Searching for Text in MongoDB

CREATING A SIMPLE TEXT QUERY

Axel Sirota
MACHINE LEARNING ENGINEER
@AxelSirota

What We Are Going to Do: Do You Remember the Movie About the Band?

The End Result

```
> db.rent.find({$text: {$search: "spacious", $caseSensitive: true}}, {"name":1, "description":1,
"price":1, "_id":0, score: {$meta: "textScore"}}).sort({score: {$meta:
"textScore"}}).limit(1).pretty()
{
    "name" : "Private bath ❤️2 beds❤️Modern Home 2 in Times SQ",
    "description" : "My apartment is spacious and brand new  with new electronics and
furnitures ) . surrounded by restaurants and bars, blocks away from Times SQ.   This private
room has a  private and spacious bathroom, one queen size bed and full size bed. /
Spacious and VERY cozy beds  / Very RESPONSIVE hosts that live next door  / Spacious
living room and kitchen  No elevator. 3 flights up.  No more than 2 suitcases per person.",
    "price" : "$120.00",
    "score" : 1.0208333333333333
}
```

The $text Evaluation Operator

An Example Document

```
> db.rent.findOne()
{
    "_id" :
ObjectId("5e8936cd6347c0057a053363"),
    "id" : 2060,
    "listing_url" : "https://www.airbnb.com/
rooms/2060",
    "scrape_id" :
NumberLong("20200313233810"),
    "last_scraped" : "2020-03-14",
    "name" : "Modern NYC",
    "price" : “$100",
    "space" : "Lovely, spacious, sunny 1 BR
apartment in 6th Floor elevator building steps
from gorgeous Fort Tryon Park/The Cloisters ....
",
    "experiences_offered" : "none",
    "neighborhood_overview" : "",
    ...
}
```

- **# Listing URL**

- **# The neighborhood**
- **# Price!**

- **# Description**

Searching for a Neighborhood

```
> db.rent.find({"neighbourhood_cleansed":
"Tribeca"}).count()
193
> db.rent.find({"neighbourhood_cleansed":
"TribeCa"}).count()
0
```

◀ **# Without “C”**

◀ **# With “c”**

◀ **# It is exact match**

Searching for Tribeca


```
{
   "name" : "Tribeca Sunset Views",
   "space" : "This is an oversized studio. When you first walk in there is the bedroom which is private to the rest of the apartment as you can't see the bedroom from the rest of the apartment. You will go down a long hallway where you will first see the bathroom, then the kitchen that opens up to the dining/living room. ....",
   "neighbourhood_cleansed" : "Battery Park City"
}
```


- **# We wouldn't get this result if searching for exact match**
- **# However the name has "Tribeca"**
- **#The *real* neighborhood is not Tribeca.**

Examples of Full Text Search

Search Tribeca

Get back all apartments that have Tribeca somewhere

(startsWith search)

Search berries

Get back blueberries, raspberries, strawberries, etc...

(Stem search)

Creating a Text Index

For full text search we need an index

```
> db.rent.createIndex({space : "text"})
{
    "createdCollectionAutomatically" : false,
    "numIndexesBefore" : 1,
    "numIndexesAfter" : 2,
    "ok" : 1
}
```

But as the One ring, **only one can exist**

```
> db.rent.createIndex({space : "text"})
{
    "ok" : 0,
    "code" : 85,
    "codeName" : "IndexOptionsConflict"
}
```

Searching for Tribeca

```
> db.rent.find({$text: {$search: "Tribeca"}},
{"_id":0, "neighbourhood_cleansed":1, "price":1,
"name":1})
{ "name" : "Architect Loft,  btw Soho & Tribeca,
Private Roof", "neighbourhood_cleansed" :
"Tribeca", "price" : "$240.00" }
{ "name" : "Stunning Sundrenched Tribeca Loft",
"neighbourhood_cleansed" : "Tribeca", "price" :
"$217.00" },
...,
{ "name" : "Accessible King - Roll-in shower -
Tribeca Hotel", "neighbourhood_cleansed" :
"Tribeca", "price" : "$300.00" }
> db.rent.find({$text: {$search: "Tribeca"}},
{"_id":0, "neighbourhood_cleansed":1, "price":1,
"name":1}).count()
908
```

- **# Note the $text operator!**
- **# We do full text search on *"Tribeca"***
- **# Multiple results! Useful?**

Fields of the $text Operator

$search	**language**
CaseSensitive	**discriticSensitive**

What Does $search Do?

- Tokenizes
- Removes stop words
- Stems
- Sets to lowercase

Understanding $search

```
> db.rent.find({$text: {$search: "Tribeca"}}, {"_id":0, "name":1}).limit(5)
{ "name" : "Architect Loft,  btw Soho & Tribeca, Private Roof" }
{ "name" : "Stunning Sundrenched Tribeca Loft" }
{ "name" : "Stylish Suite & Excellent Service in Tribeca, NY" }
{ "name" : "Stylish Hotel & Excellent Service in Tribeca, NY" }
{ "name" : "Big, Bright, Tribeca Studio w/ Doorman & Elevator" }
> db.rent.find({$text: {$search: "TriBeca"}}, {"_id":0, "name":1}).limit(5)
{ "name" : "Architect Loft,  btw Soho & Tribeca, Private Roof" }
{ "name" : "Stunning Sundrenched Tribeca Loft" }
{ "name" : "Stylish Suite & Excellent Service in Tribeca, NY" }
{ "name" : "Stylish Hotel & Excellent Service in Tribeca, NY" }
{ "name" : "Big, Bright, Tribeca Studio w/ Doorman & Elevator" }
> db.rent.find({$text: {$search: "TriBeCa"}}, {"_id":0, "name":1}).limit(5)
{ "name" : "Architect Loft,  btw Soho & Tribeca, Private Roof" }
{ "name" : "Stunning Sundrenched Tribeca Loft" }
{ "name" : "Stylish Suite & Excellent Service in Tribeca, NY" }
{ "name" : "Stylish Hotel & Excellent Service in Tribeca, NY" }
{ "name" : "Big, Bright, Tribeca Studio w/ Doorman & Elevator" }
```

Logical Operators in $search: AND, OR and NOT

```
> db.rent.find({$text: {$search: "\"bedroom apartment\""}}, {"_id":0, "name":1, "description":1}).limit(5)
{ "name" : "Fabulous Apartment with Soaking Tub", "description" : "Stylish and newly renovated 1-bedroom apartment ..." }
{ "name" : "Cute Apartment with Great Bathroom!", "description" : "Experience the city like a real New Yorker! This unique and modern one-bedroom apartment ..." }
{ "name" : "BROOKLYN LUXURY minutes 2 Manhattan & Williamsburg", "description" : "KING BED apartment is about 800 sq. ft. of perfection. Professionally designed for you comfort and with all the things you want and need.  Originally a two bedroom apartment ..." }
{ "name" : "Newly Renovated Near Central Park", "description" : "The perfect home away from home in NYC! Stylish, newly-renovated 1-bedroom apartment ... " }
```

Searching with AND

- **# Note the \" bedroom apartment \"!**
- **# We search for "bedroom" AND "apartment"**
- **# We only got 4 rentals!**

```
> db.rent.find({$text: {$search: "\"bedroom apartment\" sunny"}}, {"_id":0, "name":1}).limit(5)
{ "name" : "Fabulous Apartment with Soaking Tub" }
{ "name" : "Cute Apartment with Great Bathroom!" }
{ "name" : "BROOKLYN LUXURY minutes 2 Manhattan & Williamsburg" }
{ "name" : "Newly Renovated Near Central Park" }

> db.rent.find({$text: {$search: "sunny bedroom apartment"}}, {"_id":0, "name":1}).limit(5)
{ "name" : "Fabulous Apartment with Soaking Tub" }
{ "name" : "Cute Apartment with Great Bathroom!" }
{ "name" : "BROOKLYN LUXURY minutes 2 Manhattan & Williamsburg" }
{ "name" : "Giant Sunny Room at Brooklyn&Breakfast" }
```

Mixing with OR

- **# And "sunny"?? "Giant Sunny Room at Brooklyn&Breakfast" does not appear!**

- **# There exists rentals with "sunny"!**

If the $search string includes a phrase and individual terms, text search will only match the documents that include the phrase

“sunny bedroom apartment”
==
“apartment” OR “bedroom” OR “sunny”

“sunny \” bedroom apartment\” ”
==
“(sunny OR bedroom OR apartment)” AND “bedroom apartment”

Removing Elements: NOT

“bedroom apartment”
==
“bedroom” OR “apartment”

“bedroom -apartment”
==
“bedroom” AND NOT “apartment”

“apartment-wrecker -bedroom”
==
“apartment-wrecker” AND NOT
“bedroom”

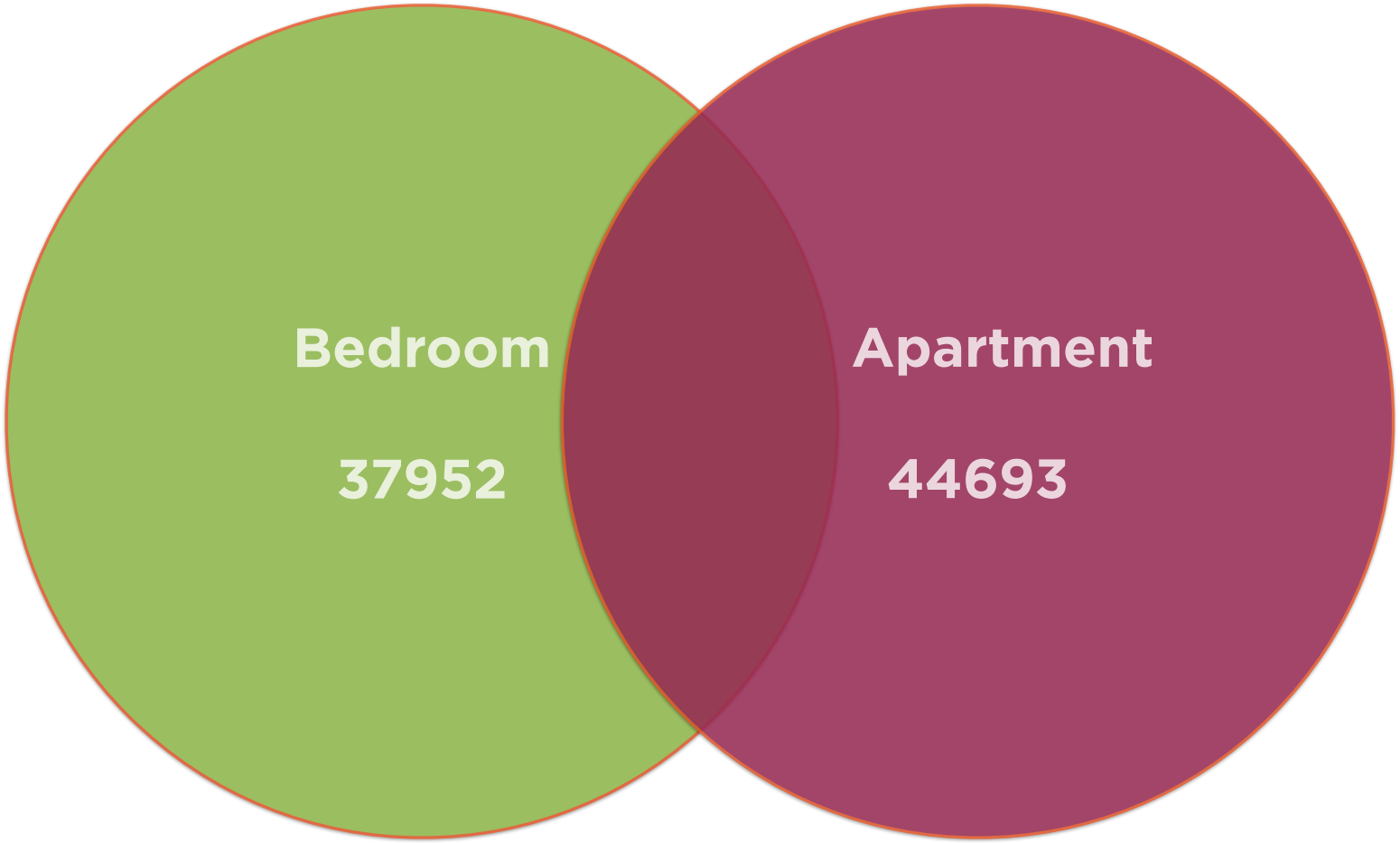
Another Example with NOT

```
> db.rent.find({$text:{$search: "bedroom apartment"}}).count()
49320
> db.rent.find({$text:{$search: "bedroom"}}).count()
37952
> db.rent.find({$text:{$search: "-bedroom apartment"}}).count()
11368


> db.rent.find({$text:{$search: "apartment"}}).count()
44693
> db.rent.find({$text:{$search: "bedroom -apartment"}}).count()
4627
```

Bedrooms and Apartments

Bedrooms and Apartments



Bedrooms and Apartments

Bedrooms and Apartments

Demo

Look for rentals with the $text operator

Play around with the logical operators

Summary

Learned about the $text operator and its corresponding fields

Searched for text from a set of token and from an exact phrase

Performed logical OR, AND and NOT conditions on a full text query