

# Электронная библиотека "Кинолетопись России"

Ю.А. Бухштаб, Н.Н.Евтеева  
Институт прикладной математики им. М.В.Келдыша РАН Москва,

Россия

## Резюме

В работе описываются основные направления работ, связанных с созданием информационной базы и разработкой программных средств в рамках проекта электронной библиотеки "Кинолетопись России".

*Ключевые слова: электронная библиотека, интермедиа, поиск изображений.*

## 1. ОБЩИЕ ПРИНЦИПЫ

На протяжении последнего года группа разработчиков в составе объединенного коллектива сотрудников Института прикладной математики им. М.В.Келдыша РАН и сотрудников Российского государственного архива кинофотодокументов (РГАКФД), при участии американской кино-телевизионной компании "Абамедиа", осуществляют комплекс работ, целью которых является создание, на базе коллекции документальных фильмов РГАКФД, электронной библиотеки - "Кинолетопись России".

В РГАКФД хранится самое крупное собрание кино фото документе в по истории России и других стран СНГ. В его фондах насчитывается около 190 000 единиц хранения кинодокументов (коробок с пленкой). Эта коллекция, представляющая собой иллюстрированную историю России начиная с середины прошлого века, представляет огромный интерес не только для средств массовой информации, но и для информационного обслуживания фундаментальных научных исследований, прежде всего исторических. Кроме того, коллекция архива является практически неисчерпаемым источником иллюстративного материала для образовательных целей.

Информационный фонд электронной библиотеки "Кинолетопись России" представляет собой совокупность документов, содержащих описания наиболее интересных фильмов из фондов РГАКФД. Каждый документ состоит из текстов, последовательно описывающих планы видеоряда, сопровождаемые тайм-кодами и соответствующие описываемым планам стоп-кадры или видеофрагменты, которые, если это необходимо, сопровождаются звуком. Кроме того, описание каждого фильма должно

сопровождаться дополнительными сведениями, такими как режиссер фильма, оператор, дата выпуска, краткая аннотация и т.д. Таким образом, ознакомившись с описанием фильма, содержащемся в электронной библиотеке "Кинолетопись России", пользователь получает достаточно полную информацию как о содержании фильма, так и о его изобразительном ряде.

К настоящему моменту был создан и опробован пилотный фрагмент библиотеки, содержащий описания всех кинодокументов за 1938 год.

## 2. БАЗОВОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ.

В рамках осуществления описываемого проекта ведутся работы по разработке инструментальных программных средств для создания, ведения и использования электронной библиотеки "Кинолетопись России".

Прежде всего, была разработана технология заполнения информационного фонда рассматриваемой электронной библиотеки и реализовано соответствующее программное обеспечение. Такой подход позволил начать цикл трудоемких работ по фактическому заполнению фонда электронной библиотеки в максимально короткие сроки.

Также было разработано ядро, реализующее механизм полнотекстового поиска, базирующийся на полностью инвертированных файлах. Этот механизм поддерживает запросы, использующие маскируемые ключи, их комбинации, меры близости, а также обеспечивает возможность задавать зону поиска.

При его проектировании большое внимание уделялось минимизации пространства, требуемого для индексных файлов. При этом организуется специальный индекс словаря слов. Элементы словаря, являющиеся синонимами, связаны между собой, поддерживаются родо-видовые связи. Также обеспечивается перестройка индексной структуры (с учетом иерархии информационного фонда) при обновлении, удалении или добавлении документов.

Кроме механизма полнотекстового поиска, разрабатываются программные компоненты, поддерживающие возможность использования гипертекстовой (точнее гипермедийной) технологии. Применение этой технологии позволяет организовывать связи между различными фрагментами информации, благодаря которым пользователь может легко находить и изучать взаимосвязанные материалы.

Центральной компонентой гипертекстовых программных средств является интерпретатор, реализующий модульный интерфейс гнезд управления, которые могут произвольно встраиваться пользователем в содержательный материал, как текстовый, так и графический. При этом обеспечивается возможность связать активное гнездо не только с другим фрагментом, но и с любой прикладной задачей. В этом случае при выходе из гнезда инициализируется запуск этой задачи, и обеспечивается возможность возврата из нее для продолжения текущего сеанса работы.

Предопределенная на стадии проектирования библиотеки структура гиперсвязей может динамически перестраиваться, в зависимости от результатов логической обработки различных событий. Важно отметить, что такой механизм позволяет отслеживать связи между различными фрагментами документов, обеспечивая автоматическую модификацию одних фрагментов в зависимости от изменений в другом.

### **3. ПОИСК МУЛЬТИМЕДИЙНЫХ ДАННЫХ.**

Важной частью проводимых работ являются исследования в области разработки систем поиска видеоинформации. В качестве первого шага предполагается разработать и опробовать алгоритмы автоматического выделения информативных стоп-кадров. Эти алгоритмы могут основываться на анализе сравнительно простых вычислимых свойств видеоряда, таких как количество однородных кадров, смена планов и т.д. Следует отметить, что возможность выделения стоп-кадров с помощью компьютера позволит автоматизировать работу квалифицированного персонала, требующую значительных затрат времени. Кроме того, в процессе анализа видеоряда могут быть определены наборы примитивов, в терминах которых описываются поисковые образы соответствующих видеофрагментов. Эти поисковые образы могут дополнять текстовые описания, помогая осуществлять более точный поиск.

В рамках настоящего проекта предполагается использовать

методы поиска в неоднородной многоуровневой базе знаний для организации взаимодействия модулей, функциональным назначением которых является анализ различных примитивов и мета-данных. Этот механизм может использоваться при формировании как запроса так и поисковых образов. В результате содержание изображения (стоп-кадра) может быть описано с помощью подходящих примитивов или даже на более высоком уровне абстракции, например, на базе комбинации различных примитивов.

При этом должна быть обеспечена возможность настройки монитора, под управлением которого осуществляется интерпретация визуальной информации, на тематическое подпространство, в рамках которого может быть обеспечена информационная поддержка при решении конкретной задачи. При настройке на тематическое подпространство (фиксированное для данного приложения или задаваемое пользователем, а в более сложных случаях определяемое специализированной экспертной системой) устанавливается среда, ассоциированная с данным подпространством. Для каждого такого тематического слоя будет определяться управляющий фрейм, содержащий информацию о соответствующих примитивах и доступных процедурах их обработки. Таким образом, настройка монитора на специфику новой предметной области будет заключаться в дополнении базы управляющих фреймов. Разумеется, должен быть разработан соответствующий языковой интерфейс, предусматривающий возможность задания операций создания и модификации этой базы.

### **Авторы:**

Ю.А. Бухштаб, к.ф.-м.н., зав. сектором Института прикладной математики им. М.В.Келдыша. Адрес: Москва 125315, ул. Усиевича д. 16 кв.40 E-mail: kikom@glasnet.ru

Н.Н. Евтеева, к.ф.-м.н., с.н.с. Института прикладной математики им.М.В.Келдыша.  
Адрес: Москва 121471, Гвардейская ул. д.6 кв.112  
Digital Library "Russian Film Chronicle\*" Yuri A.

Bukhshtab, Natalia N. Evteeva

The main approaches to information fund creation and software development for the digital library, based on the collection of film documents of Russian State Film and Photo Archive, are described in this paper.