

Информационная система  
"Интеллектуальная система проверки и исправления  
почтовых адресов клиентов банка"

**Руководство оператора**

Москва 2007

## **Оглавление**

<b>ОГЛАВЛЕНИЕ .....</b>	<b>2</b>
<b>ГЛОССАРИЙ .....</b>	<b>3</b>
<b>1. НАЗНАЧЕНИЕ СИСТЕМЫ .....</b>	<b>3</b>
<b>2. УСЛОВИЯ ПРИМЕНЕНИЯ.....</b>	<b>3</b>
<b>3. ОСНОВНЫЕ ФУНКЦИИ СИСТЕМЫ.....</b>	<b>3</b>
<b>4. ОБОБЩЕННАЯ СТРУКТУРА СИСТЕМЫ И ПОРЯДОК ЕЁ ФУНКЦИОНИРОВАНИЯ .....</b>	<b>5</b>
<b>5. ПОДГОТОВКА К РАБОТЕ ОПЕРАТОРА .....</b>	<b>6</b>
<b>6. ПОРЯДОК РАБОТЫ ОПЕРАТОРА.....</b>	<b>7</b>

## Глоссарий

**Адресный объект** – отдельный структурный элемент почтового адреса: регион, район, город, поселок городского типа, сельский населенный пункт, улица, дом, квартира.

**АРМ оператора** – автоматизированное рабочее место работника службы эксплуатации информационных систем банка, осуществляющего в режиме off-line подготовку запросов на проверку адресов клиентов банка и загрузку в базу данных соответствующей ИС банка после проверки исправленных адресов.

**ИС банка** – информационная система, осуществляющая обработку банковской информации, в том числе обработку почтовых адресов клиентов банка.

**КЛАДР** – общероссийский классификатор адресов. Разработан ФГУП ГНИВЦ ФНС России. Описание классификатора утверждено Приказом ФНС России № САЭ-3-13/594 от 17.11.2005 года.

**CSV** (Comma Separated Value) – формат представления информации, в котором в качестве разделителя полей выступает запятая.

**SOA** (Service-Oriented Architecture) – сервисно-ориентированная архитектура. Парадигма, предназначенная для проектирования, разработки и управления дискретных единиц логики (сервисов) в вычислительной среде.

**XML-сообщение** – информационное сообщение, написанное на языке XML. Язык XML (Extensible Markup Language) – расширенный язык разметки документов (<http://www.w3.org/XML/>).

## 1. Назначение системы

Информационная система "Интеллектуальная система проверки и исправления почтовых адресов клиентов банка" предназначена для выявления и исправления ошибок в почтовых адресах клиентов банка и формирования «правильных» адресов в соответствии с классификатором адресов КЛАДР.

## 2. Условия применения

Разрабатываемая система реализуется в виде Web-сервиса в соответствии с сервисно-ориентированной архитектурой (SOA). В режиме on-line система взаимодействует с ИС банка в формате XML-сообщений. Протоколом взаимодействия ИС банка с Web-сервисом создаваемой системы является SOAP1.1/HTTP с поддержкой SOAP-стиля Document и SOAP-кодировки Literal.

Взаимодействие с ИС банка в режиме off-line осуществляется оператором ИС банка в режиме чтения/записи файлов в форматах XML или CSV.

## 3. Основные функции системы

Система должна обеспечить выполнение следующих основных функций:  
получение запросов на проверку правильности заполнения почтовых адресов клиентов банка от информационных систем (ИС) банка в формате XML;  
распознавание структуры адреса, представленного сплошной строкой или группой сцепленных строк (при наличии в сплошной строке явных разделительных

символов, принятых для отделения адресных объектов в сплошной строке друг от друга: пробелов, запятых и др.), и его представление в виде отдельных полей для адресных объектов;

- выявление и исправление опечаток и ошибок в адресных объектах
- замена устаревших наименований адресных объектов, подвергшихся переименованию, на их актуальные наименования;

- проверка существования почтового адреса путем его поиска в базе данных классификатора КЛАДР (исключая уровень квартир);

- извлечение индекса почтового отделения, обслуживающего адрес;

- формирование выходного сообщения, содержащего структурированный исправленный адрес, включающий индекс почтового отделения, а также сообщения об ошибках;

- формирование протокола о результатах проверки почтовых адресов.

В процессе обработки входной информации должны выявляться и исправляться следующие типовые ошибки в адресах:

- орфографические ошибки в написаниях наименований регионов, населенных пунктов, улиц и т.д.;

- отсутствие почтового индекса;

- отсутствие наименований адресных объектов;

- отсутствие наименований типов адресных объектов;

- написание наименований адресных объектов (например, улиц) строчными буквами;

- написание наименований адресных объектов (например, регионов и районов) в родительном, а не в именительном падеже;

- использование при написании значений адресных объектов символов латинского алфавита: E, T, O, P, A, H, K, X, C, B, M, сходных по написанию с кириллицей;

- использование нестандартных, различных и неоднозначных сокращений для наименований типов адресных объектов: например, проезд - прд или пр., проспект - просп. или пр-т, ст. – станция или станция и др;

- неодинаковое позиционирование типов адресных объектов (например, улиц) по отношению к их наименованиям (до или после наименования);

- дублирование одного и того же наименования населенного пункта несколько раз;

- присутствие лишних знаков препинания (тире, запятые, кавычки).

Кроме того, в процессе обработки входной информации должны выявляться следующие ошибки в адресах:

- неправильное написание наименований некоторых регионов, населенных пунктов;

- наличие не существующих наименований населенных пунктов.

Обработка входной информации должна осуществляться с использованием классификатора адресов КЛАДР версии 4.0.

Исправление орфографических ошибок выполняется только при наличии однозначного варианта исправления, в остальных случаях система только выявляет факт ошибки в соответствующем поле адреса.

Восстановление отсутствующих адресных объектов (почтового индекса, наименований адресных объектов, наименований типов адресных объектов) производится только при осуществимости таких операций с помощью классификатора КЛАДР. В остальных случаях система только выявляет факт наличия ошибки в соответствующем поле адреса.

#### 4. Обобщенная структура системы и порядок её функционирования

Система в общем виде состоит из следующих основных программных модулей и баз данных (рис. 1):

- модуль организации сетевого взаимодействия;
- модуль распознавания структуры адреса;
- модуль выявления опечаток;
- модуль проверки правильности адреса;
- база данных правил структуризации адреса;
- база специальных хранимых структур данных;
- онтология географических наименований;
- АРМ оператора.

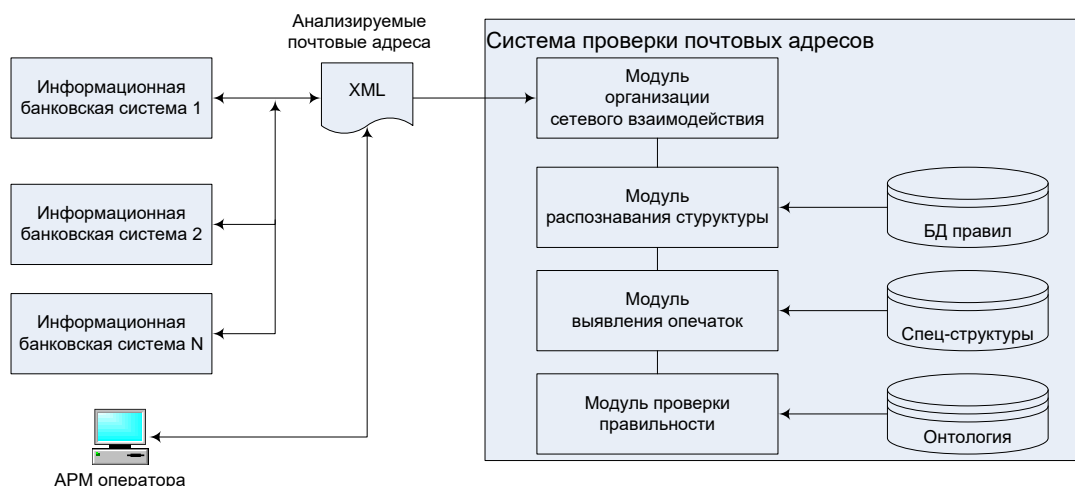


Рис. 1. Обобщенная структурная схема системы

**Модуль организации сетевого взаимодействия** обеспечивает информационное взаимодействие с ИС банка, заинтересованными в проверке почтовых адресов. Модуль реализует доставку Системе почтового адреса или группы из нескольких почтовых адресов от ИС банка и возврат исправленного адреса/адресов с сопутствующим информационным сообщением на естественном языке.

**Модуль распознавания структуры адреса** выполняет выделение структурных элементов адреса, представленного сплошной строкой или группой сцепленных строк, таких как наименование региона, города, улицы, номер дома, квартиры и др. Модуль использует набор правил, хранящихся в базе данных. База данных правил, представляет собой список правил выявления в сплошных адресных строках структурных элементов адреса. Данный список формируется Модулем настройки распознавателя.

**Модуль настройки распознавателя** использует обучающую выборку эталонных адресов, в которых человеком-оператором явно выделены структурные элементы адреса.

**Модуль выявления опечаток** реализует алгоритм нечеткого сопоставления строк и применяется для выявления опечаток, допущенных оператором при формировании адреса. Данный модуль использует специальные хранимые структуры для организации быстрого поиска «наиболее подходящей» строки-наименования из имеющихся эталонных наименований. Специальные хранимые

структуры формируются на основе внешнего авторитетного источника наименований - классификатора адресов КЛАДР.

**Модуль проверки правильности** обеспечивает выявление несоответствия проверяемого адреса имеющейся онтологии географических наименований, полученной Модулем конвертирования из актуальной версии классификатора КЛАДР. Онтология хранит как эталонные адреса, так и правила выявления несоответствий.

**АРМ оператора** обеспечивает функционирование системы в режиме off-line. С использованием АРМ оператора осуществляется запуск процесса проверки адресов, отображение информации о ходе и результатах проверки, отображение информации о причинах невозможности исправления адреса.

Кроме того, обслуживание системы обеспечивает **АРМ администратора**, который выполняет функции администратора «Подсистемы обработки почтовых адресов», администратора «Онтологии» и администратора «Набора правил извлечения».

Описанная система может функционировать в двух режимах: on-line и off-line.

Работа в режиме on-line подразумевает оперативное автоматическое взаимодействие системы с ИС банка в процессе повседневной работы, когда в ИС банка заводятся учетные записи о новых клиентах или модифицируется информация о почтовых адресах у существующих клиентов.

Работа в режиме off-line используется, когда необходимо выполнить проверку и исправления группы адресов в пакетном режиме. Для этого оператор ИС банка должен выгрузить группу адресов конкретной ИС банка в файл в формате XML или CSV и запустить процесс проверки адресов. В ходе обработки указанного файла на экране АРМ оператора отображается индикатор прогресса (объем обработанных адресов). В случае успешного завершения процесса проверки система должна выдать на экран АРМ оператора протокол о результатах проверки, а также сформировать ответный XML или CSV - файл в виде исправленного набора адресов с информационными сообщениями (комментариями) на естественном языке для каждого из таких адресов, позволяющими судить о характере обнаруженных ошибок и действиях системы по их исправлению. При невозможности уникальной идентификации адреса система должна выдать в поле <Validity> выходного сообщения значение AddressInvalid и комментарий на естественном языке для каждого из таких адресов о причинах невозможности исправления адреса.

## 5. Подготовка к работе оператора

На подготовительном этапе осуществляется установка программного обеспечения АРМ оператора.

Установка производится с помощью инсталляционных программ ClientSetap.msi и Setap.exe. Предварительно пользователь должен определиться с местом установки и хранения программы-клиента на диске, например, D:\Program Files\AdClient\.

Запустить мастер установки ClientSetap.msi. Указать имя папки для установки программы и доступность для пользователя (все или только текущий пользователь).

Нажать кнопку Next. Идет процесс установки. По окончании установки нажать кнопку Close.

Удаление программы-клиента осуществляется стандартным методом с помощью меню «Установка и удаление программ» «Панели управления» операционной системы Windows.

## 6. Порядок работы оператора

Как уже отмечалось, работа оператора ИС банка используется в режиме off-line, когда необходимо выполнить проверку и исправления группы адресов в пакетном режиме. Для этого оператор ИС банка с помощью средств АРМ оператора должен выгрузить группу адресов конкретной ИС банка в файл в формате XML или CSV и начать процесс проверки адресов.

С этой целью необходимо запустить программу «Система проверки почтовых адресов» (Client.exe). В появившемся основном окне программы (рис. 2) в меню «Файл» выбрать пункт «Открыть».

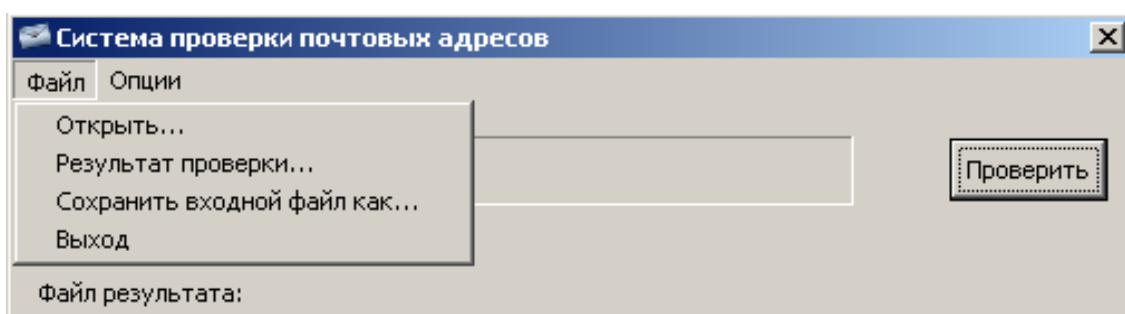


Рис. 2 Основное окно системы проверки адресов

Затем в диалоговом окне (рис. 3) выбрать соответствующий XML или CSV-файл с пакетом адресов конкретной ИС банка, которые необходимо проверить, и нажать кнопку «Открыть».

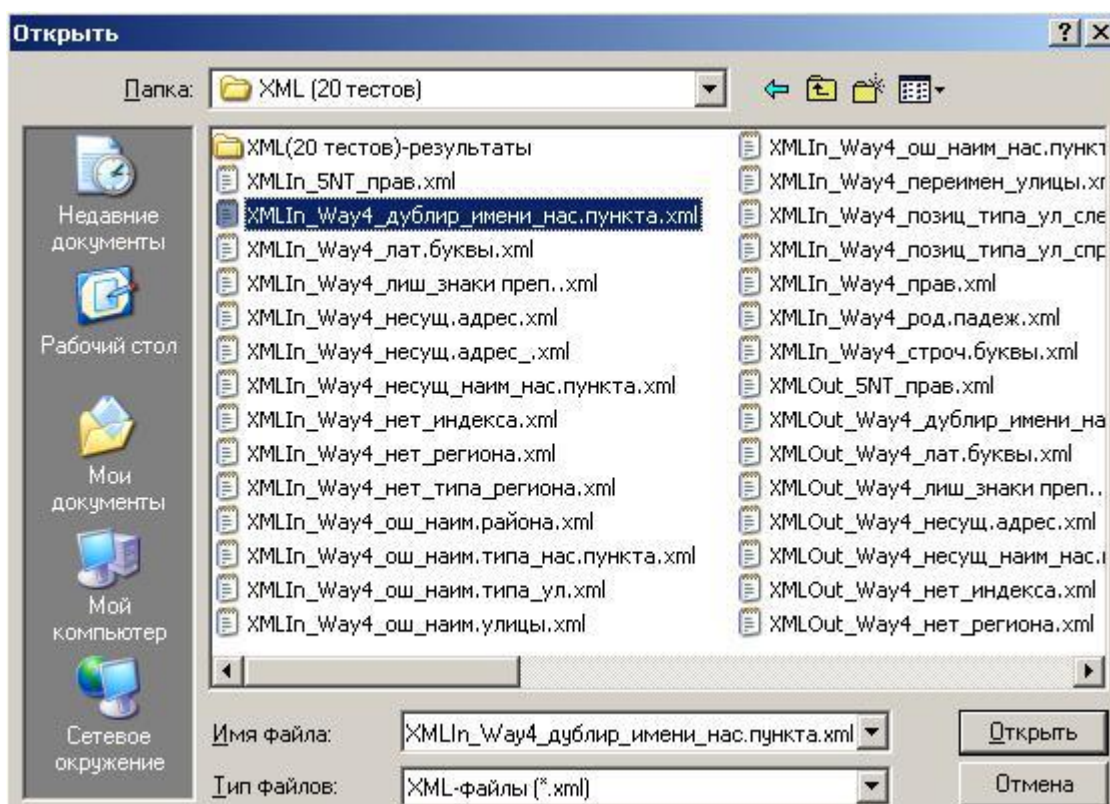


Рис. 3 Окно выбора файла адресов для проведения проверки

При необходимости предварительного преобразования входного файла из формата XML в формат CSV (или обратно) нужно сначала открыть этот файл, а

потом в меню Файл выбрать пункт «Сохранить входной файл как». В появившемся окне выбрать требуемый формат сохранения: XML или CSV.

Во вновь появившемся основном окне программы в меню «Файл» выбрать пункт «Результаты проверки» и указать имя файла и путь, где необходимо сохранить результаты проверки (Рис. 4).

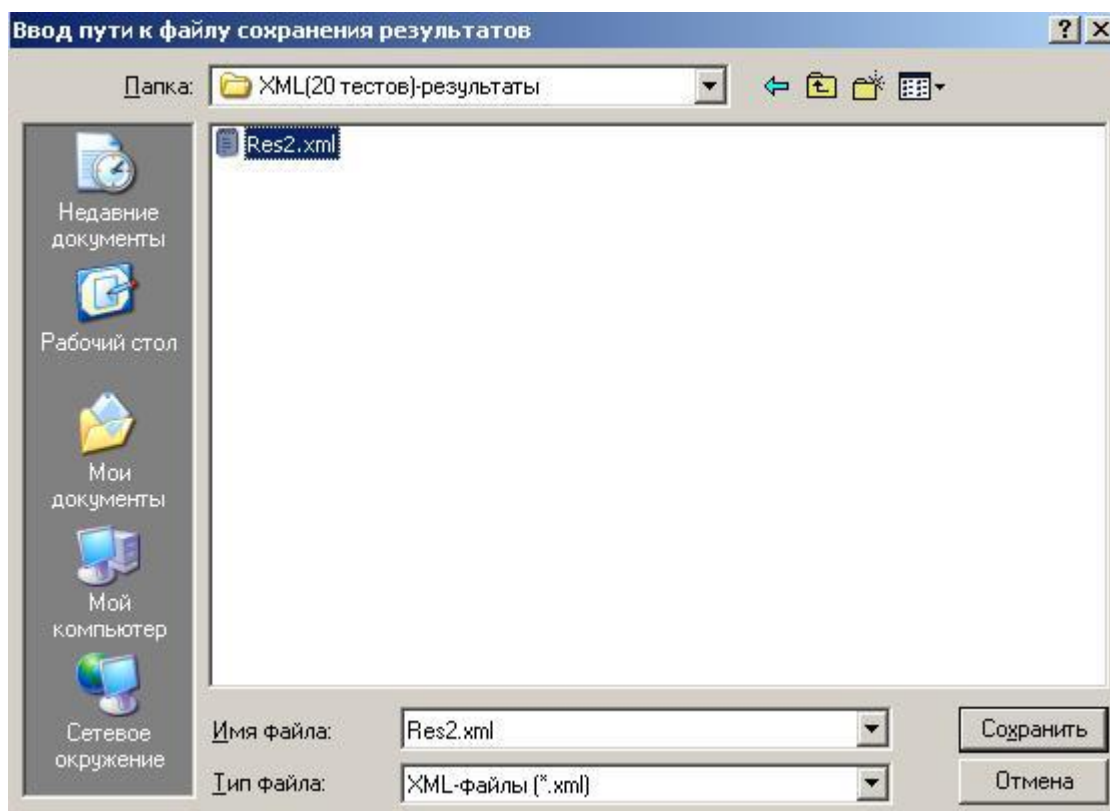


Рис. 4 Окно задания имени и типа файла результатов проверки

Одновременно задать тип файла (XML или CSV). Нажать кнопку «Сохранить».

В меню «Опция» основного окна программы выбрать пункт «Настройки» (Рис. 5) и указать (при необходимости) место на сервере, где установлен Web-сервер и место расположения файла с отчетом клиента. Кроме того, ввести значение предельного времени обработки одного адреса в секундах (при необходимости), задать размер проверяемой порции адресов (не более 100) и проставить галочку в поле «Сохранить неверные адреса в отдельный», если это требуется. Нажать кнопку ОК.



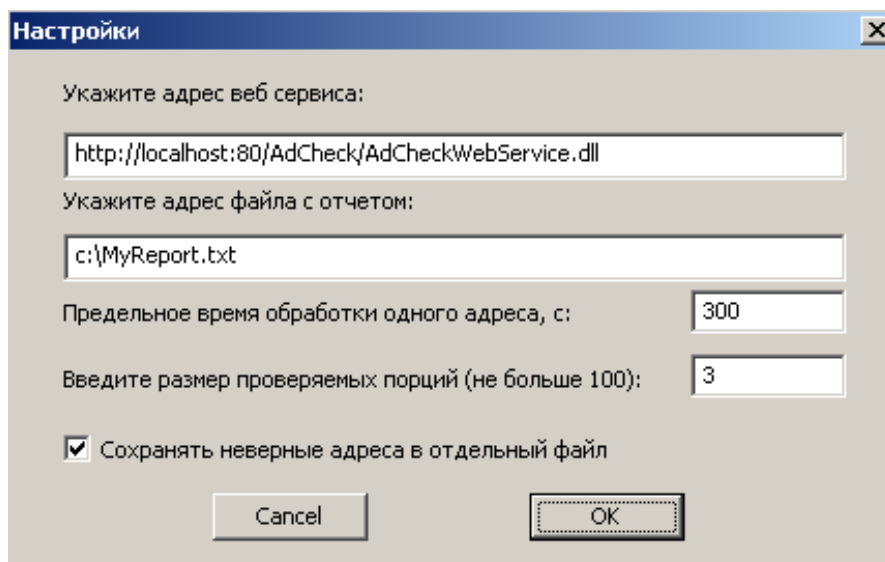


Рис. 5 Настройки программы проверки адресов в пакетном режиме

Далее в появившемся основном окне программы необходимо нажать кнопку «Проверить», т.е. запустить основную процедуру проверки пакета адресов. На экране появится визуальное отображение хода процесса проверки. Об окончании процесса свидетельствует появление на экране сообщения «Проверка завершена». При нажатии кнопки ОК на экран выдается «Протокол проверки пакета адресов» (рис. 6).

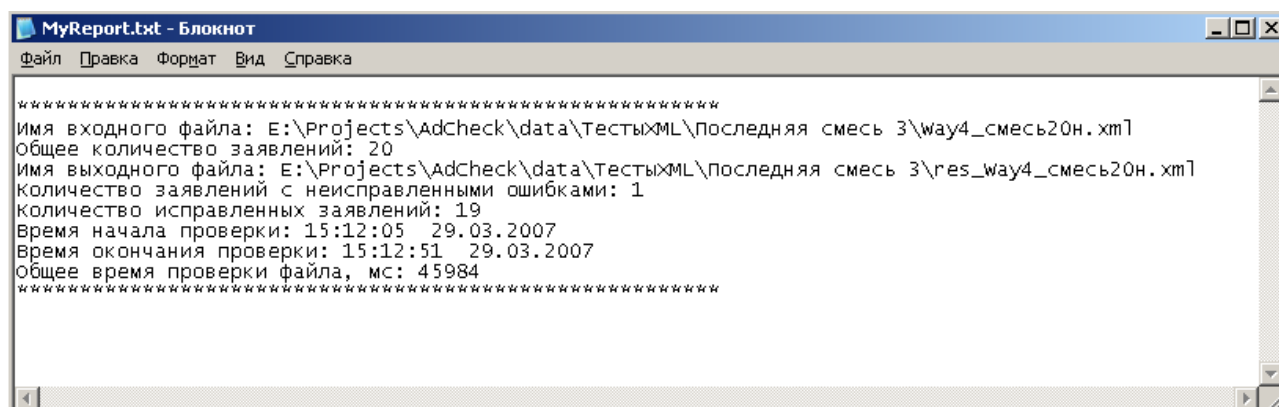


Рис. 6 Протокол результатов проверки пакета адресов

Результат проверки автоматически сохраняется в выходном XML-файле под заданным именем (рис. 5) (указывается в п. 5 Протокола). В выходном файле для каждого адреса выдаются:

- выходной (исправленный) адрес;
- информационное сообщение (комментарий системы) о результате проверки адреса;
- входной адрес, поступивший в систему для проверки.

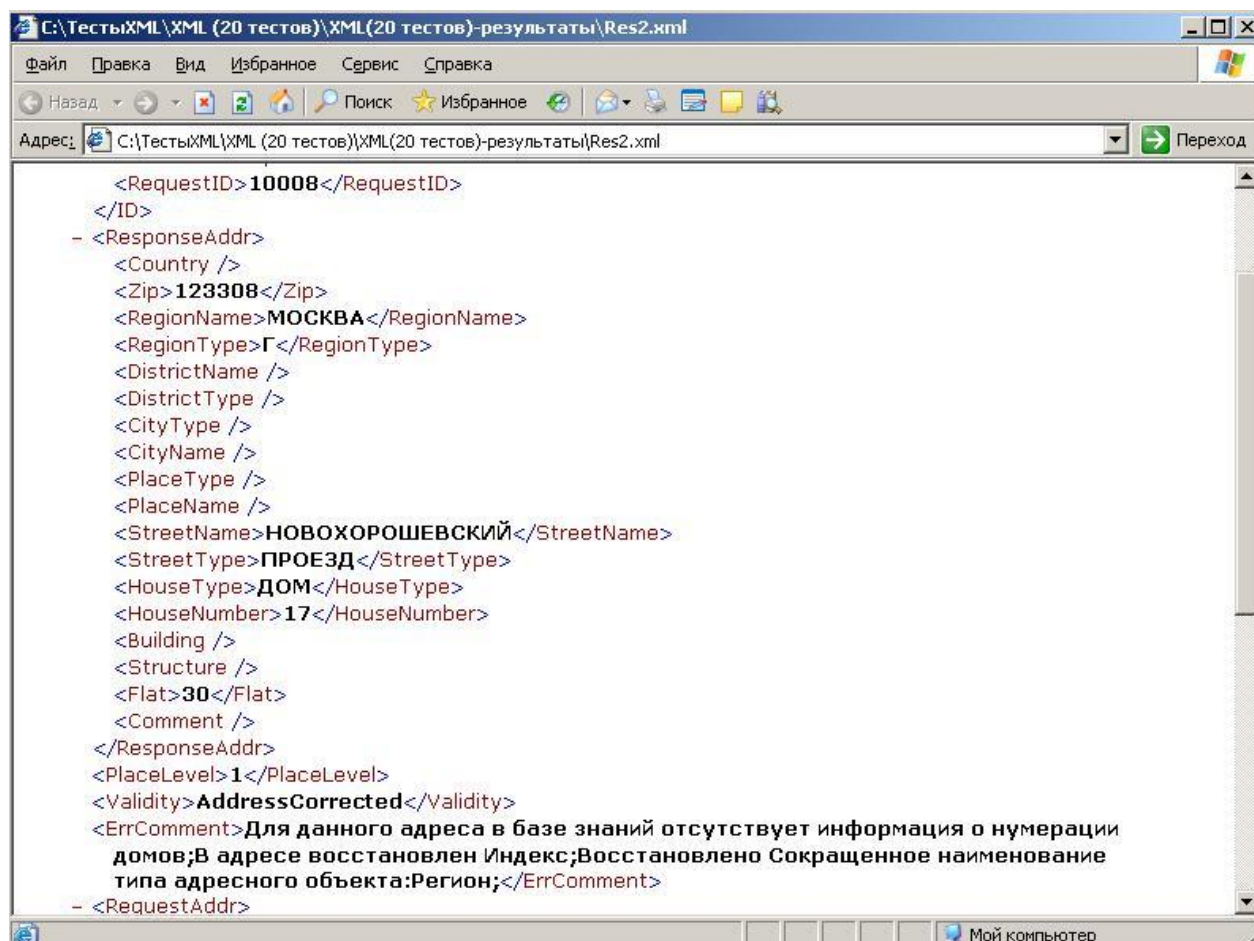


Рис. 7 Содержание выходного файла

Получив ответный XML или CSV - файл в виде исправленного набора адресов, оператор должен загрузить этот файл в базу данных соответствующей ИС банка, обновляя/дополняя её содержимое. Выполнение этой операции осуществляется с помощью программных средств, не входящих в состав системы.