

Как решение предложена унифицированная формальная модель распределенной информационной системы и технология интеграции ее компонентов, которая позволяет модернизировать и масштабировать высоконагруженные системы, ориентируясь на поставленные перед ней задачи без изменения общей логики их функционирования.

Для повышения эффективности предлагаемого решения осуществлен переход от синхронного взаимодействия между компонентами распределенной системы к асинхронному. Кроме того, для эффективного распределения нагрузки и уменьшения времени обработки запросов конечных пользователей используется зеркалирование (mirroring) находящейся на серверах информации.

В данной работе предложена технология интеграции компонентов распределенных информационных систем, основанная на использовании многопараметрических статистических моделей, что обеспечивает системных администраторов высоконагруженных распределенных информационных систем возможностью видеть полную картину сетевого взаимодействия, а также позволяет принимать эффективные меры по распределению нагрузок, масштабированию либо переработке существующей инфраструктуры или отдельных компонентов распределенных информационных систем.

Литература

1. Фаулер М. Архитектура корпоративных программных приложения.— М.: Издательский дом “Вильямс”, 2007. – 544 с.

МОДЕЛЬ ВЫЦВЕТАНИЯ ГОЛОСОВ В ВЕРСИОННЫХ СИСТЕМАХ

Резваткина А. А., Катков Ю. В.

*Санкт-Петербургский государственный университет информационных технологий
механики и оптики, Санкт-Петербург, Российская Федерация,
e-mail: linka.a.re@gmail.com, katkov.juriy@gmail.com*

В мире используется ресурсы, содержащие динамичные данные. Такими данными необходимо управлять. Удобную работу обеспечивает система, запоминая различные версии. Версионность контента позволяет расширить возможности применения системы в нескольких направлениях.

Во-первых, можно позволить пользователям редактировать основной контент ресурса, не опасаясь его «испортить».

Во-вторых, хранение любого динамически развивающегося контента, править который могут несколько пользователей.

Версионная система становится защищенной от сбоев при наличии множества версий, среди которых есть верная версия и средство выбора верной версии [1].

Средством выбора верной версии может выступать экспертная оценка. В системе веб-ресурса, экспертной оценкой можно считать статистически среднюю оценку пользователей, участвующих в оценивании статьи.

При появлении возможности оценивать статью версионной системы возникает задача расчета среднего значения рейтинга статьи. Вики — динамичная система, способная кардинально изменить свое содержание за одно или несколько изменений. Способы расчета средних значений рейтинга для статичных систем, в данном случае

являются неактуальными. Возникает необходимость обнуления оценок для версий, в значительной степени отличающихся от конечной ревизии.

Сброс голосов существующей версии рассматриваемой вики-системы носит название “истекшие рейтинги”. Такие рейтинги автоматически сбрасывают показания через 30 изменений на странице, относительно последнего голосования пользователя [2]. Недостаток модели - отсутствие зависимости обнуления голосов от важности изменений.

Для сравнения содержания версионной системы предлагается использовать алгоритмы поиска нечетких дубликатов. Значение выхода алгоритма (“степень похожести версий”) следует связать с коэффициентом выцветания оценки.

Основные типы методов поиска нечетких дубликатов:

1) Синтаксические методы — выбор последовательностей символов, слов и предложений. Основным подходом является шинглирование — очищение от разметки, представление в виде цепочек символов — шинглов. Каждой цепочке сопоставляется хеш-код, равенство кодов говорит о высоком сходстве цепочек [3].

Синтаксические методы выбраны в качестве основы будущей модели.

2) Лексические методы — выбор представительных языковых единиц. Создание словаря стоп-слов и словаря значимых слов, который подразумевает исключение первого и должен давать устойчивость к незначительным изменениям.

3) Анализ количества изменившихся символов. Метод вычисления расстояния Левенштейна - минимальное число элементарных операций редактирования, необходимых для преобразования одной строки в другую.

Минусы данного подхода в низкой устойчивости к малым исправлениям, изменению позиции абзацев и увеличение длины и количества строк приводит к квадратичному росту вычислительных затрат [4].

Предложена модель выцветания голосов в версионных системах, основанная на поиске нечетких дубликатов и корректировке оценки (голоса) пользователя в зависимости от степени схожести версий. Работа находится на окончании стадии проектирования, после которой будет произведена пробная реализация модели применительно к вики-системе MediaWiki[5]. Результаты работы могут быть внедрены как в больших общественных вики (проекты Фонда Викимедиа, Wikia), так и в малых корпоративных и обучающих вики, где важно слежение за качеством содержимого.

Литература

1. Distributed Version Control Systems: A Not-So-Quick Guide Through [Электронный ресурс]. Режим доступа (свободный): <http://www.infoq.com/articles/dvcs-guide>.
2. Article feedback tool [Электронный ресурс]. Режим доступа (свободный): http://en.wikipedia.org/w/index.php?title=Wikipedia%3AArticle_Feedback_Tool&action=historysubmit&diff=457623055&oldid=456548940.
3. C. Hoad, J. Zobel. Methods for identifying versioned and plagiarized documents. - Journal of the American society for information science and technology. Т. 54. С. 203–215, 1 February 2003. Режим доступа (свободный): <http://onlinelibrary.wiley.com/doi/10.1002/asi.10170/full>.
4. Зеленков Ю.Г., Сегалович И.В. Сравнительный анализ методов определения нечетких дубликатов для Web-документов. Труды девятой всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" - 2007. Режим доступа (свободный): http://rcdl2007.pereslavl.ru/papers/paper_65_v1.pdf.
5. Сайт проекта MediaWiki [Электронный ресурс]. Режим доступа (свободный): <http://mediawiki.org>.