

**Комплекс программных систем для обработки,
хранения и выдачи электронных документов
в Российской государственной библиотеке**

**Software Complex for Processing, Storage
and Circulation of Electronic Documents
at Russian State Library**

**Комплекс програмних систем для обробки,
зберігання й видачі електронних документів
у Російській державній бібліотеці**

А. И. Вислый

Российская государственная библиотека, Москва, Россия

О. Н. Шорин

ООО «Шофт», Москва, Россия

Alexander Visly

Russian State Library, Moscow, Russia

Oleg Shorin

«Shoft» Joint-stock Company, Moscow, Russia

О. І. Віслий

Російська державна бібліотека, Москва, Росія

О. М. Шорін

ТОВ «Шофт», Москва, Росія

Являясь второй по величине библиотекой мира, РГБ хранит и предоставляет доступ к огромному количеству электронных материалов. В контексте данной деятельности необходимо решать ряд проблем. Во-первых, электронные экземпляры отсканированных документов необходимо надёжно хранить. Во-вторых, необходимо обеспечить читателям безопасный доступ к ним. И, в-третьих, необходимо автоматизировать все технологические процессы по обработке поступающих в библиотеку документов. Компания «Шофт» разработала и внедрила в РГБ комплекс программных средств, позволяющий решить перечисленные проблемы. В докладе будут рассмотрена часть этих программных средств, каждое из которых может использоваться самостоятельно.

Russian State Library, the second world largest library, provides access to and stores the immense amount of electronic materials. Within this context, the Library has to solve a range of problems. Firstly, the electronic copies of scanned originals have to be reliably stored. Secondly, patrons are to be provided with safe access to them. Thirdly, each technological process of handling acquired documents is to be automated. «Shoft» Company has designed and implemented the software complex to solve these problems. The paper describes several of the software tools with the possibility of independent implementation of each of them.

Будучи другою по величині бібліотекою світу, РДБ зберігає й надає доступ до величезної кількості електронних матеріалів. У контексті цієї діяльності необхідно вирішувати низку проблем. По-перше, електронні екземпляри відсканованих документів необхідно надійно зберігати. По-друге, необхідно забезпечити читачам безпечний доступ до них. І, по-третє, необхідно автоматизувати усі технологічні процеси по обробці документів, що надходять до бібліотеки. Компанія «Шофт» розробила й впровадила в РДБ комплекс програмних засобів, що дозволяють розв'язати перераховані проблеми. У доповіді буде розглянуто частину цих програмних засобів, кожен з яких може використовуватися самостійно.

1. Введение

Являясь второй по величине библиотекой мира, Российская государственная библиотека хранит и предоставляет доступ к огромному количеству электронных материалов, а также обладает разветвлённой структурой, отдельные части которой географически распределены. Следуя послед

ним тенденциям в библиотечном деле, РГБ осуществляет обработку книг, диссертаций и других документов, а также предоставляет доступ к ним в своих читальных залах и через интернет. Эта деятельность выявила ряд проблем. Во-первых, электронные экземпляры отсканированных документов необходимо надёжно хранить. Во-вторых, необходимо обеспечить читателям безопасный доступ к этим материалам. И, в-третьих, необходимо автоматизировать все технологические процессы по обработке поступающих в библиотеку документов, независимо от их формы представления. Компания «Шофт» разработала и внедрила в РГБ комплекс программных средств, позволяющий решить описанные выше проблемы.

Комплекс состоит из пяти систем:

- система защищённого просмотра электронных документов DefView;
- комплекс программных средств MARCTools для работы с библиографическими записями;
- система хранения библиографических записей Метакс;
- АИС «Электронный депозитарий» ;
- АИС «Контроль технологических процессов обработки документов».

В настоящей работе будут подробно рассмотрены некоторые из этих систем.

2. Система защищенного просмотра электронных документов DefView

Вступившая в силу с 1 января 2008 года четвёртая глава Гражданского кодекса РФ ограничивает предоставление библиотеками экземпляров произведений в цифровой форме, защищённых авторских правом, во временное безвозмездное пользование. Согласно новому закону теперь это возможно только при условии исключения возможности создания копий этих произведений в цифровой форме. Использование программных средств защищённого просмотра документов (система DefView), разработанных компанией «Шофт», позволяет библиотекам в полной мере обеспечить выполнение требований ГК РФ.

На данный момент система DefView предназначена для обеспечения безопасного постраничного доступа к документам в формате Portable Document Format (PDF).

Основные свойства системы DefView:

- имеет трехзвенную архитектуру: клиент (DefView), сервер приложений (DefView Server), сервер хранения документов (PDF Server);
- позволяет постранично просматривать документ на стороне клиента;
- предотвращает несанкционированное копирование страниц документа, запрещая на уровне приложения снимок экрана и не кэшируя их в постоянной памяти компьютера пользователя;
- предоставляет набор операций над страницами (просмотр, печать, отложенная печать, загрузка) исходя из прав пользователя на документ;
- предоставляет функции предпросмотра, масштабирования, поворота и инвертирования цветов страниц документа, а также быстрой навигации по ним
- обеспечивает возможность просмотра документа в полноэкранном режиме;
- позволяет осуществлять поиск слов и фраз внутри документа с учетом морфологии русского языка и затем выделять найденные результаты на страницах документа.

Трехзвенная архитектура обладает рядом преимуществ. Во-первых, она позволяет обеспечить большую безопасность и сохранность данных за счёт того, что доступ к ним осуществляется только через сервер хранения, с которым пользователи напрямую не взаимодействуют. Во-вторых, обеспечивает расширяемость системы, позволяя разрабатывать серверы хранения, работающие с произведениями в других форматах, и подключать их к единому серверу приложений.

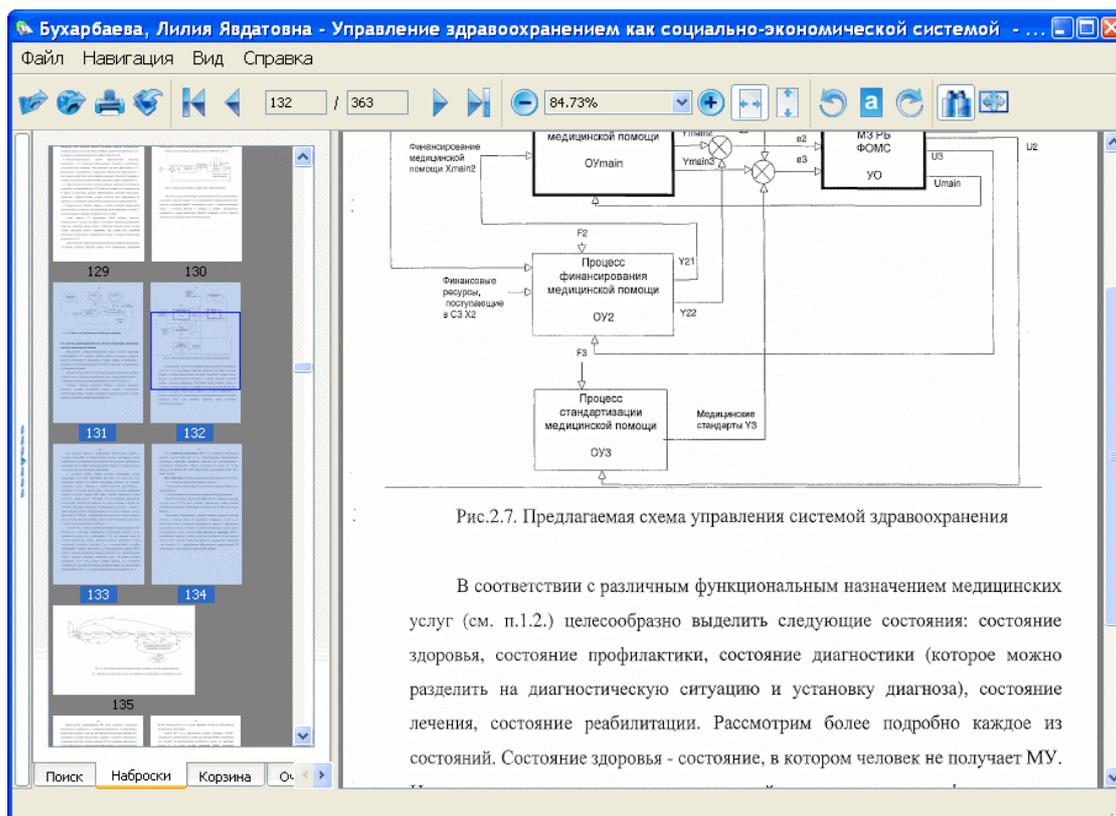


Рис. 1. Главное окно клиентского приложения DefView

Главное окно клиентского приложения DefView состоит из меню, панели инструментов, панели поиска, области отображения страниц документа и строки состояния. Приложение имеет интуитивно понятный пользовательский интерфейс.

На панели инструментов расположены кнопки для выполнения различных функций программы, которые разделены по функциональности на пять блоков: операции с документами, навигация по документу, масштабирование документа, вид документа и вид окна программы. Все кнопки имеют всплывающие подсказки, в которых отображается, какое действие будет выполнено при нажатии на ту или иную кнопку. Некоторые кнопки имеют соответствующие сочетания клавиш.

С использованием кнопок панели инструментов пользователь может быстро перейти на первую, предыдущую, следующую, последнюю и произвольную страницу документа. Также он может выбрать один из фиксированных масштабов отображения документа или установить масштаб «по ширине страницы» или «страница целиком». Пользователь может повернуть страницу, а также инвертировать цвета.

В главном меню дублируются все кнопки, представленные на панели инструментов, а также имеются дополнительные. На панели поиска и области отображения страниц документа доступно контекстное меню.

На рисунке 1 приведено главное окно клиентского приложения с активированной областью отображения набросков. С помощью набросков пользователь получает возможность предпросмотра страниц документа и быстрого перехода с одной страницы на другую. Пользователь может выделить несколько набросков для осуществления групповых операций над страницами. Также в набросках графически выделена область текущей страницы, которая показана в области отображения страниц документа.

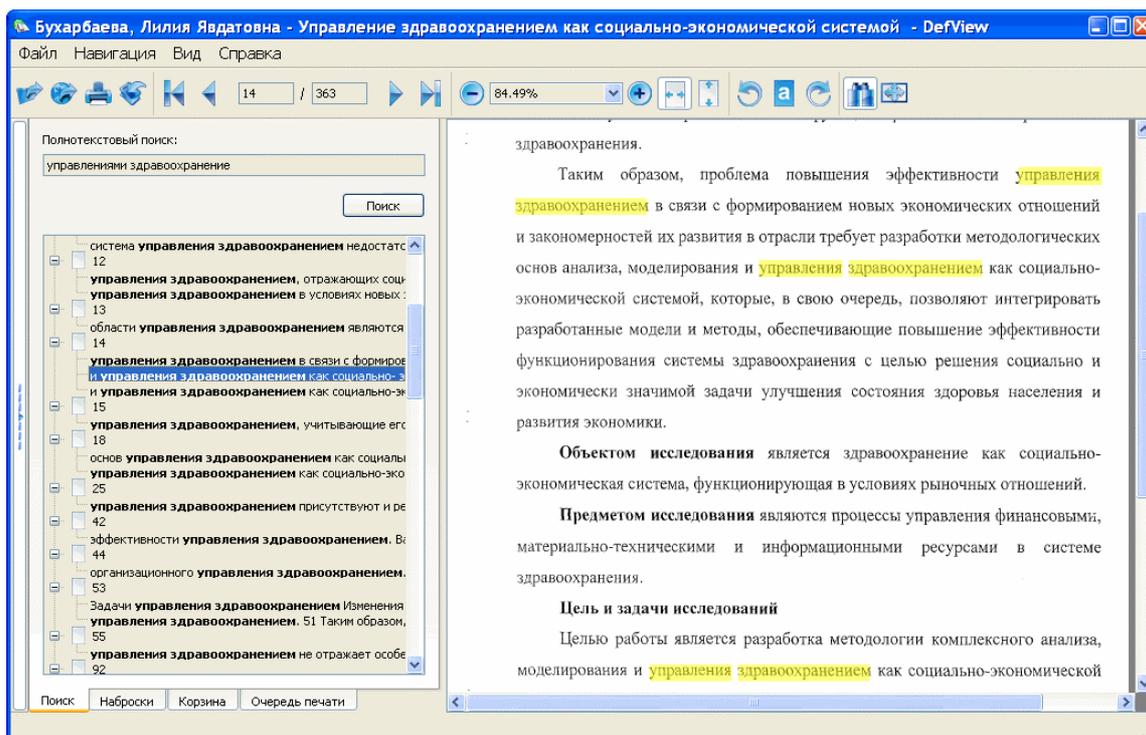


Рис. 2. Поиск фраз в документе

На рисунке 2 приведено главное окно клиентского приложения с открытой панелью поиска. Клиентское приложение позволяет осуществлять поиск слов и фраз внутри документа. Поиск осуществляется с учетом морфологии русского языка. В качестве результатов поиска выдаются сниппеты, в которых жирным шрифтом выделена найденная фраза. При клике мышкой на сниппете происходит переход на соответствующую страницу. На самой странице найденный текст подсвечивается желтым цветом.

С помощью клиентского приложения пользователь может сформировать заказ на печать выбранных им страниц. Администратор виртуального читального зала имеет возможность распечатать отобранные пользователем страницы.

3. Программный комплекс для решения задачи поиска и устранения дуплетных библиографических записей CompareMARC

В процессе обработки и хранения большого количества материалов возникает задача поиска и устранения дуплетных библиографических записей. Для решения этой задачи компанией «Шофт» был разработан программный комплекс CompareMARC, который позволяет отобразить библиографические записи, с большой долей вероятности являющиеся дуплетами, в графическом режиме сравнить их друг с другом с возможностью создания новых записей на основе существующих.

Программный комплекс CompareMARC состоит из двух программ. Первая программа предназначена для формирования массива потенциально дуплетных библиографических записей. Эта программа позволяет библиотекарю реализовать собственную функцию, на основании которой принимается решение о потенциальной дуплетности записей. С помощью реализованной функции множество библиографических записей разбивается на подмножества, в каждом из которых сгруппированы потенциально дуплетные записи.

Сформированные подмножества передаются на вход утилите, которая в графическом режиме позволяет сравнить библиографические записи.

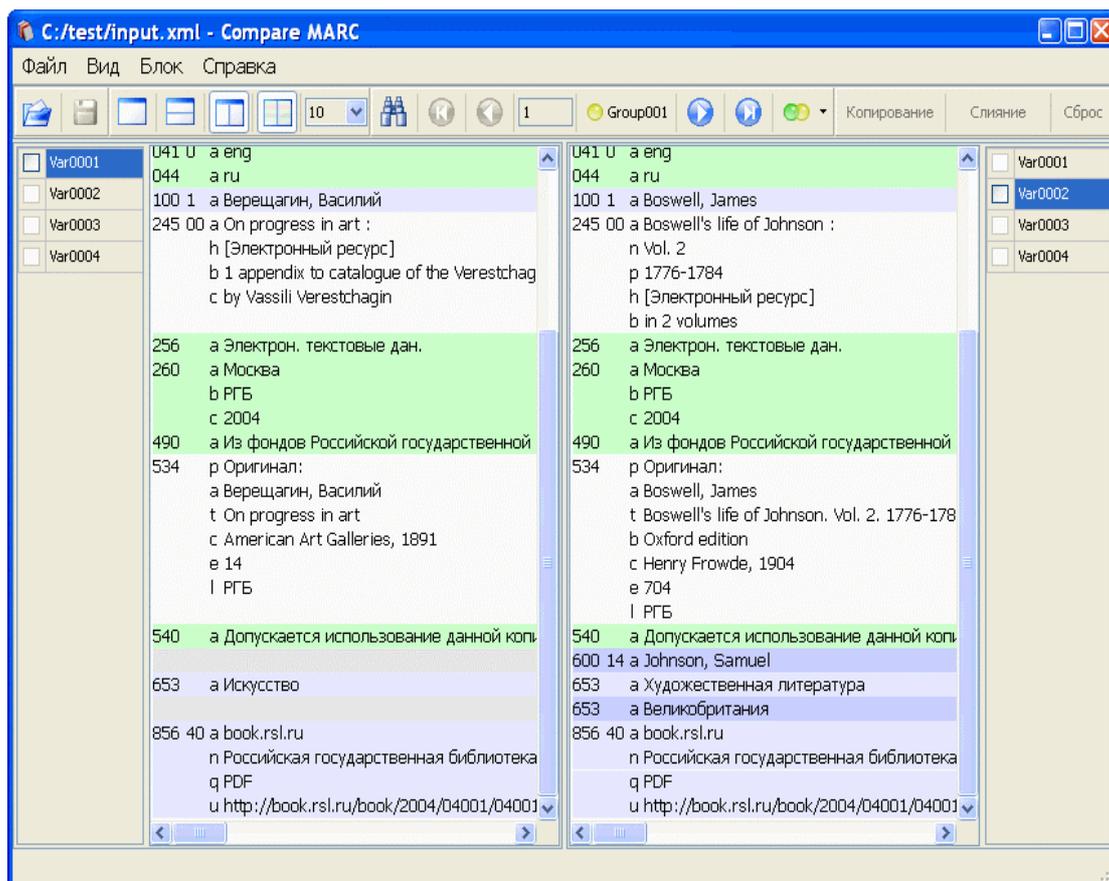


Рис. 3. Попарное сравнение потенциально дуплетных записей

На рисунке 3 приведено главное окно графической утилиты в режиме попарного сравнения потенциально дуплетных библиографических записей. В режиме попарного сравнения различными цветами отображаются поля записи, которые совпадают, которые не совпадают и которые отсутствуют в одной из записей.

Библиотекарь из подмножества записей может выбрать одну или несколько записей, которые будут целиком скопированы в итоговый файл в качестве уникальных записей, не имеющих дуплетов. В случае, если ни одна из записей из подмножества не является полным описанием документа, библиотекарь может перейти в режим слияния записей, в котором ему предлагается создать новую уникальную запись на основе имеющихся. На рисунке 4 приведено окно утилиты в режиме слияния.

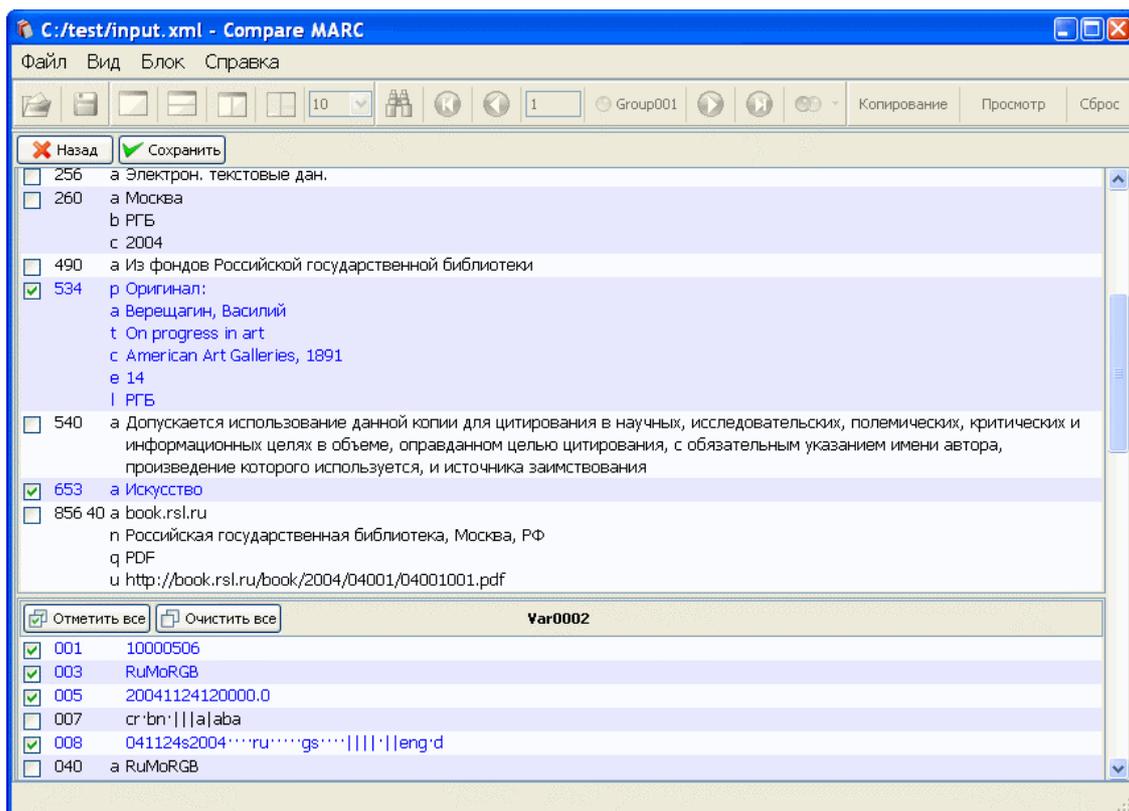


Рис. 4. Режим слияния библиографических записей

В режиме слияния библиографических записей библиотекарь имеет возможность выбора отдельных полей из разных записей, из которых будет составлена новая уникальная библиографическая запись.

4. Система хранения библиографических записей Метакс

Система Метакс предназначена для хранения библиографических записей и осуществления поиска по ним.

Основные свойства Метакс:

- позволяет загружать библиографические записи в хранилище как в обычном, так и в пакетном режиме;
- обеспечивает работу сразу нескольких пользователей с хранилищем;
- предоставляет возможность поиска по библиографическим записям с учётом морфологии русского языка;
- позволяет скачивать библиографические записи из хранилища в одной из четырёх кодировок: UTF8, WIN1251, KOI8-R, CP866.

Система Метакс имеет ряд преимуществ по сравнению с аналогичными системами. Во-первых, она содержит встроенный полнотекстовый поиск по всем библиографическим записям с учетом морфологии русского языка. Во-вторых, реализовано автоматическое разбиение библиографической записи на поля (индексирование). В-третьих, поиск по отдельным полям осуществляется существенно быстрее аналогов.

Главное окно клиентского приложения (рис. 5) системы Метакс состоит из верхнего меню, панели инструментов, главного меню (слева), рабочей области и строки состояния.

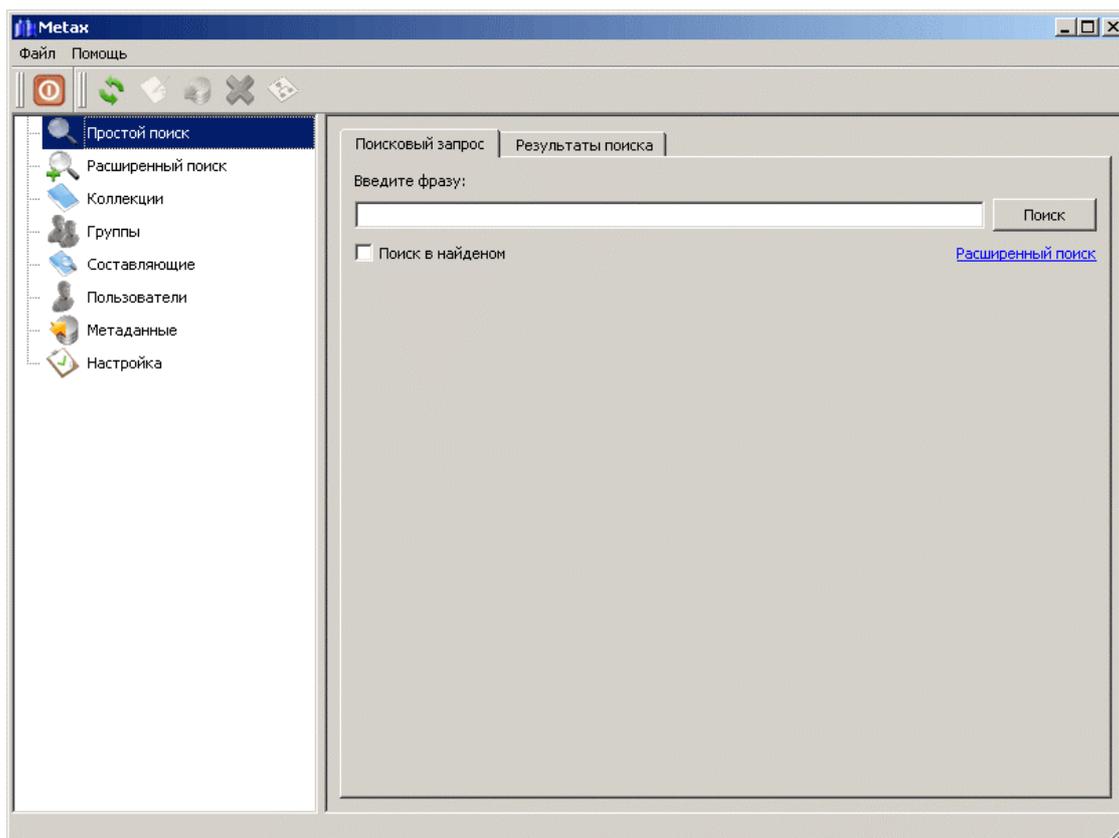


Рис. 5. Главное окно клиентского приложения системы Метакс

5. Заключение

В настоящей работе рассмотрены несколько программных средств, которые принимают участие в процессе автоматизации процесса обработки, хранения и выдачи электронных документов в Российской государственной библиотеке.

Во-первых, сделан обзор основных возможностей программного комплекса DefView, который предназначен для обеспечения безопасного доступа к электронным материалам, хранящимся в РГБ. Использование программного средства защищённого просмотра документов DefView, разработанного компанией «Шофт», позволяет библиотекам в полной мере обеспечить выполнение требований ГК РФ, вступивших в силу с 1 января 2008 года и ограничивающих предоставление библиотеками экземпляров произведений в цифровой форме, защищённых авторских прав, во временное безвозмездное пользование. Во-вторых, рассмотрено программное средство Comrag-MARC, позволяющее отобразить библиографические записи, с большой долей вероятности являющиеся дуплетами, в графическом режиме сравнить их друг с другом с возможностью создания новых записей на основе существующих. В-третьих, приведено описание системы хранения библиографических записей Метакс, позволяющей осуществлять поиск по ним.