

УДК 502.5:004.891

А. А. Платонов, Б. Х. Санжапов

АВТОМАТИЗИРОВАННАЯ СИСТЕМА СБОРА И АНАЛИЗА ДАННЫХ ПРИ ПЛАНИРОВАНИИ РАЗВИТИЯ ГОРОДСКИХ ТЕРРИТОРИЙ

Предлагается структура и основные пути реализации системы сбора, анализа и обработки информации на естественном языке, относящейся к экологическому мониторингу, в целях планирования безопасного развития городских и крупных сельских поселений. Также рассматривается интеграция данной системы с системой поддержки принятия решений в данной области.

Ключевые слова: геоинформационные системы, экологический мониторинг, системы поддержки принятия решений, обработка текстов на естественном языке.

The authors offer the structure and ways of realization of system of collection, analysis and processing of information in natural languages. It refers to environmental monitoring in order to plan safe development of urban and large agricultural settlements. The authors also consider the integration of this system with decision support system in this area.

Key words: geo-information system, environmental monitoring, decision support system, processing of texts in natural language.

Бурный рост экономик развитых стран вызвал объективную необходимость введения экономного механизма природопользования и охраны окружающей среды. Учитывая общемировую и российскую экологическую действительность, существующие тенденции и угрозы, можно сказать, что общество и государство попытались совместить формирование государственно-рыночного механизма, охраны окружающей природной среды и обеспечения экологической безопасности в единый экономический механизм экологически устойчивого развития.

Понятия «устойчивость» и «безопасность» являются важнейшими характеристиками экономики как единой системы. Под «устойчивостью» экономики как системы подразумевается прочность и надежность ее элементов, способность выдерживать внутренние и внешние нагрузки, создание условий повторяемости процесса и сохранения условий для жизнедеятельности человека. В то же время «безопасность» — это состояние объекта (региона, города, любой территории) с точки зрения способности к нормальному развитию в условиях нейтрализации внутренних и внешних угроз, а также действия непредсказуемых и трудно прогнозируемых факторов, неблагоприятно влияющих на состояние объекта. Чем устойчивее любая экономическая система, чем жизнеспособнее экономика с точки зрения ее безопасности, тем больше возможностей обеспечить экологическую безопасность.

Структуру развития любого региона России во многом определяют: природно-ресурсная база, исторически унаследованная производственная инфраструктура, особенности климата, количество и качество земельных ресурсов, исторически сложившаяся специализация, качество и количество трудовых ресурсов, ближайшее внешнее окружение (соседние регионы). При этом выбор цели развития региона во многом влияет на формирование рационального механизма природопользования в целом.

Появляется необходимость переходить к моделям социально-экономического развития, которые учитывали бы экологическую и иную специфику региона и тем самым обеспечивали в нем равновесие между экономикой и экологией, и устойчивое экономическое, экологическое, социальное развитие. Становится очевидным, что не может быть универсального для всех регионов экологически-ориентированного экономического механизма, поскольку каждый регион имеет свой набор и комбинацию природных ресурсов, уровень вовлечения их в хозяйственную деятельность и состояние окружающей естественной среды обитания.

Следовательно, необходимо разработать (проанализировать) программы развития регионов при условии обеспечения экологической безопасности для равновесия между экономикой и экологией, и устойчивого экономического, экологического, социального развития.

Встающие в настоящее время проблемы анализа текущей экологической ситуации в масштабе России и ее отдельных регионов, а также планирования развития городских и областных территорий делают актуальной задачу объединения фактов и статистики из разнородных источников в единую, централизованную систему¹. К тому же для удобства обработки и расширения возможностей анализа перспективным является привязка таких данных к карте реальной местности, что приводит к необходимости использования, с одной стороны, современной геоинформационной системы, а с другой — построения анализатора текстов на естественном языке и разработке модели для хранения извлеченных знаний, поддерживающей интеграцию с такой системой.

Развитие геоинформационных систем (ГИС) привело к появлению большого количества разнообразных по возможностям и целям баз данных, среди которых выделяются специализированные системы для задач экологического мониторинга и перспективного планирования развития городских территорий.

Существуют достаточно развитые системы, обладающие многофункциональными интерфейсами прикладных программ (API), причем многие из них могут с успехом применяться как на федеральном, так и на региональном уровне [1]. История использования таких систем в задачах регионального управления Волгоградской областью показывает, что существующие системы развивались от локальных (однопользовательских) к интранет-версиям. Такой подход демонстрирует очевидные преимущества с точки зрения безопасности и гибкости (поскольку системы разрабатываются, в значительном большинстве, собственными отделами разработки). В то же время, он обладает принципиальными имманентными недостатками, существенно затрудняющими дальнейшее их развитие, а именно:

- закрытость большей части данных для общего доступа;

- отсутствие открытых, стандартных и документированных API для работы с данными;

- недостаточная актуальность данных, вызванная их устареванием и отсутствием механизмов общественного редактирования на основе wiki-технологий;

¹ Электронный атлас Волгоградской области [Электронный ресурс] // URL: <http://maps.volgnet.ru> (дата обращения: 28.03.2014).

слабая масштабируемость систем, вызванная изначальным проектированием их как локальных и внутрисетевых, что приводит к затруднениям в интеграции с системами федерального значения, а также ставит под вопрос предоставление доступа к ним значительному количеству пользователей.

Описанные выше недостатки отсутствуют в современных открытых ГИС, таких как OpenStreetMap. Подобные системы обеспечивают высокую актуальность данных и простые средства для работы с ними, несмотря на ряд собственных специфических недостатков:

возможная недостоверность вносимых пользователями данных;

сложности коммерческого использования данных, связанные со спецификой лицензирования;

проблема наполнения ГИС специфическим прикладным контентом, таким как данные об экологической обстановке (в силу ограниченного доступа к этой информации и низкой востребованности ее пользователями).

Целесообразным представляется создание гибридного решения [2, 3], объединяющего в себе преимущества описанных выше типов систем и снижающих влияние этих недостатков на конечный продукт.

В предлагаемом проекте информационной системы картографические данные Волгоградского региона берутся из открытой ГИС, доступ к ним осуществляется через стандартные API этой системы. Данные об экологической обстановке вносятся вручную либо при помощи *Модуля извлечения знаний* из открытых источников. Далее они обрабатываются при помощи *Модуля поддержки принятия решений* и визуализируются на карте вновь при помощи стандартного API.

Таким образом, на верхнем уровне структура предлагаемой системы может быть описана при помощи диаграммы потоков данных (рис. 1).

Рассмотрим перечисленные модули более подробно. Реализация двух из них — хранилища данных и визуализации — хорошо известна и описана в многочисленной литературе. Как наиболее широко используемый, предлагается вариант реляционной СУБД в качестве основного хранилища, а для визуализации — использование встроенных в ГИС средств. Наибольший интерес представляет реализация трех оставшихся основных модулей системы.

Функциональность поискового модуля должна обеспечивать поиск разнородной информации об экологической обстановке региона по конкретным критериям (список которых формируется исследователями). Такая неструктурированная или слабо структурированная информация включает: открытые базы данных муниципальных ведомств, новости СМИ, научные статьи и различные специализированные системы «народного мониторинга» [4—6]. В качестве основы для данного сборщика данных может выступать любая из основных современных поисковых систем (Google, Yandex), обладающая необходимым для создания модуля API. Конкретная реализация зависит от выбора системы, однако возможности всех доступных в настоящее время систем удовлетворяют таким требованиям.

Для создания модели региона, пригодной для дальнейшего наполнения фактами и их последующего анализа представляется необходимым построение иерархической модели для классификации фактов по их географической принадлежности (рис. 2).

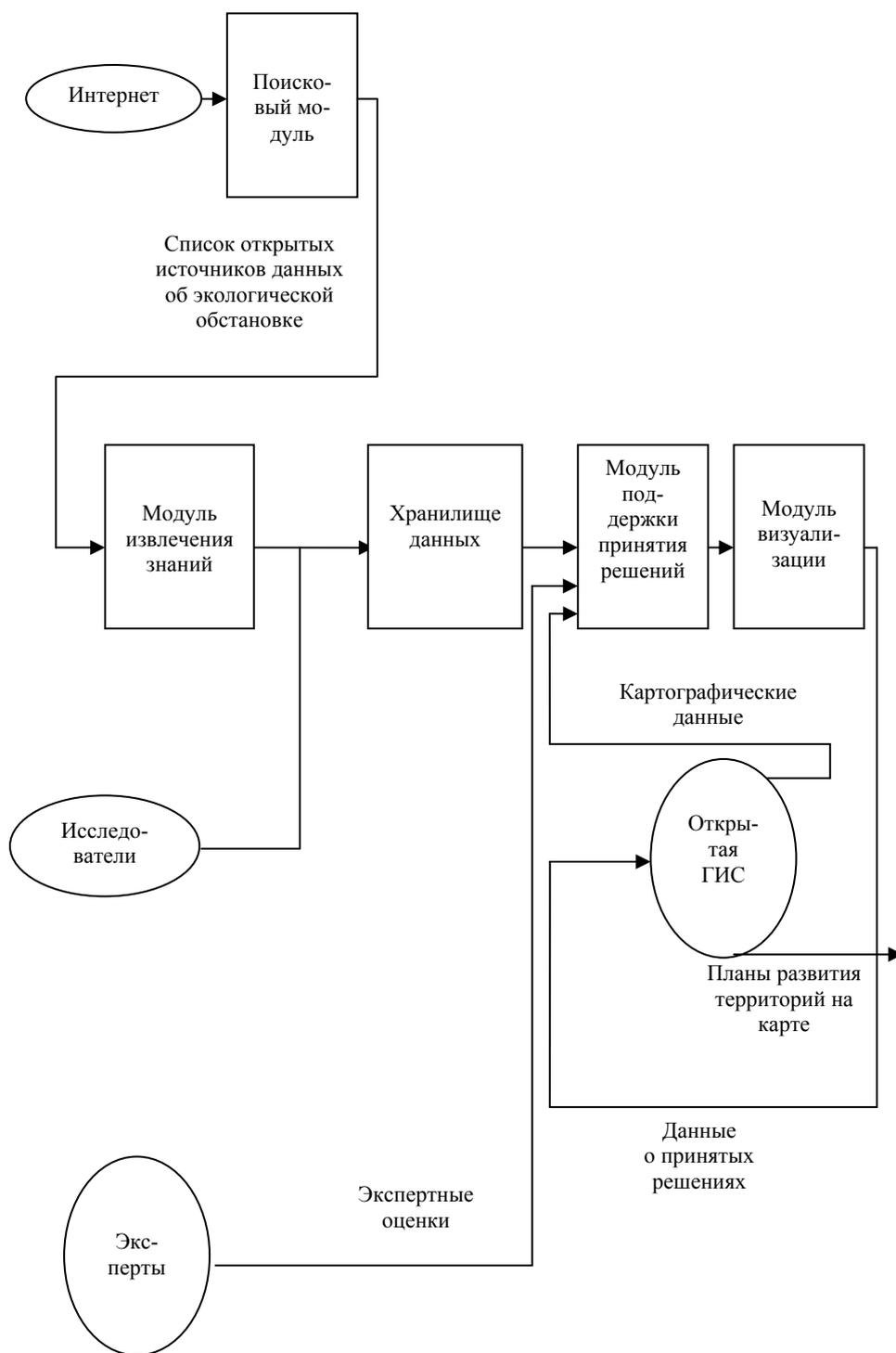


Рис. 1

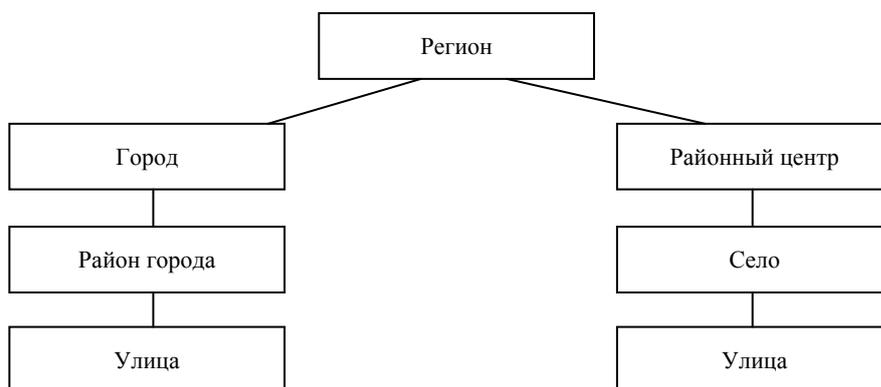


Рис. 2

Иерархическое хранилище данных такого вида далее заполняется конкретными названиями соответствующих объектов вкпе с их географическими координатами. Далее в хранилище выделяются «слои», относящиеся к тому или иному исследуемому показателю. При этом добавление новых фактов, относящихся к данному показателю, вызывает его изменение и на вышестоящих уровнях иерархии [7].

Кроме того, перспективным представляется введение временной составляющей в вышеописанную модель данных, что позволит не только анализировать статическое состояние, но и производить пространственно-временной анализ для выявления краткосрочных и длительных трендов, а также проводить интерполяцию и прогноз по имеющимся (возможно, разрозненным) данным [8, 9].

Разработка модуля извлечения знаний должна производиться в соответствии с известными методиками извлечения данных (data mining). В настоящее время нет универсальных способов извлечения и обработки неструктурированной информации, хотя существует программное обеспечение общего назначения, реализующее, в числе прочего, также и данную функциональность, например, Томита [2]. Отдельной важной задачей при проектировании данного модуля является привязка полученных фактов к координатной сетке. Хотя такая задача не имеет решения в общем случае, уже разработаны и используются методы ее решения, дающие удовлетворительный результат [10, 11].

Модуль поддержки принятия решений должен обеспечивать следующие основные функции:

- реализацию преобразования входных данных (информация об экологической обстановке с привязкой к географическим координатам) в удобный для их анализа и обработки формат;

- ввод планируемых изменений в систему (планируемые к застройке территории, выделение площадей под полигоны бытовых и промышленных отходов, создание или закрытие потенциально опасных производственных предприятий) [12];

- ввод экспертных оценок влияния данных изменений на экологическую обстановку;

- обработку экспертных оценок с использованием алгоритмов нечеткой логики;

прогноз изменений экологической обстановки с привязкой к координатам; обратное преобразование данных в формат ГИС для последующей передачи в модуль визуализации.

Разработана структура и намечены пути реализации системы, позволяющей проводить сбор информации об экологической ситуации на естественном (русском) языке с привязкой к геолокациям [13, 14], переводить эту информацию в машинно-читаемый вид, имеющей интерфейс для подключения внешних экспертных групп и позволяющей визуализировать результаты анализа при помощи ГИС-систем.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Jaiswal A., Pezanowski S., Mitra P., Zhang X., Xu S., Turton I., Klippel A., MacEachren A. M.* GeoCAM: A geovisual analytics workspace to contextualize and interpret statements about movement // *Journal of Spatial Information Science*. 2011. Vol. 3. Pp. 65—101. doi: 10.5311/JOSIS.2011.3.55
2. *Санжапов Б. Х., Садовникова Н. П.* Согласование целей при эколого-экономическом обосновании градостроительного проекта с учетом ограничений на значения характеристик, входящих в систему средств, в условиях нечеткой информации // *Вестник ВолгГАСУ. Сер.: Стр-во и архит.* 2011. Вып. 21(40). С. 151—159.
3. *Санжапов Б. Х., Садовникова Н. П.* Поддержка принятия решений при планировании развития городских территорий на основе экосистемного подхода // *Вестник ВолгГАСУ. Сер.: Стр-во и архит.* 2013. Вып. 31(50). Ч. 2. С. 577—584.
4. *Finkel J., Grenager T., Manning C.* Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling // *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 2005. P. 363—370. doi: [10.3115/1219840.1219885](https://doi.org/10.3115/1219840.1219885)
5. *Hall M. M., Jones C. B.* Quantifying spatial prepositions // *Proceedings of the 16-th ACM SIGSPATIAL international conference on Advances in geographic information systems*. 2008. P. 1. doi: [10.1145/1463434.1463507](https://doi.org/10.1145/1463434.1463507).
6. *Bateman J. A., Hois J., Ross R., Tenbrink T.* A linguistic ontology of space for natural language processing // *Artificial Intelligence*. 2010. Vol. 174. Iss. 14. P. 1027—1071.
7. *Wong S.-M. J., Dras M.* Exploiting parse structures for native language identification // *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. P. 1600—1610.
8. *Hu Y., Boyd-Graber J., Satinoff B.* Interactive topic modeling // *Proceedings of the 49-th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. P. 248—257.
9. *Solka J. L.* Text data mining: theory and methods // *Statistics Surveys*. 2008. Vol. 2. P. 94—112.
10. *Pu S., Vosselman G.* Knowledge based reconstruction of building models from terrestrial laser scanning data // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009. Vol. 64. Iss. 6. P. 575—584.
11. *Bishr Y.* Overcoming the semantic and other barriers to GIS interoperability // *International Journal of Geographical Information Science*. 1998. Vol. 12. Iss. 4. P. 299—314. doi: [10.1080/136588198241806](https://doi.org/10.1080/136588198241806).
12. *Benz U. C., Hofmann P., Willhauck G., Lingenfelder I., Heynen M.* Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information // *ISPRS Journal of photogrammetry and remote sensing*. 2004. Vol. 58. Iss. 3—4. P. 239—258.
13. *Apel M.* From 3d geomodelling systems towards 3d geoscience information systems: Data model, query functionality, and data management // *Computers & Geosciences*. 2006. Vol. 32. Iss. 2. P. 222—229.
14. *Jackson M. J., Mason D. C.* The development of integrated geo-information systems // *International Journal of Remote Sensing*. 1986. Vol. 7. Iss. 6. P. 723—740. doi: [10.1080/01431168608954729](https://doi.org/10.1080/01431168608954729).

1. Jaiswal A., Pezanowski S., Mitra P., Zhang X., Xu S., Turton I., Klippel A., MacEachren A. M. GeoCAM: A geovisual analytics workspace to contextualize and interpret state-ments about movement // *Journal of Spatial Information Science*. 2011. Vol. 3. Pp. 65—101. doi: 10.5311/JOSIS.2011.3.55
2. Sanzhapov B. Kh., Sadovnikova N. P. Soglasovanie tseley pri ekologo-ekonomicheskom obosnovanii gradostroitel'nogo proekta s uchetom ogranicheniy na znacheniya kharakteristik, vkhodyashchikh v sistemu sredstv, v usloviyakh nechetkoy informatsii // *Vestnik Volgogradskogo gosudarstvennogo arhitekturno-stroitel'nogo universiteta*. Seriya: Stroitel'stvo i arhitektura. 2011. Vyp. 21(40). S. 151—159.
3. Sanzhapov B. Kh., Sadovnikova N. P. Podderzhka prinyatiya resheniy pri planirovanii razvitiya gorodskikh territoriy na osnove ekosistemnogo podkhoda // *Vestnik Volgogradskogo gosudarstvennogo arhitekturno-stroitel'nogo universiteta*. Seriya: Stroitel'stvo i arhitektura. 2013. Vyp. 31(50). Ch. 2. S. 577—584.
4. Finkel J., Grenager T., Manning C. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling // *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 2005. P. 363—370. doi: 10.3115/1219840.1219885
5. Hall M. M., Jones C. B. Quantifying spatial prepositions // *Proceedings of the 16-th ACM SIGSPATIAL international conference on Advances in geographic information systems*. 2008. P. 1. doi: 10.1145/1463434.1463507.
6. Bateman J. A., Hois J., Ross R., Tenbrink T. A linguistic ontology of space for natural language processing // *Artificial Intelligence*. 2010. Vol. 174. Iss. 14. P. 1027—1071.
7. Wong S.-M. J., Dras M. Exploiting parse structures for native language identification // *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. P. 1600—1610.
8. Hu Y., Boyd-Graber J., Satinoff B. Interactive topic modeling // *Proceedings of the 49-th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. P. 248—257.
9. Solka J. L. Text data mining: theory and methods // *Statistics Surveys*. 2008. Vol. 2. P. 94—112.
10. Pu S., Vosselman G. Knowledge based reconstruction of building models from terrestrial laser scanning data // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009. Vol. 64. Iss. 6. P. 575—584.
11. Bishr Y. Overcoming the semantic and other barriers to GIS interoperability // *International Journal of Geographical Information Science*. 1998. Vol. 12. Iss. 4. P. 299—314. doi: 10.1080/136588198241806.
12. Benz U. C., Hofmann P., Willhauck G., Lingenfelder I., Heynen M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information // *ISPRS Journal of photogrammetry and remote sensing*. 2004. Vol. 58. Iss. 3—4. P. 239—258.
13. Apel M. From 3d geomodelling systems towards 3d geoscience information systems: Data model, query functionality, and data management // *Computers & Geosciences*. 2006. Vol. 32. Iss. 2. P. 222—229.
14. Jackson M. J., Mason D. C. The development of integrated geo-information systems // *International Journal of Remote Sensing*. 1986. Vol. 7. Iss. 6. P. 723—740. doi: 10.1080/01431168608954729.

© Платонов А. А., Санжапов Б. Х., 2014

Поступила в редакцию
в июне 2014 г.

Ссылка для цитирования:

Платонов А. А., Санжапов Б. Х. Автоматизированная система сбора и анализа данных при планировании развития городских территорий // Интернет-вестник ВолгГАСУ. Сер.: Строительная информатика. 2014. Вып. 11(32). Ст. 5. Режим доступа: <http://www.vestnik.vgasu.ru/>

For citation:

Platonov A. A., Sanzhapov B. Kh. [Automated data collection and analysis system for planning urban territories expansion]. *Internet-Vestnik VolgGASU*, 2014, no. 11(32), paper 5. (In Russ.). Available at: <http://www.vestnik.vgasu.ru/>