



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...

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] Pathiyar Cherumanal et al. 2021. Evaluating Fairness in Argument Retrieval (CIKM'21).

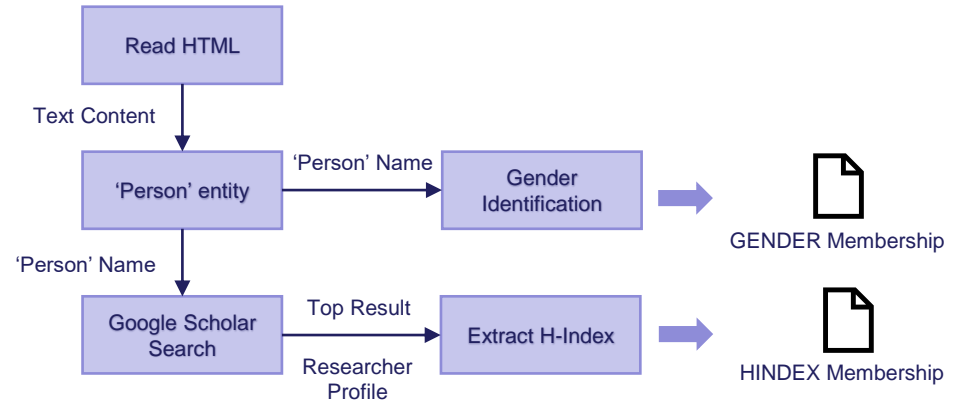
[2] Pathiyar Cherumanal et al. 2022. RMIT at TREC 2021 Fair Ranking Track.

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



Step 1: Membership Generation

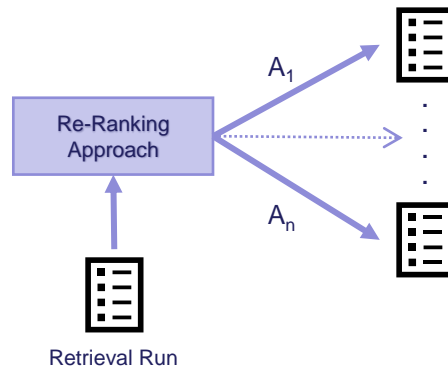
- Parsed documents using BeautifulSoup
- Entity Recognition using SpaCy
- Custom framework made available¹



[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) $A_1 \dots A_n$ (e.g., GENDER and HINDEX for Researcher-related (R) Topic-Type)



Re-Ranking Approaches:

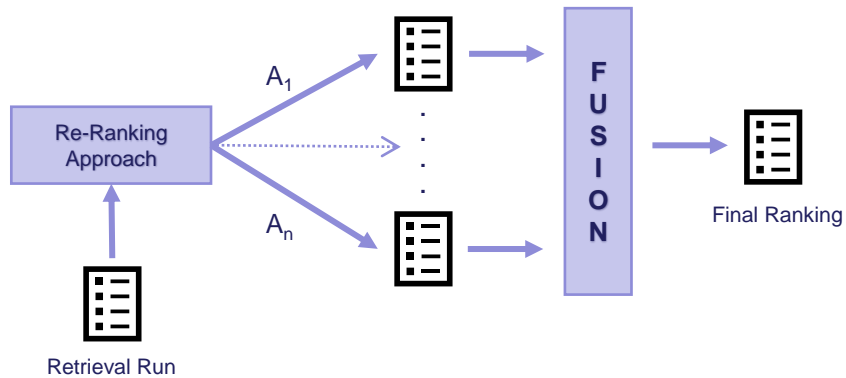
- PM-2
- Linear Combination

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



Step 2: Re-Ranking

- Diversified rankings from multiple attributes were fused using RRF¹.



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 $\lambda = 0.5$

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

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

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

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

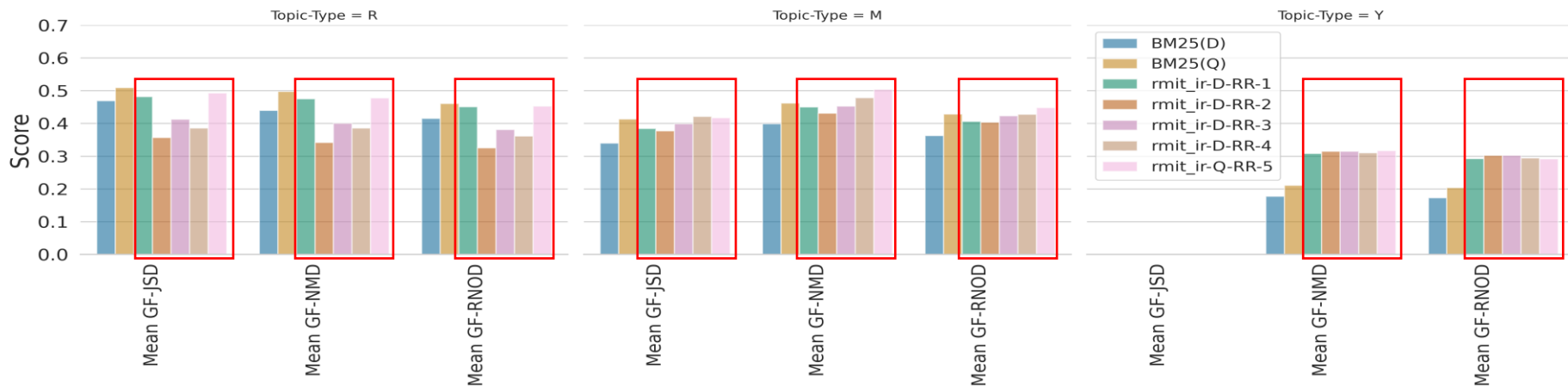
$$L_M = \text{RRF}(\text{LC}(r, \text{ORIGIN}), \text{LC}(r, \text{RATINGS}))$$

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

[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 ...

RMIT UNIVERSITY RMIT_IR at the NTCIR-17 FairWeb-1 Task

Sachin Pathiyan Cherumanal, Karan B. Dasika Hettichchi, Johanne B. Trappas, Fark Scholtes, Damiano Spina
(sachin.pathiyan.cherumanal, karan_b@student.rmit.edu.au, dasika@hettichchi, johanne_b.trappas, fark.scholtes, damiano.spina@rmit.edu.au)

Background

In information retrieval, **relevance** and **diversity** have been studied side-by-side over the recent years. Following some of our previous work¹, the aim was as follows:

"Studying whether search results diversification (SD) techniques and ranking fusion can help achieve better results along **number** and **ordinal** fairness attributes."

Methodology

Step 1: Re-ranking approaches applied to each fairness attribute $A_1, A_2, \dots, \text{CANDIR}$ and HINDEX for Relevance-related (R) Top-3000.
(\otimes an explicit SD technique called **PM2**, and
(\oplus) a **Linear Combination (LC)** run inspired by an implicit SD technique called **Hybrid Angular Relevance**.

Step 2: For each re-ranking approach, the diversified rankings from multiple attributes were fused using SD².

Membership Generation

A membership generation framework³ used for creating the membership files for each attribute of the top-ages: Researcher-related (SR), Movie-related (MR), and YouTube-related (Y).

Submitted Runs

Run Name	Description
rmit_r-Q6B-1	LC of top 50 relevance and fairness with $\lambda = 0.8$
rmit_r-Q6B-2	PM2 with $\lambda = 0.9$
rmit_r-Q6B-3	PM2 on top 50 with $\lambda = 0.9$
rmit_r-Q6B-4	LC of relevance and fairness with $\lambda = 0.9$
rmit_r-Q6B-5	LC of top 50 relevance and fairness with $\lambda = 0.8$

For all the runs, statistical significance was calculated using randomized Tukey HSD test with $B = 5, 000$ trials, $\alpha = 0.05$.

Results

Relevance: Only our LC-based runs showed improvements over the ordinal baseline.

R-Traps: The best-performing LC-based run (i.e., rmit_r-Q6B-5) outperforms the ordinal baseline only along nominal fairness measures.

M-Traps: Both PM2 and LC-based runs outperform their respective ordinal baselines across all fairness metrics.

V-Traps: rmit_r-Q6B-5 performed the best along CF-RIND; however, PM2 performed the best along CF-IR-ICD.

Overall, there was NO statistical significance observed in our results (DO NOT overinterpret) indicating that search results diversification (SD) techniques and ranking fusion can help achieve better results along nominal and ordinal fairness attributes.

Contact:

Sachin Pathiyan Cherumanal

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



SCAN ME

