Overview of the NTCIR-18 FairWeb-2 Task

Sijie Tao, Tetsuya Sakai
Waseda University, Japan

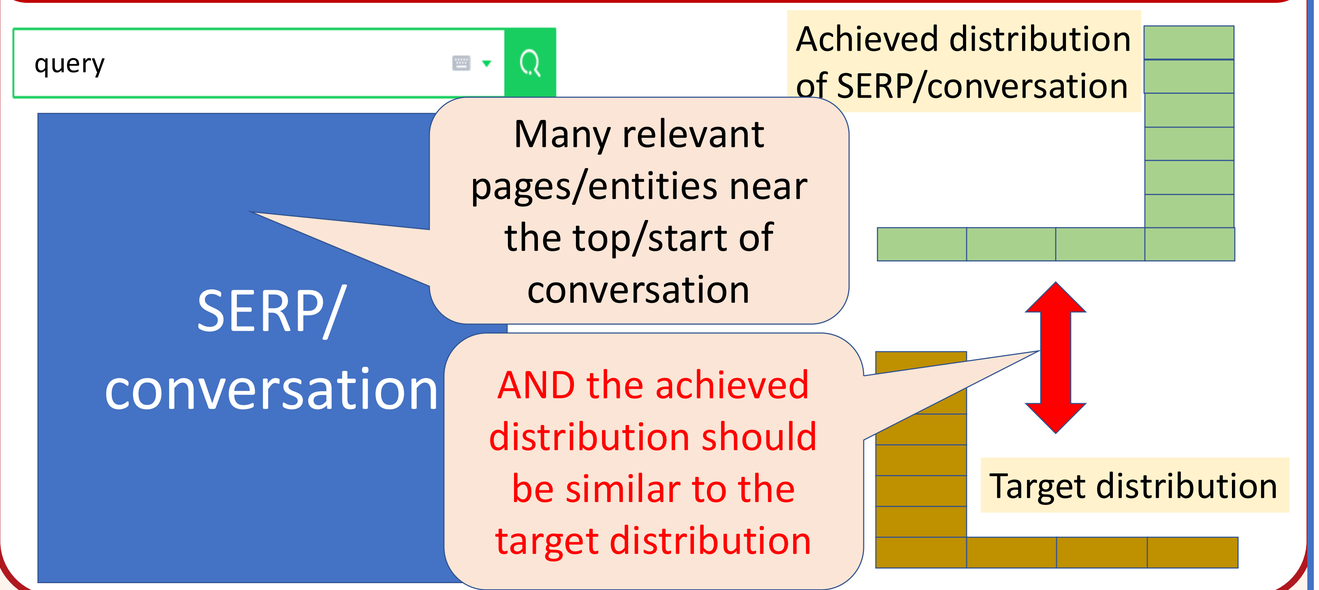
Nuo Chen
The Hong Kong Polytechnic University, P.R.C.

Haitao Li, Yiteng Tu
Tsinghua University, P.R.C.

Junjie Wang, Hanpei Fang, Yuxiang Zhang
Waseda University, Japan

Maria Maistro
University of Copenhagen, Denmark

Web Search & Conversational Search that Consider Group Fairness



Task Specifications

INPUT:

- Search topics that describe information needs about entities
- Attribute sets and a target distribution for each of them

OUTPUT:

- **WS subtask:** runs contain relevant documents & group-fair wrt each attribute set
- **CS subtask:** plain-text user-system conversations contain relevant entities & group-fair wrt each attribute set

Topic/entity types and attribute sets: R, M, Y

Topic/entity type	Attribute sets
Researchers	HINDEX (ordinal, 4 groups) PRONOUN (nominal, 3 groups)
Movies	RATINGS (ordinal, 4 groups) ORIGIN (nominal, 8 groups)
Youtube contents	SUBSCS (ordinal, 4 groups)

Ex.“Rogue One” is a relevant **M** entity to an **M** topic “Star Wars Movies”

Test Collection Construction

Target Corpus (WS subtask): ChuWeb21D-60

- 49.8 million web pages and available online!

Topics:

- Organisers created 54 topics, 18 topics for each topic type

Submitted runs:

WS subtask

Team name	Institution	#runs
IITUH18	Indian Institute of Technology(BHU)	1
RSLFW	Waseda University	5
THUIR	Tsinghua University	6*
AMS42	University of Amsterdam	5
ORG	ORGANISERS	5
Total #runs		23

CS subtask

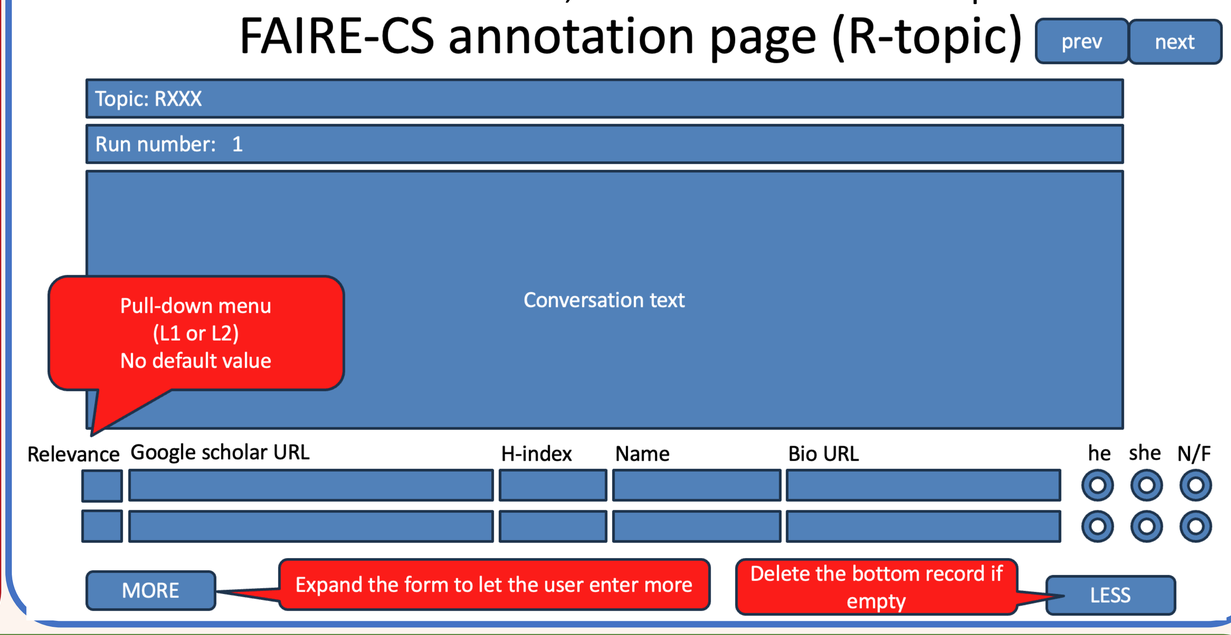
Team name	Institution	#runs
COPWA	Waseda University	3
ORG	ORGANISERS	1
Total #runs		4

*As the winner of Fairweb-1, THUIR submitted five runs and one additional REV run.

Entity annotation:

- Organisers served as gold assessors
- WS subtask: pooling depth = 25, max 3 relevant entities per document
- CS subtask: all runs annotated, max 10 relevant entities per conversation

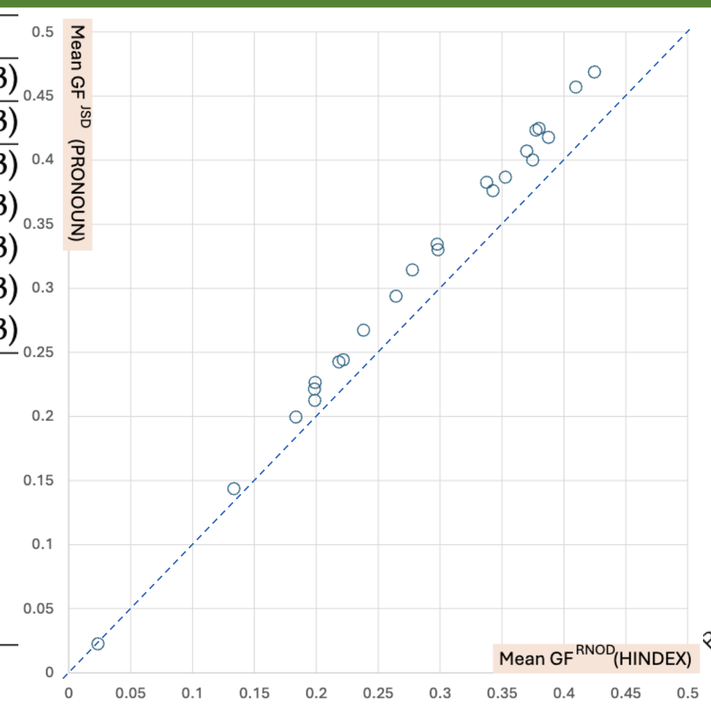
FAIRE-CS annotation page (R-topic)



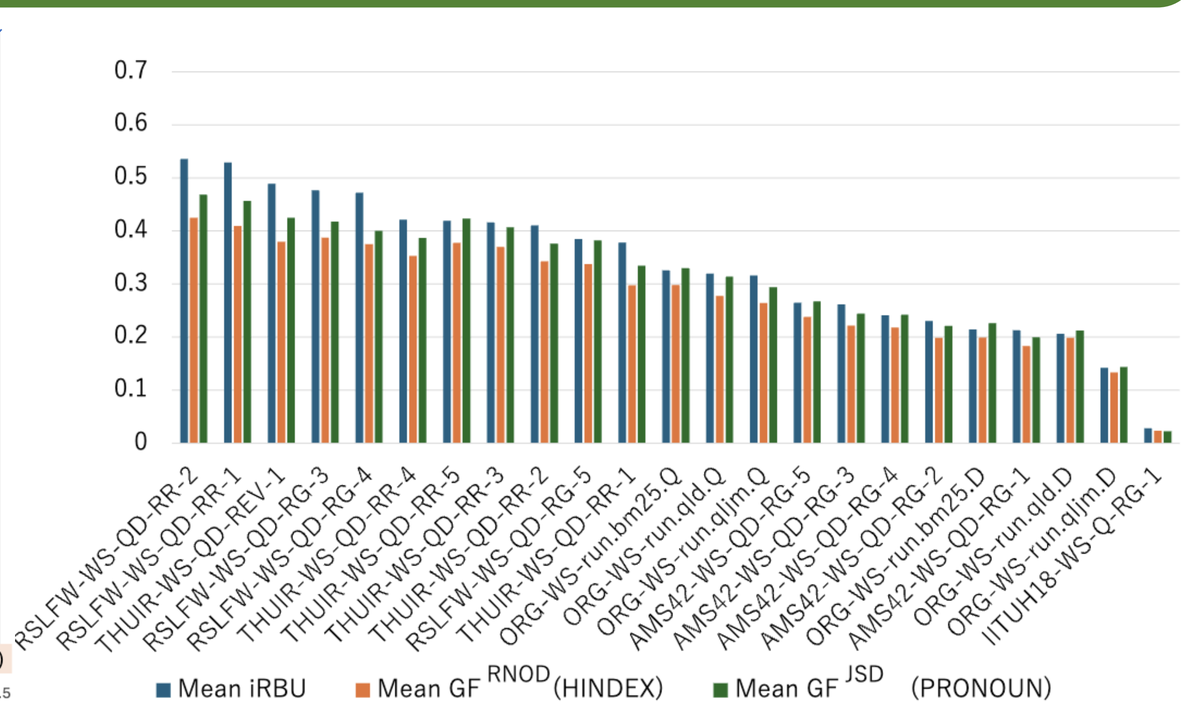
Official Results (R topics)

Rank	Run	Mean GFR
1	RSLFW-WS-QD-RR-2	0.4764 (>18-23)
2	RSLFW-WS-QD-RR-1	0.4651 (>20-23)
3	THUIR-WS-QD-REV-1	0.4314 (>22-23)
4	RSLFW-WS-QD-RG-3	0.4274 (>22-23)
5	RSLFW-WS-QD-RG-4	0.4157 (>22-23)
6	THUIR-WS-QD-RR-5	0.4067 (>22-23)
7	THUIR-WS-QD-RR-3	0.3978 (>22-23)
8	THUIR-WS-QD-RR-4	0.3868 (>23)
9	THUIR-WS-QD-RR-2	0.3765 (>23)
10	RSLFW-WS-QD-RG-5	0.3685 (>23)
11	THUIR-WS-QD-RR-1	0.3367 (>23)
12	ORG-WS-run.bm25.Q	0.3181 (>23)
13	ORG-WS-run.qld.Q	0.3039 (>23)
14	ORG-WS-run.qljm.Q	0.2914 (>23)
15	AMS42-WS-QD-RG-5	0.2568

Mean GFR score over the R topics (top 15 runs)



Intersectional group fairness: mean scores over the R topics



Mean scores over the R topics

Evaluation Measures

WS subtask: Group Fairness and Relevance (GFR)

$$GFR(L) = \sum_{k=1}^{|L|} Decay_{L@k} \left(w_0\, Utility_{L@k} + \sum_{m=1}^{M} w_m\, DistrSim^m_{L@k} \right)$$

How useful was the top k of the SERP: ERR/iRBU

Similarity between achieved distribution@k and target distribution

Probability that users will reach k and finally get satisfied

CS subtask: Group Fairness and Relevance of Conversation(GFRC)

$$\mathrm{GFR}(C) = \alpha R(C) + (1-\alpha) GF(C)$$

$$\mathrm{R}(C) = \frac{1}{\mathcal{N}} \sum_{i=1}^{T} \sum_{n_{ij} \in S_i} pw(n_{ij}) g(n_{ij})$$

Position-based weight of nugget n

Relevance score of nugget n (ERR at our task)

$$\mathrm{GF}(C) = \frac{1}{\mathcal{N}'} \sum_{i=1}^{T} PW(S_i) \sum_{m=1}^{M} w_m DistrSim^m(D^m(S_i) || D^m_*)$$

Position-based weight of turn S

Rank	Run	Mean GFRC
1	COPWA-CS-QD-MN-2	0.447591 (>2-4)
2	COPWA-CS-QD-MN-3	0.057489
3	ORG-CS-D-MN-1	0.032935
4	COPWA-CS-D-MN-1	0.000000

Run	Relevance
COPWA-CS-QD-MN-2	0.005664
ORG-CS-D-MN-1	0.000635
COPWA-CS-QD-MN-3	0.000515
COPWA-CS-D-MN-1	0.000000

Rank	Run	Mean GF^{JSD} (PRONOUN)
1	COPWA-CS-QD-MN-2	0.738476 (>2-4)
2	COPWA-CS-QD-MN-3	0.098337
3	ORG-CS-D-MN-1	0.061363
4	COPWA-CS-D-MN-1	0.000000

Rank	Run	Mean GF^{NMD} (HINDEX)
1	COPWA-CS-QD-MN-2	0.666162 (>2-4)
2	COPWA-CS-QD-MN-3	0.090909
3	ORG-CS-D-MN-1	0.045455
4	COPWA-CS-D-MN-1	0.000000

Mean GFRC, Relevance, and GF scores over the R topics

Conclusions and Future Work

WS subtask: **RSLFW** runs are the winners!
CS subtask: **COPWA-CS-QD-MN-2** is the winner!

What’s next?

- A new RAG task accepted at NTCIR-19! Stay tuned!
- R2C2 (RAG Responses: Confident and Correct?)