



# Mechanizm zamienników Regex

Poradnik Użytkownika eSZOK

# Mechanizm stosowanie zamienników Regex w Kancelarii

## I. Do czego służy mechanizm zamienników?

### 1. Mechanizm zamienników

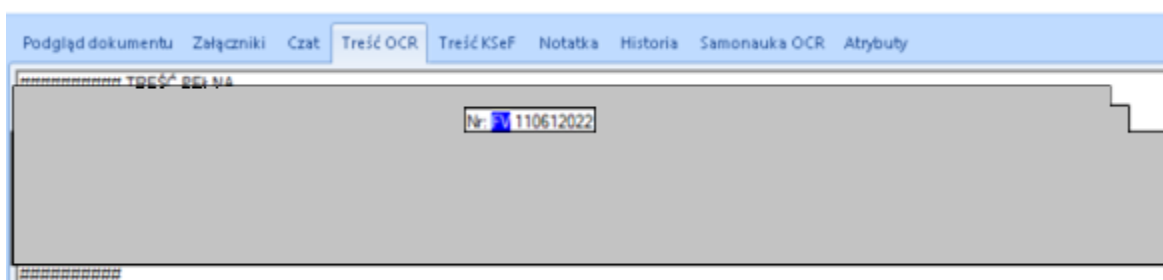
Mechanizm zamienników w Kancelarii służy do samodzielnego wprowadzania schematów rozpoznawania danych tzw.: Regex'ów.

Jeśli system odczytywania danych (OCR), błędnie odczyta informację na dokumencie wprowadzonym do Kancelarii, a mechanizm samonauki nie podpowiada poprawnego wariantu, należy przejść do zakładki Treść OCR:

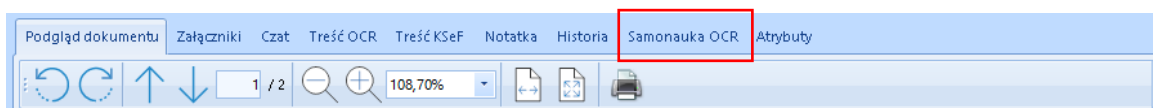


Należy sprawdzić czy w treści znajduje się szukany fragment treści np. nr dokumentu, który został błędnie odczytany.

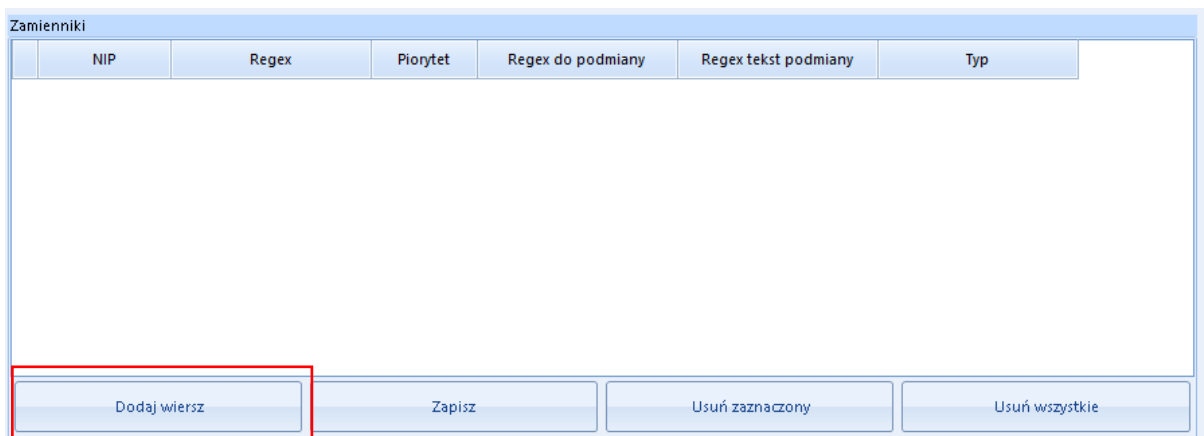
.....  
**Nr: FV 1/06/2022**  
.....



Po odnalezieniu w treści numeru dokumentu, przechodzimy do zakładki Samonauka OCR



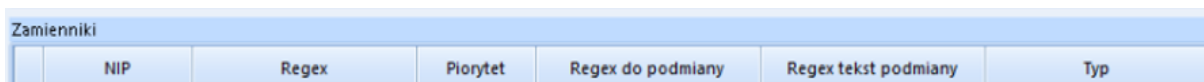
Na dole zakładki Samonauka znajduje się sekcja Zamienniki. Chcąc dodać nowy zamiennik klikamy w **Dodaj wiersz**:



NIP	Regex	Priorytet	Regex do podmiiany	Regex tekst podmiiany	Typ
-----	-------	-----------	--------------------	-----------------------	-----

**Dodaj wiersz**    Zapisz    Usuń zaznaczony    Usuń wszystkie

W Zamiennikach zostanie dodany nowy wiersz, w poszczególnych polach należy uzupełnić informacje o regexie do podmiiany:



NIP	Regex	Priorytet	Regex do podmiiany	Regex tekst podmiiany	Typ
-----	-------	-----------	--------------------	-----------------------	-----

Co oznaczają poszczególne pola:

**NIP** - wstawione automatycznie, na podstawie rozpoznanego nr NIP na fakturze

**Regex** – to pole uzupełniamy samodzielnie wzorem rozpoznania informacji w Treści OCR [Informacje o wzorach regexów można znaleźć w internecie a także istnieje możliwość testowania zamienników np. [www.regex101.com](http://www.regex101.com), [www.regexr.com](http://www.regexr.com) ].

**Priorytet** – kolejność wyszukiwania zamiennika, została automatycznie ustawiona na zero

**Regex do podmiiany**- pole to uzupełniamy samodzielnie podając wzór wyszukiwania w regexie miejsc, które mają zostać zastąpione.

**Regex tekst podmiiany** – w pole to należy wpisać poprawny znak tj. tekst/cyfrę/symbol, który zastępuje błędnie rozpoznany znak .

**Typ**- rodzaj rozpoznanej informacji, do wyboru rozwijanej list.

## 2. Przykład

Dla przykładu zastosowania regexów rozpatrzmy przypadek błędnie odczytanego nr dokumentu tj.: **FV 110612022**.

Poprawny nr dokumentu to **FV 1/06/2022**. Chcemy by w miejsce „1” został podstawiony znak „/”.

Najpierw zapisujemy wzór rozpoznanego numeru (czyli pole Regex) tj.: [A-Z][A-Z][\d\*202\d .

Zamienniki						
	NIP	Regex	Płorytety	Regex do podmiany	Regex tekst podmiany	Typ
+	8731011086	[A-Z][A-Z][\d*202\d	0	((?<=\d*)1(?=202\d)) ((?<=\d*)1(?=\d{2}1202\d))	/	Numer Dokumentu

Następnie musimy wyznaczyć miejsca (Regex do podmiany), w których znajduje się błędnie podstawiona „1” czyli ((?<=\d\*)1(?=202\d))|((?<=\d\*)1(?=\d{2}1202\d)) .

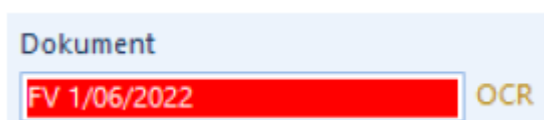
W kolejnym polu należy wpisać tekst podmiany (Regex tekst podmiany) czyli znak, który zastąpi „1” w naszym przypadku to „/”.

Ostatnim krokiem jest wskazanie typu danych, w tym przypadku jest **Numer Dokumentu**. Wpisany zamiennik zapisujemy przyciskiem **Zapisz**.

Jeżeli informacje zostały wprowadzone np. dwa razy, można je usunąć za pomocą Usuń zaznaczony lub Usuń wszystkie.



Następnie należy PPP usunąć dane OCR faktury i jeszcze raz zaczytać dane OCR, wówczas nr dokumentu zostanie poprawnie podstawiony, zgodnie z zapisanym regexem.



## II. Wyrażenia regularne

1. Wyrażenia regularne to wzorce, dzięki którym możemy sprawdzać czy dany ciąg znaków ma określony format. Poniżej słowniczek wyrażeń do zastosowania przy tworzeniu regexów:

- 1) [A-z] - wyszukuje dowolną ilość liter w zakresie od a do z, bez względu na ich wielkość
- 2) [A-Z]- wyszukuje dowolną ilość liter w zakresie od A do Z, duże litery

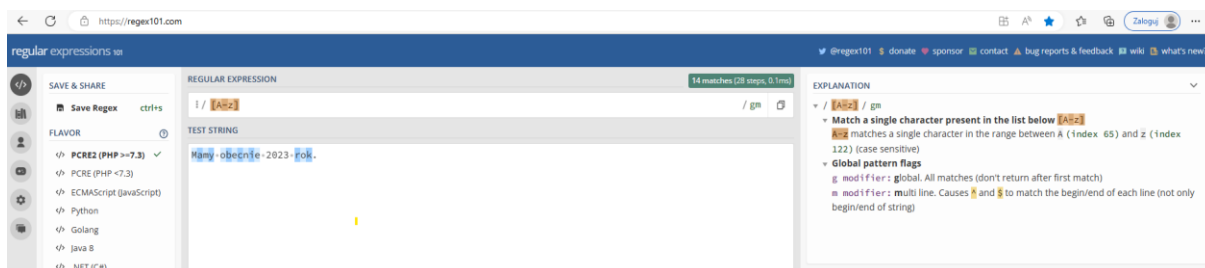
- 3) [0–9] – wyszukuje dowolną ilość cyfr w zakresie od 0 do 9
- 4) . – wyszukuje dowolną ilość znaków (w tym spacje)
- 5) \d – wyszukuje dowolną ilość cyfr o nieokreślonej ilości znaków
- 6) \D – wyszukuje wszystkie znaki, które nie są cyframi (w tym spacje)
- 7) \W – wyszukuje wszystko poza literą, cyfrą lub podkreśleniem
- 8) \d\* – wyszukuje dowolny ciąg liczbowy, o nieokreślonej ilości znaków
- 9) \d+ – wyszukuje dowolny ciąg liczbowy, o nieokreślonej ilości znaków
- 10) \d{4} – dowolna liczba o długości 4 cyfr
- 11) \d{3,} – dowolna liczba o długości 3 cyfr lub więcej
- 12) \d{2,4} – wyszukuje liczbę o długości od 2 do 4 znaków
- 13) .{5} – wyszukuje dowolny ciąg znaków o długości 5 znaków (w tym spacje)
- 14) (?<=obecnie )2022 – wyszukuje ciąg znaków 2022, przed którymi jest wyraz „obecnie” (nie jest on brany do wyniku)
- 15) 2022(?= rok) – wyszukuje ciąg znaków 2022, za którym jest wyraz „rok” (nie jest on brany do wyniku)
- 16) m|n - | jest operatorem OR, w tym przypadku wyszukuje liter m lub n
- 17) ^M – wyszukuje znak M na początku podanego ciągu znaków
- 18) k\$ – wyszukuje znak k na końcu podanego ciągu znaków
- 19) (c|) – szuka c lub pustego znaku
- 20) e? – szuka e lub pustego znaku
- 21) (?<=<img src="")[^"]\*(?=") - Wyszukiwanie linków obrazków

2. Zastosowanie wyrażeń regularnych wzorców na 2 przykładach wraz z możliwością testowania na stronie [www.regex101.com](http://www.regex101.com).

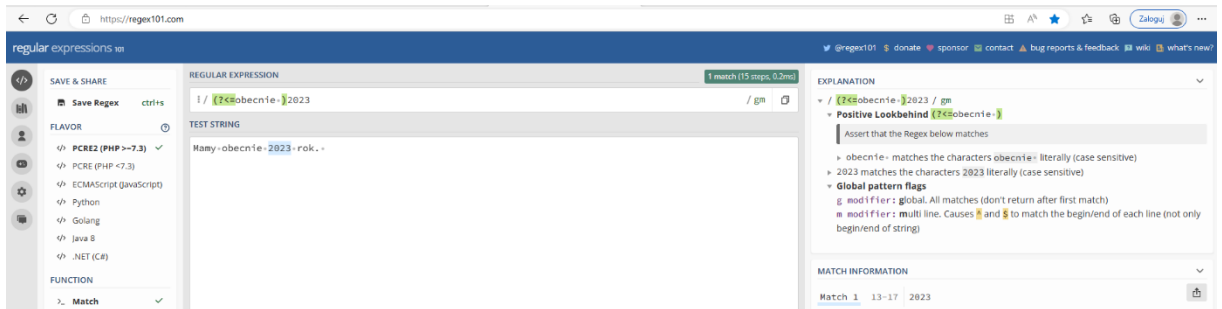
Przykład 1.

Tekst: *Mamy obecnie 2023 rok.*

Wpisując wyrażenie [A-z] program zaznacza nam jakie litery znalazł w tekście.



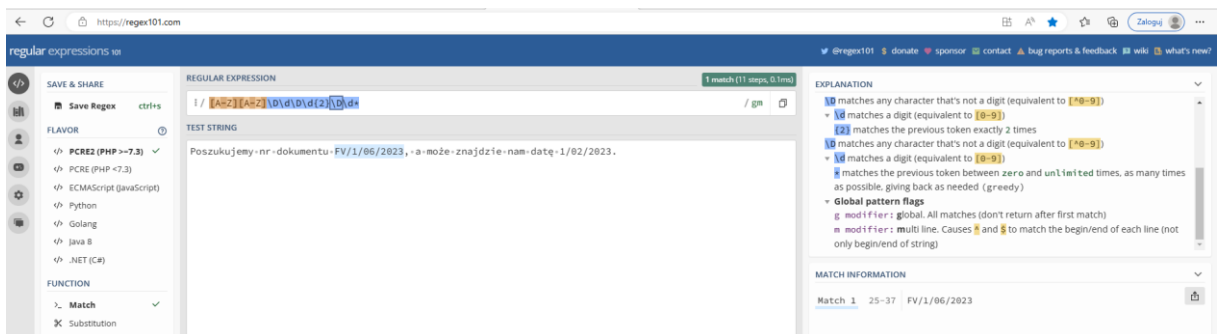
Wpisując wyrażenie `(?<=obecnie )2023`, znajdzie nam ciąg znaków jaki jest przed słowem obecnie.



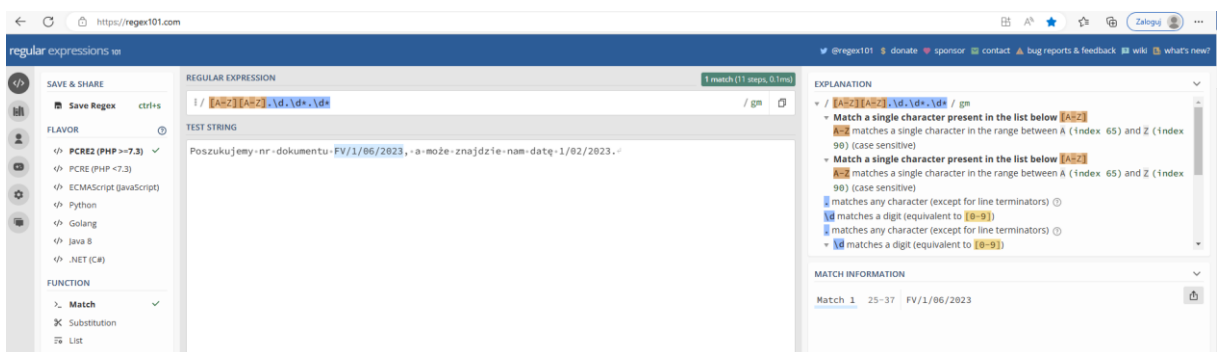
Przykład 2.

Tekst: *Poszukujemy nr dokumentu FV/1/06/2023, a może znajdzie nam datę 1/02/2023.*

Jeżeli chcemy odnaleźć w tekście nr dokumentu: **FV/1/06/2023**, to wykorzystując powyższe wyrażenia możemy stworzyć poszukiwany regex: `[A-Z][A-Z]\D\d\D\d{2}\D\d*`



Możemy poszukiwany nr dokumentu zapisać w innej postaci by odnaleźć poszukiwany format znaków: `[A-Z][A-Z].\d.\d*.\d*`



Jak widzimy dzięki regularnym wyrażeniom podanym w słowniku możemy stworzyć odpowiedni format poszukiwanego ciągu znaków.

Data utworzona poradnika: 14.02.2023