

References

1. *Guseva E.N.* Innovation strategy, factor and condition development libraries // Shkol'naja biblioteka. 2012. No 6/7.
2. *Golubeva N.L.* Project activities children's library: an educational aspect // Vestnik BAE. 2016. No 1.
3. *Dyachenko V.K.* Sotrudnichestvo v obuchenii: O kollektivnom sposome uchebnoj raboty [Cooperation in learning: collective mode of learning]. Moscow, 1991. 192 p.

УДК 004.6

Э.В. КУЗЬМИНА, Н.Г. ПЬЯНКОВА, М.Ф. ТИТОРЕНКО

**МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ
ЭЛЕКТРОННЫХ ФОЛЬКЛОРНЫХ РЕСУРСОВ**

Кузьмина Эвелина Вячеславовна, кандидат педагогических наук, доцент кафедры математики и информатики Краснодарского филиала Финансового университета при Правительстве РФ (Краснодар, ш. Нефтяников, 32), lin37@yandex.ru

Пьянкова Нина Геннадьевна, кандидат педагогических наук, доцент кафедры математики и информатики Краснодарского филиала Финансового университета при Правительстве РФ (Краснодар, ш. Нефтяников, 32), ninapyankova@mail.ru

Титоренко Марина Федоровна, доктор исторических наук, завкафедрой философии, истории и права Краснодарского филиала Финансового университета при Правительстве РФ (Краснодар, ш. Нефтяников, 32), MFTitorenko@fa.ru

Аннотация. В статье рассматриваются методы интеллектуальной обработки и визуализации фольклорной информации, представленной в разных машинных форматах. Был проработан функционал для интеллектуализации работы пользователя с фольклорными ресурсами. Предложена матрица соответствия методов трансформации данных типу фольклорного ресурса. Рассмотрены основные средства трансформации данных для оснащения аналитических систем обработки фольклорных ресурсов. **Ключевые слова:** интеллектуальная обработка, фольклорное творчество, электронные фольклорные ресурсы, визуализация, форматы данных.

UDC 004.6

E.V. KUZMINA, N.G. PYANKOVA, M.F. TITORENKO

**INTELLIGENT PROCESSING TECHNIQUES
OF ELECTRONIC RESOURCES OF FOLKLORE**

Kuzmina Evelina Vyacheslavovna, PhD (pedagogical sciences), associate professor of mathematics and computer science department at the Krasnodar branch of Financial university under the Government of the Russian Federation (Krasnodar, Neftyanikov av., 32), lin37@yandex.ru

Pyankova Nina Gennadiyevna, PhD (pedagogical sciences), associate professor of mathematics and computer science department at the Krasnodar branch of Financial university under the Government of the Russian Federation (Krasnodar, Neftyanikov av., 32), ninapyankova@mail.ru

Titorenko Marina Fedorovna, PhD (historical sciences), head of department of philosophy, history and law of the Krasnodar branch of Financial university under the Government of the Russian Federation (Krasnodar, Neftyanikov av., 32), MFTitorenko@fa.ru

Abstract. The article deals with intellectual processing techniques and visualization folkloric information provided in different formats machine. It was elaborated functionality for intellectualization of user experience with folklore resources. We have developed a matrix that presents methods of transformation of folklore

data, depending on the type of electronic resource. We have considered the main means of transformation of data to work for analytical folklore resources processing systems.

Keywords: intellectual processing, folk creativity, folklore, electronic resources, visualization, data formats.

Традиционная культура любого народа утверждает его собственную самодостаточность и исключительность. Особое место в ней занимает народное творчество: фольклор, нравы, традиции, обряды. В них переплетаются отзвуки далеких времен и современные требования к нравственности и духовной культуре молодежи. Через отдельные элементы народного творчества формируются национальные идеалы, моральные и нравственные принципы, регулируются нормы социальных отношений, формируется этническое самосознание. В этой связи отметим, что Кубань богата образцами народного творчества, которые могут стать объединяющим началом в условиях тотальной идеологизации.

На Кубани русско-украинские фольклорные традиции трансформировались в качественно новый результат, имеют своеобразный «кавказский оттенок», отражают в народной памяти исторические события, нравы, нормы поведения, словесные стереотипы социального опыта, эстетические, культурологические и аксиологические проблемы, позволяют вырабатывать ценностные ориентиры по отношению к жизненным обстоятельствам [1, с. 5].

Самобытная народная культура (в том числе казачья), несмотря на свои глубинные корни и ценностные ориентиры, все чаще предается забвению. А провалы в исторической памяти этноса, как известно, обрекают его на духовную незащищенность, утерю своего лица и, в конечном итоге, на исчезновение. Живая фольклорная историческая традиция постепенно исчезает. Рассматриваемая проблема нашла отражение в Международной конвенции «Об охране нематериального культурного наследия» (принята 17 октября 2003 г.).

В настоящее время изучением казачьей культуры (фольклорного и песенного наследия, исторических традиций взаимодействия с различными этносами) занимается Центр традиционной культуры кубанского казачества. За долгие годы работы в экспедициях специалистами Центра собран богатейший полевой материал, который недоступен даже ученым, за исключением 2-3 сотрудников Центра. Условия хранения полевого материала вызывают беспокойство, так как не соответствуют требованиям сохранения архивных материалов. Аналогичная проблема характерна для многих регионов (например, в Москве под угрозой закрытия находится Центр русского фольклора). Данный материал необходимо переводить в электронные форматы для дальнейшей систематизации и интеллектуальной обработки фольклорной информации.

В настоящее время сформирован пласт фольклорных web-ресурсов, отличительной особенностью которых является собирательный подход к представлению фольклорной информации. Сайты фольклорных ресурсов имеют недостаточно развитый набор функционала для интеллектуализации работы пользователя с web-ресурсом. Наличие разной типовидовой фольклорной информации, присутствие аудио- и видеоресурсов на сайте требует принципиально новых подходов к процессам аналитико-синтетической переработки информации и использованию интеллектуальных технологий [2, с. 135].

На сайтах фольклорных ресурсов могут быть применены следующие интеллектуальные технологии: технология инженерии знаний, программные системы извлечения и формализации знаний из неструктурированной и слабо структурированной информации, системы машинного обучения, формы визуализации информации, контента и знаний, программные системы принятия решений и идентификации ситуаций, репозитории открытых данных (linked open data), решение проблем семантической интероперабельности и глобальной идентификации объектов. Данные технологии применимы как к отдельным web-ресурсам, так и к хранилищам данных.

Особенно актуальным вопрос организации хранилищ данных становится при формировании единого фольклорного виртуального пространства на основе специализированного хостинга для учета и анализа базовых духовных ценностей народа [3, с. 40].

Главное преимущество хранилищ данных в интеллектуальных технологиях – это достоверность анализа фольклорных данных, которая реализуется за счет скорости доступа к фольклорным ресурсам, использования фольклорных данных различных типов (текстовая, аудио-, видео-, графическая информация) и форматов компьютерных файлов.

В настоящее время существуют два направления аналитических технологий, которые могут быть применены к семантической обработке фольклорных ресурсов.

Первое направление – это загрузка данных в аналитическое приложение непосредственно из фольклорных источников, где они содержатся, семантическая обработка реализуется средствами самого аналитического приложения. Это могут быть стандартные средства поиска по ключевым словам, индексация, фильтрация, сортировка, запросы к таблицам на основе реляционной алгебры [4, с. 104].

Второе направление – это использование хранилищ данных, которые представляют собой разновидность системы хранения, предназначенной для анализа данных на основе целостности, непротиворечивости, хронологии и высокой скорости обработки запроса.

Рассмотрим преимущества и недостатки загрузки данных в аналитическое приложение непосредственно из фольклорных источников для семантической обработки фольклорной информации.

Технология загрузки данных в аналитическое приложение непосредственно из фольклорных источников имеет определенное преимущество перед технологиями хранилищ данных. Аналитический процесс становится проще и дешевле, не нужен сложный процесс перегрузки исходных данных из различных источников в аналитическое приложение.

Моделирование аналитических приложений должно основываться на разных типах фольклорных ресурсов, отражающих поэзию, песни, народную музыку, танцы. Данные фольклорных ресурсов фиксируются в определенных форматах, для которых существуют различные аналитические приложения. В текстовом формате фиксируется фольклорная эпическая проза, поэзия, тексты фольклорных песен. В видео- и звуковых форматах фиксируются визуальные образы народных традиций и обрядов.

Нами будут рассмотрены основные средства трансформации, которыми могут быть оснащены аналитические системы обработки фольклорных ресурсов. Целесообразно провести ранжирование методов трансформации данных по эффективности обработки фольклорных ресурсов различных типов.

Для фольклорных ресурсов, зафиксированных в текстовых форматах, методы трансформации данных ранжируются по убыванию эффективности применения следующим образом: преобразование упорядоченных данных, кластеризация, сортировка, группировка.

Метод преобразования упорядоченных данных дает возможность оптимизировать данные в текстовом формате и произвести их группировку по временному периоду [5, с. 131]. Для фольклорной текстовой информации временные периоды могут быть определены для двух концептов: дата сбора или фиксации фольклорного текста и дата принадлежности текста предполагаемому хронологическому периоду.

Аналитические задачи в сфере фольклористики, например, функционирование единиц фонетической системы говора в текстах устного народного творчества, фольклорных произведений разных периодов, обновление традиций оказываются чрезвычайно протяженными во времени и подчиняются закономерностям обработки данных, зависящих от времени (временные ряды).

В процессе обработки временных рядов решаются определенные аналитические задачи: классификация состояния фольклорных объектов и выявления закономерностей их развития, объясняющих динамику фольклорных процессов для любых временных интервалов (даты и времени). Временные фольклорные ряды могут быть одномерными и многомерными [6, с. 34].

Одномерный временной ряд должен содержать один признак фольклорного объекта, отражающий динамику во времени.

Многомерный ряд может содержать два и более признака фольклорного объекта. Значение временного ряда определяется только в фиксированные моменты времени (отсчеты), равностоящие друг от друга.

Выделим два типа трансформации временных рядов: скользящее окно и преобразование даты и времени. Скользящее окно применяется для классификации фольклорных объектов, чтобы преобразовать значения временного ряда в таблицу для реляционной обработки. Преобразование даты и времени сводится к приведению их к виду, пригодному для визуального анализа.

Кластеризация фольклорных текстов предполагает выявление групп, имеющих одинаковое семантическое ядро среди заданного множества фольклорных документов. Следует отметить, что группы формируются на основе попарного сравнения описания документов, причем никакие характеристики этих групп не задаются заранее. Основная проблема кластеризации фольклорных документов заключается в таком разнесении по группам, при котором элементы каждой группы были бы достаточно сходны с другими, чтобы в некоторых случаях можно было бы пренебречь индивидуальными различиями между ними. Поиск информации в систематизированном фольклорном массиве документов будет намного проще, нежели в несистематизированном, так как документы, для которых профили не имеют сходства с предписанием поиска, не будут включаться в углубленный поисковый процесс.

Задачи классификации фольклорных текстовых документов сводятся к следующему: формирование классов на основе характеристик фольклорных объектов; отнесение объектов, выделенных из текстовых фольклорных документов, к сформированным классам; определение атрибутов для описания классов текстовых фольклорных документов [7, с. 54].

Формирование классов фольклорных документов должно производиться на основе сопоставления фольклорных объектов, включенных в фольклорные документы. Данные объекты в теории документального поиска принято называть ключевыми словами. Однако современный подход к интеллектуализации обработки информации требует их перевода в плоскость реляционной концепции, которая должна быть основой интеллектуальной обработки хранилищ фольклорных данных.

Формирование кластеров фольклорных документов, относящихся к определенному классу (фольклорная проза, поэзия, тексты песен и др.), должно выполняться с учетом интересов потребителей и опыта, накопленного в ходе предшествующих поисков.

Сортировка для фольклорной информации позволяет представить порядок следования записей в порядке, определенном пользователем. Для фольклорных текстов этот метод применим для ранжирования ключевых слов и выводов текстов в соответствии с частотой встречаемости этих ключевых слов. Для усиления интеллектуальных функций систем обработки фольклорных данных необходимо сортировку основывать на одном из трех методов: формирование частотной модели; модели, учитывающей различительную силу термина; модели, основанной на динамической оценке информативности.

Частотная модель учитывает те термины, которые имеют самую высокую частоту встречаемости в отдельных фольклорных документах. В то же время суммарная частота их встречаемости в массиве фольклорных ресурсов невелика.

Другой подход – учет различительной силы термина, учитывает те фольклорные термины, которые позволяют различать фольклорные тексты. При этом ценен именно тот термин, присутствие которого делает фольклорный документ непохожим на все другие документы.

Модель, основанная на информативности, должна использовать оценку релевантности, полученную от пользователя в ходе его работы с системой интеллектуальной обработки фольклорных текстов.

Метод группировки также может быть использован для анализа фольклорной информации, так как она всегда разобщена, разбавлена посторонними данными, разбросана по отдельным фольклорным текстам. Группировка позволяет объединить фольклорную инфор-

мацию в минимальное количество фольклорных объектов и их атрибутов. Иногда высокая степень детализации атрибутов может ухудшить результаты анализа фольклорного объекта. Детальные характеристики фольклорного процесса, например народного праздника, подвержены колебаниям под действием случайных факторов, поэтому характеристики, взятые по отдельности за небольшой по протяженности временной период, не дадут объективной оценки развития фольклорного процесса. Для получения более обоснованной оценки этого процесса необходимо сгруппировать показатели выделенных временных интервалов. Анализ признаков фольклорных объектов также должен быть основан на группировке, чтобы исключить случайные вариации значений признаков исследуемых объектов.

Для фольклорных ресурсов, зафиксированных в видеоформатах, методы трансформации данных ранжируются по убыванию эффективности применения следующим образом: преобразование упорядоченных данных, кластеризация, слияние, группировка, настройка набора данных.

Технология преобразования упорядоченных данных для звуковых и видеоформатов представления фольклорной информации совпадает с технологией обработки текстовой информации, приведенной выше.

Метод кластеризации для видео- и звуковой фольклорной информации требует дополнительных приложений в виде анализаторов видео- и звукового ряда.

Технология слияния должна рассматриваться в двух аспектах: анализ звукового ряда без семантики и анализ звукового ряда с семантическим содержанием. Слияние позволяет объединить два фольклорных объекта по одноименным характеристикам или дополнить один объект характеристиками другого объекта, которые отсутствуют в дополняемом. Слияние применяется для решения задач, когда выбранную информацию об одном фольклорном объекте необходимо дополнить выборочными данными о другом фольклорном объекте. В методе слияния можно применить две операции: объединение и дополнение. При объединении к характеристикам исходного фольклорного объекта добавляются все характеристики другого объекта. При дополнении к исходной выборке добавляются только те характеристики, которые отсутствовали в исходной. Метод слияния фольклорных данных является главным способом обогащения данных.

Метод группировки применительно к видео- и звуковой фольклорной информации так же, как и метод кластеризации, требует дополнительных программных приложений в виде звуковых и видеоанализаторов. Для видео- и звуковой информации метод группировки не является приоритетным, так как эти виды информации содержат большое количество фольклорных объектов и требуют их дробления для дальнейшего анализа. Например, данный метод может быть использован, когда необходимо определить соответствие песенного звукоряда фольклорной прозе или поэзии определенного региона в данный хронологический период.

Настройка набора данных позволяет изменять имена, типы, метки и назначение атрибутов фольклорных объектов в полученной выборке фольклорных данных для их дальнейшей обработки математическими методами [8, с. 89].

Можно предложить матрицу выбора технологии интеллектуализации для определенных типов данных, отражающих фольклорное творчество (см. таблицу).

Важнейшей составляющей аналитического процесса является визуализация, позволяющая представить фольклорные данные в виде, который обеспечивает эффективную работу пользователя. Способ визуализации должен максимально представить изменение данных о фольклорном объекте, содержащуюся в нем информацию, закономерности. Способ визуализации должен определяться в зависимости от типа фольклорного ресурса (текстовая, видео-, звуковая).

Для анализа фольклорных ресурсов могут быть использованы следующие средства визуализации: графики, диаграммы, таблицы, OLAP-анализ, матрицы классификации, визуализация контроля процесса обучения модели, древовидные визуализаторы, визуализация связей и карты.

Таблица

Матрица соответствия методов трансформации данных типу фольклорного ресурса

Метод	Типы ресурсов		
	Аудиоинформация	Видеоинформация	Текстовая информация
Преобразование упорядоченных данных	да	да	да
Кластеризация	да	да	да
Сортировка	нет	нет	да
Группировка	нет	нет	да
Слияние	да	да	нет
Настройка набора данных	да	да	нет

Важнейшей составляющей аналитического процесса является визуализация, позволяющая представить фольклорные данные в виде, который обеспечивает эффективную работу пользователя. Способ визуализации должен максимально представить изменение данных о фольклорном объекте, содержащуюся в нем информацию, закономерности. Способ визуализации должен определяться в зависимости от типа фольклорного ресурса (текстовая, видео-, звуковая).

Для анализа фольклорных ресурсов могут быть использованы следующие средства визуализации: графики, диаграммы, таблицы, OLAP-анализ, матрицы классификации, визуализация контроля процесса обучения модели, древовидные визуализаторы, визуализация связей и карты.

Графики позволяют визуально представить общий характер данных, закономерности, тенденции и периодические изменения. Тем самым можно выявить недостающие факты, отсутствующие в фольклорных данных. Графические методы более эффективно сопоставляют данные. Достаточно построить графики двух исследуемых фольклорных процессов в одной системе координат, чтобы оценить степень их сходства и различия. Например, если нужно сравнить количество фольклорных объектов в определенные исторические периоды, то можно воспользоваться графиком.

Если нужно сравнить динамику появления в фольклорных источниках нескольких фольклорных объектов, то могут быть использованы диаграммы: круговые или лепестковые. Круговую диаграмму можно использовать, если нужно показать долю, которую вносит то или иное значение в общую сумму. Например, частоту встречаемости фольклорных единиц фонетической системы говора в текстах устного народного творчества. Лепестковую диаграмму можно использовать в случае, если необходимо представить абсолютные значения развития нескольких фольклорных объектов за разные исторические периоды.

OLAP-анализ может быть использован при визуализации реальных фольклорных процессов, в которых присутствует большое количество фольклорных объектов, находящихся в разнообразных отношениях, с каждым из которых связаны несколько числовых характеристик. В основе OLAP-анализа лежит многомерное представление данных, которые разделены на количественные и качественные характеристики. Например, если исследуемым процессом являются фольклорные праздники, то выделяются такие фольклорные объекты, как традиции, фольклорные коллективы, регионы, в которых проходили праздники, даты, место проведения и т.д.

Матрица классификации может применяться для оценки качества классификационных моделей фольклорных ресурсов. В данном случае классификационная модель в качестве входных данных имеет признаки или атрибуты классифицируемых фольклорных объектов,

а на выходе должна формироваться метка класса, которому будет принадлежать фольклорный объект, имеющий данные признаки.

Визуализация контроля обучения модели основывается на машинном обучении с использованием нейронных сетей, деревьев решений [9, с. 22]. В процессе обучения модель получает свойства, необходимые для преобразования данных. Так как фольклорные ресурсы относятся к плохо формализуемым источникам информации, то разработка модели интеллектуального анализа является приоритетной для трансформации фольклорных данных на реляционную платформу.

Древовидные визуализаторы могут быть использованы в случае, если результатом аналитической обработки фольклорных данных являются иерархические уровни, которые разделены на предков и потомков. Для слабоструктурируемой фольклорной информации могут быть использованы следующие виды применения деревьев: деревья решений и деревья ассоциативных правил. Деревья решений могут применяться для представления любой иерархической последовательности правил вида «если... то...». Например, если в информации, зафиксированной в файлах звуковых форматов, выделяется определенное количество заданных гласных звуков, то фольклорное произведение можно отнести к определенному региону.

Визуализация связей используется, когда необходимо определить степень взаимной зависимости между различными фольклорными объектами. Фольклорные объекты должны сравниваться между собой по какому-либо признаку. Например, можно установить диаграммы связей между фольклорной активностью регионов России. В качестве признаков могут быть использованы численность постоянного населения, процент сельского населения и зафиксированное количество фольклорных источников за определенный период.

Карты являются наиболее предпочтительным способом представления информации о фольклорных ресурсах по географическим регионам. Например, распределение показателей фольклорной прозы или поэзии по различным регионам, данные, отражающие миграцию фольклорных произведений и динамику их распространения.

Рассмотренные возможности методов интеллектуальной обработки электронных фольклорных ресурсов позволили выявить зависимости между типом электронного ресурса и способом их обработки. Предпочтительными способами обработки фольклорной информации, зафиксированной в различных форматах, являются: преобразование упорядоченных данных и кластеризация.

Визуализация является логическим продолжением интеллектуального анализа фольклорных данных и должна быть использована при создании интеллектуальных систем.

Предложенный подход к интеллектуализации обработки фольклорной информации, зафиксированной в электронных носителях, позволяет говорить о возможности использования реляционного подхода при создании фольклорных интеллектуальных систем. В перспективе реляционный подход позволит обеспечить совместимость фольклорных интеллектуальных систем с хранилищами данных и технологиями извлечения знаний в базах данных Knowledge и DataMining.

Литература

1. *Титоренко М.Ф.* Социальная организация казачества верхнего Прикубанья: автореф. дис. ... д-ра ист. наук // Российская академия государственной службы при Президенте Российской Федерации. М., 2005.
2. *Кузьмина Э.В., Пьянкова Н.Г., Салий В.В.* Подходы к интеллектуализации web-ресурсов фольклорного творчества // *Культурная жизнь Юга России.* 2016. № 2(61). С. 133-137.
3. *Кузьмина Э.В., Пьянкова Н.Г.* Использование технологий хранилищ данных для интеллектуализации фольклорных информационных ресурсов // *Информационные ресурсы России.* 2016. № 6. С. 39-41.
4. *Салий В.В., Кузьмина Э.В.* Применение формализованных методов аналитико-синтетической переработки информации в библиотечно-библиографической деятельности // *Культурная жизнь Юга России.* Приложение. 2015. № 1 (1). С. 102-105.

5. Пьянкова Н.Г. Проблемы информатизации юридической деятельности // Гуманитарные знания и естественные науки: современные проблемы и перспективы развития: Материалы II Межвузовской научно-практической конференции; под общ. редакцией Е.В. Королюк. 2014. С. 130-132.

6. Третьякова Н.В. К вопросу о моделировании ситуаций и принятию управленческих решений // Новая наука: современное состояние и пути развития. 2016. № 63 (86). С. 32-36.

7. Салий В.В., Шапошников В.Л., Пьянкова Н.Г., Кузьмина Э.В. Методы анализа больших объемов слабоструктурируемой информации. Краснодар, 2017.

8. Сидоренко В.С., Кузьмина Э.В. Возможности адаптации типовой модели подготовки информатиков для социально-культурной сферы // Культурная жизнь Юга России. 2011. № 42. С. 88-89.

9. Николаева И.В. Применение искусственных нейронных сетей для прогнозирования динамики экономических показателей // Сфера услуг: инновации и качество. 2012. № 8. С. 22.

References

1. Titorenko M.F. The social organization of the Kuban Cossacks top: thesis ... PhD (history): 07.00.03. Moscow, 2005.

2. Kuz'mina Je.V., Salij V.V. Approaches to intellectualization folklore of web-resources // Kul'turnaja zhizn' Juga Rossii. 2016. № 2(61).

3. Kuz'mina Je.V., P'jankova N.G. Using the technology of data warehouses for the intellectualization of folklore Information Resources // Informacionnye resursy Rossii. 2016. № 6.

4. Salij V.V., Kuz'mina Je.V. The use of formal methods of analytical and synthetic processing of information in the library and bibliographic activity // Kul'turnaja zhizn' Juga Rossii. 2015. № 1 (1).

5. P'jankova N.G. Informatization Problems of legal activity // Humanities and natural sciences: current problems and prospects of development, 2014.

6. Tret'jakova N.V. The modeling of situations and management decisions // Novaja nauka: sovremennoe sostojanie i puti razvitija. 2016. № 6-3 (86).

7. Salij V.V., Shaposhnikov V.L., P'jankova N.G., Kuz'mina E.V. Metody analiza bol'shih objemov slabostrukturirujemoj informacii [Methods of analysis of large amounts of loosely structured information]. Krasnodar, 2017.

8. Sidorenko V.S., Kuz'mina Je.V. Possibilities of adaptation typical informatics training model for the social and cultural sphere // Kul'turnaja zhizn' Juga Rossii. 2011. № 42.

9. Nikolaeva I.V. Application of artificial neural networks to predict the dynamics of economic indicators // Sfera uslug: innovacii i kachestvo. 2012. № 8.

