SQL & FINAL REVIEW

COMPUTER SCIENCE MENTORS 61A

November 28–December 9, 2022

This worksheet is to be used in both week 15 (SQL) and week 16 (Final Review). Please save this worksheet after week 15's section and remember to bring it for week 16.

1 SQL

SQL (Structured Query Language) is a declarative programming language that allows us to store, access, and manipulate data stored in databases. Each database contains tables, which are rectangular collections of data with rows and columns. This section gives a brief overview of the small subset of SQL used by CS 61A; the full language has many more features.

1.1 Creating Tables

1.1.1 `SELECT`

`SELECT` statements are used to create tables. The following creates a table with a single row and two columns:

```
sqlite> SELECT "Adit" AS first, "Balasubramanian" AS last;
Adit|Balasubramanian
```

`AS` is an "aliasing" operation that names the columns of the table. Note that built-in keywords such as **`AS`** and **`SELECT`** are capitalized by convention in SQL. However, SQL is case insensitive, so we could just as easily write **`as`** and **`select`**. Also, each SQL query must end with a semicolon.

1.1.2 `UNION`

`UNION` joins together two tables with the same number of columns by "stacking them on top of each other". The column names of the first table are kept.

```
sqlite> SELECT "Adit" AS first, "Balasubramanian" AS last UNION
...> SELECT "Gabe", "Classon";
Adit|Balasubramanian
Gabe|Classon
```

1.1.3 CREATE TABLE

To create a named table (so that we can use it again), the **CREATE TABLE** command is used:

```
CREATE TABLE scms AS
    SELECT "Adit" AS first, "Balasubramanian" AS last UNION
    SELECT "Gabe", "Classon";
```

The remaining examples will use the following `team` table:

```
CREATE TABLE team AS
    SELECT "Gabe" AS name, "cat" AS pet, 11 AS birth_month UNION
    SELECT "Adit",         "none",       10 UNION
    SELECT "Alyssa",       "dog",         4 UNION
    SELECT "Esther",       "dog",         6 UNION
    SELECT "Maya",         "dog",         3 UNION
    SELECT "Manas",        "none",       11;
```

1.2 Manipulating other tables

We can also write **SELECT** statements to create new tables from other tables. We write the columns we want after the **SELECT** command and use a **FROM** clause to designate the source table. For example, the following will create a new table containing only the `name` and `birth_month` columns of `team`:

```
sqlite> SELECT name, birth_month FROM team;
Adit|10
...
Maya|3
```

Note that the order in which rows are returned is undefined.

An asterisk `*` selects for all columns of the table:

```
sqlite> SELECT * FROM team;
Adit|none|10
...
Maya|dog|3
```

This is a convenient way to view all of the content of a table.

We may also manipulate the table columns and use **AS** to provide a (new) name to the columns of the resulting table. The following query creates a table with each teammate's name and the number of months between their birth month and June:

```
sqlite> SELECT name, ABS(birth_month - 6) AS june_dist FROM team;
Adit|4
...
Maya|3
```

1.2.1 WHERE

WHERE allows us to filter rows based on certain criteria. The **WHERE** clause contains a boolean expression; only rows where that expression evaluates to true will be kept.

```
sqlite> SELECT name FROM team WHERE pet = "dog";
Alyssa
Esther
Maya
```

Note that = in SQL is used for equality checking, not assignment.

1.2.2 ORDER BY

ORDER BY specifies a value by which to order the rows of the new table. **ORDER BY ...** may be followed by **ASC** or **DESC** to specify whether they should be ordered in ascending or descending order. **ASC** is default. For strings, ascending order is alphabetical order.

```
sqlite> SELECT name FROM team WHERE pet = "dog" ORDER BY name DESC;
Maya
Esther
Alyssa
```

1.3 Joins

Sometimes, you need to compare values across two tables—or across two rows of the same table. Our current tools do not allow for this because they can only consider rows one-by-one. A way of solving this problem is to create a table where the rows consist of every possible combination of rows from the two tables; this is called an **inner join**. Then, we can filter through the combined rows to reveal relationships between rows. It sounds bizarre, but it works.

An inner join is created by specifying multiple source tables in a **WHERE** clause. For example, **SELECT** `*` **FROM** `team` **AS** `a, team` **AS** `b;` will create a table with 36 rows and 6 columns. The table has 36 rows because each row represents one of 36 possible ways to select two rows from `team` (where order matters). The table has 6 columns because the joined tables have 3 columns each. We use **AS** to give the two source tables different names, since we are joining `team` to itself. The columns of the resulting table are named `a.name, a.pet, a.birth_month, b.name, b.pet, b.birth_month`.

For example, to determine all pairs of people with the same birth month, we can use an inner join:

```
sqlite> SELECT a.name, b.name FROM team AS a, team AS b WHERE a.name < b.name
   AND a.birth_month = b.birth_month;
Gabe|Manas
```

1.4 Aggregation

Aggregation uses information from multiple rows in our table to create a single row. Using an aggregation function such as **MAX**, **MIN**, **COUNT**, and **SUM** will automatically aggregate the table data into a single row. For example, the following will collapse the entire table into one row containing the name of the person with the latest birth month:

```
sqlite> SELECT name, MAX(birth_month) FROM team;
Manas|11
```

Note that there are multiple rows with the largest birth month. When this happens, SQL arbitrarily chooses one of the rows to use.

The **COUNT** aggregation function collapses the table into one row containing the number of rows in the table:

```
sqlite> SELECT COUNT(*) FROM team;
6
```

1.4.1 GROUP BY

GROUP BY groups together all rows with the same value for a particular column. Aggregation is performed on each group instead of on the entire table. There is then *exactly one row* in the resulting table for each group. As before, type of aggregation performed is determined by the choice of aggregation function. The following gives, for each type of pet, the information of the person with the earliest birth month who has that pet:

```
sqlite> SELECT name, pet, MIN(birth_month) FROM team GROUP BY pet;
Gabe|cat|11
Maya|dog|3
Adit|none|10
```

1.4.2 HAVING

Just as **WHERE** filters out rows, **HAVING** filters out groups. For example, the following selects for all types of pets owned by more than one teammate:

```
sqlite> SELECT pet FROM team GROUP BY pet HAVING COUNT(*) > 1;
dog
none
```

1.5 Syntax

The clauses of a **SELECT** statement always come in this order:

```
SELECT ... FROM ... WHERE ... GROUP BY ... HAVING ... ORDER BY ...;
```

The order roughly reflects the order in which the processing steps are applied. Note that all filtering of rows comes *before* aggregation. That is, aggregation is always performed after the row-by-row filtering is complete.

1. CS 61A wants to start a fish hatchery, and we need your help to analyze the data we've collected for the fish populations! Running a hatchery is expensive—we'd like to make some money on the side by selling some seafood (only older fish of course) to make delicious sushi.

   The table `fish` contains a subset of the data that has been collected. The SQL column names are listed in brackets.

`fish`

Species [species]	**Population [pop]**	**Breeding Rate [rate]**	**$/piece [price]**	**# of pieces per fish [pieces]**
Salmon	500	3.3	4	30
Eel	100	1.3	4	15
Yellowtail	700	2.0	3	30
Tuna	600	1.1	3	20

   (a) Write a query to find the three most populated fish species.

   (b) Write a query to find the total number of fish in the ocean. Additionally, include the number of species we summed. Your output should have the number of species and the total population.

   (c) Profit is good, but more profit is better. Write a query to select the species that yields the most number of pieces for each price. Your output should include the species, price, and pieces.

   (d) Write a query that returns, for each species, the difference between our hatchery's revenue versus the competitor's revenue for one whole fish.

2. In this question, you have access to two tables.

   Grades, which contains three columns: **`day`**, `class`, and `score`. Each row represents the `score` you got on a midterm for some `class` that you took on some **`day`**.

   Outfits, which contains two columns: **`day`** and `color`. Each row represents the `color` of the shirt you wore on some **`day`**. Assume you have a row for each possible day.

`outfits`

Day	Color
11/5	Blue
9/13	Red
10/31	Orange

`grades`

Day	Class	Score
10/31	Music 70	88
9/20	Math 1A	72

   (a) Instead of actually studying for your finals, you decide it would be the best use of your time to determine what your "lucky shirt" is. Suppose you're pretty happy with your exam scores this semester, so you define your lucky shirt as the shirt you wore to the most exams.

   Write a query that will output the color of your lucky shirt and how many times you wore it.

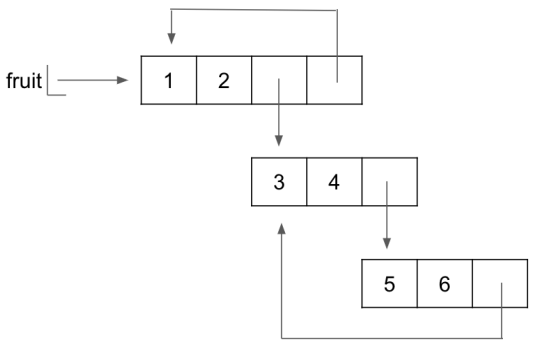
   (b) You want to find out which classes you need to prepare for the most by determining how many points you have so far. However, you only want to do so for classes where you did relatively poorly.

   Write a query that will output the sum of your midterm scores for each class along with the corresponding class, but only for classes in which you scored less than 80 points on at least one midterm. List the output from highest to lowest total score.

2 Environment Diagrams

1. Fill in each blank in the code example below so that its environment diagram is the following. You do not need to use all the blanks.

```
fruit = [1, 2, [3, 4]]
fruit._____________________________
fruit[3][2]._____________________
fruit[2][2]._____________________
fruit[3][3][2][2][2][1] = ____
```



2. Define a **non-decreasing path** as a path from the root where each node's label is greater than or equal to the previous node along the path. A **subpath** is a path between nodes X and Y, where Y must be a descendent of X (ex: Y is a branch of a branch of X).

   (a) Write a generator function `root_to_leaf` that takes in a tree `t` and yields all non-decreasing paths from the root to a leaf node, in any order. Assume that `t` has at least one node.

```
def root_to_leaf(t):
    """
    >>> t1 = Tree(3, [Tree(5), Tree(4)])
    >>> list(root_to_leaf(t1))
    [[3, 5], [3, 4]]
    >>> t2 = Tree(5, [Tree(2, [Tree(7), Tree(8)]), Tree(5, [Tree(6)])])
    [[5, 5, 6]]
    """

    if _____________________:

        _____________________________

    for _____________________________:

        if _____________________________:

            for _____________________________:

                ___________________________________
```

   (b) Write a generator function `subpaths` that takes in a tree `t` and yields all non-decreasing subpaths that end with a leaf node, in any order. You may use the `root_to_leaf` function above, and assume again that `t` has at least one node.

```
def subpaths(t):

    yield from _____________________

    for b in t.branches:

        ___________________________________
```

3. In the following problem, we will represent a bookshelf object using dictionaries.

   In the first section, we will set up the format. Here, we will directly work with the internals of the Bookshelf, so don't worry about abstraction barriers for now. Fill in the following functions based on their descriptions (the constructor is given to you):

```
def Bookshelf(capacity):
    """ Creates an empty bookshelf with a certain max capacity. """
    return {'size': capacity, 'books': {}}

def add_book(bookshelf, author, title):
    """
    Adds a book to the bookshelf. If the bookshelf is full,
    print "Bookshelf is full!" and do not add the book.
    >>> books = Bookshelf(2)
    >>> add_book(books, 'Jane Austen', 'Pride and Prejudice')
    >>> add_book(books, 'Daniel Kleppner', 'An Introduction to Mechanics
       5th Edition')
    >>> add_book(books, 'Kurt Vonnegut', 'Galapagos')
    Bookshelf is full!
    """
    if _____________________________:
        print('Bookshelf is full!')
    else:
        if author in bookshelf['books']:
            ____________________________________
        else:
            ____________________________________


def get_all_authors(bookshelf):
    """
    Returns a list of all authors who have at least one book in the
       bookshelf.
    >>> books = Bookshelf(10)
    >>> add_book(books, 'Jane Austen', 'Pride and Prejudice')
    >>> add_book(books, 'Sheldon Axler', 'Linear Algebra Done Right')
    >>> add_book(books, 'Kurt Vonnegut', 'Galapagos')
    >>> get_all_authors(books)
    ['Jane Austen', 'Sheldon Axler', 'Kurt Vonnegut']
    """
    return ____________________________________
```

Now, complete the function `most_popular_author` **without breaking the abstraction barrier**. In other words, you are not allowed to assume anything about the implementation of a Bookshelf object, or use the fact that it is a dictionary. You can only use the methods above and their stated return values.

```
def most_popular_author(bookshelf):
    """
    Returns the author with the greatest number of books on this bookshelf.
    You can assume that the bookshelf is not empty.
    >>> books = Bookshelf(100)
    >>> add_book(books, 'Orson Scott Card', 'Xenocide')
    >>> add_book(books, 'Orson Scott Card', 'Children of the Mind')
    >>> add_book(books, 'J.R.R. Tolkien', 'The Hobbit')
    >>> most_popular_author(bookshelf)
    'Orson Scott Card'
    """
    return max(_____________________________________________________,

        key=_____________________________________________________________)
```

5 Efficiency

4. Find the $\Theta(\cdot)$ runtime bound for `hiya(n)`. Remember that Python strings are immutable: when we add two strings together, we need to make a copy.

```
def hiii(m):
    word = "h"
    for i in range(m):
        word += "i"
    return word

def hiya(n):
    i = 1
    while i < n:
        print(hiii(i))
        i *= 2
```

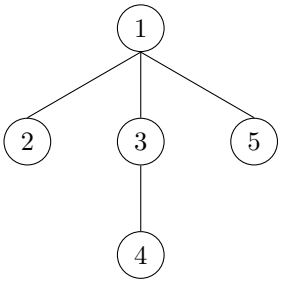
5. Implement the classes so the following code runs.

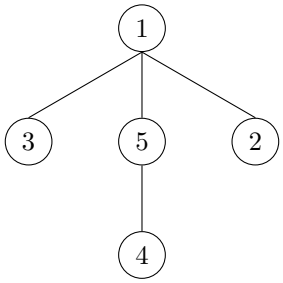
```
"""
>>> p = Plant()
>>> p.height
1
>>> p.materials
[]
>>> p.absorb()
>>> p.materials
[|Sugar|]
>>> Sugar.sugars_created
1
>>> p.leaf.sugars_used
0
>>> p.grow()
>>> p.materials
[]
>>> p.height
2
>>> p.leaf.sugars_used
1
"""

class Plant:
    def __init__(self):
        """A Plant has a Leaf, a list of sugars created so far,
        and an initial height of 1.
        """




    def absorb(self):
        """Calls the Leaf to create sugar."""




    def grow(self):
        """A Plant consumes all of its sugars to grow, each of which
        increases its height by 1.
        """
```

```
class Leaf:
    def __init__(self, plant): # plant is a Plant instance
        """A Leaf is initially alive, and keeps track of how many
        sugars it has created.
        """



    def absorb(self):
        """If this Leaf is alive, a Sugar is added to the plant's
        list of sugars.
        """
        if self.alive:


    def __repr__(self):
        return '|Leaf|'

class Sugar:
    sugars_created = 0

    def __init__(self, leaf, plant):



    def activate(self):
        """A sugar is used."""




    def __repr__(self):
        return '|Sugar|'
```

6. Implement `rotate`, which takes in a tree and rotates the labels at each level of the tree by one to the left destructively. This rotation should be modular (That is, the leftmost label at a level will become the rightmost label after running rotate). You do NOT need to rotate across different branches.

   For example, given tree `t` on the left, `rotate(t)` should mutate `t` to give us the right.






```
def rotate(t):
  """
  >>> t1 = Tree(1, [Tree(2), Tree(3, [Tree(4)]), Tree(5)])
  >>> rotate(t1)
  >>> t1
  Tree(1, [Tree(3), Tree(5, [Tree(4)]), Tree(2)])
  >>> t2 = Tree(1, [Tree(2, [Tree(3), Tree(4)]),
                    Tree(5, [Tree(6)])])
  >>> rotate(t2)
  >>> t2
  Tree(1, [Tree(5, [Tree(4), Tree(3)]),
                    Tree(2, [Tree(6)])])
  """
  branch_labels = ____________________________________

  n = len(t.branches)

  for ____________________________________________:

    ____________________________________________________

    ____________________________________________________

    ____________________________________________________
```

7. Star-Lord is cruising through space and can't afford to crash into any asteroids along the way. Let his path be represented as a (possibly nested) list of integers, where an asteroid is denoted with a 0, and stars and planets otherwise. Every time Star-lord sees (visits) an asteroid (0), he merges the next planet/star with the asteroid. In other words, construct a NEW list so that all asteroids (0s) are replaced with a list containing the planet followed by the asteroid (e.g. (planet 0)). You can assume that the last object in the path is not an asteroid (0).

```
;Doctests
scm> (collision (list 1 2 3 0 4))
(1 2 3 (4 0))
scm> (collision (list 4 3 (list 0 1) 2))
(4 3 ((1 0)) 2)
scm> (collision (list 1 -2 0 -3 4 0 -5 6))
(1 -2 (-3 0) 4 (-5 0) 6)
scm> (collision (list 1 0 0 2 3))
(1 (0 0) 2 3)

;Asteroids can merge with other asteroids too

(define (collision lst)

  (cond ((_____________________) lst)

    ((_________________________________)

      _____________________________________)

    ((_________________________________)

      (cons _____________________________

        _____________________________________))

    (else _____________________________________)
  )
)
```

8. Write a tail recursive function, `skip-list`, that takes in a potentially nested list `lst` and a filter function `filter-fn`, goes through each element in order, and returns a new list that contains all elements that pass the `filter-fn`. The returned list is *not nested*.

   Hint: `pair?` is a predicate procedure that returns true if its argument is a Scheme list and false otherwise.

```
;Doctests
scm> (skip-list '(1 (3)) even?)
()
scm> (skip-list '(1 (2 (3 4) 5) 6 (7) 8 9) odd?)
(1 3 5 7 9)

(define (skip-list lst filter-fn)
  (define (skip-list-tail _________ _________ next)

    (cond
        ((null? lst) (if (null? _________)
                         ___________________________________

                         ___________________________________))

        ((pair? ________) (___________________________________))

        ((_______________) ___________________________________)

        (else _____________________________________)
    )
  )
  (skip-list-tail ___________ ___________ ___________)
)
```

9. (Spring 2018 Final)
Implement `lambda-macro`, a macro that creates anonymous macros. A lambda-macro expression has a list of formal parameters and one body expression. It creates a *macro* with those formal parameters and that body. Assume that the symbol anon is not use anywhere else in a program that contains lambda-macro.

```
;Doctests
scm> (define mac (lambda-macro (x) `(begin ,x ,x)))
mac
scm> (mac (print 1))
1
1
(define-macro (lambda-macro bindings body)

      `(begin (_____________________________

               _____________________________

               _____________________________)

         anon))
```

10. Define a macro, `eval-and-check` that takes in three expressions and evaluates each expression in order. If the last expression evaluates to a truth-y value, return the symbol ok. Otherwise, return fail.

```
;Doctests
scm> (eval-and-check #f #f #t)
ok
scm> (eval-and-check (+ 2 3) (print 2) (> 2 3))
2
fail
scm> (eval-and-check (define x 1) (print x) (> x 0))
1
ok


(define-macro (eval-and-check expr1 expr2 expr3)

     ___________________________________________

     ___________________________________________

     ___________________________________________)
```