

## СОЗДАНИЕ АГРЕГАТОРА ИНТЕРНЕТ-СЕРВИС-ПРОВАЙДЕРОВ С ПОДСИСТЕМОЙ АВТОМАТИЗИРОВАННОЙ ОБРАБОТКИ АДРЕСНЫХ БАЗ

© 2022 А. Г. Черняк, Т. В. Корелина, А. Н. Зеленина

Воронежский институт высоких технологий (Воронеж, Россия)

*Актуальность разработки обусловлена востребованностью услуг интернет-провайдеров на территории России и связанной с этим потребностью в наличии информационного ресурса по тематике телекоммуникационных услуг. В рамках исследования была поставлена цель по созданию веб-сайта-агрегатора провайдеров, для полноценного функционирования которого требуется подсистема обработки адресных баз. В задачи исследования входит проектирование системы управления контентом сайта, определение необходимых модулей административной панели, разработка алгоритма обработки адресных баз и его реализация.*

*Основными преимуществами разработанной программной системы являются высокая точность распознавания адресов, удобный способ вывода информации для проверки результатов контент-менеджером, гибкое управление системой из интуитивного веб-интерфейса, низкие системные требования.*

*Ключевые слова: агрегатор, интернет-провайдер, автоматизация, обработка адресных баз, РНР.*

### Введение

На 2022 год количество пользователей интернета достигло 4,9 млрд человек по всему миру, что почти на 200 млн человек больше, чем годом ранее.<sup>1</sup>

Такой масштабный прирост пользователей приводит к возникновению на рынке большого числа поставщиков услуги доступа в интернет для конечных пользователей – интернет-провайдеров третьего уровня (Tier 3).

Для примера, в России, на текущий момент около 4 тысяч компаний, имеющих лицензии на предоставление телематических услуг связи и имеющих ненулевую финансовую отчетность за 2021 год. В основном, это небольшие региональные провайдеры, предоставляющие доступ к интернету на ограниченной территории, например, в каком-либо областном центре, но среди них есть и компании, работающие по всей России.

Если учесть, что в РФ сейчас около 1300 населённых пунктов с населением свыше 10 тысяч жителей, то получается, что на 1 российский город приходится около трёх провайдеров, при этом количество провайдеров

в крупных городах - от 10 и более (в Москве их больше двухсот).

Такое многообразие, с одной стороны, предоставляет потребителю более свободный выбор, но в то же время влечёт и дополнительные трудности, связанные с выбором компании, качество услуг и сервиса которой будут удовлетворять потребностям конкретного клиента.

### Постановка задачи для решения проблемы

Основная проблема, на которую было обращено внимание, это проблема определения доступных для подключения провайдеров интернета по конкретному адресу. На просторах сети можно обнаружить сайты, предлагающие посетителю ввести свой адрес, и обещающие после этого выдать список доступных на адресе провайдеров, однако, по результатам анализа нескольких таких ресурсов становится ясна неполноценность предоставляемых клиенту сведений. Для примера рассмотрим адрес: г. Воронеж, ул. Адмирала Чурсина, 12. Это большой многоквартирный дом, сданный в эксплуатацию в 2020 году, и

Черняк Александр Геннадьевич – Воронежский институт высоких технологий, студент.  
Корелина Татьяна Викторовна – Воронежский институт высоких технологий, канд. техн. наук, доцент.

Зеленина Анна Николаевна – Воронежский институт высоких технологий, [snakeans@gmail.com](mailto:snakeans@gmail.com).

имеющий по результатам обзона местных компаний, 5 доступных к подключению провайдеров. На рисунке 2 можно увидеть

скриншот из системы поиска провайдеров, где виден соответствующий результат.

## ПРОБЛЕМА ПОИСКА ПРОВАЙДЕРА ДЛЯ КОНЕЧНОГО ПОЛЬЗОВАТЕЛЯ

- ~4000 активно действующих телекоммуникационных компаний в России
- Усреднённо: 3-4 провайдера на 1 населённый пункт с населением свыше 10 тысяч человек (1300 таких н. п. в РФ)
- Фактически: от 10 провайдеров на 1 крупный город
- ~200-300 провайдеров в Москве, ~600 провайдеров суммарно в населённых пунктах Московской области
- От 4 до 10 провайдеров на 1 многоквартирный дом (в крупных городах)

2

Рисунок 1. Анализ проблемы

## КОЛИЧЕСТВО ПРОВАЙДЕРОВ В ОДНОМ МНОГОКВАРТИРНОМ ДОМЕ

На примере адреса  
г. Воронеж, ул. Адмирала Чурсина, 12:

г Воронеж, ул Адмирала Чурсина, д 12  
5 провайдеров доступно (всего 6 записей)

- Телеком-Сервис
- Мегафон
- ТТК
- Ростелеком
- JustLan
- NetByNet Технические ограничения

3

Скриншот из разработанной системы

Рисунок 2. Реальное количество провайдеров по указанному адресу

Однако, если указать данный адрес на любом из ресурсов, которые пользователь может свободно обнаружить в поисковиках Яндекс или Google, мы увидим следующую

картину: самые популярные сайты по тематике поиска интернет-провайдеров отображают ноль (рис. 3), одного (рис. 4), и четырех, но с ошибкой (рис. 5) провайдеров, доступных к подключению.

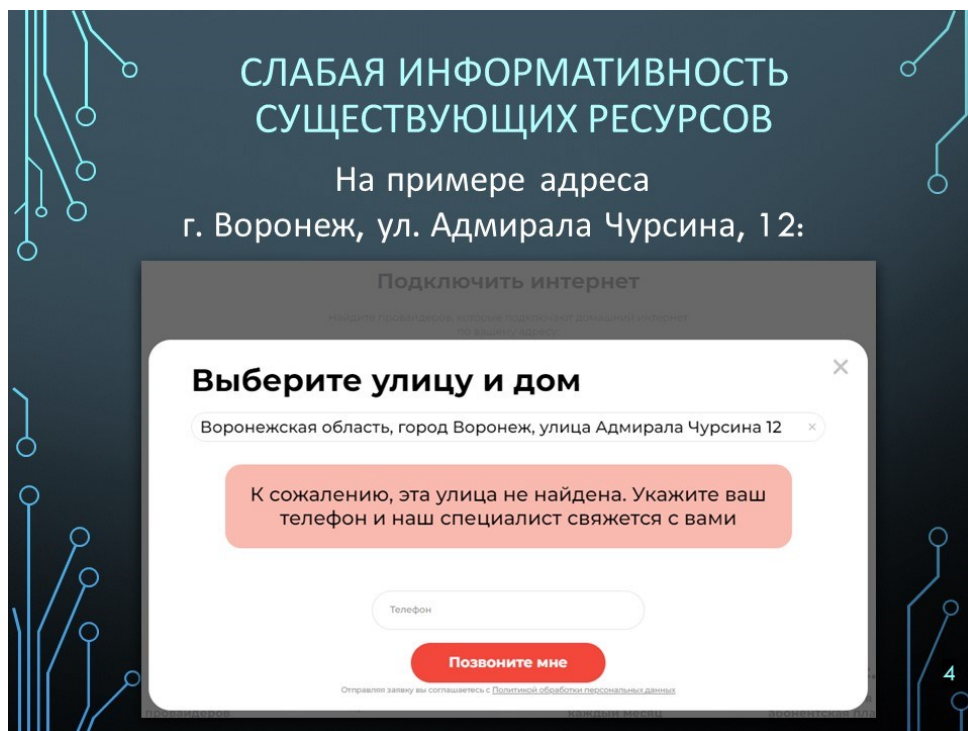


Рисунок 3. Пример полного отсутствия информации о провайдерах на указанном адресе

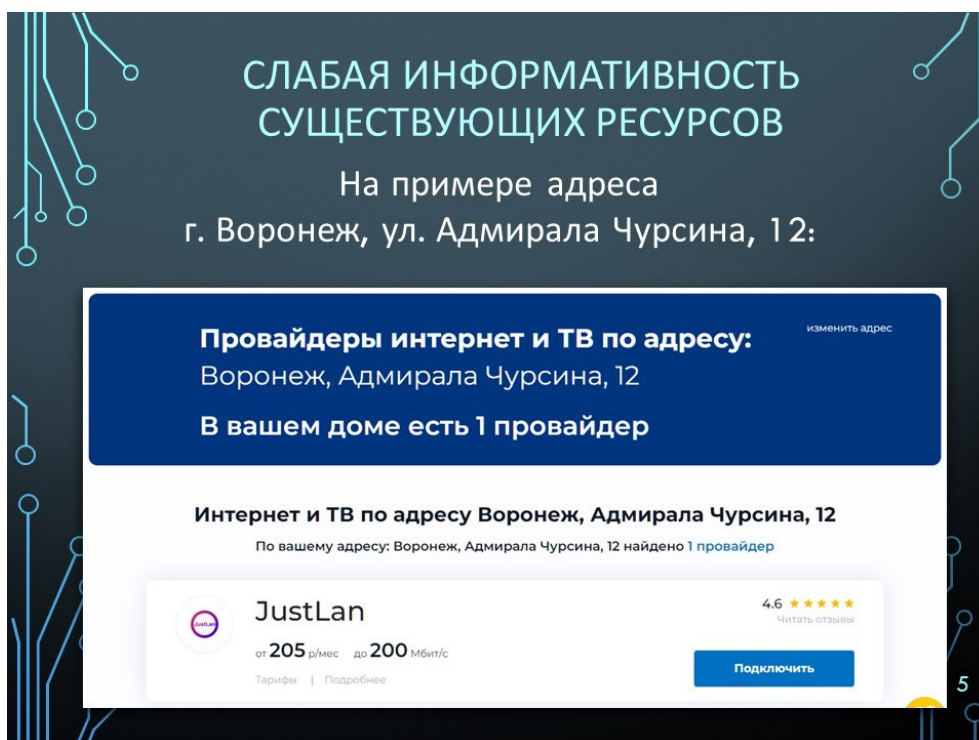


Рисунок 4. Пример неполной информации, найден один провайдер

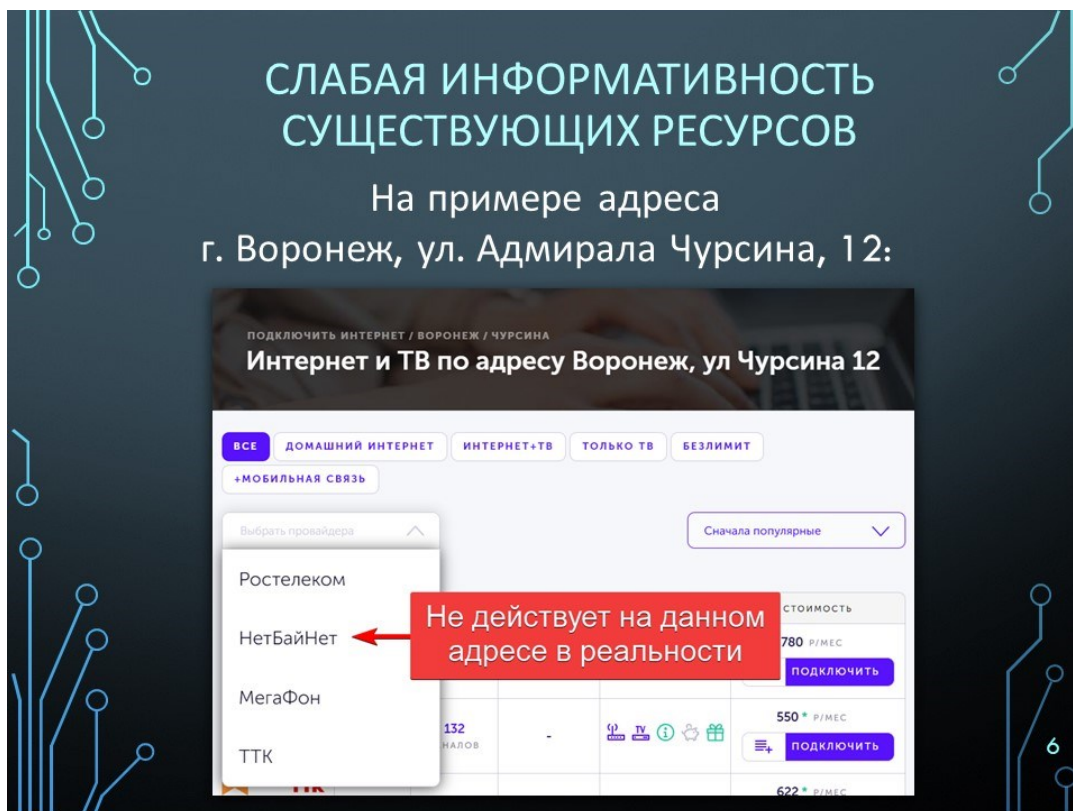


Рисунок 5. Пример неполной и некорректной информации, данный провайдер на адресе не действует

Неполнота доступной информации очевидна, при этом она не обусловлена разницей в исходных данных – по открытой самими агрегаторами информации, перечисленные провайдеры являются партнёрами перечисленных в примере провайдеров, и получают информацию непосредственно от них в том же виде.

В рамках данного исследования были обнаружены причины низкой точности сведений по значительному количеству адресов в любом городе, вне зависимости размера города, числа активных в нём провайдеров, новизны и размера дома, или каких-либо других факторов. Проблема, по результатам проведённого анализа, оказалась, скорее, технической: большинство таких сайтов получают информацию о наличии провайдеров на адресе от самих провайдеров, которые предоставляют списки адресов. Однако, эти списки (образцы которых также получены в рамках исследования), представляют собой таблицы с сотнями тысяч и даже миллионами строк в формате excel. При этом, написание адресов

в таких таблицах варьируется от провайдера к провайдеру, и далеко не всегда хоть как-нибудь структурировано или приведено к единому виду – нередко адрес написан одной строкой, причём в разных версиях одной и той же базы адрес может быть записан по-другому. В результате таких разночтений и объёма баз, существующие агрегаторы провайдеров не справляются с задачей по сопоставлению одних и тех же адресов, но записанного по-разному в разных базах, друг с другом, что и приводит к низкой точности результатов поиска на существующий агрегаторах провайдеров.

В рамках данного исследования была поставлена задача разработать новый агрегатор и специальную систему распознавания адресов, позволяющую вовремя увидеть и минимизировать ошибки, что позволит составить намного полную базу провайдеров по адресу.

Алгоритм работы системы распознавания адресов отображен на рисунке 6.



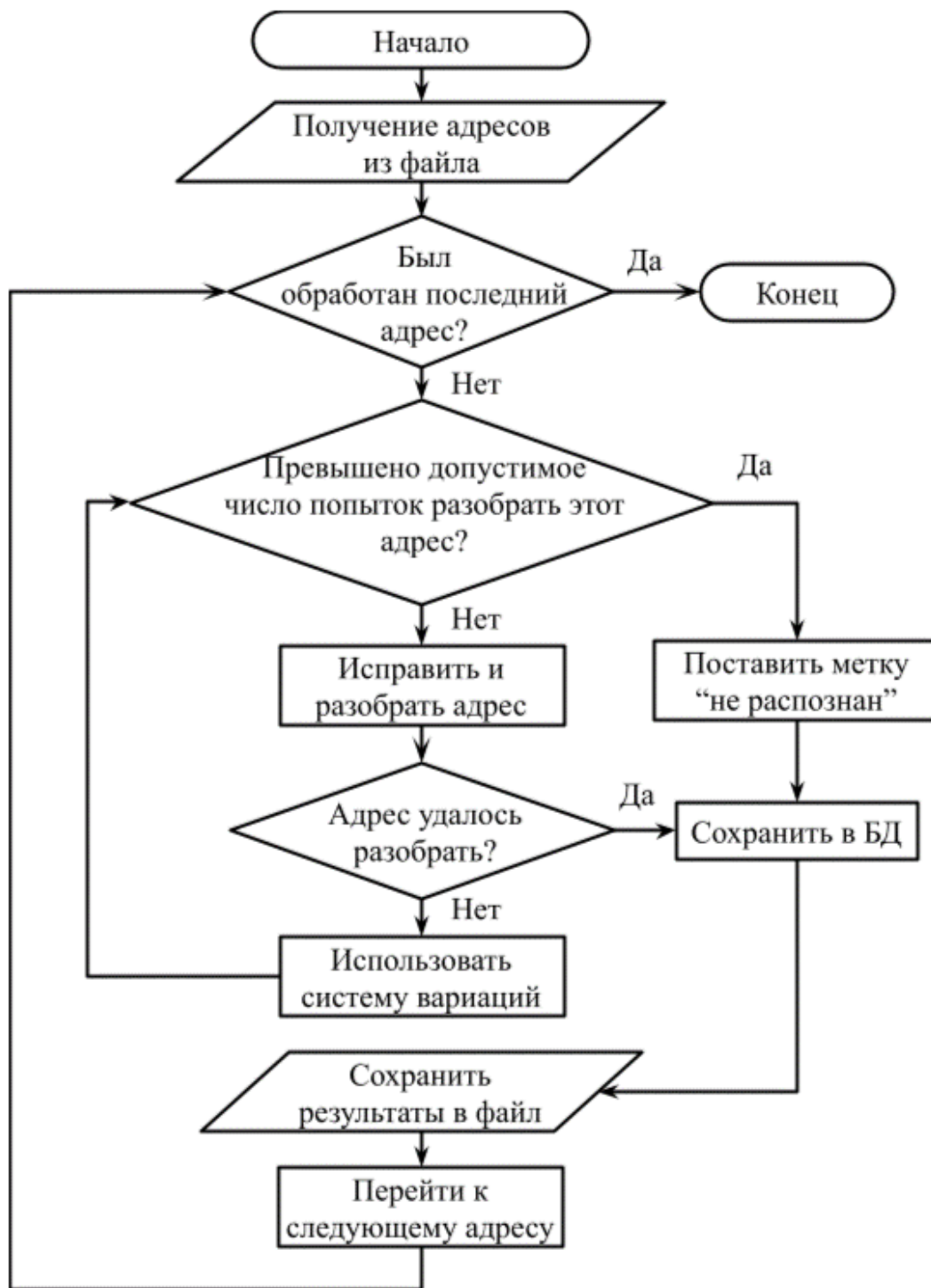


Рисунок 6. Концептуальная схема алгоритма обработки адресов

После загрузки файла с адресами, эти адреса по очереди перебираются парсером, и сохраняются в базу данных. Также результат разбора каждой строчки сохраняется в xls-таблицу, помечая ошибки, что позволяет проверить корректность распознавания адресов

из файла и поправить ошибки, после чего загрузить файл с исправленными адресами заново. Доступ к результатам осуществляется из административной панели сайта, из раздела «История обработки», вид которого представлен на рисунке 7.



Рисунок 7. История запусков процесса обработки

Для предоставления информации о провайдерах был разработан сайт «Провайдеры Города» (в каждом городе – свой заголовок),

главная страница раздела «Провайдеры Воронежа» показана на рисунке 8.

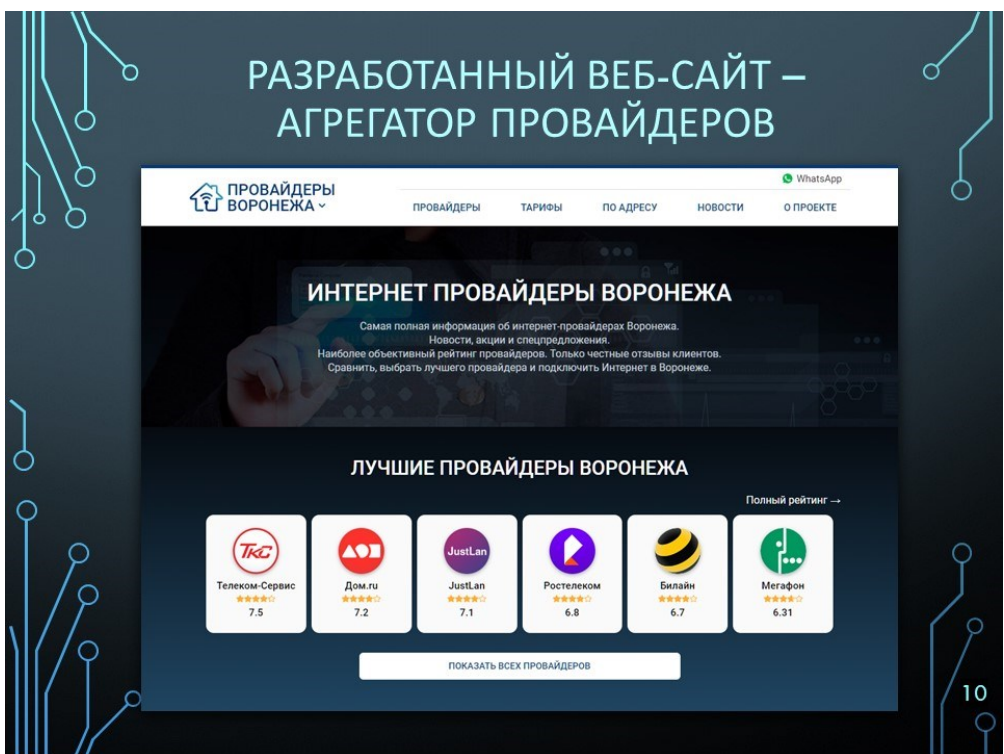


Рисунок 8. Главная страница раздела «Провайдеры Воронежа»

## Заключение

В результате создания продвинутого обработчика адресных баз стало возможным создание полноценной базы адресов провайдеров с долей неточностей до 0,5 %. Учитывая, что на существующих в сети агрегаторах, по данным нашего исследования, доля ошибок составляет 3-5 %, можно говорить о 5-10-кратном увеличении точности базы, в сравнении с аналогичными решениями.

При использовании партнерских программ провайдеров, достигнутые результаты позволят увеличить экономическую эффективность проекта в таких же пределах, поскольку возможность подключить абонента или передать его для подключения провайдеру прямым образом зависит от точности адресной базы.

## СПИСОК ИСТОЧНИКОВ

1. Зандстра, Мэтт. PHP: объекты, шаблоны и методики программирования, 5-е изд.: Пер. с англ. – СПб.: ООО “Диалектика”, 2019. – 736 с.: ил. – Парал. тит. англ. ISBN 978-5-907144-54-5 (рус.).

2. Локхарт, Джош. Современный PHP. Новые возможности и передовой опыт [Текст] / Джош Локхарт ; пер. с англ. Р. Н. Рагимов. – М. : ДМК Пресс, 2016. – 304 с. : ил. – ISBN 978-1-49190-501-2 (англ.). – ISBN 978-5-97060-184-6 (рус.).

3. Прохоренок Н. А. HTML, JavaScript, PHP и MySQL. Джентльменский набор Web-мастера. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2010. – 912 с.: ил. + Видеокурс (на CD-ROM) – (Профессиональное программирование) ISBN 978-5-9775-0540-6.

4. Прохоренок, Н. А. HTML, JavaScript, PHP и MySQL. Джентльменский набор Web-мастера. – 5-е изд., перераб. и доп. / Н. А. Прохоренок, В. А. Дронов. – СПб.: БХВ-Петербург, 2019. – 912 с.: ил. – (Профессиональное программирование) ISBN 978-5-9775-3986-9.

5. Форта, Бен. Освой самостоятельно регулярные выражения. 10 минут на урок: Пер. с англ. – М. : Издательский дом “Вильямс”, 2005. – 184 с. : ил. – Парал. тит. англ. ISBN 5-8459-0713-6 (рус.).

6. Новиков Б. А. Основы технологий баз данных: учеб. пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. – 2-е изд. – М.: ДМК Пресс, 2020. – 582 с. ISBN 978-5-97060-841-8 (ДМК Пресс) ISBN 978-5-6041193-5-8 (Постгрес Профессиональный, электронное издание).

7. Моргунов, Е. П. PostgreSQL. Основы языка SQL: учеб. пособие / Е. П. Моргунов; под ред. Е. В. Рогова, П. В. Лузанова. – СПб.: БХВ-Петербург, 2018. – 336 с.: ил. ISBN 978-5-9775-4022-3 (БХВ-Петербург) ISBN 978-5-6041193-2-7 (Постгрес Профессиональный, электронное издание).

8. Роберт С. Мартин, Джеймс В. Ньюкирк, Роберт С. Косс. Быстрая разработка программ. Принципы, примеры, практика. – Вильямс, 2004, ISBN 5-8459-0558-3, ISBN 0-13-597444-5.

9. Одинцов И. О. Профессиональное программирование. Системный подход. – 2-е изд. – СПб.: БХВ-Петербург, 2004. – 624 с. – ISBN 5-94157-457-6.

## CREATING AN AGGREGATOR OF INTERNET SERVICE PROVIDERS BASED ON THE AUTOMATED ADDRESS BASES PROCESSING SUBSYSTEM

© 2022 A. G. Chernyak, T. V. Korelina, and A. N. Zelenina

*Voronezh Institute of High Technologies (Voronezh, Russia)*

*The relevance of the development proposed in this article is preconditioned by a high demand for the services of Internet providers in Russia and the related need for an information resource on the subject of telecommunications services. Thus, the goal was to create a website-aggregator of providers, which involved developing a subsystem for processing address databases. The objectives of the study include designing a site content management system, determining the necessary modules for the administrative panel, developing an algorithm for processing address databases and implementing it. The main advantages of the developed software system are the high accuracy of address recognition, a convenient way to display information for checking the results by the content manager, flexible system management from an intuitive web interface, and low system requirements.*

*Keywords: aggregator, Internet provider, automation, processing of address databases, PHP.*