RMIT_IR at the NTCIR-17 FairWeb-1 Task

Sachin Pathiyan Cherumanal, Kaixin Ji, Danula Hettiachchi, Johanne R. Trippas, Falk Scholer, Damiano Spina

What's next…

RMIT UNIVERSITY

Motivation

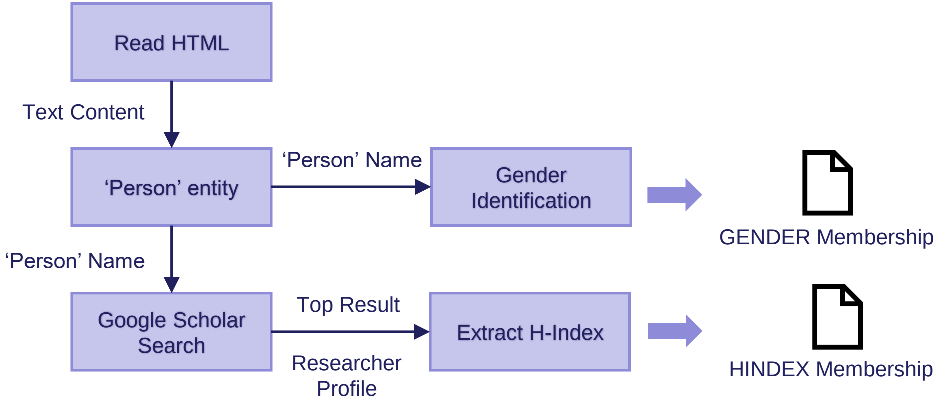
Fairness and Diversity have been studied side-by-side over the recent years, especially for multi-attribute fairness[1,2,3].

"Exploring whether search results diversification (SRD) techniques and ranking fusion can help achieve fairer results along nominal and ordinal fairness attributes."

[1] Pathiyan Cherumanal et al. 2021. Evaluating Fairness in Argument Retrieval (CIKM'21).
[2] Pathiyan Cherumanal et al. 2022. RMIT at TREC 2021 Fair Ranking Track.
[3] Pathiyan Cherumanal et al. 2023. RMIT CIDDA IR at the TREC 2022 Fair Ranking Track

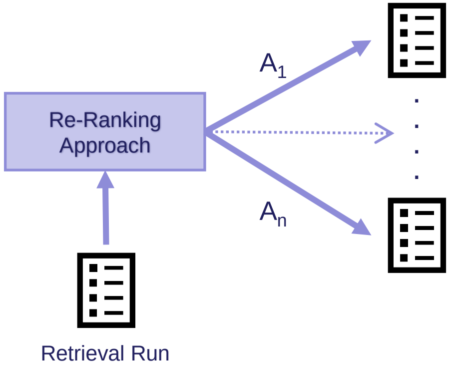
Step 1: Membership Generation

- Parsed documents using BeautifulSoup
- Entity Recognition using SpaCy
- Custom framework made available[1]



[1] https://github.com/rmit-ir/fairweb-1

Step 2: Re-Ranking



- Retrieval (r) – BM25(Q), BM25(D)
- Re-ranking approaches applied to each fairness attribute (A) A_1 … A_n (e.g., GENDER and HINDEX for Researcher-related (R) Topic-Type)

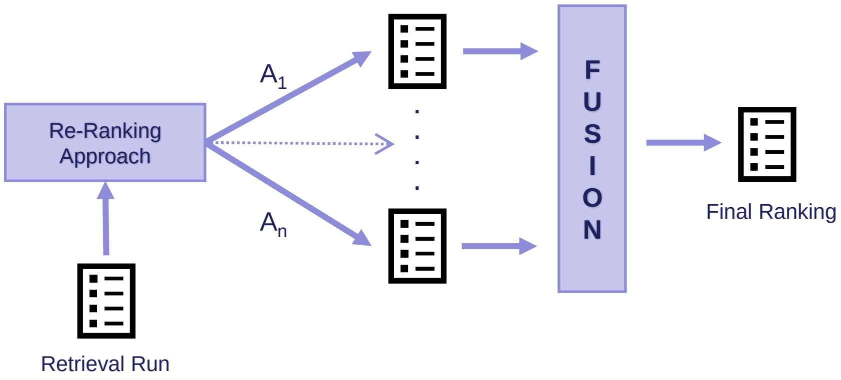
Re-Ranking Approaches:

- PM-2
- Linear Combination

$$LC = ((1 - \lambda) * R) + (\lambda * F)$$

Step 2: Re-Ranking

- Diversified rankings from multiple attributes were fused using RRF[1].



Run Name	Description
rmit_ir-D-RR-1	Linear combination of top 50 relevance and fairness with $\lambda = 0.9$.
rmit_ir-D-RR-2	PM2 with $\lambda = 0.9$
rmit_ir-D-RR-3	PM2 on top 50 with $\lambda = 0.9$
rmit_ir-D-RR-4	Linear combination of relevance and fairness with $\lambda = 0.9$
rmit_ir-Q-RR-5	Linear combination of top 50 relevance and fairness with $\lambda = 0.5$

$$P_R = \mathrm{RRF}\left(\text{PM-2}(r, \text{GENDER}), \text{PM-2}(r, \text{HINDEX})\right)$$

$$P_M = \mathrm{RRF}\left(\text{PM-2}(r, \text{ORIGIN}), \text{PM-2}(r, \text{RATINGS})\right)$$

$$P_Y = \text{PM-2}(r, \text{SUBSCS})$$

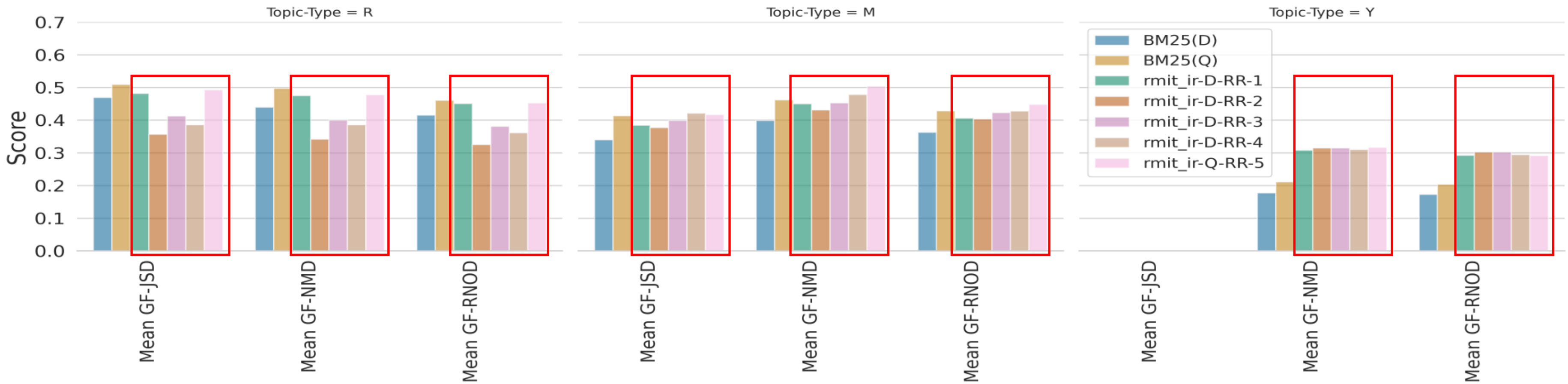
$$L_R = \mathrm{RRF}\left(\mathrm{LC}(r, \text{Gender}), \mathrm{LC}(r, \text{HINDEX})\right)$$

$$L_M = \mathrm{RRF}\left(\mathrm{LC}(r, \text{ORIGIN}), \mathrm{LC}(r, \text{RATINGS})\right)$$

$$L_Y = \mathrm{LC}(r, \text{SUBSCS})$$

[1] Pathiyan Cherumanal et al. 2023. RMIT CIDDA IR at the TREC 2022 Fair Ranking Track

Results



- R-Topic: All our submitted runs performed poorly compared to the retrieval baseline.
- M-Topic: LC-based runs outperform PM-2 and baselines.
- Y-Topic: All outperform retrieval baseline, and *rmit_ir-Q-RR-5* best run out of our submitted runs.

However, not all runs showed statistically significant improvement over the retrieval baselines.

Thank you

See you at the Poster session …

Contact:

Sachin Pathiyan Cherumanal

Email: sachin.pathiyan.cherumanal@student.rmit.edu.au

SCAN ME