

## Всероссийская перепись населения: современные методы сбора и обработки информации

Статья рекомендована С. Б. Шапошником 20.05.2016



**МАНЖУЛА Олег  
Владимирович**

*Заместитель начальника  
управления информационных  
ресурсов и технологий  
Росстата*

### Аннотация

В статье рассматриваются ключевые задачи проведения Всероссийской переписи населения на основе использования новых информационно-коммуникационных технологий с применением мобильных устройств и сети Интернет, а также экспертная система обоснования выбора методов сбора и обработки информации Всероссийской переписи населения с учетом географических и социально-экономических особенностей регионов.

### Ключевые слова:

**Всероссийская перепись населения, информационно-коммуникационные технологии, методика, экспертная система, мобильные средства, сеть Интернет, метод, качественная оценка, ресурсное обеспечение.**

Проведение Всероссийской переписи населения (далее — ВПН) на основе применения новых информационно-коммуникационных технологий призвано обеспечить три аспекта эффективности информатизации процессов сбора и обработки статистической информации о населении:

- *социальный*, когда каждый гражданин сможет пройти данную процедуру таким способом, который ему подходит в большей мере;
- *организационный*, предполагающий оптимальное распределение ресурсов (материальных, трудовых, временных) с учетом особенностей каждого территориального образования;
- *экономический*, предполагающий сокращение финансовых и временных ресурсов, уменьшение числа возможных ошибок при движении информационных потоков, снижение рисков в результате технических сбоев.

С учетом перечисленных аспектов проведения ВПН, международного опыта [1] и сложившихся практик Росстата [2] предлагаются следующие методы сбора и обработки информации в ходе переписи населения.

1. *Последовательный метод* сбора и обработки информации, основанный на разделении этапов проведения интернет-переписи и переписи путем обхода квартир (домов) проживания переписчика. Сущность метода состоит в том, что на первом этапе проводится только интернет-перепись. Обход жилых помещений и опрос респондентов осуществляется на втором этапе переписи. Таким образом можно сократить количество помещений, требующих обхода переписчиками, привлечь интернет-аудиторию к участию в переписи. Последовательный метод является наименее затратным, так как к респондентам, переписавшимся самостоятельно, переписчику идти

не нужно. Однако указанный метод наиболее продолжительный по времени: предполагается, что будет отведено две-три недели на проведение интернет-переписи, несколько дней на автоматизированную обработку ее данных и доведение этой информации до переписчиков, а затем две недели на обход оставшихся домохозяйств переписчиками и неделя на контрольные мероприятия.

2. *Одновременный метод* сбора и обработки информации переписи населения, предполагающий одномоментное проведение переписи путем обхода помещений и интернет-переписи. Сущность метода состоит в том, что переписчики осуществляют обход всех помещений и предоставляют респондентам выбор способа опроса. Это более затратный метод, по сравнению с первым. Период переписи составляет две недели и одна неделя — контрольные мероприятия.

3. *Локализованный метод* сбора и обработки информации переписи населения, учитывающий разделение помещений на группы для прохождения интернет-переписи и опроса путем обхода помещений. Сущность метода заключается в том, что до начала переписи выделяются наиболее перспективные с точки зрения наличия сети Интернет и других социально-экономических условий помещения. Респондентам, живущим в них, сразу предлагается пройти перепись через интернет. Другой группе предлагается на выбор — через интернет или путем непосредственного опроса. Метод позволяет увеличить число респондентов, прошедших перепись самостоятельно, однако требует дополнительных действий с их стороны, что может привести к отказу от участия в переписи. Период переписи также составляет две недели и одна неделя — контрольные мероприятия.

В результате анализа достоинств и недостатков указанных выше методов сделан вывод о том, что последовательный метод сбора и обработки информации переписи населения с разделением этапов интернет-переписи и переписи путем обхода помещений значительно снижает нагрузку на переписчиков за счет сокращения количества домохозяйств, требующих проведение опроса на месте. Вместе с тем в этом случае увеличивается время проведения переписи, возникает необходимость проведения в середине переписи автоматизированной обработки данных о респондентах, прошедших интернет-перепись, а также риск выявления респондентов, «не успевших» пройти опрос on-line, и риск дублирования заполненных электронных анкет. Поэтому два других метода направлены на минимизацию перечисленных недостатков и предлагаются к использованию в условиях, когда их применение компенсирует перечисленные риски. Неоднозначность выбора метода сбора и обработки информации обуславливает необходимость разработки экономико-математических моделей решения задачи обоснования выбора метода сбора и обработки информации с учетом географических, социально-экономических и финансовых особенностей регионов России.

## **Методика выбора метода сбора и обработки информации**

Такая методика при проведении ВПН в каждом регионе России базируется на анализе качественных факторов, связанных с анализом социальных,

географических и прочих факторов и необходимости обоснования ресурсного обеспечения (рис. 1).

В данной методике с помощью экспертной системы осуществляется обоснование выбора метода сбора и обработки информации при проведении ВПН с учетом качественных факторов, а модели позволяют осуществить реализацию выбранного метода в отдельном регионе России.

Для решения первой задачи — выбора одного из возможных методов сбора и обработки информации с учетом анализа качественных оценок различных географических и социально-экономических факторов регионов России — обычно применяются экспертные системы [3–5]. Для решения второй задачи — расчета ресурсного обеспечения — строятся математические модели с учетом риска неполноты информации об экономических процессах и их параметрах, что часто определяется ограниченностью ресурсов (финансовых, временных, технических и др.) [6].

Решение задачи выбора метода сбора и обработки информации основано на определении типа региона по набору качественных оценок выявленных факторов (табл. 1), для которого может быть применен тот или иной метод сбора и обработки информации и проверки ресурсных ограничений.

Таким образом, методика выбора метода сбора и обработки статистической информации в рамках проведения переписи населения разработана на основе метода построения экспертных систем с использованием аппарата теории нечетких множеств. Предлагается осуществить качественную оценку выявленных факторов, позволяющих интерпретировать по шкалам количественные данные и определить степень принадлежности к значениям

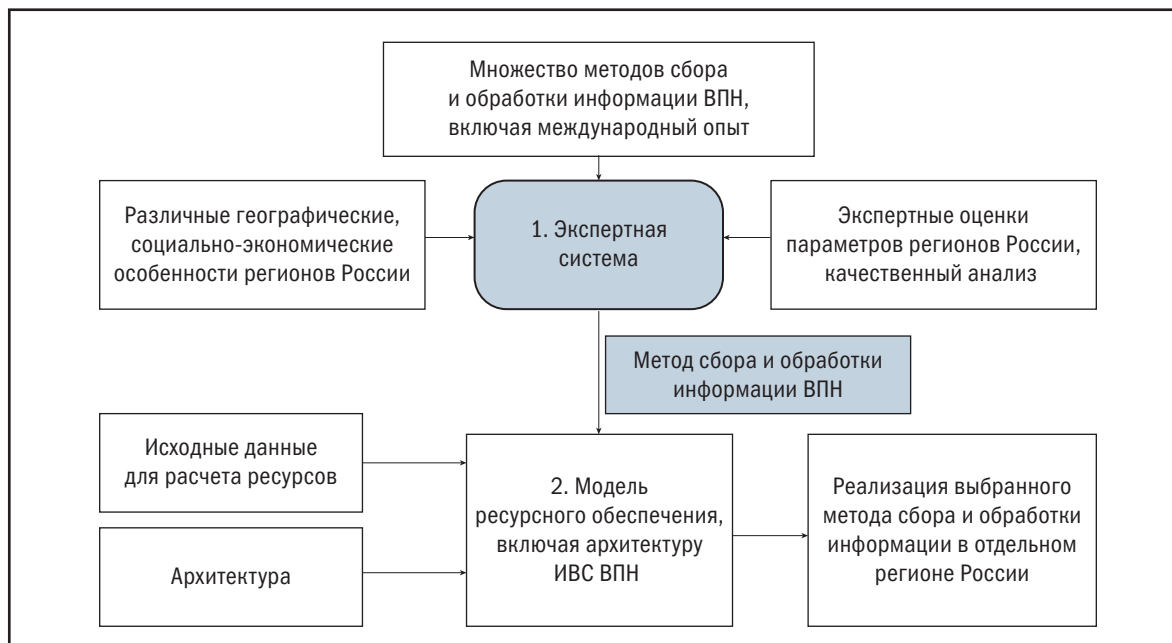


Рис. 1. Структурная блок-схема методики выбора метода сбора и обработки информации при проведении ВПН в каждом регионе России

лингвистических переменных, а далее осуществить оценку классов факторов (готовность населения к переписи с использованием ИКТ, особенности региона, экономическая среда, ИКТ-инфраструктура региона) и целевой переменной — выбора метода сбора и обработки информации при проведении переписи населения [5, 6] и выполнении проверки бюджетных ограничений по затратам на сбор, передачу и обработку информации о населении в каждом регионе России.

### Экспертная система выбора метода сбора и обработки информации Всероссийской переписи населения с учетом географических и социально-экономических особенностей регионов

Задача принятия решений — одна из самых распространенных в любой предметной области [7]. Она сводится к выбору одной или нескольких альтернатив из некоторого набора. Для того чтобы сделать такой выбор, необходимо четко определить цель и критерии (показатели качества), по которым будет проводиться оценка некоторого набора альтернативных вариантов. Основные критерии, на основании качественного анализа характеристик которых будет построен выбор метода проведения переписи населения в отдельном субъекте Российской Федерации, определены в таблице 1.

Предлагаемая экспертная система обеспечивает качественную интерпретацию множеств параметров. В соответствии с таблицей 1 выбор

Факторы		Принципы
Готовность населения к участию в переписи с использованием ИКТ	Готовность населения к участию в переписи	<ol style="list-style-type: none"> <li>1. Обеспечение полноты охвата населения в городской и сельской местности</li> <li>2. Исключение двойного счета при условии применения нескольких методов опроса населения</li> <li>3. Защита конфиденциальных (персональных) данных на всех носителях информации и этапах обработки данных</li> <li>4. Повышение доверия населения к важному государственному мероприятию — ВПН</li> <li>5. Обеспечение качества полученных данных</li> <li>6. Оперативность сбора и обработки информации</li> <li>7. Достоверность обработки информации</li> </ol>
	Возраст	
Особенности региона	Уровень образования	
	Мотивация участия в переписи	
	Площадь региона	
Экономическая среда	Удаленность региона, муниципального образования	
	Труднодоступность отдельных районов	
	Уровень жизни (низкий, средний, высокий)	
ИКТ-инфраструктура	Материальное стимулирование участия в переписи населения	
	Развитие инфраструктуры (качество оптоволоконной связи)	
	Надежность инфраструктуры	
Квалификация и численность персонала	Используемые технические средства и технологии	

Табл. 1. Факторы и принципы, оказывающие влияние на эффективность ВПН и получение достоверных сведений о населении

метода сбора и обработки информации переписи населения в отдельном регионе России имеет следующие ключевые факторы (подцели решения задачи):

- A – готовность населения к участию в переписи с ИКТ;
- B – экономическая среда;
- C – ИКТ-инфраструктура;
- D – особенности региона.

Выше определены показатели, влияющие на значения факторов при решении задачи выбора метода сбора и обработки информации переписи населения. Общая структура взаимосвязи целевых переменных, увязывающая анализ всех факторов, влияющих на определение их значений с использованием метода оценки нечетких переменных, представлена в виде графа «и-или» (рис. 2).

Граф «и-или» позволяет представлять в графическом виде функциональные зависимости целевых переменных от выявленных факторов [8, 9]. В этом графе каждая вершина-фактор, за исключением конечных, представляет собой целевую переменную, а определяющие его зависимые вершины-аргументы – связанные с ней подчиненные факторы. Условие конъюнкции (совместность анализа факторов) обозначается пересекающейся дугой, а условие дизъюнкции (независимость влияния факторов на целевую переменную) никак не обозначено, причем если какой-либо фактор-аргумент определяется другими факторами-аргументами, то он становится подцелью. Каждой целевой переменной (подцели) в графе «и-или» соответствует некоторая задача (подзадача), которая не может быть решена, пока не будут получены значения нижестоящих подцелей (решены подзадачи). Важность рассмотрения подцелей может быть задана с помощью некоторого весового коэффициента. Таким образом, концептуальный анализ в виде графа «и-или» отражает в обобщенной форме процесс решения характерных для переменной задач, в данном случае оценки факторов, влияющих на выбор метода сбора и обработки информации.

Представленному на рисунке 2 графу «и-или» соответствуют следующие подзадачи.

A – оценка готовности населения к участию в переписи с применением ИКТ на основе следующих характеристик:

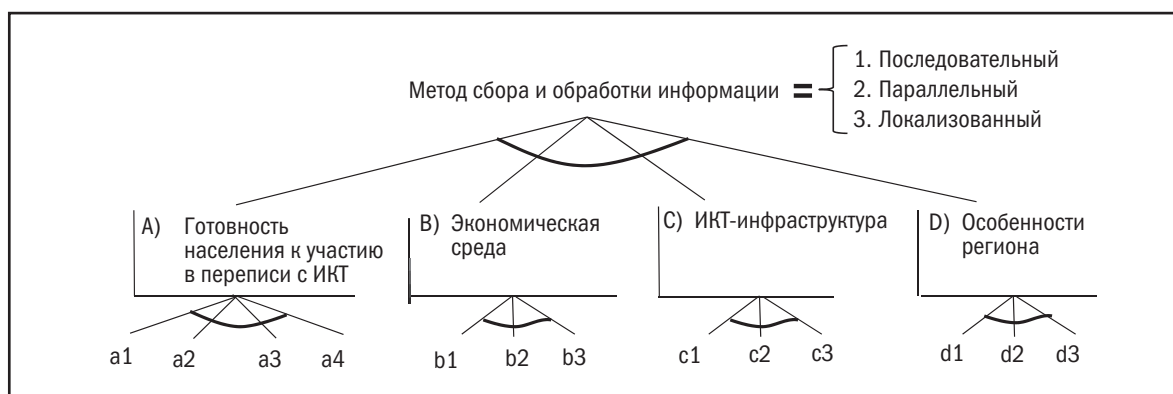


Рис. 2. Граф «и-или» анализа факторов, влияющих на выбор метода сбора и обработки информации при проведении переписи населения в отдельном регионе России

a1 — доверие населения к ИКТ (компьютерная грамотность населения, сложность заполнения электронных анкет, доверие респондентов к переписи населения, обеспечение конфиденциальности персональных данных);

a2 — возраст: молодежь в возрасте 18–24 лет (работающая молодежь, учащаяся молодежь), взрослое население в возрасте 25–60 и в возрасте старше 60 лет);

a3 — уровень образования (высшее и послевузовское профессиональное, неполное высшее образование, среднее или начальное профессиональное, среднее (полное) общее, основное общее и ниже);

a4 — мотивация участи в интернет-переписи (ИП) (желание/нежелание пускать переписчика в дом, личная заинтересованность самостоятельно пройти ИП, комфортные условия заполнения анкет).

В — оценка экономической среды на основе следующих характеристик:

b1 — уровень жизни (низкий, средний, высокий);

b2 — наличие точки доступа;

b3 — материальное стимулирование участия в переписи населения (100–200 руб., положенных на счет мобильного телефона, очень эффективное стимулирующее средство, что обойдется дешевле любой рекламы и планшетных компьютеров).

С — оценка ИКТ-инфраструктуры региона на основе следующих характеристик:

c1 — развитие инфраструктуры (доля домохозяйств, имеющих доступ к интернету в регионе, ежегодный прирост использования сети интернет в регионе, организация ИП со стороны региональной власти);

c2 — надежность инфраструктуры (риск хакерских атак, взлома сайтов интернет-переписи, появления сайтов «двойников» переписи, наличие устойчивой связи);

c3 — используемые технические средства и технологии.

Д — оценка особенности региона на основе следующих характеристик:

d1 — площадь региона;

d2 — удаленность региона (муниципального образования);

d3 — труднодоступность отдельных районов.

Вследствие качественного характера задачи оценивания факторов для реализации описанного графа «и-или» предлагается использовать статическую экспертную систему, основанную на обработке лингвистических переменных.

Оценки по перечисленным подцелям формируются путем классификации параметров с помощью таблиц решений, в которой каждая строка есть правило, определяющее зависимость значений анализируемых характеристик от конъюнкции факторов.

Набор признаков, который функционально определяет характеристику подцели А — оценки готовности населения, показан в таблице 2.

Характеристики подцелей определяются путем сравнения оцениваемых показателей с требуемыми условиями проведения переписи населения. Неудовлетворительному значению при этом присваивается «0», удовлетворительному значению — «1». В качестве примера общую оценку значения

«готовность населения к участию в интернет-переписи и ПК-переписи» из таблицы 2 можно представить в виде следующей таблицы решений (табл. 3).

В соответствии с приведенной ниже формулой (2), конъюнкция ответов X на признаки характеристики «готовность населения к участию» дает значение Y — готово или не готово население в данном регионе к участию в интернет-переписи и ПК-переписи с коэффициентом уверенности КУА  $\geq 60$ . Двойная черта отделяет признаки от оценки. Аналогично строятся таблицы решений для других характеристик подцелей рисунка 2. Таблица решений

Признаки		Значения			
a1) доверие населения к ИКТ	a11) компьютерная грамотность населения	a111) достаточная для самозаполнения анкет (1)		a112) недостаточная для самозаполнения анкет (0)	
	a12) сложность заполнения электронных анкет	a121) несложная для самозаполнения (1)		a122) сложная для самозаполнения (0)	
	a13) доверие респондентов к переписи населения	a131) имеется (1)		a132) отсутствует (0)	
	a14) обеспечение конфиденциальности ПД	a141) — соответствует требованиям (1)		a142) — не соответствует требованиям (0)	
a2) возраст	a21) возраст 18–24 года	a22) возраст 25–60 лет	a23) старше 60 лет		
a3) уровень образования	a31) высшее	a32) неполное высшее	a33) среднее/ профессиональное	a34) основное общее и ниже	
a4) мотивация	a41) личная заинтересованность	a42) нежелание пускать переписчика в дом	a43) комфортные условия заполнения анкеты		

Табл. 2. Признаки, оказывающие влияние на фактор А — готовность населения к участию в интернет-переписи и переписи с использованием планшетных компьютеров в отдельном регионе России

a1) доверие населения к ИКТ	a2) возраст	a3) уровень образования	a4) мотивация	Оценка готовности	Коэффициент уверенности
1	1	1	1	удовлетв.	100
1	1	1	0	удовлетв.	75
1	1	0	1	удовлетв.	75
1	0	1	1	удовлетв.	75
0	1	1	1	удовлетв.	75
1	1	0	0	неудовлетв.	50
0	0	1	1	неудовлетв.	50
1	0	1	0	неудовлетв.	25
0	1	0	1	неудовлетв.	25
0	0	0	0	неудовлетв.	0

Табл. 3. Оценка готовности населения к участию в интернет-переписи и ПК-переписи в отдельном регионе России

выбора метода сбора и обработки информации при проведении переписи населения в отдельном регионе России в целом показана в таблице 4.

Характер функций принадлежности для различных оцениваемых показателей может иметь разный вид в зависимости от степени их важности. Таким образом, коэффициент уверенности можно рассматривать как весовую значимость оцениваемого показателя.

При решении задачи выбора метода сбора и обработки информации переписи населения использовалась процедура нечеткого вывода. В качестве решения задачи предлагается выбор тех значений целевой переменной метода сбора и обработки информации, для которых получается максимальное значение коэффициента уверенности из минимальных коэффициентов уверенности, характеризующих подцели.

В соответствии с условиями конъюнкции (совместности) параметров на рисунке 2 коэффициент уверенности рассчитывается с условием объединения параметров по формуле:

$$KU_{факт} = \min\{KU_1, KU_2, KU_3, \dots, KU_{i-1}, KU_i\} \quad (1)$$

где  $KU_i$  – оценка уверенности  $i$ -го параметра.

В соответствии с условиями конъюнкции (совместности) подцелей на рисунке 2 коэффициент уверенности выбранного региона России  $KU_{рег}$  рассчитывается по следующей формуле:

А – готовность населения к участию в переписи с ИКТ	В – экономическая среда	С – ИКТ-инфраструктура	Д – особенности региона	Коэффициент уверенности
1	1	1	1	100
1	1	1	0	75
1	1	0	1	75
1	0	1	1	75
0	1	1	1	75
1	1	0	0	50
0	0	1	1	50
1	0	1	0	50
0	1	0	1	50
1	0	0	0	25
0	1	0	0	25
0	0	1	0	25
0	0	0	1	25
0	0	0	0	0

Табл. 4. Оценка решений выбора метода сбора и обработки информации при проведении переписи населения в отдельном регионе России



$$КУ_{рез} = \min\{КУ_A^{факт}, КУ_B^{факт}, КУ_C^{факт}, КУ_D^{факт}\} \quad (2)$$

где  $КУ_A^{факт}$  — коэффициент уверенности оценки фактора А — готовности населения к участию в переписи с ИКТ,  $КУ_B^{факт}$  — коэффициент уверенности оценки фактора В — экономическая среда,  $КУ_C^{факт}$  — коэффициент уверенности оценки фактора С — ИКТ-инфраструктура региона;  $КУ_D^{факт}$  — коэффициент уверенности оценки фактора D — особенности региона.

При этом предлагается разбить регионы России в зависимости от величины итоговой оценки на три группы с помощью следующих правил:

- если  $КУ_{рез} \geq 60$ , то регион готов к проведению переписи методом переписи населения с применением сети интернет в первую очередь (последовательный метод);
- если  $25 \leq КУ_{рез} < 60$ , то регион готов к интернет-переписи частично, сбор информации о населении предлагается осуществлять с применением мобильных средств (параллельный, локализованный методы — в зависимости от особенностей региона и наличия ресурсов);
- если  $КУ_{рез} < 25$ , то регион не готов к применению методов сбора и обработки информации через сеть интернет и с применением мобильных средств, является труднодоступным, рекомендуется проводить перепись населения традиционным способом или с использованием других вспомогательных средств сбора информации (телефонные и/или радио каналы).

\* \* \*

В соответствии с концепцией проведения ВПН [2], для достижения одной из целей переписи — удержания стоимости ее проведения не выше среднемировых значений — необходимо обеспечить максимальный охват населения путем проведения интернет-переписи в каждом регионе России, а оставшуюся часть населения переписывать с применением планшетных компьютеров, что в целом приведет к уменьшению стоимости переписи в расчете на одного опрашиваемого.

Данную работу целесообразно организовать и провести на этапе формирования общего организационного плана Всероссийской переписи населения. Определение метода сбора и обработки информации переписи с помощью разработанной экспертной системы позволяет выбрать наиболее подходящие методы, на основе которых в дальнейшем в каждом регионе определяются требования к расчету ресурсного обеспечения, количеству привлекаемого персонала по каждой категории и рассчитываются потребности для формирования бюджетов проведения переписи населения в каждом регионе России.

С помощью разработанной экспертной системы можно осуществлять эффективный выбор метода сбора и обработки информации переписи населения в отдельно взятом регионе Российской Федерации. Алгоритм выбора

метода сбора и обработки информации переписи населения был реализован О. О. Манжулой в программе, разработанной на языке Visual C#, результат тестирования которой подтвердил работоспособность алгоритма.

Новизна при решении задачи выбора метода сбора и обработки информации заключается в построении базы знаний продукционных правил, используемых для определения наиболее предпочтительного метода в отдельном регионе России с учетом географических и социально-экономических особенностей, что позволяет оптимизировать расходы на проведение переписи, добиться максимального охвата населения, а также обеспечить максимальную достоверность данных о населении.

#### ЛИТЕРАТУРА

1. **Методология переписи: основные результаты проведенного ЕЭК ООН опроса по практике проведения странами переписей и первоначальные предложения в отношении рекомендаций КЕС по проведению цикла переписей 2020 года. Записка Целевой группы ЕЭК ООН по методологии переписи.** Европейская экономическая комиссия. Конференция европейских статистиков. Группа экспертов по переписям населения и жилищного фонда. Пятнадцатое совещание. Женева, 30 сентября – 3 октября 2013 года. Пункт 2 предварительной повестки дня.
2. СУРИНОВ А. Е. **Всероссийская перепись населения: опыт и перспективы** // Доклад на научно-практической конференции. Росстат, 27–28 ноября 2012 г. URL: [http://www.gks.ru/free\\_doc/new\\_site/perepis2010/croc/perepis\\_itog1612.htm](http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itog1612.htm).
3. БОЛОТОВА Л. С. **Системы искусственного интеллекта: модели и технологии, основанные на знаниях.** М.: Финансы и статистика, 2012.
4. РЫБИНА Г. В. **Основы построения интеллектуальных систем.** М.: Финансы и статистика (Инфра-М), 2010.
5. ТЕЛЬНОВ Ю. Ф., ТРЕМБАЧ В. М. **Интеллектуальные информационные системы.** М.: МЭСИ, 2010.
6. ВИНТИЗЕНКО И. Г., ЧЕРКАСОВ А. А. **Математические методы в экономике** // Вестник Адыгейского государственного университета. Серия 5: Экономика. Вып. № 2/2011.
7. АНДРЕЙЧИКОВ А. В., АНДРЕЙЧИКОВА О. Н. **Анализ, синтез, планирование решений в экономике.** М.: Финансы и статистика, 2000.
8. ТЕЛЬНОВ Ю. Ф. **Реинжиниринг бизнес-процессов: компонентная методология.** М.: Финансы и статистика, 2004.
9. РОМАНОВ А. Н., ОДИНЦОВ Б. Е. **Информационные системы в экономике: 2-е изд.** М.: Вузовский учебник, 2008.