

EECS 498 System Design of a Search Engine

Winter 2026

Lab 2: Introduction

GSI: Christopher Davis
IA: Rickey Dong

Nicole Hamilton
[https://web.eecs.umich.edu/~nham/
nham@umich.edu](https://web.eecs.umich.edu/~nham/nham@umich.edu)

agenda

1. Introduction / announcements
2. New assignments
3. Speed dating
4. Voting

agenda

1. Introduction / announcements
2. New assignments
3. Speed dating
4. Voting



Christopher Davis

chcdavis@umich.edu

About Me

- MS CS SUGS student (2026)
- BS CS + Honors Math (2025)
- Took this course in W25
- Worked for Google Scholar
 - Focused on ranking, but also some indexing and front-end. Maybe more on this when we get to ranking.
 - Updating codebase to C++20

Office Hours

Chris:

Tues/Thurs 12:30-2:30 BBB Atrium (pending room assignment)

By appointment (virtually) most other times M-F.

Just email chcdavis@umich.edu with **EECS498W26** in title.

Rickey:

Mon 3:30-4:30 (1637 BBB Table #1)

Wed 2:30-4:00 (1270A Leinweber)

agenda

1. Introduction / announcements
2. New assignments
3. Speed dating
4. Voting

New Assignments

- Good job on the first homework!

New Assignments

- Good job on the first homework!
- There are two new autograder assignments out.
 - Well, technically three

New Assignments

- Good job on the first homework!
- There are two new autograder assignments out.
 - Well, technically three
- UTF-8
 - Encode and decode UTF-8 characters
 - Notorious and difficult (and common) problem; lots of edge cases
 - Lecture next week introducing, but should read up on the format early to get a general idea.

New Assignments

- Good job on the first homework!
- There are two new autograder assignments out.
 - Well, technically three
- UTF-8
 - Encode and decode UTF-8 characters
 - Notorious and difficult (and common) problem; lots of edge cases
 - Lecture next week introducing, but should read up on the format early to get a general idea.
- String/Vector
 - Create your own well-optimized versions of RAI string/vector

Strings, Vectors, Etc.

- Common complaint:
 - “Why are we wasting time doing this? Wouldn’t we just use `std::xyz` in the real world?”

agenda

1. Introduction / announcements
2. New assignments
3. Speed dating
4. Voting

speed dating

First, how are we doing with team matching?

Reminder: should have teams formed by 18 January (Sunday)

speed dating

Today, we continue our speed-dating.

1. Break up in small groups of 3 or 4 for a short 5-to-10-minute discussion.
2. Collect some results.
3. Form new groups with people you haven't met yet and repeat.

What would you like to know about someone to know if they're a good match?

The -121 Section

Whether they drink Bubble Tea
Whether they are Performative
Motivation for success
Why they took the class
Good listeners
How interested they are
How much work they want
Whether the same pace

The -051 Section

How committed they are
Past experience on large projects
Working with people
Communication style
Track record on 281
Where do you live
How many extracurricular activities
Availability
Good chemistry
Expectations A or A+

We need a cartoon mascot for a class tee shirt. What animal should it be?

The -121 Section

Burmese Rescue Dog

Fat Squirrel

Eagle

C++ Mascot (a Rat)

Hawk

Spider

Lady bug

Octopus

The -051 Section

Road Runner

Panda

Cute spider

Raccoon

Owl

Cheetah

Mole peeking out of a hole

Golden retriever

Are there ways you might use AI ethically and productively in a class like this?

The -121 Section

Anything that doesn't jeopardize
own learning
Translating programs
Ask questions
What does this code do?
Generate testcases
Help with documentation
Explain concepts
Practice exams
Presentation slides

The -051 Section

Generate shell scripts / menial tasks
How to use Google Cloud Platform
Understand a concept
Refresh how to use a new language
Understand someone else's code
Exam prep / example questions
Suggestions / ideas / progress
Brainstorming

How could you decide whether a link is worth crawling or a page is worth indexing?

The -121 Section

- A link is worth crawling if it appears on lots of pages
- Some Top level domains are better than others, .gov > .biz
- Distance from the seed list
- Authority and Hubs
- Accessibility of the page
- Avoid pages with ? parameters in the URL

The -051 Section

- TLD .gov or .edu > .biz
- Correct spelling
- # of links pointing to the page
- # of slashes in the URL
- List of credible websites
- Filetypes
- Duplicate content
- # of words in the title or on the page
- Links in the middle of the text

When you look at search result, how do you decide whether it's a good match?

The -121 Section

Lots of occurrences of the search words

Similar words on the page

Semantic similarity

Keywords in the title

Words near the top

<H1> vs <p>

Important words

The -051 Section

All the terms are on the page and in the title or near the top of the page

Length of page, longer CSS

Recent

When you look at search result, how do you decide whether it's a good match?

The -121 Section

The -051 Section

If you distribute your engine across a lot of machines, should it have a master?

The -121 Section

The -051 Section

What do you think will be the hardest part of building the engine?

The -121 Section

The -051 Section