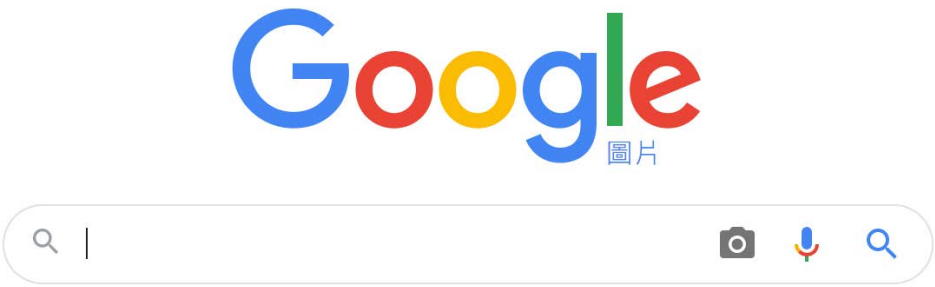
Course Overview

Kuan-Yu Chen (陳冠宇)

2020/09/18 @ TR-514, NTUST

Information Retrieval

- Information Retrieval is a broad area of computer science
  - It mainly focuses on proving the **users** with easy access to **information** of **their interest**



- If we focus on "text" data, IR deals with the representation, storage, organization of, and access to information items
  - documents, Web pages, online catalogs, structured records, multimedia objects

The Goal of IR.

- Early goals of the IR area
  - Indexing text
  - Searching for useful documents in a collection

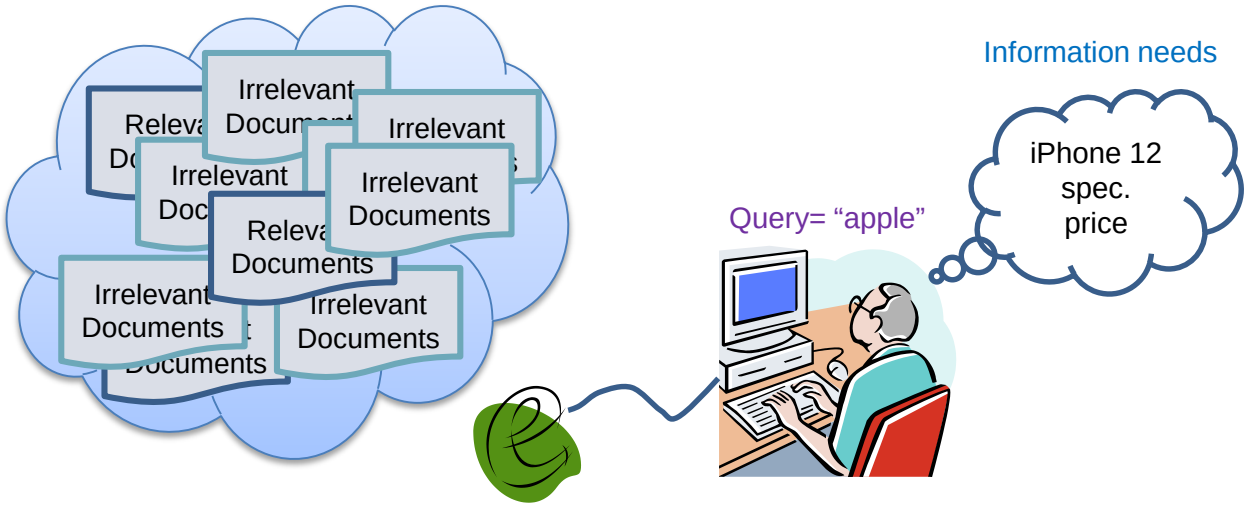
- Nowadays, research in IR
  - Modeling
  - Web search
  - Text classification
  - Systems architecture
  - User interfaces
  - Data visualization
  - Filtering
  - Languages

The Goal of IR..

- In terms of research, the area may be classified into two distinct and complementary points
  - Computer-centered
    - Building efficient indexes/representations
    - Processing user queries with high performance
    - Developing ranking algorithms
  - Human-centered
    - Studying the behavior of the user
    - Understanding the information need

IR Problems.

- Users of modern IR systems, such as search engine users, have **information needs** of varying complexity
  - A full description of the user information need is not a good query to be submitted to the IR system
  - Instead, the user translate this information need into a query
    - A set of ***keywords***, or ***index terms***, which summarize the user information need
  - The key goal of the IR system is to retrieve information that is useful or relevant to the user's information need



IR Problems..

- That is, the IR system must rank the information items according to a degree of relevance to the user query

- The definition of the IR problem
  - ***The key goal of an IR system is to retrieve all the items that are relevant to a user query, while retrieving as few nonrelevant items as possible***

- The notion of **relevance** is of central importance in IR

About “Relevance”

- Relevance is a personal assessment that depends on the task being solved and its context

- Relevance can change with
  - Time
  - Location
  - Device
  - ….

Until now, no IR system can provide perfect answers to all users all the time!

Tentative Grading

- Homework: 85%
  - Programming with C/C++/Python/Matlab
    - HW1: 15%
    - HW2: 15%
    - HW3: 10%
    - HW4: 15%
    - HW5: 15%
    - HW6: 15%
    - For each HW, you should submit codes and a report, and the report accounts for 3 points of the score
- Attending the Invited Speech: 5%
- Final Project: 15%
  - You should present your work in class

Tentative Syllabus

Date	Syllabus	Homework
9/18	Course Overview	
9/25	Break for Rocling2020	
10/2	Holiday for Moon Festival	
10/9	Holiday for National Day	
10/16	Classic Models	Homework-1 (deadline: 10/29 23:59)
10/23	Extended Probabilistic Models	Homework-2 (deadline: 11/5 23:59)
10/30	Evaluation & Benchmark Collections	Homework-3 (deadline: 11/12 23:59)
11/6	Latent Semantic Analysis	
11/13	Topic Models	Homework-4 (deadline: 11/26 23:59)
11/20	Search Results Diversification	
11/27	Pseudo-Relevance Feedback & Query Models	Homework-5 (deadline: 12/10 23:59)
12/4	Talk	Submit Your Member List!
12/11	Representation Learning for Information Retrieval	
12/18	Supervised Retrieval Models & Information Retrieval in Practice	Homework-6 (deadline: 12/31 23:59) & Submit Your Paper Title!
12/25	Break for Your Final Project	
1/1	Holiday for Founding Anniversary	
1/8	Presentation-1	
1/15	Presentation-2	

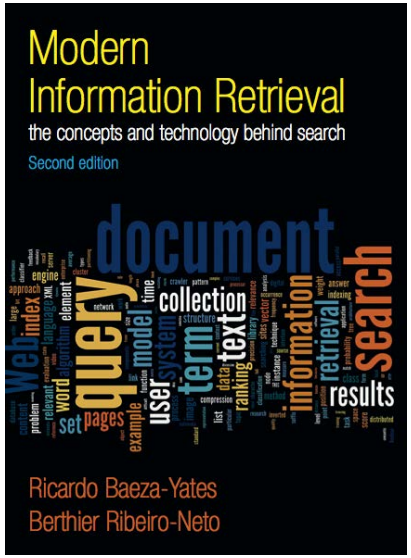
Instructor & TA

- Instructor: 陳冠宇
  - T4-508, Thursday 9:00~16:00
  - kychen@mail.ntust.edu.tw
  - (02) 2737-6377
  - http://faculty.csie.ntust.edu.tw/~kychen/courses/2020_Fall_IR/2020_IR.html

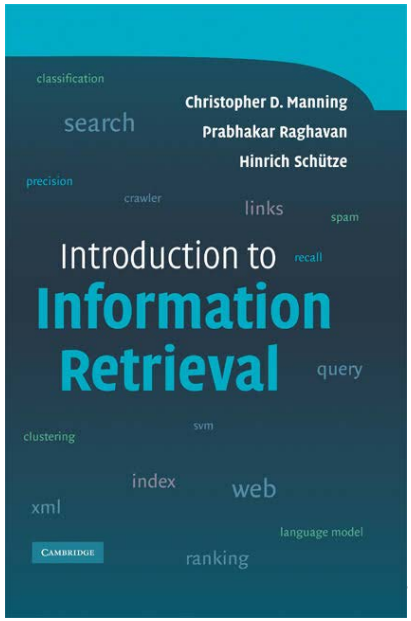
- TA:林崇恩 & 余福浩 (E1-222-3)

References.

- R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition), ACM Press, 2011



- C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008



References..

- Conferences
  - ACM Annual International Conference on Research and Development in Information Retrieval (SIGIR)
  - International Joint Conferences on Artificial Intelligence (IJCAI)
  - ACM Conference on Information Knowledge Management (CIKM)
  - Annual Meeting of the Association for Computational Linguistics (ACL)
  - International Conference on Learning Representations (ICLR)

- Journals
  - Journal of the American Society for Information Science (JASIS)
  - ACM Transactions on Information Systems (TOIS)
  - Information Processing and Management (IP&M)
  - ACM Transactions on Asian Language Information Processing (TALIP)
  - Information Retrieval Journal (IRJ)

Questions?

kychen@mail.ntust.edu.tw