Reguläre Ausdrücke

Quelle Oracle.com

Reguläre Ausdrücke

- Inhalt
  - Grundlagen
  - Metazeichen
  - Zeichenklassen
  - Funktionen
  - Ausdrücke
  - Übung

Reguläre Ausdrücke Grundlagen

Grundlagen

Grundlagen

- Reguläre Ausdrücke (regular expressions, regexp) sind in der Anwendungsentwicklung weit verbreitet und erlauben sehr mächtige Operationen mit Zeichenketten.
- Beispiele für solche Operationen sind pattern matching oder komplexes Find & Replace.
- Ihren Ursprung haben Reguläre Ausdrücke in den Skriptsprachen der UNIX-Welt (Perl).

Grundlagen

- Die kompakte und standardisierte Syntax legt es nahe, komplexe Operationen auf Zeichenketten auch in der Datenbank mit Regulären Ausdrücken durchzuführen.
- Die Anwendungsbeispiele
  - Formatprüfungen
    - IP-Adressen, e-Mail-Adressen, Bankleitzahlen, KFZ-Kennzeichen, ...
  - Zeichenketten extrahieren
    - Extraktion der Postleitzahl aus einer Adresse
  - Komplexes Find & Replace
    - Telefonnummern umformatieren von 089 ... in +49 (0) 89 ...

Grundlagen

- Ein regulärer Ausdruck repräsentiert ein Muster (pattern).
- Anhand dieses Musters werden dann in der zu durchsuchenden Menge (eine Datei oder einer Tabelle in der Datenbank passende Zeichenketten gefunden (pattern matching).
- Zeichen, die beim Suchen direkt übereinstimmen müssen, werden auch als solche in einem regulären Ausdruck notiert.

Metazeichen

- Zusätzlich können Metazeichen notiert werden.

Zeichen	Bedeutung
.	... steht für ein beliebiges Zeichen.
[ABC]	Eckige Klammern beschreiben eine Auswahl. Dieses Beispiel steht also für das Zeichen A,B oder C.
[A-Z]	Der Bindestrich innerhalb eckiger Klammern bestimmt einen Bereich. Daher steht dieses Beispiel für alle Zeichen des lateinischen Alphabets.
[^A]	Das Dach (^) negiert die Auswahl. Daher werden hier alle Zeichen des lateinischen Alphabets außer "A" angesprochen.
+	... kennzeichnet eine Mengenangabe. Das "+" steht für **ein oder mehrere** Vorkommen des Zeichens oder Metazeichens.
*	... steht **für Null bis viele Vorkommen** des Zeichens oder Metazeichens.
{m,n}	... steht für "m" bis "n" Vorkommen des Zeichens oder Metazeichens.
^	... bezeichnet, dass das Zeichen oder Metazeichen am Anfang der Zeichenkette vorkommen muss.
$	... bezeichnet, dass das Zeichen oder Metazeichen am Ende der Zeichenkette vorkommen muss.
()	... dienen zur Gruppierung von Regulären Ausdrücken. Speziell für Find & Replace ist dies wichtig, denn bei der Ersetzung können die Gruppen dann mit dem Backslash ("\1" bis "\n") einzeln angesprochen werden

Zeichenklassen

- Darüber hinaus stehen ab Oracle10g Release 2 auch die sog. Zeichenklassen bereit

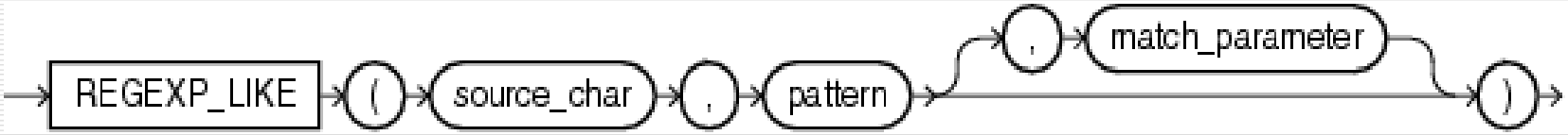
Zeichen	Bedeutung
[:digit:]	... steht für die Ziffern von "0" bis "9".
[:lower:]	... steht für kleingeschriebene Buchstaben. Dies beinhaltet die Zeichen von "A" bis "Z" zzgl. sprachspezifischer Sonderzeichen (ä,ü,ö).
[:upper:]	... steht für großgeschriebene Buchstaben. Dies beinhaltet die Zeichen von "A" bis "Z" zzgl. sprachspezifischer Sonderzeichen (ä,ü,ö).
[:alpha:]	... steht für die Vereinigung von [:lower:] und [:upper:] .
[:blank:]	... steht für alle Arten von Leerzeichen, also Tabulator und Blanks.

Funktionen

- Die Oracle-Datenbank stellt fünf Funktionen zum Umgang mit regulären Ausdrücken bereit.
- Die Funktionen können in jeder SQL-Abfrage und in jedem PL/SQL-Block verwendet werden
  - **REGEXP_LIKE** stellt fest, ob ein Muster in der Zeichenkette existiert.
  - **REGEXP_SUBSTR** extrahiert die zum regulären Ausdruck passende (Teil-)Zeichenkette.
  - **REGEXP_INSTR** gibt die Zeichenposition zurück, an der die zum Ausdruck passende Teilzeichenkette beginnt.
  - **REGEXP_REPLACE** ersetzt die zum Ausdruck passende Teilzeichenkette durch eine andere.
  - **REGEXP_COUNT** gibt die Anzahl des Vorkommens eines Musters an

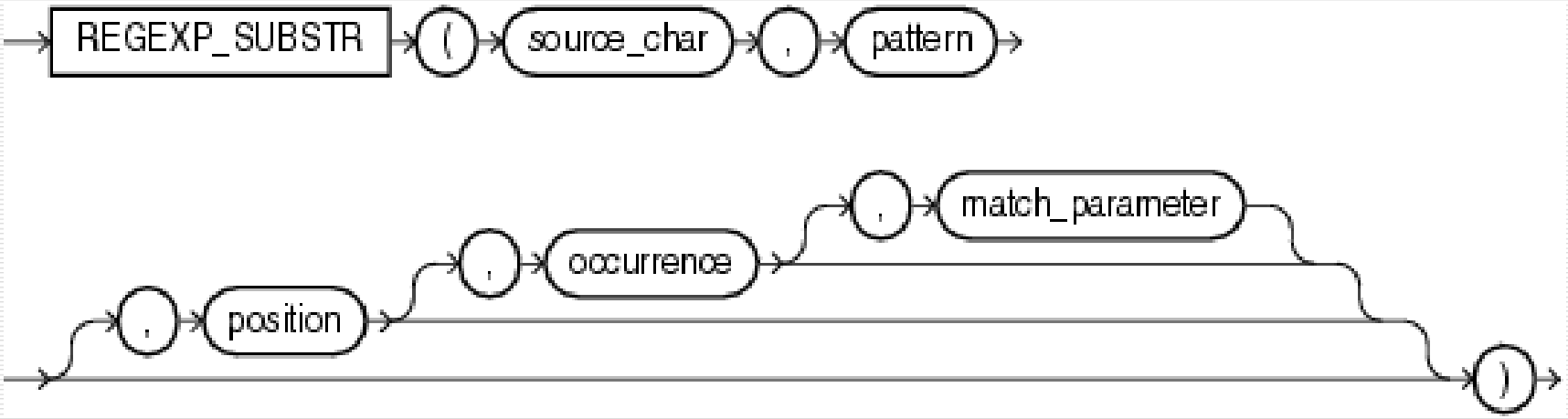
Funktionen

- Verwendung REGEXP_LIKE



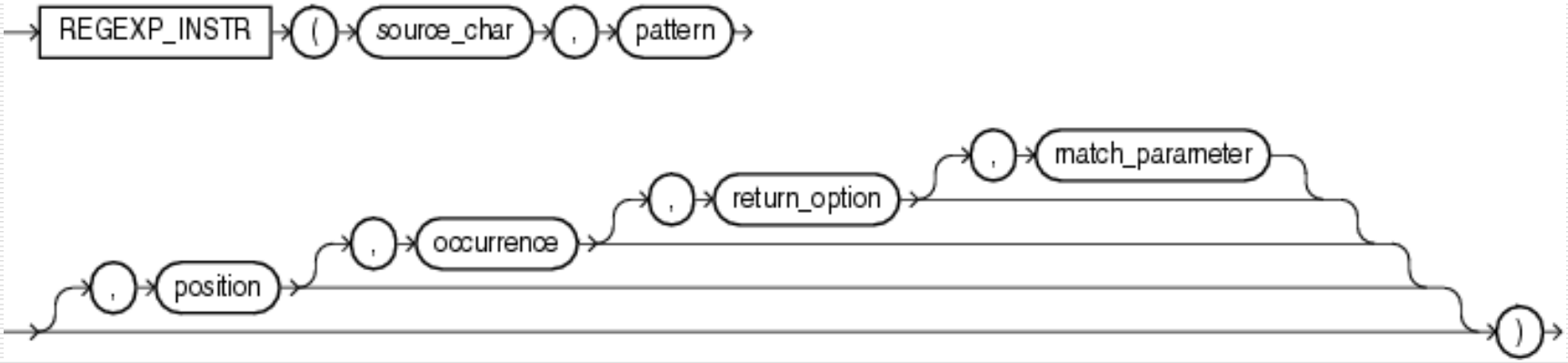
Funktionen

- Verwendung REGEXP_SUBSTR



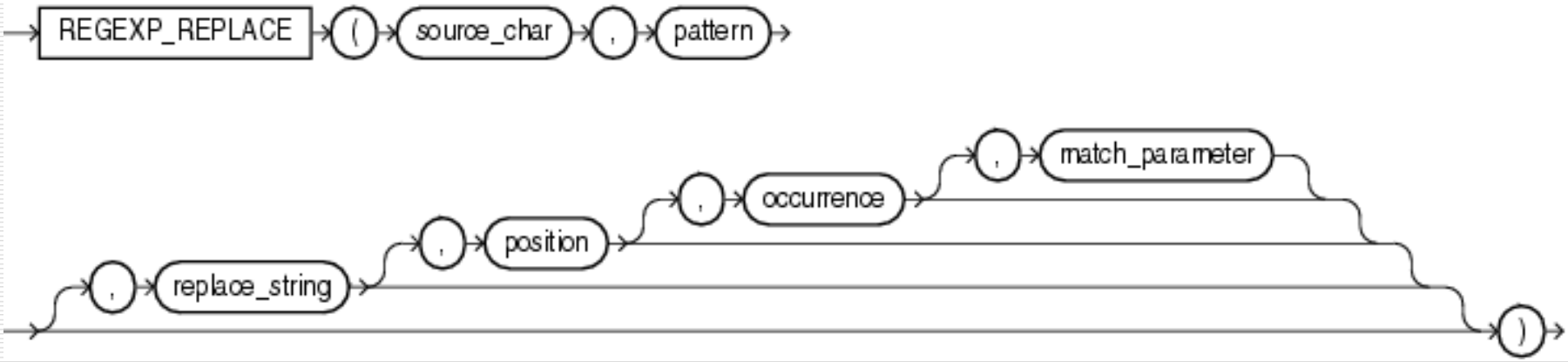
Funktionen

- Verwendung REGEXP_INSTR



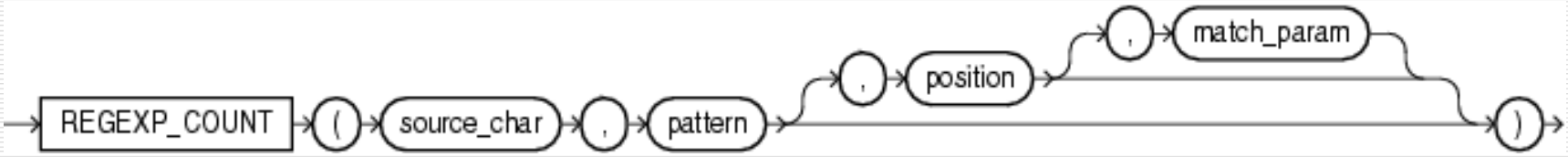
Funktionen

- Verwendung REGEXP_REPLACE



Funktionen

- Verwendung REGEXP_COUNT



Beispiele

- Vorkommen des Buchstaben „W“

```
SQL> select last_name from employees where
    regexp_like(last_name,'[W]');
LAST_NAME
-------------------------
Walsh
Weiss
Whalen
```

Beispiele

- Vorkommen der Buchstaben „W“ oder „V“

```
SQL> select last_name from employees where
    regexp_like(last_name,'[WV]');
LAST_NAME
-------------------------
Vargas
Vishney
Vollman
Walsh
Weiss
Whalen
```

Beispiele

- Vorkommen der Buchstaben „ss“

```
SQL> select last_name from employees where
    regexp_like(last_name,'(ss)');
LAST_NAME
-------------------------
Bissot
Russell
Weiss
SQL> select last_name from employees where
    regexp_like(last_name,'(s{2})');
LAST_NAME
-------------------------
Bissot
Russell
Weiss
```

Beispiele

- Vorkommen der Buchstaben „B“ oder „R“ und „ss“
- .* unbekannte Anzahl von Zeichen
- {n} Anzahl der Zeichen

```
SQL> select last_name from employees where
    regexp_like(last_name,'[BR].*(s{2})');
LAST_NAME
-------------------------
Bissot
Russell
```

Beispiele

- Beginn mit „W“ und enden mit „s“ oder „n“
- .* unbekannte Anzahl von Zeichen
- ^ Führende Zeichenkette
- $ Endende Zeichenkette

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[W].*[sn]$');
LAST_NAME
-------------------------
Weiss
Whalen
```

Beispiele

- Beginn mit „W“ und enden mit „ss“
- .* unbekannte Anzahl von Zeichen
- ^ Führende Zeichenkette
- $ Endende Zeichenkette

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[W].*(ss)$');
LAST_NAME
-------------------------
Weiss
```

Beispiele

- Beginn mit „W“ und enden mit „ss“ oder „en“
- .* unbekannte Anzahl von Zeichen
- ^ Führende Zeichenkette
- $ endende Zeichenkette
- | oder

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[W].+((ss)|(en))$');
LAST_NAME
-------------------------
Weiss
Whalen
```

Beispiele

- Beinhalten von „W“ oder „T“ gefolgt von „a“ oder „o“

```
SQL> select last_name from employees where
    regexp_like(last_name,'[WT][ao]');
LAST_NAME
-------------------------
Taylor
Taylor
Tobias
Walsh
```

Beispiele

- Beginn mit „V“ bis „Z“
- ^ Beginn

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[V-Z]');
LAST_NAME
-------------------------
Vargas
Vishney
Vollman
Walsh
Weiss
Whalen
Zlotkey
```

Beispiele

- Beginnen mit „W“ oder „T“ und enden mit „r“ oder „s“

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[WT].*[rs]$');
LAST_NAME
-------------------------
Taylor
Taylor
Tobias
Tucker
Weiss
```

Beispiele

- Beginnen mit „W“ oder „T“ nicht gefolgt von einem „a“ und enden mit „r“ oder „s“

```
SQL> select last_name from employees where
    regexp_like(last_name,'^[WT][^a].*[rs]$');
LAST_NAME
-------------------------
Tobias
Tucker
```

Beispiele

- Lösche die letzten numerischen Werte

```
SQL> select email from employees where rownum < 3;
EMAIL
-------------------------
ABANDA6
ABULL5

SQL> update employees set email=regexp_replace(email,'[1-9]$','');
107 Zeilen wurden aktualisiert

SQL> select email from employees where rownum < 3;
EMAIL
-------------------------
ABANDA
ABULL
```

Beispiele

- Anders: Lösche die letzten numerischen Werte

```
SQL>  update employees set
    email=regexp_replace(email,'[.*[:digit:]]{1}$','') ;

SQL> select email from employees where rownum < 3;

EMAIL
-------------------------
ABANDA
ABULL
```

Beispiele

- Ausschneiden bis „lo“ oder „al“

```
SQL> select last_name,regexp_substr(last_name,'.*(lo|al)') from
     employees;

LAST_NAME                 REGEXP_SUBSTR(LAST_NAME,'
------------------------- -------------------------
Taylor                    Taylo
Taylor                    Taylo
Tobias
Vollman
Walsh                     Wal
Weiss
Whalen                    Whal
Zlotkey                   Zlo
```

Beispiele

- Verwendung als Constraint

```
alter table kunden
add constraint ch_email_gueltig check
(regexp_like(email, '^[A-Za-z0-9._%-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,4}$'))
```

Neuerungen in 11g

- RGEXP_COUNT
  - Zählen von Zeichenvorkommen
- REGEXP_INSTR/REGEXP_SUBSTR
  - Zusatzparameter für die Einschränkung eines Suchpatterns

Beispiele

- Zählen von Zeichenvorkommen

```
SQL> select last_name,regexp_count(last_name,'(s)') from employees;
LAST_NAME                 REGEXP_COUNT(LAST_NAME,'(S)')
------------------------- -----------------------------
Urman                                                 0
Vargas                                                1
Vishney                                               1
Vollman                                               0
Walsh                                                 1
Weiss                                                 2
Russell                                               2
Whalen                                                0
Zlotkey                                               0
```

Beispiele

- Zählen von Zeichenvorkommen

```
SQL> select last_name,regexp_count(last_name,'(ss)') from employees;
LAST_NAME                 REGEXP_COUNT(LAST_NAME,'(SS)')
------------------------- ------------------------------
Urman                                                  0
Vargas                                                 0
Vishney                                                0
Vollman                                                0
Walsh                                                  0
Weiss                                                  1
Russell                                                1
Whalen                                                 0
Zlotkey                                                0
```

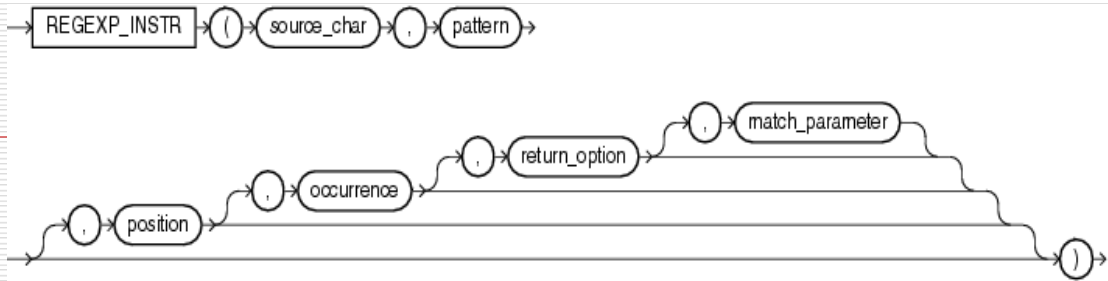
Beispiele

- Zählen von Zeichenvorkommen

```
SQL> select last_name,regexp_count(last_name,'(ss)$') from employees;
LAST_NAME                 REGEXP_COUNT(LAST_NAME,'(SS)')
------------------------- ------------------------------
Urman                                                  0
Vargas                                                 0
Vishney                                                0
Vollman                                                0
Walsh                                                  0
Weiss                                                  1
Russell                                                0
Whalen                                                 0
Zlotkey                                                0
```

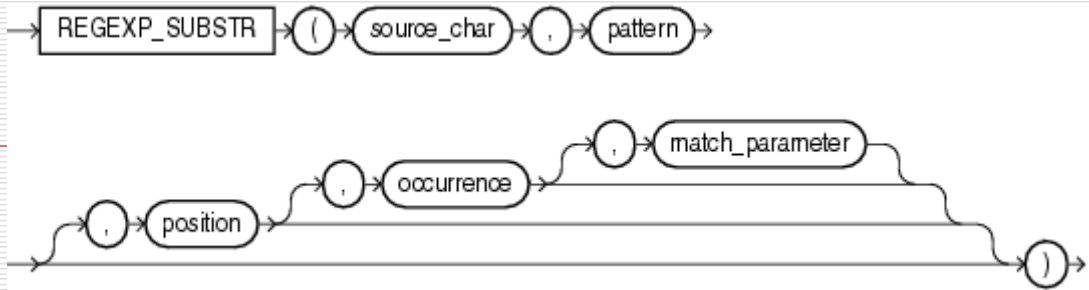
REGEXP_INSTR

```
SQL> select regexp_instr(last_name,'(A)(be)(l)',1,1,0,'i',1) from
employees where last_name='Abel';
REGEXP_INSTR(LAST_NAME,'(A)(BE)(L)',1,1,0,'I',1)
------------------------------------------------
                                               1
SQL> select regexp_instr(last_name,'(A)(be)(l)',1,1,0,'i',2) from
employees where last_name='Abel';
REGEXP_INSTR(LAST_NAME,'(A)(BE)(L)',1,1,0,'I',2)
------------------------------------------------
                                               2
SQL> select regexp_instr(last_name,'(A)(be)(l)',1,1,0,'i',3) from
employees where last_name='Abel';
REGEXP_INSTR(LAST_NAME,'(A)(BE)(L)',1,1,0,'I',3)
------------------------------------------------
                                               4
```



REGEXP_SUBSTR

```
SQL> select regexp_substr(last_name,'(A)(be)(l)',1,1,'i',1) from
employees where last_name='Abel';
REGEXP_SUBSTR(LAST_NAME,'
-------------------------
A
SQL> select regexp_substr(last_name,'(A)(be)(l)',1,1,'i',2) from
employees where last_name='Abel';
REGEXP_SUBSTR(LAST_NAME,'
-------------------------
be
SQL> select regexp_substr(last_name,'(A)(be)(l)',1,1,'i',3) from
employees where last_name='Abel';
REGEXP_SUBSTR(LAST_NAME,'
-------------------------
l
```



Fragen

Reguläre Ausdrücke

- Zusammenfassung
  - Grundlagen
  - Funktionen
  - Ausdrücke
  - Beispiele
  - Übung