

Sachin Pathiyan Cherumanal, Kaixin Ji, Danula Hettiachchi, Johanne R. Trippas, Falk Scholer, Damiano Spina
 {sachin.pathiyan.cherumanal, kaixin.ji}@student.rmit.edu.au, {danula.hettiachchi, j.trippas, falk.scholer, damiano.spina}@rmit.edu.au

Background

In information retrieval, **fairness** and **diversity** have been studied side-by-side over the recent years.

Following some of our previous work^{1,2}, the aim was as follows:

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

[1] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. 2022. RMIT at TREC 2021 Fair Ranking Track. Proceedings of TREC'21.

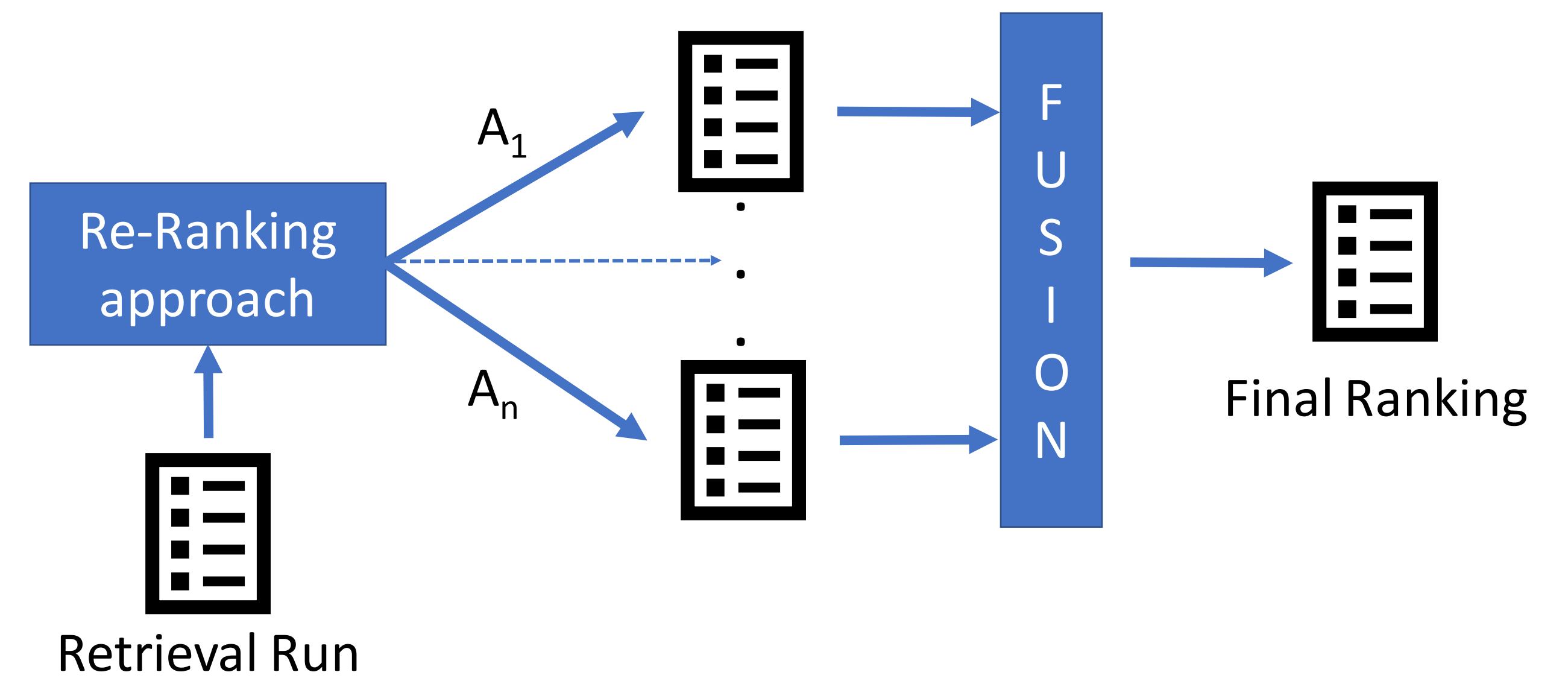
[2] Sachin Pathiyan Cherumanal, Marwah Alaofi, Reham Abdullah Altalhi, Elham Naghizade, Falk Scholer, and Damiano Spina. 2023. RMIT CIDDA IR at the TREC 2022 Fair Ranking Track. Proceedings of TREC'22.

Methodology

Step 1: Re-ranking approaches applied to each fairness attribute A_1, \dots, A_n (e.g., GENDER and HINDEX for Researcher-related (R) Topic-Type):

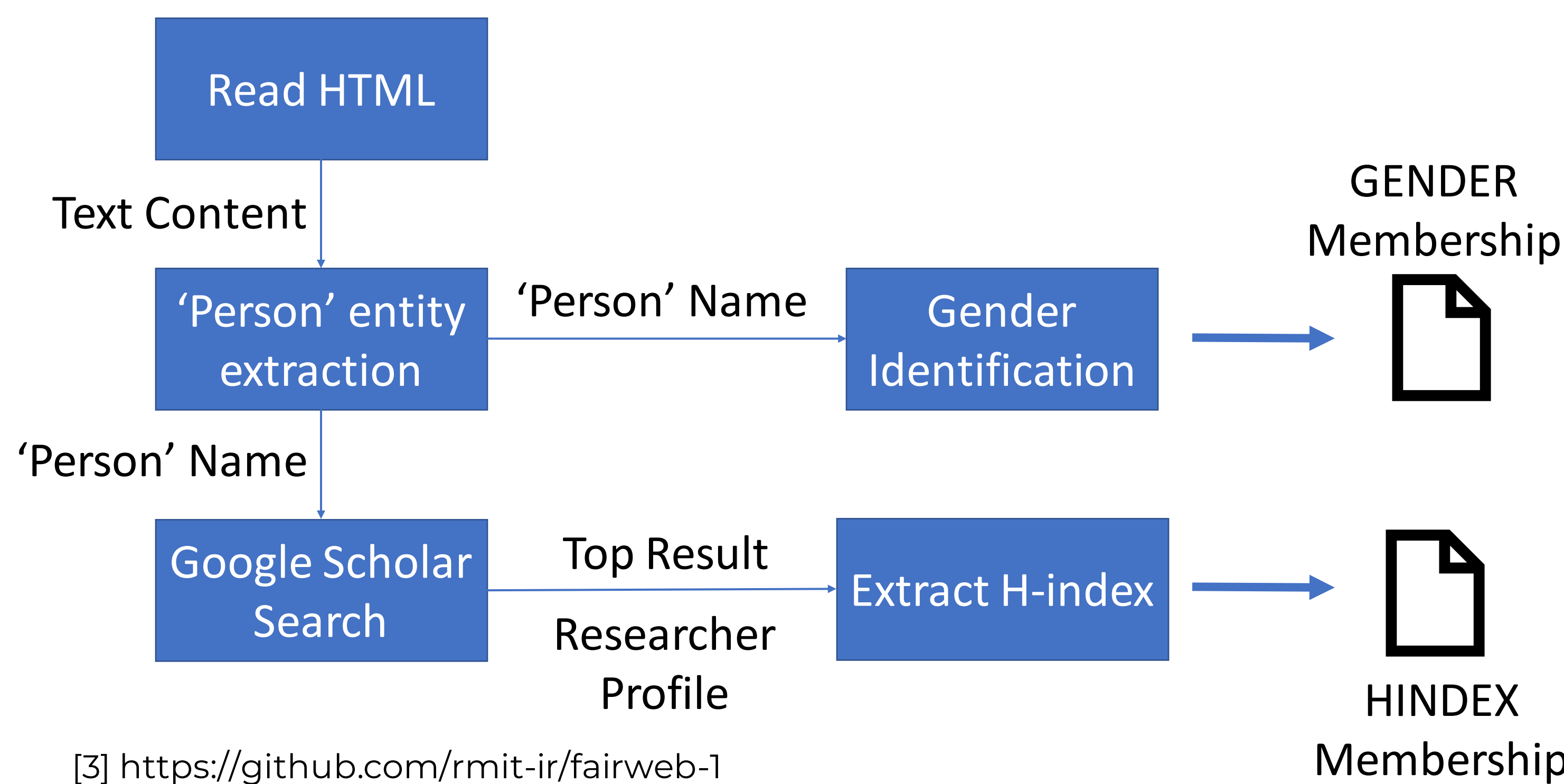
- (i) an explicit SRD technique called **PM-2**, and
- (ii) a **Linear Combination (LC)** run inspired by an implicit SRD technique called Maximal Marginal Relevance.

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



Membership Generation

A membership generation framework³ used for creating the membership files for each attribute of the topic-types: Researcher-related (R), Movie-related (M), and YouTube-related (Y).



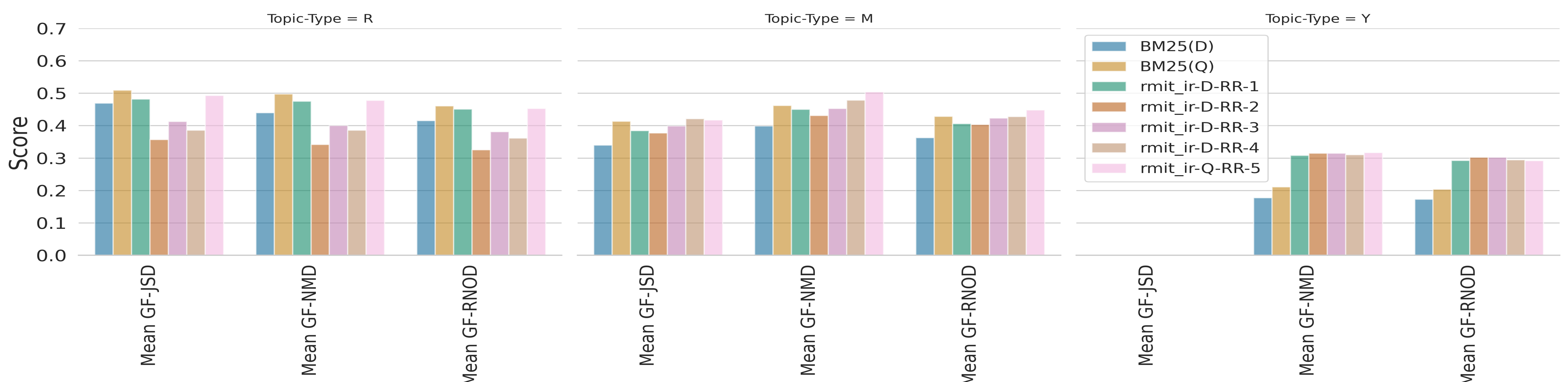
[3] <https://github.com/rmit-ir/fairweb-1>

Submitted Runs

| Run Name | Description |
|----------------|---|
| rmit_ir-D-RR-1 | LC 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 | LC of relevance and fairness with $\lambda = 0.9$ |
| rmit_ir-Q-RR-5 | LC of top 50 relevance and fairness with $\lambda = 0.5$ |

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 retrieval baseline.

R-Topic. The best-performing LC-based run (i.e., rmit_ir-Q-RR-5) outperforms the retrieval baseline only along nominal fairness measures.

M-Topic. Both PM-2 and LC-based runs outperform their respective retrieval baseline across all fairness metrics.

Y-Topic. rmit_ir-Q-RR-5 performed the best along GF-NMD however, PM-2 performed the best along GF-RNOD.

Overall, there was NO statistical significance observed, so our results DO NOT conclusively indicate that search results diversification (SRD) techniques and ranking fusion can help achieve fairer results along nominal and ordinal fairness attributes.