



**Исида-Информатика**

# **ISIDA Retriever**

## **Руководство по установке в окружении Linux**

Код документа: 3921-2.5.25. Версия документа: 4  
Дата редакции документа: 24.06.2020. Количество листов: 10

Витебск, 2020 г.

## Содержание

<b>1.</b>	<b>НАЗНАЧЕНИЕ ДОКУМЕНТА</b> .....	<b>3</b>
<b>2.</b>	<b>УСТАНОВКА И ПЕРВИЧНАЯ НАСТРОЙКА СИСТЕМЫ</b> .....	<b>3</b>
2.1.	УСТАНОВКА ВЕБ-СЕРВЕРА .....	4
2.2.	УСТАНОВКА PHP-ИНТЕРПРЕТАТОРА .....	4
2.3.	УСТАНОВКА СУБД.....	5
2.4.	УСТАНОВКА ПРОГРАММНЫХ КОМПОНЕНТОВ ПРЕОБРАЗОВАНИЯ ФОРМАТОВ ФАЙЛОВ И ПОСТРОЕНИЯ ПОЛНОТЕКСТОВОГО ИНДЕКСА .....	5
2.4.1	Установка программных компонентов преобразования форматов файлов .....	5
2.4.2	Интеграция с компонентами преобразования файлов графических форматов .....	6
2.4.3	Установка программных компонентов поискового индекса .....	6
2.4.4	Интеграция с программными компонентами поискового индекса .....	7
2.5.	УСТАНОВКА И НАСТРОЙКА ЦЕНТРАЛЬНЫХ КОМПОНЕНТОВ СИСТЕМЫ .....	7
2.5.1	Установка Mediawiki.....	7
2.5.2	Установка центральных компонентов .....	8
2.5.3	Обновление БД.....	8
2.5.4	Настройка конфигурационных файлов.....	8
2.5.5	Запуск регламентов .....	8

© ЗАО «ЛМА»			
<i>Код</i>	<i>Описание</i>	<i>Лист</i>	<i>Листов</i>
3921-2.5.25	ISIDA Retriever. Руководство по установке в окружении Linux.	2	10
Принадлежность			
<i>Код</i>	<i>Наименование</i>	<i>Версия</i>	
3921	ISIDA Retriever	8.*	

## 1. Назначение документа

Документ содержит общие указания по установке прикладного программного обеспечения *ISIDA Retriever* (далее - Система) в окружении *Linux*.

Документ предназначен для информационно-технологического администратора.

## 2. Установка и первичная настройка Системы

Установка программных компонентов Системы состоит из следующих шагов:

1. Установка веб-сервера.
2. Установка *php*-интерпретатора.
3. Установка СУБД.
4. Установка программных компонентов преобразования форматов файлов.
5. Установка программных компонентов поискового индекса.
6. Установка и настройка базовых программных компонентов Системы.

Требования к серверному программному комплексу (поддерживаемые версии программного обеспечения приводятся в документе «Структура программного решения»):

1. ОС – *Linux Red Hat / CentOS / Fedora Core / Debian / Ubuntu*.
2. *PHP*.
3. Web-сервер (*Apache HTTP Server* или *Microsoft Internet Information Services*).
4. *Mediawiki*.
5. *MySQL* или *PostgreSQL*.
6. *Apache Solr*.
7. *SUN Open Office* или *LibreOffice* (для преобразования файлов форматов *Microsoft Office* или/и *Sun Open Office*).
8. *Ghostscript* (для преобразования файлов формата *Adobe PDF*).
9. *ImageMagick*.
10. *Vips (libvips)*.
11. *ffmpeg* и кодеки *flv, libx264* (для файлов документов видео-форматов).
12. Кодек *mp3* (для файлов документов видео-форматов и аудио-форматов).
13. Аудио- и видео-кодеки, соответствующие аудио- и видео-кодекам медиа-документов (при публикации документов, представленных файлами видео-форматов и аудио-форматов).
14. Программное обеспечение работы из командной строки с системой управления версиями файлов *Git* (при публикации документов на основе файлов из *Git*).

## 2.1. Установка веб-сервера

Процесс установки веб-сервера подробно описывается в соответствующей документации к выбранному веб-серверу и не является предметом данного документа.

Требуется включение модуля *mod\_rewrite*.

## 2.2. Установка php-интерпретатора

Процесс установки php-интерпретатора не является предметом данного документа.

Требуется включение следующих расширений к php-интерпретатору:

1. *curl*
2. *dom*
3. *exif*
4. *fileinfo*
5. *filter*
6. *gd*
7. *iconv*
8. *json*
9. *ldap*
10. *libxml*
11. *mbstring*
12. *pdo*
13. *pdo\_mysql*
14. *pgsql*
15. *simplexml*
16. *soap*
17. *sockets*
18. *tidy*
19. *tokenizer*
20. *xml*
21. *xmlreader*
22. *xmlrpc*
23. *xmlwriter*
24. *zip*
25. *zlib*

## 2.3. Установка СУБД

Рекомендации по установке выбранной СУБД приводятся в соответствующей документации к данной СУБД.

## 2.4. Установка программных компонентов преобразования форматов файлов и построения полнотекстового индекса



### Внимание!

Все описываемые в разделе манипуляции производятся в консоли на сервере.

Программные компоненты преобразования форматов файлов состоят из нескольких частей, предназначенных для преобразования (конвертирования) различных типов файлов.

Компоненты *Libre Office*, необходимы для получения текстового представления документов формата *Microsoft Word*, *Microsoft Excel*, *Microsoft PowerPoint*<sup>1</sup> и конвертирования файлов в формат *Adobe PDF*.

Программные компоненты *ImageMagick* и *Ghostscript* необходимы для получения *PDF*-представления документов, представленных графическими форматами.

В свою очередь, программные компоненты *libvips* необходимы для получения представления документов, представленных большеформатными графическими файлами.

Конвертеры являются наиболее требовательными по ресурсам и окружению частью Системы.

### 2.4.1 Установка программных компонентов преобразования форматов файлов

Программные компоненты преобразования форматов файлов устанавливаются следующей командой:

Для **Red Hat / CentOS**

```
sudo yum install unoconv ghostscript libvips imagemagick
```

Для **Fedora Core**

```
sudo dnf install unoconv ghostscript libvips imagemagick
```

---

<sup>1</sup> К таким файлам относятся файлы с расширениями: *doc*, *docx*, *xls*, *xlsx*, *ppr*, *pptx*, *rtf*, *txt*.

**Для Debian / Ubuntu**

```
sudo apt-get install unoconv ghostscript libvips imagemagick
```

**2.4.2 Интеграция с компонентами преобразования файлов графических форматов**

Выполните настройку Системы для работы с компонентами преобразования файлов графических форматов. См. раздел "Файл конфигурации" документа «ISIDA Retriever. Руководство по установке и первичной настройке», пункт "Настройки программных компонентов преобразования форматов файлов".

При наличии в используемой прикладной конфигурации доступа к функциям публикации и просмотра видео-документов, скопируйте содержимое каталога *retriever.dir* из архива *\*.util\*.zip* в каталог Системы.

**2.4.3 Установка программных компонентов поискового индекса**

Построение поискового индекса и непосредственно поиск осуществляется программным продуктом *Apache Solr*. Функционирует *Apache Solr* под управлением другого программного продукта – *Jetty* (<http://www.eclipse.org/jetty/>).

Дистрибутив *Apache Solr* можно скачать с сайта производителя: <http://lucene.apache.org/solr/>.

Установите программный продукт *Apache Solr*, следуя инструкциям производителя.

Распакуйте архив *\*.util\*.zip*. Содержимое каталога *solr.dir* поместите в каталог установки *Apache Solr*. Должно получиться примерно следующее:

```

.|solr {версия}
    server
        solr\
            documents
            conf
            data
            core.properties
            dictionary
            conf
            data
            core.properties

```

Каталог *data* – это каталог с данными *Apache Solr*. Для высоконагруженных систем следует разместить его в разделе файловой системы с достаточным

количеством свободного места. Однако, при неудачном размещении, в будущем этот каталог можно будет перенести в другое место.



### **Внимание!**

Большое значение на быстродействие Apache Solr, тем самым и на быстроту отклика системы на поисковые запросы, имеет быстродействие файловой системы мощность процессора компьютера, на котором работает Apache Solr. При использовании подсветки искомых слов в результатах поиска нагрузка на процессор многократно возрастает.

В том случае, если потребуется перенести хранилище данных Apache Solr, отредактируйте файл `solr {версия}\server\solr-webapp\webapp\WEB-INF\web.xml`. Снимите комментарии с секции с атрибутом `solr/home` и укажите реальный путь к каталогу с данными Apache Solr (к примеру: `C:\Programs\solr-data`).

Apache Solr установлен успешно, если по адресу `http://{solr_server}:{solr_port (по умолчанию 8983)}/solr` открывается приветственная страница Apache Solr.

## **2.4.4 Интеграция с программными компонентами поискового индекса**

Выполните настройку Системы для работы с Apache Solr. См. раздел «Файл конфигурации» документа «ISIDA Retriever. Руководство по установке и первичной настройке», пункт «Настройки полнотекстового индекса».

## **2.5. Установка и настройка центральных компонентов Системы**

Процедура установки и настройки центральных компонентов Системы состоит из следующих этапов:

1. Установка *Mediawiki*.
2. Установка центральных компонентов Системы.
3. Создание БД в рамках используемого экземпляра СУБД.
4. Настройка конфигурационного файла.
5. Определение регламентов.
6. Интеграция с *ionCube Loader* (при необходимости).

### **2.5.1 Установка Mediawiki**

Процесс установки *Mediawiki* заключается в

- распаковке архива с дистрибутивом `mediawiki-*.tar.gz` в каталог приложений веб-сервера,
- описание, при необходимости, нового приложения в рамках веб-сервера.

- выполнение процедуры установки *Mediawiki* (см. соответствующую документацию к программному продукту *Mediawiki*).

В процессе установки *Mediawiki*:

- Движок базы данных – *InnoDB*.
- Кодовая таблица базы данных – *UTF-8*.

### 2.5.2 Установка центральных компонентов

Установка центральных компонентов представляет собой распаковку архива с дистрибутивом Системы в каталог с установленным программным продуктом *Mediawiki*.

При совпадении имен файлов файлы из дистрибутива Системы должны заменить имеющиеся файлы.

В каталоге *rt-utils* переименуйте файлы:

*pdftopng.64.linux* ► *pdftopng*  
*wkhtmltopdf.linux* ► *wkhtmltopdf*

### 2.5.3 Обновление БД

Обновление таблиц в базе данных, созданной ранее при установке *Mediawiki*, может быть произведено при помощи любой утилиты, позволяющей выполнять *sql*-скрипты.

Обновление осуществляется с помощью *sql*-скрипта *rt-install/sql/{dbms}/{dbms}\_create\_db\_1.sql*.

### 2.5.4 Настройка конфигурационных файлов

См. раздел "Файлы конфигурации" документа «ISIDA Retriever. Руководство по установке и первичной настройке».

### 2.5.5 Запуск регламентов

В рамках Системы для выполнения некоторых задач, в числе которых:

1. полнотекстовое распознавание графических образов документов,
2. обновление индекса документов,
3. проверка срока действия документов,



4. создание версий документов для печати,
  5. создание отображения документов
- и др. требуется запуск специализированных регламентов.

Регламенты представляют собой исполняемые скрипты формата \*.sh. Размещаются регламенты в каталоге *rt-schedule*.

Запуск регламентов может быть осуществлен с помощью *cron* (*Linux*).

Таблица 2.5.5-1

Регламент	Назначение	Рекомендуемая частота запуска
<i>runJobs</i>	Запуск на выполнение очереди заданий. В рамках выполнения очереди производится: полнотекстовое распознавание графических образов документов, обновление индекса документов, создание версий документов для печати, создание отображения документов. Регламент прекращает работу после полной очистки очереди заданий.	Каждые 5 минут. Повторный запуск регламента при работе предыдущего экземпляра осуществлен не будет.
<i>datebookEventCheck</i>	Проверка событий ежедневника системы.	Ежеминутно.
<i>taskEventCheck</i>	Проверка сроков исполнения заданий.	Ежедневно.
<i>expiryContractCheck</i>	Информирование пользователей за настраиваемое количество дней об истечении срока действия договора	Ежедневно.
<i>importDars</i>	Импорт результатов пакетного сканирования.	Выполняется вручную.
<i>syncLDAP</i>	Синхронизация учетных данных с LDAP.	Ежедневно.
<i>expiryContractCheck</i>	Регламент, предназначенный для контроля сроков договора. Выполняет информирование пользователей за настраиваемое количество дней об истечении срока действия договора.	Регламент запускается один раз в сутки.
<i>dalayedRecognition</i>	Регламент, выполняющий отложенное распознавание ранее помещенных документов.	Каждые 15 минут.

Регламент	Назначение	Рекомендуемая частота запуска
<i>checkAccessionDate</i>	Регламент, выполняющий проверку наступления у редакций документа даты вступления в силу и установку признака актуальности соответствующей редакции.	Ежедневно.

Для подготовки к запуску регламентов следует:

1. Переименовать файл запуска регламента. К примеру: *runJobs.sample.sh* в *runJobs.sh*.
2. Исправить пути в файле запуска регламента.
3. Создать задание в рамках *cron*<sup>2</sup>.



### Внимание!

Так как поиск документов осуществляется по БД полнотекстового индекса, поиск по измененным значениям реквизитов документов может быть осуществлен только при завершении работы регламента выполнения очереди заданий *runJobs*. Аналогично дело обстоит и с процессом обновления полнотекстового индекса для всех документов. Процесс же обновления полнотекстового индекса для одного документа, осуществляемый с закладки «Операции», осуществляется в режиме *on-line*.

<sup>2</sup> Возможно использование любых аналогов предлагаемых программных средств выполнения регламентных процедур.