

Oracle Text

Operator CONTAINS - Podstawy

1. Utwórz w swoim schemacie kopię tabeli CYTATY ze schematu ZTPD.

2. Znajdź w tabeli CYTATY za pomocą standardowego operatora LIKE cytaty, które zawierają zarówno słowo ' optymista ' jak i ' pesymista ' ignorując wielkość liter.

AUTOR	TEKST
-----	-----
Winston Churchill	Pesymista widzi trudności w każdej...

3. Utwórz indeks pełnotekstowy typu CONTEXT na kolumnie TEKST tabeli CYTATY przy domyślnych preferencjach dla tworzonego indeksu.

4. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają zarówno słowo ' optymista ' jak i ' pesymista ' (ignorując wielkość liter w tym i kolejnych zapytaniach ze względu na charakterystykę indeksu).

AUTOR	TEKST
-----	-----
Winston Churchill	Pesymista widzi trudności w każdej...

5. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowo ' pesymista ', a nie zawierają słowa ' optymista '.

AUTOR	TEKST
-----	-----
Oscar Wilde	Pesymista to ktoś, kto postawiony...

6. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowa ' optymista ' i ' pesymista ' w odległości maksymalnie 3 słów.

AUTOR	TEKST
-----	-----

7. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowa ' optymista ' i ' pesymista ' w odległości maksymalnie 10 słów.

AUTOR	TEKST
-----	-----
Winston Churchill	Pesymista widzi trudności w każdej...

8. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowo ' życie ' i jego odmiany. Niestety Oracle nie wspiera stemmingu dla języka polskiego. Dlatego zamiast frazy '\$ życie ' „poratujemy się” szukaniem frazy ' życi% '.

AUTOR	TEKST
-----	-----
...	

6 rows selected

9. Zmodyfikuj poprzednie zapytanie, tak by dla każdego pasującego cytatu wyświetlony został stopień dopasowania (SCORE).

10. Zmodyfikuj poprzednie zapytanie, tak by wyświetlony został tylko najlepiej pasujący cytat (w przypadku „remisu” może zostać wyświetlony dowolny z najlepiej pasujących cytatów).

AUTOR	TEKST	DOPASOWANIE
Herodot	Całe nasze życie to...	11

11. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowo ‘problem’ za pomocą wzorca z „literówką”: ‘probelm’.

AUTOR	TEKST
Paulo Coelho	Kiedy już rozwiążesz jakiś problem...
Albert Einstein	Nie jestem ... nad problemem...

12. Wstaw do tabeli CYTATY cytat Bertranda Russella 'To smutne, że głupcy są tacy pewni siebie, a ludzie rozsądni tacy pełni wątpliwości.'. Zatwierdź transakcję.

13. Znajdź w tabeli CYTATY za pomocą operatora CONTAINS cytaty, które zawierają słowo ‘głupcy’. Jak wyjaśnisz wynik zapytania?

AUTOR	TEKST
-------	-------

14. Odszukaj w swoim schemacie tabelę, która zawiera zawartość indeksu odwróconego na tabeli CYTATY. Wyświetl jej zawartość zwracając uwagę na to, czy słowo ‘głupcy’ znajduje się wśród poindeksowanych słów.

15. Indeks CONTEXT utworzony przy domyślnych preferencjach nie jest uaktualniany na bieżąco. Możliwa jest synchronizacja na żądanie (poprzez procedurę) lub zgodnie z zadaną polityką (poprzez preferencję ustawioną przy tworzeniu indeksu: po zatwierdzeniu transakcji, z zadanym interwałem czasowym). Można też przebudować indeks usuwając go i tworząc ponownie. Wadą tej opcji jest czas trwania operacji i czasowa niedostępność indeksu, ale z tej opcji skorzystamy ze względu na jej prostotę.

16. Sprawdź czy po przebudowaniu indeksu słowo ‘głupcy’ pojawiło się w indeksie odwróconym, a następnie powtórz zapytanie z punktu 13.

AUTOR	TEKST
Bertrand Russell	To smutne, że głupcy są tacy pewni siebie ...

17. Usuń indeks na tabeli CYTATY, a następnie samą tabelę CYTATY.

Zaawansowane indeksowanie i wyszukiwanie

1. Utwórz w swoim schemacie kopię tabeli QUOTES ze schematu ZTPD.
2. Utwórz indeks pełnotekstowy typu CONTEXT na kolumnie TEXT tabeli QUOTES przy domyślnych preferencjach.
3. Tabela QUOTES zawiera teksty w języku angielskim, dla którego Oracle Text obsługuje stemming. Sprawdź działanie operatora CONTAINS dla wzorców:
 - 'work'
 - '\$work'
 - 'working'
 - '\$working'
4. Spróbuj znaleźć w tabeli QUOTES wszystkie cytaty, które zawierają słowo 'it'. Czy system zwrócił jakieś wyniki? Dlaczego?
5. Sprawdź jakie stop listy dostępne są w systemie. Odpytaj w tym celu perspektywę słownikową CTX_STOPLISTS. Jak myślisz, którą system wykorzystywał przy dotychczasowych zapytaniach?
6. Sprawdź jakie słowa znajdują się na domyślnej stop liście. Odpytaj w tym celu perspektywę słownikową CTX_STOPWORDS.
7. Usuń indeks pełnotekstowy na tabeli QUOTES. Utwórz go ponownie wskazując, że przy indeksowaniu ma być użyta dostępna w systemie pusta stop lista.
8. Ponów zapytanie o wszystkie cytaty, które zawierają słowo 'it'. Czy tym razem system zwrócił jakieś wyniki?
9. Znajdź w tabeli QUOTES cytaty zawierające słowa 'fool' i 'humans'.
10. Znajdź w tabeli QUOTES cytaty zawierające słowa 'fool' i 'computer'.
11. Spróbuj znaleźć w tabeli QUOTES cytaty zawierające słowa 'fool' i 'humans' w jednym zdaniu. Zinterpretuj komunikat o błędzie.
12. Usuń indeks pełnotekstowy na tabeli QUOTES.
13. Utwórz grupę sekcji bazującą na NULL_SECTION_GROUP, zawierającą dodatkowo obsługę zdań i akapitów jako sekcji.
14. Utwórz ponownie indeks pełnotekstowy na tabeli QUOTES wskazując utworzoną grupę sekcji obsługującą zdania i akapity.
15. Sprawdź czy teraz działają wzorce odwołujące się do zdań szukając najpierw cytatów zawierających w tym samym zdaniu słowa 'fool' i 'humans', a następnie 'fool' i 'computer'.
16. Znajdź w tabeli QUOTES wszystkie cytaty, które zawierają słowo 'humans'. Czy system zwrócił też cytaty zawierające 'non-humans'? Dlaczego?

17. Usuń indeks pełnotekstowy na tabeli QUOTES. Utwórz preferencję dla leksera (używając BASIC_LEXER), wskazującą, że myślnik ma być traktowany jako część indeksowanych tokenów (składnik słów tak jak litery). Utwórz ponownie indeks pełnotekstowy na tabeli QUOTES wskazując utworzoną preferencję dla leksera.

18. Ponów zapytanie o wszystkie cytaty, które zawierają słowo 'humans'. Czy system tym razem zwrócił też cytaty zawierające 'non-humans'?

19. Znajdź w tabeli QUOTES wszystkie cytaty, które zawierają frazę 'non-humans'. Wskazówka: myślnik we wzorcu należy „escape’ować” („skorzystać z sekwencji ucieczki”).

20. Usuń swoją kopię tabeli QUOTES i utworzoną preferencję.