независимо; схему информационных связей между выделенными в модели элементами; число разновидностей исходной, промежуточной и результирующей информации; частоту использования различных видов информации; установить последовательность решения задач и их связь и различными данными; определить объем информации, циркулирующей в системе и ряд других. Это позволяет создавать эффективные (по памяти и скорости) алгоритмы обработки очень больших объемов данных.

На примере проектирования подсистемы технической подготовки производства КИС ОАО «Белшина» описан алгоритм обработки триангулированной матрицы, что особенно эффективно при ее больших размерах и ограничениях на объем памяти и быстродействие.

Литература / References

1. Найдич, А.И. Big Data: проблема, технология, рынок / А.И. Найдич // Компьютер Пресс, 2012. — № 9. — 102 с.
 Naydich, A.I. Big Data: problema, tekhnologiya, rynok // Kompyuter Press, 2012. — № 9. — 102 s.
2. Денисов, А.А. Современные проблемы системного анализа: Информационные основы: учебное пособие / А.А. Денисов. — СПб.: Изд-во СПбГТУ, 2005. — 295 с.
 Denisov, A.A. Sovremennye problem sistemnogo analiza: Informatsionnye osnovy: uchebnoye posobiye / A.A. Denisov. — SPb.: Izd-vo SPbGTU, 2005. — 295 s.
3. Свиридов, А.С. Методика проведения предпроектного обследования с целью проектирования информационной сети предприятия / А.С. Свиридов. — М.: Телекоммуникации, 2004. Вып. 4. С. 27—30
 Sviridov, A.S. Metodika provedeniya predproyektного ob sledovaniya s tselyu proyektirovaniya informatsionnoy seti predpriyatiya / A.S. Sviridov. — М.: Telekommunikatsii, 2004. — Вып. 4. — С. 27—30.
4. Трофимов, С.А. CASE-технологии: практическая работа в Rational Rose / С.А. Трофимов. — М.: «Издательство БИНОМ», 2001. — 272 с.
 Trofimov, S.A. CASE-tekhnologii: prakticheskaya rabota v Rational Rose / S.A. Trofimov. — М.: "Izdatelstvo BINOM", 2001. — 272 s.
5. Губин, М.В. Модели и методы представления текстового документа в системах информационного поиска. Диссертация на соискание ученой степени кандидата физико-математических наук / М.В. Губин. — Санкт-Петербург: 2005. — 89 с.
 Gubin, M.V. Modeli i metody predstavleniya tekstovogo dokumenta v sistemakh informatsionnogo poiska. Dissertatsiya na soiskaniye uchenoy stepeni kandidata fiziko-matematicheskikh nauk / M.V. Gubin. — Sankt-Peterburg: 2005. — 89 s.
6. Барсегян, А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, Olap / 2-е изд., перераб. СД / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. — СПб: BHV-Санкт-Петербург, 2008. — 384 с.
 Barsegyan, A.A. Tekhnologii analiza dannykh: Data Mining, Visual Mining, Text Mining, Olap / 2-e izd., pererab. SD / A.A. Barsegyan, M.S. Kupriyanov, V.V. Stepanenko, I.I. Kholod. — SPb: BHV-Sankt-Peterburg, 2008. — 384 s.
7. Аникин, В.М. Аналитические модели детерминированного хаоса / В.М. Аникин. — М.: Физматлит, 2007. — 328 с.
 Anikin, V.M. Analiticheskiye modeli determinirovannogo khaosa / V.M. Anikin. — М.: Fizmatlit, 2007. — 328 s.
8. Брайчевский, С.М. Современные информационные потоки: актуальная проблематика / С.М. Брайчевский, Д.В. Ландэ // Научно-техническая информация. — Сер. 1. — 2005. — Вып. 11. — С. 21—33.
 Braychevskiy, S.M. Sovremennye informatsionnye poroki: aktualnaya problematika / S.M. Braychevskaya, D.V. Lande // Nauchno-tekhnicheskaya informatsiya. — Ser. 1. — 2005. — Вып. 11. — С. 21—33.
9. Горбатов, В.А. Фундаментальные основы дискретной математики. Информационная математика / В.А. Горбатов. — М.: Высш. шк., 2000. — 544 с.
 GorbatoV, V.A. Fundamentalnye osnovy diskretnoy matematiki. Informatsionnaya matematika / V.A. GorbatoV. — М.: Vyssh. shk., 2000. — 544 s.
10. Ландэ, Д.В. Поиск знаний в Internet. Профессиональная работа: Пер. с англ. / Д.В. Ландэ. — М.: Издательский дом «Вильямс», 2005. — 272 с.
 Lande, D.V. Poisk znaniy v Internet. Professionalnaya rabota: Per. s angl. / D.V. Lande. — М.: Izdatelskiy dom "Vilyams", 2005. — 272 s.
11. Самойленко, А.П. Data Mining: Учебный курс. CD / А.П. Самойленко, В.А. Дюк. — СПб: Питер, 2001. — 368 с.
 Samoylenko, A.P. Data Mining: Uchebnykurs. CD / A.P. Samoylenko, V.A. Dyuk. — SPb: Piter, 2001. — 368 s.
12. Кузьминов, В.В. Разработка и исследование алгоритмов управления технической подготовкой производства в условиях возмущающих воздействий (на примере шинной промышленности): автореферат диссертации на соискание ученой степени кандидата технических наук / В.В. Кузьминов. — Киев: 1990. — 17 с.
 Kuzminov, V.V. Razrabotka i issledovaniye algoritmov upravleniya tekhnicheskoy podgotovkoy proizvodstva v usloviyakh vozmushchayushchikh vozdeystviy (na primere shinnoy promyshlennosti): avtoreferat dissertatsii na soiskaniye uchenoy stepeni kandidata tekhnicheskikh nauk / V.V. Kuzminov. — Kiev: 1990. — 17 s.
13. Донец, Г.А. Алгебраический подход к проблеме раскраски плоских графов / Г.А. Донец, И.З. Шор. — Киев: Наукова думка, 1982. — 144 с.
 Donets, G.A. Algebraicheskij podkhod k problem raskraski ploskikh grafov / G.A. Donets, I.Z. Shor. — Kiev: Naukova dumka, 1982. — 144 s.
14. Столярский, Г.Е. Алгоритм и программные средства эффективной организации информационных структур в развивающихся АСУ: автореферат диссертации на соискание ученой степени кандидата технических наук / Г.Е. Столярский. — Киев: 1986. — 23 с.
 Stolyarskiy, G.E. Algoritm i programnye sredstva effektivnoy organizatsii informatsionnykh struktur v razvivayushchikhsya ASU: avtoreferat dissertatsii na soiskaniye uchenoy stepeni kandidata tekhnicheskikh nauk / G.E. Stolyarskiy. — Kiev: 1986. — 23 s.