

Wyrażenia regularne i walidacja danych

Czas: 2 × 45 min

praca indywidualna lub w parach

Cel: Nauczenie się wyszukiwania danych według wzorca, wykrywania i poprawiania błędnych e-maili, telefonów i kodów pocztowych oraz blokowania złych danych ograniczeniem CHECK z REGEXP.

Po zajęciach potrafisz:

- zapisuje wzorce z kotwicami ^ \$, klasami znaków, alternatywą i kwantyfikatorami
- wykrywa błędne dane zapytaniem z NOT REGEXP
- normalizuje dane funkcjami REGEXP_REPLACE, REGEXP_SUBSTR i REGEXP_INSTR
- zabezpiecza tabelę ograniczeniem CHECK z wyrażeniem regularnym
- rozróżnia porównanie bez i z rozróżnianiem wielkości liter (COLLATE utf8mb4_bin)

Sprzęt i materiały

- komputer w pracowni z XAMPP (MariaDB 10.4) lub MySQL 8.4
- baza sklep zaimportowana z pliku `sklep.sql` (folder Pliki_do_zadan)
- plik `r2_07_zgloszenia_brudne.sql` z folderu Pliki_do_zadan (tworzy bazę rejestracja)

Oprogramowanie

- **XAMPP (Apache + MariaDB + phpMyAdmin)** — panel sterowania XAMPP, konsola: przycisk Shell albo `C:\xampp\mysql\bin\mysql.exe -u root`
apachefriends.org/download.html
- **phpMyAdmin** — zakładka SQL do uruchamiania zapytań; wchodzi w skład XAMPP
localhost/phpmyadmin
- **VS Code lub Notepad++** — do pisania pliku `.sql` z zapytaniami
code.visualstudio.com

Bezpieczeństwo

- Dane poprawiaj tylko w kopii tabeli `zgloszenia_czyste` (Z9). Tabela `zgloszenia` zostaje bez zmian, żeby można było porównać wyniki i powtórzyć czyszczenie.
- Przed każdym UPDATE uruchom SELECT z tym samym warunkiem i sprawdź, które wiersze się zmieniają.
- Dane w pliku są fikcyjne. Prawdziwych danych osobowych z formularzy nie kopiuj na prywatne komputery ani do zrzutów ekranu w sprawozdaniu.

Przebieg zadania

1 Wzorce na bazie sklep

kolumna REGEXP 'wzorzec' zwraca 1, gdy wzorzec pasuje do dowolnego fragmentu tekstu. Kotwice `^` i `$` wymuszają dopasowanie od początku i do końca.

1. Zaimportuj `sklep.sql` i `r2_07_zgloszenia_brudne.sql`. Utwórz plik `r2_07_nazwisko.sql`.
2. Z1. W bazie sklep policz produkty, których nazwa zawiera cyfrę. Następnie wypisz produkty, których nazwa zaczyna się od „Laptop” albo „Monitor” (jeden wzorzec z alternatywą `|`).
3. Z2. Wypisz monitory z przekątną zapisaną jako dwie cyfry i cudzysłów (np. 27") oraz samą przekątną wyciętą funkcją REGEXP_SUBSTR.
4. Z3. Wypisz adresy e-mail klientów w formacie „litera.litera@...” (kropka między dwoma ciągami liter przed znakiem @).
5. Z4. Porównaj wyniki `'Nowak' REGEXP '^nowak$'` i `'Nowak' COLLATE utf8mb4_bin REGEXP '^nowak$'`. Wyjaśnij różnicę.

2 Wykrywanie błędów (baza rejestracja)

Wzorzec e-maila do tego zadania (uproszczony, ale wystarczający do wyłapania typowych pomyłek):

```
^[a-z0-9._%+-]+@[a-z0-9.-]+[.][a-z]{2,}$
```

1. Z5. `USE rejestracja;` Wypisz zgłoszenia z e-mailem niezgodnym ze wzorcem z ramki oraz osobno e-maile z dwiema kropkami obok siebie. Dlaczego zgłoszenia 4 i 19 też trafiły na listę?
2. Z6. Wypisz zgłoszenia z kodem pocztowym w złym formacie (poprawny: dwie cyfry, myślnik, trzy cyfry).
3. Z7. Dla każdego zgłoszenia pokaż telefon, same cyfry telefonu (REGEXP_REPLACE usuwający wszystko poza cyframi) i ich liczbę. Które numery nie dadzą się poprawić automatycznie?
4. Z8. Wypisz osoby, których imię lub nazwisko nie ma postaci „wielka litera + małe litery” (użyj COLLATE utf8mb4_bin i polskich liter w klasach znaków). Funkcją REGEXP_INSTR wskaż pozycję cyfry w nazwisku.

3 Czyszczenie danych w kopii

1. Z9. Utwórz kopię: `CREATE TABLE zgloszenia_czyste LIKE zgloszenia;` i przepis� dane `INSERT ... SELECT`. Jednym UPDATE popraw imiona i nazwiska (usuń spacje, pierwsza litera wielka, reszta mała) oraz e-maile (usuń spacje z brzegów, małe litery). Ile wierszy się zmieniło?
2. Z10. Ujednolicić telefony do 9 cyfr: najpierw zostaw same cyfry, potem z numerów zaczynających się od 48 lub 0048 i mających po nich 9 cyfr zostaw 9 ostatnich (RIGHT). Numery, które nadal nie mają 9 cyfr, ustaw na NULL.
3. Z11. Ujednolicić kody pocztowe: jeśli po usunięciu znaków innych niż cyfry zostaje 5 cyfr, zapisz je w postaci 00-000. Który kod został błędny i dlaczego?
4. Z12. Po czyszczeniu znajdź duplikaty zgłoszeń (ten sam e-mail więcej niż raz).

4 Ograniczenia CHECK

CHECK z REGEXP działa w MariaDB od 10.2 i w MySQL od 8.0.16. Wartość NULL nie narusza ograniczenia.

1. Z13. Dodaj do `zgloszenia_czyste` ograniczenie `chk_kod` sprawdzające format kodu pocztowego. Co się stało i dlaczego? Ustaw błędne kody na NULL i dodaj ograniczenie ponownie. Spróbuj wstawić zgłoszenie z kodem 12345 — zapisz kod błędu.
2. Z14. Błędne e-maile ustaw na NULL (do ręcznego wyjaśnienia), a potem dodaj ograniczenie `chk_email` (wzorzec z ramki i brak dwóch kropek). Sprawdź wstawienie e-maila `jan.test@example` (błąd) i `jan.test@example.pl` (poprawne).

W sprawozdaniu umieść

- ☐ plik `.sql` z poleceniami Z1–Z14
- ☐ zrzuty ekranu wyników Z5, Z7 oraz błędów z Z13 i Z14
- ☐ tabela błędów wykrytych w danych i sposobu ich poprawienia
- ☐ odpowiedzi na pytania kontrolne

Pytania kontrolne

1. Czym różni się `nazwa LIKE '%7%'` od `nazwa REGEXP '7'`, a czym `nazwa REGEXP '^7$'`?
2. Dlaczego bez `COLLATE utf8mb4_bin` wzorzec `^[A-Z][a-z]+$` nie wykryje imienia „piotr”?
3. Dlaczego wzorzec e-maila w REGEXP nie gwarantuje, że adres istnieje? Jak aplikacja może to sprawdzić?
4. Dlaczego przed dodaniem CHECK trzeba najpierw poprawić istniejące dane?

Kryteria oceny

Kryterium	Punkty
Wzorce na bazie sklep (Z1–Z4)	4
Wykrywanie błędnych danych (Z5–Z8)	5
Czyszczenie danych w kopii (Z9–Z12)	5
Ograniczenia CHECK z REGEXP (Z13–Z14)	3
Odpowiedzi na pytania kontrolne	3
Razem	20

Sprawozdanie przygotuj na wzorze dołączonym do zadania (plik .docx o tym samym numerze).