CS 5180 – Assignment #2
Maximum Points: 100 pts.

Bronco ID: |__|__|__|__|__|__|__|__|__|

Last Name: _____

First Name: _____

Note 1: Your submission header must have the format as shown in the above-enclosed rounded rectangle.
Note 2: Homework is to be done individually. You may discuss the homework problems with your fellow students, but you are NOT allowed to copy – either in part or in whole – anyone else's answers.
Note 3: Your deliverable should be a .pdf file submitted through Gradescope until the deadline. Do not forget to assign a page to each of your answers when making a submission. In addition, source code (.py files) should be added to an online repository (e.g., Github) to be downloaded and executed later.
Note 4: All submitted materials must be legible. Figures/diagrams must have good quality.
Note 5: Please use and check the Canvas discussion for further instructions, questions, answers, and hints. The bold words/sentences provide information for a complete or accurate answer.

1. [40 points]. Based on the `documents` collection below, write a **single MongoDB query** to retrieve the requested information. The `num_chars` field does not include spaces and punctuation marks.

```
documents
{
"_id": 1, "title": "Exercise", "text": "Baseball is played during summer months.", "num_chars": 34,
"date": {"$date": "2024-09-03"}, "category": "Sports",
terms: [{term: "baseball", count: 1, num_chars: 8}, {term: "is", count: 1, num_chars: 2},
        {term: "played", count: 1, num_chars: 6}, {term: "during", count: 1, num_chars: 6},
        {term: "summer", count: 1, num_chars: 6}, {term: "months", count: 1, num_chars: 6}]
}
{
"_id": 2, "title": "California", "text": "Summer is the time for picnics here. Picnics time!", "num_chars": 40,
 "date": {"$date": "2024-09-04"}, "category": "Sports",
terms: [{term: "summer", count: 1, num_chars: 6}, {term: "is", count: 1, num_chars: 2},
        {term: "the", count: 1, num_chars: 3}, {term: "time", count: 2, num_chars: 4},
        {term: "for", count: 1, num_chars: 3}, {term: "picnics", count: 2, num_chars: 7},
        {term: "here", count: 1, num_chars: 4}]
}
```

{

"_id": 3, "title": "Discovery", "text": "Months, months, months later we found out why.", "num_chars": 36,

"date": {"$date": "2024-09-05"}, "category": "Seasons",

terms: [{term: "months", count: 3, num_chars: 6}, {term: "later", count: 1, num_chars: 5},

{term: "we", count: 1, num_chars: 2}, {term: "found", count: 1, num_chars: 5},

{term: "out", count: 1, num_chars: 3}, {term: "why", count: 1, num_chars: 3}]

}

{

"_id": 4, "title": "Arizona", "text": " Why is summer so hot here? So hot!", "num_chars": 25,

"date": {"$date": "2024-09-06"}, "category": "Seasons",

terms: [{term: "why", count: 1, num_chars: 3}, {term: "is", count: 1, num_chars: 3},

{term: "summer", count: 1, num_chars: 6}, {term: "so", count: 2, num_chars: 2},

{term: "hot", count: 2, num_chars: 3}, {term: "here", count: 1, num_chars: 4}]

}

a. [5 points]. How many documents are in `documents`? *{4}*.
b. [5 points]. How many documents in `documents` have the term "summer"? Requirement: query `documents` by using `terms`. *{3}*.
c. [5 points]. List the `text` of documents linked to the category "Sports". *{ Baseball is played during summer months., Summer is the time for picnics here. Picnics time!}*
d. [5 points]. What distinct `terms` are in `documents`? Requirement: query `documents` by using `terms`. *{baseball, during, for, found, here, hot, is, later, months, out, picnics, played, so, summer, the, time, we, why}*
e. [5 points]. List the `title` and `date` of documents registered after '09/04/2024'. *{Discovery, 09/05/2024; Arizona, 09/06/2024}*
f. [5 points]. List the `title` and `num_chars` of documents that have `num_chars` greater than 30 but lower than 40 sorted by `num_chars` in descending order. *{Discovery, 36; Exercise, 34}*
g. [5 points]. How many terms (considering repetitions) are in the document "Arizona"? Requirement: query `documents` by using `title`. Requirement: query `documents` by using `terms`. *{8}*
h. [5 points]. How many times (considering repetitions) does the term "months" occur in `documents`? Requirement: query `documents` by using `terms`. *{4}*

2. [40 points]. Now, consider that you are creating a `users` collection from scratch. Write the corresponding C~~R~~UD operation to obtain the data as presented.

a. [5 points]. Creating the user, Leslie.

```
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Pawnee"
}
```

b. [5 points]. Updating the user, Leslie.

```
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Linkiee"
}
```

c. [5 points]. Creating the user, Ron.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee"
}
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": " Linkiee"
}
```


d. [5 points]. Updating the user, Leslie.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee"
}
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Linkiee",
  "hobbies": ["scrapbooking", "guitar", "hiking"]
}
```


e. [5 points]. Updating the user, Ron.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee",
  "jobHistory": [
    {"title": "Deputy Director", "yearStarted": 2004},
    {"title": "City Councillor", "yearStarted": 2012}]
}
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Linkiee",
  "hobbies": ["scrapbooking", "guitar", "hiking"]
}
```


f. [5 points]. Updating the user, Leslie.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee",
  "jobHistory": [
    {"title": "Deputy Director", "yearStarted": 2004},
    {"title": "City Councillor", "yearStarted": 2012}]
}
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Linkiee",
  "hobbies": ["scrapbooking","guitar"]
}
```


g. [5 points]. Updating the user, Ron.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee",
  "jobHistory": [
    {"title": "Deputy Director", "yearStarted": 2004},
    {"title": "City Councillor", "yearStarted": 2012},
    {"title": "Manager", "yearStarted": 2014}]
}
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Linkiee",
  "hobbies": ["scrapbooking","guitar"]
}
```


h. [5 points]. Deleting the user, Leslie.


```
{
  "_id": 2,
  "first_name": "Ron",
  "last_name": "Swandaughter",
  "cell": "8125559347",
  "city": "Pawnee",
  "jobHistory": [
    {"title": "Deputy Director", "yearStarted": 2004},
    {"title": "City Councillor", "yearStarted": 2012},
    {"title": "Manager", "yearStarted": 2014}]
}
```

3. [20 points]. Complete the Python program (db_connection_mongo.py) by using PyMongo. Use the driver program index_mongo.py to trigger the operations (**do not change it**). Use the provided sample output to validate your implementation. Add the link to an online repository as the answer to this question.

   a) [6 points]. Create a document.

      Input: {*id, text*, *title*, *date,* and *category*}

   b) [4 points]. Update a document.

      Input: {*id, text*, *title*, *date,* and *category*}

   c) [4 points]. Delete a document.

      Input: {*id*}

   d) [6 points]. Output the inverted index ordered by term.

      Output: { *'term': ' document_i_title: count, document_j_title: count'*}

Sample output: Input data in red.

######### Menu ##############

#a - Create a document

#b - Update a document

#c - Delete a document.

#d - Output the inverted index ordered by term.

#q - Quit

Enter a menu choice: d

{}

Enter a menu choice: a

Enter the ID of the document: 1

Enter the text of the document: Baseball is played during summer months.

Enter the title of the document: Exercise

Enter the date of the document: 2024-09-03

Enter the category of the document: Sports

Enter a menu choice: d

{'baseball': 'Exercise:1', 'during': 'Exercise:1', 'is': 'Exercise:1', 'months': 'Exercise:1', 'played': 'Exercise:1', 'summer': 'Exercise:1'}

Enter a menu choice: a

Enter the ID of the document: 2

Enter the text of the document: Summer is the time for picnics here. Picnics time!

Enter the title of the document: California

Enter the date of the document: 2024-09-04

Enter the category of the document: Sports

Enter a menu choice: d

{'baseball': 'Exercise:1', 'during': 'Exercise:1', 'for': 'California:1', 'here': 'California:1', 'is': 'Exercise:1, California:1', 'months': 'Exercise:1', 'picnics': 'California:2', 'played': 'Exercise:1', 'summer': 'California:1, Exercise:1', 'the': 'California:1', 'time': 'California:2'}

Enter a menu choice: a

Enter the ID of the document: 3

Enter the text of the document: Months, months, months later we found out why.

Enter the title of the document: Discovery

Enter the date of the document: 2024-09-05

Enter the category of the document: Seasons

Enter a menu choice: d

{'baseball': 'Exercise:1', 'during': 'Exercise:1', 'for': 'California:1', 'found': 'Discovery:1', 'here': 'California:1', 'is': 'Exercise:1, California:1', 'later': 'Discovery:1', 'months': 'Exercise:1, Discovery:3', 'out': 'Discovery:1', 'picnics': 'California:2', 'played': 'Exercise:1', 'summer': 'Exercise:1, California:1', 'the': 'California:1', 'time': 'California:2', 'we': 'Discovery:1', 'why': 'Discovery:1'}

Enter a menu choice: a

Enter the ID of the document: 4

Enter the text of the document: Why is summer so hot here? So hot!

Enter the title of the document: Arizona

Enter the date of the document: 2024-09-06

Enter the category of the document: Seasons

Enter a menu choice: d

{'baseball': 'Exercise:1', 'during': 'Exercise:1', 'for': 'California:1', 'found': 'Discovery:1', 'here': 'Arizona:1, California:1', 'hot': 'Arizona:2', 'is': 'Arizona:1, Exercise:1, California:1', 'later': 'Discovery:1', 'months': 'Exercise:1, Discovery:3', 'out': 'Discovery:1', 'picnics': 'California:2', 'played': 'Exercise:1', 'so': 'Arizona:2', 'summer': 'California:1, Arizona:1, Exercise:1', 'the': 'California:1', 'time': 'California:2', 'we': 'Discovery:1', 'why': 'Arizona:1, Discovery:1'}

Enter a menu choice: c

Enter the document ID to be deleted: 3

Enter a menu choice: d

{'baseball': 'Exercise:1', 'during': 'Exercise:1', 'for': 'California:1', 'here': 'Arizona:1, California:1', 'hot': 'Arizona:2', 'is': 'Exercise:1, Arizona:1, California:1', 'months': 'Exercise:1', 'picnics': 'California:2', 'played': 'Exercise:1', 'so': 'Arizona:2', 'summer': 'California:1, Arizona:1, Exercise:1', 'the': 'California:1', 'time': 'California:2', 'why': 'Arizona:1'}

Enter a menu choice: b

Enter the ID of the document: 4

Enter the text of the document: Why is summer so hot here? This is a bad time!

Enter the title of the document: Arizona

Enter the date of the document: 2024-09-07

Enter the category of the document: Seasons

Enter a menu choice: d

{'a': 'Arizona:1', 'bad': 'Arizona:1', 'baseball': 'Exercise:1', 'during': 'Exercise:1', 'for': 'California:1', 'here': 'Arizona:1, California:1', 'hot': 'Arizona:1', 'is': 'California:1, Exercise:1, Arizona:2', 'months': 'Exercise:1', 'picnics': 'California:2', 'played': 'Exercise:1', 'so': 'Arizona:1', 'summer': 'Arizona:1, California:1, Exercise:1', 'the': 'California:1', 'this': 'Arizona:1', 'time': 'Arizona:1, California:2', 'why': 'Arizona:1'}

Important Note: Answers to all questions should be written clearly, concisely, and unmistakably delineated. You may resubmit multiple times until the deadline (the last submission will be considered).

NO LATE ASSIGNMENTS WILL BE ACCEPTED. ALWAYS SUBMIT WHATEVER YOU HAVE COMPLETED FOR PARTIAL CREDIT BEFORE THE DEADLINE!