



# Les expressions régulières



## Définition

- Une chaîne de caractères, qui décrit, selon une syntaxe précise, un ensemble de chaînes de caractères possibles.
- Ainsi, si les caractères correspondent, la valeur de l'expression est vraie : le texte recherché est trouvé.
- La grammaire permet de faire des expressions extrêmement complexe, ce qu'il faut éviter...



# Grammaire

- Il existe différentes grammaires pour les expressions régulières:
  - Standard POSIX
  - Python
  - Tcl
  - Perl (PCRE)
  - PHP (Très proche de POSIX)
  - ICU
  
- Le standard POSIX est le plus répandu dans les outils de recherche

# La base POSIX

- **^** représente le début d'une ligne, **\$** la fin
- **.** représente un caractère
- **|** représente l'alternative « ou »
- **[]** représente un ensemble de caractère
- **[^]** ensemble de caractère à exclure
- **[-]** représente un intervalle de valeur
- **{min,max}** est un quantificateur
  - **\*** signifie 0 ou plus {0,}
  - **+** signifie 1 ou plus {1,}
  - **?** signifie 0 ou 1 {0,1}

**^**téléphone**\$**

t.l.phone

t(e|é)l(e|é)phone

t[eé]l[eé]phone

t[^aiouy]l[eé]phone

[0-9][0-9][0-9][0-9]

[0-9]{4}

# Exemples

- Un numéro de téléphone :

- $(0|\backslash+33)$
- $[1-9]$
- $[0-9]\{8\}$ 
  - ->  $(0|\backslash+33)[1-9][0-9]\{8\}$

- Prise en compte des séparateurs:

- $(0|\backslash+33)( *) [1-9]( *[0-9]\{2\})\{4\}$

+33 1 23 4567 89

- $(0|\backslash+33)( *) [1-9]( (( *) | (\backslash.*) | (-*)) [0-9]\{2\})\{4\}$

01.23-45 6789

0826 792 792



# Les limites

- Recherche de 2 mots

`.*(bonjour).*(salut).*`

Attention l'ordre compte...

`.*(bonjour|salut).*`

L'un ou l'autre ou les deux...

Mais sur la même ligne !!

1 fichiers trouvés - Recherche:.\*(NUM\_DOSSIER).\*(46317).\* dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

.\*(NUM\_DOSSIER).\*(46317).\* x

Principal Options Dates Attributs Scripting Fichiers compressés Expert

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le .\*(NUM\_DOSSIER).\*(46317).\* Expression régulière Aa ?

Examiner: F:\\_CARVED\_\Export\eDossier [Sous-dossiers] Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 30/03/2014 Avant: 01/09/2014

Nom	Emplacement	M
0061401015073_fin.xml	F:\_CARVED_\...\0061401015073\	06/01/

Sommaire Concordances Texte Miniatures Rapports

F:\\_CARVED\_\Export\eDossier\0061401015073\0061401015073\_fin.xml (6 Ko, 06/01/2014)

```
<Champ Nom="NUM_DOSSIER"><String>46317</String></Champ>
```

Prêt Trouvé: 1 élément (5.05 Ko) Sélectionné: 1 élément (5.05 Ko)



0 fichiers trouvés - Recherche:.\*(46317).\*(NUM\_DOSSIER).\* dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

.\*(46317).\*(NUM\_DOSSIER).\* x

Principal Options Dates Attributs Scripting Fichiers compressés

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le .\*(46317).\*(NUM\_DOSSIER).\* Expression régulière Aa ?

Examiner: F:\\_CARVED\_\Export\eDossier  Sous-dossiers Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 30/03/2014 Avant: 01/09/2014

Expert

Départ Arrêt

Nom	Emplacement
<i>Aucun fichier ou dossier trouvé correspondant aux critères de recherche</i>	

Sommaire Concordances Texte Miniatures Rapports

**Statistiques de recherche**

Trouvé:	0 éléments (0.00 Ko)
Cherché:	10 535 éléments (44.68 Mo)
Recherche en attente:	0 éléments (0.00 Ko)
Vérifié:	61 231 éléments (5.38 Go)
Etat:	Terminé (21 secs)
Terminé:	27/10/2017 10:40:22

**⚠ Avertissement généré durant la recherche**

Terminé Trouvé: 0 éléments (0.00 Ko) Sélectionné: 0 éléments (0.00 Ko)



9 889 fichiers trouvés - Recherche:.\*(46317|NUM\_DOSSIER).\* dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

Principal Options Dates Attributs Scripting Fichiers compressés

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le .\*(46317|NUM\_DOSSIER).\* Expression régulière Aa ?

Examiner: F:\\_CARVED\_\Export\eDossier  Sous-dossiers Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 30/03/2014 Avant: 01/09/2014

Nom	Emplacement	N
3651301015072_fin.xml	F:\_CARVED_\...\3651301015072\	31/
3651301015071_fin.xml	F:\_CARVED_\...\3651301015071\	31/
3651301015070_fin.xml	F:\_CARVED_\...\3651301015070\	31/
3651301015069_fin.xml	F:\_CARVED_\...\3651301015069\	31/
3651301015068_fin.xml	F:\_CARVED_\...\3651301015068\	31/
3651301015067_fin.xml	F:\_CARVED_\...\3651301015067\	31/
3651301015066_fin.xml	F:\_CARVED_\...\3651301015066\	31/
3651301015065_fin.xml	F:\_CARVED_\...\3651301015065\	31/
3651301015064_fin.xml	F:\_CARVED_\...\3651301015064\	31/
3651301015063_fin.xml	F:\_CARVED_\...\3651301015063\	31/

Sommaire Concordances Texte Miniatures Rapports

F:\\_CARVED\_\Export\eDossier\0061401015074\0061401015074\_fin.xml (3 Ko, 06/01/2014)

```
<Champ Nom="NUM_DOSSIER"><String>47760</String></Champ>
```

Sélectionné 1 élément (2.93 Ko) Trouvé: 9 889 éléments (42.07 Mo) Sélectionné: 1 élément (2.93 Ko)



1 fichiers trouvés - Recherche: 46317 NUM\_DOSSIER dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

46317 NUM\_DOSSIER x

Principal Options Dates Attributs Scripting Fichiers compressés

Nom fichier: \*.xml Expression DOS Aa ?  
Contenant le: 46317 NUM\_DOSSIER Mot entier Aa ?  
Examiner: F:\\_CARVED\_\Export\eDossier [Sous-dossiers] Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 30/03/2014 Avant: 01/09/2014

Expert  
Départ Arrêt

Nom	Emplacement	M
0061401015073_fin.xml	F:\_CARVED_\...0061401015073\	06/01

Sommaire Concordances Texte Miniatures Rapports

F:\\_CARVED\_\Export\eDossier\0061401015073\0061401015073\_fin.xml (6 Ko, 06/01/14)

```
20 <Champ Nom="NUM_DOSSIER"><String>46317</String></Champ>
```

Prêt Trouvé: 1 élément (5.05 Ko) Sélectionné: 1 élément (5.05 Ko)



# Les classiques

- Recherche de mots clés dans des courriels entre 2 dates

1 fichiers trouvés - Recherche: 46317 NUM\_DOSSIER dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

46317 NUM\_DOSSIER x

Principal Options **Dates** Attributs Scripting Fichiers compressés

Vider tout

Expert

Départ Arrêt

Modifié:  Après: 01/01/2014  Avant: 01/09/2014

Créé:  Après: 01/04/2017  Avant: Aujourd'hui

Dernier accès:  Après: Aujourd'hui  Avant: Aujourd'hui

Nom	Emplacement	Modifié
0061401015073_fin.xml	F:\_CARVED_\...0061401015073\	06/01/2014 15:14:32

Sommaire Concordances Texte Miniatures

F:\\_CARVED\_\Export\eDossier\0061401015073\0061401015073\_fin.xml

20 <Champ Nom="NUM\_DOSSIER"><String>46317</Strir

Sélectionné 1 élément (5.05 Ko) Trouvé: 1 élément (5.05 Ko) Sélectionné: 1 élément (5.05 Ko)

3 503 fichiers trouvés - Recherche:.\*(2014)-(0[1-8])-([0-9]{2}).\* dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

.\*(2014)-(0[1-8])-([0-9]{2}).\* x

Principal Options Dates Attributs Scripting Fichiers compressés

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le .\*(2014)-(0[1-8])-([0-9]{2}).\* Expression régulière Aa ?

Examiner: F:\\_CARVED\_\Export\eDossier\  Sous-dossiers Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 01/01/2014 Avant: 01/09/2014

Nom	Emplacement	Modifié
1881401018797_fin.xml	F:\_CARVED_\...\1881401018797\	07/07/2014 11
1881401018796_fin.xml	F:\_CARVED_\...\1881401018796\	07/07/2014 11
1881401018795_fin.xml	F:\_CARVED_\...\1881401018795\	07/07/2014 11
1881401018794_fin.xml	F:\_CARVED_\...\1881401018794\	07/07/2014 11
1881401018793_fin.xml	F:\_CARVED_\...\1881401018793\	07/07/2014 11
1881401018792_fin.xml	F:\_CARVED_\...\1881401018792\	07/07/2014 11
1881401018791_fin.xml	F:\_CARVED_\...\1881401018791\	07/07/2014 11
1881401018790_fin.xml	F:\_CARVED_\...\1881401018790\	07/07/2014 11
1881401018789_fin.xml	F:\_CARVED_\...\1881401018789\	07/07/2014 10
1881401018787_fin.xml	F:\_CARVED_\...\1881401018787\	07/07/2014 10

Sommaire Concordances Texte Miniatures Rappo...

F:\\_CARVED\_\Export\eDossier\0061401015074\0061401015074\_fin.xml (3)

```

5      <Creation><date>2014-01-06</date></Creation>
14     <Datenum><date>2014-01-06</date></Datenum>
16     <Datevideo><date>2014-01-06</date></Datevideo>
  
```

Prêt Trouvé: 3 503 éléments (14.73 Mo) Sélectionné: 1 élément (2.93 Ko)

1 fichiers trouvés - Recherche: "NUM\_DOSSIER" AND "46317" AND "(2014)-(0[1-8])-(0-9){2}" dans F:\\_CARVED\_\Export\eDossier\ (\*.xml)

Fichier Edition Affichage Recherche Outils Fenêtre Aide

"NUM\_DOSSIER" AND "46317" AND "... x

Principal Options Dates Attributs Scripting Fichiers compressés Expert

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le "NUM\_DOSSIER" AND "46317" AND "(2014)-(0[1-8])-(0-9){2}" Booléen RegEx Aa ?

Examiner: F:\\_CARVED\_\Export\eDossier\  Sous-dossiers Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 01/01/2014 Avant: 01/09/2014

Nom	Emplacement	Modifié
0061401015073_fin.xml	F:\_CARVED_\...\0061401015073\	06/01/2014 15:14

Sommaire Concordances Texte Miniatures Rappo...

F:\\_CARVED\_\Export\eDossier\0061401015073\0061401015073\_fin.xml (6 Ko)

```

5 <Creation><date>2014-01-06</date></Creation>
14 <Datenum><date>2014-01-06</date></Datenum>
16 <Datevideo><date>2014-01-06</date></Datevideo>
20 <Champ Nom="NUM_DOSSIER"><String>46317</String></Ch:

```

Prêt Trouvé: 1 élément (5.05 Ko) Sélectionné: 1 élément (5.05 Ko)

272 fichiers trouvés - Recherche: "atrimouille" AND "Système" AND ".\*(video).\*(2014)-(0[1-8])-([0-9]{2})" dans F:\\_CARVED\_\Export\leDossier...

Fichier Edition Affichage Recherche Outils Fenêtre Aide

"atrimouille" AND "Système" AND ".\*(video).\*(2014)-(0[1-8])-([0-9]{2})" x

Principal Options Dates Attributs Scripting Fichiers compressés

Nom fichier: \*.xml Expression DOS Aa ?

Contenant le "atrimouille" AND "Système" AND ".\*(video).\*(2014)-(0[1-8])-([0-9]{2})" Booléen RegEx Aa ?

Examiner: F:\\_CARVED\_\Export\leDossier\  Sous-dossiers Simple texte

Taille (Ko) > 0 < 0 Modifiés: Après: 01/01/2014 Avant: 01/09/2014

Nom	Emplacement	Modifié
0581401015514_fin.xml	F:\_CARVED_\...\0581401015514\	27/02/2014 10
0581401015513_fin.xml	F:\_CARVED_\...\0581401015513\	27/02/2014 09
0581401015512_fin.xml	F:\_CARVED_\...\0581401015512\	27/02/2014 09
0581401015511_fin.xml	F:\_CARVED_\...\0581401015511\	27/02/2014 09
0581401015510_fin.xml	F:\_CARVED_\...\0581401015510\	27/02/2014 09
0581401015509_fin.xml	F:\_CARVED_\...\0581401015509\	27/02/2014 09
0581401015508_fin.xml	F:\_CARVED_\...\0581401015508\	27/02/2014 09
0581401015507_fin.xml	F:\_CARVED_\...\0581401015507\	27/02/2014 09
0581401015506_fin.xml	F:\_CARVED_\...\0581401015506\	27/02/2014 09
0481401015496_fin.xml	F:\_CARVED_\...\0481401015496\	17/02/2014 10

Sommaire Concordances Texte Miniatures Rappo...

F:\\_CARVED\_\Export\leDossier\0581401015514\0581401015514\_fin.xml (8)

```

15 <Operateurnum><string>atrimouille</string></Operateurnum>
16 <Datevideo><date>2014-02-27</date></Datevideo>
17 <Operateurvideo><string>Système</string></Operateurvideo>
  
```

Prêt Trouvé: 272 éléments (1.15 Mo) Sélectionné: 1 élément (5.98 Ko)

# Conclusions

- Ne pas écrire des regex trop complexes.
- Tester les regex sur un jeu de test (quitte à se le créer).
- Les regex ne sont pas adaptées à la recherche de mots clés précis.
- Les regex servent pour trouver une information structurée (N° tel, Date, N° CB...)
- On peut facilement combiner la recherche par mot clés avec des regex pour un résultat très efficace