

ПРИМЕНЕНИЕ СВЯЗАННЫХ СПИСКОВ ДЛЯ НАВИГАЦИИ В ФOLKCOHOMИЧЕСКИХ СТРУКТУРАХ

В последнее время в интернет-технологиях находят применение системы фолксонимического упорядочения данных - категоризации информации посредством произвольно выбираемых тегов. Этот способ представления обладает рядом преимуществ по сравнению с более традиционными типами классификаций: таксономической (иерархической) и фасетной. Наблюдаемая тенденция является прямым следствием ограничений таксономии (если объект можно привязать только к одному узлу, становится невозможным описать все его необходимые качества) и фасетизации (необходимость существования заранее продуманной и слабо расширяемой системы тегов).

Фолксонимия предоставляет пользователю ряд дополнительных преимуществ: возможность использовать наглядные средства навигации сразу по всему пространству тегов, а также актуализирующийся в реальном масштабе времени набор категорий классификатора. Пользователь оказывается избавлен от входного барьера [1], поскольку для начала работы (классифицирования) не нужно обладать знанием о существующем классификаторе. При этом обратная связь от использования фолксонимии мгновенна: при назначении объекту тега, пользователь сразу может получить информацию о том, что еще помечено тем же тегом, сколько людей пометили этот объект, и как теги пересекаются.

Фолксонимические системы применяются в интернет-ресурсах с обширной базой документов. Пример типичного использования тегов документов - агент чтения новостей Google Reader. Новостные статьи, читаемые в рамках Google Reader, можно «тегировать» - маркировать уникальными ключевыми словами или фразами, чтобы упростить процесс персональной категоризации и группировки входящего контента. Кроме произвольных тегов, контент также можно маркировать как «избранное», в качестве дополнительного тега или маркера для идентификации важных статей. Функция фильтрации позволяет выбирать только те потоки, которые соответствуют введенным ключевым словам (фильтр применяется только к заголовкам фидов).

В Google Reader интегрированы также средства поиска RSS/блогов. Их можно весьма эффективно использовать для нахождения релевантных RSS-источников и фидов на требуемые темы. Как только поиск заканчивается, все релевантные результаты отображаются с опцией их немедленного маркирования тегами и подписки на каждый из них.

Использование тегов документов упрощает навигацию по лентам новостей, позволяя отображать документы, за которыми закреплены определенные ключевые слова. Однако средства, применяемые для обзора пространства тегов (по аналогии с просмотром структуры каталогов файловой системы) весьма ограничены. В более простом варианте пользователю предоставляется одноранговый список тегов, из которого предлагается сделать выбор, чтобы дальше посмотреть относящиеся к тегу документы. Такой подход не позволяет отслеживать структуру сложного графа, который представляет собой развитая фолксонимическая система. Вместе с тем отображение фолксонимических данных в виде графа не получило широкого распространения из-за ряда топологических и эргономических недостатков. Вместо отображения графа тегов в современных интернет-ресурсах применяется отображение их взвешенного списка, известное как облако тегов (tag cloud).

Частота упоминаний ключевых слов отображается в виде изображения соответствующих тегов в формате гиперссылки. Размер изображения тем больше, чем выше релевантность данного слова (термина, имени). 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, а ее одноуровневая организация не предполагает ни развитой навигации, ни отслеживания внутренних связей между тегами [1, 2].

Нами разработана экспериментальная система фолксономической навигации по документам, использующая для визуализации категорий структуру связанных списков.

Навигация по системе связанных списков, по нашему мнению, перспективна для фолксономического представления, изначально была реализована в менеджере рабочего пространства системы NextStep для визуализации файловой системы. В режиме просмотра (browsing) файловый менеджер предоставляет пользователю два средства навигации: Icon Path и View Area. Icon Path представляет траекторию от корня файловой системы к текущему каталогу либо выбранному файлу. View Area отображает содержание файловой системы. Она представляет графическим способом среду выдвинутого на первый план объекта в виде последовательности списков содержимого элементов из Icon Path [3, 4].

В разрабатываемой системе данные о документах и их информационных тегах организованы в базу данных в виде таблицы. Каждая запись в таблице соответствует информации об одном документе, а ее поля заполняются свойствами этого документа. Для хранения информации о выборе пользователем документов и их информационных тегов при навигации по связанным спискам использована структура данных, приведенная на рисунке 1.

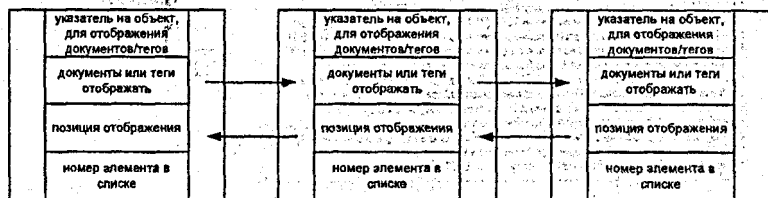


Рис. 1. Структура данных в виде двунаправленного списка

Структура хранит цепочку выбора пользователя [4] и обеспечивает легкое передвижение по ней. Если пользователь возвращается на несколько шагов назад и осуществляет выбор в другом направлении, часть прежней цепочки удаляется и создается новая.

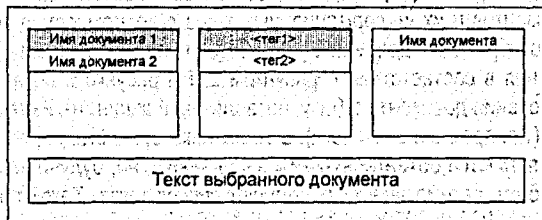


Рис. 2. Навигация по тегам

На рисунке 2 показана цепочка из трех списков документов/тегов; один из которых является активным. По мере продвижения пользователя по базе изменяются отображаемые элементы путем манипуляций с их видимостью и/или координатами расположения на форме. Вследствие того, что тег <тег> указан только для первого объекта, в третьем списке отображается только его название. При выборе тега <тег2> у второго объекта имени первого и второго объектов отображаются в третьем списке, т.к. выбранным тегом промаркированы оба. Аналогичный результат в третьем списке будет получен при выборе тега <тег2> у первого объекта.

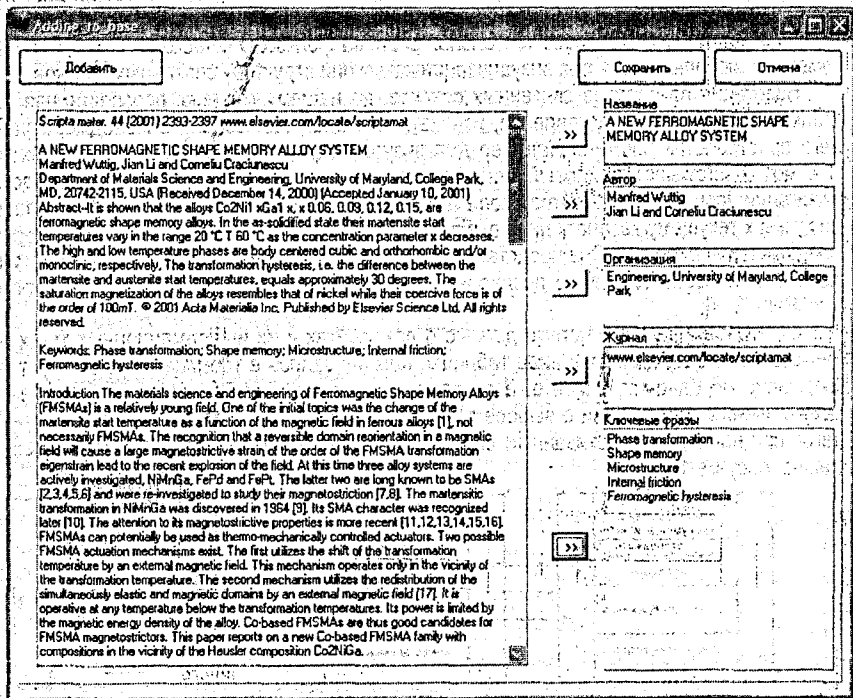


Рис. 3. Добавление документа в базу

Для апробации предложенного интерфейсного решения реализована программная система ведения архива документов с интерфейсом, позволяющим ориентироваться по спискам документов и их информационных тегов с возможностью возврата по цепочке выбора. Для отображения документов и их тегов использована метафора View Area, модифицированная в соответствии с рисунком 2. На рисунке 3 показан диалог первоначального включения документа в базу, позволяющий выделить из текста документа ключевые слова (т.е. произвольные теги), а также некоторые стандартные метаданные. Метаданные не являются обязательными к заполнению, но, будучи введены при внесении документа в базу, автоматически пополняют список тегов. Такой подход к заполнению метаданных может быть отнесен к нечеткой фасетизации и является целесообразным

при работе с документами стандартного типа (в приведенном примере – архивом научных статей), т.к. стимулирует пользователя к изначальной тегированию информации, поиск по которой впоследствии весьма вероятен. Рисунок 4 демонстрирует возможность циклического перемещения по спискам на примере четырех панелей, первая из которых отображает список тегов, вторая – список статей, соответствующих выбранному тегу либо всех статей (если тег не выбран), третья панель отображает теги выбранного документа, сужая тем самым поиск, а четвертая является копией второй, показывая список документов, соответствующих тегу, выбранному в третьей панели.

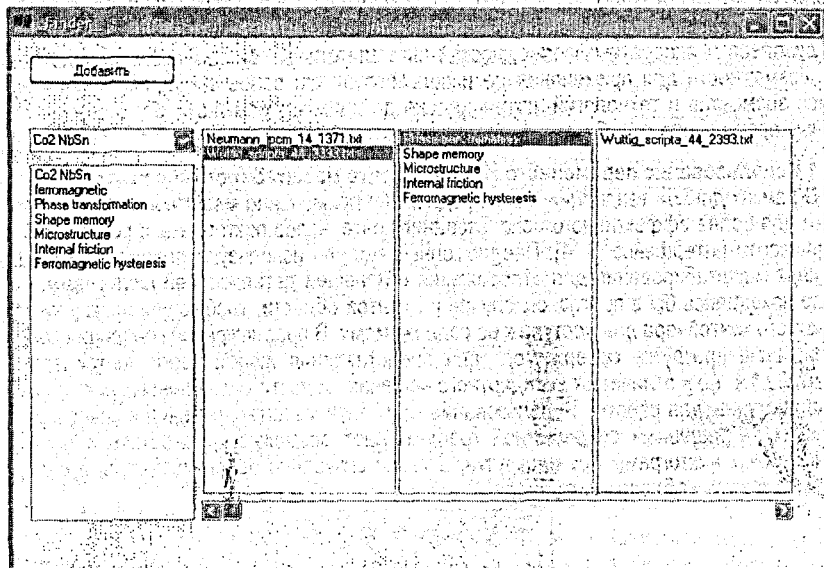


Рис. 4. Реализация системы связанных списков для архива документов

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. 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>
3. Leeuw D. GWorkspace User Guide. <http://qnustep.made-it.com/GWorkspace/GWorkspace.html>. March 29, 2006
4. Раскин Дж. Интерфейс: новые направления в проектировании компьютерных систем. – СПб.: Символ-Плюс, 2003.