

Задание к модулю Windows Services

Служба обработки результатов потокового сканирования

При массовом сканировании документов довольно часто используется следующая схема работы:

- Потоковый сканер (сканер с автоподачей бумаги) выдает последовательность изображений, которые сохраняются в папку на локальном компьютере или в сети (чаще всего это последовательность вида `img_001.jpeg`, `img_002.jpeg`, ...)
- За этой папкой (или папками – их может быть несколько, если используется несколько потоковых сканеров) следит специальная программа, которая:
 - Разделяет поток изображений на отдельные документы
 - Собирает и сохраняет их в виде многостраничного файла (обычно это PDF, многостраничный TIFF или XPS)
 - Передает их для дальнейшей обработки (пересылает по почте, заносит в СЭД, просто складывает в другую папку)

Для того, чтобы разделять последовательность листов на отдельные документы, обычно используют листы или наклейки со штрих-кодами. В самом простом сценарии это выглядит так:

- Между сканируемыми документами кладется листок со штрих-кодом (его подбирают так, чтобы он гарантированно не встретился в документах)
- Как только служба обрабатывающая последовательность встречает изображение нужного ей штрих-кода, она понимает, что предыдущий документ закончился и начинается новая последовательность страниц.

В данном задании мы создаем упрощенный вариант такой службы.

Пункты со звездочкой (*) являются необязательными – делать их только по согласованию с ментором.

Задание 1

Разработать службу «склейки», результатов работы потокового сканера в единый PDF (или XPS) файл.

Служба должна уметь следующее:

- Следить за одной или несколькими папками на диске или в сети
- При появлении в папках изображений, подходящих именами под шаблон `<префикс>_<номер>.<png|jpeg|...>` собирать последовательность страниц и сохранять их как единый документ в отдельную папку.
- При запуске просматривать папки для мониторинга и при наличии там подходящих файлов также объединять их в документ.

В качестве признака «конца документа» предлагается использовать следующие:

- «Перескок» нумерации документов (т.е. `img_001.jpeg`, `img_002.jpeg`, `img_004.jpeg` – это два документа из 2 и 1 страниц)
- Истек таймаут появления следующей страницы
- (*) На очередной странице обнаружен штрих-код с определенным значением

Примечание! Прежде чем приступать к заданию, обсудите с ментором следующие моменты:

- В каком формате будут сохраняться документы:
 - PDF (можно, например, использовать связку библиотек [PdfShar+MigraDoc](#))
 - XPS (используя стандартный [FixedDocument](#))
 - Многостраничный TIFF (например, как описано в <http://stackoverflow.com/questions/398388/convert-bitmaps-to-one-multipage-tiff-image-in-net-2-0>)
- Будете ли реализовывать функционал «разделение по штрих-коду» (например, на основе библиотеки [ZXing.Net](#))
- На базе какого фреймворка вы будете реализовывать службу: System.ServiceProcess или Topshelf.

Задание 2

Проверьте, что ваша служба корректно обрабатывает следующие ситуации:

- Исходный файл, который вам нужно обработать занят другим процессом (обычно делают несколько попыток открыть и только после их истечения – ошибка)
- Пришла команда остановки службы во время длительной операции (например, собирания многостраничного файла из множества страниц)
- Одна или несколько страниц для многостраничного файла оказались «битыми» (например, неверный формат). Возможная реакция – перемещение всей последовательности в отдельную папку «битых», для дальнейшей ручной корректировки