


Préambule

 Les expressions ici utilisent la syntaxe reconnue par python qui suit le standard POSIX (à quelques exceptions près ¹⁾).

Capture

Les parenthèses () de groupement permettent de délimiter un ensemble d'alternatives et d'opérer des captures. Le plus souvent si nous réalisons une capture nous voudrions pouvoir y faire référence via un nom que nous aurons choisi.

En python la syntaxe qui permet cela est la suivante

```
(?P<nom_du_groupe_de_capture>...)
```

La sous-expression régulière (ici symbolisée par ...) qui aura été trouvée sera accessible par le nom **nom_du_groupe_de_capture**

Dès lors nous pouvons y faire référence dans la suite de l'expression régulière elle-même :

```
(?P=nom_du_groupe_de_capture)\\
```

Mais le plus souvent nous voudront surtout y faire référence lors d'un post formatage. Tout de suite un petit exemple.

Numéro de facture

Soit le texte extrait suivant

```
RueDuCommerce SA
Facture 44-50, avenue du Capitaine Glarner
93585 Saint-Ouen Cedex
EzDEV
```

2

```
Rue Adolphe Pégoud
90130 PETIT-CROIX
    Numéro de facture : 56497          Référence client : 100132
    Date de la facture : 04/11/2015
```

Cet exemple illustre bien l'un des cas que vous retrouvez le plus souvent. Il s'agit de récupérer un numéro (ici de facture).

Nous pourrions être tentés de faire une recherche avec le filtre de champ

```
Numero de facture : .*
```

puis un post formatage

```
@keepdigits(@self)
```

Ce qui pourrait fonctionner si nous spécifions correctement la longueur du champ.

Soit le filtre de champ suivant

```
Numero de facture : (?P<numero>\d+)
```

Pour ne garder que le numero on fera ceci dans le champ post de formatage :

```
@caststr(@<numero>)
```

Recherche multi-ligne

Prenons cet exemple de texte extrait

```
Tricatel  
120, Avenue Foch  
94000  
Créteil
```

| | Facture N° | Date de la facture |
|-----------------------------|---|--------------------|
| Adresse Client : | | 04/11/2015 |
| | 681476 | |
| Gastro Discount | | |
| Numéro de client | 456 | |
| La Défense | | |
| Paiement d0 | 04/12/2015 | |
| 75000 Paris | | |
| Modalité de paiement | 30 jours | |
| France | | |
| Emis par | Jacques Tricatel | |
| Contact client | Noël Grobouchon | |
| Date de la vente/prestation | 20/10/2015 | |
| Adresse de livraison | Gastro Discount Lyon, ZAC De la Malbouf | |
| 69100 VILLEURBANNE | | |
| Date de livraison | Du 05/11/2015 au 10/11/2015 | |

Si l'on souhaite récupérer le numéro de facture on pourrait le faire assez simplement en utilisant la macro @relativeto. Ce serait une bonne solution mais il existe une alternative qui ne nous obligera

pas à créer deux champs d'analyse.

Sur un modèle d'analyse un peu touffu voilà quelque chose que l'on aimerait éviter !

Nous allons créer un seul champ de type variable, nous allons restreindre la zone de recherche via les positions X/Y.

1)

Article wikipedia sur les expression rationnelles

https://fr.wikipedia.org/wiki/Expression_rationnelle#Python

From:

<http://wiki.ezdev.fr/> - **EzGED Wiki**

Permanent link:

<http://wiki.ezdev.fr/doku.php?id=cold:tutoriaux:regex&rev=1446654003>

Last update: **2023/03/17 09:56**

