**Техническое задание на создание утилиты**

**«Загрузчик документов»**

**1. Назначение и цели создания утилиты**

**1.1 Назначение утилиты.**

Утилита предназначена для быстрого наполнения ИС «Банк документов» файлами Заказчика в форматах pdf, tif, doc, docx, rtf.

**1.2 Цели создания утилиты**

Автоматизация создания карточек документов на основе реестра документов.

Сортировка файлов с текстами документов из одной папки в структуру папок в соответствии с требованиями предъявляемыми ИС «Банк документов» для загрузки документов.

**2. Функции утилиты**

Утилита должна выполнять следующие функции:

- принимать данные о значениях атрибутов из реестра документов и данные о месте хранения исходных файлов с текстами документов;

- на основе данных реестра, создавать карточки документов в формате \*.a2 по определенным требованиям;

- выгружать заполненные карточки документов в определяемый Пользователем каталог;

- копировать файлы с текстами документов в отдельные папки. Структура дерева папок должна соответствовать требованиями.

**3. Технические требования к утилите.**

**3.1. Требования в утилите в целом**

3.1.1. Утилита «Загрузчик документов» должна:

* обеспечивать выполнение всех требований, указанных в настоящем техническом задании;
* обеспечивать возможность запуска на локальном компьютере Заказчика без установки дополнительного программного обеспечения.

3.1.2. Требования по безопасности

Информационная безопасность:

* журнализация действий с указанием ошибок и их причин при их возникновения.
* отсутствие изменений в реестре документов и исходных файлах с текстами документов (включая их перемещение).

3.1.3 Требования к кодировке

Все файлы формата a.2 должны быть в кодировке ANSI

**3.2. Требования формированию карточек.**

Исходные данные для формирования карточек берутся из реестра документов. В реестре перечислены атрибуты каждого документа, такие как:

- Наименование

- Обозначение

- Номер

- дата принятия

- дата начала действия

- статус

- и др

Количество атрибутов для одного документа до 90 штук.

Количество документов, значения атрибутов которых перечислены в реестре, до 50 000.

Результатом работы утилиты должен являться набор файлов (карточек) в формате \*.а2, каждый из которых содержит значения атрибутов по одному документу. Каждому атрибуту назначается номер, соответствующий номеру атрибута в ИС «Банк Документов».

Пример содержания карточки:

1: Система стандартов безопасности труда. Средства индивидуальной защиты глаз и лица при сварке и аналогичных процессах. Общие технические условия

6: 12.4.254-2013

15: 23.09.2014

7927: ГОСТ 12.4.254-2013

7774: ГОСТ

7772: действ

7860: ГОСТ 12.4.254-2013.pdf

….

N: «…»

Где N – номер атрибута, : - разделитель, «…» - значение атрибута.

Каждая созданная карточка должна иметь имя в формате – «N.а2», где N – уникальный номер карточки в рамках обротки фонда документов текущего Заказчика. То есть, если обработка фонда идет в несколько этапов (был сбой или надо загружать документы группами), каждой номера карточек не должны повторяться. Для другого заказчика повторение номеров карточек допустимо.

Данные реестра должны загружаться в утилиту либо непосредственно копированием данных из реестра документов Заказчика, либо из путем автоматического разбора данных промежуточного файла xls созданного по определенному шаблону и наполненного данными из реестра документов Заказчика.

Соответствие номера атрибута и названия атрибута назначаются либо в интерфейсе утилиты, либо в промежуточном файле xls

Атрибут 7860 должен присутствовать только если к данной карточке будет прикрепляться документ в формате pdf и tif. Его значение = имени прикрепляемого файла.

**3.3 Требования к обработке файлов с текстами документов в формате pdf и tif (вкладка скан-копия).**

Файлы заказчика в форматах pdf и tif размещаются в одной или нескольких папках. Путь к папке с файлами должен указываться из интерфейса утилиты. Имя файла и его расширение указывается либо в промежуточном файле xls, либо формируется утилитой автоматически из наименования или/и номера документа (изменяемый параметр, настраиваемый в утилите).

В результате работы утилиты каждый файл должен быть скопирован в отдельную папку. Имя папки - N\_attach, где N = уникальному номеру карточки соответствующей данному файлу.

Вместе с самим файлом в папке должен формироваться дополнительный файл.а2 в котором:

1: ГОСТ 12.4.254-2013.pdf – имя вложенного файла для отображения в ИС = 7860: из карточки

13: ГОСТ 12.4.254-2013.pdf – имя реального файла

Путь создания папок N\_attach = путь создания карточек.

**3.4 Требования к обработке файлов с текстами документов в формате doc, docx, rtf.**

Файлы заказчика в форматах doc, docx, rtf размещаются в одной или нескольких папках. Путь к папке с файлами должен указываться из интерфейсе утилиты. Имя файла и его расширение указывается либо в промежуточном файле xls, либо формируется утилитой автоматически из наименования или/и номера документа (изменяемый параметр, настраиваемый в утилите).

В результате работы утилиты каждый файл должен быть скопирован в отдельную папку. Имя папки N\_text, где N = уникальному номеру карточки соответствующей данному файлу.

При копировании файла его имя должно меняться по шаблону N\_7845.doc/docx/rtf, где N – уникальный номер карточки соответствующей данному файлу.

Рядом с doc-файлом должен создаваться файл mht сконвертированный из doc с именем N\_7778.mht

**3.5 Требования к обработке файлов с текстами документов в формате pdf (вкладка текст).**

Файлы заказчика в формате pdf размещаются в одной или нескольких папках. Путь к папке с файлами должен указываться из интерфейса утилиты. Имя файла и его расширение указывается либо в промежуточном файле xls, либо формируется утилитой автоматически из наименования или/и номера документа (изменяемый параметр, настраиваемый в утилите).

В результате работы утилиты каждый файл должен быть скопирован в отдельную папку. Имя папки - N\_text, где N = уникальному номеру карточки соответствующей данному файлу.

**3.6 Требования к обработке файлов вложений (вкладка вложенные файлы).**

Файлы заказчика в любых форматах размещаются в одной или нескольких папках. Путь к папке с файлами должен указываться из интерфейса утилиты. Имя файлов и их расширение указывается либо в промежуточном файле xls, либо формируется утилитой автоматически из наименования или/и номера документа (изменяемый параметр, настраиваемый в утилите).

Файлов вложений может быть несколько!

В результате работы утилиты файлы вложений должны быть скопированы в отдельную папку. Имя папки - N\_attach, где N = уникальному номеру карточки соответствующей данному файлу.

Вместе с файлами в папке должны формироваться дополнительные файлы с расширением «а2» в которых:

1: ГОСТ 12.4.254-2013.pdf – имя вложенного файла для отображения в ИС

13: ГОСТ 12.4.254-2013.pdf – имя реального файла

Путь создания папок N\_attach = путь создания карточек.

В результате в папке N\_attach должен быть такой список файлов:

Вложение1.a2

Вложение1.pdf

Вложение2.a2

Вложение2.pdf

**…**

ВложениеN.a2

ВложениеN.pdf

**3.7. Требования к интерфейсу.**

В интерфейсе системы должна быть возможность верифицировать и изменять данные загруженные из промежуточного xls или введенных пользователем данных.