Сервис (служба) автоматического занесения документов

Подобного рода службы используются в системах электронного документооборота (СЭД) для ввода (захвата, capture) документов из различных источников. Как правило, подобного рода службы могут:

• Мониторить различные источники документов (почтовые ящики, факсы, сканеры, просто файловые папки)

• Преобразовывать документы (конвертировать в различные форматы, собирать несколько изображений в один документ, …)

• Извлекать метаданные (с помощью распознавания или, например, данных из письма или факсимильного сообщения)

• Заносить подобранные файлы в СЭД в виде документа и запускать процесс обработки (например, высылать уведомление делопроизводителю)

В данном задании мы создаем упрощенный вариант такой службы.

Задание 1. Разработать простую службу мониторинга файловой системы и занесения документов в некую систему управления документами (далее СЭД). Служба должна уметь следующее:

• Следить за одной или несколькими папками на диске или в сети • При появлении в этих папках документов, переносить документы в СЭД (можно в одну и ту же папку)

• При запуске просматривать папки для мониторинга и при наличии там файлов переносит их тоже В качестве СЭД может выступать любая (на ваш выбор) система работы с документами/файлами. Например, это может быть OneDrive или OneDrive for Business (с ним можно работать через SDK: <https://dev.onedrive.com/sdks.htm>)

Задание 2. Дополните службу функционалом «склейки» последовательности изображений в один общий файл (такое требуется для поддержки работы с потоковыми сканерами, которые не умеют формировать многостраничные документы сами). Т.е. ваша служба должна для отдельных папок следить не за появлением одиночных файлов, а за последовательностями из image-файлов. При этом вам нужно будет самостоятельно решить (и согласовать с ментором) следующие аспекты:

• Как будут упорядочиваться файлы для сборки в один документ (например, по дате создания или по имени: 001.jpeg, 002.jpeg, …)?

• Что будет служить признаком окончания последовательности, или, иначе говоря, что уже все страницы одного документа представлены (например, можно именовать Doc01-001.jpeg, Doc01-002.jpeg, Doc02-001.jpeg, … или раскладывать по разным папкам и ждать заполнения папки не более определенного количества минут, …)?

• В какой формат будете конвертировать документы: PDF (например, с помощью http://www.pdfsharp.com), XPS (например, с помощью стандартного FixedDocument) или многостраничный TIFF (например, как описано в <http://stackoverflow.com/questions/398388/convert-bitmaps-to-one-multipage-tiff-image-in-net-2-0>)

Задание 3 Проверьте, что ваша служба корректно обрабатывает следующие ситуации:

• СЭД недоступна (служба должна делать несколько попыток отправить документ с некоторым интервалом, но не «зацикливаться»)

• Отсылаемый файл занят другим процессом (опять-таки: 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 В любом случае – все ситуации должны логгироваться.