

CS3245

Information Retrieval

Lecture 0: Course Organization

0



Live Q&A
<https://pollev.com/jin>



Why should you care about

INFORMATION RETRIEVAL?

A keystone of today's tech




Google Search

I'm Feeling Lucky


← ×

RECENT



Pikachu


last seen at 6:21 PM



Slowpoke

last seen at 7:43 PM

4



Bulbasaur

last seen at 7:41 PM

amazon

Account & Lists
Orders

1-16 of over 10,000 results for "information retrieval"

Department

Books

- Computer Network Administration
- Network Storage & Retrieval Administration
- Computers & Technology Databases & Big Data
- Web Development & Design
- See more

Kindle Store

- Computer Databases
- Data Storage & Retrieval
- Computers & Technology
- Computer Programming
- Word Processing
- See more



Introduction to Information Retrieval

by Christopher D. Manning, Prabhakar Raghavan, et al. | Jul 7, 2008

★★★★☆ 30

Hardcover

\$24.75 to rent

\$63.90 to buy

Only 4 left in stock - order soon.

More Buying Choices

\$28.59 (58 used & new offers)

eTextbook

Beginner's Guide to Managing...

★★★★☆ 8

Beginner's Guide to Managing...
prime

Beginner's Guide to Managing...
prime

Beginner's Guide to Managing...
prime

clarksville, tn
×




Clarksville
Tennessee

Sunny - 48°F
3:15 PM

Directions

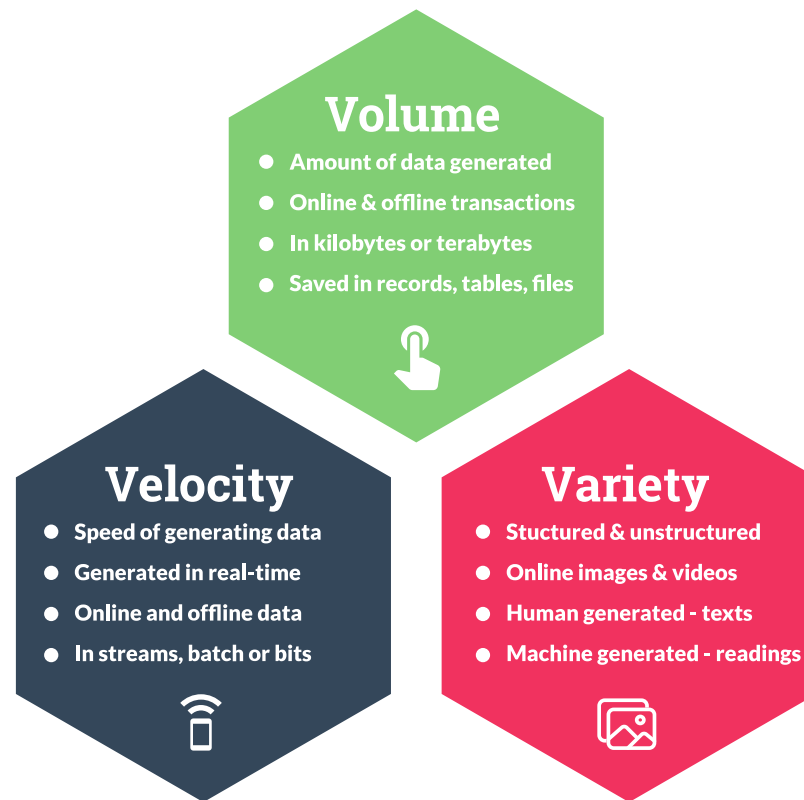
★ SAVE
📍 NEARBY
📱 SEND TO YOUR PHONE
🔗 SHARE

Information Retrieval

3

A necessity in the era of big data

The 3V's of Big Data



<https://blog.hurree.co/blog/the-pros-and-cons-of-big-data>

Part of an ongoing evolution



V 1.0

Directories

- Organized by a human
- Broad subject categories organized by hierarchy
- **Select and read**



V 2.0

Search Engines

- Information retrieval (poor man's NLP) over billions of websites and pages
- Algorithms personalize the searching experience and rank results based on the search criteria
- **Free text search**



V 3.0

Conversational AI

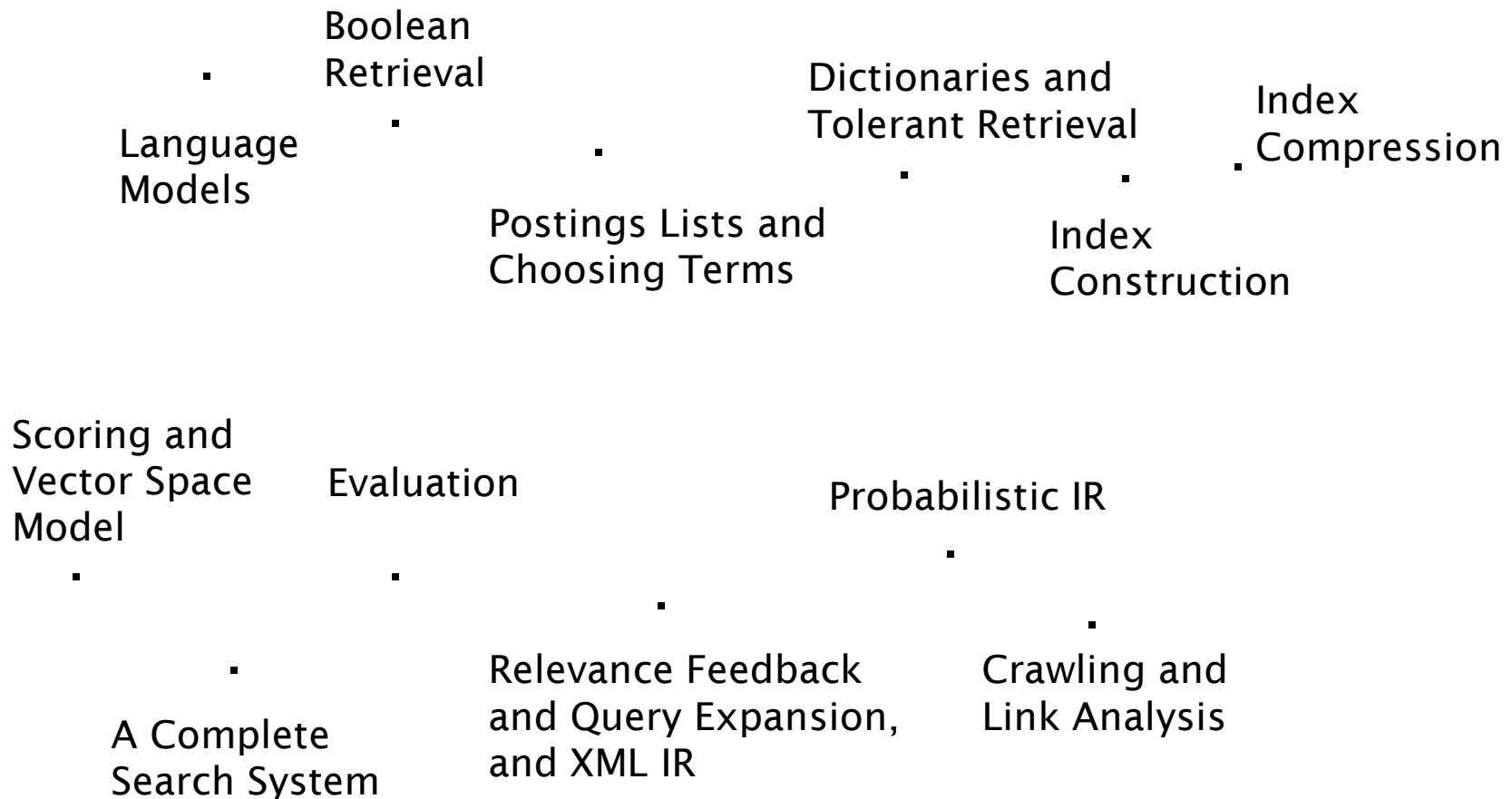
- Advanced natural language understanding using machine learning and computational linguistics
- Knows user history and conversation state based on previous interactions
- Knowledge summarization and task completion
- **Natural language conversations**



<https://medium.com/@sidjreddy/conversational-artificial-intelligence-in-the-context-of-information-revolution-a3257867d50b>

More than just a piece of technology

Focus of the course





And now for the

COURSE ORGANIZATION

Lecturer

Zhao Jin zhaojin@comp.nus.edu.sg

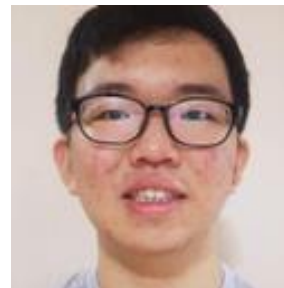
(Just call me “**Jin**” for short)

- Undergraduate, PhD, Instructor, Lecturer and Assistant Dean in SoC
- Programming Methodology, Software Engineering Projects, Information Retrieval
- Travelling, Music and Games



Teaching Assistants

- Jonathan Alain Gabriel
 - E1154491@u.nus.edu
- Lee Zhi Xuan
 - e0564887@u.nus.edu
- Ou Jin Bin
 - E0902046@u.nus.edu
- Teoh Jun Jie
 - e0564824@u.nus.edu



Course activities

- **12** (physical) **lectures** (Weeks 1-13 except Week 10)
 - 12-2pm on the Fridays
 - Recorded on a best effort basis
 - LT15 (AS6 Level 1)
- **6** (physical, **optional**) **tutorials** (odd weeks)
 - 4 groups: 10-11am/11am-12nn (Thu & Fri)
 - Not recorded (solutions to be released afterwards)
 - For discussions, clarifications and **participation marks**
 - SR 9 (COM1-02-09)



Course activities

- **4 homework assignments**
 - maximum group size: **1/2/2/4**
- **12 (online) quizzes** (via Canvas)
- **1 (physical, open-book) final exam**
- **2 (online, optional) help sessions** (via Zoom)
 - 3-5pm on the **Friday of Week 2** for technical issues
 - 10am-12pm on the **Monday of Exam Week 1** for exam matters + Past year paper
 - Recorded on a best effort basis.



Grading

Component	Percentage	Remarks
Participation	5%	Forum + Tutorial (optional)
4 Homework assignments	40% (5%/10%/10%/15%)	Due in Weeks 4/7/11 and Reading Week
12 Quizzes	5% (0.5% each)	Due weekly before the subsequent lecture
Final Exam	50%	30 Apr (Tue) , 5-7pm

Course web sites

<https://www.comp.nus.edu.sg/~cs3245>

- Official website
- General information
- Course materials
- Homework assignments

<https://canvas.nus.edu.sg/>

- Canvas course
- Announcements
- Recordings
- Internal resources (e.g., past year papers)
- Homework submissions



When using a search engine to find our course materials, make sure you find our site for 23/24 Sem II.



CANVAS
BY INSTRUCTURE

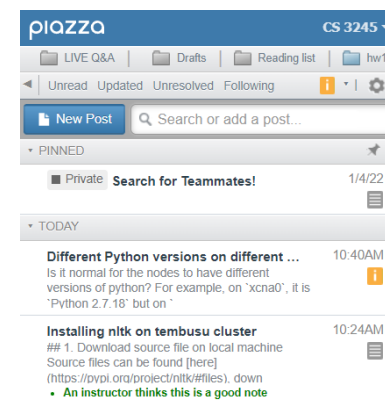
Course web sites

<https://bit.ly/cs3245-2320-forum>

- Piazza forum
- Discussion (+ Teammate recruitment)
- **Participation counts!**

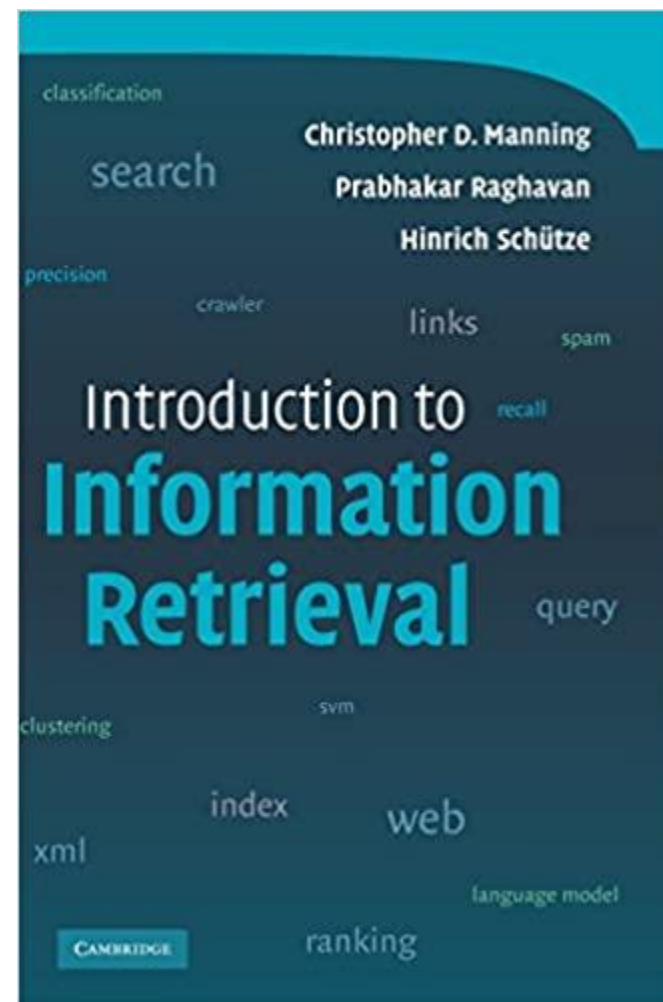
<https://pollev.com/jin>

- PollEverywhere
- Live Q&A during lectures



Required Textbook

- <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>





Python

- To be used for all homework assignments.
- Easy to transition to and comes with many useful libraries (e.g., for text processing and statistics).
- Some materials to be released for your reference.

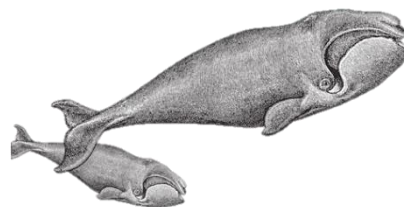
- <https://www.python.org/>



NLTK (Natural Language Toolkit)

- To be used for all homework assignments
 - Some restrictions apply.
- Coded in Python (surprise!)
 - Yes, we know that this is a course on **information retrieval** and not **natural language processing**, but the two are forever intertwined.

- <https://www.nltk.org/>



SoC Compute Cluster + Slurm

- A high-performance computing cluster in SoC
 - For running (and grading) your homework submissions, which can be quite compute intensive.
 - Python Version: 3.10.12
 - Managed by Slurm Workload Manager

- Activation

- <https://mysoc.nus.edu.sg/~myacct/services.cgi>

← Login w/ your
NUSNET account

The SoC Compute Cluster is a Unix based compute cluster on which you can run your compute intensive jobs. You can also use this to do parallel computing experiments.

This service is currently enabled

Disable

SoC Compute Cluster + Slurm

■ Login

- Turn on VPN if you are off-campus
- Use ssh or any other similar tools
- Server: xlog[0-2].comp.nus.edu.sg
- Login w/ Your **SoC Unix** account

```
C:\Users\dcszjin>ssh zhaojin@xlog0.comp.nus.edu.sg
zhaojin@xlog0.comp.nus.edu.sg's password:
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.4.0-126-ge
```

Apply for one @
<https://mysoc.nus.edu.sg/~newacct/>
if you don't have one.

■ User guides

- Compute Cluster Guide:
<https://dochub.comp.nus.edu.sg/cf/guides/compute-cluster/start>
- Slurm Quick Start Guide:
<https://dochub.comp.nus.edu.sg/cf/guides/compute-cluster/slurm-quick>

To-dos

- **Install** Python + NLTK on your computer if you plan to work on your homework assignments locally.
 - Remember to match the Python version in the Cluster.
- **Familiarize** yourself with the Cluster + Slurm
 - Install NLTK on your account in the Cluster
 - Try to run Python programs in the cluster using Slurm
- Any technical issues...
 - Post in the forum
 - Attend the technical help session

Facebook Rule

- Discussion is acceptable and encouraged.
- However, you **should not take any written (electronic or otherwise) record away from a discussion.**
- After the discussion, **do something else for at least a half-hour** (Facebook / Instagram / Telegram, or doing an assignment for a different class), before switching back to CS3245.
- This will assure that you are able to reconstruct what you learned from the discussion, **by yourself, using your own brain.**

Freedom of Information Rule

A faint silhouette of a person walking a dog is visible in the background of the slide, positioned behind the title.

- If you collaborate with other students for your assignments but are not making a group submission with them, you must **always fill in the name(s) of your collaborators on your assignment.**
- You will be assessed for the parts for which you claim is your own contribution.
- Failing to do so will be considered as plagiarism.

No-Sponge Rule

- For homework assignments which group submissions are allowed, **every member in the group must actively contribute.**
- Members of the group have the responsibility to **NOT**
 - tolerate anyone who is putting forth no effort
 - let anyone who is making a good faith effort "fall through a crack"
- We want to know about dysfunctional group situations as early as possible.



Use of AI Tools

- You are allowed to use AI tools for learning purposes.
- However, for all homework assignments, you must **implement the core logic on your own.**

Plagiarism

- **Source code plagiarism** is a serious violation of NUS code of student conduct.
 - <https://studentconduct.nus.edu.sg/wp-content/uploads/2023/03/NUS-Code-of-Student-Conduct.pdf>
- We adopt a "no mercy" policy when it comes to disciplinary action on plagiarism.
 - Both parties, the student who copied, and the student who allowed others to copy, will be penalized equally.
- Bottom Line: write your own code and do **NOT** share it with anyone!