Министерство науки и высшего образования Российской Федерации

ФГАОУ ВО «УрФУ имени первого Президента России Б. Н.Ельцина»

Институт радиоэлектроники и информационных технологий – РтФ

Школа бакалавриата

Департамент информационных технологий и автоматики

Использование FTS в PostgreSQL

Отчет по лабораторной работе №2

|  |  |
| --- | --- |
| Преподаватель | С.Г. Мировода |
| Студенты группы  РИ-460004 | Дьячкова А.С |
| Ваганова Е.А  Козырев Н.В |
|  |

Екатеринбург

2020

**1. Тип tsvector**

Выполните запрос

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog.');

В ответ будет возвращён список [лексем](https://en.wikipedia.org/wiki/Lexeme)

to\_tsvector

'brown': 3 'dog': 9 'fox': 4 'jump': 5 'lazi': 8 'quick': 2

Задание 1

1. Изучите документацию к функции to\_tsvector

*to\_tsvector разбирает текстовый документ на фрагменты, сводит фрагменты к лексемам и возвращает значение tsvector, в котором перечисляются лексемы и их позиции в документе. При обработке документа используется указанная конфигурация текстового поиска или конфигурация по умолчанию.*

1. Вызовите эту функцию для следующей строки: Съешь ещё этих мягких французских булок, да выпей чаю

*'булок':6 'вып':8 'ещ':2 'мягк':4 'съеш':1 'французск':5 'ча':9 'эт':3*

1. Почему в векторе нет слова да?

*Некоторое слова распознаются как стоп-слова и игнорируются как слова, фигурирующие в тексте настолько часто, что искать их бессмысленно. В нашем примере это «Да». Если фрагмент не воспринимается ни одним словарём из списка, он так же игнорируется.*

**2. Тип tsquery**

Выполните по очереди запросы

-- №1

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('fox');

***Ответ: True***

-- №2

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('foxes');

***Ответ: True***

-- №3

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('foxhound');

***Ответ: False***

Задание 2

1. Что означают символы @@

*Оператор @@ - используется для сопоставления*

1. Почему второй запрос возвращает true, а третий не возвращает

*Так как один из словарей понимает, что слово foxes – это слово fox во множественном числе.*

1. Выполните запрос

SELECT to\_tsvector('Russian', 'Съешь ещё этих мягких французских булок, да выпей чаю.')

@@ to\_tsquery('Russian','булка');

Почему слово булка не находится?

*Потому что слово булок и булка имеют разную лексему*

4. Используйте функцию select ts\_lexize('russian\_stem', 'булок'); для того чтобы понять почему.

5. Замените в предложении слово булок, на слово пирожков Выполните запросы

--№1

SELECT to\_tsvector('Russian', 'Съешь ещё этих мягких французских пирожков, да выпей чаю.')

@@ to\_tsquery('Russian','пирожки');

--№2

SELECT to\_tsvector('Russian', 'Съешь ещё этих мягких французских пирожков, да выпей чаю.')

@@ to\_tsquery('Russian','пирожок');

Почему первый запрос возвращает true, а второй не возвращает?

*Потому что слово пирожков и пирожок имеют разную лексему*

**3. Операторы**

Выполните запрос

-- И

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('fox & dog');

***Ответ: True***

- ИЛИ

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('fox | rat');

***Ответ: True***

- отрицание

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog')

@@ to\_tsquery('!clown');

***Ответ: True***

- группировка

SELECT to\_tsvector('The quick brown fox jumped over the lazy dog' )

@@ to\_tsquery('fox & (dog | rat) & !mice');

***Ответ: True***

Задание 3

1. Приведите аналогичные запросы для любого предложения на русском

И

SELECT to\_tsvector('Russian','Защита диплома в УрФУ состоится не смотря ни на что')

@@ to\_tsquery('Диплома & УрФУ');

ИЛИ  
SELECT to\_tsvector('Russian','Защита диплома в УрФУ состоится не смотря ни на что')  
@@ to\_tsquery('Диплома | СИНХ');

Отрицание  
SELECT to\_tsvector('Russian','Защита диплома в УрФУ состоится не смотря ни на что')  
@@ to\_tsquery('!СИНХ');

Группировка  
SELECT to\_tsvector('Russian','Защита диплома в УрФУ состоится не смотря ни на что')  
@@ to\_tsquery('Состоится & (Диплома | СИНХ)');

1. Почему для английского языка не нужно указывать язык в первом аргументе и какой анализатор используется если никакой не указан?

**4. Поиск фраз**

Изучите документацию по [операторам](https://www.postgresql.org/docs/current/functions-textsearch.html) FTS Выполните запрос

SELECT to\_tsvector('Russian', 'Съешь ещё этих мягких французских булок, да выпей чаю.')

@@ to\_tsquery('Russian','мягких<2>булок');

Задание 4

1. Что означает число 2 в операторе <->

*Число 2 означает, что слово «Мягких» находится через слово от слова «Булок». Это особенно полезно при поиске точных последовательностей лексем, так как оператор фразового поиска поддерживает порядок лексем.*

1. Модифицируйте запрос так, чтобы можно было найти фразу съешь ещё

*SELECT to\_tsvector('Russian', 'Съешь ещё этих мягких французских булок, да выпей чаю.') @@ to\_tsquery('Russian','Съешь<->ещё');*

1. Для чего нужно использовать функцию phraseto\_tsquery

*Для поиска фраз, игнорируя знаки препинания*

**5. Утилиты**

1. Приведите примеры использования функций ts\_debug и ts\_headline

*SELECT ts\_debug('Russian','Защита диплома в УрФУ состоится не смотря ни на что');  
SELECT ts\_headline('Russian', 'Защита диплома в УрФУ состоится не смотря ни на что' ,  
to\_tsquery('ЕГЭ & России'));*