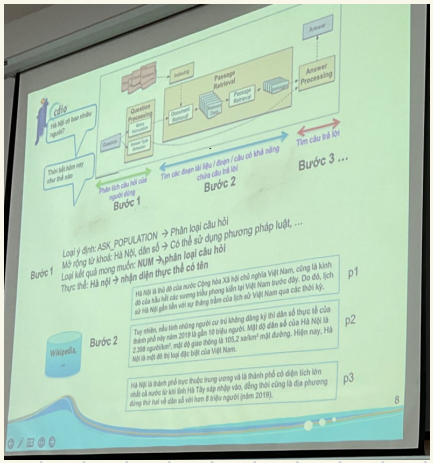
KTNLVB 070923

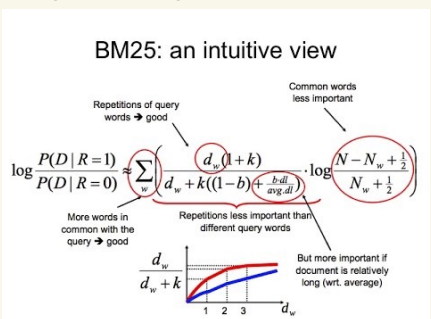
Tìm kiếm thông tin trên văn bản
B1: Phân loại câu hỏi
Nhận diện các thực thể trong câu hỏi
Mở rộng câu truy vấn
B2: Truy tìm kiếm thông tin

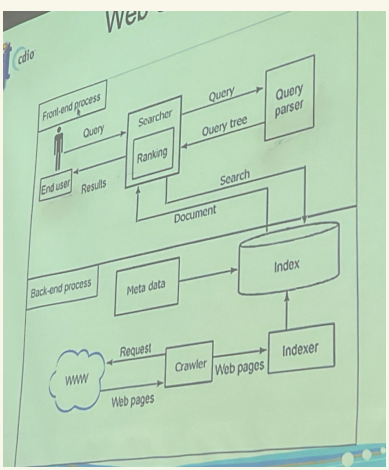


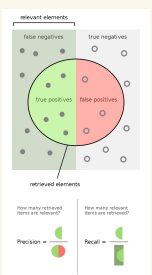
TF-IDF càng cao thì từ đó càng quan trọng
corpus: kho ngữ liệu

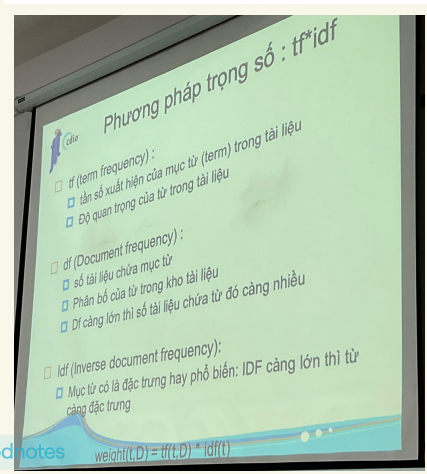
Tính điểm liên quan của D và Q
D tập từ khoá của document
Q là tập từ khoá của câu truy vấn

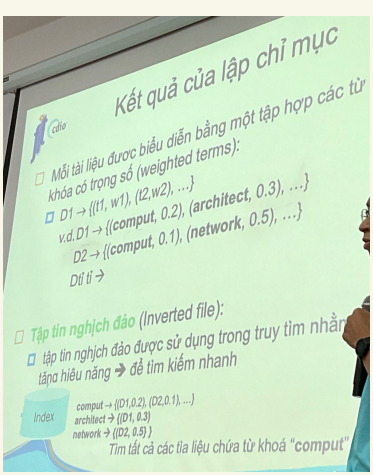
Các cách tìm:
- Theo biểu thức logic (ví dụ Q = Computer ^ Network, thì D phải có bao gồm cả 2 từ đó
- BM25 - Những tài liệu càng ngắn thì càng được ưu tiên











MAP (Mean Average Precision)

$$MAP = \frac{1}{n}\sum_{Q_i}\frac{1}{|R_i|}\sum_{D_j \in R_i}\frac{j}{r_{ij}}$$

- r_{ij} = thứ hạng rank of the j-th relevant document for Q_i
- $|R_i|$ = #rel. doc. for Q_i
- n = # test queries

Q1	Q2	j
1	4	1
5	8	2
10		3

$$MAP = \frac{1}{2}[\frac{1}{3}(\frac{1}{1}+\frac{2}{5}+\frac{3}{10})+\frac{1}{2}(\frac{1}{4}+\frac{2}{8})]$$

sb_Webui_controller
Photoshop Beta’s Generative Fill (uncropt)