

**На правах рукописи**

**ТИМОФЕЕВ Кирилл Александрович**

**СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ  
ВЗАИМОСВЯЗАННЫХ ДОКУМЕНТОВ И ПРОЦЕССОВ  
НА ОСНОВЕ СТРУКТУРНО-ЛОГИЧЕСКИХ МОДЕЛЕЙ  
ИХ ОТНОШЕНИЙ**

**Специальность 05.13.10 – Управление в социальных  
и экономических системах**



**А В Т О Р Е Ф Е Р А Т**  
**диссертации на соискание ученой степени**  
**кандидата технических наук**

**Уфа 2007**

Работа выполнена на кафедре автоматизированных систем управления  
Уфимского государственного авиационного технического университета

Научный руководитель      д-р техн наук, проф  
**КУЛИКОВ Геннадий Григорьевич**

Официальные оппоненты      д-р. техн наук, проф  
**КАБАЛЬНОВ Юрий Степанович**  
  
канд техн. наук  
**ИБАТУЛЛИНА София Мухамедовна**

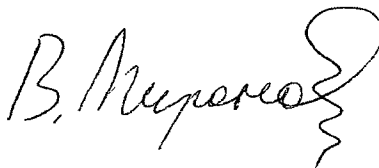
Ведущая организация      Республиканский научно-технологический и  
информационный комплекс «Баштехинформ»  
Академии наук Республики Башкортостан

Защита состоится 12 ноября 2007 года в 14 00 часов  
на заседании диссертационного совета Д212.288.03  
при Уфимском государственном авиационном техническом университете  
по адресу: 450000, Уфа-центр, ул К Маркса, 12

С диссертацией можно ознакомиться в библиотеке университета

Автореферат разослан «10» Октября 2007 г

Ученый секретарь  
диссертационного совета  
д-р техн. наук, проф



**В.В. Миронов**

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** Управление организацией в современных условиях требует от менеджмента способности анализировать возникающие варианты развития событий, находить, взвешивать и обоснованно принимать наилучшее решение. Сложность и ответственность решаемых задач по управлению с одной стороны и возможности информационных технологий с другой стороны стимулируют развитие автоматизированных систем поддержки принятия решений. Главной целью создания систем поддержки принятия решений является повышение эффективности использования возможностей объекта управления, достигаемое через

- повышение степени обоснованности и эффективности принимаемых решений,
- повышение оперативности управления,
- снижение трудозатрат на менеджмент при выполнении вспомогательных процессов,
- упрощение и ускорение труда работников организации

Известно, что одним из основных способов существенного улучшения основных показателей предприятия, повышения его конкурентоспособности и снижения издержек в последние годы является процессный подход к организации управления. Процессы, составляющие иерархические структуры управления, становятся предметом теоретических и практических исследований. Сегодня недостаточно лишь грамотно описать процессы, необходимо еще управлять ими в самой системе поддержки процессного управления. Общей чертой деятельности практически для всех организаций является также то, что почти все виды работ и операций документируются. Документ при этом является основанием для принятия управленческих решений. Учитывая большое количество видов документов и важные функции, выполняемые документом в процессе управления организацией, необходимо максимально эффективно наладить работу с документами.

Сфера деятельности крупной организации может охватывать несколько десятков направлений, каждое из которых состоит из процессов и подпроцессов и их общее количество может достигать не одной тысячи, а число видов документов, используемых при их выполнении, достигать нескольких сотен. Например, специфика банковской деятельности в области документооборота в части процессов заключается в необходимости скорейшего рассмотрения всех поступающих документов в установленные регламентами сроки и подготовки ответов на них либо выполнения необходимых действий в предельно короткие промежутки времени (до нескольких минут).

При совместном использовании системы автоматизации делопроизводства и документооборота и системы поддержки процессного управления возникает необходимость организации взаимодействия между этими системами, т.е. отнесения создаваемых и получаемых документов к технологическим процессам, т.к. в

соответствии с принципами процессного управления исполнение процесса не продолжится, пока в контрольной точке не будет зарегистрирован необходимый документ В программном аспекте это приводит к необходимости создания отдельной подсистемы, обеспечивающей связывание документов и технологических процессов

В этих условиях актуальна задача разработки метода определения взаимосвязи документов и бизнес-процессов, т.е. их взаимной классификации, привязки каждого документа к соответствующему процессу Предлагаемый метод должен позволять автоматизировать процесс прикрепления документа из системы автоматизации документооборота к соответствующему выполнению технологического процесса в системе поддержки процессного управления Исследованию различных аспектов классификации текстовых данных посвящены работы как отечественных, так и зарубежных авторов В УГАТУ подобные исследования проводятся профессорами Ю С Кабальновым, Л Р Черняховской, доцентом В Б Ивановым

**Цель работы** Целью работы является разработка метода определения взаимосвязей между документами и технологическими процессами и определения правил их взаимодействия в информационной системе поддержки процессного управления в единой информационной среде

**Задачи исследования.** Для достижения поставленной цели в работе сформулированы и решены следующие задачи

- 1) Разработать структурно-логические модели взаимосвязей документов и технологических процессов, основанные на имеющейся статистике их отношений,
- 2) Разработать алгоритм формирования репозитория структурно-логических моделей взаимодействующих документов и технологических процессов,
- 3) Разработать алгоритм определения взаимосвязей документов и технологических процессов на основе разработанного репозитория,
- 4) Разработать программное обеспечение для информационной системы поддержки процессного управления банковской деятельностью, автоматизирующее классификацию документов и технологических процессов на основе структурно-логических моделей их отношений
- 5) Произвести количественную оценку качества классификации, производимой с использованием программного обеспечения, реализующего предложенный метод

**Методика исследования** Выполненные в работе исследования базируются на основных положениях системного подхода и методах автоматизированного проектирования информационных систем При анализе функций системы поддержки процессного управления, методов классификации документов и при описании предлагаемых моделей использованы методы системного анализа, в качестве средства моделирования применена методология структурного анализа SADT При разработке метода классификации документов и анализе свойств предлагаемой системы использованы методы общей теории множеств и теории вероятности

### **Основные результаты, выносимые на защиту**

- структурно-логическая модель взаимосвязей документов и технологических процессов на основе статистики их соотношения,
- алгоритм формирования базы структурно-логических моделей в форме репозитория в гетерогенной среде хранилищ данных,
- алгоритм классификации отношений документов и технологических процессов в информационной системе поддержки процессного управления

### **Научная новизна работы заключается в**

- 1) Формализованной структурно-логической модели взаимосвязей документов и технологических процессов, позволяющей производить анализ их отношений на основе пересечения подмножеств ключевых, семантических и количественных атрибутов, характеризующих документы и технологические процессы
- 2) Схеме и соответствующих алгоритмов организации взаимосвязи хранящихся в автономных базах данных документов и технологических процессов
- 3) Алгоритме определения взаимосвязей документов и технологических процессов в контуре процессного управления, основанном на разработанных моделях

### **Практическая значимость.** Практическую значимость представляют

1. Метод определения взаимосвязей документов и технологических процессов, доведенный до инженерной методики, позволяющий автоматически выполнять связывание документов и соответствующих им технологических процессов.

2. Программное обеспечение для системы поддержки процессного управления, автоматизирующее предложенный метод определения взаимосвязи документов и технологических процессов с использованием разработанных алгоритмов и обеспечивающее повышение производительности труда, в том числе за счет уменьшения затрат на осуществление операций по соотношению документов в системе поддержки процессного управления и системе автоматизации делопроизводства и документооборота

**Внедрение результатов** Основные результаты диссертационной работы использованы при разработке системы поддержки процессного управления в Национальном банке Республики Башкортостан Банка России и в учебном процессе Уфимского государственного авиационного технического университета

**Апробация работы** Основные положения и результаты работы докладывались на следующих научно-технических конференциях межрегиональной научно-практической конференции "Повышение функциональной роли банковской системы через улучшение качества ее деятельности Управление бизнес-процессами в Банке России и кредитных организациях" (Уфа, 17–18 января 2005 г), «Развитие банковского дела в Омском регионе история и современность» (Омск, 2005), "Повышение функциональной роли банковской системы через улучшение качества ее деятельности Управление бизнес-процессами в Банке России и кредитных организациях" (Уфа, 17–18 января 2006 г), 8-й

Международной конференции по проблемам информатики и информационных технологий CSIT'2006, Карлсруэ (Германия), ноябрь 2006

**Публикации** Основные материалы диссертационной работы опубликованы в 8 научных трудах

**Структура работы.** Диссертационная работа состоит из 128 страниц машинописного текста, включающего в себя введение, четыре главы, заключение, приложения, рисунки, таблицы и список литературы из 107 наименований.

Автор выражает благодарность к.э.н Хуснутдинову А.А. за консультации при написании работы

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** дано описание объекта исследования, обосновывается актуальность темы диссертационной работы, сформулированы цель исследования, направления и методы ее достижения, описаны научная новизна и практическая значимость полученных результатов

**В первой главе** дается обзор принципов ведения документооборота и процессного управления. На рис. 1 приведена общая схема использования документов и описаний технологических процессов в рамках эксплуатации системы поддержки принятия решения в условиях процессного управления организацией.

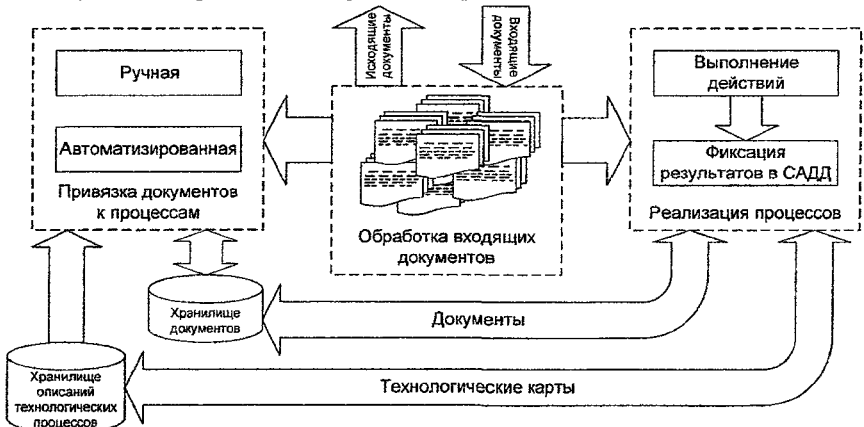


Рисунок 1 Общая схема взаимосвязи документов и технологических процессов

Рассматриваются существующие методы выделения схожих объектов среди имеющихся наборов исходных данных, приведен обзор методов классификации текстовых документов, используемых при этом, проводится сравнительный анализ различных методов классификации. Принимается во внимание тот факт, что для организационно-распорядительных документов, содержащих текстовую информацию,

характерно отсутствие четкой структуры и формальных признаков, по которым можно было бы однозначным образом определить вид документа. Из этого делается вывод о необходимости использования текста документа как набора слов и лексических единиц, из которых он состоит и которые характеризуют процесс, для проведения определения взаимосвязей документов и технологических процессов.

Задача определения взаимосвязи документов и технологических процессов состоит в классификации объектов (документов) по типам объектов другого рода (процессам), информация о которых содержится в трудно формализуемых атрибутах – текстах документов и описаниях процессов. Сложность задачи определения взаимосвязей документов и технологических процессов состоит в том, что должен производиться анализ и связывание разнородных объектов – документов и технологических процессов, существующие же методы классификации соотносят однородные объекты. Другими словами, необходимо отличить документ и определить соответствующий ему технологический процесс в условиях отсутствия в описании процесса информации, пригодной для связывания с документом. В связи с вышеизложенным представляется необходимой разработка нового метода определения взаимосвязей документов и технологических процессов, иными словами – классификация отношений между документами и процессами. Решение задачи осложняется большой размерностью структуры системы документооборота и процессного управления. Деятельность организации может быть представлена тысячами процессов и сотнями видов документов, используемых при их реализации.

**Во второй главе** описывается метод определения взаимосвязей документов и технологических процессов. Поскольку в документе достаточно информации для отнесения его человеком к нужному процессу, существует теоретическая возможность создания системы автоматического соотнесения документов и технологических процессов. Для реализации этой возможности предложено создать модель, описывающую частотные характеристики слов, составляющих документы, учитывающую их соотнесение с технологическими процессами.

В работе используются следующие основные понятия.

*Документ* – совокупность связанных лексических единиц, состоит из формализованной части – заголовка, и неформализованной – тела документа. *Термин* – лексическая единица, состоящая из букв и цифр, ограниченная с обеих сторон символом «пробел», «табуляция», «перевод строки» или «граница абзаца» или перечисленными символами слева и знаком препинания справа. Если к термину с любой стороны присоединен знак препинания, при выделении терминов знаки препинания удаляются. *Технологическая карта* – формализованное представление технологического процесса в виде набора атрибутов. *Реализация процесса* – исполнение технологического процесса, заданного описанием в виде технологической карты и имеющее наступившее иницирующее событие.

В работе документ  $D$  рассматривается как набор значащих терминов  $T$ , его составляющих, получаемый после отбрасывания шумовых терминов по словарю

слов-слов  $V$  Принадлежность термина к шумовым определяется по его вхождению в словарь шумовых терминов, состоящий из форм обращения, приветствий, предлогов и т.д. Таким образом  $D = T \setminus V$ , где  $D$  – документ,  $T$  – термины, составляющие документ,  $V$  – словарь шумовых терминов. Каждый термин  $T, i \in T, i \in T$  имеет свойства, характеризующие его значение в документе: это зона документа  $Z$ , к которой относится термин, и количество вхождений этого термина в каждую зону  $K$ . Документ как множество терминов можно описать следующим образом  $D = T(z, k) \setminus V$ . Множество документов образует базу документов  $B = \{D_i\}, i=1 \dots |B|$ .

Технологические процессы представляют собой описание выполняемых операций, для которых указаны сведения о документах, используемых при их исполнении, время начала действия описания процесса и время окончания действия описания. Так, можно записать  $P = \langle O_p, D_o, R_D, t_n, t_k \rangle$ , где  $P$  – технологический процесс,  $O_p = \{O_j\}$  – множество операций  $p$ -го процесса,  $D_o = \{D_j\}$  – документ,  $R_D = \{R_j\}$  – роль документа  $D$  в операции,  $t_n, t_k$  – дата начала и окончания действия данного описания технологического процесса. В операции процесса документ может играть следующие роли: документ инициирует выполнение процесса, инициировать выполнение операции, использоваться в результате выполнения процесса, завершать выполнение операции и завершать выполнение процесса.

Связь документа и технологического процесса определяет, с одной стороны, к каким процессам относится документ, и с другой – какие документы используются в процессе. Так, можно записать

$$L_{ij} = \langle D_i, P_j, W_{ij} \rangle, \quad (1)$$

где  $L_{ij}$  – связь  $i$ -го документа и  $j$ -го технологического процесса,  $D_i$  –  $i$ -ый документ из множества документов  $D$ ,  $P_j$  –  $j$ -ый технологический процесс из базы описаний технологических процессов  $P$ ,  $W_{ij}$  – множество весов терминов, составляющих документ  $D_i$  в процессе  $P_j$ . Определение весов терминов описано ниже.

Таким образом, структурно-логическая модель взаимосвязи документов и технологических процессов описывается следующим кортежем

$$M_j = \langle \langle T(z, k) \setminus V \rangle, \langle O_p, D_o, R_D, t_n, t_k \rangle, W_{ij} \rangle \quad (2)$$

Совокупность моделей взаимосвязей документов и технологических процессов составляет репозиторий (базу моделей документов процессов)  $M' = \{M_i\}, i=1 \dots |M|$ .

Формирование базы прецедентов, состоящей из моделей документов и процессов в их взаимосвязи, производится по следующему алгоритму

1 Из хранилища системы поддержки процессного управления отбирается множество всех технологических процессов, когда-либо выполнявшихся и имеющих статус завершенных. Завершенность исполнения процесса необходима для обеспечения качества обучения, т.к. по незавершенному процессу документ может быть отнесен во время его исполнения к другому процессу, что негативно скажется на модели документов процессов, если включить этот документ в обработку.

$$E = \{E_i\}, i=1 \dots K_p \quad (3)$$

где  $E$  – множество реализаций процессов,  $E_i$  – реализация  $i$ -го процесса,  $K_p$  – количество описаний процессов.



2 Для каждого элемента полученного множества реализаций технологических процессов  $E$  определяются связанные документы в разрезе ролей, т.е. формируется множество документов, используемых в реализациях процессов

$$D' = \langle D_b, R_b, t_b, t_c \rangle, i = I \ |M| \quad (4)$$

где  $D'$  – множество документов, относящихся к реализациям всех процессов,  $D_i$  – множество документов  $i$ -го технологического процесса,  $R_i$  – роль документа в запусках  $i$ -го процесса

3 Из полученного множества документов формируется множество терминов, составляющих эти документы. При выделении терминов производится их фильтрация в соответствии со следующими правилами

- шумовые термины отбрасываются в соответствии с принадлежностью термина к словарю шумовых терминов, формируемым экспертами до начала формирования моделей документов процессов,

- все термины приводятся к нижнему регистру,

- термины, встречающиеся в разных зонах документа, учитываются по зонам

4 Вычисляется вес термина  $W_i$  в документах процесса как отношение количества его вхождений в документы процесса  $K_p$  к количеству документов данного процесса  $K_D$   $W_i = K_p / K_D$  (5)

В результате обработки всех документов, имеющих связи с выполнявшимися технологическими процессами, из выделенных терминов и их весов формируется матрица взаимосвязей процессов и документов (т.н. матрица процесс-термин) размерностью  $N \times M$ , по строкам которой расположены все термины, выделенные в документах процессов, а по столбцам – сами процессы, где  $N$  – количество технологических процессов, когда-либо запускавшихся и завершённых,  $M$  – количество уникальных терминов, выделенных в этих процессах. Таблица 1 иллюстрирует пример матрицы процесс-термин

Таблица 1 Матрица процесс-термин

№	Процесс 1	Процесс 2	Процесс 3	..	Процесс N
Термин 1	0,91	0,233	0,523		0,324
Термин 2	0,835	0,667	0,445		0,233
Термин 3	0,322	0,923	0,645		0,176
Термин M	0,697	0,433	0,921		0,46

Определение технологического процесса, к которому относится обрабатываемый документ, предложено производить по алгоритму, основанному на использовании полученной базы моделей взаимосвязей документов и технологических процессов и приведенному на рис 2

Алгоритм заключается в следующем.

- производится формирование множества терминов рассматриваемого документа  $T_b$ . При этом различаются зоны документа, также как и при формировании моделей взаимосвязей документов и процессов, и используются те

же правила выделения и отсеивания терминов, что и на этапе построения моделей документов процессов, описанные выше,

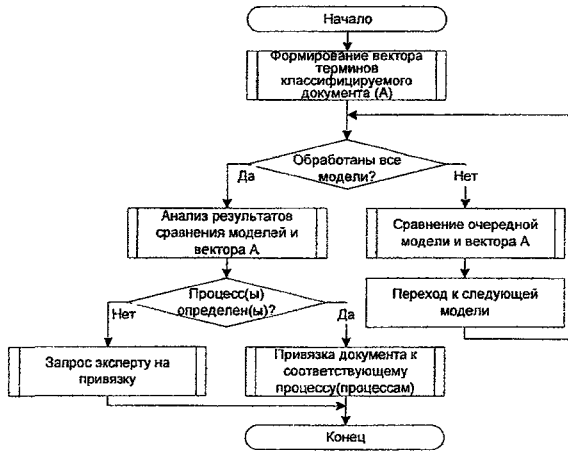


Рисунок 2 Блок-схема алгоритма классификации

– из базы прецедентов последовательно выбираются все модели  $M_i \in M$  и по описанному ниже алгоритму определяется степень их корреляции с полученным на предыдущем шаге набором терминов  $T_0$ . Результатом сравнения является множество вычисленных степеней корреляции документа и моделей,

– производится анализ результатов сравнения набора терминов связываемого документа  $T_0$  со всеми моделями и принятие решения об отнесении документа к соответствующему процессу (нескольким процессам) либо о необходимости произведения классификации экспертом,

– после осуществления привязки документа к какому-либо технологическому процессу номер документа и номер соответствующего ему процессу заносятся в список классифицированных документов, ожидающих подтверждения корректности классификации. Подтверждение корректности классификации заключается в завершении того запуска процесса, к которому был прикреплен классифицируемый документ. В случае завершения выполнения процесса без изменения привязки документа, т.е. подтверждения корректности произведенной классификации, документ считается правильно классифицированным и удаляется из списка документов, ожидающих подтверждения классификации, в отличие от случая неподтверждения корректности, в этом случае необходимо скорректировать те модели, которые позволили сделать вывод о неправильной принадлежности документа к процессу, и те, к которым документ на самом деле должен относиться. Для этого реквизиты документа заносятся в список документов, ожидающих обработки для коррекции моделей.

Определение степени корреляции набора терминов обрабатываемого документа  $T_d$  и модели  $M_i$  осуществляется по следующему алгоритму Производится последовательный перебор терминов связываемого документа  $T_i \in T_d$  и вычисление количества их совпадений с терминами модели Данное значение складывается из количества совпадений терминов из наименования регистрационной карточки документа в системе документооборота, отражающего его наименование, с полем технологической карты, отражающим наименование документа, относящегося к процессу и количества совпадений терминов, выделенных из тела классифицируемого документа, с соответствующими терминами моделей документов процессов В обоих случаях учитывается вес термина в модели, а для различия значимости зон документа введены коэффициенты значимости зон документа  $K_{z_i}$ , их значения получаются экспериментальным путем и основаны на том, что название документа в РК заполняется экспертом исходя из накопленных знаний и опыта работы, соответственно наличие документа с таким названием с большей достоверностью должно являться основанием для отнесения документа к процессу, чем его содержимое

Результат  $R_i$  оценки взаимосвязи документа и процесса рассчитывается следующим образом

$$R_i = K_{z1} \sum_{j=1}^L K_j W_j + K_{z2} \sum_{j=1}^L K_j W_j, \quad (6)$$

где  $K_{z1}$  и  $K_{z2}$  – коэффициенты значимости зон документа,  $L$  – количество совпавших терминов для каждой зоны,  $K_j$  – количество совпадений  $j$ -го термина,  $W_j$  – его вес в  $i$ -ой модели После расчета всех  $W_i$  матрица нормируется

$$W_i = \frac{W_i}{\max(W_i)} \quad (7)$$

Полученные значения оценки степени принадлежности документа к процессам можно представить графиком, приведенным на рис 3

Возможны следующие варианты распределения оценок степени принадлежности документов к процессам

- На графике есть одна точка максимума, однозначно определяющая принадлежность документа процессу – ее значение на 20% больше значений остальных точек Граница однозначности может быть изменена по результатам апробации метода

- Некоторое количество точек образует группу наиболее вероятных вариантов принадлежности, остальные точки на графике расположены ниже границы и не рассматриваются далее

- Среди оценок степени принадлежности документов к процессам нет ни однозначного максимума, ни группы точек, пригодных для дальнейшей обработки

Количественные параметры, такие как минимальное превосходство максимального значения над остальными, количество вариантов в группе, граница попадания значений в группу устанавливается экспериментальным путем



Рисунок 3 Пример распределения результатов поиска

Для определения эффективности разработанной системы определения взаимосвязей документов и технологических процессов осуществляется оценка показателей качества классификации. В некоторых публикациях для определения ошибки классификации текстовых документов предлагается подход, при котором исследуемая выборка по отношению к процессу  $P_i$  ( $i = 1 \dots N$ ) разбивается на четыре непересекающихся множества и представляется в виде таблицы частот сопряженности размером  $2 \times 2$  – таблица 2)

Таблица 2

	Документ принадлежит процессу $P_i$	Документ не принадлежит процессу $P_i$
Документ отнесен к процессу $P_i$	$a_i$	$b_i$
Документ не отнесен к процессу $P_i$	$c_i$	$d_i$

В ячейках таблицы расположены следующие элементы

- $a_i$  – количество документов, относящихся к процессу  $P_i$  и отнесенных в результате классификации к этому процессу,
- $b_i$  – количество документов, не относящихся к процессу  $P_i$ , но отнесенных в результате классификации к этому процессу – так называемая ошибка первого рода,
- $c_i$  – количество документов, относящихся к процессу  $P_i$ , но не отнесенных в результате классификации к этому процессу – так называемая ошибка второго рода,
- $d_i$  – количество документов, не относящихся к процессу  $P_i$  и не отнесенных в результате классификации к этому процессу

Данные величины рассчитываются после завершения исполнения процесса. Если процесс завершился без изменения привязки документа экспертом, то увеличивается значение  $a_i$ , в противном случае увеличивается значение  $b_i$ , или  $c_i$ , в зависимости от того, правильно ли был отнесен документ к процессу. Для всей совокупности процессов рассчитываются следующие интегральные величины  $A = \sum_{i=1}^N a_i$ ,  $B = \sum_{i=1}^N b_i$ ,  $C = \sum_{i=1}^N c_i$ ,  $D = \sum_{i=1}^N d_i$ . Вероятности этих гипотез несовместимы и образуют полную группу событий, т.е.  $P(A) + P(B) + P(C) + P(D) = 1$ . На основании

этих величин рассчитываются коэффициент точности классификации ( $a$ ) и ошибка классификации ( $e$ ) по следующим формулам

$$a = \frac{A}{A+C} 100\%, e = \frac{C}{A+C} 100\% \quad (8, 9)$$

В данной работе качество классификации оценивается коэффициентами точности и ошибки классификации по контрольной выборке. Оценка качества определения взаимосвязей документов и технологических процессов должна производиться при практическом использовании предложенного метода для выявления снижения адекватности репозитория и принятия соответствующих мер (повторного его формирования, изменения коэффициентов зон документов и т.д.)

В третьей главе описывается реализация предложенного метода в виде модуля системы поддержки процессного управления, функционирующего по схеме, приведенной на рис. 1. Входящий документ регистрируется в системе документооборота, проходит классификацию и соотносится с каким-либо технологическим процессом, который затем выполняется. По результатам реализации процесса производится корректировка моделей взаимосвязи документа и процесса. Реализация производилась на программно-технологической платформе Национального банка Республики Башкортостан Банка России.

В настоящее время в Банке России технологической основой системы электронного документооборота является система САДД «Дело» 8.6.1. Система поддержки процессного управления СОУД ТУ реализована на базе аналитического комплекса «Прогноз» - специализированного инструментария создания систем поддержки принятия решений и информационно-аналитических систем в банковской сфере, различных областях экономики, финансов и производства. В связи с открытостью системы СОУД ТУ, наличием в АК «Прогноз» мощного инструмента для создания приложений, наличием в САДД «Дело» API-интерфейса, в качестве платформы для разработки системы классификации выбрана среда АК «Прогноз».

Реализация предложенного метода произведена в виде двух подсистем – подсистемы обучения (формирования репозитория) и подсистемы классификации. Фрагмент предлагаемой информационной модели приведен на рис. 4. Фрагменты программ на языке FORE (язык АК «Прогноз»), хранимых процедур на языке PL/SQL и часть экранных форм приведены в диссертационной работе. При практической реализации метода использовано хранилище описаний технологических процессов системы СОУД ТУ и хранилище документов системы САДД «Дело». Взаимодействие этих систем с целью обеспечения информационной безопасности производится от имени специально созданного пользователя. В диссертационной работе приведены требования к серверам и рабочим станциям, используемым комплексами «Прогноз» и «Дело».

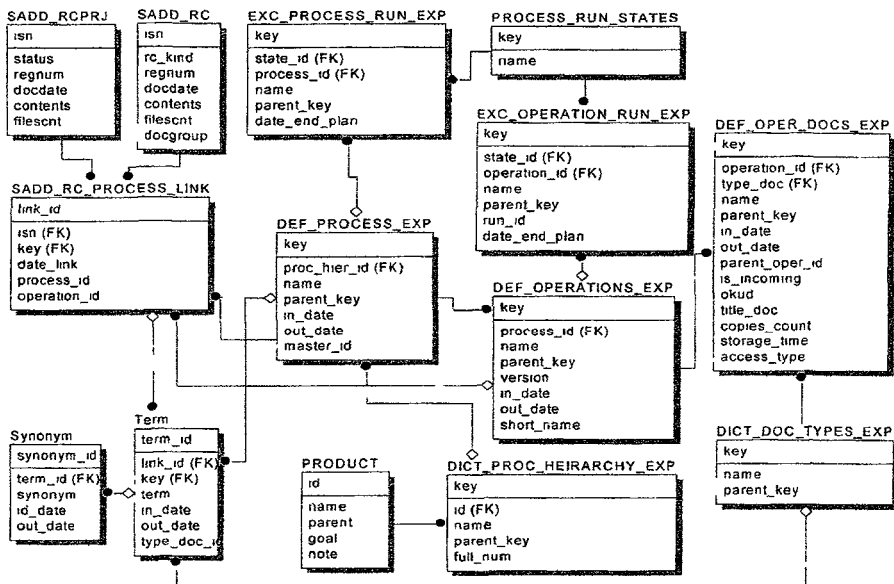


Рисунок 4 Фрагмент логической модели разработанного модуля

В четвертой главе приводится описание проведенного эксперимента, осуществленного на базе Национального банка Республики Башкортостан Банка России. Целью проведения эксперимента является исследование адекватности и корректности модели взаимосвязей документов и технологических процессов и алгоритмов ее построения и применения, а также проверка на практике предположения, лежащего в основе построения модели документа процесса, о том, что выделенная совокупность терминов и их статистические характеристики могут быть использованы для определения технологических процессов, к которым относится тот или иной документ.

Для получения наиболее достоверных данных и исследования влияния количества документов, на которых строятся их модели, на качество классификации, произведена серия испытаний, отличающихся наборами входных данных. Проведены следующие испытания:

1) Произведено формирование репозитория на полном объеме обучающей выборки по четырем выбранным технологическим процессам и тестовая классификация на полном объеме тестовых документов,

2) Произведено формирование репозитория на полном объеме обучающей выборки только по одному технологическому процессу и тестовая классификация на полном объеме тестовых документов,

3) Формирование репозитория производится в два этапа на разных половинах обучающей выборки, после каждого этапа производится тестовая классификация на полном объеме тестовых документов

Для проведения эксперимента отобрана группа из 4 технологических процессов, относящихся к направлению деятельности «Информатизация деятельности»

- предоставление сервисов информационно-телекоммуникационной системы (ИТС),
- поддержка сервисов ИТС;
- техническая эксплуатация оборудования и средств связи, вычислительной техники, организационной техники и средств механизации кассовой работы,
- ведение договоров

Выбор этих процессов обусловлен наличием в их исполнении большого количества типов документов это информационные сообщения, приказы, распоряжения, договоры. Проведение эксперимента на этих процессах позволит сделать выводы о применимости предложенных модели и алгоритмов для полного набора технологических процессов и их документов

В обучающую выборку включены документы, привязанные к процессам и созданные в течение предыдущего календарного года (т.е. с 01.01.2006 по 31.12.2006). По выбранным технологическим процессам за указанный период времени отобрано 389 документов, хранящихся в электронном виде в хранилище САДД, из которых всем условиям удовлетворяют 346 (по 43 документам имеются незавершенные запуски процессов). Таким образом, в ходе экспериментальных исследований должны быть обработаны 346 документов по 4 технологическим процессам, из числа которых часть документов должна быть выделена для построения моделей, а оставшиеся должны быть использованы для проверки системы классификации на основании полученных моделей.

Распределение документов на обучающую и тестовую выборки осуществлялось случайным образом, количество документов отражено в таблице 3

Таблица 3

Параметры Испытание	Количество процессов в обучающей выборке	Количество документов в обучающей выборке	Количество документов в тестовой выборке
1	4	173	173
2	1	47	299
3	2	173	173

Ниже приведены обобщенные результаты работы модуля при определении взаимосвязи документов и технологических процессов в рамках первой серии экспериментов. В ходе обработки документов обучающей выборки (привязанных к выбранным технологическим процессам) для формирования моделей документов процессов выделено 1184 значащих уникальных термина, среднее количество терминов в документе составляет 192. В среднем каждый выделенный термин встречается в ~80% обработанных документов. В данной матрице представлены

уникальные термины после произведения замены синонимов. В случае наличия синонимов они заменяются на термин с максимальным приоритетом из справочника синонимов.

После проведения классификации документов получены результаты, в графическом виде представленные на рис 5

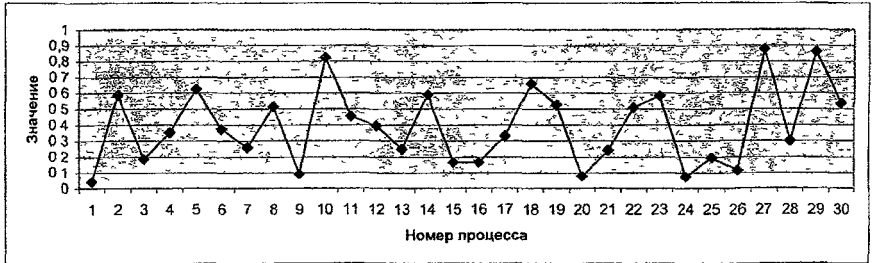


Рисунок 5 – Результаты классификации 1 документа

Из рисунка возможно сделать вывод о принадлежности классифицируемого документа к процессу с номером 27, а также, возможно, к процессам 10 и 29, что определяется анализом того, есть ли у этих трех процессов общие документы. Если есть – документ должен быть отнесен ко всем трем процессам, иначе – к тем, которые имеют общие документы (т.е. 27 и 10 или 27 и 29 или только 27).

Обобщенные результаты классификации документов тестовой выборки по первому виду испытаний приведены в таблице 4, результаты классификации документов трех групп приведены в таблице 5.

Таблица 4

термин \ процесс	Количество документов, относящихся к процессу	Количество документов, верно отнесенных системой к процессу	Количество документов, неверно отнесенных системой к процессу	Ошибка классификации
предоставление сервисов ИТС	49	46	3	6,1%
поддержка сервисов ИТС	96	90	6	6,2%
техническая эксплуатация оборудования и средств связи	19	17	2	10,5%
Ведение договоров	9	8	1	11,1%

Показатели качества классификации (коэффициент точности классификации и ошибка классификации), рассчитанные по приведенному в главе 2 алгоритму, для первого вида испытаний составили  $a = 93\%$ ,  $c = 7\%$

Таблица 5

Показатель качества \ Группа	Количество верно классифицированных документов	Количество неверно классифицированных документов	Количество неклассифицированных документов	Коэффициент точности классификации	ошибка классификации
1	46	3	0	93%	7%
2	32	15	0	68%	23%
3	44	10	2	91,3%	5,8%



Отсутствие в первых двух испытаниях неклассифицированных документов говорит о том, что наличие в классифицируемых документах терминов, присущих документам этого процесса, позволяет хоть и с малой точностью (всего 68% во втором случае), но произвести классификацию даже при малом объеме обучающей выборки. Проведенные испытания показали, что предлагаемая в данной работе модель взаимосвязей документов и технологических процессов, алгоритм ее формирования и метод классификации пригодны к использованию на практике для классификации документов по технологическим процессам. Проведенные 3 вида испытаний показали, что качество получаемой в результате обучения модели и точность классификации зависят от размеров обучающей выборки связанных документов и технологических процессов.

## **ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ**

Основные результаты работы состоят в следующем

1 Разработана формализованная структурно-логическая модель для анализа соответствия документов и технологических процессов, отличающаяся от существующих моделей включением в рассмотрение всех терминов, составляющих документы, и использованием формальных описаний технологических процессов

2 Разработан алгоритм формирования базы структурно-логических моделей документов и технологических процессов при их взаимодействии, позволяющий проанализировать имеющуюся статистику соотношения документов и описаний процессов и на ее основе сформировать модели, описывающие их отношения

3 Разработан метод определения взаимосвязей документов и технологических процессов, доведенный до инженерной методики, позволяющий автоматически выполнять связывание документов и соответствующих им технологических процессов

4 Разработано программное обеспечение для системы поддержки процессного управления, автоматизирующее предложенный метод определения взаимосвязей документов и технологических процессов с использованием разработанных алгоритмов и обеспечивающее повышение производительности труда, в том числе за счет уменьшения затрат времени на осуществление операций по соотношению документов в системе поддержки процессного управления и системе автоматизации делопроизводства и документооборота

5 Произведенный эксперимент подтвердил достоверность получаемых моделей и пригодность метода для практического использования. Количественная оценка качества автоматически получаемых взаимосвязей превышает 90%, что является приемлемым результатом

## ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### *Публикации в изданиях, рекомендуемых ВАК:*

1 О создании методики классификации документов по технологическим процессам / К А Тимофеев // Вестник УГАТУ Научный журнал Уфимского государственного авиационного технического университета, 2006 Т 8 № 2(18) С 55 – 57

### *Прочие публикации:*

2. Нейросетевая система принятия решения о кредитоспособности заемщика / К А Тимофеев // Повышение функциональной роли банковской системы через улучшение качества ее деятельности Управление бизнес-процессами в Банке России и кредитных организациях Сб научн тр Уфа Нефтегазовое дело, 2005 С. 172–175

3 Международные стандарты управления качеством и документацией / К А Тимофеев // Развитие банковского дела в Омском регионе история и современность матер межрег. науч -практич. конф – Омск, 2005 С 325–329

4 Управление организацией документооборот и процессные технологии / К А Тимофеев, А А. Хуснутдинов // Развитие банковского дела в Омском регионе история и современность матер. межрег науч -практ конф Омск, 2005 С 329–333

5 Технология Workflow в управлении ТУ Банка России / Г Г Куликов, К.А Тимофеев, А А Хуснутдинов // Повышение функциональной роли банковской системы через улучшение качества ее деятельности Управление бизнес-процессами в Банке России и кредитных организациях Сб научн тр М Наука, 2006 С. 172–175

6 Интеграция системы поддержки процессного управления с существующими документообразующими системами / К А Тимофеев // Автоматизированные системы обработки информации и управления . Сборник тр шк аспирантов Уфа УГАТУ, 2006 Т 1 С. 27–33

7 Оптимизация документооборота с использованием процессных технологий и нейросетевого классификатора документов по процессам / Г Г Куликов, К А Тимофеев, А А Хуснутдинов // Матер 8-й Междунар конф по проблемам информатики и информационных технологий CSIT'2006 Т 2 С.69–72 (статья на англ языке)

8 Подходы к созданию методики классификации документов по технологическим процессам в кредитных организациях и Банке России / К А Тимофеев // Рынок розничных банковских услуг в регионе проблемы и перспективы матер межрег науч -практ. конф Омск, 2006 С 267-271

Тимофеев Кирилл Александрович

СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ  
ВЗАИМОСВЯЗАННЫХ ДОКУМЕНТОВ И ПРОЦЕССОВ  
НА ОСНОВЕ СТРУКТУРНО-ЛОГИЧЕСКИХ МОДЕЛЕЙ  
ИХ ОТНОШЕНИЙ

Специальность 05 13 10 - Управление в социальных  
и экономических системах

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
кандидата технических наук

Подписано в печать 08 10 2007 Формат 60x84 1/16  
Бумага офсетная Печать плоская Гарнитура Times  
Усл печ л 1,0 Усл кр-отт 1,0 Уч –изд л 0,9  
Тираж 100 экз Заказ № 528.

ГОУ ВПО Уфимский государственный авиационный технический университет  
Центр оперативной полиграфии  
450000, Уфа-центр, ул. К.Маркса, 12