RMIT at the NTCIR-13 We Want Web Task
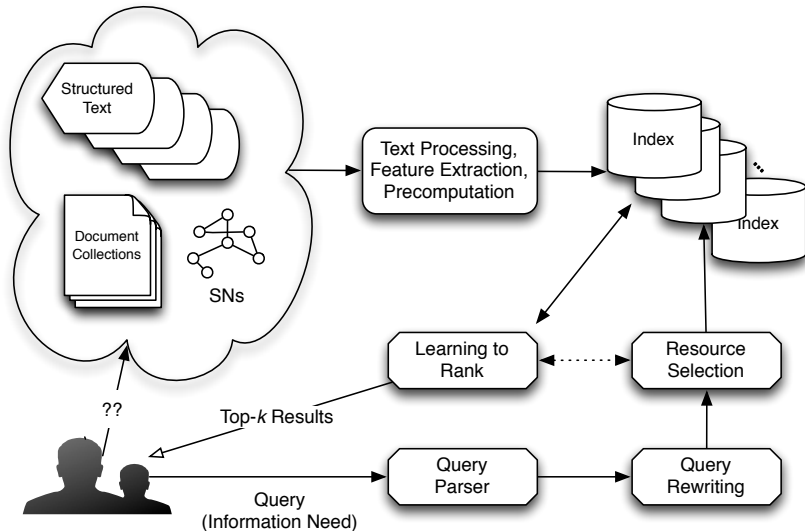
Luke Gallagher

with Joel Mackenzie, Rodger Benham, Ruey-Cheng Chen,
Falk Scholer, and J. Shane Culpepper

School of Science (Computer Science)
RMIT University

NTCIR '17 (December 8, 2017)

WWW English Subtask

- RMIT submitted four systems for the English subtask

- Classic effectiveness techniques:
    - Term dependencies (FDM, SDM)
    - Query Expansion
    - Field extents (`title`, `inlink`, `body`)

- Static document features:
    - PageRank
    - Spaminess

Victor Lavrenko and W. Bruce Croft. In: *Proc. SIGIR*. 2001.

D. Metzler and W. B. Croft. In: *Proc. SIGIR*. 2005.

System Configurations: English Subtask

- **RMIT-1:** SDM Fields $+$ RM3 Query Expansion $(10, 50, 0.6)$

- **RMIT-2:** Linear combination of **RMIT-1** $+ 0.25 \times$ PageRank Priors

- **RMIT-3:** FDM $+$ RM3 Query Expansion $(20, 10, 0.8)$

- **RMIT-4:** n-gram Fields $+$ RM3 Query Expansion $(10, 50, 0.6)$

Post-retrieval spam filtering was applied to all systems except **RMIT-1**. Documents with a spam score less than 70 were removed from retrieved results.

Gordon V. Cormack, Mark D. Smucker, and Charles L. Clarke. In: *Inf. Retr.* (2011).

Structured Fields-Based Query

"big red house"

```
#weight(
   α₁ #combine(big.title red.title house.title)
   α₂ #combine(big.inlink red.inlink house.inlink)
   α₃ #weight(
       β₁ #combine(big.body red.body house.body)
       β₂ #combine(#1(big.body red.body)
            #1(red.body house.body))
       β₃ #combine(#uw8(big.body red.body)
             #uw8(red.body house.body))
     )
   )
```

RMIT-1 values were $(\alpha_1, \alpha_2, \alpha_3) = (0.20, 0.05, 0.75)$, and $(\beta_1, \beta_2, \beta_3) = (0.8, 0.1, 0.1)$. Tuned on CW09B 200 topics.

English Subtask Results (CW12B)

System	ERR@k		NDCG@k		RBP@p
	@5	@10	@5	@10	@0.9
RMIT-3	0.5065	0.5207	0.3977	0.3968	0.7670+0.0242
RMIT-2	0.5285	0.5378	0.4186	0.4069	0.7533+0.0228
RMIT-4	**0.5635**	**0.5728**	0.4402	0.4249	0.7422+0.0270
RMIT-1	0.5548	0.5712	**0.4670**	**0.4783**‡	**0.8438+0.0221**‡
Post-hoc analysis of submissions					
BM25	0.4760	0.4879	0.3718	0.3713	0.6509+0.1919‡
R3-NQE	0.4955	0.5096	0.3884	0.3879	0.7560+0.0348
R2-NQE	0.5279	0.5403	0.4161	0.4125	0.7537+0.0408
R4-NQE	0.5533	0.5637	0.4276	0.4071	0.7238+0.0456
RBC-14	**0.5819**	**0.5951**	**0.4817**‡	0.4776‡	**0.8263+0.0025**‡
R1-NQE	0.5743	0.5884	0.4723†	**0.4877**‡	0.8220+0.0453‡

Holm corrected pairwise statistical tests, with $\dagger$ and $\ddagger$ indicating significance at $p = 0.05$ and $p = 0.01$ respectively relative to **RMIT-3**.

Topic	RMIT-1	BM25	Δ	Query
83	0.4129	0.7189	-0.3060	jetstar airlines hong kong
88	0.3179	0.5943	-0.2764	mexico climate
57	0.3209	0.5893	-0.2684	axle ratio
54	0.4505	0.7025	-0.2519	anime pillow
41	0.2281	0.4676	-0.2395	autumn
71	0.5948	0.0000	+0.5948	dog food for allergies
46	0.6958	0.1423	+0.5535	musical note
45	0.6399	0.0812	+0.5586	commendatory term
30	0.9458	0.3898	+0.5561	robot
28	0.5113	0.0000	+0.5113	typing practice

Post-hoc Query Expansion Analysis

- Query Expansion was used in all submitted systems

- What happens if we turn it off?

- Use the same system configuration without Query Expansion

				NDCG@10				
System A	System B	Win	Tie	Loss	$\frac{Win}{Loss}$	$\sum Win$	$\sum Loss$	$\frac{\sum Win}{\sum Loss}$
R1-NQE	RMIT-1	32	39	29	1.103	16.033	15.391	1.042
R2-NQE	RMIT-2	29	45	26	1.115	12.957	11.518	1.125
R3-NQE	RMIT-3[†]	11	70	19	0.579	4.943	6.909	0.715
R4-NQE	RMIT-4[‡]	12	58	30	0.400	4.883	13.655	0.358

Conclusions and Future Work

- We Want Web task helps to drive research in other sub-fields and vice-versa

- This round we focused on classic retrieval techniques that are known to be effective

- Aim to participate in Chinese subtask in future rounds

- Use more sophisticated techniques in future (LTR, Duet Matching)

Thank You!