Day 26
Working with inconsistencies in text data

Aditya Shenoy

Two most common inconsistency in text data :

- Differences in the **case** of characters.
- **Spaces** in a string.

We can deal with these issues by using functions to change character case or remove spaces .

Converting Case :

Easiest ways to handle inconsistencies in case is to convert data to either all upper or all lower case.

```sql
SELECT lower('hElLo woRlD 1!');
```

```
>> hello world 1!
```

```sql
SELECT upper('hElLo woRlD 1!');
```

```
>> HELLO WORLD 1!
```

Trimming spaces :

The **trim function** can be used to remove spaces from one or both ends of a string.

Trim() trims spaces from both ends.

```
SELECT trim(' sql ');
```

```
>> 'sql'
```

rtrim() and **ltrim()** trims spaces from end of string and beginning of string respectively.

```
SELECT rtrim(' sql '),
       ltrim(' sql ');
```

```
>> ' sql'
   'sql '
```

Aditya Shenoy

Using ILIKE to take care of both :

Suppose, we want to count the occurrence of 'Apple' from the following set of inputs :

```
SELECT *
FROM fruit;
```

	fav_fruit	
▶	apple	<········· #1
	Apple	<········· #2
	apple	<········· #3
	banana	
	banana	
	APPLES	<········· #4
	apple	<········· #5
	banana	
	apple	<········· #6
	BANANA	
	APPLE	<········· #7
	apple	<········· #8

There are **8** occurrences of '**Apple**' inputted in **6** different ways.

```
-- Using ILIKE


SELECT *
  FROM fruit
 -- ILIKE for case insensitive
 WHERE fav_fruit ILIKE '%apple%';
```

OUTPUT :

```
fav_fruit
-----------
apple
Apple
 apple
APPLES
apple
apple
APPLE
apple
```