

Содержание:

ВВЕДЕНИЕ

Первое десятилетие XXI века отмечается ростом количества людей, использующих компьютеры, резким увеличением объема информации, получаемой через компьютерные сети и Интернет. Повсеместное применение компьютеров и их объединение в локальные, региональные и глобальные сети привело к активному развитию, распространению и повсеместному использованию информационно-коммуникационных технологий (ИКТ) и формированию информационного пространства. Информационное пространство – это инфокоммуникационная среда, гарантированно удовлетворяющая информационные и методические потребности ее субъектов, служб при подготовке, принятии и реализации решений в любой сфере деятельности.

В рамках данного исследования рассматриваются, к которым относятся: – это персона или группа, которая предоставляет информацию без требования сохранения ее конфиденциальности, – информация или отношения незащищенные от публичного раскрытия. Открытые источники относятся к среде общедоступной информации, и не имеют ограничения в доступе для физических лиц. Развитие информационных и коммуникационных технологий существенным образом изменили структуру и информационно-технические характеристики потоков информации. Анализ характера изменений объемов информации, циркулирующей через транснациональные инфо-коммуникационные сети показывает, что рост объема информации имеет экспоненциальную характеристику.

Под распределенной обработкой информации понимается комплекс операций с информацией (традиционно описываемый термином "обработка информации"), проводимый на независимых, но связанных между собой вычислительных машинах, предназначенных для выполнения общих задач.

Системы распределенной обработки информации (или распределенные вычислительные системы) в виде многомашинных вычислительных комплексов и компьютерных сетей представляют собой одну из наиболее прогрессивных форм организации средств вычислительной техники.

Появление и широкое распространение систем распределенной обработки информации обусловлено, с одной стороны, ускоренным развитием микроэлектроники, снижением стоимости вычислительных средств, увеличением их производительности при уменьшении габаритов, а с другой стороны повышением требований к производительности, надежности и эффективности вычислительных систем, предъявляемых сферами их применения.

Цель работы анализ управления распределенной информацией и его перспектив.

Объект исследования управление распределенной информацией.

Предмет исследования принципы управления распределенной информацией, характеристики распределенных систем баз данных, управление распределенной информацией, его перспективы, Однородные и неоднородные распределенные системы, методы построения распределенных баз данных, проблемы управления распределенной информацией и пути их решения.

Задачи исследования вытекают из поставленной цели:

проанализировать принципы управления распределенной информацией;

описать модели распределенных баз данных;

рассмотреть направления развития управления распределенной информацией.

Методы исследования: изучение и анализ научной литературы, статей по теме работы в периодических изданиях, научных источников в сети Интернет, анализ и сравнение модели распределенных баз данных с другими моделями.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ОБРАБОТКИ ИНФОРМАЦИИ

1.1. Понятие информации и ее обработки информации

Информатика — это основанная на использовании компьютерной техники дисциплина, изучающая структуру и общие свойства информации, а также

закономерности и методы её создания, хранения, поиска, преобразования, передачи и применения в различных сферах человеческой деятельности.

Обработка информации – получение одних информационных объектов из других информационных объектов путем выполнения некоторых алгоритмов [15].

Обработка является одной из основных операций, выполняемых над информацией, и главным средством увеличения объёма и разнообразия информации. Средства обработки информации — это всевозможные устройства и системы, созданные человечеством, и в первую очередь, компьютер — универсальная машина для обработки информации. Компьютеры обрабатывают информацию путем выполнения некоторых алгоритмов. Живые организмы и растения обрабатывают информацию с помощью своих органов и систем.

Понятие обработки информации является весьма широким. Ведя речь об обработке информации, следует дать понятие инварианта обработки. Обычно им является смысл сообщения (смысл информации, заключенной в сообщении). При автоматизированной обработке информации объектом обработки служит сообщение, и здесь важно провести обработку таким образом, чтобы инварианты преобразований сообщения соответствовали инвариантам преобразования информации.[10]

Цель обработки информации в целом определяется целью функционирования некоторой системы, с которой связан рассматриваемый информационный процесс. Однако для достижения цели всегда приходится решать ряд взаимосвязанных задач.

К примеру, начальная стадия информационного процесса – рецепция. В различных информационных системах рецепция выражается в таких конкретных процессах, как отбор информации (в системах научно-технической информации), преобразование физических величин в измерительный сигнал (в информационно-измерительных системах), раздражимость и ощущения (в биологических системах) и т.п.

Процесс рецепции начинается на границе, отделяющей информационную систему от внешнего мира. Здесь, на границе, сигнал внешнего мира преобразуется в форму, удобную для дальнейшей обработки. Для биологических систем и многих технических систем, например читающих автоматов, эта граница более или менее четко выражена. В остальных случаях она в значительной степени условна и даже расплывчата. Что касается внутренней границы процесса рецепции, то она

практически всегда условна и выбирается в каждом конкретном случае исходя из удобства исследования информационного процесса. [12]

Следует отметить, что независимо от того, как "глубоко" будет отодвинута внутренняя граница, рецепцию всегда можно рассматривать как процесс классификации.

Формализованная модель обработки информации

Обратимся теперь к вопросу о том, в чем сходство и различие процессов обработки информации, связанных с различными составляющими информационного процесса, используя при этом формализованную модель обработки. Прежде всего заметим, что нельзя отрывать этот вопрос от потребителя информации (адресата), от семантического и прагматического аспектов информации. Наличие адресата, для которого предназначено сообщение (сигнал), определяет отсутствие однозначного соответствия между сообщением и содержащейся в нем информацией.

Совершенно очевидно, что одно и то же сообщение может иметь различный смысл для разных адресатов и различное прагматическое значение.

Предположим, что с каждым конкретным потребителем информации связано некоторое множество I , элементами которого являются пары смысл-значение. Существует множество X сообщений, элементами которого могут быть символы, слова, фразы, значения физических величин и процессов – словом, любые знаки. Чтобы из сообщения X могла быть извлечена информация I , должно существовать некоторое отображение j

являющееся результатом действия по крайней мере трех факторов:

- 1) договоренности между отправителем и потребителем, что позволяет "осмысливать" сообщение;
- 2) наличием конкретной цели у адресата;
- 3) той ситуацией, в которой находится адресат.

Последние два фактора определяют значение сообщения. Отображение j называется правилом интерпретации сообщения. Оно может быть общим, понятным для многих потребителей информации, либо известным лишь паре отправитель—потребитель, а для других потребителей информации незнание правила j приводит к тому, что даже воспринятое сообщение не поддается интерпретации или ведет к ложной интерпретации.

1.2 Распределительная система управления

Распределенные системы управления (СУ) решают задачи автоматизации технологических процессов (ТП), когда эти процессы имеют распределенный территориальный и алгоритмический характер. Многие ТП, а следовательно и системы управления ими, характеризуются большим объемом сенсорной информации, поступающей от объектов управления, и быстрой реализацией алгоритма.

Такие ТП можно встретить в химической, машиностроительной и других отраслях, а так же научных комплексах. Автоматизация распределенных и быстропротекающих процессов строится на основе локальных систем управления (ЛСУ) функционально законченных частей распределенных ТП и телекоммуникационной системы, имеющей различную архитектурную реализацию (круговую, линейную и т.д.). [8]

Как и в любой СУ, время реакции распределенной СУ на управляющее воздействие является определяющим критерием при разработке систем данного типа. На рис.1 представлена упрощенная структура распределенной СУ, где НС набор сенсоров, СОД система обработки данных, ЛСУ локальная система управления, СИУ силовые исполнительные устройства. В СОД, на основе обработанных данных из НС формируются события и аларм-сообщения, которые поступают в ЛСУ данного узла и (или) передаются по сети Network в другие узлы.

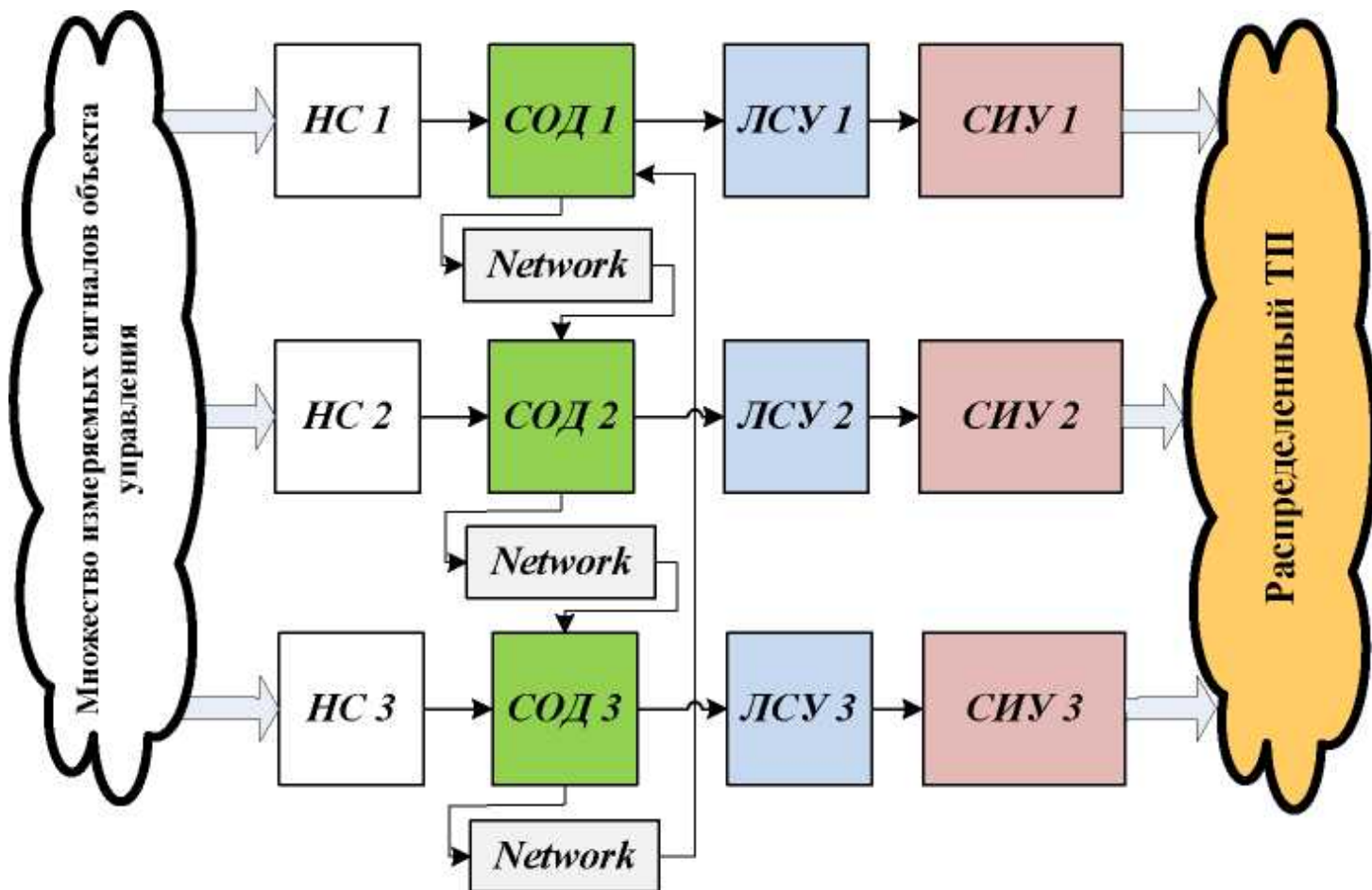


Рис. 1 Структура распределенной системы управления с кольцевой телекоммуникационной сетью.

Пусть время распространения информации $t_{propagation}$ от одной ЛСУ к другой определяется как:

$$t_{propagation} = t_{is} + t_{ip} + t_{fxd} \quad (1)$$

где t_{is} время обработки исходной информации (например, сигнала, связанного с температурным параметром объекта управления) интеллектуальными сенсорами, t_{ip} время обработки информации подсистемой сбора и обработки данных с интеллектуальных датчиков и подготовки для передачи по телекоммуникационной сети, t_{fxd} время передачи сообщений по телекоммуникационной сети.

Поскольку в уравнении (1) время t_{is} и t_{fxd} является практически константами, то можно оптимизировать $t_{propagation}$ только за счет уменьшения t_{ip} . Стоит отметить, что время t_{fxd} можно уменьшить за счет сокращения объема передаваемой информации используя методы сжатия данных без потерь. Вместе с тем, чем больше степень сжатия информации, тем больше на это тратится время

[1].

Одним из решений уменьшения времени при сжатии является аппаратная реализация алгоритма. В работе [2] представлена реализация алгоритма сжатия данных без потерь на базе Field-Programmable Gate Array (FPGA).

Для уменьшения времени используются скоростные промышленные интерфейсы типа Ethernet с оптической средой распространения сигнала, скорость которых может достигать десятков Gb/s. Другое направление уменьшения t_{зд} сжатие протокольных данных при передаче. В работе [3] приводится метод оптимизации UDP сетевой передачи путем использования аппаратной реализации на FPGA при приеме биржевых данных. Вместе с аппаратной обработкой пакетов на основе FAST протокола (так же на FPGA) данная оптимизация позволяет уменьшить время доставки и обработки пакетов почти в 4 раза [3].

Одной из важнейших сетевых технологий, является распределенная обработка данных. Персональные компьютеры стоят на рабочих местах, на местах возникновения и использования информации. Они соединены каналами связи. Это дало возможность распределить их ресурсы по отдельным функциональным сферам деятельности и изменить технологию обработки данных в направлении децентрализации. Распределенная обработка данных позволила повысить эффективность удовлетворения изменяющейся информационной потребности информационного работника и, тем самым, обеспечить гибкость принимаемых им решений.

Преимущества распределенной обработки данных:

большое число взаимодействующих пользователей, выполняющих функции сбора, регистрации, хранения, передачи и выдачи информации;

снятие пиковых нагрузок с централизованной базы путем распределения обработки и хранения локальных баз данных на разных ЭВМ;

обеспечение доступа информационному работнику к вычислительным ресурсам сети ЭВМ;

обеспечение симметричного обмена данными между удаленными пользователями.

Формализация концептуальной схемы данных повлекла за собой возможность классификации моделей представления данных на иерархические, сетевые и реляционные. Это отразилось в понятии архитектуры систем управления базами

данных и технологии обработки. Архитектура СУБД описывает ее функционирование как взаимодействие процессов двух типов: клиента и сервера.

1.3 Характеристика распределительной обработки информации

Важным направлением по уменьшению времени реакции распределенной СУ является метод опережающей обработки событий. Под событиями в системе управления понимается набор именованных сигналов, активное состояние которых влияет на выполнение алгоритма управления ТП. События могут быть внешними по отношению к конкретной ЛСУ (полученными из другой подсистемы), или внутренними, сформированными на основании сенсорной информации данной системы.

В распределенной СУ формирование и обработка событий может быть в ЛСУ1, а реализация алгоритма в ЛСУ3. Таким образом, между формированием события и реализацией алгоритма присутствует телекоммуникационная сеть с транспортной задержкой. Основная идея опережающей обработки событий заключается в том, что даже если появление очередного события нарушает временной порядок ряда событий, необходимо передать это событие в сеть, то есть, если сеть свободна ее необходимо загрузить. [1]

Примером такого подхода может служить предлагается в работе адаптивный метод опережающей обработки событий, поступающих от сенсоров, с использованием буферизации. Метод основан на алгоритме, при котором обрабатываются те события, которые пришли первыми, даже если они нарушают определенный порядок. Для восстановления необходимого порядка поступления событий в систему управления через сеть, введен буфер, в котором задерживаются события на определенное время. Такой метод существенно сокращает время обработки и дает возможность более равномерно загрузить телекоммуникационную сеть. Необходимо заметить, что в случае аппаратной реализации обработки информации с сенсоров на FPGA, обработку событий можно распараллелить.

Современным инструментом обработки потоков данных в реальном времени является технология data mining. Эта технология "извлечения" данных имеет хорошо отработанные алгоритмы и программные решения и применяется во

многих отраслях. Многие сетевые решения data mining, связанные с анализом данных, из-за больших скоростей, реализованы на FPGA. Обработка потока входных данных в СОД распределенных СУ выполняется в реальном времени и критична к значению переменной t_{ip} . Поэтому решения по обработке и анализу чаще всего выполняются на FPGA. Можно выделить несколько основных направлений data mining на платформе FPGA:

построение деревьев решений (decision trees) ;

поиск и анализ данных по шаблону (pattern) ;

сигнатурный и корреляционный анализ.

Для формирования событий в СОД распределенной СУ чаще всего используется анализ данных по шаблону [9].

Шаблон может представлять собой любую информацию, организованную в виде произвольной структуры данных. Интерес представляют многовариантные взаимоотношения этих данных, в результате анализа которых можно получить новые знания или сформулировать, например, прогноз поведения технологического процесса. Известны различные алгоритмы и способы сравнения в потоке данных. В работе предложен алгоритм построчного поиска и сравнения, а авторы работы реализуют этот алгоритм поиска методом построения конечного автомата, который ищет одновременно по нескольким шаблонам, совершая простые операций.

Рассмотрим механизм работы шаблона на FPGA. Пусть в качестве шаблона задано множество переменных $B = \{B_1, B_2, B_3, \dots, B_n\}$, где каждая переменная принимает двоичное значение равное значению выходного сигнала соответствующего датчика $S_1, S_2, S_3, \dots, S_n$ (рис.2). Каждая переменная из множества B может принимать произвольную разрядность двоичных чисел, ограниченную технологическими возможностями FPGA. Поток данных поступающий от сенсоров обозначим через множество $X = \{X_1, X_2, X_3, \dots, X_n\}$. Анализ потока данных на базе шаблона можно выполнить через сравнение соответствующих переменных множеств B и X на равенство.

B

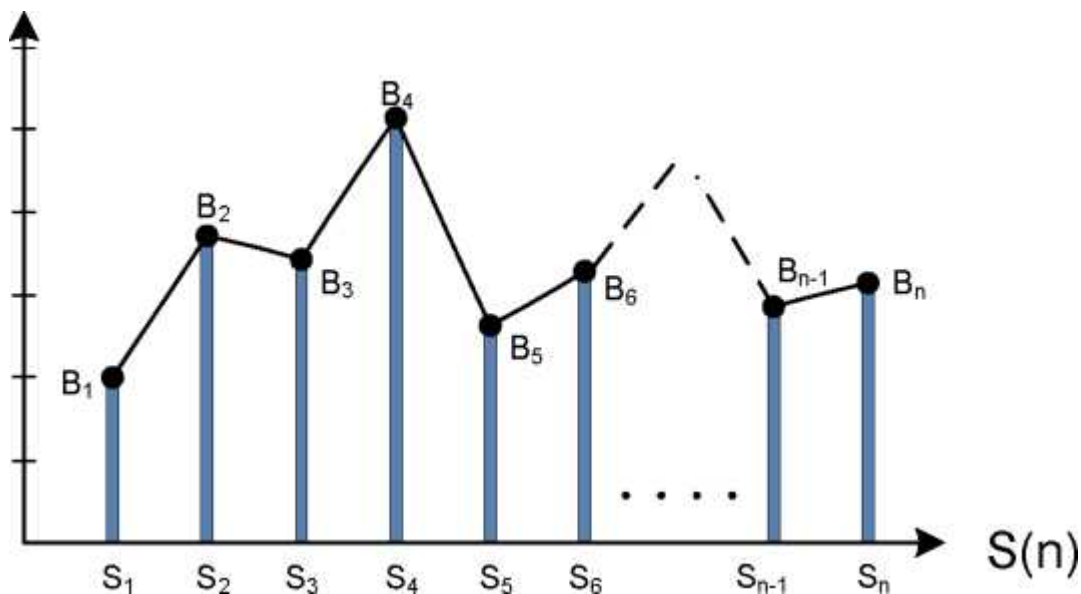


Рис.2 Шаблон для анализа и обработки потока информации с n сенсоров

С помощью булевых функций можно выразить равенство и неравенство (с определением больше или меньше). Определим условие побитного равенства для переменных из множеств B и X как

$$r_i = b_i \& X_i + \bar{b}_i \& \bar{X}_i, \quad (2)$$

где $i = 0, 1, 2, m-1$, m разрядность переменных, r_i условие равенства i го разряда. Обратное

утверждение (побитное неравенство) можно записать следующим образом:

$$r_i = b_i \& X_i + \bar{b}_i \& \bar{X}_i = b_i \odot X_i \quad (3)$$

Таким образом, условие R равенства переменных из множеств B и X определяется как

$$R = r_0 \& r_1 \& r_2 \dots \& r_n \quad (4)$$

Реализация выражений (2), (3) и (4) в базисе FPGA позволяет за один такт обработки сравнить переменные шаблона множества B с потоком данных от сенсоров и сформировать событие [9], при этом количество шаблонов может быть сколь угодно большим (ограничения определяются емкостью FPGA). Это позволяет существенно сократить время t_{is} обработки информации в СОД.

Современные FPGA, используемые в аппаратуре СОД распределенных СУ содержат достаточное число ресурсов не только для реализации набора быстрых

компараторов при поиске по нескольким шаблонам, но и для определения корреляционной зависимости между отдельными переменными.

ГЛАВА 2. РАЗНОВИДНОСТИ ОБРАБОТКИ ИНФОРМАЦИИ

2.1 Технология обработки текстовой информации

В компьютере для работы с текстами используются ПС (программные средства), называемые текстовыми редакторами (ТР). Существует большое количество разнообразных текстовых редакторов, различающихся по своим возможностям – от очень простых до многофункциональных ПС, называемых издательскими системами. Наиболее популярный среди пользователей IBM – совместных компьютеров текстовый редактор Word for Windows.

ТР это самостоятельная компьютерная программа, используемая специально для ввода и редактирования текстовых данных. Мощные ТР с широкими возможностями по форматированию текста называют таксовыми процессорами. [13]

Технологии обработки текстов одними из наиболее распространенных технологий обработки информации. Текст это последовательность символов, к которым относятся буквы, пробел, знаки препинания, цифры, знаки арифметических операций и т.п. Набираемый на клавиатуре текст воспроизводится на экране монитора в рабочем поле редактора. Текстовый курсор указывает то место на экране, на котором пользователь может оказывать воздействие с помощью редактора. В текстовых документах, созданных на компьютере с помощью текстового редактора, используются разные шрифты.

Современные ТР имеют достаточно большие наборы шрифтов и у каждого шрифта есть свое название. Например: Calibri (Основной текст), Batang и др. Буквы шрифта могут быть разного начертания: полужирное жирное, курсивное. ТР(текстовые редакторы) дают возможность управлять размером символов. При этом в памяти приходится хранить не только коды символов но и указания на способ их изображения. Это в свою очередь увеличивает память. Почти все редакторы используемые и распространение в нашей стране позволяет нам использовать так русский так и латинский алфавит, так же они все имеют режим орфографического

контроля текста. Для этого в машине хранится большой словарь. Благодаря этому при написании ошибки ТР автоматически исправляет ошибки или предоставляет возможные варианты.

Рабочие поле ТР это экран дисплея, на котором отображаются все действия, выполняемые ТР. Важным элементом среды ТР является интерфейс это те средства , с помощью которых пользователь может общаться с ТР и управлять им. На сегодняшний день наиболее предпочтительным является интерфейс в форме меню, из которого специальным маркером (выделенным цветом) можно выбрать те или иные команды ТР. Одновременно с меню на экране высвечивается строка состояния, в которой дается информация о текущем состоянии ТР (режим работы, позиция курсора и пр.). [2]

Текст, обрабатываемый с помощью ТР, хранится в оперативной памяти и визуально может быть представлен в виде рулона бумаги (разделенного на страницы в некоторых ТР), длина и ширина которого в большинстве случаев не позволяют целиком наблюдать его на экране. Таким образом, экран можно считать своеобразным окном, через которое пользователь просматривает текст. Для перемещения этого окна по тексту используются специальные клавиши. Есть ТР, позволяющие открывать несколько таких окон «над» соответствующим количеством текстов.

Таким образом, наблюдали своего рода «эволюцию» программ, позволяющих пользователю работать с текстовыми файлами: от простого текстового редактора, предлагающего минимальный необходимый набор функций по созданию и обработке текстовых файлов, с помощью которого можно производить лишь базовые операции с текстом, до современного текстового процессора Microsoft Word, содержащего огромный набор всевозможных функций и возможностей, позволяющих создавать документы очень высокого качества. Соответственно, разница между документами, созданными в этих редакторах, будет огромна. Понятие «ТР (текстовый редактор)» не удовлетворяет возможностям таких программ как Microsoft Word в данном случае вводится новый термин текстовый процессор. Постоянное развитие и совершенствование ТП максимально приближает их по возможностям к издательским программам

2.2 Технология обработки числовой информации в профильных курсах информатики

В современном обществе, которое называется «информационным», умение пользоваться компьютерными технологиями становится все более значимым для многих профессий. Поэтому изучение технологий обработки числовой информации позволит учащимся более успешно не только продолжить обучение в вузе, но и социализироваться в обществе. Технология обработки числовой информации входит в обязательный минимум содержания основного среднего образования в соответствии с большинством школьных программ. На изучение технологии обработки числовой информации в 10 классе отводится 22 часа.

Умение обрабатывать информацию с помощью компьютерных технологий входит в понятие информационной компетенции. Под компетенцией С.В. Тришина и А.В. Хуторской понимают некоторое «первоначально принятое обществом правило к образованию и надлежащей подготовке специалиста, необходимое для высококачественного выполнения им собственных профессиональных обязанностей». [5]

Информационная компетенция – это умение самостоятельно находить, анализировать, отбирать, редактировать и передавать нужную информацию при помощи устных и письменных коммуникативных информационных технологий.

Исходя из требований информационного общества, несомненно, что для достижения конкретного уровня информационной компетентности специалисту необходимо:

- постоянно приобретать новейшие знания и умения в области информационно-коммуникационных технологий;
- развивать собственные коммуникативные и умственные возможности;
- осуществлять интерактивный диалог в едином информационном пространстве.

Информационная компетенция обеспечивает навыки и опыт деятельности учащегося по отношению к информации, содержащейся в учебных предметах и образовательных сферах, а также в окружающем мире. Владение информационной компетенцией подразумевает умение решать следующие задачи:

- 1) Освоение и систематизации знаний, имеющих отношение к средствам моделирования, информационным процессам в разнообразных концепциях (технологических, биологических, социальных);

- 2) Уметь при помощи действительных предметов (компьютер, модем, факс, принтер, копир и т.д.) и информационных технологий (аудио и видеозапись, электронная почта, СМИ, Сеть интернет) собственноручно находить, исследовать и выбирать данные, формировать, изменять, сохранять и передавать ее;
- 3) Уметь конструировать математическую модель, алгоритм, создавать программы на языке программирования;
- 4) Уметь находить пути доступа к базам данных и средствам информационного обслуживания;
- 5) Понимать отличие форм и способов представления данных в вербальной, графической и числовой формах;
- 6) Знать о существовании доступных источников информации и уметь ими пользоваться;
- 7) Уметь оценивать существующие сведения с разных точек зрения, использовать способы разбора информации;
- 8) Уметь применять имеющиеся сведения при решении различных задач. Одной из важнейших составляющих информационной компетенции является умение работать с числовой информацией. Учитывая, что старшая школа является профильной, следует обратить особое внимание на подбор практических задач для каждого профиля при изучении темы «Обработка числовой информации». [3]

Содержание таких задач должно отвечать тем реальным ситуациям, с которыми школьники могут столкнуться при выборе профессии соответствующего профиля. В соответствии с приказом Министерства Образования от 6 октября 2009 г. № 413-ФЗ в старшей школе основными являются следующие профили:

1. физико-математический;
2. социально-экономический;
3. информационно-технологический;
4. социально-гуманитарный;
5. естественно-научный (биология и географии). При изучении темы «Обработка числовой информации» для физико-математического профиля на практических занятиях целесообразно предложить задачи следующих типов: – построение

графиков функций; – использование встроенных функций MS Excel; – моделирование физических экспериментов; – выполнение сложных расчетов с использованием макросов

Примеры: 1. Используя математические функции ОСТАТОК, ОТБР, получить сумму цифр числа от 0 до 999. 2. На отдельном листе рабочей книги построить график функции $y = -4x^2 + 3$ для всех целых $x \in [-10, 10]$. Для социально-экономического профиля на практических занятиях целесообразно предложить задачи следующих типов:

- выполнить финансовые вычисления в MS Excel;
- использовать финансовые функции для экономических расчетов;
- осуществлять сортировку, фильтрацию, консолидацию, подведение итогов и сводные отчеты в базах данных, организованных на основе списков в MS Excel;

Примеры задач: 1. Рассчитать размер комиссии каждого продавца в зависимости от объема его продаж.

2. Рассчитать общий объём продаж, количество месяцев продаж и средний объём выручки в месяц для каждого товара, используя функции СУММ, СЧЁТ и СРЗНАЧ. Для информационно-технологического профиля на практических занятиях целесообразно предложить задачи следующих типов:

- работа в MS Excel с применением макросов;
- работа с БД;
- осуществлять сортировку, фильтрацию, консолидацию, подведение итогов и сводные отчеты в базах данных, организованных на основе списков в MS Excel;

Примеры: 1. Создать макрос для преобразования всех букв в тексте в указанном диапазоне ячеек в прописные;

2. Отфильтровать данные в списке.

3. Объединить таблицы, которые находятся в разных книгах: Для социально-гуманитарного профиля на практических занятиях целесообразно предложить задачи следующих типов: – просматривать, дополнять и обновлять данные с

помощью электронных форм; – находить и извлекать только лишь необходимые сведения с помощью запросов; – анализировать либо печатать данные в установленном макете с помощью отчетов.

Пример:

1. Используя функции категории "Текстовые" (СЦЕПИТЬ, ДЛСТР, ПСТР), выполнить задание. В ячейке A2 ввести своё имя, в ячейке A3 должно появиться приветствие "Здравствуйте, Имя!".
2. Разработать таблицу, содержащую следующие сведения об абитуриентах: фамилия, оценки за экзамены по математике, русскому и иностранному языкам, сумма баллов за три экзамена и информацию о зачислении: если сумма баллов больше или равна проходному баллу и оценка за экзамен по математике — 4 или 5, то абитуриент зачислен в учебное заведение, в противном случае — нет.

Для естественно-научного профиля на практических занятиях целесообразно предложить задачи следующих типов: – построение графиков, диаграмм и гистограмм функций; – использование встроенных функций MSExcel; – использование пакета анализа данных;

Пример:

1. С помощью биномиального распределения определить, какова вероятность того, что 7 из 10 бригадиров, сдающих тест на технику безопасности, его пройдут, если вероятность успеха для каждого из бригадиров равна $\frac{1}{2}$ (т. е. каждый бригадир знает половину материала).
2. С помощью инструментов пакета анализа данных сгенерировать массив случайных чисел, построить интервальный вариационный ряд, вычислить частоты и построить графики их распределения. Таким образом, подобранные с учетом профиля практические задания по обработке числовой информации с помощью MSExcel будут способствовать формированию не только информационной компетенции учащихся, но и начальному погружению их специфику будущей профессии

2.3 Преимущества и недостатки распределительной обработки данных

Проблемы распределенных баз данных

Исходя из определения Дэйта, распределенную базу данных в общем случае можно рассматривать как слабосвязанную сетевую структуру, узлы которой представляют собой локальные базы данных. Локальные базы данных автономны, независимы и самоопределены; доступ к ним обеспечивается от различных поставщиков. Связи между узлами — это потоки тиражируемых данных. Топология DDB варьируется в широком диапазоне: возможны варианты иерархии, структур типа звезда и т. д. В целом топология DDB определяется географией информационной системы и направленностью потоков тиражирования данных.

Рассмотрим теперь проблемы реальных распределенных баз данных. Проблемы централизованных СУБД существуют и здесь, однако децентрализация добавляет новые:

а) Какова общая модель данных распределенной системы? Мы должны иметь единую концептуальную схему всей сети. Это обеспечит логическую прозрачность данных для пользователя, в результате чего он сможет формировать запрос ко всей базе, находясь за отдельным терминалом (т. е. как бы работая с централизованной базой данных).

б) Необходима схема, определяющая местонахождение данных в сети. Это обеспечит прозрачность размещения данных, благодаря которой пользователь может не указывать, куда переслать запрос, чтобы получить требуемые данные.

в) Распределенные базы данных могут быть однородными или неоднородными по аппаратным и программным средствам. Проблему неоднородности сравнительно легко решить, если распределенная база является неоднородной по аппаратным средствам, но однородной по программным средствам (одинаковые СУБД в узлах). Если же в узлах распределенной системы используются разные СУБД, необходимы средства преобразования структур данных и языков. Это должно обеспечить прозрачность преобразования в узлах распределенной базы данных.

г) Управление словарями. Для обеспечения всех видов прозрачности в распределенной базе данных нужны программы, управляющие многочисленными справочниками или словарями.

д) Методы выполнения запросов в распределенной базе данных отличаются от аналогичных методов централизованных СУБД, так как отдельные части запроса нужно выполнять в месторасположении соответствующих данных и передавать частичные результаты на другие узлы; при этом должна быть обеспечена координация всех процессов.

е) В распределенной базе данных нужен сложный механизм управления одновременной обработкой, который, в частности, должен обеспечивать синхронизацию при обновлениях информации, это гарантирует непротиворечивость данных.

ж) Развитая методология распределения и размещения данных, включая разбиение, является одним из основных требований к распределенной базе данных.

База данных физически распределяется по узлам компьютерной информационной системы при помощи фрагментации и репликации (тиражирования) данных. [7]

4 Особенности распределенных баз данных

В сегодняшнем быстро меняющемся компьютерном мире сосуществуют по крайней мере три основные идеологии: клиент - сервер, Web и распределенные объекты (DCOM, CORBA). Внутри каждого направления также существует большое количество решений и стандартов от разных производителей. Сегодняшняя ситуация вызывает очень большую озабоченность независимых разработчиков и потребителей: Какую технологию выбрать и что будет со мной и моим бизнесом, если я приму неправильное решение? При этом очевидно, что цена ошибки будет весьма высока, кроме того большие средства уже вложены в разработку и эксплуатацию уже существующих систем.

Клиент-сервер

Термин "клиент-сервер" означает такую архитектуру программного комплекса, в которой его функциональные части взаимодействуют по схеме "запрос-ответ". Если рассмотреть две взаимодействующие части этого комплекса, то одна из них (клиент) выполняет активную функцию, т. е. инициирует запросы, а другая (сервер) пассивно на них отвечает. По мере развития системы роли могут меняться, например некоторый программный блок будет одновременно выполнять функции сервера по отношению к одному блоку и клиента по отношению к другому.

Любая информационная система должна иметь минимум три основные функциональные части - модули хранения данных, модули обработки и модули интерфейса с пользователем. Каждая из этих частей может быть реализована независимо от двух других. Например, не изменяя программ, используемых для хранения и обработки данных, можно изменить интерфейс с пользователем таким образом, что одни и те же данные будут отображаться в виде таблиц, графиков

или гистограмм. Не меняя программ представления данных и их хранения, можно изменить программы обработки, например изменив алгоритм полнотекстового поиска. И наконец, не меняя программ представления и обработки данных, можно изменить программное обеспечение для хранения данных, перейдя, например, на другую файловую систему.

В классической архитектуре клиент-сервер приходится распределять три основные части приложения по двум физическим модулям. Обычно ПО хранения данных располагается на сервере (например, сервере базы данных), интерфейс с пользователем - на стороне клиента, а вот обработку данных приходится распределять между клиентской и серверной частями. В этом-то и заключается основной недостаток двухуровневой архитектуры, из которого следуют несколько неприятных особенностей, сильно усложняющих разработку клиент-серверных систем.

При разбиении алгоритмов обработки данных необходимо синхронизировать поведение обеих частей системы. Все разработчики должны иметь полную информацию о последних изменениях, внесенных в систему, и понимать эти изменения.

Это создает большие сложности при разработке клиент-серверных систем, их установке и сопровождении, поскольку необходимо тратить значительные усилия на координацию действий разных групп специалистов. В действиях разработчиков часто возникают противоречия, а это тормозит развитие системы и вынуждает изменять уже готовые и проверенные элементы.

Технология клиент-сервер по праву считается одним из "китов", на которых держится современный мир компьютерных сетей. Но те задачи, для решения которых она была разработана, постепенно уходят в прошлое, и на сцену выходят новые задачи и технологии, требующие переосмысления принципов клиент-серверных систем.

ГЛАВА 3.ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ОБРАБОТКИ ОТКРЫТОЙ ИНФОРМАЦИИ В УСЛОВИЯХ АКТИВНОГО РАЗВИТИЯ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ

ТЕХНОЛОГИЙ

По данным исследовательской компания Digital Universe, что в 2011 году мировой объем цифровой информации вырос до 1.8 зеттабайт (1.8 триллионов гигабайт). Через 5 лет, в 2016 году, данный показатель увеличится в 9 раз, а через 10 лет – уже в 50 раз. К 2020 году мировое сообщество произведет количество информации, которое будет превышать 35 зеттабайт. Резко увеличится, приблизительно в 75 раз, количество форматов данных, так называемых "контейнеров", а штат IT-специалистов, управляющих аппаратными и информационными средствами, оперирующими данными, увеличится всего в 1.5 раза [3].

Изменения, вызванные стремительным развитием ИКТ, способствующие более оперативному и полному доступу к большим объёмам открытой информации и сделали актуальными вопросы повышения эффективности поиска и обработки требуемой информации. В целях получения требуемой информации, для принятия рационального управленческого решения, задача поиска в информационном пространстве представляется перебором многомерного потока информации. В данных условиях простой перебор информации является неэффективным, так как, с достаточной вероятностью достоверности его еще необходимо своевременно и полно обработать.

В соответствии с теорией управления начальным и необходимым этапом принятия решения является оценка обстановки (текущей и перспективной). Чем сложнее является система, тем большего количества параметров текущей обстановки следует проанализировать. Для своевременного принятия решения и повышения эффективности работы руководителя необходимым является создание органов обработки информации (ООИ), призванных анализировать требуемые параметры в интересах принятия управленческого решения руководителем. По мере увеличения объёма информации, циркулирующего через ИКТ, выходные документы средств накопления и обработки информации постоянно усложняются и увеличиваются в объёме.

Однако возможности человека имеют свои ограничения, в частности средняя скорость чтения современного человека составляет 400 – 600 знаков в минуту при условии 100 % усвоения материала. Следовательно, в данных условиях традиционные задачи, такие как поиск и анализ растущего потока информации, требуют новых решений. С целью повышения эффективности работы организации применяется автоматизация трудоёмких процессов, в том числе и такого этапа

деятельности организаций как обработка информации.

Эффективность решения задач обработки информации напрямую зависит от используемых методов и способов её реализации. В соответствии с видами представления информации рассматриваются и методы её обработки: структурные, статистические логические (семантические). Самостоятельное существование каждого из указанных методов обработки во многом условно. Практическое решение задач обработки приводит к комбинированному применению методов обработки. Способами обработки информации являются ручной, механизированный, автоматизированный, автоматический.

Они определяются в зависимости от применяемых технических средств и степени участия человека в решении задач обработки. С целью повышения эффективности работы организации активно применяется автоматизация трудоёмких процессов, в том числе и такого этапа деятельности организаций как обработка информации. Повышение эффективности работы, в данном случае, возможно на основе применения математического аппарата теории массового обслуживания [1].

Однако, современная обстановка обуславливает необходимость применения существующих методов обработки информации в сочетании со способами в целом.

В результате объединения этих двух компонентов открывается перспектива применения новых комбинаций проверенных методов обработки и постоянно совершенствуемых органами обработки информации приёмов работы, что, в свою очередь, требует постоянной адаптации средств информационного поиска и обработки информации, к изменению обстановки, а также максимально-возможной автоматизации данных процессов с применением теории массового обслуживания.

Учитывая постоянное усложнение информационных потоков должны развиваться методы и способы для поискового исследования информации. Анализируя проблемные вопросы руководитель должен понимать, что необходимо принять такое решение, которое не только исправит поведение управляемой системы (приблизит реальное, практическое поведение системы к требуемому), но и не позволит в будущем системе выходить за области допустимых значений в аналогичных ситуациях [4].

Анализ процедуры информационного поиска показывает несоответствие её информационно-лингвистического обеспечения современным реалиям. В условиях необходимости охвата и обобщения огромных динамических разнородных информационных потоков, а также непредсказуемости характера представляющих

интерес данных, одним из наиболее перспективных путей решения данной проблемы является постепенный отказ от использования в качестве единственного критерия отбора наличия в них определенных слов и развитие новых подходов к их обработке.

Самостоятельным направлением, при решении данных задач в современной системе обработки информации любой системы, является совершенствование процедуры информационного поиска, а так же разделения информационных массивов. В современных поисковых системах тексты автоматически индексируются по набору составляющих эти тексты слов.

Такое представление текстов как простого набора слов имеет большое количество очевидных недостатков, затрудняющих поиск релевантных текстов, таких как: избыточность в пословном индексе используются слова-синонимы, выражающие одни и те же понятия; слова текста считаются независимыми друг от друга, что не соответствует свойствам связного текста; многозначность слов поскольку многозначные слова могут рассматриваться как дизъюнкция двух или более понятий, выражающих различные значения многозначного слова, то маловероятно что все элементы этой дизъюнкции интересуют пользователя.

Данные недостатки могут быть устранены за счёт применения концептуального индексирования, то есть такое индексирование, когда текст индексируется не по словам, а по понятиям, которые обсуждаются в данном тексте. Для того чтобы реализовать на практике схему автоматического концептуального индексирования и концептуального поиска необходимо иметь ресурс, описывающий систему понятий данной предметной области, то есть онтологию в данной предметной области.

Улучшение качества информационного поиска возможно за счет различных способов применения онтологических ресурсов, в их числе использование традиционных информационно-поисковых тезаурусов в комбинации с разного рода статистическими моделями; тезауруса для автоматического индексирования в булевских моделях поиска документов, в задаче автоматической рубрикации, автоматического аннотирования. Тезаурусы несут дополнительную семантику, определяя связи между терминами. Отношения, свойственные для тезаурусов: синонимия, иерархическое отношение и ассоциация.

Основанные на интеллектуальных технологиях управления знаниями, новые методы направлены на более эффективное выявление ожидаемой информации из

огромных объемов открытой информации с одной стороны, и её комплексный анализ для идентификации тех знаний, которые скрыты в логическом смысле информации и обычно недоступны для осознания человеком в малый промежуток времени, с другой.

Одним из действенных инструментов описания предметной области обработки информации являются тезаурусы, как специальная терминология, полно охватывающая понятия, определения и термины требуемой сферы деятельности ООИ, способствующие правильной оценке обстановке и выработке требуемых данных.

Тезаурус включает вводную часть, состоящую из титульного листа и введения, основную часть (лексико-семантический указатель) и дополнительные части (систематический, пермутационный, иерархический и другие указатели и списки специальных категорий лексических единиц). Обязательными составными частями являются вводная часть и лексико-семантический указатель.

Лексико-семантический указатель содержит дескрипторные и аскрипторные статьи. Дескрипторная статья состоит из заглавного дескриптора, списка дескрипторов и аскрипторов, семантически связанных с ним, с обозначением видов связи. Аскрипторная статья состоит из аскриптора и заменяющих его при обработке и поиске информации дескрипторов или комбинации дескрипторов.

Для организации эффективного управления, необходимо стремиться затрачивать минимальное количество времени на выработку грамотного управленческого решения, в любой момент времени цикла управленческой деятельности. 2 В целях реализации данного вопроса, в ходе управленческой деятельности, рациональным является внедрение методик, примененных на практике [5].

Для повышения эффективности работы ООИ перспективным является разработка методики по созданию тезариуса. Методика работы по созданию тезауруса включает следующие этапы: определение тематического охвата тезауруса (области его применения); сбор массива лексических единиц (терминов), выделение специфических для данной области ключевых слов, представляющих терминологию области; формирование словника тезауруса; построение словарных статей и указателей, У выбранных в качестве дескрипторов слов устраняется многозначность. Между дескрипторами устанавливаются системные (парадигматических) связи. Дескрипторы объединяются в предметные классы, представляющие собой основные разделы данной предметной области.

оформление тезауруса; его экспертиза и регистрация

ЗАКЛЮЧЕНИЕ

Важно отметить, что распределение (или разделение) не идентично параллелизму. Распределение видов обработки информации состоит в том, чтобы поручить их вычислительным машинам, наилучшим образом приспособленным к этому. Параллелизм же подразумевает понятие одновременности обработки информации. При этом распределение позволяет в ряде ситуаций проводить эффективную параллельную обработку информации при выполнении больших объемов параллельных вычислений.

Таким образом, в общем случае распределение не подразумевает параллелизма, но возможность "распараллелить" распределенную обработку информации существует. Возможность взаимодействия вычислительных систем при реализации распределенной обработки информации определяют как их способность к совместному использованию данных или к совместной работе с использованием стандартных интерфейсов. Взаимодействие подразумевает понятие "открытых систем", то есть систем, способных к коммуникации в неоднородной среде.

Взаимодействие между программами с точки зрения хронологии последовательно приобретает следующие формы:

обмен (программы различных систем посылают друг другу сообщения, как правило, файлы);

разделение (имеется непосредственный доступ к ресурсам нескольких машин, например, совместное использование файлов);

совместная работа (машины играют в реализации программы взаимодополняющие роли).

При использовании сетевых информационных технологий становится возможной реализация территориального распределения производства. Для администрации фирмы становится безразлично, где именно находится производство: в этом здании, за 100 м или за 10 000 км. Появляются совсем другие проблемы, такие как межконтинентальное снабжение, поясное время и т.д., поскольку становится возможным планетарное распределение промышленного производства.

Могут создаваться транснациональные компании, реализующие мировой товарный экспорт внутри фирмы. При этом метрополия, вложив 5 – 7 % от суммы оборота в экономику другой страны, получает возможность контролировать 50 – 60 % ее экономики. Объясняется это тем, что за счет вложения наукоемких технологий, страна-метрополия получает возможность оказывать влияние и даже осуществлять контроль за экономическим и политическим развитием другой страны.

СПИСОК ЛИТЕРАТУРЫ

1. Арсеньев Б.П., Яковлев С.А. Интеграция распределенных баз данных. - СПб.: Лань, 2011. - 461с.
2. Бураков П.В., Петров В.Ю. Введение в системы баз данных: Учебное пособие. - СПб: СПбГУ ИТМО, 2010. - 128 с.
3. Голицина О.Л., Максимов Н.В., Попов И.И. Базы данных: Учебное пособие. - 2-е изд., испр. и доп. - М.: Форум, Инфра-М, 2009. - 400 с.
4. Власов Т.Н. и др. Освоение офисных технологий. Обработки текстовой информации // Материалы дистанц. Курса. Центр информатизации образования КК ИПК РО ; Красноярск, 2013. – 594с.
5. Камилова Р.Ш., Муртилова К.М-К. Принципы выбора программного обеспечения автоматизации бухгалтерского учета. Актуальные вопросы современной экономики. 2014. №4. С. 388-392.
6. Котенев Е.В., Игнатъев И.Ю. Метод анализа в управленческой деятельности руководителя (начальника). // Современные проблемы и перспективы развития гуманитарных, технических, общественных, естественных наук и промышленной безопасности. Сб. статей по итогам всероссийской научно-практической конференции с международным участием, 16-17 октября 2014 года, г. СанктПетербург. – СПб.: Изд-во «КультИнформПресс», 2014 – с. 99-102.
7. Котенев Е.В. Методика в управленческой деятельности в современных условиях. // Научный взгляд на современный этап развития общественных, технических, гуманитарных и естественных наук. Актуальные проблемы. Сб. статей по итогам всероссийской научно-практической конференции, 2-3 сентября 2014 года, г. СанктПетербург. – СПб.: Изд-во «КультИнформПресс», 2014. – с. 80-81.
8. Карпова, Т.С. Базы данных: Модели, разработка, реализация: Учебное пособие. - СПб.: Питер, 2012. - 303 с.
9. Корнеев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. - М.:Нолидж, 2010. - 352 с.

10. Кузнецов С.Д. Основы баз данных. 2-е изд. М.: Бином, 2017. - 488 с.
11. Камилова Р.Ш., Муртилова К.М-К. Принципы выбора программного обеспечения автоматизации бухгалтерского учета. Актуальные вопросы современной экономики. 2014. №4. С. 388-392.
12. Мурлин А.Г., Мурлина В.А., Янаева М.В. Технология обработки информации средствами текстового редактора. Учебно-практическое пособие / Краснодар, 2014. - 385с.
13. Шахбанов Р.Б., Бабаева З.Ш. Когнитивная наука в экономике и образовании. // Научное обозрение. Серия1:Экономика и право. 2013.№1-2. С.170-176