



NYU

**TANDON SCHOOL
OF ENGINEERING**

Text Operations and Metadata

Torsten Suel
CSE Department
NYU Tandon



NEW YORK UNIVERSITY



NYU

TANDON SCHOOL
OF ENGINEERING

■ **Text operations in Oracle**

- **like: pattern matching, char-oriented**
- **contains: index-based, word-oriented**



NEW YORK UNIVERSITY



■ LIKE

- ... Where Person.Name LIKE ‘_essi%’
- Matches: ‘Bessie’, ‘Jessie’, ‘Jessica’
... but not ‘Dressica’
- _ (underscore) means one arbitrary CHAR
- % means an arbitrary STRING
- Can choose to normalize to upper or lower case
- Where P.Description LIKE ‘%pen%’ matches ‘Appendix’
- Usually used with varchar, on small strings





■ CONTAINS

- **Where CONTAINS (P.Description, 'T30', 1) > 0**
- **Word-oriented**
- **Used for clob data type, on larger pieces of text**
- **Much more powerful (includes ranking)**
- **Faster than LIKE in many cases because it uses index**
- **But must build index first (Create index ...)**





■ Metadata

- The dictionary or catalog stores information about the database itself.
- This is data about data or 'metadata'.
- Almost every aspect of a DBMS uses this metadata
- The dictionary holds:
 - Description of database objects (tables, users, rules, views, indexes,..)
 - Information about who is using which data (locks)
 - Schemas and mapping
 - The dictionary itself





NYU

TANDON SCHOOL
OF ENGINEERING

■ Metadata

- **DBMS table structure is more than the user schema.**
- **Most DBMS systems use tables internally too**
- **Good orthogonality**
- **Once table security is finished, everything is secure.**



NEW YORK UNIVERSITY



■ Query Exercises in mySQL

- Based on the following schema:

ACTOR(AID, FIRSTNAME, LASTNAME)

MOVIE(MID, MNAME, BUDGET, GROSS)

ACTED_IN(AID, MID, STARRING, WAGE)





NYU

TANDON SCHOOL
OF ENGINEERING

1. List all tables that exist in the text schema

```
SELECT *  
  
FROM INFORMATION_SCHEMA.TABLES  
  
WHERE TABLE_SCHEMA = 'TEST'  
  
LIMIT 0 , 30
```



NEW YORK UNIVERSITY



Result

TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE	ENGINE	VERSION	ROW_FORMAT	TABLE_ROWS
def	test	acted_in	BASE TABLE	InnoDB	10	Compact	0
def	test	actor	BASE TABLE	InnoDB	10	Compact	4
def	test	movie	BASE TABLE	InnoDB	10	Compact	0





NYU

TANDON SCHOOL
OF ENGINEERING

■ 2. List all tables that have more than 3 columns.

```
SELECT TABLE_NAME  
  
FROM INFORMATION_SCHEMA.COLUMNS  
  
WHERE TABLE_SCHEMA = 'TEST'  
  
GROUP BY TABLE_NAME  
  
HAVING COUNT(*) > 3
```












NEW YORK UNIVERSITY



NYU

TANDON SCHOOL
OF ENGINEERING

Result

					TABLE_NAME
<input type="checkbox"/>	 Edit	 Inline Edit	 Copy	 Delete	acted_in
<input type="checkbox"/>	 Edit	 Inline Edit	 Copy	 Delete	movie



NEW YORK UNIVERSITY



NYU

**TANDON SCHOOL
OF ENGINEERING**

3. List the first name and last name of all actors whose last name is equal to the name of an attribute in MOVIE

```
SELECT FIRSTNAME, LASTNAME  
  
FROM TEST.ACTOR, INFORMATION_SCHEMA.COLUMNS I  
  
WHERE TABLE_SCHEMA = 'TEST'  
  
AND I.TABLE_NAME = 'MOVIE'  
  
AND LASTNAME = I.COLUMN_NAME
```



NEW YORK UNIVERSITY



The Table Actor

					aid	firstname	lastname
<input type="checkbox"/>	Edit	Inline Edit	Copy	Delete	1	DAVID	BUDGET
<input type="checkbox"/>	Edit	Inline Edit	Copy	Delete	2	TOM	GROSS
<input type="checkbox"/>	Edit	Inline Edit	Copy	Delete	3	JOHN	SMITH
<input type="checkbox"/>	Edit	Inline Edit	Copy	Delete	4	TOM	MIKE














NYU

TANDON SCHOOL
OF ENGINEERING

Result

					FIRSTNAME	LASTNAME
<input type="checkbox"/>	 Edit	 Inline Edit	 Copy	 Delete	DAVID	BUDGET
<input type="checkbox"/>	 Edit	 Inline Edit	 Copy	 Delete	TOM	GROSS



NEW YORK UNIVERSITY



NYU

TANDON SCHOOL
OF ENGINEERING

More Exercises:

- 1. List names of the tables that contain a column called 'AID'.
- 2. List the tables that have the most columns.
- 3. Or list the tables with more rows than columns.



NEW YORK UNIVERSITY