

**Служба мониторинга как средство повышения
эффективности работы
распределенной библиотечной сети АРБИКОН**

**Monitoring Service as a Means of Operation Efficiency Increase
in the ARBIKON Distributed Library Network**

Р. Т. Усманов

*Институт корпоративных библиотечных информационных систем
Санкт-Петербургского государственного политехнического университета,
Санкт-Петербург, Россия*

А. А. Кузнецов

*Факультет технической кибернетики Санкт-Петербургского государственного
политехнического университета, Санкт-Петербург, Россия*

Rustam Usmanov

*Institute of Cooperative Library Information Systems of St.Petersburg State Polytechnical University,
St.Petersburg, Russia*

Alexey Kuznetsov

*Faculty of Technical Cybernetics of St.Petersburg State Polytechnical University,
St.Petersburg, Russia*

Рассмотрены принципы создания службы онлайн-мониторинга состояния Z39.50-серверов, входящих в пространство АРБИКОН. Определены способы динамического реконфигурирования системы для оптимизации времени отклика на поисковые запросы.

В нашей стране процесс формирования распределённых библиотечных сетей выявил несколько специфических проблем. Обширные территории, дороговизна использования качественных каналов связи, большой багаж несовместимых форматов баз данных, желание сэкономить на дорогом, но качественном программном обеспечении и ещё ряд других причин привели к тому, что некоторые члены консорциумов не полностью удовлетворяют корпоративному профилю АРБИКОН в области стандартов на обеспечение доступности библиографических баз данных. В связи с этим, требуется разработка дополнительных решений для повышения качества функционирования распределенной библиотечной сети.

В докладе рассмотрены основные проблемы, которые явились предпосылками к разработке службы автоматического мониторинга.

Если объединить схожие проблемы и обобщить понятия, то для формирования критериев эффективности распределенной библиотечной сети АРБИОН можно выделить три основных фактора:

- проблема доступности Z-ресурса;
- проблема несоответствия Z-ресурса параметрам заявленного корпоративного профиля;
- проблема большой временной задержки при доступе к Z-ресурсу.

Здесь вводится новое понятие «Z-ресурс», описывающее совокупность программно-аппаратных средств оконечного узла распределенной библиотечной сети, представляющего собой базу данных, расположенную на одном из серверов Z39.50.

Для измерения эффективности работы распределённой сети АРБИКОН введём несколько новых категории понятий, относящихся к количественной оценке эффективности Z-ресурсов и оценки эффективности распределённых сетей в целом.

Введена следующая система индикаторов:

- 1) Индикатор текущей доступности Z-ресурса – «I_{тд}».
- 2) Индикатор интегральной доступности Z-ресурса – «I_{ид}».

3) Индикатор полноты соответствия Z-ресурса – «I_C».

Для определения индикатора полноты соответствия Z-ресурса корпоративному профилю распределённой библиотечной сети предложены следующие индикаторы:

3.1) Поддержка необходимой версии протокола Z39.50 – «S_Z».

3.3) Поддержка требуемых кодировок.

3.4) Поддержка требуемых форматов записи.

3.5) Поддержка требуемых точек доступа.

На основании анализа данных журнала пользовательских запросов программы широковещательного Z-сервера за 2005 год, индикатор учитывает ранжирование значимости точек доступа.

4) Индикатор времени доступа к Z-ресурсу – «I_t».

5) Индикатор эффективности функционирования распределённой сети – «Net».

Существующая распределённая система, в которой осуществляется широковещательный поиск не предусматривает обратной связи, поэтому, при возникновении ситуации, когда Z-ресурс недоступен, техническая служба АРБИКОН может и не узнать, что ряд партнёров не только не обеспечивают надлежащего корпоративного сервиса, но и вообще не предоставляют доступ к своим ресурсам.

На основании анализа данных, полученных при анализе журнала широковещательного Z-сервера, был сделан вывод, что построение аналитической модели в терминах теории массового обслуживания не представляется возможным, так как интервалы времени между различными поисками не подчиняются экспоненциальному распределению, для которого существуют расчётные формулы. Исследование по методу Колмогорова-Смирнова на предмет соответствия экспоненциальному распределению не дали положительного результата – вероятность такой гипотезы «Asymp. Sig. (2-tailed)» равна 0. Поэтому было применено имитационное моделирование. В качестве инструмента имитационного моделирования выбран пакет AnyLogic, агентный подход.

Для создания агентной имитационной модели был предложен ряд иерархически связанных элементов модели:

- широковещательный сервер;
- консорциумы;
- серверы Z39.50;
- базы данных.

Структура распределённой сети динамически строится на основании последних реальных данных о доступности и параметрах Z-ресурсов, полученных в результате работы программы Z-робота. Создана специальная подпрограмма на Java (при помощи библиотеки DBXML) позволяющая прямо из модели обращаться к файлу базы данных Z-робота для получения результатов последнего опроса реальных серверов Z39.50. Модель не только содержит реальное количество баз данных, серверов и консорциумов, но и каждый из этих элементов имеет реальные характеристики в текущий момент времени (см. рис. 2).

При щелчке на базе данных или сервере Z39.50 происходит показ полной информации о выбранном объекте, полученный при работе Z-робота.

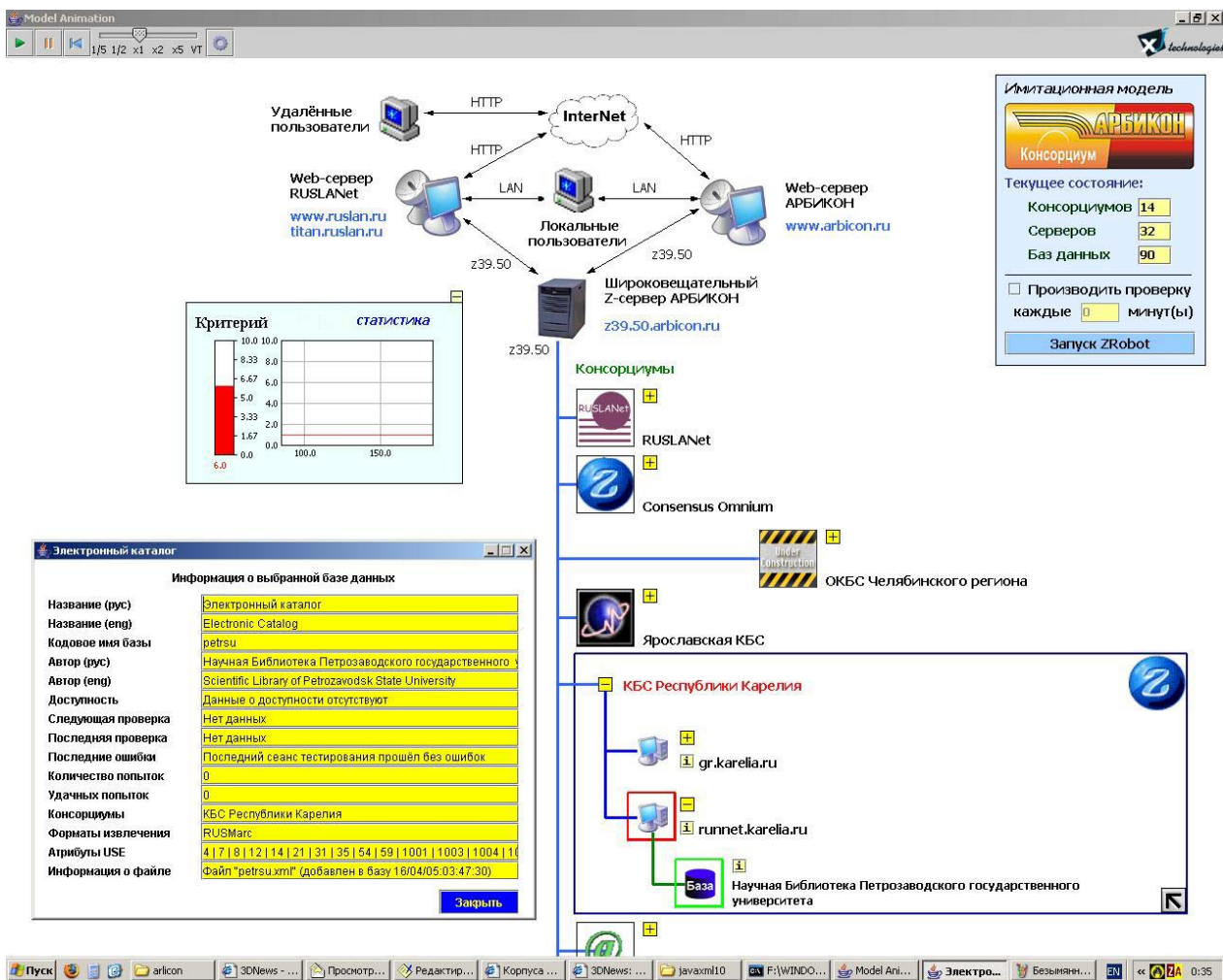


Рис. 2. Пользовательский интерфейс модели

Для ускорения реакции на запрос широковещательного сервера был предложен новый метод поиска. Для этого в цикл работы широковещательного сервера введён ряд программных продуктов, позволяющих заранее исключать из поиска заведомо неработающие сервера. Предложенный комплекс ПО, интегрированный в действующую инфраструктуру распределённой сети, изображён на рис. 3.

Состав ПО:

- Z-робот (AZSpy – Another Z39.50 Spy), программа периодически подключающаяся ко всем Z-ресурсам распределённой сети для проверки их работоспособности и соответствия параметрам корпоративного профиля;
- конфигурационный XML-файл, в котором задаются параметры работы программы Z-робота (например, количество одновременно опрашиваемых серверов, период сброса статистики, время таймаута, через которое Z-ресурс признаётся неработоспособным);
- описание серверов (DBXML) – база данных, в которой хранятся информация обо всех Z-ресурсах (сервер, порт, база данных) и результаты их тестирования программой Z-робота;
- дополнительное ПО – набор программ для начальной загрузки информации о Z-ресурсах в базу данных описания серверов из уже имеющихся файлов формата XML и анализа журнала широковещательного сервера распределённой сети;
- модель АРБИКОН – модель распределённой библиотечной сети,
- база данных статистики – база данных содержащая статистическую информацию о различных распределениях времен доступа реальной распределённой сети, информация в которую загружается при помощи разработанного дополнительного ПО;

– модуль генерации описания источника – программа, которая при изменении параметров сети (например, Z-ресурс стал недоступен), генерирует новый файл XML описания источника, не содержащий недоступных Z-ресурсов;

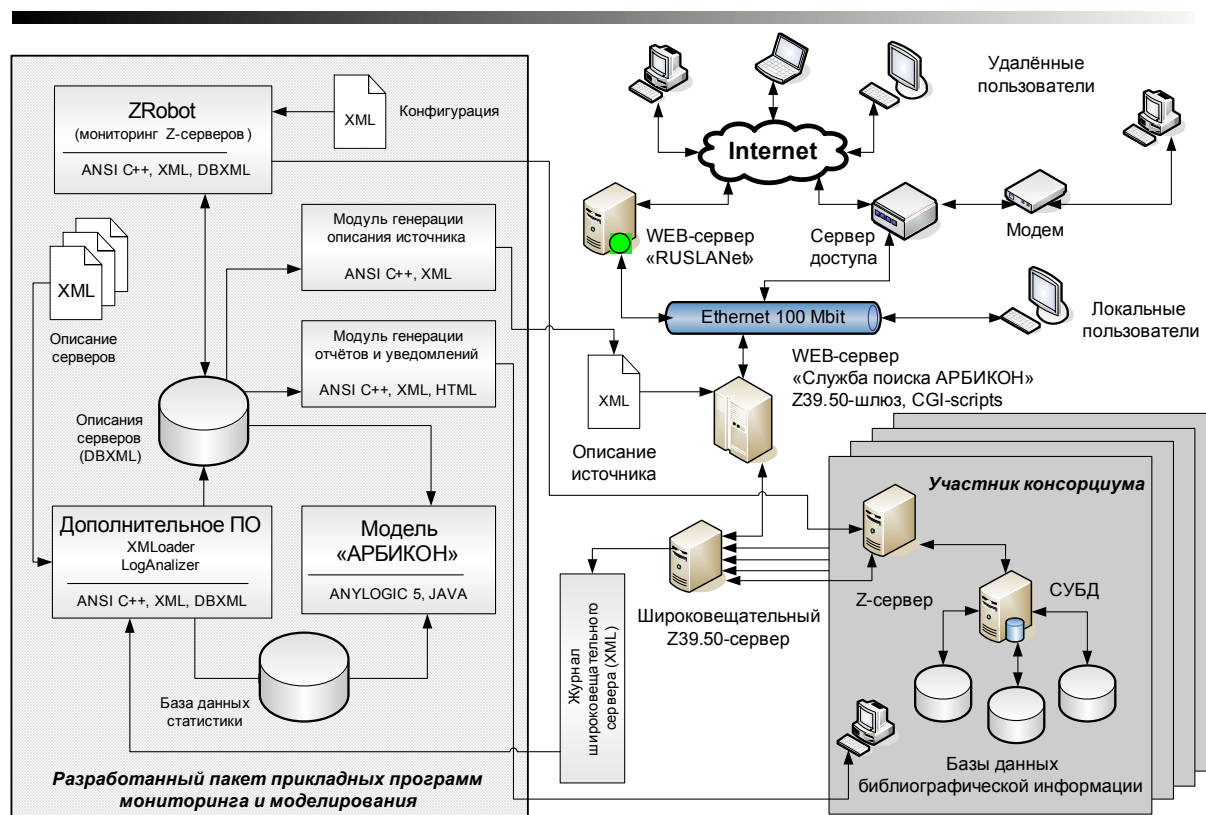


Рис. 3. Схема взаимодействия комплекса ПО и распределённой информационно-библиотечной сети

– модуль генерации отчётов и уведомлений – программа, создающая и обновляющая статистику по z-ресурсам в формате HTML для публикации в WEB, а также рассылающая информационные письма администраторам недоступных z-ресурсов.

Робот представляет собой многопоточное приложение, каждый поток программы по оригинальному алгоритму осуществляет проверку одного из Z-ресурсов (одна из баз на одном из серверов Z39.50). Процедуру мониторинга Z-роботом распределённой сети можно гибко настраивать под текущие нужды. Реализовано два типа проверки состояния Z-ресурсов: полная и сокращённая. Собранная о Z-ресурсах информация записывается в базу данных описания серверов, на основании которой обновляется структурная организация агентов в интерактивной модели сети. Так же, на основании информации, сохраненной в базе данных, после завершения тестирования, методом выгрузки через специальную программу, создаётся XML-файл, отражающий структуру и состояние распределённой сети в данный момент времени..

Далее, путём поэтапного извлечения из глобального XML-файла информации только о доступных Z-ресурсах каждого из консорциумов, получаем локальные XML-файлы, на основе которых генерируются списки серверов для страниц поискового портала. Поскольку данные списки теперь не содержат недоступные серверы, то уже не приходится ожидать минутного таймаута для получения результатов поиска. Практика тестирования технологии на медицинском портале показала работоспособность предложенной концепции и практически мгновенное получение запрошенной информации при двух неработающих серверах. Ранее, в худшем случае, при таймаутах на серверах, мы получали 40 секундные задержки при выполнении тех же запросов.

В результате внедрения предложенного комплекса ПО планируется получить сокращение среднего времени поиска с текущих 37,7 сек., до 1-2 сек.

Комплекс ПО используется технической службой АРБИКОН. В принципе, это комплекс может быть использован любой сетью Z39.50 серверов – консорциумом, крупной организацией, ведомственной сетью и пр.