

Reminder

Kuan-Yu Chen (陳冠宇)

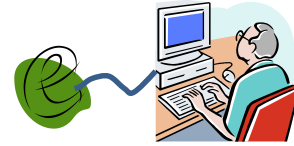
2020/12/04 @ TR-313, NTUST

HW5

#	Team Name	Score ?
1	80847002S_羅天宏	0.59666
2	好想要不勞而獲喔	0.57917
3	너는나만큼(👉這首歌好聽)	0.56469
4	我家宏哥修4次	0.55121
5	快樂ㄟ甘蔗man😁	0.54701
6	jjj_liao	0.53853
7	B10615034_黃柏翰	0.53850
8	Enilesab	0.53813
9	For Fun	0.53489
10	是在跟我開什麼玩笑	0.53312
11	SummerSHIT	0.53301
12	test1111112312	0.53301
13	hongyun	0.52940
14	0.52574
15	壞ㄌ	0.52519
📍	Baseline: Rocchio	0.52495
16	Test	0.52276
17	M10915201_陳牧凡	0.48970
18	M10815048	0.48965
19	M10815111_謝公耀	0.48939
20	TESTT1235	0.48938

#	△pub	Team Name	Score ?	Entries	Last
1	—	80847002S_羅天宏	0.56026	15	12m
2	—	好想要不勞而獲喔	0.55615	35	1h
3	—	너는나만큼(👉這首歌好聽)	0.55208	24	1h
4	▲5	For Fun	0.53006	15	7h
5	▼1	我家宏哥修4次	0.52816	49	1d
6	▲2	Enilesab	0.51904	46	4d
7	—	B10615034_黃柏翰	0.50731	58	10h
8	▼3	快樂ㄟ甘蔗man😁	0.50719	19	6h
9	▼3	jjj_liao	0.50701	19	2d
10	▲5	壞ㄌ	0.50404	31	12h
11	—	SummerSHIT	0.49889	3	4d
12	—	test1111112312	0.49889	25	4d
📍		Baseline: Rocchio	0.49729		
13	▼3	是在跟我開什麼玩笑	0.49726	2	5d
14	▼1	hongyun	0.49183	10	1d
15	▲1	Test	0.48618	34	18h
16	▼2	0.47316	2	10h
17	▲9	B10515025	0.47157	14	10h
18	▼1	M10915201_陳牧凡	0.46785	13	11h
📍		FYI: BM25 (k1=0.8 b=0.7)	0.46639		
19	▼1	M10815048	0.46593	18	16h
20	▼1	M10815111_謝公耀	0.46570	16	18h

KL-Divergence Measure



- Kullback-Leibler (KL)-Divergence measure is a LM-based IR method

$$KL(q||d_j) = \sum_{w \in V} P(w|q) \log \frac{P(w|q)}{P(w|d_j)} \propto - \sum_{w \in V} P(w|q) \log P(w|d_j)$$

- A query is treated as a **probabilistic model** rather than simply an **observation**
- KL-divergence supports us to achieve a better result by considering **both** query and document models

About Final Project

- Please submit your member list today!
 - 2~4 team members
 - Choose a paper

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	Statistical 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	

Hand in Retroactively

- Homework 1~4 can be handed in retroactively
 - You can get 4 points if you can outperform “baseline”
 - Show a picture to prove the result
 - Source codes and report should be uploaded
 - The **HARD** deadline is 1/11
- Please follow our rules
 - **Don't cheat yourself, your friends, and me!**
 - **Don't create multiple accounts!**
 - Implement the model or measurement by **YOUSELF!**

Questions?



kychen@mail.ntust.edu.tw