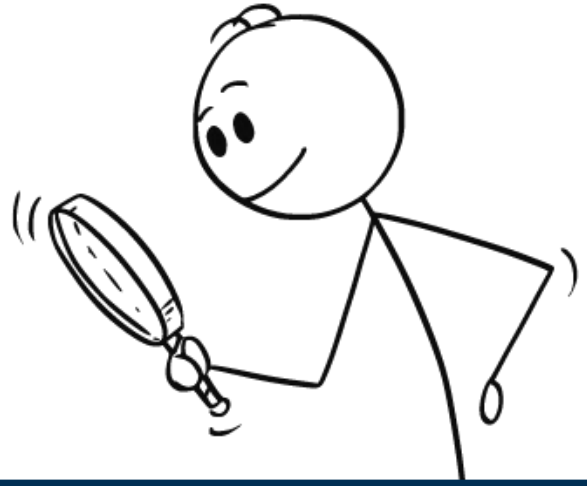


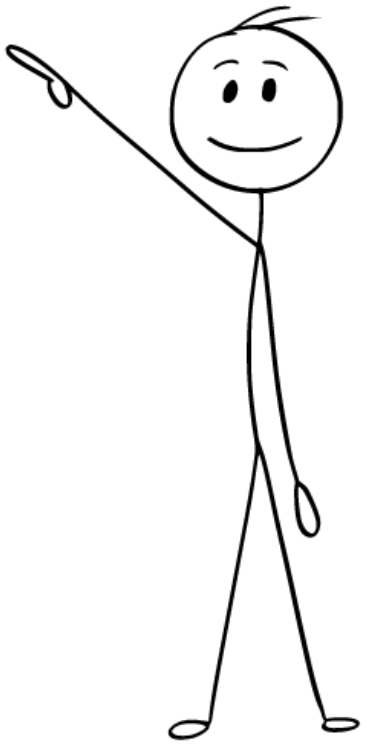
CSE 4/535
Information Retrieval

Sayantan Pal
PhD Student, Department of CSE
338Z Davis Hall

University at Buffalo
Department of CSE

COURSE STAFF

1. Instructor – Sayantan Pal, PhD Student, Department of CSE, 338Z Davis Hall
2. Teaching Assistant – Xixian Yang, PhD Student, Department of CSE, UB, 338Y Davis Hall
3. Graders
    - Debasmit Roy, MS Student Department of CSE, UB
    - Shalini Agarwal, MS Student Department of CSE, UB



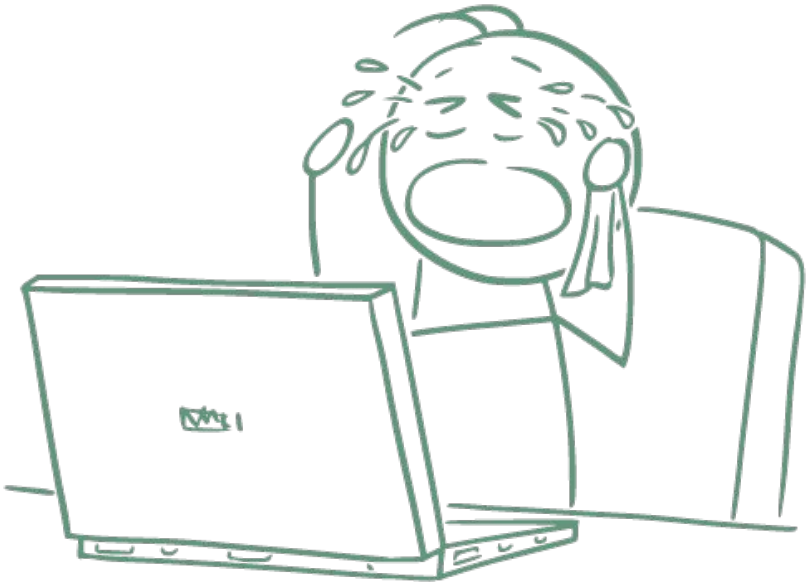
Logistics

1. Website: All course materials, links, schedule, extra resources
2. Piazza – All discussion for the course is hosted here – check regularly
3. Brightspace & UB-HUB: Assignment submission, grades
4. Timberlake Servers: Project Submission (More information coming soon)

Note: Please keep class discussions on Piazza (private/anonymous posts exist) Always include [CSE 435] or [CSE 535] in the subject line when emailing

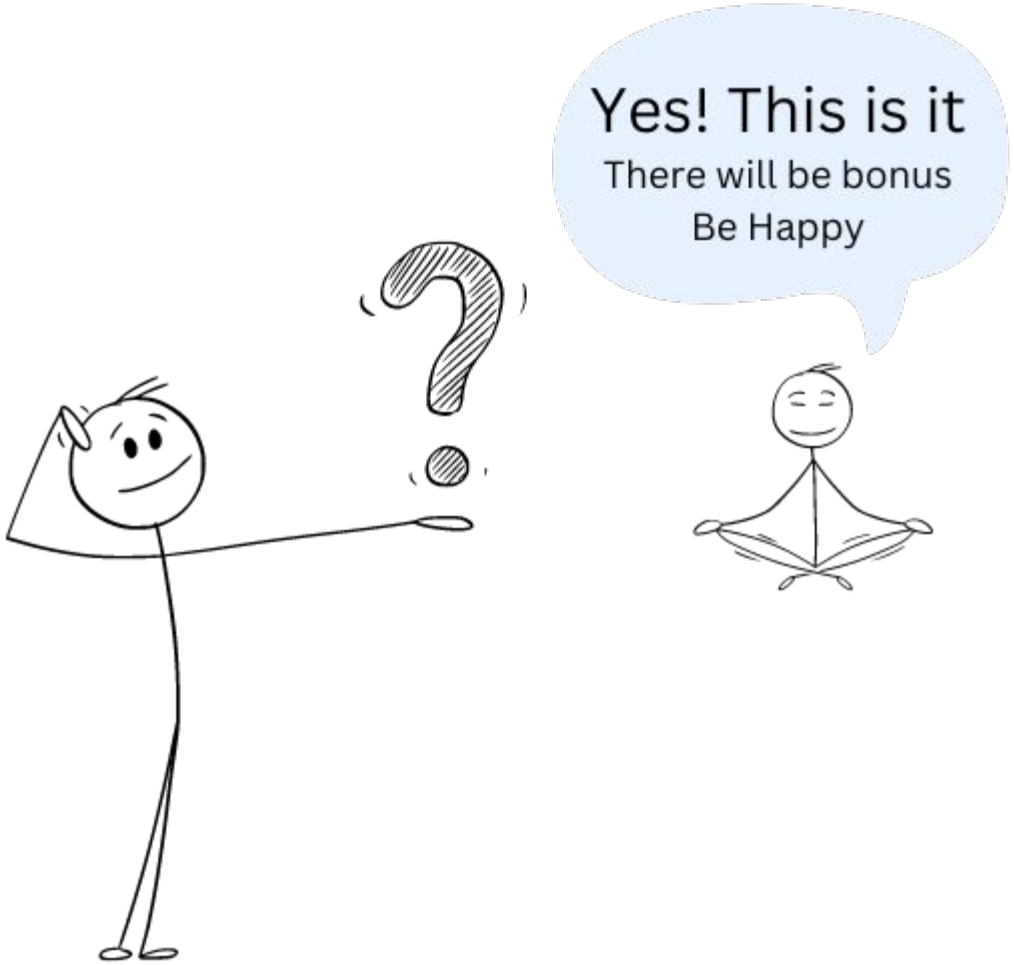
Responsibilities

1. Attend lectures and join the party of knowledge!
2. Dive into books and reference materials!
3. Pop into office hours or light up the Piazza stage with your queries.
4. Tackle the course project/assignments or stock up on tissues!
5. Midterms - Keep those tissues handy
6. **DO NOT** violate AI, tissues won't save you

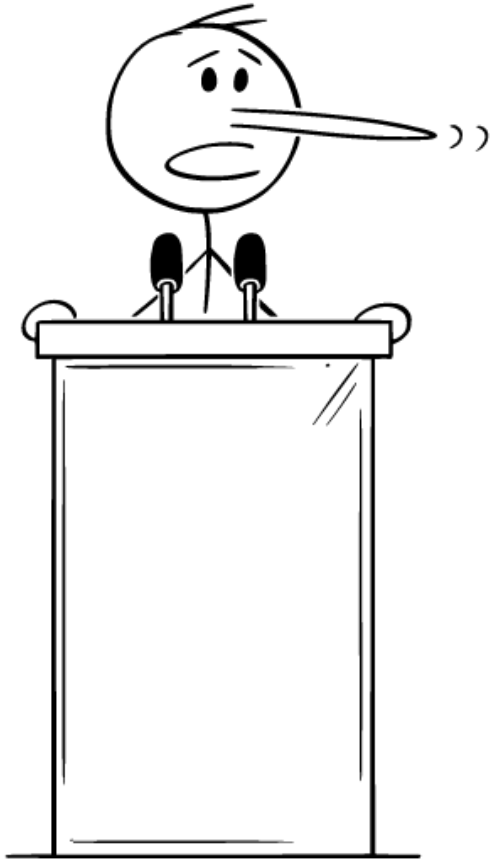


Grading

1. Projects 60%
2. Midterms 40%
3. Will it be curved? Depends on your performance

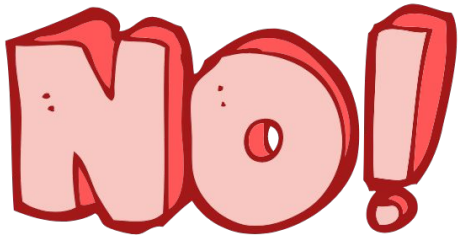


Academic Integrity



Collaboration, AI, Extra Resources

1. Work together to brainstorm ideas
2. Explain concepts to each other
3. Discuss course content
4. Include a list of your collaborators on all submitted work

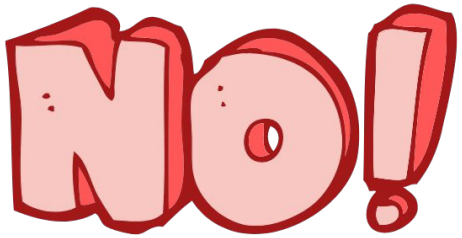


1. Write solutions when working together
2. Describe the details of solutions to others
3. Leave your work in a place where it is accessible by another student

When in doubt, ask a course staff!

Resource Policy

1. Use materials provided by course staff (Piazza, Class, OH)
2. Use materials from the course textbook or readings
3. Cite all materials you reference for written work and code

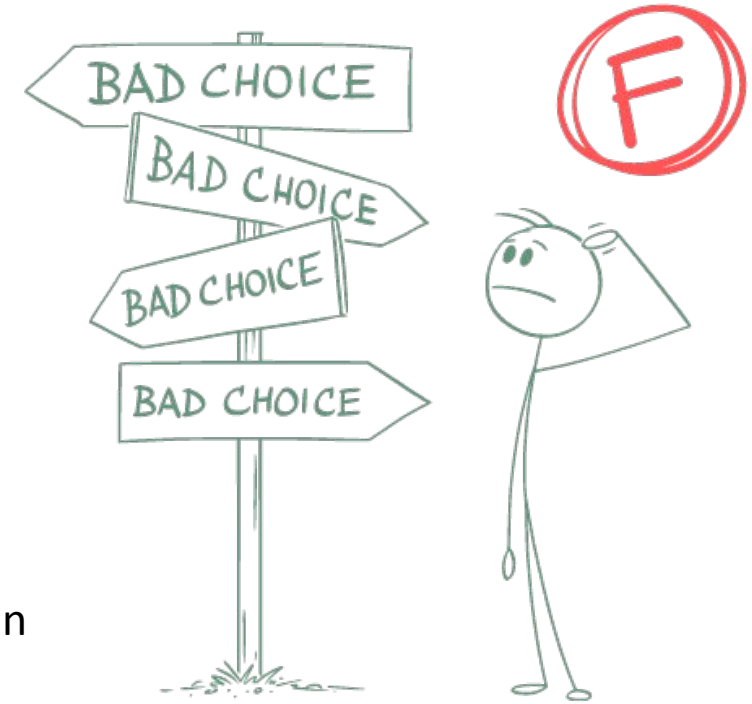


1. Use codes from Github or other sources without citing
2. Use ChatGPT or similar LLMs
3. Use Chegg or similar websites to find solution

When in doubt, ask a course staff!

Other ways to get an F

1. Work in a group by assigning each person to a problem
2. Copying your friend's homework because you forgot
3. Sharing your work with your friend
4. Submitting work without citations
    - Citing outside work will help you avoid AI repercussions
    - We grade you on the work you did, but you won't get an AI violation



Other ways to get an F

You are liable/punishable if someone else submits your work as their own.

Ways to avoid an F

Don’t cheat…

Ways to avoid an F(Amnesty Policy)

Don't cheat… but we understand mistakes are made.

We will grant amnesty for any AI violation **IF** you tell us about it **BEFORE** we discover it

Why does Academic Integrity Matter?

Solutions may exist due to the simplicity of the problems

- Exercises try to force you to think a certain way
- Learning requires simplified/limited problems

You will not understand the design process from a solution

- Experience solving problems isn't obtained from reading solutions
- Anyone (who can read and write) can copy-paste

Exact solutions to every problem don't always exist

- Stack Overflow/ StackExchange (and similar platforms) cannot do your job
- Open source solutions may not do what you need
- Depending on licensing, you can't always use open source solutions in closed source

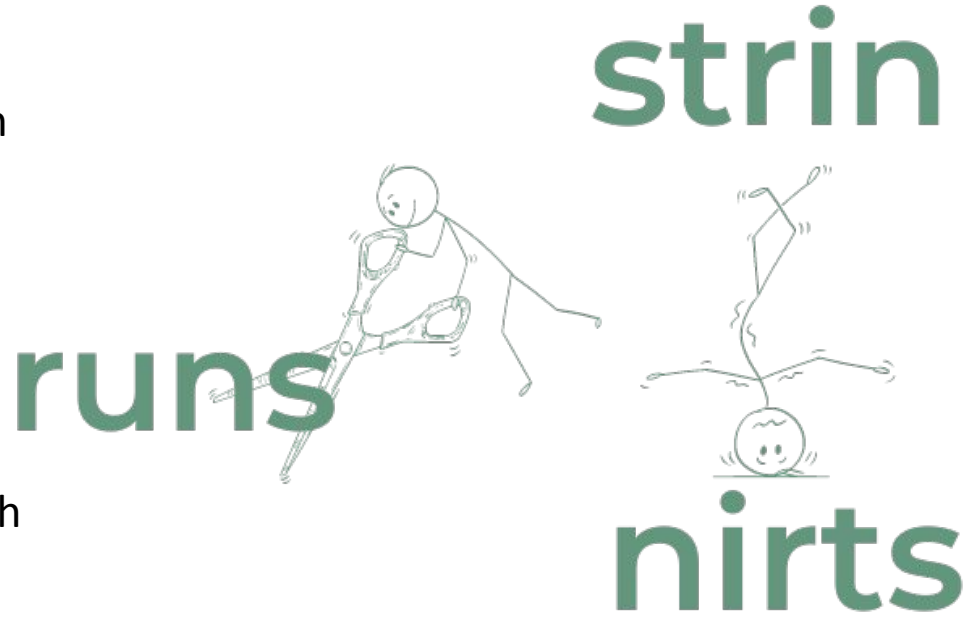
Why does Academic Integrity Matter?

It doesn't just hurt you…but it also hurts the credibility of UB and its graduates!

Coursework

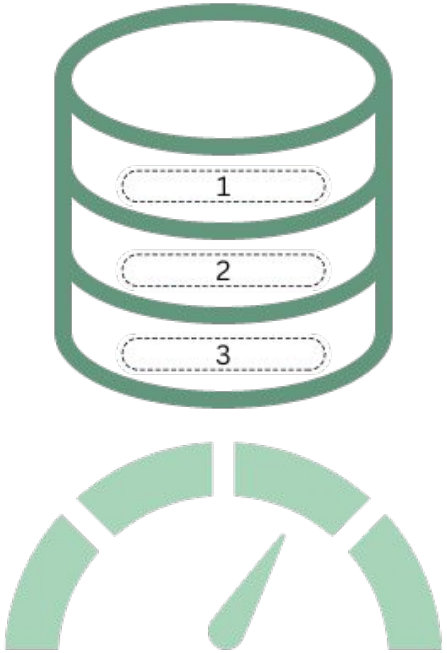
What is this course about

- **Fundamental concepts of IR**: The course kicks off with an introduction to IR, setting the stage by presenting conceptual models and the Boolean Model, which is an early and foundational approach to information retrieval.
- **Text Processing & Analysis**: Subsequent weeks dive into essential text processing tasks like tokenization and techniques for text analysis, such as stemming and using stop lists. This helps in preparing and understanding textual data, ensuring efficient and meaningful retrieval.



What is this course about

- **Indexing & Search Efficiency**: The course emphasizes the importance of indexing by discussing index construction, compression, and techniques to enhance retrieval efficiency. Indexing is pivotal in enabling fast and accurate search.
- **Advanced Text Analysis & Ranking**: Advanced concepts like text properties, scoring, and ranking in IR systems are also explored. The course introduces embeddings and latent semantic indexing, key for capturing semantic information in texts.

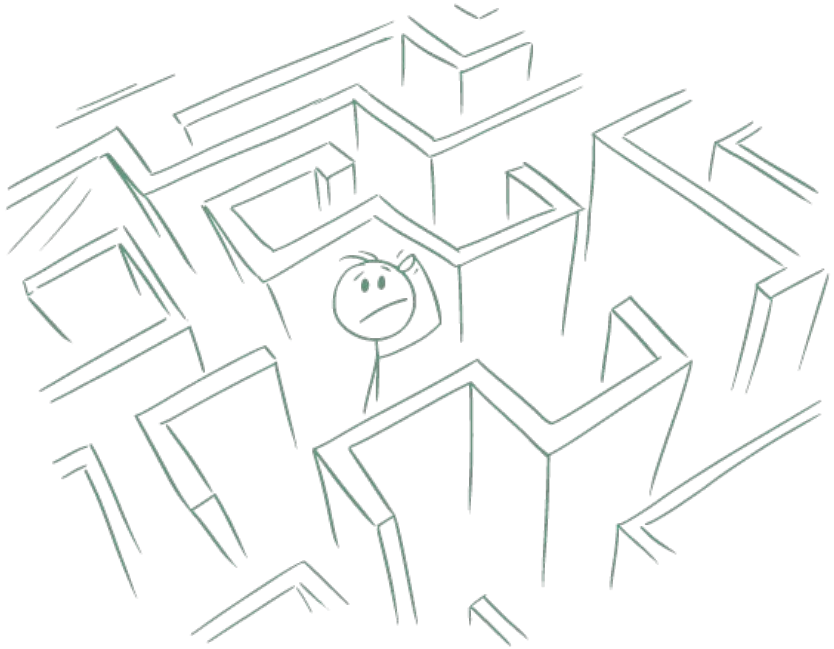


What is this course about

- **Deep Learning & Modern IR**: With a nod to the contemporary state of the field, the course delves into deep learning for IR, discussing transformers and their significant role in modern retrieval systems.
- **Evaluation & Refinement**: Understanding how well an IR system works is crucial. The course, therefore, covers evaluation in IR and methods to refine search results, such as relevance feedback and query expansion.

What is this course about

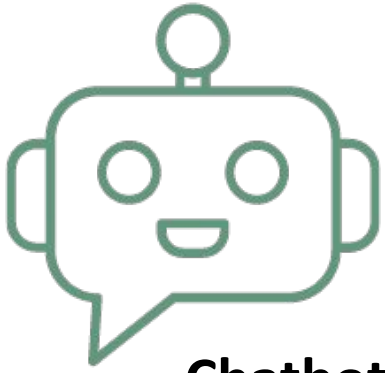
- **Advanced Topics & Applications**: The course delves into machine-learned ranking techniques, explores the intricacies of web search and crawling, introduces social network analysis, and covers cutting-edge areas like word vector models and large language models.
- **Recommendation Systems**: Branching slightly from traditional search, the course touches upon recommendation systems, a closely related field focusing on suggesting items to users.



Applications of IR

Search System

QnA



Chatbots

How can you get the most from the course?

Be eager to learn about an emerging technology in high demand

Focus on opportunities to learn, and grades will come naturally

Work hard to learn new skills and knowledge

Don’t be afraid to dive in and learn new languages/libraries

Be attentive in class

Work on the project yourself

Exercise your collaborative skills only during the final project

How should you assess success in this course?

Not by grade…

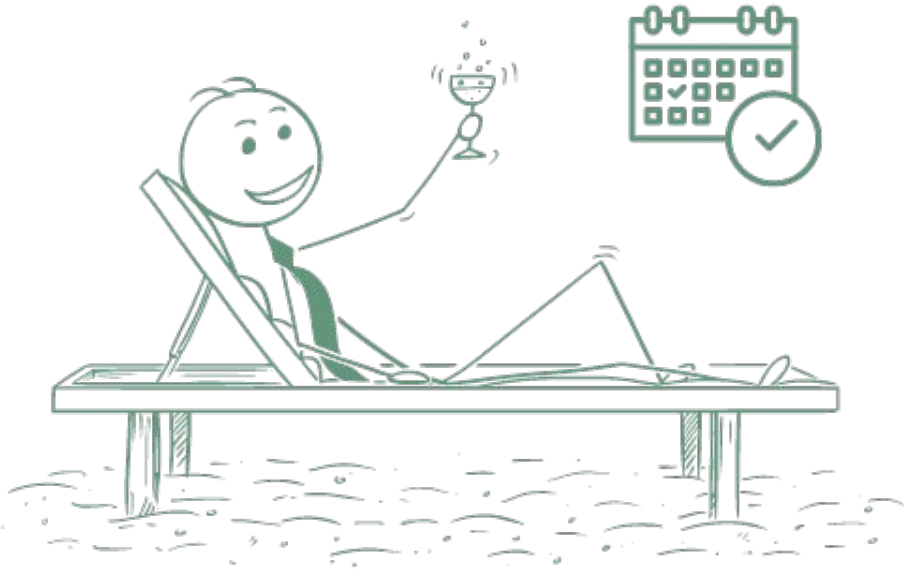
By new concepts you learn about Information Retrieval

By new skills you develop to solve retrieval related problems

By new knowledge you gain about new tools, search engines, retrieval techniques etc.

…do these and the grade will come too

You decide…



References

1. Slides provided by Dr. Eric Mikida (DIC) - Fall 2022
2. Materials provided by Dr. Rohini K Srihari