

ОСОБЕННОСТИ ИНТЕГРАЦИИ СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПЕРЕВОДА В CM SERVICES ДЛЯ УПРАВЛЕНИЯ ЛОКАЛИЗАЦИЕЙ

Анашкин Р.В., Беневоленский С.Б., Кирьянов А.А.

ООО «Связь-Строй»

Рассмотрена задача перевода контента различных информационных интернет-ресурсов на иностранные языки. Существующие интернет-порталы, решающие подобные задачи, имеют ряд недостатков. В данной работе разработан программный комплекс (ПК) управления переводом, лишенный этих недостатков и обеспечивающий взаимодействие со специализированным программным обеспечением, таким как системы управления контентом (CMS) Drupal, Joomla и 1С-Битрикс. Эти CMS обладают необходимыми компонентами, наиболее важными из которых являются высокая модульность и поддержка необходимых протоколов коммуникации. При этом выполнен анализ проблем интеграции системы автоматизированного перевода в CM Services для управления локализацией. Учитывались недостатки существующих решений, заключающиеся в отсутствии возможности предоставить онлайн интерфейс переводчика с поддержкой памяти переводов и организовать обсуждение перевода по сегментам. Устранены недостатки систем управления локализации за счет внедрения специализированного модуля «Интеграция», который предназначен для перевода контента широко распространенных CMS. По полученным в процессе тестирования модуля данным сделан вывод, что серверная подсистема способна обрабатывать запросы от 100 000 обращений в день.

Ключевые слова: перевод контента, программный комплекс управления переводом, интеграция системы автоматизированного перевода, онлайн интерфейс переводчика, поддержка памяти переводов.

FEATURES OF INTEGRATION OF AUTOMATED TRANSLATION IN CM SERVICES TO MANAGE THE LOCALIZATION

Anashkin R.V., Benevolenskiy S.B., Kiryanoff A.A.,

ООО "Svyaz-Stroy"

The problem translate the various Internet information resources in foreign languages. Existing Internet portals to solve such problems have a number of drawbacks. In this paper we developed the program complex (PC) translation management without these drawbacks and provides interaction with specialized software, such as content management systems (CMS) Drupal, Joomla and 1C-Bitrix. These CMS have the necessary components, the most important of which are the high modularity and support necessary communication protocols. In this case, the analysis of the problems of integration of automated translation in CM Services to manage the localization. Taken into account the shortcomings of existing solutions is the lack of ability to provide online translation interface with support for translation memories and organize discussion of transfer segment. Address the deficiencies of the localization of management systems through the introduction of a specialized unit "Integration", which is intended to translate the common CMS. According to the obtained in the process of testing the module data, it was concluded that the server subsystem can handle requests from 100,000 hits per day.

Keywords: software, translation memory, content management services, website localization.

Введение. В настоящее время актуальна задача перевода контента различных информационных интернет-ресурсов на иностранные языки. Такое повышение актуальности обусловлено возрастанием интенсивности международных связей, процессами глобализации мировой экономики.

Цель исследования. Разработать программный комплекс (ПК) управления переводом, обеспечивающий взаимодействие со специализированным программным обеспечением, таким как системы управления контентом (CMS) Drupal, Joomla и 1С-Битрикс.

Материал и методы исследования. Существующие интернет-порталы, решающие подобные задачи [3; 4], имеют такие недостатки, как отсутствие возможности предоставить онлайн интерфейс переводчика с поддержкой памяти переводов и организовать обсуждение перевода по сегментам. В данной работе поставлена задача разработать программный комплекс (ПК) управления переводом, лишенный этих недостатков и обеспечивающий взаимодействие со специализированным программным обеспечением, таким как системы управления контентом (CMS) Drupal, Joomla и 1С-Битрикс. Эти программы обладают необходимыми компонентами, наиболее важными из которых являются высокая модульность и поддержка необходимых протоколов коммуникации.

Нами разработан ПК [5], который предназначен для перевода контента широко распространенных CMS (Joomla, Drupal и 1С-Битрикс) интернет-ресурсов через специальный модуль «Интеграция». Алгоритм процедуры запросов клиентской подсистемы, реализуемой всеми клиентскими интеграционными единицами, на первом этапе определяет тип протокола (REST или SOAP).

Основными интерфейсами построения системы межсерверного взаимодействия являются SOAP и REST [1; 6]. Программная логика этого модуля подразделяется на две полуавтономные подсистемы, условно именуемые «Клиент» и «Сервер». Сервер реализован как стандартная подсистема ПК (с наличием специальных элементов управления на веб-страницах интерфейса административной панели, откуда можно управлять зарегистрированными внешними системами и осуществлять вызов алгоритмов непосредственно из управляющего кода фреймворка Django [2] и т.п.). Клиентская подсистема для каждой из интегрированных CMS представляет собой отдельную программу, которая непосредственно взаимодействует с кодом самой CMS и предоставляет графическую пользовательскую консоль для неё, а не для ПК.

На рисунке 1 изображена блок-схема алгоритма процедуры запросов клиентской подсистемы, которая должна реализовываться всеми клиентскими интеграционными единицами. На первом этапе определяется тип протокола (REST или SOAP).

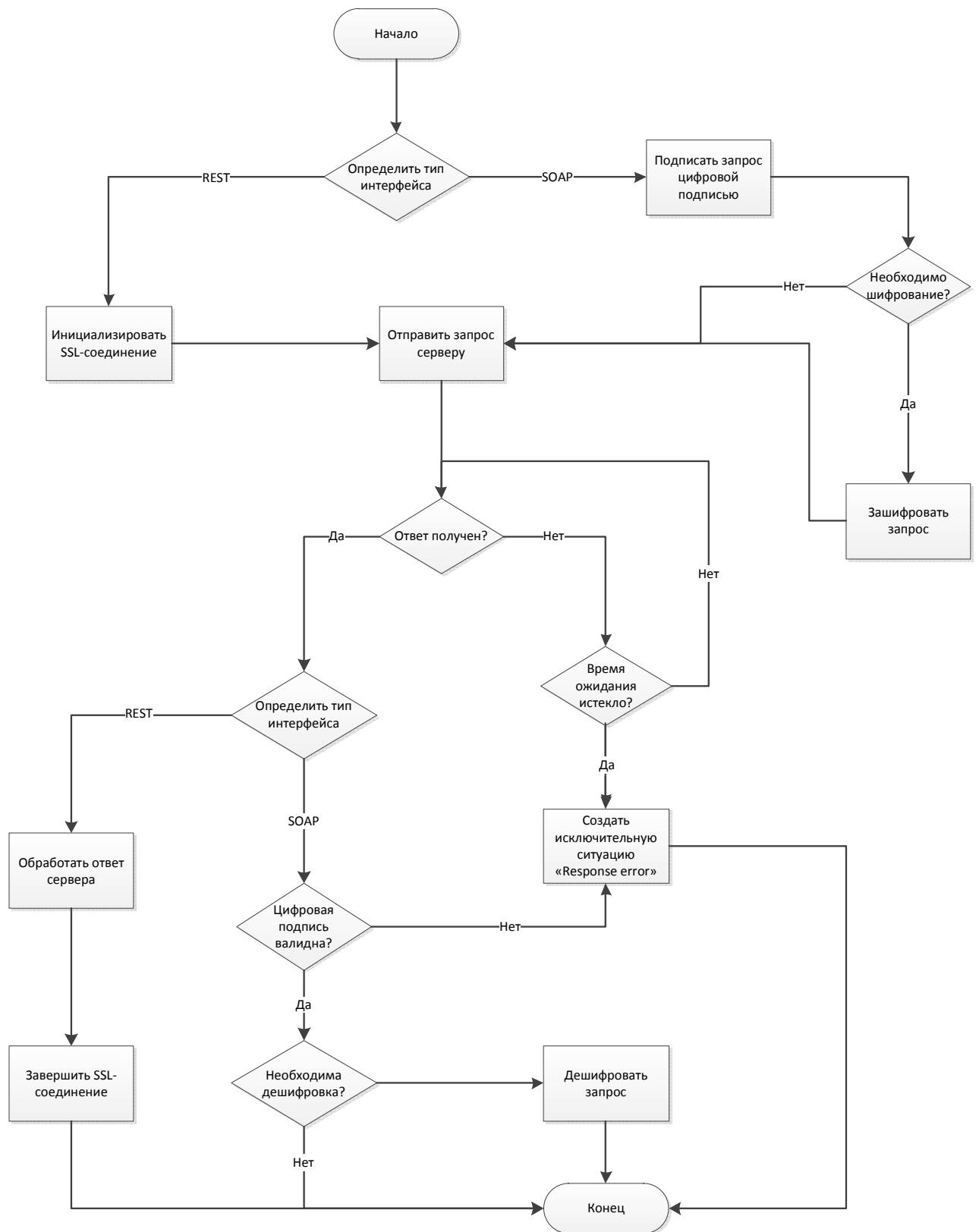


Рисунок 1. Блок-схема алгоритма процедуры запросов клиентской подсистемы.

В случае REST происходит инициализация закрытого SSL-соединения. В случае же SOAP - генерация запроса и внедрение в него цифровой подписи. Также опционально (в зависимости от параметров, заданных пользователем и/или системой) может происходить полное шифрование всего запроса. Затем происходит обращение клиентской подсистемы к серверному интерфейсу.

Если ответа от сервера нет, но время ожидания ещё не истекло, программа входит в ждущий режим. Когда ответа от сервера по-прежнему нет и время ожидания подходит к концу, подсистема создаёт исключительную ситуацию (на которую CMS должна каким-либо образом отреагировать) и аварийно завершается. В случае же успешного возвращения данных сервером алгоритм проводит деинициализацию соединения (при REST) или (при SOAP) проверку подлинности и дешифровку (опционально) и завершается, возвратив данные вызывающему алгоритму. Роль клиента предусматривает осуществление коммуникации при помощи унифицированных согласованных запросов REST-API между серверным модулем ПК (а через него и со всеми остальными необходимыми для выполнения задачи модулями) и подсистемами CMS, для которой клиент реализован. Модуль «Интеграция» позволяет осуществлять авторизацию в ПК при помощи аккаунтов популярных социальных сетей.

Для успешной разработки и полноценного тестирования работы программного модуля «Интеграция» необходимо наличие, помимо Серверной (основанной на плагине `django-piston` и изначально встроенной в ПК), также и работающей Клиентской подсистемы. С этой целью были разработаны полнофункциональные клиентские полуавтономные программные подсистемы для внешних CMS (Joomla, Bitrix, Droopal). На рисунке 2 приведен алгоритм работы данного клиента.

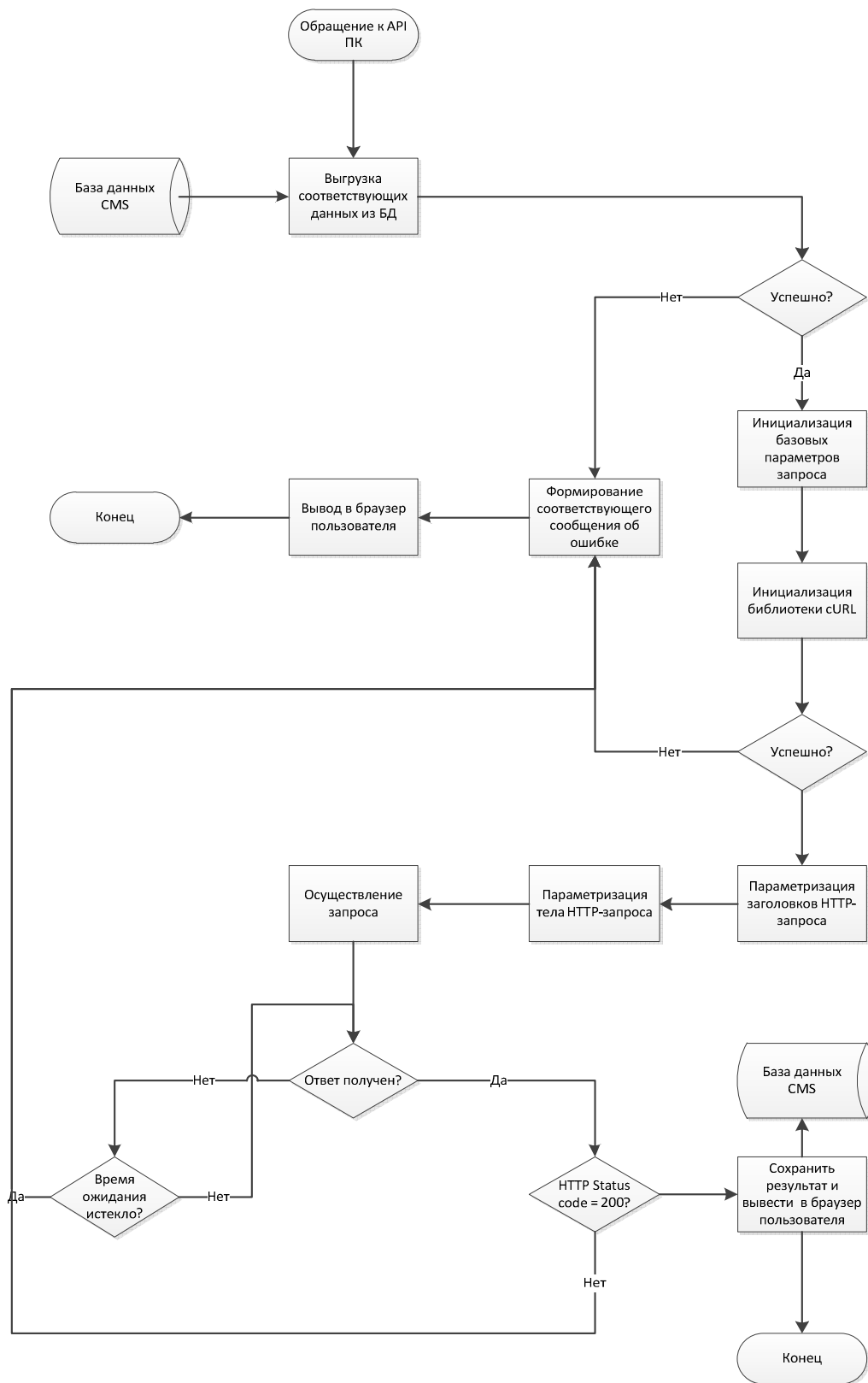


Рисунок 2 – Алгоритм работы клиентской подпрограммы модуля «Интеграция»

На начальном этапе происходит инициализация необходимых для осуществления запроса к API данных (уникальный идентификатор заявки, уникальный идентификатор элемента, целевой язык перевода, дата и время создания заявки и её поступления к переводчику). Далее происходит запуск и инициализация библиотеки PHP-привязки

библиотеки к URL (т.к. именно она является основным инструментом программного межсерверного взаимодействия). Если инициализация библиотеки не удалась, то происходит формирование сообщения об ошибке и его передача в web-браузер пользователя. В случае успеха происходит инициализация самого запроса (установка URL-ресурса, настройка заголовков BasicHTTPAuth и т.д.) и отправка его серверной подсистеме ПК. Затем алгоритм входит в цикл ожидания ответа от серверной подсистемы, выйти из которого он может в двух случаях: при истечении времени ожидания (в этом случае генерируется отчёт об ошибке и алгоритм завершается) или при получении ответа. В последнем случае данные, полученные от сервера, интерпретируются, сохраняются в базу данных CMS и отображаются в web-интерфейсе CMS.

Программный модуль «Интеграция» также располагает функционалом, позволяющим пользователям вместо создания учётной записи непосредственно в ПК использовать для аутентификации и авторизации уже существующую учётную запись какого-либо из указанных сервисов: mail.ru, yandex.ru, facebook.com, vkontakte.ru, twitter.com.

Модуль «Интеграция» принимает в качестве входных данных запросы в формате SOAP или REST API, передаваемые по протоколу HTTP, и выводит XML-документы, передаваемые по протоколу HTTP.

Результаты исследования и их обсуждение. ПК и модуль «Интеграция» были подвергнуты тестированию для исследования эффективности функционирования. При этом анализировалась динамика статистической взаимосвязи между сложностью алгоритма генерации страницы и стабильностью работы ПК. В качестве критерия стабильности было выбрано значение времени ответа серверной подсистемы в каждой транзакции (т.е. одиночной последовательности атомарных действий вида «клиент запрашивает данные сервера – ПК генерирует данные – сервер отправляет данные клиенту»).

Для имитации алгоритма был выбран модуль поиска по базе данных TMX (TranslationMemoryExchange — Обмен памятью переводов), запросы к которому инициировались в момент обращения к странице. Результаты тестирования при поиске по базе данных размером 1МВ приведены на рисунках 3 и 4.

На основании полученных данных можно сделать вывод, что серверная подсистема ПК является в достаточной мере устойчивой и способна обрабатывать запросы от 100 000 обращений в день, однако время обработки запроса в значительной мере зависит от размера файла базы памяти переводов, хранящей переводы сегментов в формате TMX. Пиковые значения, как правило, относятся к настройкам web-сервера Apache и возникают в те моменты, когда происходит запуск дополнительных экземпляров его процесса, что происходит в результате балансировки нагрузки.

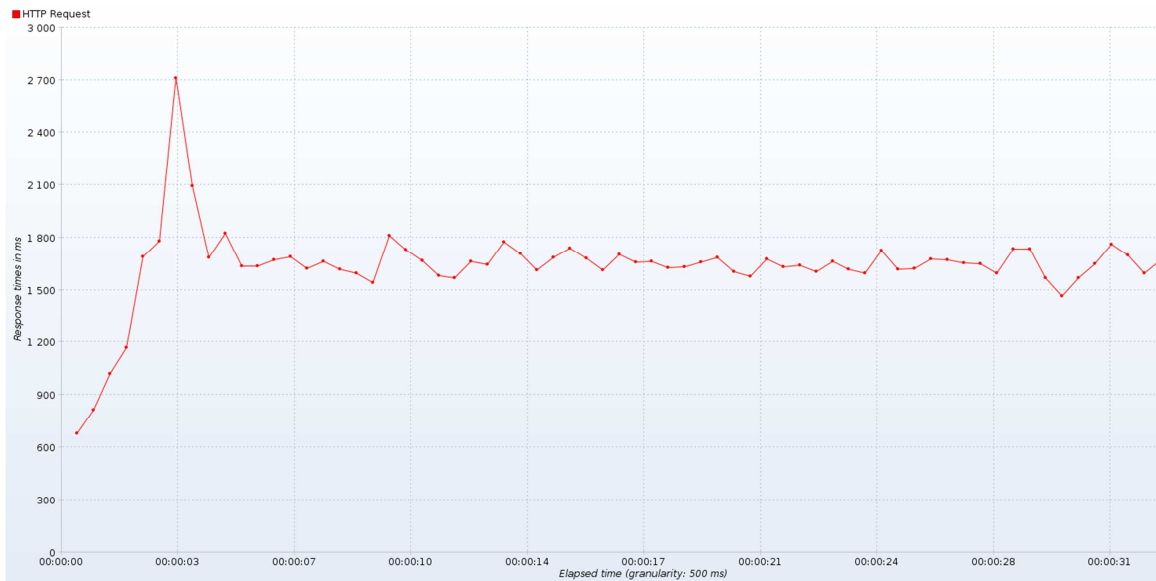


Рисунок 3 - 100 клиентов в секунду производят обращения к ПК.

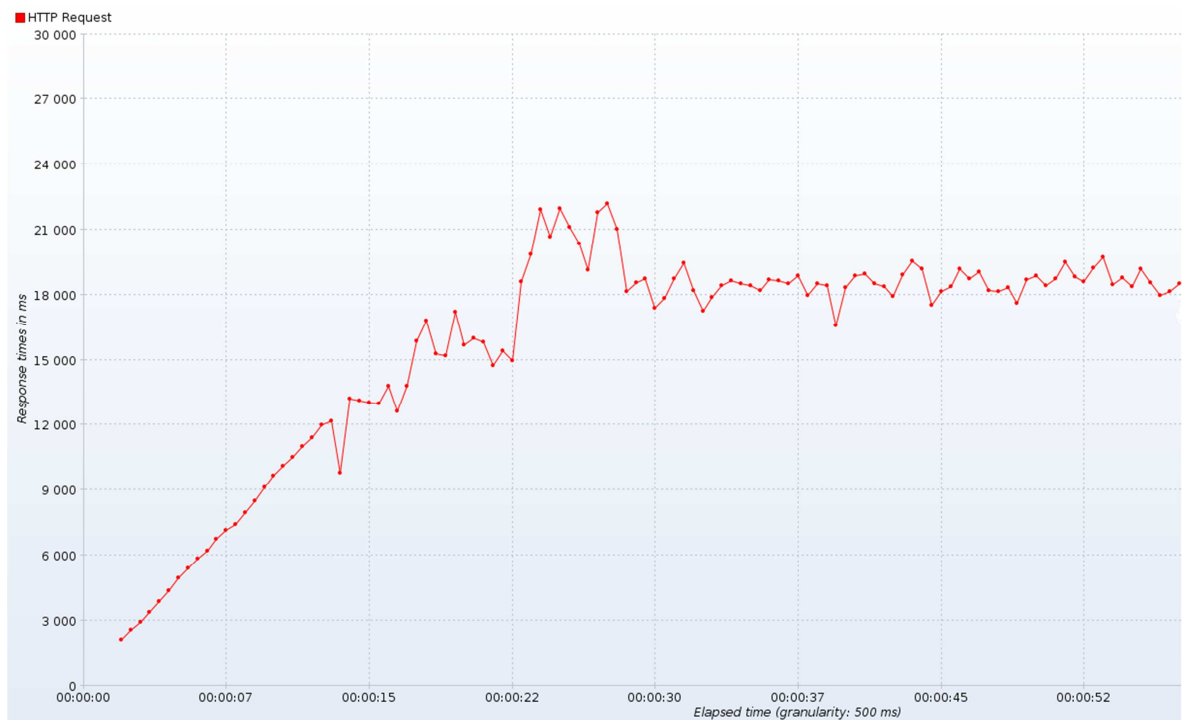


Рисунок 4 - 1000 клиентов в секунду производят обращения к ПК.

Выводы. Таким образом, разработан программный комплекс (ПК) управления переводом, обеспечивающий взаимодействие со специализированным программным обеспечением, таким как системы управления контентом (CMS) Drupal, Joomla и 1С-Битрикс.

Работа выполняется при поддержке Министерства образования и науки РФ (государственный контракт № 07.524.11.4020).

Список литературы

1. Введение в службы RESTful с использованием WCF [Электронный ресурс]. - Режим доступа: <http://msdn.microsoft.com/ru-ru/magazine/dd315413.aspx> (дата обращения: 08.04.13).
2. Головатый А., Каплан-Мосс Дж. Django. Подробное руководство = Django. The definitive guideto / пер. с англ. — СПб. : Символ-Плюс, 2010. — 560 с. — (High Tech). — ISBN 978-5-93286-187-5
3. Качественный и быстрый перевод текстов онлайн [Электронный ресурс]. - Режим доступа: <http://mrtranslate.ru/perevod.php> (дата обращения: 08.04.13).
4. Kochereshkin K. Transifex помогает пользователям всего мира локализовать продукты ROSA [Электронный ресурс]. – Режим доступа: <http://www.rosalab.ru/blogs/transifex-pomogaet-polzovatelyam-vsego-m> (дата обращения: 08.04.13).
5. Translateok.com [Электронный ресурс]. – Режим доступа: <http://advisor.wmtransfer.com/SiteDetails.aspx?url=translateok.com&tab=r1> (дата обращения: 08.04.13).
6. SOAP [Электронный ресурс]. - Режим доступа: <http://ru.wikipedia.org/wiki/SOAP> (дата обращения: 08.04.13).

Рецензенты:

Марсов В.И., д.т.н., профессор кафедры «Автоматизация производственных процессов» Московского автомобильно-дорожного государственного технического университета (МАДИ), г. Москва.

Галушкин А.И., д.т.н., профессор, начальник лаборатории «Интеллектуальные информационные системы» Федерального государственного научного учреждения «Центр информационных технологий и систем органов исполнительной власти», г. Москва.