

УДК 004.9:656

А. И. Фарафонов, аспирант, Ижевский государственный технический университет имени М. Т. Калашникова

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО ГЕОКОДИРОВАНИЯ В ТРАНСПОРТНОЙ КОМПАНИИ

На сегодняшний день службы такси являются важной частью транспортной системы современного города. В России такси как вид транспорта стал популярен сравнительно недавно. На данный момент услуга такси является массовой, дневной объем заказов составляет сейчас до 10 % населения города. Такой рост спроса потребовал значительно увеличить степень автоматизации работы службы такси. Одно из направлений такого улучшения – использование геокодирования – сопоставления географических координат заказам.

Традиционно службы такси делят города на сектора, внутри которых и распределяются заказы. Для того чтобы правильно определить сектор, оператор службы такси должна была знать город и выбрать сектор вручную. Ручной ввод адреса требует много времени, приводит к большому количеству ошибок и опечаток.

Предлагается модель адреса, имеющая иерархическую структуру и отражающая все необходимые аспекты (рис. 1).

Систему геокодирования можно представить в виде трех слоев (рис. 2).

На **слое данных** размещаются 3 базы данных различного назначения. База данных заказов хранит исход-

ные данные: заказы, адреса, маршруты поездок. База данных адресов хранит информацию об адресах.

Слой логики содержит базовые подсистемы: модуль извлечения адресов из заказов, модуль автодополнения и модуль геокодирования.

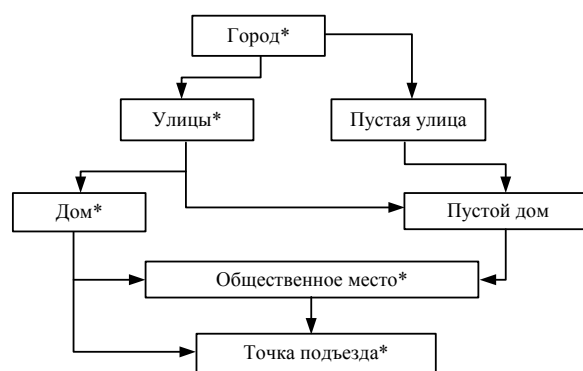


Рис. 1. Модель адреса

Сервисы предоставляют различным программам к функциям слоя логики: прямое и обратное геокодирование, автодополнение с различными параметрами и фильтрами, ведение персонифицированных списков адресов для каждого клиента.

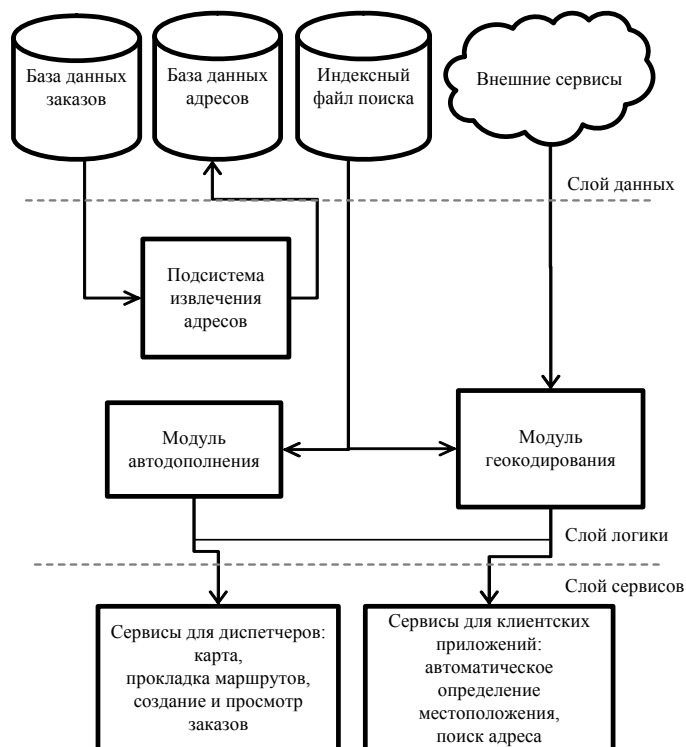


Рис. 2. Структура системы геокодирования

Для извлечения адресов используются данные с GPS терминалов водителей, которые передаются в момент посадки в машину. Адрес преобразуется из строки в структуру (см. рис. 1), из него выделяются компоненты улицы, номера дома.

Адреса различного типа используются с разной частотой. Такие адреса, как торговые центры, могут набирать тысячи заказов в день, частные дома, в которых проживает одна семья, – один заказ в несколько месяцев. Для редко используемых адресов применяется быстрый старт – адрес попадает в базу после первого использования. После 4 заказов от разных клиентов в действие вступает алгоритм кластеризации [1].

Алгоритм кластеризации выделяет несколько кластеров и отбрасывает шумовые точки. Каждому выявленному кластеру приписывается адрес с комментарием на основе соответствующих заказов. В дальнейшем при геокодировании используется центр масс точек, входящих в данный кластер. При наличии двух кластеров, соответствующих одному адресу, выбирается кластер большей мощности.

Важное требование к базе адресов в работе алгоритма фильтрации адресов – точность и однозначность.

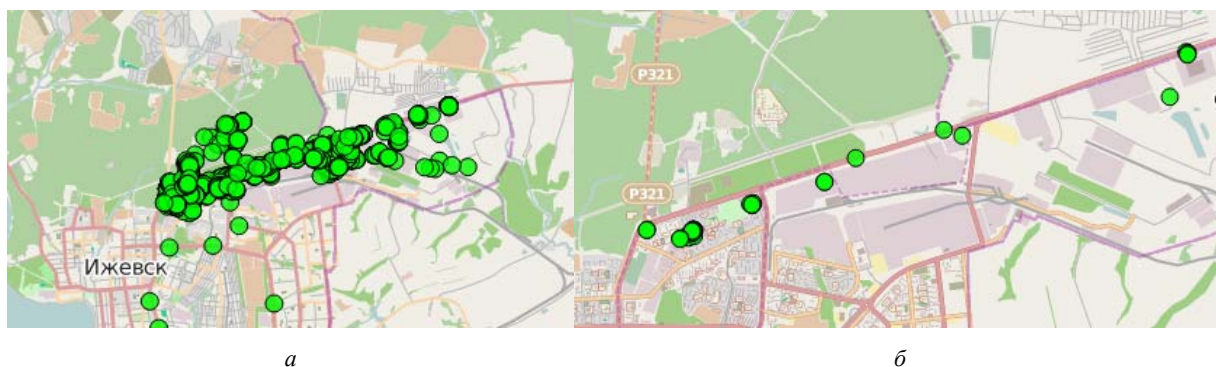


Рис. 3. Множество точек, соответствующих улице из справочника (а) и одному из ошибочных вариантов (б)

Требования к времени выполнения запросов

Виды запросов	Содержание запросов	Время выполнения
Основные запросы	Геокодирование, обратное геокодирование, автодополнение по префиксу	Не более 10 мс
Информационно-справочные запросы	Список всех улиц, список всех домов на улице	Не более 100 мс
Аналитические запросы	Список всех заказов за период на определенных улицах	Не более 1 с

На данный момент достигнуты следующие важные цели:

- доля правильно геокодированных заказов достигла 100 %;
- время ввода заказа оператором сокращено с 30 до 10 сек.;
- снято требование наличия карты города для запуска службы такси;
- доля заказов с комментариями выросла с 20 до 50 %.

Разработанные алгоритмы позволяют выделять объекты на карте города, автоматически переносить объекты на карте в случае изменения адреса. Также собранная информация позволяет сохранять веро-

Первый уровень фильтрации – приведение всех строк к каноническому виду:

- для улиц распознаются ключевые слова и сокращения: улица, ул., проезд, переулок и др.;
- для зданий распознаются несколько типовых шаблонов;
- для подъездов распознаются шаблоны типа -й подъезд, к шлагбауму, на парковку и др.

Во всех словах производится фонетическая замена (о-а, е-и...) [2].

Второй уровень фильтрации – исправление незначительных опечаток. Для этого вычисляется попарное расстояние редактирования («расстояние Левенштайна») [3].

Третий уровень фильтрации – пространственная фильтрация. Для каждого объекта строится поверхность плотности [4]. Таким образом, для выяснения степени совпадения двух объектов достаточно выяснить объем, ограниченный пересечением этих поверхностей (рис. 3).

К времени выполнения запросов в этих подсистемах автодополнения и геокодирования выдвигаются жесткие требования (таблица).

ятный характер геокодирования: в случае не точно определенных результатов варианты предлагаются на основе функций вероятности, что значительно улучшает качество таких подсказок. Перспективным направлением анализа собираемых данных является создание схем проезда и маршрутов движения по городу с учетом изменяющихся дорожных условий.

Также на основе созданной системы можно собирать и анализировать как краткосрочную (пробки, загруженность дорог, временные заторы и препятствия, доступность парковочных мест), так и долгосрочную (проблемные участки дорог, недогруженные и перегруженные парковки) информацию.

Библиографические ссылки

1. A density-based algorithm for discovering clusters in large spatial databases with noise / Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. – AAAI Press, 1996. – Pp. 226–231.

2. The Soundex Indexing System // The U. S. National Archives and Records Administration, 2007. – URL:

Получено 17.06.14

<http://www.archives.gov/research/census/soundex.html> (дата обращения: 11.02.2015).

3. Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Докл. Академии наук СССР. – М. : Наука, 1965. 163.4:845-848.

4. Wand M. P., Jones M. C. Kernel Smoothing. – London : Chapman & Hall /CRC Press, 1995.

УДК 004.932.7 + 621.397

А. В. Самохвалов, кандидат технических наук, Ижевский государственный технический университет имени М. Т. Калашникова

КОМПРЕССИЯ КОНТУРНОГО И КОДИРОВАНИЕ МАСКИРОВАННОГО ИЗОБРАЖЕНИЙ

Контурная информация играет существенную роль при восприятии человеческим зрением изображений. Следовательно, «неконтурная» информация (области одного цвета или плавно меняющегося цвета) является визуально избыточной

и может быть частично или полностью исключена из изображения [1, 2]. На данной идее основывается предложенный ниже контурный метод сжатия изображений, структурная схема которого представлена на рис. 1.

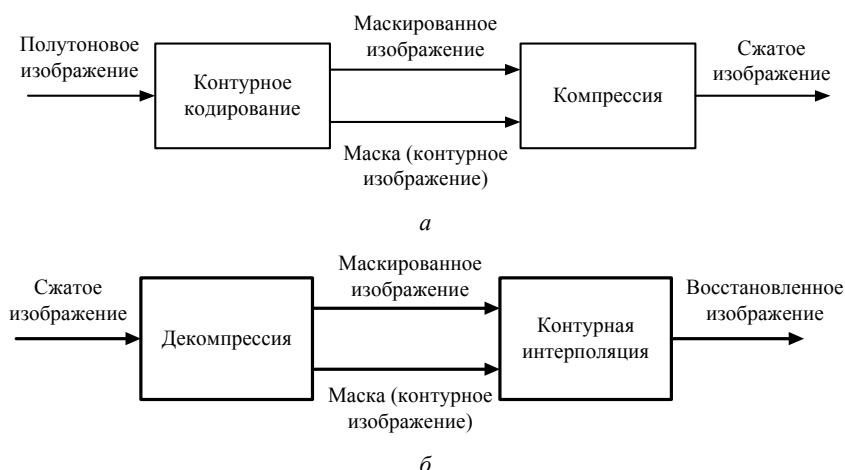


Рис. 1. Структурная схема контурного метода сжатия изображений: а – компрессия; б – декомпрессия

Основная идея сжатия полутоновых изображений с помощью метода контурных кодирования и интерполяции заключается в следующем. Сначала на полутоновом изображении (рис. 2) выделяются контуры (рис. 3): формируется двоичное изображение, в котором «0» означает отсутствие контура, а «1» – наличие контура в соответствующем пикселе исходного изображения.

Далее полученные контуры исходного изображения утолщаются. Таким образом, получена контурная маска (рис. 4) (результат утолщения контуров). На следующем этапе происходит маскирование изображения (рис. 5): полученная маска накладывается на исходное изображение (операция умножения исходного изображения и маски). В результате получа-

ем маскированное исходное изображение, на котором хорошо просматриваются околоконтурные области (пиксели исходного изображения, лежащие на контурах и вблизи контуров, не претерпевающие никаких изменений), а также пустые (потерянные) области. От исходного изображения остается очень малая часть (5–20 % для большинства изображений) значащих пикселей. Используя на данном этапе стандартные методы сжатия, можно получить высокий коэффициент компрессии, при том что достаточно важная контурная составляющая изображения остается без изменений.

В статье рассмотрены этапы компрессии контурного изображения и эффективного кодирования маскированного изображения для дальнейшего его сжатия.