

УДК 528:004.9

ГЕОСТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ ДЛЯ ОРГАНИЗАЦИИ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В СФЕРЕ ПЕЧАТИ И СРЕДСТВ МАССОВОЙ ИНФОРМАЦИИ НА ТЕРРИТОРИИ РЕСПУБЛИКИ БАШКОРТОСТАН

С. В. Павлов¹, О. А. Ефремова², А. Ф. Атнабаев³

¹psvgis@mail.ru, ²efremova-oa@yandex.ru, ³aaf1981@mail.ru,

¹ Заместитель председателя Комитета Государственного Собрания-Курултая РБ
по промышленности, инновационному развитию и предпринимательству

^{2,3} ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступила в редакцию 15.05.2017

Аннотация. Описываются ключевые моменты исследования, целью которого является решение проблемы анализа территориально распределенной информации о количестве подписчиков печатной продукции Агентства по печати и средствам массовой информации Республики Башкортостан, а также осуществление информационной поддержки принятия решений на основе этих данных за счет внедрения использования ГИС-технологий в органы исполнительной власти.

Ключевые слова: геоинформационная система (ГИС); космические снимки; прогнозирование паводковой ситуации; система поддержка принятия решений

ВВЕДЕНИЕ

Одной из важнейших задач по управлению жизнедеятельностью субъекта Российской Федерации является контроль и регулирование в сфере средств массовой информации. На территории Республики Башкортостан в составе Правительства эти функции выполняет Агентство по печати и средствам массовой информации (в дальнейшем АПиСМИ), к основным задачам которого относятся:

– реализация основных направлений государственной политики Республики Башкортостан в сфере массовых коммуникаций, издательско-полиграфической деятельности, распространения печатной продукции;

– реализация мер, обеспечивающих законодательную защиту прав и интересов организаций в сфере массовых коммуникаций, издательско-полиграфической деятельности, распространения печатной продукции, находящихся в ведении Агентства;

– содействие экономической эффективности работы организаций, осуществляющих деятельность в сфере массовых коммуникаций, издательско-полиграфической деятельности, распространения печатной продукции [1].

В состав Агентства по печати и средствам массовой информации Республики Башкортостан

входят одно казенное предприятие (КП РБ Редакции газет и журналов) и семь редакционно-издательских комплексов (ГУП РБ РИК).

В процессе работы у сотрудников АПиСМИ накапливаются огромные объемы различных статистических данных, которые им необходимо обрабатывать. Для повышения качества информационной поддержки принятия решений сотрудникам АпиСМИ необходимо иметь актуальные и непротиворечивые данные, уметь правильно их анализировать, а также предоставлять результаты анализа не только в текстовом, но и в графическом виде, для более наглядного представления (графики, диаграммы, таблицы и др.).

Одной из важных задач, возникающих в процессе организации поддержки принятия решений в сфере средств массовой информации, является получение, анализ, классификация и визуализация территориально распределенной информации о количестве подписчиков в районах Республики Башкортостан. Одним из результатов анализа информации для поддержки принятия решений являются статистические данные, представляемые наглядно в виде тематических карт. Для представления результатов в виде картографического материала, а именно тематических карт, необходимо использовать геоинформационные системы, т.к. они специа-

лизируются на обработке и представлении, как статистических, так и пространственных данных в виде карт. Именно ГИС-технологии позволяют наиболее точно анализировать территориально распределенную информацию и визуализировать ее в простой и доступной форме.

Исходя из вышесказанного, основной целью исследования является решение проблемы анализа территориально распределенной информации о количестве подписчиков печатной продукции Агентства по печати и средствам массовой информации Республики Башкортостан, а также осуществление информационной поддержки принятия решений на основе этих данных за счет использования ГИС-технологий.

Таким образом, для достижения поставленной цели требуется решить следующие задачи:

- 1) провести системный анализ предметной области;
- 2) осуществить формализацию бизнес процессов организации поддержки принятия решений в сфере средств массовой информации Республики Башкортостан;
- 3) разработать структуру информационной модели поддержки принятия решений в сфере средств массовой информации Республики Башкортостан;
- 4) провести анализ существующих методов классификации статистических данных и разработку модифицированного метода классификации данных о количестве подписчиков в районах Республики Башкортостан;
- 5) выполнить реализацию предложенных методов и алгоритмов.

СИСТЕМНЫЙ АНАЛИЗ ДЕЯТЕЛЬНОСТИ АГЕНТСТВА ПО ПЕЧАТИ И СРЕДСТВАМ МАССОВОЙ ИНФОРМАЦИИ РЕСПУБЛИКИ БАШКОРТОСТАН

За период с 1990 по 2016 г. республиканские средства массовой информации претерпели значительные изменения. За 24 года башкирская пресса превратилась в мощную общественную силу, объединенную единым информационным контентом. Национальная печать сегодня успешно выполняет возложенные на нее задачи, обеспечивая право населения на достоверную информацию на родных языках об общественно-политических, социально-экономических событиях в республике, о культуре населяющих ее народов, об их истории и перспективах на будущее.

Всему этому предшествовала большая работа, так в 90-е гг. прошлого века происходило

становление современного варианта национальной прессы, велась работа по регистрации средств массовой информации, лицензированию полиграфпредприятий и издательств. За эти годы в Республике Башкортостан появились 24 новых периодических печатных издания, в том числе 13 республиканских средства массовой информации на башкирском языке (журналы «Акбузат», 1991 г.; «Шонкар», 1995 г.; «Тамаша», 1995 г.); на русском языке (газета «Истоки», 1990 г.; журналы «Бельские просторы», 1998 г.; «Рампа», 1994 г.); на татарском языке (газета «Омет», 1991 г.; журнал «Тулпар», 1994 г.); на чувашском языке (газета «Урал саси», 1990 г.); на марийском языке (газета «Чолман», 1993 г.); на удмуртском языке (газета «Ошмес», 1999 г.); на трех языках – башкирском, русском и английском стали издаваться журналы «Ватандаш» (1996 г.) и «Панорама Башкортостана» (2007 г.).

В рамках системного анализа предметной области установлено, что Агентство по печати и средствам массовой информации Республики Башкортостан является республиканским органом исполнительной власти, реализующим в пределах своей компетенции государственную политику и регулирование в сфере массовых коммуникаций, издательско-полиграфической деятельности, распространения печатной продукции. Структура Агентства по печати и средствам массовой информации Республики Башкортостан включает (рис. 1): внутреннюю структуру управления Агентства (руководитель, заместитель руководителя, организационно-производственный отдел, отдел государственной службы, отдел печатных средств массовой информации, финансово-экономический отдел, отдел электронных средств массовой информации), а также взаимодействующие с ней организации (печатные издания, телеканалы, радиоканалы, службы доставки, типографии, сетевые партнеры). Одним из результатов системного анализа является разработанная функциональная модель информационной поддержки принятия решений с использованием ГИС Агентства по печати и средствам массовой информации РБ в нотации BPMN (рис. 2).

На основе полученных результатов – структуры Агентства по печати и средствам массовой информации, функциональной и логической моделей – была разработана информационная модель специализированных пространственных данных ГИС Агентства по печати и средствам массовой информации (рис. 3) и спроектирован прототип ГИС АПиСМИ.

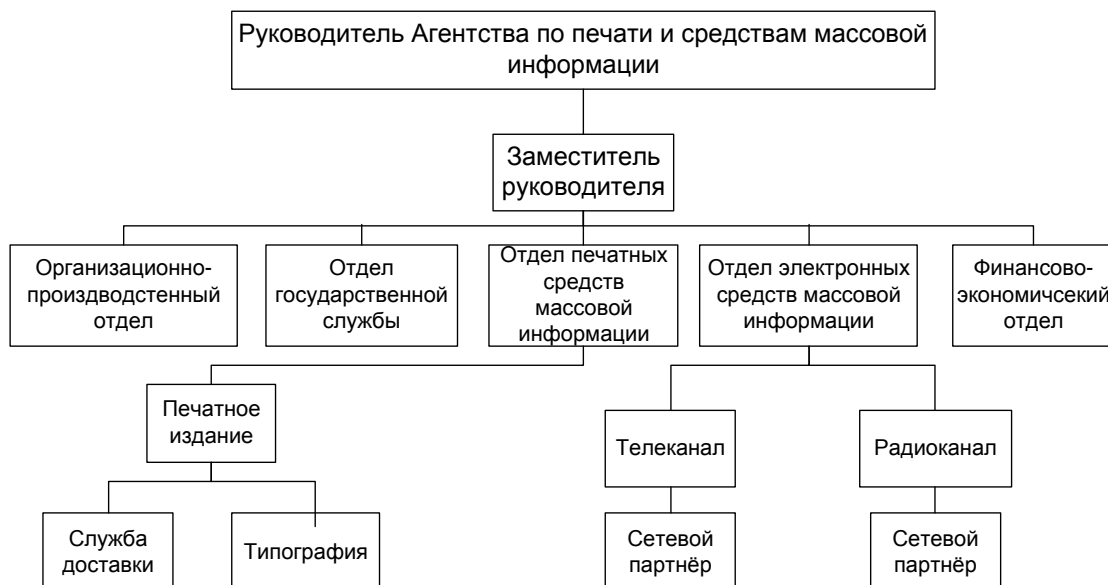


Рис. 1. Структура Агентства по печати и средствам массовой информации Республики Башкортостан

Основными задачами ГИС АПиСМИ являются:

1) надежное и своевременное предоставление информации об объектах печатных изданий, расположенных на территории РБ;

2) предоставление информации о количестве подписчиков печатной продукции;

3) поддержка принятия решений связанных с деятельностью районных отделений Агентства по печати и средствам массовой информации.

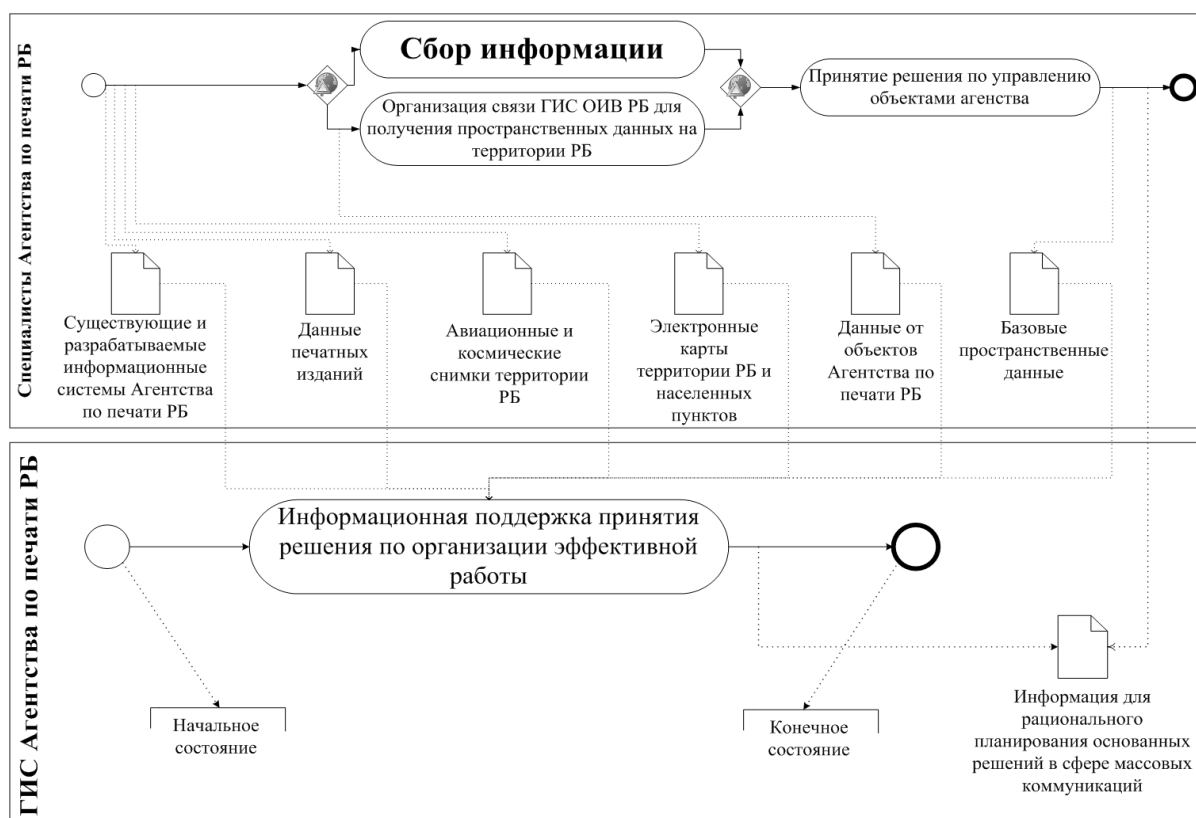


Рис. 2. Функциональная модель информационной поддержки принятия решений с использованием ГИС Агентства по печати и средствам массовой информации Республики Башкортостан

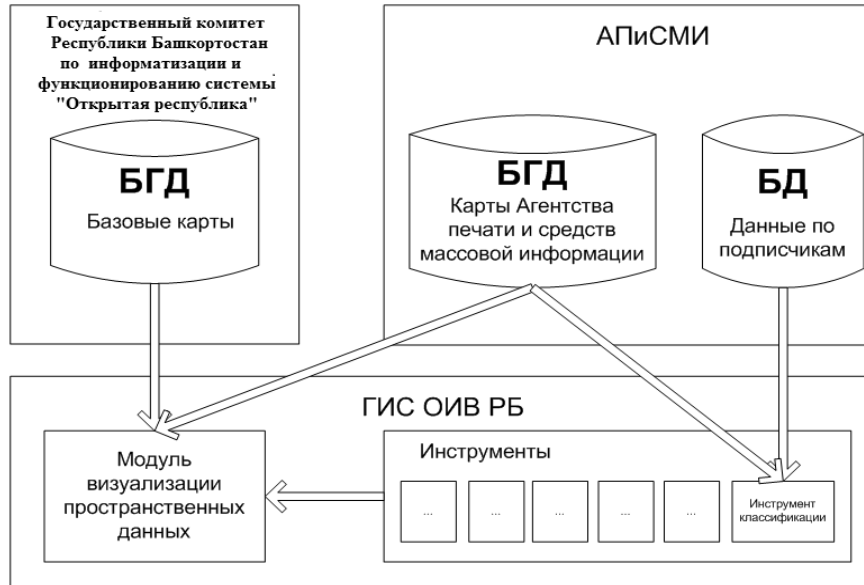


Рис. 3. Логическая структура данных ГИС Агентства по печати и средствам массовой информации Республики Башкортостан

ГИС Агентства по печати и средствам массовой информации Республики Башкортостан представляет собой сложную многофункциональную систему, логическая структура которой представлена на рис. 4. Поскольку разрабатываемая ГИС Агентства по печати и средствам мас-

совой информации РБ базируется на программно-аппаратном комплексе Государственного комитета Республики Башкортостан по информатизации и вопросам функционирования систем «Открытая Республика», а также системном и специальном программном обеспечении,

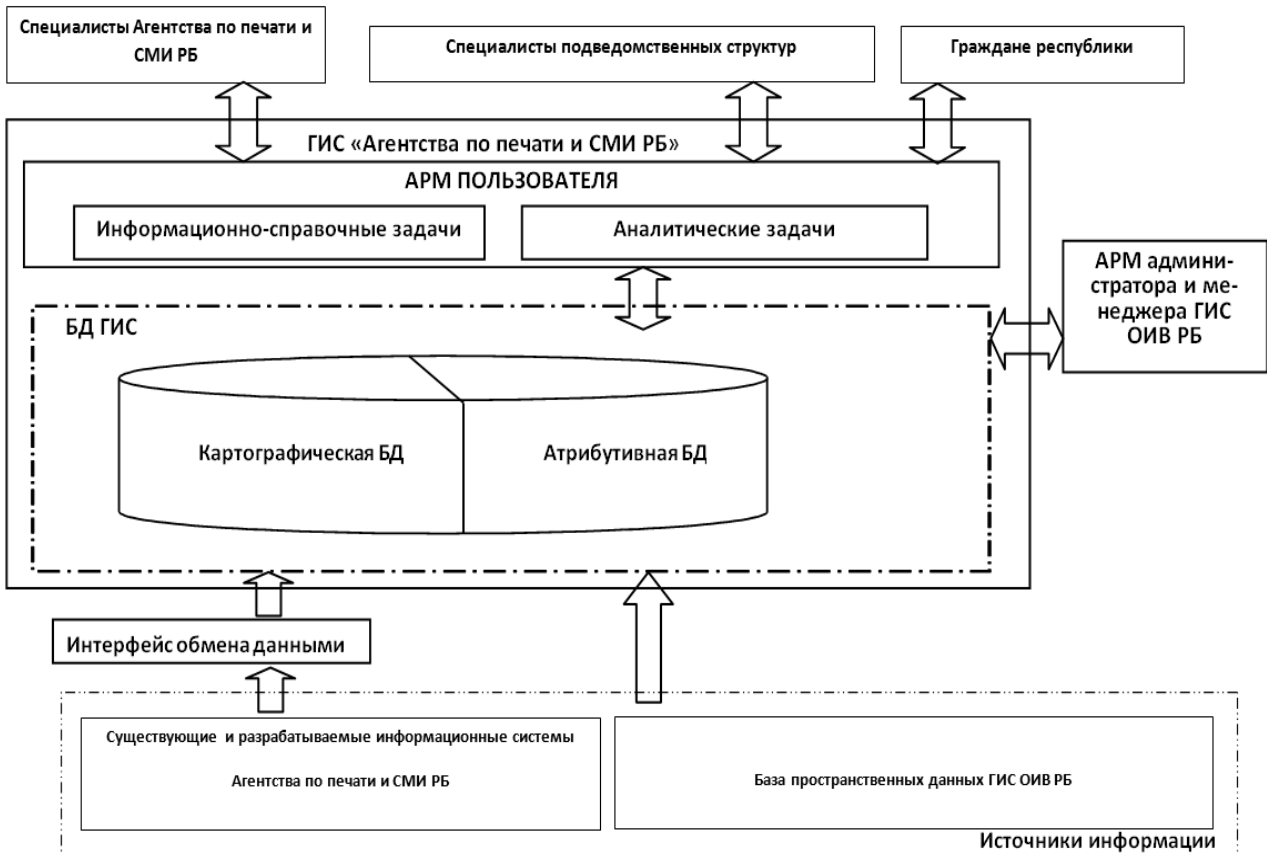


Рис. 4. Логическая структура ГИС «Агентства по печати и СМИ РБ»

то при разработке первой очереди ГИС учитывалась совместимость разрабатываемой ГИС АПиСМИ и существующей Геоинформационной системы органов исполнительной власти республики Башкортостан (в дальнейшем ГИС ОИВ РБ). ГИС АПиСМИ является функциональной частью ГИС ОИВ РБ и обеспечивает единство подходов к хранению, передаче и обработке информации, архитектуре, номенклатуре технических средств и средств общесистемного программного обеспечения.

АНАЛИЗ И КЛАССИФИКАЦИЯ ДАННЫХ АПИСМИ

Основываясь на данных, полученных в результате системного анализа Агентства по печати и средствам массовой информации Республики Башкортостан, можно сделать вывод о том, что на сегодняшний день ключевой задачей, для Агентства в рамках АПиСМИ, требующей реализации средствами ГИС является задача классификации данных по количеству подписчиков. Существуют стандартные методы классификации [2, 3]:

- метод равных интервалов;
- метод заданных интервалов;
- квантиль;
- метод естественных границ;
- метод стандартного отклонения.

Кратко охарактеризуем методы и их достоинства и недостатки. Метод равных интервалов разбивает диапазон значений атрибута на поддиапазоны равного размера. Метод заданных интервалов позволяет указать размер интервала, который будет использован для определения последовательности классов с одинаковым диапазоном значений. По сути, является тем же методом, что и равные интервалы, только в случае равных интервалов задается количество классов, а в случае заданного интервала задается шаг значений. При использовании метода квантиль каждый класс содержит одинаковое число объектов. Такая классификация хорошо подходит для линейно распределенных данных, так как в каждый класс назначается одинаковое количество данных. Здесь не бывает пустых классов, или классов, содержащих слишком малое или слишком большое количество значений. Поскольку объекты сгруппированы по принципу их одинакового количества в каждом классе, полученная карта может ввести в заблуждение. Похожие объекты могут попасть в разные классы, а объекты с существенно разными значениями могут оказаться в одном.

В методе естественных границ классы основаны на естественном группировании данных: границы классов определяются таким образом, чтобы сгруппировать схожие значения и максимально увеличить различия между классами. Объекты делятся на классы, границы которых устанавливаются там, где встречаются относительно большие различия между значениями данных. Классификация методом стандартного отклонения, также называемого среднеквадратическим отклонением, показывает, насколько значения атрибутов объектов отличаются от среднего значения. Границы классов строятся с равными диапазонами значений, пропорциональными стандартному отклонению – обычно в интервалах 1, 0,5 или 0,25 среднеквадратического отклонения, используя средние значения и стандартное отклонение от среднего [4, 5].

В рамках исследования проведен анализ того, какую статическую информацию хотят видеть сотрудники Агентства по печати и средствам массовой информации Республики Башкортостан. В результате анализа выявлено, что для общего анализа ситуации, зачастую необходимо определить не отклонение от среднего, а всего лишь несколько лучших и несколько худших позиций. Такая статистика не требует тяжелых вычислений и итоговый вариант включает в себя три класса объектов: первый класс содержит несколько объектов с наибольшим количеством подписчиков печатных изданий (10 объектов), второй класс содержит объекты с наименьшим количеством подписчиков (10 объектов) и в третий класс входят все оставшиеся объекты.

Исходя из этого, предложен модифицированный метод классификации данных, идея которого заключается в том, что берется информация о количестве подписчиков печатных изданий в каждом районе республики и в результате анализа отображается информация в каких районах самое большое и самое малое количество людей оформивших подписку, т.е. интересуют только два крайних класса.

Формально задачу можно описать следующим образом.

Электронные карты территории РБ представляет собой совокупность слоев, каждый из которых содержит простые и сложные объекты, описываемые или определяемые своими координатами. Исходя из этого, цифровая карта местности может быть представлена в виде множества тематических слоев, необходимых для описания и анализа всей требуемой пространственной информации:

$$Map = \{ C_i \}, \quad i = \overline{1, n}. \quad (1)$$

Под слоем понимается совокупность однотипных пространственных объектов, относящихся к одной теме (классу объектов), в пределах некоторой территории и в единой системе координат:

Так, для точечных объектов (например, издательств) справедливо соотношение:

$$C_{Pnt} = \{((x^{Pnt_j}, y^{Pnt_j}), Atr^{Pnt_j})_j, \quad j = \overline{1, n_{Pnt}}\}. \quad (2)$$

А для полигональных объектов (например, районы РБ) справедливо соотношение:

$$C_{Pol} = \{(\{(x, y)_q\}_t, label_{Pol_j}, Atr^{Pol_j})_j, \quad j = \overline{1, n_{Pol}}, \quad t = \overline{1, b_j}, \quad q = \overline{1, c_t}\}, \quad (3)$$

где $Atr = \{atr_1, atr_2, \dots, atr_n\}$ – совокупность атрибутивных характеристик каждого объекта, а (x, y) – координаты объектов.

Атрибутами данных объектов могут являться множества X – количество подписчиков и G – года, для которых есть данные по подписчикам:

$$X = \{x_1, x_2, \dots, x_n\}, \quad (4)$$

где x_n – это количество подписчиков, а $n \in [1, \infty]$;

$$G = \{g_1, g_2, \dots, g_n\}, \quad (5)$$

где g_n – это года, а $n \in [1, \infty]$.

К точечным объектам относятся издательства Y :

$$Y = \{y_1, y_2, \dots, y_n\}, \quad (6)$$

где y_n – это издательства, а $n \in [1, \infty]$.

А к полигональным – районы республики R :

$$R = \{r_1, r_2, \dots, r_n\}, \quad (7)$$

где r_n – это районы республики, а $n \in [1, \infty]$.

Исходя из введенных соотношений (1)–(7), справедливыми является соотношения:

$$\{(y, \{g_1, g_2, \dots, g_n\}, \{x_1, x_2, \dots, x_n\}) : x \in X, y \in Y, \quad g \in G, n \in [1, \infty]\}, \quad (8)$$

$$\{(r, \{y_1, y_2, \dots, y_n\}) : r \in R, y \in Y, n \in [1, \infty]\}. \quad (9)$$

Соотношение (8) показывает, что для каждого издательства существует некоторый набор лет, в которые это издательство работало, и в каждый год было свое число подписчиков.

Из соотношения (9) видно, что для каждого района существует некоторый набор издательств, которые работают в нем.

РЕАЛИЗАЦИЯ ПРЕДЛОЖЕННЫХ МЕТОДОВ И АЛГОРИТМОВ

В рамках реализации предложенного модифицированного метода классификации разработан алгоритм, который представлен на рис. 5 и реализован в виде картографического веб-сервиса с использованием сервис-ориентированного подхода. Сервис-ориентированный подход – это концепция построения архитектуры корпоративной информационной системы из слабосвязанных между собой частей на основе сервисов – отдельных компонентов с фиксированными интерфейсами, выполняющих определенные функции. Ключевым понятием сервис-ориентированного подхода являются интерфейсы. В интерфейсе сервиса определены параметры обращения к нему и описан результат, то есть интерфейс определяет суть сервиса, а не технологию его реализации. Сервис-ориентированный подход – это концепция построения архитектуры корпоративной информационной системы из слабосвязанных между собой частей на основе сервисов – отдельных компонентов с фиксированными интерфейсами, выполняющих определенные функции. Ключевым понятием сервис-ориентированного подхода являются интерфейсы. В интерфейсе сервиса определены параметры обращения к нему и описан результат, то есть интерфейс определяет суть сервиса, а не технологию его реализации. Сервис-ориентированный подход предлагает единую схему взаимодействия сервисов независимо от того, находится ли сервис в том же самом приложении, в другом адресном пространстве многопроцессорной системы, на другой аппаратной платформе в корпоративной intranet-сети или в приложении, развернутом на ИТ-площадке партнера.

На рис. 6 представлен результат работы разработанного веб-сервиса, который анализирует данные на основе предлагаемого выше модифицированного метода, и визуализация результатов анализа для Республики Башкортостан, по данным о количестве подписчиков всех региональных изданий, в результате получены 3 класса: в первый класс входит 10 изданий с лучшим показателем, в третий – 10 изданий с худшим показателем количества подписчиков, а во второй класс включены все остальные издания, так называемые «среднячки».

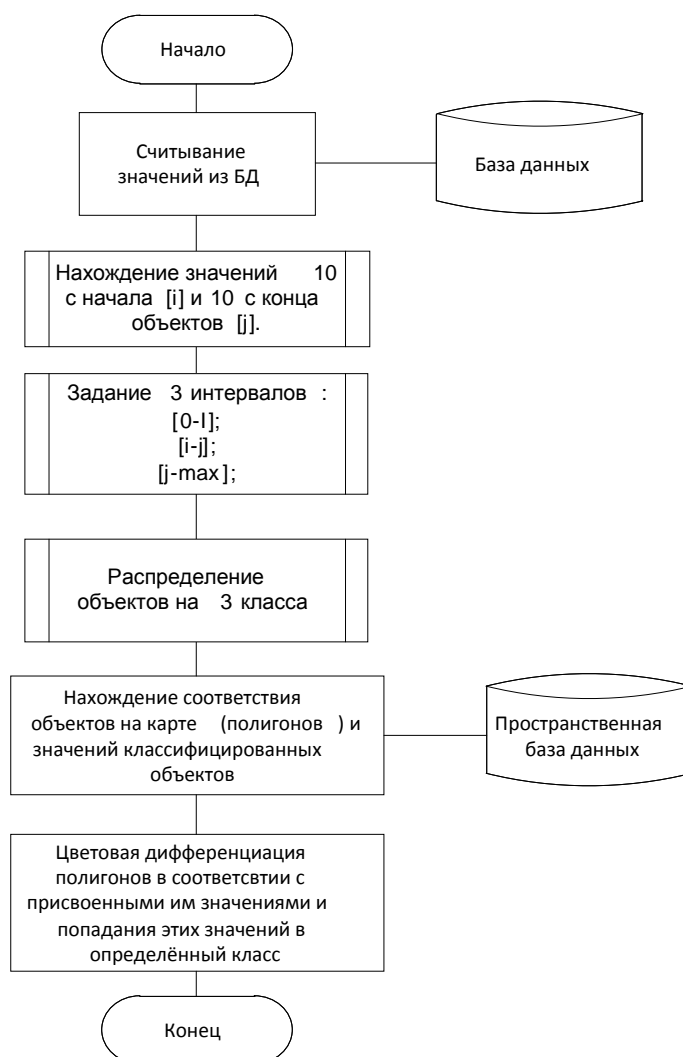


Рис. 5. Блок-схема алгоритма классификации

ЗАКЛЮЧЕНИЕ

В ходе работы проведен системный анализ Агентства по печати и средствам массовой информации, в результате которого были разработаны структура Агентства, функциональная модель информационной поддержки принятия решений, логическая структура данных, а так же логическая структура Геоинформационной системы Агентства по печати и средствам массовой информации.

Для организации поддержки принятия решений разработан модифицированный алгоритм анализа и классификации данных о подписчиках печатных изданий в районах Республики Башкортостан. Алгоритм внедрен в ГИС АПИСМИ и применен к реальным данным. Результатом работы алгоритма является карта районов республики, на которой районы разделены на клас-

сы и дифференцированы по цветовому признаку в соответствии с количеством подписчиков в каждом районе, что позволило снизить затраты времени на анализ информации о подписчиках и сделать выводы о качестве работы районных отделений АПИСМИ, тем самым повысить качество работы Агентства и подведомственных ему организаций.

СПИСОК ЛИТЕРАТУРЫ

1. Гвоздев В. Е., Семененко Д. В. Информационная поддержка анализа состояния территориальных систем по разнотипным признакам // Вестник УГАТУ. 2009. Т. 12, № 1, С. 9–16. [V. E. Gvozdev, D. V. Semenenko, "Information support of the analysis of the state of territorial systems on different types of signs", (in Russian), in *Vestnik UGATU*, vol. 12, no. 1(30), pp. 9-16, 2009.]

2. Павлов С. В., Ефремова О. А., Ямалов И. У. Интеграция пространственной информации в Геоинфор-

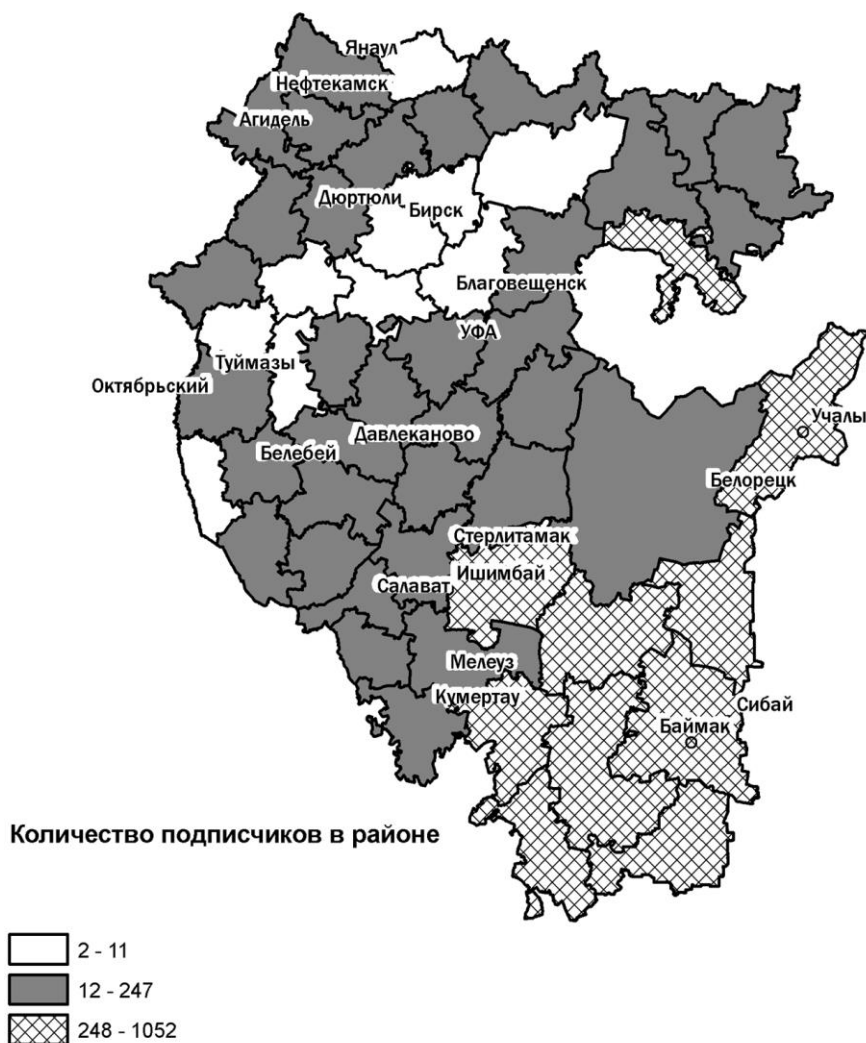


Рис. 6. Классификация тиража региональных изданий Республики Башкортостан модифицированным методом

мационной системе органов исполнительной власти на основе сервис-ориентированной архитектуры // Вестник УГАТУ. 2013. Т. 17, №. 5. С. 129–139. [S. V. Pavlov, O. A. Efremova, I. U. Yamalov, "Integration of spatial information in the Geoinformation System of Executive Authorities on the Basis of Service-Oriented Architecture", (in Russian), in *Vestnik UGATU*, vol. 17, no. 5(58), pp. 129-139, 2013.]

3. **Статистическое** исследование территориальных систем / Гузайров М. Б. [и др.]. М.: Машиностроение, 2008. 187 с. М. Guzairov, et. al. [*Statistical study of territorial systems*. М.: Mashinostroenie, 2008.]

4. **Geospatial Analysis-A Comprehensive Guide**, 3rd edition; © 2006–2009; de Smith, Goodchild, Longley. URL: http://www.spatialanalysisonline.com/HTML/?classification_and_clustering.htm (дата обращения: 29.01.2014). [(2014, Jan. 01). *Geospatial Analysis-A Comprehensive Guide*. [Online]. Available: www.spatialanalysisonline.com/HTML/?classification_and_clustering.htm]

5. **Дэйвисон М.** Многомерное шкалирование. Методы наглядного представления данных. М.: Финансы и статисти-

стика, 1987. 254 с. [M. Davison, *Multidimensional scaling. Methods of visual representation of data*, М.: Finance and Statistics, 1987.]

ОБ АВТОРАХ

ПАВЛОВ Сергей Владимирович, проф. каф. геоинформ. систем. Дипл. математик по вычислительн. математике (БГУ, 1977). Д-р техн. наук (УГАТУ, 1998). Иссл. обл. геоинформационных систем.

ЕФРЕМОВА Оксана Александровна, доц. каф. геоинформ. систем Дипл. инж. по автоматизир. сист. обраб. инф. и управл. (УГАТУ, 1999). Канд. техн. наук (УГАТУ, 2003). Иссл. в обл. геоинформационных систем.

АТНАБАЕВ Андрей Фарагатович, доц. каф. геоинформ. систем. М-р техн. и технол. (УГАТУ, 2004). Канд. техн. наук (УГАТУ, 2007). Иссл. в обл. применения ГИС-технологий при разработке корпоративных систем и обработки данных ДЗЗ.

METADATA

Title: Geostatistical data processing for the organization of support of decision-making in the sphere of press and mass media on the territory of the Republic of Bashkortostan

Authors: S. V. Pavlov¹, O. A. Efremova², A. F. Atnabaev³

Affiliation: ¹Deputy Chairman of Committee of the State Assembly-Kurultay RB

^{2,3}Ufa State Aviation Technical University (UGATU), Russia.

Email: ¹psvgis@mail.ru, ²efremova-oa@yandex.ru, ³aaf1981@mail.ru

Language: Russian.

Source: Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 21, no. 2 (76), pp. 103-111, 2017. ISSN 2225-2789 (Online), ISSN 1992-6502 (Print).

Abstract: Describes the key points of the study, the purpose of which is to solve the problem of the analysis of territorially distributed information about the number of subscribers of printed materials of the Agency for press and mass media of the Republic of Bashkortostan, as well as the implementation of informational support of decision-making on the basis of these data due to the introduction of GIS technology in the Executive branch

Key words: geographic information system (GIS); satellite images; forecasting of flooding; a system to support decision-making

About authors:

PAVLOV, Sergey Vladimirovich, Prof., Dept. of Geoinformation Systems. Dipl. Mathematician (Bashkir State Univ., 1977). Dr. of Tech. Sci. (UGATU, 1998).

EFREMOVA, Oksana Aleksandrovna, Associate Professor, Dept. of Geoinformational Systems. Cand. (PhD) Tech. Sci. (UGATU, 2003).

ATNABAEV, Andrey Faragatovich, docent., Dept. of Geographic Information System. Dipl. master (USATU., 2004). Cand. of Tech. Sci. (UGATU, 2007). Researcher in the use of GIS technology in the development of enterprise systems