



ΠΕΡΙΕΧΟΜΕΝΑ:

1. Ταίριασμα Συμβολοσειρών
 1. LIKE
 2. REGEXP

Κωνσταντίνος Λ.

Ασημένιος Χορηγός Μαθήματος

Μαρία Νάνου

Χάλκινος Χορηγός Μαθήματος

ΜΑΘΗΜΑ 1.4: Ταίριασμα Συμβολοσειρών

Ασχολούμαστε με μια ειδική οικογένεια συνθηκών (WHERE clause):

- Με τις οποίες μπορούμε να αναζητήσουμε για συμβολοσειρές που έχουν μια συγκεκριμένη μορφή
- Λέμε ότι κάνουμε **“ταίριασμα συμβολοσειρών” (string matching)**
- Συντακτικό:

```
WHERE string_column [NOT] LIKE pattern
```

- **string column:** Στήλη (ή υπολογιζόμενη στήλη) που είναι συμβολοσειρά
- **pattern:** συμβολοσειρά που περιέχει σύμβολα και τους **χαρακτήρες μπαλαντέρ (wildcards)**:
 - **_** (underscore): Ταιριάζει με ακριβώς ένα χαρακτήρα
 - **%**: Ταιριάζει με πολλούς χαρακτήρες (0,...,N)

1.1. Χρήση του underscore:

-- example 1.1 (DB: world, script: like.sql)

```
SELECT *
FROM country
WHERE Code LIKE 'GR_';
```

-- example 1.2

```
SELECT *
FROM country
WHERE Code LIKE '__C';
```

1.1. LIKE

1.2. Χώρες που αρχίζουν από 'Α'

-- example 1.3

```
SELECT *
FROM country
WHERE name LIKE 'A%';
```

1.3. Χώρες που περιέχουν 'ΑΝ' και τελειώνουν με 'Ε'

-- example 1.4

```
SELECT *
FROM country
WHERE name LIKE '%AN%E';
```

Άσκηση 1: πίνακας city

- 1) Μετρήστε πόσες πόλεις αρχίζουν από Α
- 2) Εμφανίστε τις πόλεις που το 2ο γράμμα τους είναι Ρ και τελειώνουν σε Α
- 3) Εμφανίστε τις πόλεις που αρχίζουν από Μ ή Ν περιέχουν τουλάχιστον ένα Ο και τελειώνουν σε Α.
- 4) Εμφανίστε τις πόλεις που το όνομα τους αποτελείται από ακριβώς 5 χαρακτήρες.

Η **REGEXP** (< Regular Expression = Κανονική Έκφραση):

WHERE string_column [NOT] **REGEXP** pattern

- Αναζητά πρότυπα όπως η LIKE αλλά έχει πολύ πιο πλούσια εκφραστική δυνατότητα.



A. Ιδιότητες: αρχίζει - περιέχει - τελειώνει:

Ιδιότητα	Πρότυπο
Αρχίζει	'^str'
Περιέχει	'str'
Τελειώνει	'str\$'

όπου str είναι μία συμβολοσειρά

2.1. Πόλεις που αρχίζουν από 'ar'

-- examples 2.1 (DB: world, script: Regexp)

SELECT * FROM city **WHERE** name **REGEXP** '^ar';

2.2. Πόλεις που περιέχουν 'me'

SELECT * FROM city **WHERE** name **REGEXP** 'me';

2.3. Πόλεις που τελειώνουν με 'nia'

SELECT * FROM city **WHERE** name **REGEXP** 'nia\$';

B. Ταίριασμα με ένα χαρακτήρα:

Ιδιότητα	Πρότυπο
Ακριβώς ένας χαρακτήρας	.

2.4. Πόλεις που το 2ο γράμμα τους είναι 'r'

SELECT * FROM city **WHERE** name **REGEXP** '^.r';

Σημείωση:

- Πάντα οι κανονικές εκφράσεις διαβάζονται από αριστερά προς τα δεξιά και εκφράζουν τη διάταξη της συμβολοσειράς

Άσκηση 2: πίνακας city

Επιστρέψτε τις πόλεις που

- Αρχίζουν με 'A', ακολουθούν 2 χαρακτήρες και τελειώνουν με 'n'
- Περιέχουν το 'b' και το 'e' με αυτή τη σειρά και μεταξύ τους βρίσκονται 3 ακριβώς γράμματα.
- Περιέχουν τουλάχιστον 3 γράμματα
- Περιέχουν ακριβώς 3 γράμματα
- Περιέχουν το πολύ 3 γράμματα

Επανάληψη (Συνεχόμενων) Χαρακτήρων:

Επανάληψη	Τελεστής
0 ή περισσότερες φορές	'c*'
1 ή περισσότερες φορές	'c+'
το πολύ 1 φορά	'c?'
Ακριβώς n φορές	'c{n}'
Τουλάχιστον n φορές	'c{n,}'
n έως m φορές	'c{n,m}'

όπου c είναι ένας χαρακτήρας

- και μπορούμε να το εφαρμόσουμε και σε πολλούς χαρακτήρες αρκεί να τους ενθέσουμε σε παρενθέσεις, π.χ. (str)*

3.1. Πόλεις που αρχίζουν με 'b', περιέχουν το 'e' και τελειώνουν με 'a'.

-- examples 2.2 (DB: world, script: Regexp)

```
SELECT * FROM city WHERE name REGEXP '^b.*e.*a$';
```

3.2. Πόλεις που περιέχουν 'ab' τουλάχιστον 2 φορές

```
SELECT * FROM city WHERE name REGEXP 'ab.*ab';
```

3.3. Πόλεις που περιέχουν το 'abb'

```
SELECT * FROM city WHERE name REGEXP 'abb';
```

```
SELECT * FROM city WHERE name REGEXP 'ab{2}';
```

3.4. Πόλεις που το πλήθος των γραμμάτων τους είναι πολλαπλάσιο του 5.

```
SELECT * FROM city WHERE name REGEXP '^(.{5})*$';
```

Άσκηση 3: πίνακας city

Επιστρέψτε τις πόλεις που

- Αρχίζουν με 'b' και έχουν μήκος τουλάχιστον 8
- Περιέχουν 3 συνεχόμενες εμφανίσεις του 'a'
- Περιέχουν τουλάχιστον πέντε 'a' (συνεχόμενα ή μη)
- Έχουν τις ιδιότητες του ερωτήματος 1 και του ερωτήματος 3
- Έχουν τις ιδιότητες του ερωτήματος 1 και του ερωτήματος 3 και τελειώνουν με 'r'

Ομάδες Χαρακτήρων:

Περιορισμός	Επιλογή χαρακτήρων	Μη Επιλογή Χαρακτήρων
Επιλογή από σύνολο χαρακτήρων	Π.χ. [abc]	Π.χ. [^abc]
Εύρος ψηφίων	Π.χ. [0-9]	Π.χ. [^0-9]
Εύρος χαρακτήρων	Π.χ. [a-z]	Π.χ. [^a-z]

Παρατηρήσεις:

- Η παράσταση θα γίνει ταίριασμα με ένα χαρακτήρα. Έτσι π.χ. το [abc] μπορεί να ταιριάζει με τα a ή b ή c

Ειδικοί Χαρακτήρες:

- Οι χαρακτήρες: ., [,], (,), {, }, ^, \$, |, + χρησιμοποιούνται στο συντακτικό των κανονικών εκφράσεων.
- Αν θέλουμε να τις ενσωματώσουμε σε ερωτήματα, πρέπει να προηγηθεί διπλό backslash \\: π.χ. \\. για την τελεία.

4.1. Πόλεις που περιέχουν τουλάχιστον 3 συνεχόμενα φωνήεντα:

```
-- examples 2.3 (DB: world, script: Regexp)
SELECT * FROM city WHERE name REGEXP '[aeuiouy]{3}'
```

4.2. Πόλεις που έχουν μια τελεία στο όνομά τους:

```
SELECT * FROM city WHERE name REGEXP '\\.'
```

Κλάσεις Χαρακτήρων

- Ταιριάζει με έναν χαρακτήρα από μία κλάση χαρακτήρων. Αντικαθιστώντας στο [[:class:]], όπου class ένα από τα:

Περιορισμός	Παράσταση
alnum	Alphanumeric characters
alpha	Alphabetic characters
blank	Whitespace characters
cntrl	Control characters
digit	Digit characters
graph	Graphic characters
lower	Lowercase alphabetic characters
print	Graphic or space characters
punct	Punctuation characters
space	Space, tab, newline, and carriage return
upper	Uppercase alphabetic characters
xdigit	Hexadecimal digit characters

4.3. Πόλεις που δεν περιέχουν μόνο αλφαβητικούς χαρακτήρες:

```
SELECT * FROM city WHERE name NOT REGEXP '^[[:alpha:]]+$'
```

4.4. Πόλεις που είναι μία λέξη, και δεν ξεκινούν από A-K και N-Z:

```
SELECT * FROM city WHERE name REGEXP '^[^A-KN-Z][a-z]+$'
```

Διάζευξη Συμβολοσειρών (|):

Ιδιότητα	Πρότυπο
str1 ή str2	'str1 str2'

5.1. Πόλεις που περιέχουν ree ή ceb:

-- examples 2.4 (DB: world, script: Regexp)

SELECT * FROM city **WHERE** name **REGEXP** 'ree|ceb'

(όλες οι ασκήσεις να γίνουν επί της ΒΔ: classicmodels)

Άσκηση 4: Πίνακας customers

Βρείτε τα ονόματα πελατών (customerName), που αρχίζουν από B και το όνομα τους περιέχει τη συμβ/ρα "Co." ή τη συμβολοσειρά "Inc." (προσοχή ότι η τελεία μπορεί και να μην υπάρχει π.χ. "Co" ή "Inc")

Άσκηση 5: Πίνακας customers

Βρείτε τα επώνυμα επαφών (contactLastName), που το επώνυμό τους ξεκινά από φωνήεν και είναι τουλάχιστον 7 γράμματα

Άσκηση 6: Πίνακας employees

Ελέγξτε αν τα e-mails των υπαλλήλων είναι έγκυρα. Πρέπει να είναι αλφαβητικοί χαρακτήρες (τουλάχιστον 2), μετά το @, μετά αλφαβητικοί χαρακτήρες (τουλάχιστον 3), μετά η ., μετά το com.

Άσκηση 7: Πίνακας employees

Ο μάνατζερ μας, αναζητά έναν υπάλληλο του οποίου το επώνυμο έχει δύο όμοια συνεχόμενα σύμφωνα, το μικρό του όνομα ξεκινά από L και το τηλέφωνο του περιέχει το 31 ή το 13 (δεν θυμάται ακριβώς).

Βοηθήστε τον μάνατζερ, δίνοντας του τα στοιχεία του υπαλλήλου που αναζητά.

Άσκηση 8: Πίνακας payments

Ομαδοποιήστε με βάση τον αριθμό πελάτη, το άθροισμα των ποσών των πληρωμών του.

- Ταξινομήστε τα αποτελέσματα σε φθίνουσα σειρά αθροίσματος ποσών.
- Στον υπολογισμό να συνεκτιμηθούν μόνο εκείνες οι πληρωμές με checkNumber που αποτελείται από δύο γράμματα ακολουθούμενο από πέντε αριθμούς.
- Επίσης θέλουμε να εμφανιστούν τελικά μόνο οι κωδικοί πελατών που το άθροισμα των ποσών είναι περισσότερο από 100.000.