

```
ct->Close();
this->Delete();
break;
case 2: this->Delete();
break;
```

При удалении записи из таблицы «Кафедры» открывается каскадная таблица «Дисциплины» и из нее удаляются все записи с требуемым кодом кафедры. Далее производится удаление записи в таблице «Кафедры».

Предложенное решение сочетает в себе достоинства как ручного, так и автоматизированного метода: позволяет самостоятельно определять действия, которые будут выполняться при удалении, модификации, добавлении записей и при этом не требует детализации последовательности действий для каждого конкретного случая, как при ручном методе. Фактически программист задает необходимые транзакции, которые позволяют работать со связанными таблицами.

Разработанный шаблон позволяет реализовать не только каскадное удаление, но и добавление, а также модификацию данных. Кроме того, он пригоден для работы с несколькими внешними ключами, что достаточно актуально в современных программных продуктах.

УДК 004.514.62

*Ненадоев И.В.*

*Научный руководитель: к.т.н., доцент Костюк Д.А.*

### ИСПОЛЬЗОВАНИЕ СТРУКТУРЫ «ОБЛАКО ТЕГОВ» ДЛЯ ДОСТУПА К АРХИВАМ ДОКУМЕНТОВ

По мере роста емкости накопителей и, соответственно, архивов документов, проблема поиска информации приобретает все более острый характер, а падение эффективности иерархических файловых систем как средства упорядочения становится все более заметным. На передний план выходят такие проблемы каталогизации, как необходимость дисциплинированного подхода к именованию и размещению файлов, вдумчивому составлению названий, а также значительные трудозатраты по разбору уже существующего хаоса.

Системы локального поиска по содержанию документов получили распространение на рабочих местах пользователей начиная с 2004 года. Разработчиками таких систем были опробованы два подхода: сканирование в процессе поиска и использование заранее составленных индексных таблиц. Поиск по первому варианту занимает слишком продолжительное время (а при использовании в локальной сети приводит к абсолютно неприемлемым задержкам); поэтому он не получил значительного распространения. Поиск с предварительным индексированием, занявший большую долю рынка, также нельзя назвать идеальным решением проблемы, поскольку операция индексирования – ресурсоемкая процедура, которая должна постоянно выполняться в фоновом режиме для обеспечения актуальности результатов поиска, тем самым заметно повышая нагрузку системы и снижая срок службы аппаратного обеспечения.

В дополнение к сказанному поиск сам по себе – отнюдь не идеальное решение. Применение вместо поиска информации визуального ориентирования в ее потоке, основанного на системе информационных тегов документов, позволяет увеличить наглядность и интуитивность интерфейса, упростить ориентирование человека в информационном пространстве.

Хотя фолксономические системы активно применяются в интернет-ресурсах с обширной базой документов, для локальных хранилищ документов их возможности не задействованы. Отчасти это вызвано их относительно недавним появлением и тем, что готовые реализации на сегодняшний день доступны лишь на языках веб-программирования.

Таким образом, облако тегов (tag cloud) – структура, называемая также взвешенным списком, используемая для визуального представления категорий – применяется в веб-технологиях. Теги, из которых формируется облако, это обычно отдельные слова, и перечисляются они, как правило, в алфавитном порядке. При этом важность тега отображается размером шрифта или цветом. Например, размер может быть функцией частоты упоминания ключевых слов. В любом случае, размер изображения тем больше, чем выше релевантность данного слова (термина, имени); и пользователь может найти нужный тег как по алфавиту, так и по релевантности. Тег обычно играет роль гиперссылки, так что активация тега осуществляет переход по заданному адресу, позволяя пользователю углубляться в поток данных [1, 2].

В памяти облако тегов хранится в виде массива взвешенных слов. Типичное облако включает от 30 до 150 тегов. Структура имеет значительно большую емкость по сравнению с такими традиционными способами отображения взвешенных массивов слов, как круговые диаграммы или гистограммы, наглядность которых резко ухудшается, когда число весовых коэффициентов превышает полтора десятка. В свою очередь, представление взвешенных массивов облаком тегов дает меньшую точность [3, 4].

Первая публикация, касающаяся облака тегов (или по крайней мере взвешенного списка) – по-видимому роман Д. Коупленда «Microserfs», вышедший в 1995 г.

Широкому кругу пользователей облако тегов оказалось впервые доступным на хостинге изображений Flickr, в виде разработки соучредителя Flickr дизайнера С. Баттерфилда [3]. В свою очередь, эта визуализация ссылок вебсайта базировалась на программе «Search Referral Zeitgeist» Дж. Фланагана, представлявшей из себя написанный на языке perl скрипт для визуализации частотного распределения источников перехода на заданную web-страницу.

Выделяют три главных типа приложений, основанных на концепции облака тегов. Типы эти отличаются скорее назначением, чем внешним видом. В первом типе вводится облако тегов для каждого объекта, тогда как во втором типе имеются глобальные облака тегов, где приоритеты объединены по всем объектам и пользователям.

В первом типе размер является функцией количества применений тега к конкретному объекту. Облако тегов первого типа полезно как средство отображения метаданных об объекте, по которому проводилось голосование среди пользователей, и при этом точные результаты не важны. Пример – музыкальный сервис Last.fm, использующий этот метод как средство отображения жанра, которым были промаркированы артист или трек.

Во втором, более широко используемом типе, размер представляет число объектов, к которым тег был применен, т.е. является отражением популярности каждого тега. Примеры этого типа облака тегов используются в хостинге изображений службы Flickr и агрегаторе блогов Technorati.

В третьем типе теги используются как метод классификации элементов контента. Теги представлены в облаке, где большие теги представляют количество элементов контента, удовлетворяющих данной категории.

Более широко та же самая визуальная техника может использоваться для визуализации нетеговых данных [1], как в облаке данных (data cloud). Облако данных представляет собой панель, использующую цвет и размер шрифта для отображения численных данных, таких как рост населения или котировки валют и др. Рисунок 1 демонстрирует пример такого применения. Размером шрифта обозначено изменение цены акции, а цвет показывает положительное или отрицательное направление изменения.

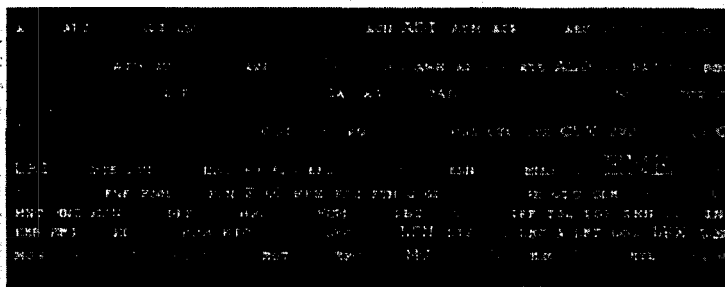


Рис. 1. Использование облака данных для отображения изменения курса акций

Облака тегов типично представляются посредством активных элементов HTML. Теги могут перечисляться в алфавитном порядке, в случайном порядке, они могут быть сортированы по значимости и др. В некоторых случаях выполняется семантическое группирование тегов так, чтобы сходные теги очутились рядом. Независимо от группирования тегов могут использоваться эвристические методы для уменьшения размера облака [3].

Для апробации предложенных навигационных решений применительно к архивам документов нами использована база текстовых документов, произвольно снабжавшаяся пользовательскими тегами. Документы добавлялись в архив постепенно, и к концу исследованного периода их число достигло 745. В подавляющем большинстве теги формировались пользователями в момент добавления документа. На рис. 2 показан фрагмент облака тегов документов, построенного в результате работы с системой.

cad dev gis gui lectures net prg sci sql srv Use vir

Рис. 2. Облако тегов документов

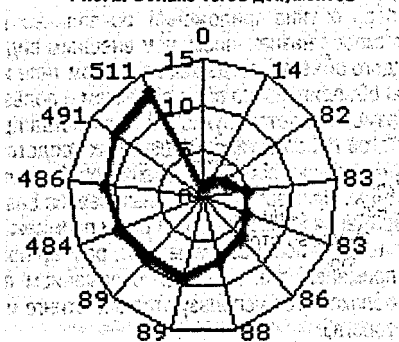


Рис. 3. Рост числа тегов за время существования системы

Рисунок 3 демонстрирует динамику роста числа тегов за время существования архива, которое указано в днях. Как можно видеть, в результате число фолксномических категорий достигло 13.

Частота использования тегов в эксперименте оказалась неравномерной. Рисунок 4 показывает число документов, промаркированных в результате каждым из 13 тегов, а также отражает для каждой категории число документов, промаркированных более чем

одним тегом. Как видно, 23% из созданного массива тегов пользуются значительной популярностью, 31% включает 5-10 документов, а остальными 46% тегов представлены уникальные и малоиспользуемые категории. Следует учитывать, что полученные соотношения могут являться функцией тематической направленности использованного архива. Для получения более точной картины частотного распределения документов по категориям необходимо дополнительное исследование с привлечением значительного количества пользователей.

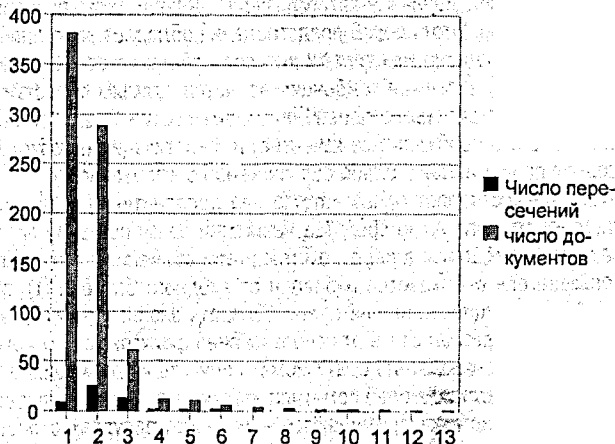


Рис. 4. Соотношение численности промаркированных тегом документов с количеством пересечений

Равномерность роста числа тегов, наблюдаемая на рисунке 3, в свою очередь свидетельствует о непрерывных коррективах, вносимых в пользовательскую систему классификации. Это делает выигрышным предложенный фолксномический подход по сравнению с таксономией и фасетизацией, требующими заранее определенной строгой иерархии каталогов и налагающими на пользователя дополнительные расходы на периодически повторяющуюся реструктуризацию архива, плавно возрастающие по мере увеличения числа хранимых документов.

#### ЛИТЕРАТУРА

1. Mathes A. Folksonomies – Cooperative Classification and Communication Through Shared Metadata. / Computer Mediated Communication - LIS590CMC <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
2. Раскин Дж. Интерфейс: новые направления в проектировании компьютерных систем. – СПб.: Символ-Плюс, 2003.
3. Aouiche K., Lemire D., Godin R. Collaborative OLAP with Tag Clouds: Web 2.0 OLAP Formalism and Experimental Evaluation WEBIST 2008. <http://arxiv.org/abs/0710.2156>
4. Kaser O., Lemire D. Tag-Cloud Drawing: Algorithms for Cloud Visualization Tagging and Metadata for Social Information Organization (WWW 2007) <http://arxiv.org/abs/cs/0703109>