

Discussion 7

Quiz 3 Review

This content was created for Information Retrieval Fall 2021 at UC Irvine by Brooke Kelsey Ryan. These slides, accompanying recordings, and any other course materials are protected by U.S. copyright law may not be reproduced, distributed, or displayed without the express written consent of the author.

© Brooke Kelsey Ryan 2021

Welcome! Information Retrieval - Discussion 7

Itinerary November 10, 2021

- Quiz 3 Review
- Preparing for the code walkthroughs

Deadlines this week

- Friday 11:59pm: Assignment 3, Milestone 1
- Friday in class: Quiz 3
- Starting today through next week: Code walkthroughs

Code Walkthroughs

How to Prepare (**with Brooke!!**)

(not necessarily applicable for meetings with other TAs, so please get info from them if you are meeting with them)

Goals of the Meeting

- Primary goal of the meeting is to **demonstrate** your **understanding** of the code and the project
- Think of it as an interview of the architects (you!)
- Goal is not to “drill” you or do any gotchas
 - But if you can’t answer a question, we can’t give you points for that one
 - Also need to make sure all team members can demonstrate understanding



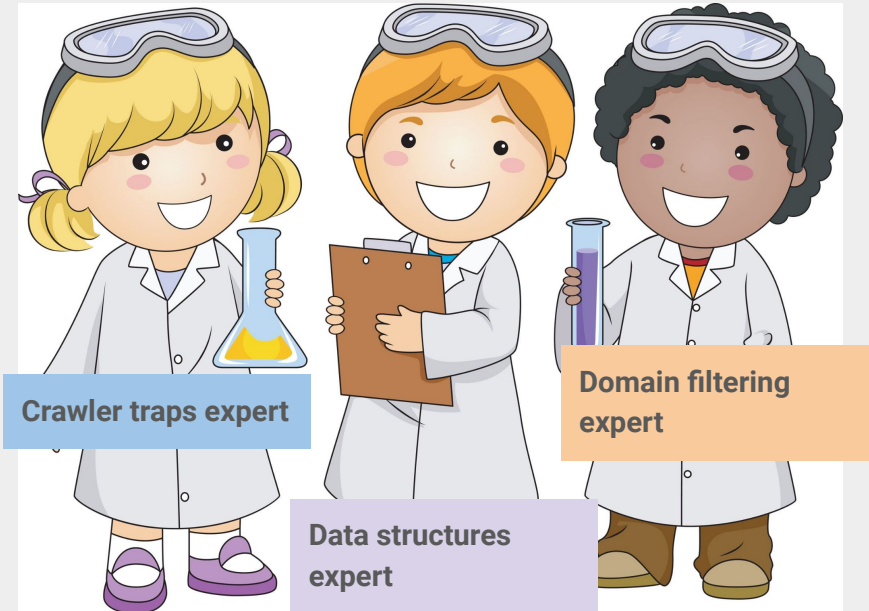
Pretend you're a podcast guest and you're showing off your teams' genius design and architecture choices!

Recommended Team Organization

This is not required, simply what I might strategize doing if I were a student. Feel free to alter what makes most sense.

- Look over the “essential” components of the assignment (ex, “**Crawler Behavior Requirements**” on Canvas)
 - For example: crawler traps, politeness, etc.
- I would recommend assigning an “expert” to each essential area
 - This can be where there were natural breaks in the work
 - Everyone in the team should have an “expert” position—otherwise you risk losing points

Example for team of 3 students:



This is not comprehensive, and you might have to have some members being ‘experts’ in multiple areas! Look at assignment and think about what are essential areas.

Example Pre-Call Preparation

- To prepare for your interview, have an “experts meeting” that morning before your Zoom call with the TA
 - Practice switching off asking the expert questions about their area
 - The other two should also “chime in” with their additional thoughts for each area
 - Be sure you can answer both high-level questions about the concept, as well as being able to “point” to where in the code this is implemented
 - Think about examples you encountered that demonstrate the concept.

Example question: How did you decide what was a “trap”?



*This would be a really great breakdown of teamwork! **Expert** can answer majority/**general approach**, and **teammates** add on to talk about **edge cases** and where in the code it was **implemented**.*

Other tips

- The call is **quick**, only **10** minutes!
- Make sure all your code is **ready to go** and that you have a designated **screen sharer**!



Questions?

Quiz 3

Review

Overview - 100pts

- 1 question from Quiz 1 material (2pts)
- 2 questions from Quiz 2 material (4pts)
- Remaining 94pts is Quiz 3 material!
- Reminder: Open-slides (only the lecture slides, not these slides unfortunately!)

Definitions

- Term-Document Matrix
- Inverted Index
 - What's the “inverted” part in an inverted index?
- Document ID

Incidence Vectors / Bitwise Operations

Given the term-document matrix to the right, what is the result of searching for:

1. Cat AND Dog

	Page 1	Page 2	Page 3	Page 4
Cat	1	0	1	0
Dog	0	0	1	1
Fish	1	1	1	0

Incidence Vectors / Bitwise Operations - **Answer**

Given the term-document matrix to the right, what is the result of searching for:

1. Cat AND Dog? - **Page 3**

Number 1	1	0	1	0	1
Number 2	1	1	1	0	0

AND	1	0	1	0	0
OR	1	1	1	0	1
XOR	0	1	0	0	1

Review
BitWise
operations



	Page 1	Page 2	Page 3	Page 4
Cat	1	0	1	0
Dog	0	0	1	1
Fish	1	1	1	0

This is the only combination that yields true!

Incidence Vectors / Bitwise Operations - Practice

Given the term-document matrix to the right, what is the result of searching for:

1. Cat AND Fish
2. Fish XOR Dog
3. Dog OR Cat
4. Dog OR Fish
5. Fish AND Dog

	Page 1	Page 2	Page 3	Page 4
Cat	1	0	1	0
Dog	0	0	1	1
Fish	1	1	1	0

Terms in an Inverted Index

Given the following sentences:

- A. Kiki is the greatest cat in the world.
- B. My cat's name is Kiki.
- C. Kiki loves to eat food and sleep.



1. How many terms will be in the inverted index?
2. What are the postings for “cat”?

Analyzing Complexity of Code

- There will be applied questions of analyzing some pseudocode or some situations
- Make sure you first **understand the concepts**
- Then **practice** applying those concepts **on problems in the course**



*Please review Big O if you don't know what this is!
Lots of great online resources out there.*

Analyzing Complexity of Code - Pseudocode Example

Now, we want to apply this to something relevant for this class, such as **building an index**.

Try this example →

What is the Big-O Complexity of `create_index` with respect to N = number of documents?

```
def function create_index(documents):  
  
    index = {}           # Maps terms to postings lists  
    doc_ids = {}        # Maps doc id numbers to URLs  
    n = 0  
  
    for document in documents:  
        if document.url not in doc_ids.values():  
            doc_ids[n] = document.url;  
        else:  
            continue      # We've seen this document already  
  
        unique_tokens = remove_duplicates(parse(document))  
  
        for all tokens in unique_tokens:  
            if token not in index:  
                list = index[token]  
                index[token].append(Posting(n))  
  
        n = n+1  
  
    return index, doc_ids
```

Analyzing Complexity of Boolean Retrieval (sorted)

Need to know Big-O complexity
using AND:

1. A 3-way merge algorithm

Query: Elt.1 AND Elt.2 AND Elt.3

List 1	1	2	7	11
List 2	1	3	11	13
List 3	2	4	6	11

Analyzing Complexity of Boolean Retrieval (unsorted)

Need to know Big-O complexity of:

1. A 3-way intersection algorithm
2. An algorithm to sort the lists

List 1	7	2	1	11
List 2	3	1	11	13
List 3	11	4	2	6

Query Optimization

Consider the following query and corresponding postings lists
(with doc ids only):

Query: kiki the fat cat

the: 4, 5, 10, 11, 14, 15, 16

kiki: 4, 11

fat: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16

cat: 11,15,16

Assume Boolean retrieval with an AND operation between the terms.

What is the best order for processing this query as fast as possible? (processing is left to right)



Miscellaneous

- Found some really great resources from UC Berkeley for “general” studying tips for Computer Science exams
 - Some of the advice is specific to the course, but much of it is generic and good advice IMO!
 - Emphasizes both **understanding concepts** but also **practice on problems!**
- [Exam Studying Guide](#)
- [CS 61A Guide for Studying for the Final](#)

Next Week's Discussion

Tentative plan for next week's discussion based on upcoming course deadlines.

- Assignment 3, Milestone 2

Recommended Homework

To best prepare for next week's session, I recommend you do the following.

- Practice experts Zoom call with your teammates for the code walkthrough
- Quiz 3 practice problems
- Keep working consistently on Assignment 3 and milestones!