

CS3245

# Information Retrieval

# 0

Lecture 0: Course Organization



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



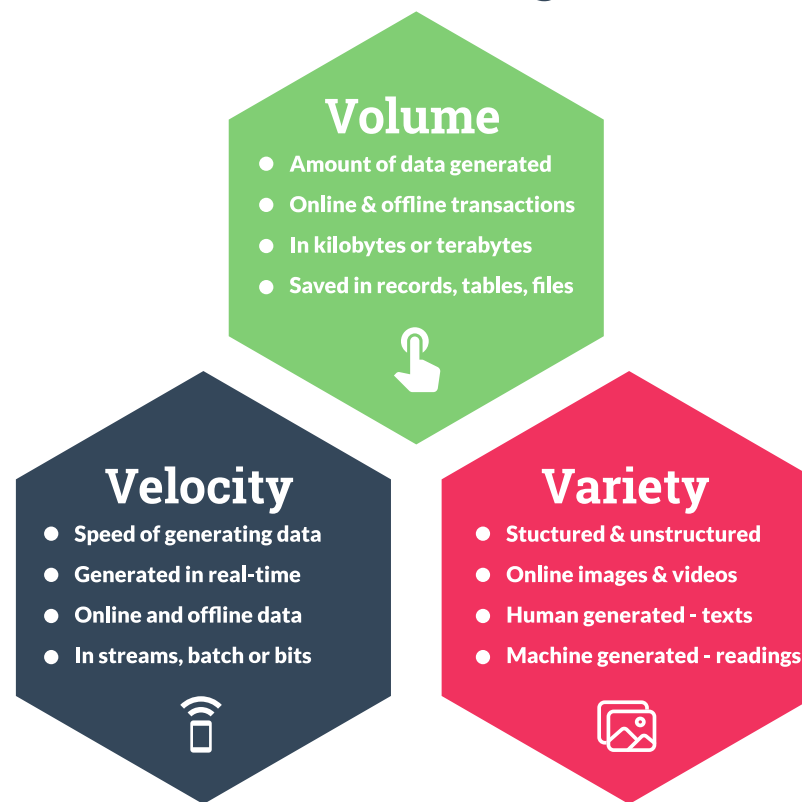
Why should you care about

# INFORMATION RETRIEVAL?

# Era of big data



## The 3V's of Big Data



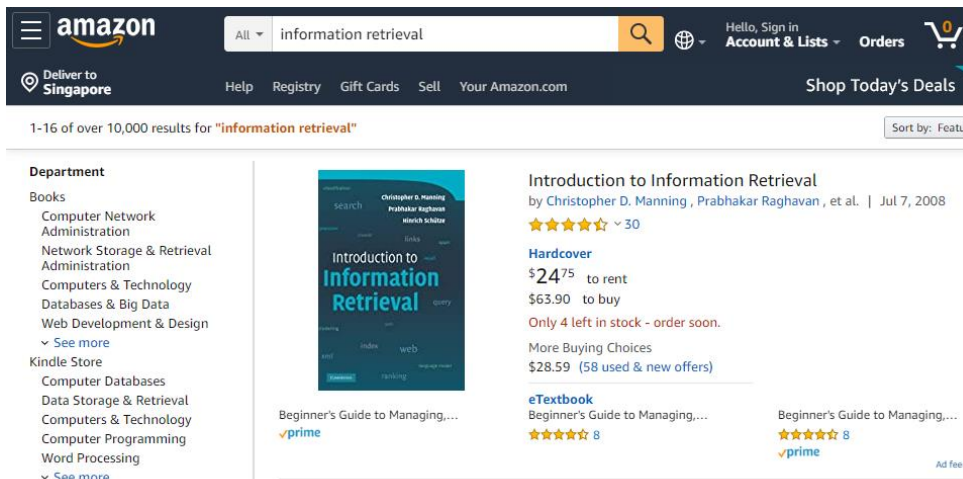
<https://blog.hurree.co/blog/the-pros-and-cons-of-big-data>

# A keystone of today's tech





Google Search    I'm Feeling Lucky



amazon    information retrieval    Hello, Sign in Account & Lists    Orders    Shop Today's Deals

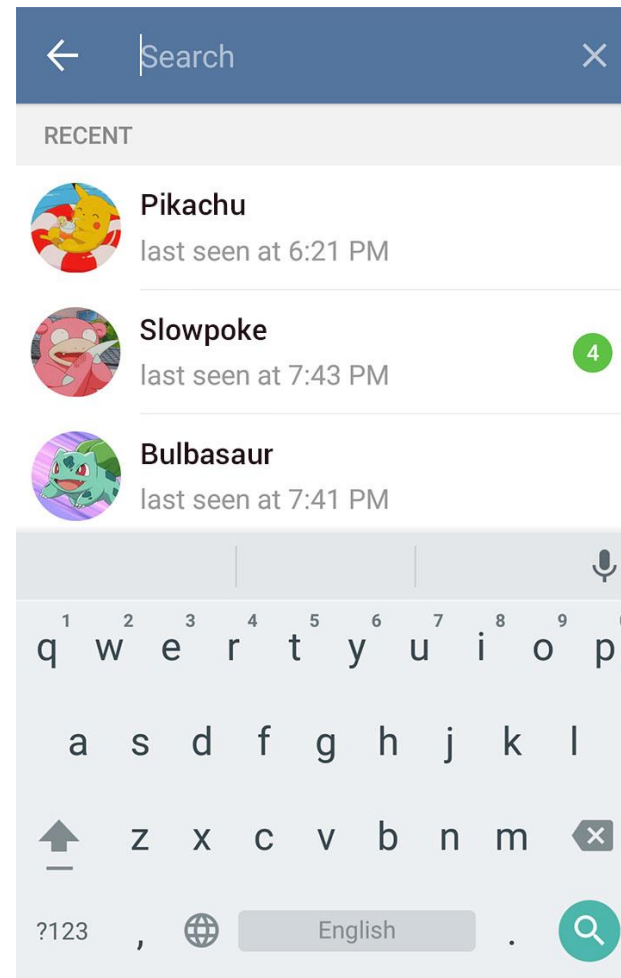
1-16 of over 10,000 results for "information retrieval"    Sort by: Feature

**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  
 prime



← Search ×

RECENT

-  **Pikachu**  
last seen at 6:21 PM
-  **Slowpoke**  
last seen at 7:43 PM    4
-  **Bulbasaur**  
last seen at 7:41 PM

q w e r t y u i o p  
 a s d f g h j k l  
 z x c v b n m  
 ?123 , English .

# 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>

# Jobs everywhere



Information retrieval in Worldwide Set alert

11,701 results

**Librarian, Research Data Specialist**  
Nanyang Technological University Singapore  
Singapore, Singapore (On-site)

3 weeks ago · 1 applicant

**Bibliotekarie med fokus på användarmötet till universitetsbiblioteket**  
Karolinska Institutet  
Stockholm, Stockholm County, Sweden (Hybrid)

1 week ago · 7 applicants

**Développeuse-Analyste ou développeur-analyste en RI - Niv 2**  
Université Laval  
Québec, QC (On-site)

3 weeks ago · 1 applicant

**Chart Retrieval Analyst I**  
Summa Health  
Akron, OH (On-site)

## Librarian, Research Data Specialist

→ ...

Nanyang Technological University Singapore · Singapore, Singapore (On-site) 3 weeks ago · 1 applicant

- Full-time · Entry level
- 10,001+ employees · Higher Education
- 1 connection · 354 company alumni
- See recent hiring trends for Nanyang Technological University. [Try Premium for free](#)
- Actively recruiting

Apply 
Save

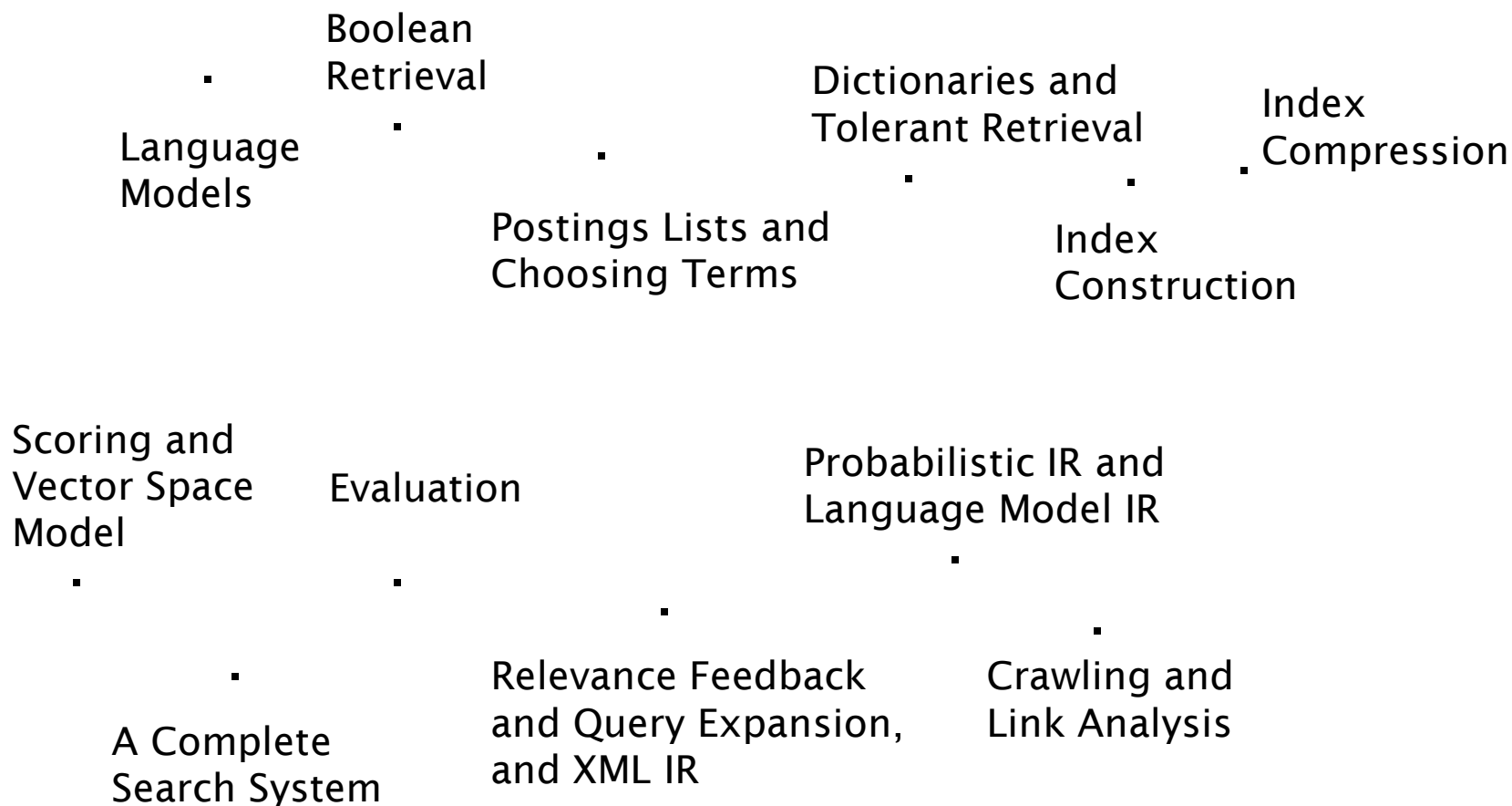
We invite applications for the position of Research Data Specialist in our Institutional Repositories Team in the Office of Information, Knowledge and Library Services (OIKLS), Nanyang Technological University (Singapore). We are looking for an individual with exposure to administering web applications and keen to learn more on research data curation and archival in an academic setting.

The successful candidate will work closely with the Team, to provide services that include repository application management and advocacy for researchers on data curation, data sharing, and metadata standards. This position will report to the Team Lead of Institutional Repositories in OIKLS.

# More than just a piece of technology

---

# Focus of the module







And now for the

# **COURSE ORGANIZATION**

# Lecturer

Zhao Jin [zhaojin@comp.nus.edu.sg](mailto:zhaojin@comp.nus.edu.sg)

(Just call me “Jin” for short)

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



# Teaching Assistants

- He Tingyu
  - [e0920777@u.nus.edu](mailto:e0920777@u.nus.edu)
- Lim Junwei
  - [e0543963@u.nus.edu](mailto:e0543963@u.nus.edu)
- Tan Jun, Nigel John
  - [e0559733@u.nus.edu](mailto:e0559733@u.nus.edu)
- Wong Man Chun
  - [e0550621@u.nus.edu](mailto:e0550621@u.nus.edu)
- Zhong Suyang
  - [e0989374@u.nus.edu](mailto:e0989374@u.nus.edu)



# Course activities

- **12** (physical) **lectures** (Weeks 1-11 & 13)
  - 12-2pm on the Fridays
  - Recorded on a best effort basis
  - LT15 (AS6 Level 1)
- **5** (physical, **optional**) **tutorials** (Weeks 3,5,7,9,11)
  - 5 groups: 10-11am/11am-12nn (Thu & Fri), 12-1pm (Thu)
  - Not recorded (solutions to be released afterwards)
  - For discussions, clarifications and **participation marks**
  - SR15 (COM3-01-25)
  - One more in Week 13 via a pre-recorded video



# Course activities

- **4 homework assignments**
  - maximum group size: **1/2/2/4**
- **1 (take-home, open-book) midterm**
- **1 (physical, open-book) final exam**
- **2 (online, optional) help sessions (via Zoom)**
  - 3-5pm on the **Friday of Week 2** for technical issues
  - 3-5pm on the **Friday of Reading Week** for exam matters
  - Recorded on a best effort basis.

# Grading



Component	Percentage	Remarks
Participation	5%	Forum + Tutorial (optional)
Homework assignments	40% (5%/10%/10%/15%)	Due in <b>Weeks 4/7/10 and Reading Week</b>
Midterm	10%	Due in <b>Recess Week</b>
Final Exam	45%	<b>25 Apr, 9-11am</b>

# 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 22/23 Sem II.



**CANVAS**  
BY INSTRUCTURE

# Course web sites

<https://bit.ly/cs3245-2220-forum>

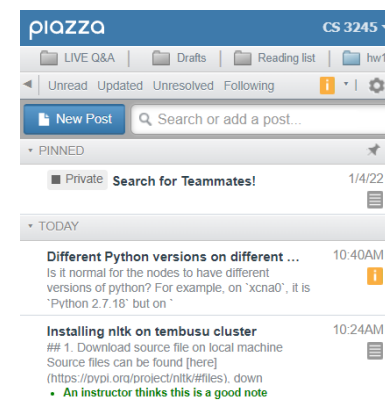
- Piazza forum
- Discussion (+ Teammate recruitment)
- **Participation counts!**

■ <https://pollev.com/jin>

- PollEverywhere
- Live Q&A during lectures

<https://www.facebook.com/cs3245>

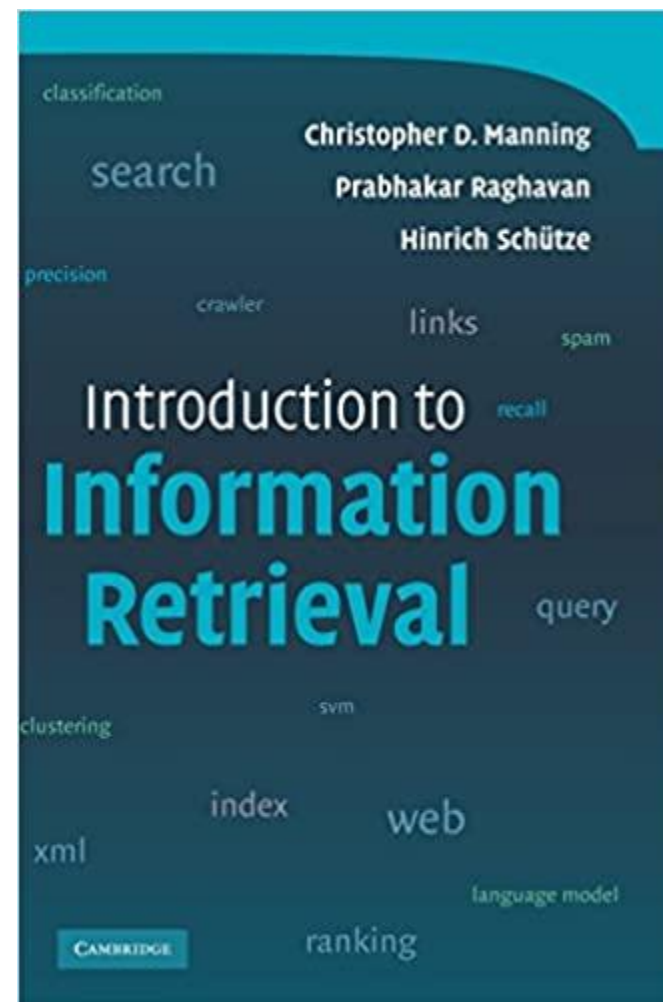
- FB page
- IR News / resources / opportunities





# Required Textbook

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





# Python

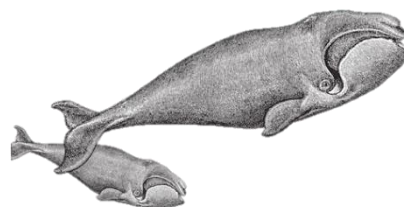
- To be used for all homework assignments.
- Easy to transition to and comes with many useful libraries (e.g., for text processing and statistics).
- Some materials to be released for your reference.
- <https://www.python.org/>



# 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: 3.8.10
  - Managed by Slurm Workload Manager
- Activation
  - <https://mysoc.nus.edu.sg/~myacct/services.cgi>

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
  - Username / Password: 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-generic x86_64)
```

- 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

# Freedom of information rule

A faint, light-colored silhouette of a person walking a dog is positioned in the upper right area of the slide, behind the main title.

- Collaboration is acceptable and encouraged.
- To assure that all collaboration is on the level, **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.

# Facebook rule

- You are free to meet with fellow students(s) and discuss assignments with them.
- Writing on a board or shared piece of paper is acceptable during the meeting; however, you **may not take any written (electronic or otherwise) record away from the meeting.**
- After the meeting, do something else for at least a half-hour (Facebook / Instagram / Telegram, or doing an assignment for a different class), before working on the assignment.
  - This will assure that you are able to reconstruct what you learned from the meeting, **by yourself, using your own brain.**
  - You will be asked to certify that you meet this requirement per assignment.