

Skanywanie wraz z automatyczną dekretacją dokumentów w aplikacji Komornik SQL

Wersja dokumentu: 1.0

Data opracowania: 05 październik 2012 roku

Autorzy: Łukasz Czarnecki, Rafał Kaczor, Piotr Kaczmarzyk

Spis treści

Spis treści.....	2
Wstęp	3
Wymagania.....	3
Rejestracja modułu rozpoznającego tekst	3
Konfiguracja aplikacji.....	6
Preferencje skanowania	8
Filtry dla modułu OCR.....	10
Typy skanowanych dokumentów	11
Zmienne.....	11
Wyrażenia regularne w zmiennych	13
Reguły	16
Predefiniowane wyrażenia regularne	17
Nazwy skanowanych dokumentów	18
Drukowanie etykiet	19
Modele skanowania oraz dekretacji dokumentów	20
Model 1. Skanowanie seryjne z ręczną dekretacją dokumentów	20
Model 2. Skanowanie seryjne z dekretacją dokumentów na podstawie kodów kreskowych.....	20
Model 3. Skanowanie seryjne z wykrywaniem znaków oraz automatyczną dekretacją dokumentów ..	20
Proces skanowania dokumentów.....	21
Lista skanowanych dokumentów	21
Skanowanie seryjne.....	24
Skanowanie dokumentów w modelu 1	25
Skanowanie dokumentów w modelu 2	26
Skanowanie dokumentów w modelu 3	27
Uwagi końcowe	28

Wstęp

Zapewnienie natychmiastowego dostępu do informacji w codziennej pracy kancelarii jest warunkiem koniecznym do prowadzenia szybkiego i skutecznego postępowania egzekucyjnego. Sterty akt egzekucyjnych w formie tradycyjnej nie ułatwiają wyszukiwania w nich rozmaitych informacji. Rozwiązaniem opisanego problemu może być prowadzenie bazy zawierającej elektroniczne odpowiedniki dokumentów znajdujących się w aktach egzekucyjnych. Proces skanowania oraz przyporządkowywania dokumentów do odpowiednich rekordów bazy danych może jednak zniechęcić każdego użytkownika aplikacji kancelaryjnej.

Dostarczamy Państwu rozwiązanie wspierające użytkowników systemu w zakresie skanowania dokumentów. Dzięki wykorzystaniu modułu rozpoznawania tekstu (OCR) możliwe jest wykorzystanie wykrytych w zeskanowanym obrazie słów do przeprowadzenia procesu automatycznej dekretacji dokumentów. Dostarczenie w pełni konfigurowalnego modułu interpretacji odnalezionych danych umożliwia użytkownikom programu dostosowanie go do indywidualnych potrzeb. Dzięki zastosowaniu modelu wykrywającego stałe wyrażenia i wzorce tekstu definiowane przez użytkowników w postaci wyrażeń regularnych oraz zastosowanie reguł postępowania w przypadku ich wykrycia użytkownicy otrzymują potężne narzędzie spełniające oczekiwania nawet tych najbardziej wymagających.

Wymagania

Do poprawnego działania modułu przeznaczonego do skanowania dokumentów wymagane jest posiadanie zainstalowanego urządzenia skanującego obsługującego proces akwizycji dokumentów przy użyciu sterowników typu TWAIN. Poza opisanym wymaganiem nie ma żadnych innych mogących mieć wpływ na przeprowadzany w aplikacji proces skanowania.

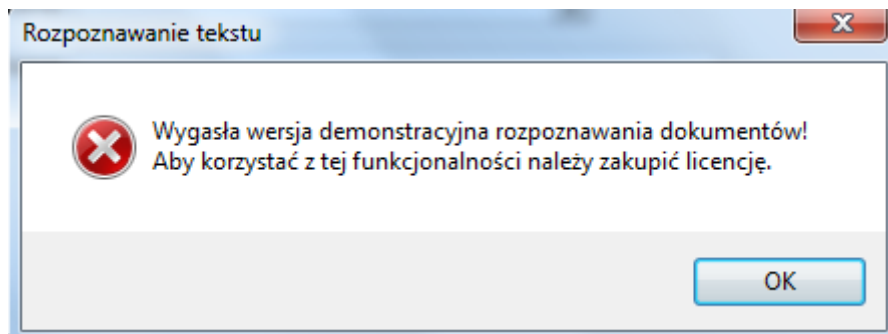
Wykrywanie tekstu w obrazach (OCR) jest bardzo złożoną operacją wymagającą do jej wykonania dużej mocy obliczeniowej. Aby zapewnić wysoką prędkość wykonywania tej operacji oraz odpowiedni komfort pracy użytkownikom trzeba o tym pamiętać. Przy wyborze stacji klienckiej, na której wykorzystywana będzie opisywana funkcjonalność należy kierować się właśnie jej mocą obliczeniową.

Rejestracja modułu rozpoznającego tekst

Twórcy aplikacji Komornik SQL udostępniłi wszystkim użytkownikom możliwość sprawdzenia funkcjonalności działania modułu inteligentnego skanowania. W tym celu powstała demonstracyjna wersja oprogramowania, działająca przez 30 dni od daty pierwszego uruchomienia.

Aby rozpoznawanie skanowanych dokumentów działało prawidłowo, należy posiadać zainstalowaną bibliotekę *LeadTools*.

Po wygaśnięciu wersji demonstracyjnej program poinformuje o tym fakcie użytkownika oraz zada pytanie czy użytkownik chce wykupić licencję dla skanowania OCR.



Użytkownikowi zainteresowanemu wykupem licencji zostanie udostępniony klucz licencyjny, który pozwoli na dalsze korzystanie z modułu inteligentnego skanowania.

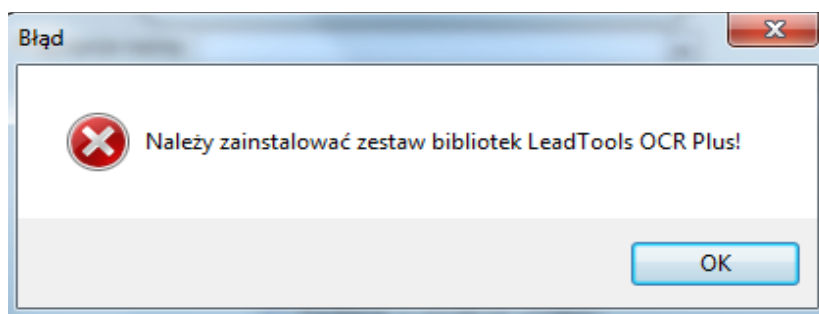
W przypadku braku zainstalowanej biblioteki LeadTools, program umożliwia jedynie dekretację dokumentów tzw. Searchable pdf (wyszukiwalny pdf).

Rejestracji modułu rozpoznającego tekst można dokonać z poziomu głównego menu po wciśnięciu przycisku „Rejestracja programu”.

Instalacja bibliotek *LeadTools*

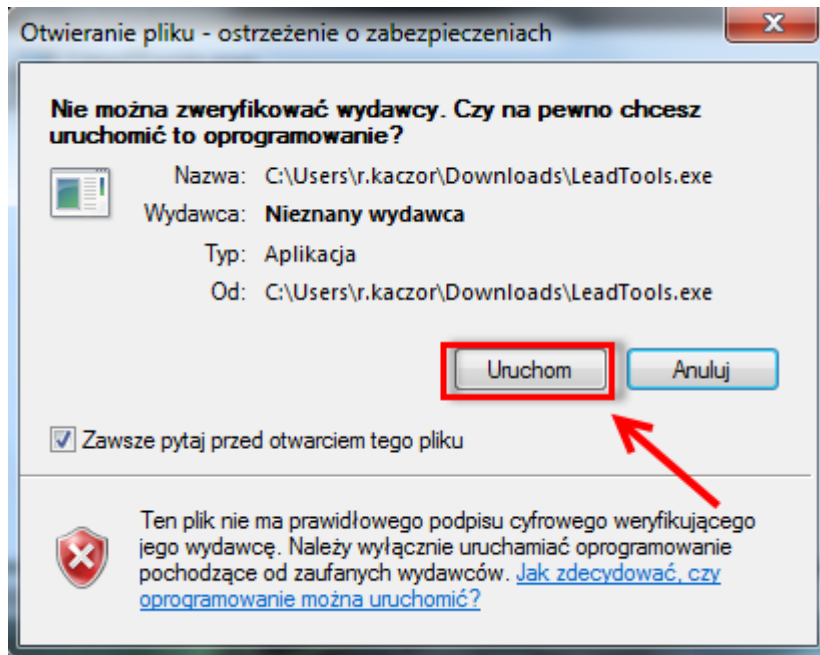
Wraz z wersją aplikacji która umożliwia automatyczne skanowanie i dekretację dokumentów, udostępniony zostanie instalator z bibliotekami *LeadTools*, którego instalacja wymagana jest do pracy z modułem.

W przypadku braku zainstalowanej biblioteki *LeadTools*, przy próbie skanowania z włączoną opcją rozpoznawania OCR, program wyświetli poniższy komunikat:

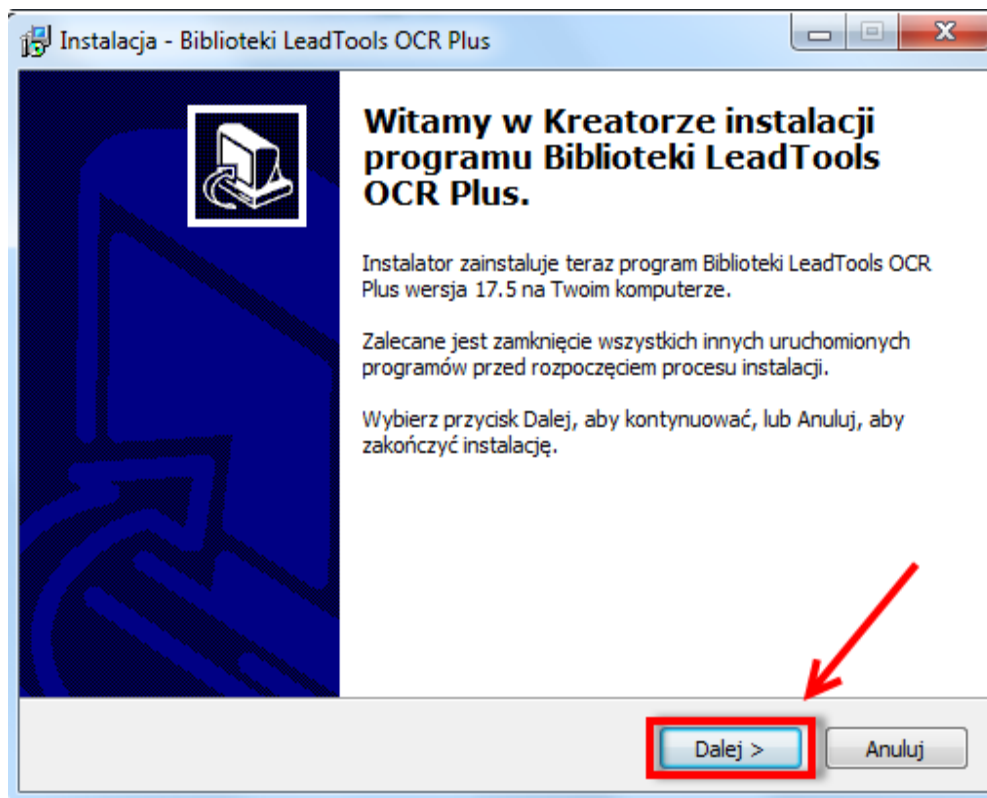


Proces instalacji biblioteki *LeadTools* jest opisany poniżej.

Po pobraniu pliku instalatora, należy odnaleźć go w miejscu gdzie odbył się zapis i dwukrotnie kliknąć plik. Pojawi się okno systemu operacyjnego informujące o zabezpieczeniach.

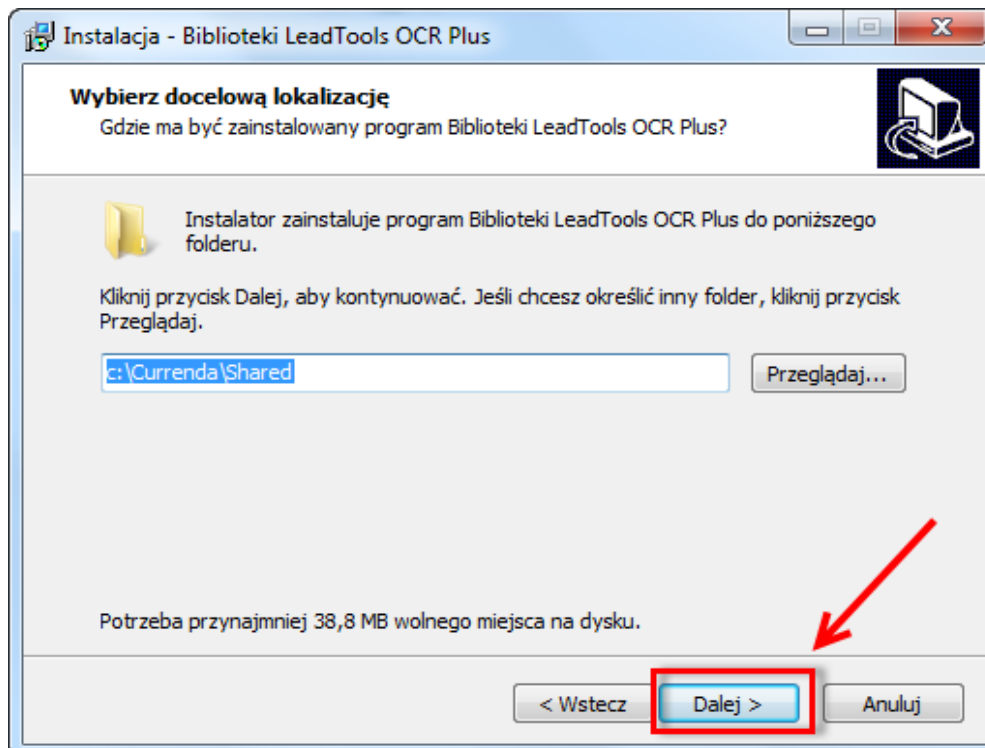


Należy kliknąć przycisk „Uruchom”

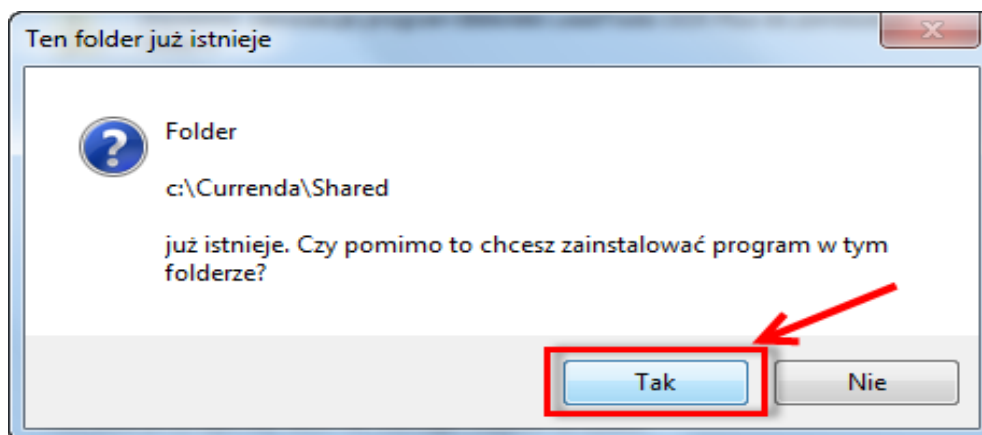


Użytkownikowi pojawi się okno instalatora biblioteki LeadTools OCR Plus.

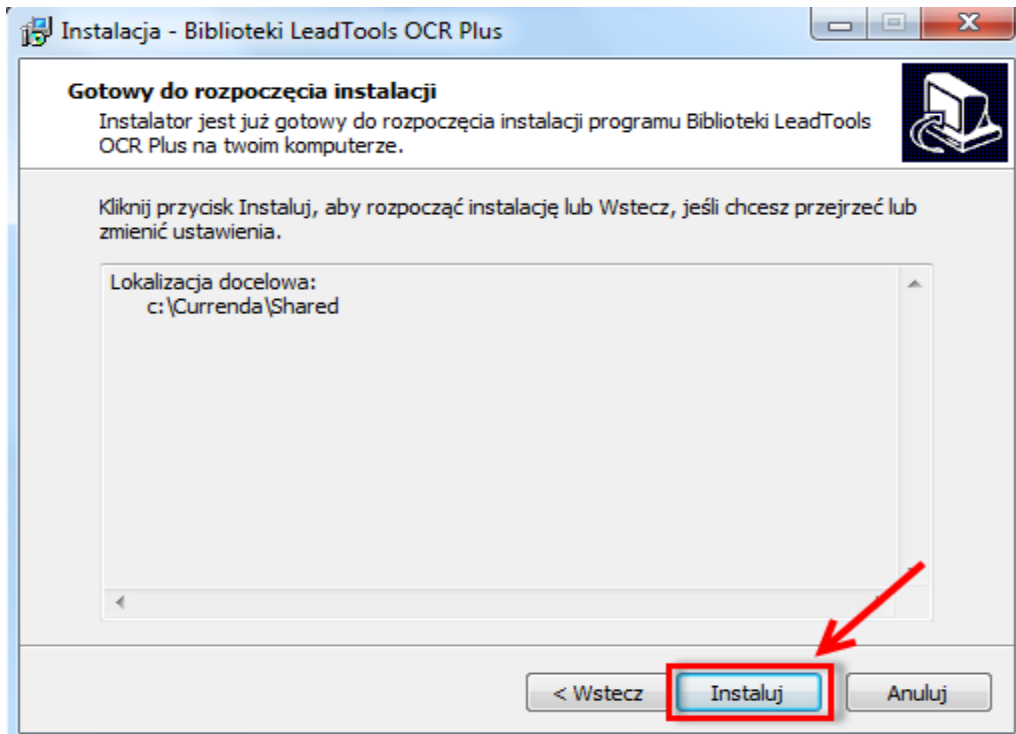
Aby kontynuować instalację biblioteki należy kliknąć przycisk „Dalej”



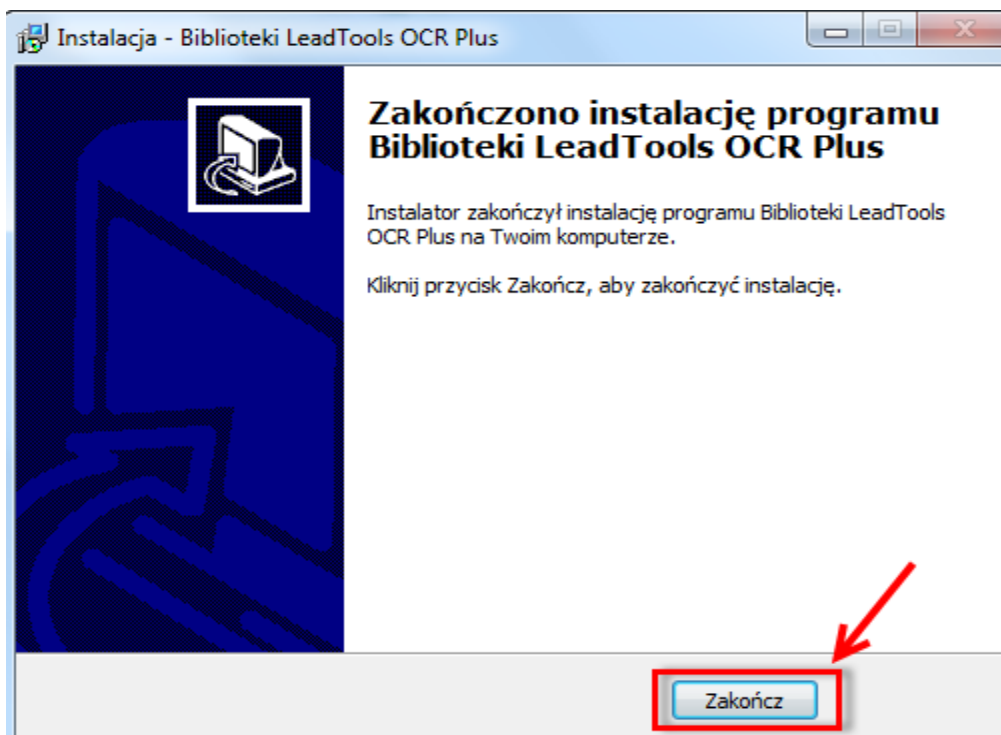
Kolejnym oknem, które się pojawi podczas instalacji, jest okno w którym należy określić lokalizację instalacji bibliotek LeadTools. Biblioteki należy umieścić w katalogu z którego uruchamiany jest program. Domyślna ścieżka w systemie operacyjnym to C:\ Currenda\Shared, i tę ścieżkę należy zaakceptować klikając przycisk „Dalej”



Następnie należy zaakceptować chęć instalacji w wybranym folderze.



Następnym krokiem jest zainstalowanie biblioteki w wybranym katalogu.



Gdy instalacja się zakończy, pojawi się okno informujące o zakończeniu instalacji biblioteki LeadTools OCR. Należy kliknąć przycisk „Zamknij” aby zamknąć okno.

Konfiguracja aplikacji

Preferencje skanowania

Konfiguracja modułu skanującego dostępna jest w module „Konfiguracja” | menu „Preferencje” | „Konfiguracja skanowania”.

Rejestracja	[L] - Domyślne źródło skanowania
Zestawienia	<input type="radio"/> Plik <input checked="" type="radio"/> Skaner
Zaległość	[L] - Parametry skanowania
Pisma	<input type="checkbox"/> Wyświetl okno wyboru źródła
Wydruki	Nazwa: FUJITSU fi-5110Cdj #2
Książka nadawcza	Typ: FUJITSU Scanner
Terminarz	Wersja: 1.9
Księgowanie	Tryb: Biało-czarny
E-Bank	[L] - Opcje urządzenia
E - Poczta	<input type="checkbox"/> Wyświetl okno skanowania
E - Wyciąg	DPI: 300 Format: PDF
E - Sąd	<input checked="" type="checkbox"/> Podajnik <input checked="" type="checkbox"/> Duplex
E - KIR (Ognivo)	<input type="checkbox"/> Automatyczny rozmiar <input type="checkbox"/> Automatyczna rozdzielczość
Skanowanie	[L] - Rozpoznawanie dokumentów (OCR)
Serwer Proxy	<input checked="" type="checkbox"/> Szukanie kodu kreskowego
Katalogi	<input checked="" type="checkbox"/> Włącz funkcję OCR (wymagana biblioteka LeadTools)
Inne	OCR DPI: 300
Konta email	<input checked="" type="checkbox"/> Automatyczna dekretacja dokumentów
E-zapytania - CPD	<input checked="" type="checkbox"/> Automatyczne pozycjonowanie i wykrywanie pustych stron
	<input checked="" type="checkbox"/> Logowanie wyników do pliku ocr.log
	Weryfikacja rozpoznania sygnatury: szukanie nazw stron
	Tolerancja daty dokumentu: 14

Wyświetlone okno zawiera wszystkie dostępne parametry wykorzystywane podczas procesu skanowania, rozpoznawania oraz dekretacji dokumentów w aplikacji *Komornik SQL*. Kolejne opcje oznaczają kolejno:

- Grupa „[L] – **Domyślne źródło skanowania**” – pozwala określić, co ma być źródłem podłączanych dokumentów:
 - „Plik” – okno „Skanowanie dokumentów” domyślnie posiada zaznaczoną opcję wprowadzania dokumentów z pliku znajdujących się na dysku twardym komputera.
 - „Skaner” - okno „Skanowanie dokumentów” domyślnie posiada zaznaczoną opcję wprowadzania dokumentów przy użyciu jednego z zainstalowanych w systemie operacyjnym urządzeń skanujących.

- Grupa „[L] – **Parametry skanowania**” – pozwala określić podstawowe opcje związane z procesem wprowadzania dokumentów przy użyciu urządzenia skanującego:
 - „*Wyświetl okno wyboru źródła*” – pozwala włączyć opcję wyboru urządzenia skanującego podczas pracy w module skanowania dokumentów. Przy włączonej opcji aplikacja udostępnia możliwość zdefiniowania skanera tak, aby był domyślnie stosowany, a okno wyboru nie pojawiało się z każdym razem.
 - „*Nazwa*” – możliwość wyboru z listy urządzenia skanującego. Na liście widoczne są wszystkie urządzenia skanujące zainstalowane w systemie operacyjnym komputera.
 - „*Tryb*” – wybór trybu skanowania. Dostępne do wyboru trzy standardowe tryby skanowania: biało czarny, skala szarości, kolorowy.
 - Pozostałe dwa pola „*Typ*” oraz „*Wersja*” są nieedytowalne i wyświetlają informację na temat urządzenia skanującego odczytane ze sterownika.

- Grupa „[L] – **Opcje urządzenia**” – pozwala określić opcje urządzenia wykorzystywane w procesie wprowadzania dokumentów przy użyciu urządzenia skanującego.
 - „*Wyświetl okno skanowania*” – pozwala włączyć lub wyłączyć okno przeznaczone do wyboru ustawień parametrów skanowania. W przypadku zaznaczonej opcji wyświetlane będzie okno umożliwiające ustawienie odpowiednich parametrów z ich wstępnym zaznaczeniem na podstawie opcji wybranych poniżej, które podlegają dalszym zmianom. W przeciwnym przypadku okno nie zostanie w ogóle wyświetlone, a użyte zostaną wybrane poniżej parametry – nie będzie możliwości poddania ich dalszym zmianom.
 - *DPI* – wybór rozdzielczości zapisu skanowanego dokumentu do bazy. Dostępne są opcje zapisu w następujących rozdzielczościach: 150, 200, 300, 600 DPI.
 - *Format* – wybór formatu zapisu skanowanego dokumentu. Dostępne formaty to *PDF, GIF, TIFF*.
 - *Podajnik* – opcja dostępna tylko dla skanerów z podajnikiem, umożliwiająca wybór źródła, z którego pobierane będą dokumenty do skanowania.
 - *Duplex* – możliwość skanowania obustronnego. Dostępna tylko dla urządzeń, które posiadają tę opcję. Włączenie tej opcji dla urządzenia skanującego nie wyposażonego w funkcję skanowania obustronnego może zakłócić proces podziału dokumentów na strony.
 - „*Automatyczny rozmiar*” – automatyczne dopasowanie rozmiaru obrazu. W przypadku użytkowania modułu rozpoznającego tekst (OCR) oraz funkcji wykrywania pustych stron zaleca się, aby automatyczny rozmiar był włączony. Funkcja ta zapobiega skanowaniu obszarów wykraczających poza faktyczny rozmiar dokumentów, co może być przyczyną błędnego wykrywania stron pustych.
 - „*Automatyczna rozdzielczość*” – automatyczne dopasowanie rozdzielczości obrazu.

- Grupa opcji „[L] – **Rozpoznawanie dokumentów (OCR)**” – pozwala określić dodatkowe parametry dotyczące procesu skanowania, rozpoznawania oraz dekretacji dokumentów wprowadzanych do systemu.

- „Szukanie kodu kreskowego” – wykrywanie i zapisywanie we wprowadzanych do systemu dokumentach kodu kreskowego. Funkcja wykorzystywana w modelu skanowania opartym o wstępną dekretację dokumentów poprzez wydruk z aplikacji zestawu kodów kreskowych oraz ich naklejeniu na korespondencję przychodzącą do kancelarii. Szczegółowy opis jest dostępny w rozdziale poświęconym modelom procesu skanowania dokumentów w aplikacji Komornik SQL.
- „Włącz funkcję OCR (wymagana biblioteka LeadTools)” – pozwala włączyć opcję rozpoznawania skanowanych dokumentów. Do działania wymagana jest instalacja wersji aplikacji zawierająca biblioteki *LeadTools*, wykupienie licencji na moduł rozpoznawania *LeadTools* oraz przeprowadzenia procesu rejestracji aplikacji zgodnie z rozdziałem „Rejestracja modułu rozpoznającego tekst”.
- „OCR DPI” – pozwala określić rozdzielczość skanowania 300, 600 DPI wykorzystywaną podczas procesu rozpoznawania dokumentów. Wykorzystanie wyższej rozdzielczości rozpoznawania, niż zapisu dokumentów do bazy danych pozwala zwiększyć trafność rozpoznawania dokumentów przy równoczesnym zapewnieniu standardowego wzrostu rozmiaru bazy skanowanych dokumentów. Należy przy tym pamiętać, iż zastosowanie różnych rozdzielczości dla procesu rozpoznawania i skanowania może wydłużyć czas przeprowadzanych operacji ze względu na wymaganą konwersję dokumentów.
- „Automatyczna dekretacja dokumentów” – automatyczna dekretacja dokumentu do sprawy w oparciu o zdefiniowane stałe wyrażenia, zmienne oraz reguły przyporządkowań. Szczegółowy opis znajduje się w dalszej części instrukcji.
- „Automatyczne pozycjonowanie i wyszukiwanie pustych stron” – automatyczne wykrywanie dokumentów wprowadzanych do systemu odwrotnie. Wykorzystywane w celu poprawnego przeprowadzenia operacji rozpoznawania tekstu. Wykrycie strony wprowadzonej odwrotnie zapewnia również konwersję dokumentu na jej poprawną formę podczas zapisu do bazy danych. Wyszukiwanie pustych stron powoduje ich wykrycie oraz pominięcie podczas zapisu danych.
- „Logowanie wyników do pliku ocr.log” – pozwala na zapis informacji o procesie rozpoznawania skanowanych dokumentów w pliku ocr.log. Plik ten zostanie utworzony w katalogu C:\Currenda\Temp.
- „Weryfikacja rozpoznania sygnatury” – opcja odpowiedzialna za dodatkową weryfikację poprawności przypisania dokumentu wg nazw lub pełnych nazw stron w przypadku wykrycia sygnatury sprawy.
- „Tolerancja daty dokumentu” – opcja pozwalająca zabezpieczyć proces wykrywania zmiennej odpowiedzialnej za datę dokumentu przed jej błędnym wykonaniem. W przypadku, gdy została zdefiniowana zmienna oraz reguła odpowiedzialna za ustalenie daty rejestrowanej czynności, program sprawdza czy data ta różni się od daty skanowania o więcej niż określoną liczbę dni. Jeśli tak program przyjmuje jako datę czynności datę skanowania.

Filtry dla modułu OCR

Proces dekretacji korespondencji przychodzącej do kancelarii na podstawie tekstu wykrytego w

zeskanowanych obrazach opiera się o zdefiniowanie pewnych elementów specyficznych dla konkretnych typów dokumentów oraz reguł zachowań podejmowanych po ich odnalezieniu. Użytkownik aplikacji ma zatem możliwość zdefiniowania takich elementów i reguł indywidualnie dla kancelarii.

W aplikacji zaimplementowana została inicjalna baza zmiennych, reguł oraz predefiniowanych wyrażeń regularnych.

Dostęp do modułu konfiguracyjnego związanego z filtrami dla modułu OCR jest możliwy poprzez uruchomienie modułu „Konfiguracja” | menu „Słowniki” | „Filtry dla modułu OCR”.

Typy skanowanych dokumentów

Ze względu na istniejący w aplikacji Komornik SQL podział na różnego rodzaju typu skanowanych dokumentów analogicznie wprowadzono podział definiowania zmiennych oraz reguł dla każdego z nich. Zatem użytkownik może wydzielić zmienne i reguły dla:

- tytułów wykonawczych
- czynności egzekucyjnych
- pism KMO
- wniosków egzekucyjnych
- protokołów z czynności
- licytacji
- typu innego
- ruchomości
- nieruchomości
- opisu i oszacowania
- uczestników

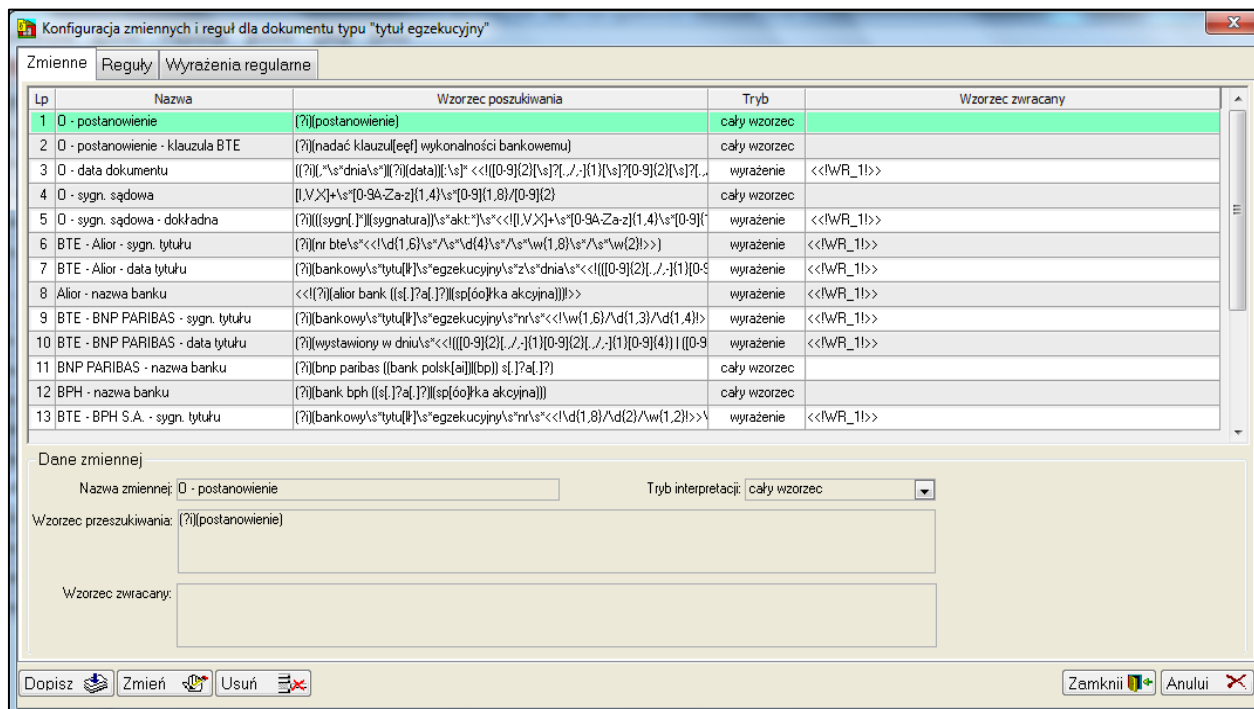
Typem, na który położono największy nacisk jest typ „czynności egzekucyjne”. Cały proces obsługi korespondencji przychodzącej opiera się właśnie o ten typ. Jednak do wykorzystania użytkownik ma całą ich resztę o ile zajdzie taka potrzeba.

Wybór odpowiedniego typu dokumentu na zakładce „Filtry dla modułu OCR” daje dostęp do okna przeznaczonego do definiowania szczegółowych informacji umożliwiających sprawne przeprowadzenia modułu dekretacji dokumentów. Wyświetlone okno zawiera trzy zakładki – dwie z nich („Zmienne” oraz „Reguły”) są specyficzne dla danego typu dokumentu, trzecia jest wspólna, gdyż zawiera informacje, które mogą być wykorzystywane w każdym z nich.

Zmienne

Moduł konfiguracyjny umożliwia definiowanie zmiennych będących w szczególności wyrażeniami regularnymi, których obsługę zaimplementowano zgodnie ze standardem wykorzystywanym na platformie *Microsoft.NET*. Wyrażenia regularne są wzorcami w postaci ciągów znaków opisującymi tekst o określonym schemacie. Umożliwiają one określenie stałych wyrażeń oraz zmiennych, podobnych do siebie tekstów, które będą następnie wyszukiwane w przetworzonym dokumencie. Samo opracowywanie wyrażeń regularnych dla osoby niebędącej osobą powiązaną z dziedziną informatyki

może się wydawać dość trudne, jednak takie podejście zapewnia maksymalne wykorzystanie ich potencjału. Opis podstawowych elementów wykorzystywanych podczas tworzenia wyrażeń regularnych znajduje się również w niniejszej instrukcji.



Do dyspozycji użytkownika, na zakładce zmienne są następujące elementy:

- lista zapisanych zmiennych podlegających dekretacji w ramach danego typu
- szczegóły zaznaczonej na liście zmiennej

Lista zapisanych zmiennych jest listą nieedycyjną. Zawiera jedynie najważniejsze informacje o danej zmiennej. Zmiana aktualnego wyboru na liście powoduje odświeżenie szczegółów zmiennej w części okna znajdującej się grupie „Dane zmiennej”.

Użytkownik ma dostęp do następujących funkcji przeznaczonych do jej modyfikacji:

- „Dopisz” – dopisanie do listy nowej pustej zmiennej.
- „Zmień” – edycja szczegółowych danych istniejącej zmiennej. Ze względu na fakt, iż dana zmienna może już być użyta w regule nie jest możliwa edycja nazwy już istniejącej zmiennej. Wszystkie pozostałe dane szczegółowe podlegają edycji.
- „Usuń” – usunięcie pozycji z listy. Nie jest możliwe usunięcie zmiennej wykorzystanej już w regule. Próba przeprowadzenia takiej operacji zakończy się wyświetleniem odpowiedniego komunikatu.

Dla każdej ze zmiennych istnieje możliwość zdefiniowania jej szczegółów:

- „Nazwa zmiennej” – edycja tego pola możliwa jest wyłącznie podczas tworzenia nowej zmiennej. Pole to jest wykorzystywane podczas tworzenia reguł zachowań w przypadku ich wykrycia. Nie

jest dopuszczalne umieszczanie w nazwie zmiennej znaków „(, ,, „)”, „<<!” , „!>>”.

- „Tryb interpretacji” – dopuszczalne są dwie wartości „wyrażenie” oraz „cały wzorzec”. Tryb „wyrażenie” powoduje, iż wykrycie zmiennej spowoduje podstawienie za nią konkretnej wartości będącej albo stałym ciągiem znaków albo elementem wzorca przeszukiwania. Elementy z wzorca przeszukiwania w takim przypadku muszą być ograniczone za pomocą znaczników „<<!” oraz „!>>”. Tryb „cały wzorzec” oznacza, iż w przypadku zdefiniowania zmiennej w regule zamiast tej zmiennej zostanie wstawiona cała wykryty tekst.
- „Wzorzec przeszukiwania” – wzorzec oparty o składnię wyrażeń regularnych wykorzystywanej na platformie *Microsoft.NET*. Wprowadzane wartości mogą być stałymi ciągami tekstu lub schematami określającymi tekst. Opis sposobu definiowania wyrażeń regularnych zostanie opisany w dalszej części instrukcji. Dopuszczalne jest wskazywanie elementów wzorca przeszukiwania, które będą zwracane zamiast zmiennej w sposób opisany w punkcie powyżej (poprzez oznaczenia go za pomocą odpowiednich znaczników).
- „Wzorzec zwracany” – wzorzec zawierający informację o tym jaka wartość zostanie podstawiona w przypadku wykrycia zmiennej.

Przykładowe definicje zmiennych wraz ze szczegółowym opisem znajdują się w rozdziale poświęconym wyrażeniom regularnym.

Wyrażenia regularne w zmiennych

Jak zostało to opisane wcześniej, proces wykrywania elementów tekstu oparty został o proces dopasowywania za pomocą wyrażeń regularnych.

Definiowanie zmiennych odbywa się poprzez edycję pola „Wzorzec przeszukiwania” w danych zmiennej. Wprowadzanie wzorca przeszukiwania umożliwia również wskazanie ograniczonego elementu całego wyrażenia, który będzie mógł zostać użyty do budowy „Wzorca zwracanego” przy równoczesnym zapewnieniu sprawdzenia całości. Do tego celu służą znaczniki „<<!” oraz „!>>”.

Przykład.

Nazwa zmiennej: **Sygnatura akt**

Tryb interpretacji: **Wyrażenie**

Wzorzec przeszukiwania:

`(?i)((sygn[.]*)(sygnatura))\s*akt:*\s*<<![IVX]+\s*[0-9A-Za-z]{1,4}\s*[0-9]{1,8}/[0-9]{2}!>>`

Wzorzec zwracany: `<<!WR_1!>>`

Tekst, w którym wyszukiwane jest dopasowanie do wzorca:

Warszawa, dnia 12 lipca 2012 roku
Sygnatura akt: IVc 1002/12
Pismo wierzyciela
Wnoszę o zajęcie wierzytelności w Urzędzie Skarbowym...

Rozpoczęcie definicji wzorca o znaków `(?i)(...)` oznacza, iż cały wzorzec znajdujący się w `(...)` będzie sprawdzany niezależnie od wielkości liter. Czyli dopuszczalne będzie wykrycie np. słowa *sygnatura* jak również słowa *Sygnatura*, *SYGNTURA*, *sygn.*, itp. Element całego wzorca `((sygn[.]*)|(sygnatura))` oznacza, iż poszukujemy elementu, w którym znajdzie się albo wzorzec `(sygn[.]*)` albo `(sygnatura)`, w dalszym ciągu niezależnie od wielkości liter. Takie zachowanie wzorca jest możliwe do wykonania dzięki użyciu znaku `|`. Przyporządkowanie tekstu do schematu `(sygn[.]*)` oznacza, iż poszukujemy stałego tekstu (ze względu na umiejscowienie go pomiędzy dwoma znakami nawiasów) *sygn* oraz opcjonalnie znaku `,`. Opcjonalnie ze względu na fakt, iż znak ten został ujęty pomiędzy nawiasami `[` oraz `]`, oznaczającymi dopuszczalny zestaw pojedynczego znaku, po którym wystąpił znak `*`. Znak ten oznacza, iż element z grupy wystąpi zero lub więcej razy. Dla pojedynczego znaku dopuszczalne jest również pominięcie nawiasów kwadratowych, więc równoznaczne z tym zapisem byłoby wprowadzenie `.`. Dodatkowo użyty został element `\s*` - `\s` oznacza w wyrażeniach regularnych wystąpienie znaku białego, np. spacji. Ponowne użycie znaku `*` oznacza, iż pomiędzy poszczególnymi elementami tego elementu dopuszczamy znak spacji zero lub więcej razy.

Podsumowując tę część wzorca - zastosowano tutaj następujące elementy definiowania wyrażeń regularnych:

- znaki `"` oraz `,` – definiowanie grupy kolejnych znaków, np. stałego tekstu
- wzorzec `(?i)(wewnętrzne wyrażenie regularne)` – zapewnienie, iż wyrażenie regularne zawarte pomiędzy nawiasami wykrywane będzie niezależnie od wielkości liter w nim wykrytych
- znaki `"` oraz `,` – definiowanie grupy dopuszczalnych znaków, np. `[abc]` będzie wyszukiwało elementy zawierające albo literę `"a"` albo `"b"` albo `"c"`. Możliwe jest również definiowanie zakresów znaków, np. `[a-z]` będzie oznaczało wszystkie małe litery, `[A-Z]` wszystkie wielkie litery, a `[0-9]` wszystkie cyfry.
- znak `*` – definiowanie, iż dana grupa (ujęta w nawiasy zwykłe lub kwadratowe) wystąpi 0 lub więcej razy w danym wzorcu
- znak `|` – definiowanie, iż w wyszukiwanym wzorcu dopuszczalna jest albo jedna grupa albo inna
- znaki `\s` – definiowanie przyporządkowania do znaków białych (np. spacji)

Znając już podstawy definiowania wyrażeń regularnych można przejść do dalszej części przykładowego wyrażenia, a mianowicie:

```
<<![IVX]+\s*[0-9A-Za-z]{1,4}\s*[0-9]{1,8}/[0-9]{2}!>>
```

Ten element został już ujęty w znaczniki `<<!` oraz `!>>` oznaczające, iż podczas interpretacji chcielibyśmy mieć możliwość odczytania dokładnie tej wartości, która się pomiędzy nimi znajduje pomimo, że całe wyrażenie jest dłuższe. Na potrzeby dopasowywania do wzorców znaczniki te zostaną usunięte, jednak program zapamięta co pomiędzy nimi zostało wykryte.

W tym wyrażeniu doszło kilka nowych elementów definiujących wyrażenia regularne:

- znak `+` – jest wykorzystywany analogicznie do znaku `*`, jednak oznacza jedno lub więcej wystąpień danej grupy lub pojedynczego znaku
- znaki `[0-9A-Za-Z]` – przedstawia sposób łączenia różnych zakresów znaków, czyli w tym

przypadku dopasowanie będzie poprawne dla dowolnej cyfry, małej lub wielkiej litery.

- znaki „{1,8}” – definiowanie liczebności wystąpień danego znaku lub grupy znaków, w tym przypadku dopuszczalne jest wystąpienie od jednego do ośmiu znaków, a w przypadku {2} oznacza, iż dany znak powinien wystąpić dokładnie dwa razy

Zatem zdefiniowane wyrażenie oznacza, iż przeszukujemy tekst pod kątem wzorca o następującym schemacie:

- znaki „l”, „V” lub „X” wystąpią jeden lub więcej razy, np. „lV”, „lX”, itp.
- może się pojawić spacja (ale nie jest to konieczne)
- wystąpi dowolna cyfra, mała lub wielka litera od jednego do czterech razy
- ponownie może pojawić się spacja
- wystąpi liczba składająca się z od jednej do ośmiu cyfr
- znak „/”
- ponownie wystąpi liczba jednak w tym przypadku dopuszczalne są wyłącznie liczby dwucyfrowe

Wszystkie elementy oznaczone we wzorcu przeszukiwania za pomocą znaczników „<<!” oraz „!>>” mogą zostać użyte do budowy elementów wartości zwracanej przez zmienną. Można ich użyć poprzez umieszczenie odwołania do elementu wzorca za pomocą znaku „<<!WR_X!>>”, gdzie X oznacza kolejny numer oznaczonego wyrażenia w pełnym wzorcu przeszukiwania. Zatem w opisywanym przypadku możliwe jest użycie wyłącznie elementu oznaczonego za pomocą nazwy <<!WR_1!>>, gdyż tylko jeden element został oznaczony w ten sposób. Jednak nic nie stoi na przeszkodzie, aby zmienne były przeznaczone jedynie do sprawdzania czy tekst o określonym schemacie został wykryty, a użytkownik chciałby zwracać stały określony tekst. Takie zachowanie jest wiele razy wykorzystywane w inicjalnej bazie zmiennych.

Podczas budowy wzorca przeszukiwania program automatycznie zlicza wykryte elementy oznaczone, a wciśnięcie prawego przycisku myszy na polu „Wzorzec zwracany” umożliwi wstawienie jednego z wykrytych elementów na podstawie budowanego dynamicznie menu.

W oparciu o przedstawione powyżej informacje oraz dane z przykładu można jasno stwierdzić, iż aplikacja podczas przetwarzania tekstu ustali, iż zgodny z wzorcem jest element:

Sygnatura akt: IVc 1002/12

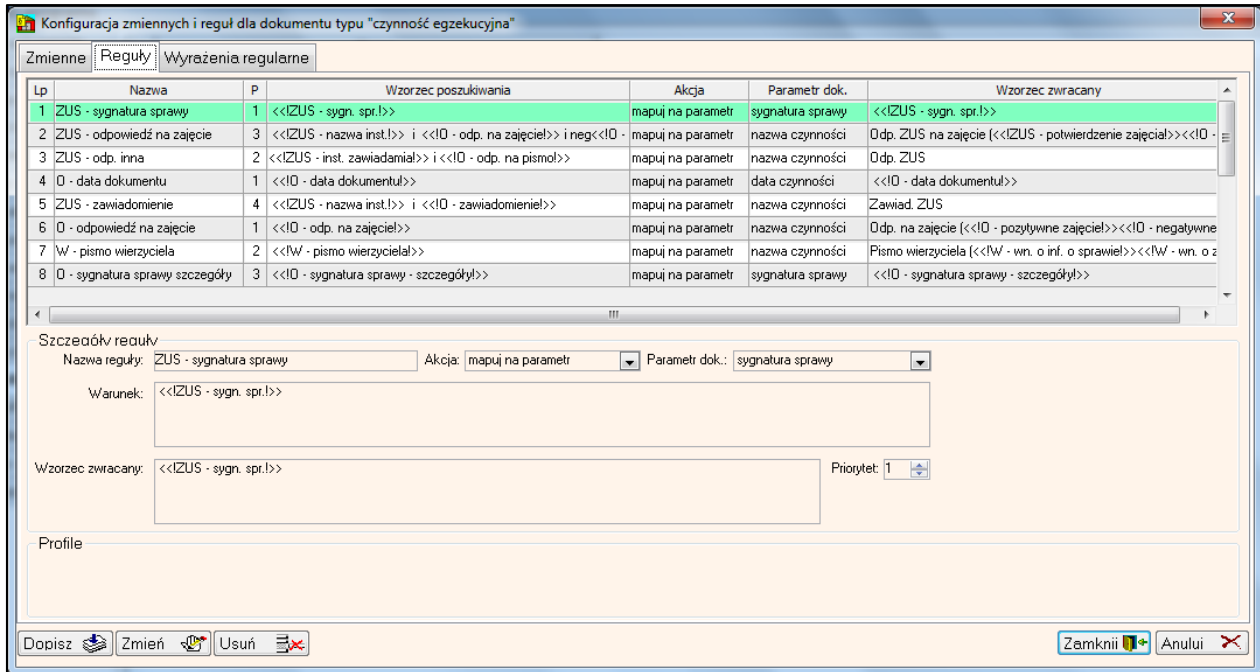
Ze względu na fakt, iż „Wartość zwracana” jest ograniczona do <<!WR_1!>> to za tą zmienną zostanie podstawiony wyłącznie wartość IVc 1002/12, która z punktu widzenia użytkownika jest najbardziej interesująca.

Dopuszczalne jest również używanie w polu „Wzorzec zwracany” innych stałych elementów tekstu w połączeniu z wykrytymi w nich wyrażeniami regularnymi.

Na stronach firmy *Microsoft* pod adresem <http://msdn.microsoft.com/en-us/library/az24scfc.aspx> można znaleźć bardzo bogatą i szczegółową dokumentację dotyczącą tworzenia wyrażeń regularnych, więc serdecznie zapraszamy do lektury.

Reguły

Aplikacja Komornik SQL do poprawnej dekretacji dokumentów wymaga zdefiniowania reguł zachowań umożliwiających wskazanie jej co ma zostać wykonane po prawidłowym wykryciu zmiennej lub zmiennych. Do tego celu przeznaczona została zakładka „Reguły” przypominająca swoim wyglądem zakładkę „Zmienne”. Okno w tym przypadku analogicznie zostało podzielone na listę zapisanych w bazie reguł oraz szczegóły aktualnie zaznaczonej na liście pozycji. Dostępne są również identyczne opcje związane z dodawaniem, edycją oraz usuwaniem reguł.



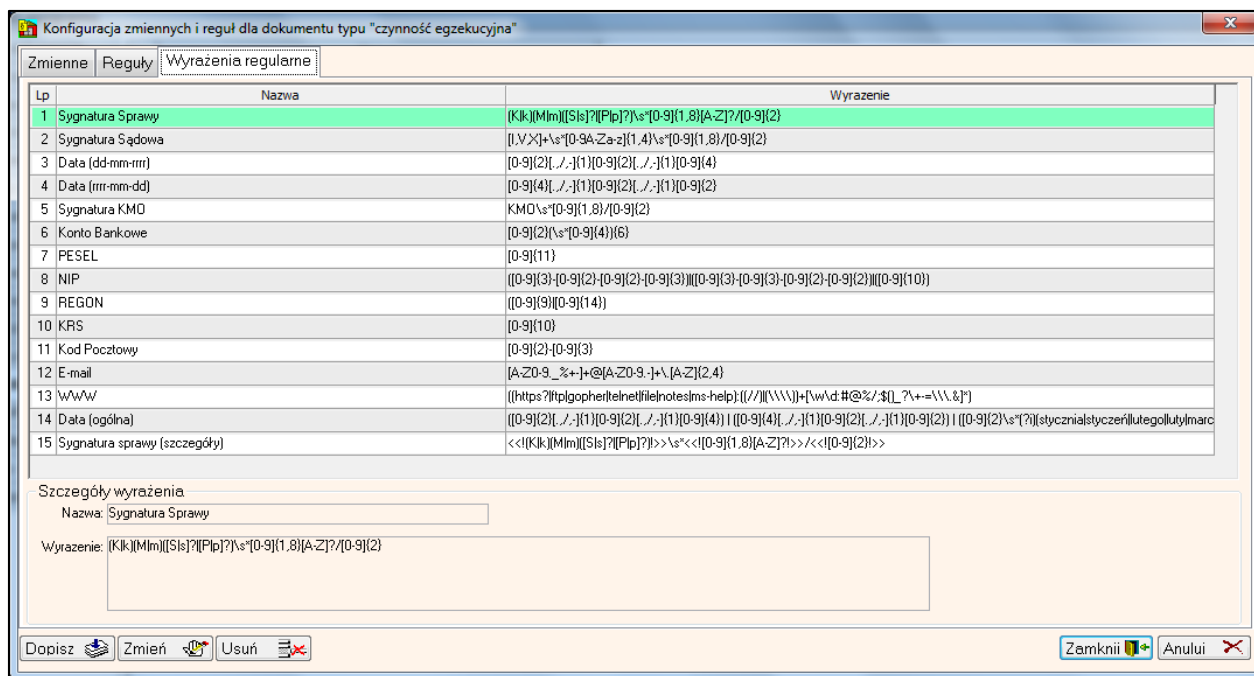
Dla każdej z reguł możliwe jest wprowadzanie zmian dotyczących jej szczegółów:

- „Nazwa reguły” – edycja tego pola możliwa jest wyłącznie podczas tworzenia nowej reguły. Pole wykorzystywane wyłącznie w celach informacyjnych do odróżnienia jednej reguły od innej.
- „Akcja” – dopuszczalne są trzy wartości „mapuj na parametr”, „powiąż sygnaturę” oraz „zmień typ dokumentu”. Opcja „mapuj na parametr” przeznaczona jest do wskazania, iż w przypadku wykrycia warunku wykonania reguły (pole „Warunek”) za zdefiniowany parametr zostanie wstawione wyrażenie zdefiniowane w polu „Wzorzec zwracany”. Kolejna opcja „powiąż sygnaturę” przeznaczona jest do wskazania elementu zmiennej, która zostanie użyta do powiązania z konkretną sprawą. Analogiczną funkcjonalnością jest wybranie akcji „mapuj na parametr” z parametrem „sygnatura sprawy (string)”. Kolejna funkcja „zmień typ dokumentu” umożliwia wskazanie, iż dana reguła będzie podstawą do ustalenia konkretnego typu dokumentu w przypadku skanowania dokumentów typu „inny”. Jeśli w tym trybie wykryta będzie poprawnie reguła z taką akcją to nastąpi automatyczne przyporządkowanie do danego typu, a następnie zostanie przeprowadzona dekretacja dokumentu na podstawie reguł dla niego zdefiniowanych.
- „Parametr dok.” – parametr, do którego zostanie podstawiona wartość znajdująca się w polu „Wzorzec zwracany” (np. nazwa czynności).

- „*Warunek*” – pole to oznacza warunek wykonania reguły. Wyrażenie to może się składać z wielu zmiennych oraz spójników logicznych. Do budowy warunku można używać menu kontekstowego (dostępnego pod prawym przyciskiem myszy) umożliwiającego wybór z listy zdefiniowanych zmiennych oraz obsługiwanych spójników logicznych. Program obsługuje budowę warunków wystąpienia w oparciu o następujące spójniki logiczne:
 - spójnik *i* – oznacza, iż muszą być spełnione (wykryte) dwie sąsiadujące do spójnika zmiennej
 - spójnik *lub* – oznacza, iż musi być spełniona (wykryta) jedna lub druga z sąsiadujących do spójnika zmiennych
 - spójnik *neg* – oznacza, iż aby spełniony został ten element warunku nie może być wykryta zmienna następująca po spójniku
- „*Wzorzec zwracany*” – wartość, która zostanie podstawiona za wskazany w regule parametr. Można zdefiniować stały ciąg lub zmienne. Zmienne mogą być dowolnymi zmiennymi zdefiniowanymi na liście zmiennych. Menu kontekstowe w tym przypadku również jest wykorzystywane do podpowiadania nazw zmiennych, które mogą zostać wstawione do budowy wartości zwracanej.
- „*Priorytet*” – pole to jest przeznaczone do wskazania, która reguła zostanie zastosowana w przypadku, gdy spełnione zostaną dwie o tych samych parametrach. W takim przypadku zostanie wykonana tylko jedna, posiadająca wpisaną największą liczbę w polu. Mechanizm ten może zostać wykorzystany do tworzenia reguł coraz bardziej uszczegóławianych. Reguła najbardziej ogólna posiada najmniejszy priorytet. Czym bardziej szczegółowa jest reguła tym powinna posiadać zdefiniowany wyższy priorytet.
- „*Profile*” – pole przeznaczone do przypisania reguły do określonego profilu (nazwy skanowanego dokumentu), w którym może być przeprowadzany proces skanowania dokumentów. Proces skanowania w danym profilu umożliwia zawężenie wykrywania reguł wyłącznie do tych, które zostały do profilu przypisane. Mechanizm taki może zostać wykorzystany do ograniczenia sprawdzanych reguł dla konkretnych, specyficznych dokumentów, dla których nie jest wymagane sprawdzanie wszystkich reguł, a użytkownicy zapewnią wstępny podział korespondencji tradycyjnej. Lista dostępnych profili jest możliwa do modyfikacji w module „Konfiguracja” | menu „Słowniki” | „Nazwy skanowanych dokumentów”.

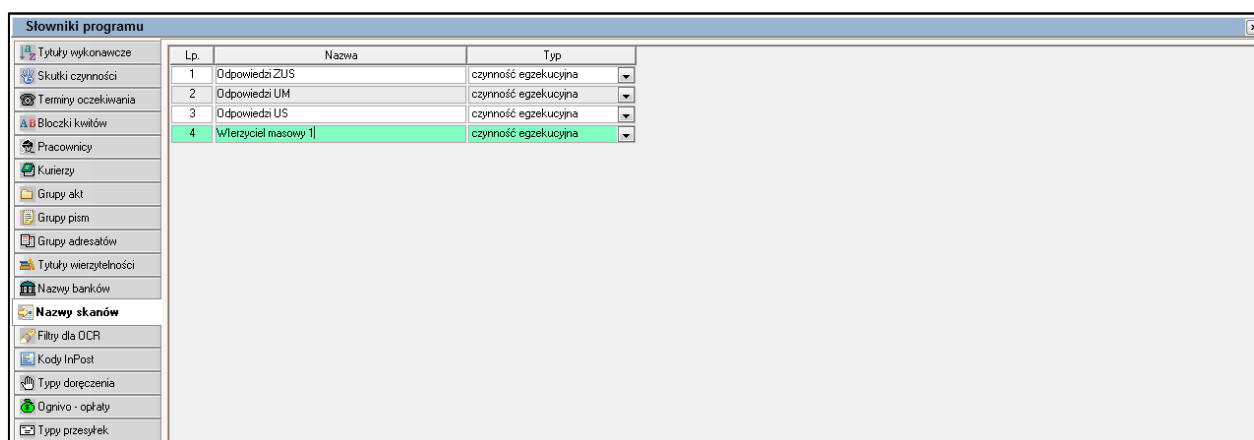
Predefiniowane wyrażenia regularne

Zakładka ta jest przeznaczona do przygotowania listy najczęściej wykorzystywanych wyrażeń regularnych podczas budowy wzorców przeszukiwania w zmiennych. Przygotowane wyrażenia regularne tj. data, sygnatura sprawy egzekucyjnej, adres e-mail, itp. będą mogły być bezpośrednio wybrane podczas budowy zmiennej z menu kontekstowego po wciśnięciu prawego przycisku myszy w polu „*Wzorzec przeszukiwania*”.



Nazwy skanowanych dokumentów

W aplikacji Komornik SQL istnieje możliwość skonfigurowania nazw dla skanowanych dokumentów. Opcja dostępna jest w module „Konfiguracja” | menu „Słowniki” | „Nazwy skanowanych dokumentów”. Program pozwala na dodanie pozycji, na nadanie im przykładowej nazwy dla skanowanego dokumentu oraz określeniu, z jakim typem dokumentu nazwa ma być powiązana. Utworzone przez użytkownika nazwy mogą być wykorzystane podczas skanowania dokumentów.



Nazwy skanowanych dokumentów wykorzystywane są również do definiowania tzw. profili dekretacji. Powyżej został opisany proces przyporządkowywania reguły do wybranych profili. Nazwy skanowanych dokumentów pełnią funkcję właśnie takich profili. Wprowadzenie nazwy na liście oraz przyporządkowanie reguły do tego profilu umożliwia przeprowadzenia procesu dekretacji na ograniczonej liście reguły przypisanych do profilu.

Drukowanie etykiet

W jednym z przygotowanych modeli procesu skanowania i dekretacji dokumentów aplikacja Komornik SQL wykorzystuje etykiety drukowane z poziomu aplikacji, którymi korespondencja przychodząca jest oklejana. Szczegółowy opis procesu skanowania dokumentów z wykorzystaniem kodów kreskowych zostanie opisany w dalszej części instrukcji. Na tym etapie opisany zostanie wyłącznie sposób przygotowywania etykiet zawierających kody kreskowe.

Konfiguracja wydruków etykiet oraz sam wydruk dostępny jest w module „Konfiguracja” | menu „Wydruki” | „Drukowanie etykiet”. Dostępne w oknie konfiguracyjnym opcje oznaczają kolejno:

- Grupa opcji „[L] – **Ustawienie kodów kreskowych**”:
 - „*Pojedyncze*” – wydruk pojedynczej etykiety poprzez podanie jej numeru
 - „*Seryjny*” – wydruk seryjny etykiet
 - „*Automatyczna numeracja etykiet*” – automatyczne nadawania numerów etykiet
 - „*Rodzaj kod*” – możliwość wybrania kodów dla pism i zwrotek
 - „*Ilość kodów*” – ilość drukowanych kodów kreskowych
 - „*Ostatni numer*” – numer ostatniego kodu
 - „*Numer początkowy*” – numer pierwszego kodu na wydruku (tylko przy odznaczonej, opcji „*automatyczna numeracja etykiet*”)
- Grupa opcji „[L] - **Ustawienia strony**”:
 - „*Margines górny*” – pozwala określić górny margines strony
 - „*Margines lewy*” – pozwala określić lewy margines strony
- Grupa opcji „[L] - **Ustawienia etykiet**”:
 - *Wysokość* – wysokość drukowanej etykiety w mm
 - *Szerokość* – szerokość drukowanej etykiety w mm
 - *Liczba kolumn* – Liczba kolumn na stronie
 - *Liczba wierszy* – Liczba wierszy na stronie

Modele skanowania oraz dekretacji dokumentów

Aplikacja Komornik SQL w ostatnich czasach została przygotowana do obsługi procesu skanowania oraz dekretacji dokumentów zgodnie z trzema zaprezentowanymi poniżej modelami. Zostały one opracowane z myślą o kancelariach, które chcą ten proces usprawnić. Mając na względzie różne specyfiki pracy kancelarii, różne modele urzędzeń skanujących oraz różne podejścia użytkowników aplikacji zaimplementowane zostały właśnie trzy różne podejścia. Każde z nich posiada swoje zalety i wady oraz każda w różny sposób wpływa na przyspieszenie pracy.

Model 1. Skanowanie seryjne z ręczną dekretacją dokumentów

Model skanowania seryjnego z ręczną dekretacją dokumentów umożliwia skanowanie seryjne korespondencji przychodzącej lub innego typu dokumentów. Początkowy etap tego procesu nie wymaga przypisywania dokumentów do żadnej ze spraw. Polega on wyłącznie na zeskanowaniu wszystkich dokumentów, które później zostają przyporządkowywane do poszczególnych spraw lub typów.

Proces ten znacząco przyspiesza sam proces zamiany dokumentacji tradycyjna na jej odpowiedniki w wersji elektronicznej, jednak wymaga sporych nakładów pracy dla zapewnienia odpowiednich przyporządkowań.

Model 2. Skanowanie seryjne z dekretacją dokumentów na podstawie kodów kreskowych

Model wykorzystujący kody kreskowe ułatwia proces przyporządkowań dokumentów w stosunku do poprzednio opisywanego modelu. Skanowanie dokumentów zostaje poprzedzone oklejeniem ich etykietami z kodami kreskowymi, które podczas procesu skanowania zostaną automatycznie rozpoznane, a wraz z obrazem zostanie zapisana wartość znakowa umożliwiająca przyporządkowanie. Następnie osoby, do których trafiają dokumenty rejestrują odpowiednie wpisy w sprawach oraz podłączają elektroniczne odpowiedniki poprzez wykorzystanie kodu kreskowego naklejonego na dokument. Użycie czytnika kodów kreskowych umożliwia automatyczne odnalezienie dokumentu w bazie skanowanych dokumentów bez konieczności żmudnego przeszukiwania listy wszystkich zeskanowanych obrazów.

Model 3. Skanowanie seryjne z wykrywaniem znaków oraz automatyczną dekretacją dokumentów

Model wykorzystuje zaawansowane techniki rozpoznawania tekstów w obrazach oraz automatycznej dekretacji dokumentów w oparciu o zdefiniowane w bazie danych zmienne oraz reguły. Zamiana zeskanowanego obrazu w tekst umożliwia poddanie go dalszej obróbce w celu przeprowadzenia jego interpretacji – odszukanie pewnych specyficznych dla danego dokumentu wyrażen na podstawie, których można określić informacje dotyczące danego dokumentu.

Model skanowania	Zalety	Wady
Skanowanie seryjne z ręczną dekreacją dokumentów	<ul style="list-style-type: none"> • możliwość seryjnego skanowania wielu dokumentów • zapewnienie odpowiedniego podziału pracy (skanowanie, przyporządkowywanie) • przyśpieszenie pracy w stosunku do standardowego podejścia skanowania dokumentów 	<ul style="list-style-type: none"> • dodatkowa praca związana z podziałem dokumentów na jedno lub wielokartkowe • proces dekreacji dokumentów wymaga żmudnego wyszukiwania dokumentów z listy
Skanowanie seryjne z dekreacją dokumentów na podstawie kodów kreskowych	<ul style="list-style-type: none"> • możliwość seryjnego skanowania wielu dokumentów • zapewnienie odpowiedniego podziału pracy (skanowanie, przyporządkowywanie) • przyśpieszenie pracy w stosunku do standardowego podejścia skanowania dokumentów 	<ul style="list-style-type: none"> • dodatkowa praca związana z podziałem dokumentów na jedno lub wielokartkowe • znaczące ułatwienie następuje dopiero po wyposażeniu stanowisk pracy użytkowników w czytniki kodów kreskowych
Skanowanie seryjne z wykrywaniem znaków oraz automatyczną dekreacją dokumentów	<ul style="list-style-type: none"> • możliwość seryjnego skanowania wielu dokumentów • automatyczne zakładanie odpowiednich rekordów związanych ze skanowanymi dokumentami w oparciu o zmienne i reguły • zapisywanie wykrytego w dokumentach tekstu w bazie skanowanych dokumentów • pełne raporty z rozpoznawania oraz dekreacji dokumentów • znaczące przyśpieszenie pracy w stosunku do standardowego podejścia skanowania dokumentów 	<ul style="list-style-type: none"> • dodatkowa praca związana z podziałem dokumentów na jedno lub wielokartkowe • rozwiązanie dodatkowo płatne ze względu na wykorzystanie zewnętrznego modułu

Proces skanowania dokumentów

Lista skanowanych dokumentów

Okno zawierające listę zeskanowanych dokumentów zawiera szereg nowych przycisków odpowiadających za poszczególne funkcje związane ze skanowaniem oraz dekreacją dokumentów.

W związku z dodatkowymi funkcjonalnościami wprowadzonymi w aplikacji uległa również zmiana zawartość kolumn listy skanowanych dokumentów.

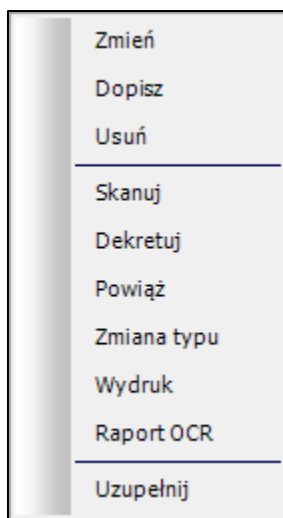
Lp	V	Typ	P	R	Data skanowania	Data dokumentu	Sprawa / Opis	Plik	Strony	Nazwa dokumentu	Kod
1	<input type="checkbox"/>	C	+		08-08-2012	02-08-2012	KMS 76/12	TIFF	2	Czynność egzekucyjna	
2	<input type="checkbox"/>	C	-		08-08-2012	01-01-1900		TIFF	2		
3	<input type="checkbox"/>	C	-		08-08-2012	01-01-1900		TIFF	2		
4	<input type="checkbox"/>	C	+	D	07-08-2012	27-07-2012	KMS 76/12	TIFF	2	Zawład. COI	
5	<input type="checkbox"/>	C	+		07-08-2012	19-06-2012	KM 3561/11	TIFF	2	Umorzenie postępowania - egzekucja bezskuteczna	

Znaczenie poszczególnych kolumn opisuje poniższa lista:

- „Lp” – liczba porządkowa.
- „V” – kolumna przeznaczona do zaznaczania pozycji dla operacji seryjnych.
- „P” – kolumna zawiera informacje o przeprowadzonym procesie dekretacji: „+” – poprawne przyporządkowanie (wykryto sygnaturę sprawy oraz zgadzają się dane stron postępowania w przypadku, gdy włączona została opcja dodatkowej walidacji poprawności danych), „-” nieudane przyporządkowanie, „?” – niepewne przyporządkowanie (wykryto sygnaturę sprawę, jednak nie zgadzają się dane stron postępowania w przypadku, gdy włączona była walidacja poprawności danych).
- „R” – kolumna zawiera informacje o przeprowadzonym procesie rozpoznawania: „R” – tekst został rozpoznany, „D” – tekst został rozpoznany, a dokument zadekretowany, „O” – oznacza, że dokument oczekuje na dekretację ręczną, „N” – oznacza, iż dokument niezadekretowany, jeżeli kolumna ta jest pusta oznacza to, iż nie przeprowadzono procesu rozpoznawania np. przy ręcznym dodaniu dokumentu.
- „Data skanowania” – data skanowania dokumentu.
- „Data dokumentu” – data dokumentu.
- „Sprawa/opis” – sygnatura sprawy egzekucyjnej, do której przyporządkowany jest dokument
- „Plik” – format plików (np. TIFF, GIF, PDF). Ze względu na fakt, iż kolejne strony dokumentów mogą być podłączane w różnych formatach wyświetlana tutaj informacja zawiera format wyłącznie pierwszej strony.
- „Strony” – liczba stron dokumentu. Uwaga. W przypadku dokumentów PDF należy pamiętać, iż każda strona może zawierać więcej stron, więc element ten można traktować jako liczbę załączników dla danego dokumentu.
- „Nazwa dokumentu” – nazwa wprowadzonego do systemu dokumentu.
- „Kod” – wartość kodu kreskowego wykrytego w dokumencie.

Poszczególne przyciski dostępne w lewym bocznym menu odpowiadają za wykonywanie następujących operacji:

- „Dopisz” – umożliwia dodanie nowego skanowanego dokumentu poprzez kreator umożliwiający wybranie typu skanowanego dokumentu, sprawy, w której zostanie dokument załączony oraz wybranie lub utworzenie nowego wpisu ściśle związanego z wprowadzonym dokumentem (np. założenie nowej czynności egzekucyjnej lub wybór już istniejącej).
- „Zmień” – zmiana istniejącego zeskanowanego dokumentu (np. dodanie kolejnej strony)
- „Dane” – wyświetlenie danych zeskanowanego dokumentu (np. informacji o czynności egzekucyjnej czy tytule wykonawczym).
- „Usuń” – usunięcie zapisanego w bazie dokumentu. Uwaga. Usunięcie dokumentu nie powoduje usunięcia wpisu, z którym jest powiązany (np. usunięcia czynności egzekucyjnej).
- „Skanuj” – uruchomienie funkcji skanowania seryjnego.
- „Dekretuj” – uruchomienie procesu dekretacji zaznaczonych dokumentów w oparciu o zdefiniowane w module „Konfiguracja” | menu „Słowniki” | „Filtry dla OCR” zmienne i reguły. Proces ten będzie poprawnie funkcjonował wyłącznie dla pozycji z listy, które posiadają w kolumnie "R" wartość „R” oznaczającą, iż dokument został rozpoznany. Pozostałe pozycje zostaną pominięte.
- „Wydruk” – przygotowanie podglądu wydruku listy dokumentów zgodnej z filtrowaniem na wyświetlonej w oknie liście
- „Seria” – wydruk seryjny zaznaczonych na liście dokumentów
- „Czyść” – czyszczenie bazy skanowanych dokumentów z dokumentów nie powiązanych ze sprawami w repertorium.
- „Zamknij” - zamknięcie okna
- „Anuluj” – zamknięcie okna



Dodatkowo wciśnięcie prawego przycisku myszy na liście skanowanych dokumentów umożliwia wybór opcji z menu kontekstowego. Część z nich jest odpowiednikami funkcji z lewego bocznego menu. Pozostałe oznaczają:

- „Powiąż” – ręczne powiązanie zapisanego w bazie dokumentu z konkretną sprawą egzekucyjną oraz miejscem w strukturze aplikacji (np. zapisaną lub nową czynnością egzekucyjną czy tytułem wykonawczym).

- „Zmiana typu” – umożliwi zmianę typu zapisanego w bazie dokumentu.
- „Raport OCR” – wyświetlenie okna zawierającego szczegółowe informacje na temat procesu rozpoznawania oraz dekretacji dokumentu (rozpoznany tekst, wykryte zmienne, reguły, przyporządkowania).
- „Uzpełnij” – w przypadku, gdy nazwa zeskanowanego dokumentu jest pusta, program ją uzupełni w oparciu o powiązane dokumenty.

Skanowanie seryjne

Moduł skanowanych dokumentów jest dostępny w module „Moduły dodatkowe” | menu „Skanowane dokumenty” | przycisk „Skanuj”. Został rozszerzony o obsługę procesu skanowania seryjnego wraz z implementacją trzech modeli dekretacji dokumentów. Dzięki takiemu rozwiązaniu użytkownik posiada możliwość seryjnego skanowania dokumentów, ręcznej lub automatycznej dekretacji oraz weryfikacji poprawności zapisu informacji dotyczących samego procesu przyporządkowywania.

The screenshot shows a dialog box titled "Skanowanie seryjne" with the following fields and options:

- Źródło danych:** Radio buttons for "Plik" (selected) and "Skaner".
- Data dokumentu:** Dropdown menu set to "08-08-2012".
- Data skanowania:** Dropdown menu set to "08-08-2012".
- Typ dokumentu:** Dropdown menu set to "czynność egzekucyjna".
- Propozycja nazwy:** Dropdown menu.
- Nazwa:** Text input field.
- Options:**
 - Rozpoznawanie tekstu - OCR
 - Każda kartka odrębnym dokumentem
 - Wyświetl dane dokumentu
 - Zadana sygnatura sprawy
- Ile kartek dokumentu:** Dropdown menu set to "1".
- Buttons:** "Skanuj" and "Anuluj".

Za uruchomienie procesu skanowania odpowiedzialny jest przycisk „Skanuj”. Po wciśnięciu przycisku pojawi się okno skanowania seryjnego. W oknie tym dostępne są następujące parametry skanowania:

- „Źródło danych” – wskazanie sposobu wprowadzania dokumentów do bazy danych. Dostępne są dwie opcje (parametr jest domyślnie ustawiany w sposób zgodny z konfiguracją aplikacji):
 - „Plik” – źródłem podłączanego dokumentu będzie istniejący plik na dysku twardym.
 - „Skaner” – źródłem podłączanego dokumentu będzie urządzenie skanujące. Pobranie

obrazu odbędzie się przy użyciu sterownika TWAIN. Podczas skanowania wykorzystane zostaną ustawienia konfiguracyjne wskazane w module „Konfiguracja” | menu „Preferencje” | „Konfiguracja skanowania”

- „Data dokumentu” – pole przeznaczone do przechowywania informacji na temat daty dokumentu. Data ta jest datą domyślną. Jeśli proces skanowania będzie uwzględniał automatyczne rozpoznawanie oraz dekretację dokumentu, a w trakcie tego procesu zostanie spełniona reguła z akcją „mapuj na parametr” oraz parametrem „data czynności” to zostanie ona zastąpiona datą wykrytą w dokumencie.
- „Data skanowania” - pole przeznaczone do odnotowania informacji na temat daty przeprowadzonego skanowania dokumentu.
- „Typ dokumentu” – dostępne opcje: *tytuł egzekucyjny, czynność egzekucyjna, pismo KMO, wniosek egzekucyjny, protokół z czynności, licytacja, inna, ruchomość, nieruchomość, opis i oszacowanie, uczestnik*. Wskazanie, jakiego typu dokumenty są właśnie wgrywane do bazy skanowanych dokumentów. Element ten jest również brany pod uwagę w przypadku automatycznej dekretacji dokumentu – zastosowane zostają tylko te reguły, które są dla niego przypisane.
- „Propozycja nazwy” – wybrany przez użytkownika profil dekretacji dokumentów (patrz rozdział „Nazwy skanowanych dokumentów”).
- „Nazwa” – domyślna nazwa skanowanego dokumentu – w przypadku skanowania z automatyczną dekretacją dokumentów, gdzie dla danej pozycji nie została spełniona reguła z mapowaniem na nazwę czynności, do założenia wpisu zostanie użyta właśnie ta nazwa.

Dodatkowo okno to zostało rozbudowane o następujące opcje związane z automatycznym rozpoznawaniem oraz dekretacją skanowanych dokumentów.

- „Rozpoznawanie tekstu – OCR” – umożliwi włączenie/wyłączenie opcji rozpoznawania skanowanych dokumentów. Funkcja dostępna wyłącznie po włączeniu funkcji rozpoznawania (OCR) w module „Konfiguracja” | menu „Preferencje” | „Konfiguracja skanowania”.
- „Każda kartka odrębnym dokumentem” – zaznaczenie tej opcji daje możliwość określenia ile kartek ma posiadać dokument – podczas seryjnego skanowania program podzieli wszystkie zeskanowane dokumenty na takie, które posiadają odpowiednie ilości stron. Uwzględniany jest również typ skanowania dokumentów: *simplex* lub *duplex*. W zależności od tego ustawienia program będzie odpowiednio przypisywał kolejne zeskanowane strony jako strony tego samego lub już następnego dokumentu.
- „Wyświetl dane dokumentu” – opcja przeznaczona do wyświetlenia danych zakładanego wpisu po przeprowadzonym procesie dekretacji.
- „Zadana sygnatura sprawy” – możliwość wskazania sygnatury sprawy, do której będą przyporządkowywane wszystkie dokumenty w ramach pojedynczego procesu skanowania.

Skanowanie dokumentów w modelu 1

Model ten zakłada rozbięcie całego procesu przyporządkowywania skanowanych dokumentów na dwa etapy:

- skanowanie seryjne
- ręczne przyporządkowywanie zeskanowanych wcześniej w sposób seryjny dokumentów do poszczególnych miejsc w strukturze aplikacji

Proces seryjnego skanowania dokumentów został opisany w podrozdziale „Skanowanie seryjne”. Tym elementem może zajmować się osoba wyznaczona wyłącznie do tego celu. Użytkownicy, którzy w następnej kolejności otrzymują wybrane dla nich dokumenty muszą przeprowadzić proces ręcznej ich dekretacji, który sprowadza się do:

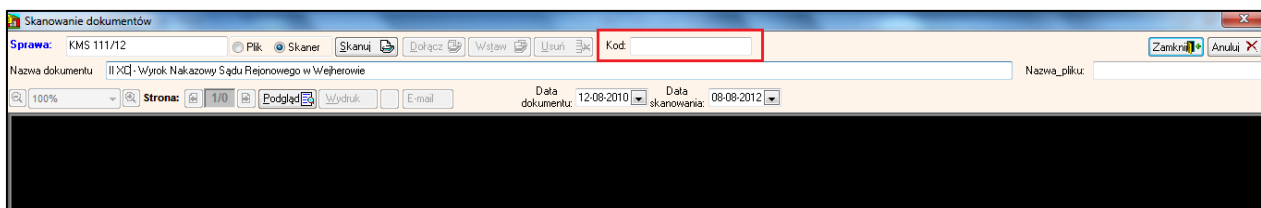
- wyszukania oraz wskazania dokumentu na liście skanowanych dokumentów
- wciśnięcia prawego przycisku myszy w celu wyświetlenia menu kontekstowego
- wybrania opcji „Powiąż” oraz przejściu kolejnych kroków kreatora umożliwiającego wskazanie sprawy, wybrania istniejącego lub dodania nowego wpisu, do którego zostanie przyporządkowany dokument (np. wybór lub utworzenie nowej czynności obcej we wskazanej sprawie egzekucyjnej)

Procedura jest powtarzana dla kolejnych dokumentów przyporządkowanych do danego pracownika kancelarii, które zostały wcześniej zeskanowane i zapisane w bazie danych. Proces ten może być bez problemu wykonywany równoległe przez wszystkich użytkowników.

Skanowanie dokumentów w modelu 2

W modelu tym wykorzystywane są kody kreskowe, którymi oklejane są dokumenty podlegające skanowaniu. Osoba wyznaczona w kancelarii za odbiór korespondencji powinna mieć przygotowane arkusze samoprzylepne z etykietami zawierającymi kody kreskowe. Wydruk takich arkuszy został opisany w podrozdziale „Drukowanie etykiet” rozdziału „Konfiguracja aplikacji”. Wszystkie pierwsze strony dokumentów zostają przez tą osobę oklejone kolejnymi etykietami. Następnymi krokami są:

- skanowanie seryjne – proces opisany został powyżej w rozdziale „Skanowanie seryjne”. Należy pamiętać, aby podczas tego procesu włączona była funkcja wykrywania kodów kreskowych w skanowanych dokumentach (moduł „Konfiguracja” | menu „Preferencje” | „Preferencje skanowania” | opcja „Szukanie kodu kreskowego”) – dzięki temu podczas skanowania w bazie danych zostanie zapisana informacja odczytana automatycznie z kodu kreskowego, umożliwiającą późniejsze powiązanie zeskanowanego dokumentu z odpowiednim wpisem w bazie danych.
- ręczne przyporządkowywanie zeskanowanych wcześniej dokumentów do poszczególnych miejsc w strukturze aplikacji



Przyporządkowywanie dokumentów odbywa się z poziomu poszczególnych spraw. Użytkownik rejestrujący daną czynność egzekucyjną, tytuł wykonawczy lub jakikolwiek dokument innego typu przeprowadza ten proces, a następnie za pomocą przycisku „Skanuj” podłącza zeskanowany wcześniej dokument. Do tego celu wystarczy posiadać oryginał dokumentu zawierający naklejony kod kreskowy. Ustawienie kursora w polu „Kod” w oknie „Skanowanie dokumentów” oraz wprowadzenie kodu za pomocą czytnika kodów kreskowych spowoduje automatyczne odnalezienie dokumentu oraz podłączenie do rejestrowanego elementu. W takim przypadku nie jest wymagane żmudne wyszukiwanie skanowanego dokumentu z listy.

Skanowanie dokumentów w modelu 3

Model ten zakłada największą automatyzację procesu skanowania i dekretacji dokumentów. Umożliwia on zamianę obrazów zawierających graficzny odpowiednik dokumentu na tekst, który może zostać poddany dalszej obróbce i interpretacji.

Ze względu na fakt, iż proces zamiany obrazu na tekst wykonywany jest na bazie zewnętrznego modułu wymagane jest poniesienie dodatkowych kosztów związanych z jego wykorzystywaniem. Jednak w celu minimalizacji tych kosztów możliwe jest przeprowadzanie rozpoznawania podczas skanowania wyłącznie na ograniczonej liczbie stanowisk komputerowych, a późniejszą dekretację dokumentów (w oparciu o zdefiniowane zmienne i reguły) przeprowadzać na pozostałych stanowiskach. Oczywiście istnieje możliwość przeprowadzania procesów rozpoznawania oraz dekretacji równolegle – wybór sposobu postępowania zależy wyłącznie od użytkowników.

Zatem proces skanowania wraz z automatyczną dekretacją, zakładając, iż użytkownicy przeprowadzili już konfigurację zmiennych oraz reguł (opisana została w rozdziale „Konfiguracja aplikacji”, podrozdział „Filtry dla OCR”) przedstawia się następująco:

Aby istniała możliwość rozpoznawania tekstu w zeskanowanych obrazach należy w pierwszej kolejności uaktywnić funkcję „*Włącz funkcję OCR (wymagana biblioteka LeadTools)*” dostępną w module „Konfiguracja” | menu „Preferencje” | „Konfiguracja skanowania”. Dzięki temu podczas skanowania dokumentów zapisywana będzie jego elektroniczna wersja w postaci umożliwiającej jej interpretację. Takie dokumenty oznaczane zostają w kolumnie „R” wartością „R” – rozpoznane.

W zależności od ustawienia „*Automatyczna dekretacja dokumentów*” proces ten jest wykonywany automatycznie po zapisaniu wszystkich dokumentów lub jest możliwy do wykonania w terminie późniejszym poprzez użycie przycisk „*Dekretuj*”.

Niezależnie od momentu, kiedy zostaje wywołana funkcja dekretacji ustalająca na podstawie zdefiniowanych reguł odpowiednie parametry dokumentu i przyporządkowuje je do odpowiednich miejsc w bazie danych. Po zakończeniu operacji następuje automatyczna zmiana danych dokumentów wyświetlanych na liście. Proces dekretacji wykonuje następujące operacje:

- wykrycie zmiennych i reguł, w tym odpowiedzialnej za sygnaturę sprawy.
- przyporządkowanie parametrów dokumentu na podstawie wykrytych zmiennych i reguł (np. ustalenie nazwy dokumentu), ewentualnie przyjęcie wartości domyślnych.
- dekretacja dokumentu do sprawy o wykrytej sygnaturze (jeśli włączona jest funkcja „*Weryfikacja*”).

rozpoznania sygnatury” następuje dodatkowe sprawdzenie poprawności stron postępowania).

- wyświetlenie danych szczegółowych zakładanych w bazie wpisów (tylko w przypadku włączonej w oknie „Skanowanie seryjne” opcji „Wyświetl dane dokumentu”.

Proces ten jest powtarzany dla wszystkich zeskanowanych lub zaznaczonych (w przypadku, gdy wykonywany był z listy) dokumentów.

Uwagi końcowe

Sam proces rozpoznawania tekstu w obrazach graficznych jest bardzo złożonym procesem, który nie gwarantuje 100% trafności rozpoznawania. Należy o tym pamiętać podczas jego wykorzystywania. Jednak dołożyliśmy wszelkich starań, aby skanowanie i dekretacja dokumentów była jak najbardziej trafna, niemniej jednak informujemy, iż nie ponosimy żadnej odpowiedzialności za jakiegokolwiek nieprawidłowości podczas przyporządkowywania danych. Dostarczony produkt jest w pełni konfigurowalny, dzięki czemu użytkownicy mogą zwiększyć trafność przyporządkowań oraz ich opisów we własnym zakresie. Mimo, iż proces ten jest złożony to jego wykonanie można zlecić osobom dbającym o systemy informatyczne kancelarii.

Wszelkie uwagi dotyczące funkcjonowania opisanego modułu prosimy kierować do działu Helpdesk za pomocą portalu serwisowego www.helpdesk.currenda.pl. Każda uwaga zostanie przeanalizowana, a te najbardziej interesujące zostaną zaimplementowane w aplikacji.