

TP1: Intro BigData / SQL

Recherche textuelle / patterns

1/ Tous les films qui ont le mot **stardust** dans leur nom.

```
SELECT title FROM movies
WHERE LOWER(title) LIKE '%stardust%'
```

2/ Compter tous les films dont le titre ne commence pas par le mot **the**.

```
SELECT COUNT(title) as count FROM movies
WHERE LOWER(title) NOT LIKE 'the%'
```

```
>> 2211
```

3/Tous les films qui ont le mot **war** dans le titre mais pas en dernière position.

```
SELECT COUNT(title) FROM movies
WHERE LOWER(title) LIKE '%war%'
AND LOWER(title) NOT LIKE '%war'
```

Distance Levenshtein

1/ La distance Levenshtein entre les mots **execution** et **intention**.

```
SELECT levenshtein('execution', 'intention')
```

```
>> 5
```

2/ Tous les films qui sont a une distance levenshtein inférieure a 9 de la chaine suivante: **a hard day nght**.

```
SELECT title from movies
WHERE levenshtein(LOWER(title), 'a hard day nght') < 9
```

```
>> title
A Hard Day's Night
Hard Target
Dead of Night
```

N-gram, similarity search (%)

1/ Tous les tri-grammes du mot **Avatar**.

```
SELECT show_trgm('Avatar')
```

```
>> show_trgm
{" a", " av", "ar « ,ata,ava,tar,vat}
```

2/ La similarité entre **VOTKA** et **VODKA**.

```
SELECT similarity('VOTKA', 'VODKA')
```

```
>> similarity
0.333333
```

3/ Tous les films dont le titre est similaire a plus de 0.1 du titre **Avatar**.

```
SELECT title FROM movies
WHERE similarity(LOWER(title), 'avatar') > 0.1

>>title
Star Wars
Dark Star
Avalon
A Star Is Born
Pet Sematary
Far and Away
Hatari!
War and Peace
Avanti!
Ada
Star 80
Avatar

SELECT COUNT(title) FROM movies
WHERE similarity(LOWER(title), 'avatar') > 0.1

>> count
12
```

Full text search

Trouver les films qui contiennent les formes grammaticales des mots **night** et **day**

```
SELECT title
FROM movies
WHERE to_tsvector(title) @@ to_tsquery('english', 'night & day');

>> title
A Hard Day's Night
Six Days Seven Nights
Long Day's Journey Into Night
```

Recherche phonétique

1/ Trouver les films qui ont des acteurs dont les noms se prononcent pareil.

```
SELECT m.title FROM movies_actors AS ma1, actors AS a1, movies AS m
WHERE ma1.actor_id = a1.actor_id
AND m.movie_id = ma1.movie_id
AND EXISTS
(SELECT name FROM actors AS a2, movies_actors AS ma2
WHERE ma2.movie_id = ma1.movie_id
AND soundex(a2.name) = soundex(a1.name)
AND a2.name <> a1.name)
```

2/ Trouver les acteurs avec un nom similaire a Robin Williams, triés par similarité (combiner %, metaphone et levenshtein)