

Why you should participate in the NTCIR-18 FairWeb-2 Task

Version 20240326

FairWeb-2 organisers:

Sijie Tao, Tetsuya Sakai, Junjie Wang, Hanpei Fang, Yuxiang Zhang
(Waseda University),

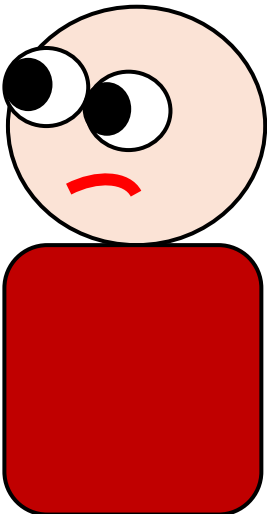
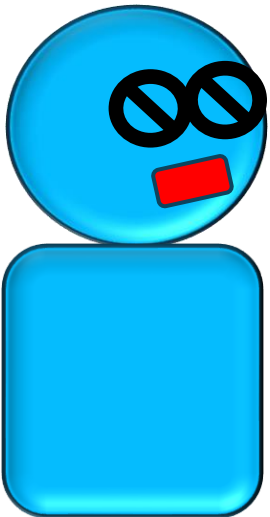
Haitao Li, Yiteng Tu (Tsinghua University, China),

Nuo Chen (The Hong Kong Polytechnic University, China),

Maria Maistro (University of Copenhagen, Denmark)

fairweb2org@list.waseda.jp

sakailab.com/fairweb2/



OUTLINE

1. Motivation
2. Web Search Subtask
3. Conversational Search Subtask
4. Summary

Web search that's "fair"

*"You are serving as a general chair of an IR conference. You want to hire **diverse** IR researchers as organisers."*

Diversity dimensions:

- Different career stage (include junior researchers, not just famous researchers)
 - Different genders
 - Different countries
- etc.

information retrieval researchers

Sign in

200

ALL IMAGES VIDEOS MAPS NEWS SHOPPING MORE

2,460,000,000 Results Any time

An information retrieval system not only occupies an important position in the network information platform, but also plays an important role in information acquisition, query processing, and wireless sensor networks. It is a procedure to help researchers **extract documents from data sets as document retrieval tools.**

Author: Binbin Yu, Binbin Yu

Publish Year: 2019

Email: yubinbin80@sina.com

[Research on information retrieval model based on ontology](#)

jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-019-1354-z

Was this helpful?

People also ask

What is information retrieval?

What is information retrieval in engineering engineering?

What is the purpose of the information retrieval community study?

When was the first information retrieval system invented?

We only get famous and **high h-index** people...

Information retrieval



Information retrieval (IR) in computing and information science is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. ...

Related people



Susan Dumais



C. J. van Rijsbergen



Ricardo Baeza-Yates



Gerard Salton



Calvin Mooers

h-index > 100

Explore more

h-index > 80

ALL IMAGES VIDEOS MAPS NEWS SHOPPING ⋮ MORE

2,460,000,000 Results Any time ▾

An information retrieval system not only occupies an important position in the network information platform, but also plays an important role in information acquisition, query processing, and wireless sensor networks. It is a procedure to help researchers **extract documents from data sets as document retrieval tools.**

Author: Binbin Yu, Binbin Yu

Publish Year: 2019

Email: yubinbin80@sina.com

[Research on information retrieval model based on ontology](#)

jwcn-urasipjournals.springeropen.com/articles/10.1186/s13638-019-1354-z

Was this helpful? 👍 👎

People also ask

What is information retrieval?

What is information retrieval in e

What is the purpose of the inform

When was the first information re

Poor gender balance...

Information retrieval



Information retrieval (IR) in computing and information science is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. ...

Related people



Susan Dumais



C. J. van Rijsbergen



Ricardo Baeza-Yates



Gerard Salton



Calvin Mooers

Data: Wikipedia
Wikipedia text under CC-BY-SA license

Feedback 👍 👎

Explore more

Let's consider group fairness

Approximates true gender: only reflects whether "he" or "she" is found in the researcher bio

Attribute set: **HINDEX**

Attribute set: **GENDER**

Target distribution: give more exposure to junior researchers!

Target distribution: equal opportunities for different genders!



Group 1

Group 2

Group 3

Group 4

$x < 10$

$10 \leq x < 30$

$30 \leq x < 50$

$50 \leq x$

Group 1

Group 2

Group 3

he

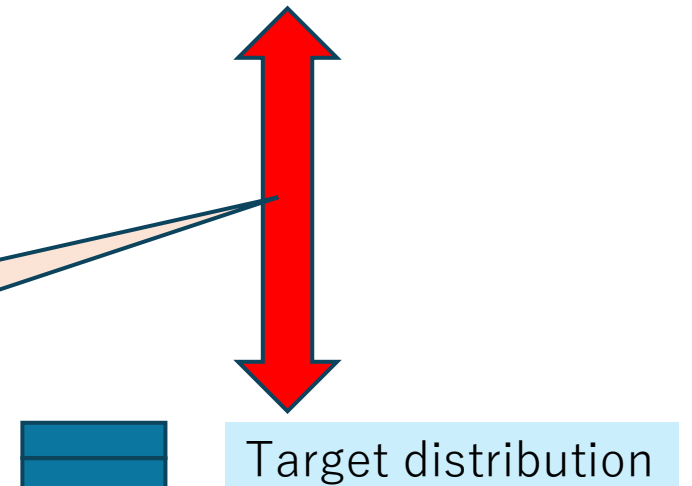
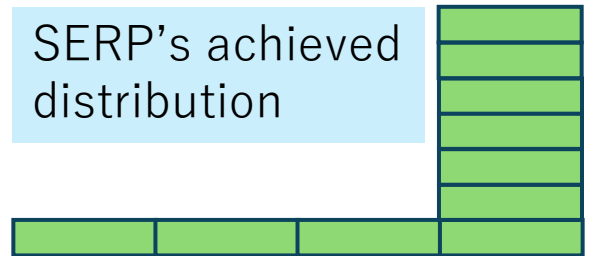
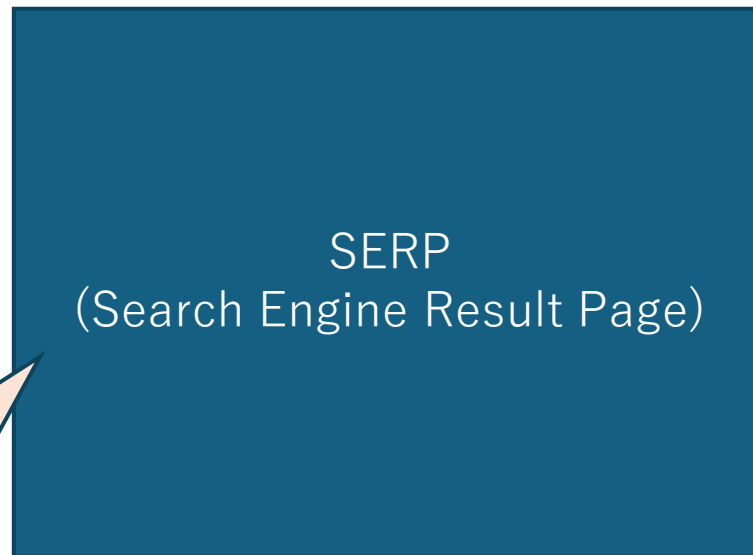
she

pronoun
not found

Web search that considers group fairness

NAVER

query

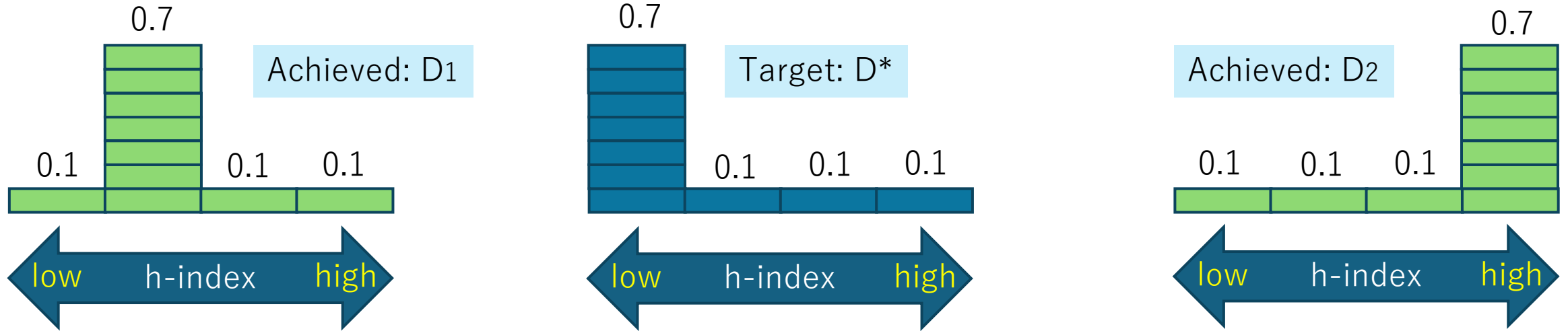


Many relevant pages near the top (traditional adhoc IR)

AND the achieved distribution should be similar to the target one

Handling ordinal groups properly

See [Sakai21ACL](#), [Sakai21CIKM1q](#)



If divergences for **nominal groups** (e.g. Jensen-Shannon Divergence) are used...

$$\text{JSD} = 0.3651$$

$$\text{JSD} = 0.3651$$

Divergences for **ordinal groups** can tell the difference

$$\text{NMD} = 0.2000$$

Closer to target

$$\text{RNOD} = 0.5477$$

Closer to target

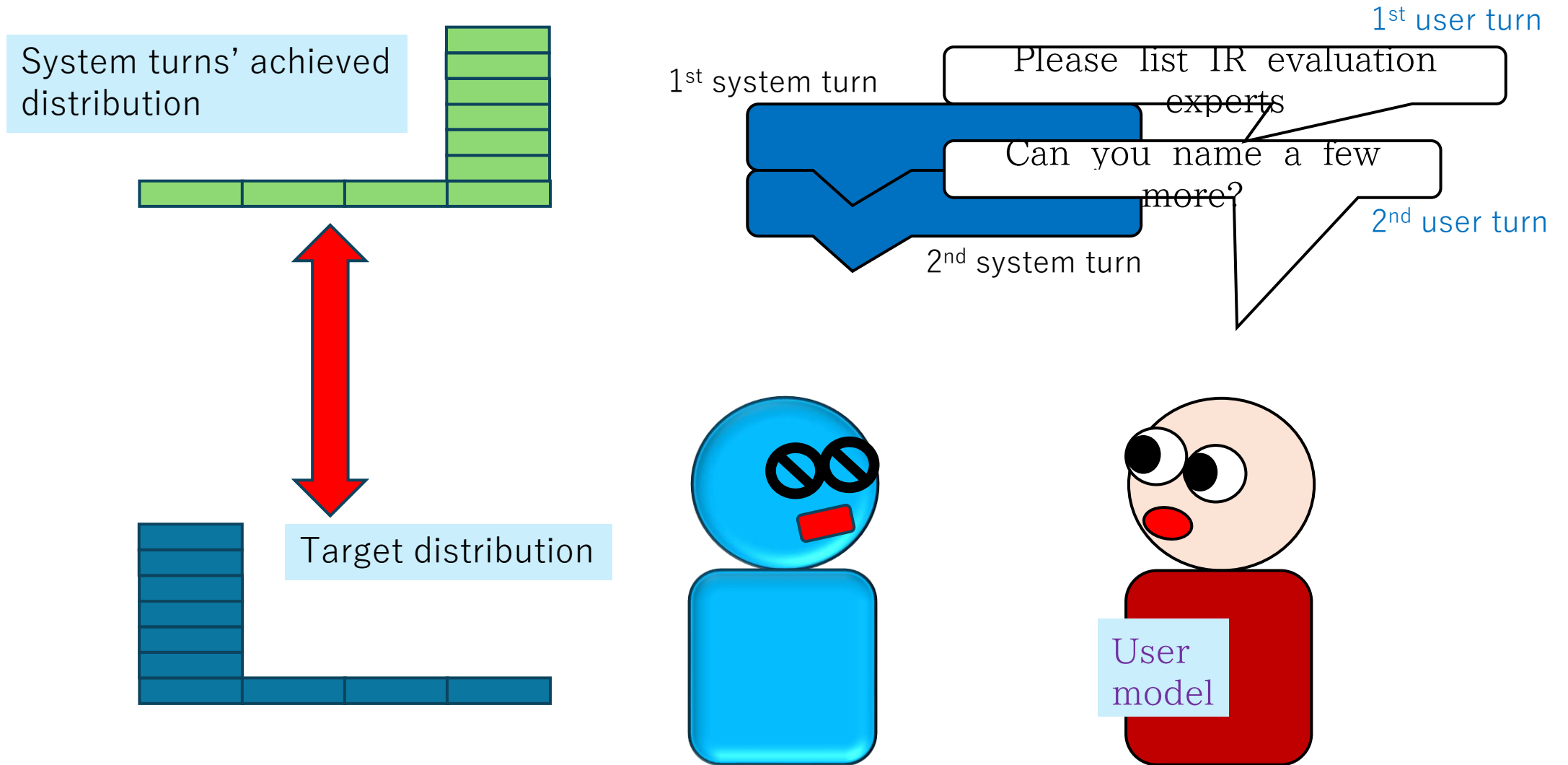
$$\text{NMD} = 0.6000$$

$$\text{RNOD} = 0.6000$$

Subtasks

- Web Search (same as FairWeb-1)
 - Input: Researcher(R)/Movie(M)/YouTube(Y) topic
 - Output: SERP (Search Engine Result Page)
where the target corpus is Chuweb21D
- Conversational Search (NEW)
 - Input: R/M/Y topic (same topic set as Web Search)
 - Output: User and system turns (English text, up to 1,250 words)

Conversational search that's "fair"



Topic types and attribute sets (same as FairWeb-1 for continuity)

M (movie): RATINGS (ordinal, 4 groups)

ORIGIN (nominal, 8 groups)

R (researcher): HINDEX (ordinal, 4 groups)

GENDER (nominal, 3 groups)

Y (YouTube): SUBSCS (ordinal, 4 groups)

#ratings on IMDb
 $x < 100$
 $100 \leq x < 10K$
 $10K \leq x < 1M$
 $1M \leq x$

Countries of origin on IMDb mapped to 8 geographic regions (one movie may cover multiple countries)

e.g. a UK-Japan movie => Asia, Europe

Example topics from FairWeb-1

- R-topic

Researcher entities

"**Researchers** who have coauthored at least one paper with Stephen E. Robertson, the inventor of BM25."

- M-topic

Movie entities

"sci-fi **movies** that feature time travel, listed in IMDb."

- Y-topic

Video entities

"**Covers of songs** written by Burt Bacharach and Hal David, available on YouTube. Only Bacharach/David collaborations (just these two writers) are relevant."

OUTLINE

1. Motivation

2. Web Search Subtask

3. Conversational Search Subtask

4. Summary

Subtask workflow

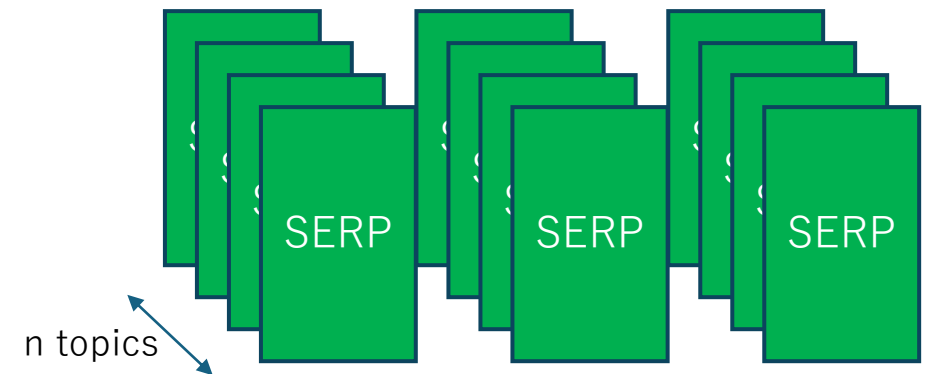
Annotators

Organisers

Participants

Release the topic set

Submit runs (SERP for each topic)



Subtask workflow

Annotators

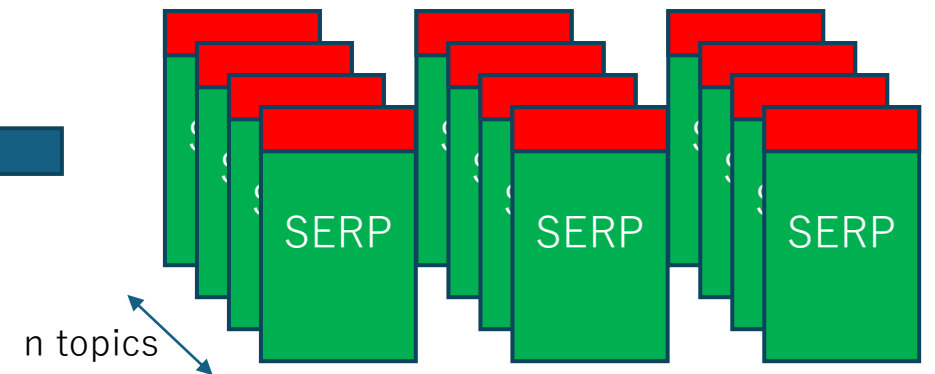
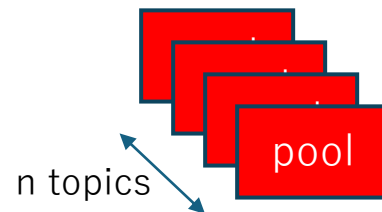
Organisers

Participants

Release the topic set

Submit runs (SERP for each topic)

Form a depth-k pool for each topic



Subtask workflow

Annotators

Organisers

Participants

Release the topic set

Submit runs (SERP for each topic)

Form a depth-k pool for each topic

Annotate up to 3 relevant entities from each document

Web page

Relevant entity

Web page

Relevant entity

Subtask workflow

Annotators

Organisers

Participants

Release the topic set

Submit runs (SERP for each topic)

Form a depth-k pool for each topic

Annotate up to 3 relevant entities from each document

Derive relevance + group membership of each page, compute evaluation measures for each run

Relevant entity

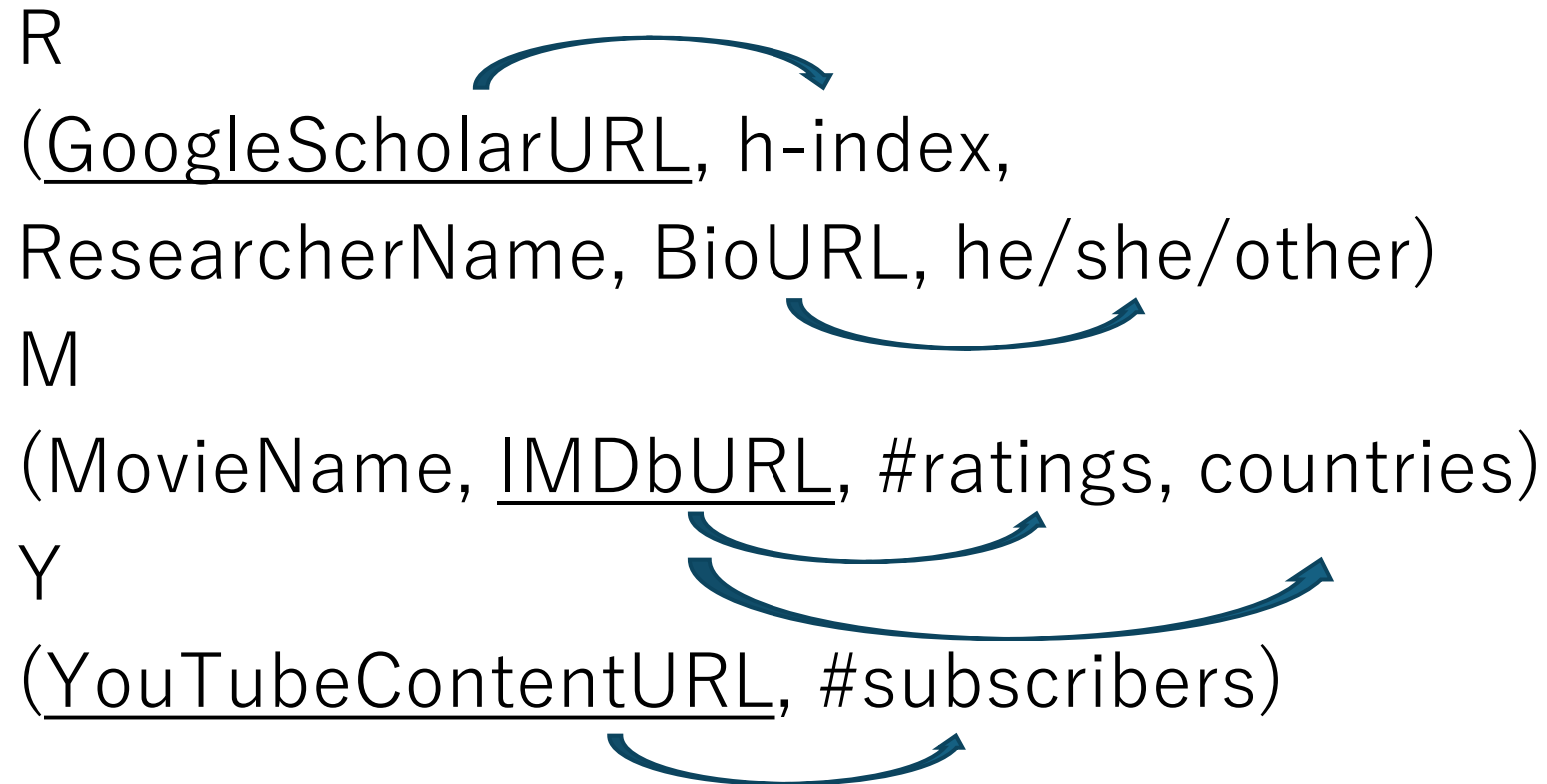
Web page

Relevant entity

Page relevance

Group membership

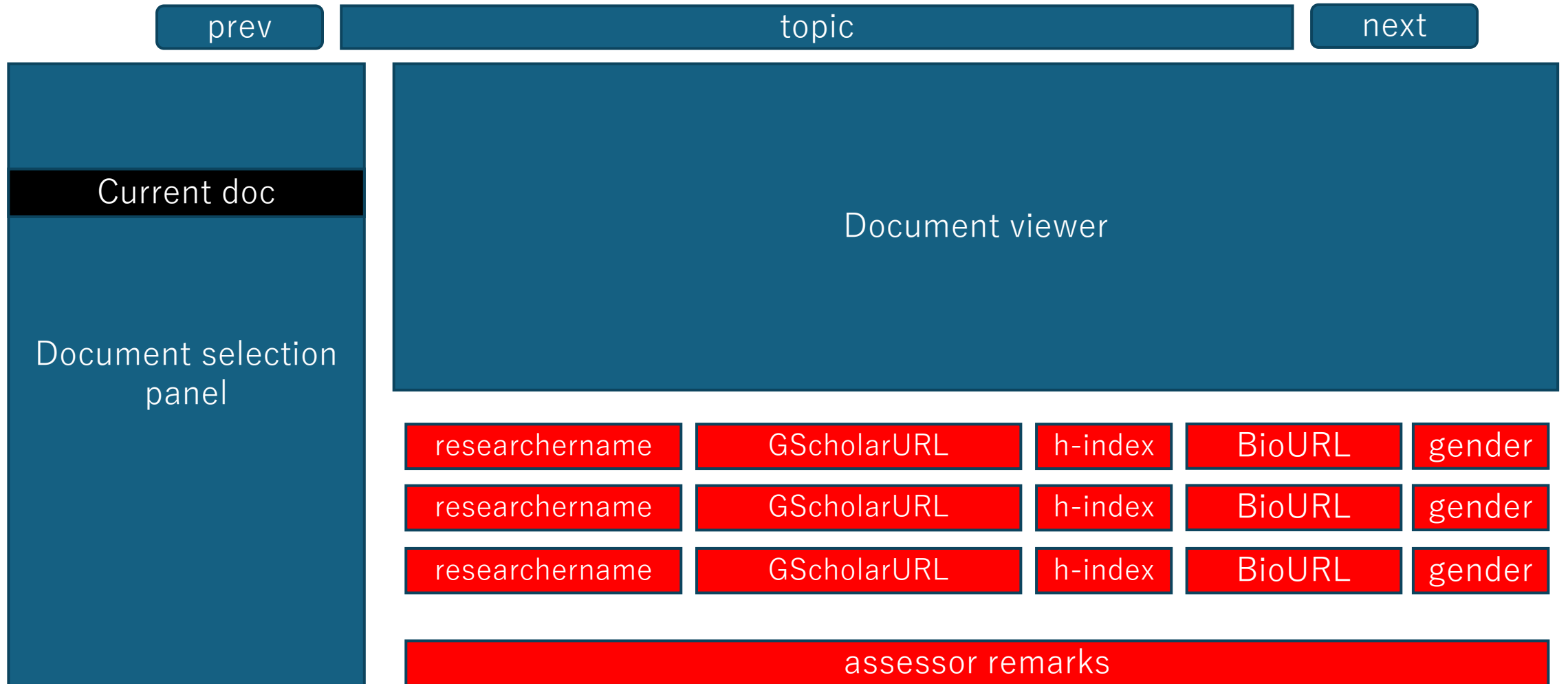
Relevant entity schema (key attributes are underlined)



Annotators will use their favourite search engines to locate BioURLs, GoogleScholarURLs, and IMDbURLs.

FAIRE: FAIRE stands for Annotation Interface for Relevant Entities

Annotation interface (R topic)



Backend records: <topicID, docID, ResearcherName, GScholarURL, h-index, BioURL, he/she/other>

Deriving page relevance

If 2 annotators each find 3 nonoverlapping relevant entities

d : page

$E(d)$: set of relevant entities extracted from d ($|E(d)| \leq 6$)

$r(e) \in \{1,2\}$: relevance level of $e \in E(d)$

Page relevance level $g(d) \in \{0,1,2\}$ defined as follows

$$g(d) = \begin{cases} 0 & (E(d) = \emptyset); \\ \max_{e \in E(d)} r(e) & (\text{otherwise}). \end{cases}$$

Max relevance level $G=2$

Page relevance level = max entity relevance level within page

Deriving page group membership

Group 1

$C = \{C_1, \dots, C_{|C|}\}$: attribute set

$F(e, C_i)$: flag that maps e to exactly one group

Hard group membership for entities

A researcher whose h-index=5 (C =HINDEX):

$$F(e, C_1)=1, F(e, C_2) = F(e, C_3) = F(e, C_4) = 0$$

Ad researcher whose bio says "he" (C =GENDER):

$$F(e, C_1)=1, F(e, C_2) = F(e, C_3) = 0$$

Group membership probabilities of d :

Uniform for nonrelevant page

$$P(d, C_i) = \begin{cases} 1/|C| & (E(d) = \emptyset); \\ \frac{|\sum_{e \in E(d)} F(e, C_i)|}{|\sum_i \sum_{e \in E(d)} F(e, C_i)|} & (\text{otherwise}). \end{cases}$$

	C1	C2	C3	C4
e1	Red	Grey	Grey	Grey
e2	Red	Grey	Grey	Grey
e3	Grey	Red	Grey	Grey

2/3

1/3

Deriving page group membership

$C = \{C_1, \dots, C_{|C|}\}$: attribute set

$ORIGIN(e) (\subseteq C)$: set of geo regions

for movie $e \in E(d)$ ($m = |ORIGIN(e)| (\geq 1)$)

For e , if **Countries of Origin** = {UK, Russia}
 UK \rightarrow Europe
 Russia \rightarrow Asia, Europe
 \Rightarrow **ORIGIN(e)** = {Asia, Europe}

Soft group membership wrt ORIGIN for movie entities

$$G(e, C_i) = \begin{cases} 1/m & (C_i \in ORIGIN(e)); \\ 0 & (\text{otherwise}). \end{cases}$$

$$P(d, C_i) = \begin{cases} 1/|C| & (E(d) = \emptyset); \\ \frac{|\sum_{e \in E(d)} G(e, C_i)|}{|\sum_i \sum_{e \in E(d)} G(e, C_i)|} & (\text{otherwise}). \end{cases}$$

Uniform for nonrelevant page

	C1	C2	C3	C4
e1	0.5		0.5	
e2	1			
e3		0.5	0.5	

1.5/3 0.5/3 1.0/3

GFR (Group Fairness and Relevance)

Sakai+23TOIS

L: SERP

$d_{L@k}$: doc at rank k in L

Probability that users will be satisfied with doc at k

$$p_{L@k}^{\text{sat}} = \frac{2^{g(d_{L@k})} - 1}{2^G}$$

Page relevance level

0, 1/4, 3/4 in our task

Probability that users will reach k and finally get satisfied

$$Decay_{L@k} = \begin{cases} p_{L@1}^{\text{sat}} & k = 1 \\ p_{L@k}^{\text{sat}} \prod_{j=1}^{k-1} (1 - p_{L@j}^{\text{sat}}) & \text{otherwise.} \end{cases}$$

$$GFR(L) =$$

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

Similarity between achieved distribution@k and target

Normalised Cumulative Utility (Sakai+08EVIA) e.g. ERR

Decay: probability that the users will abandon the SERP at rank k

ERR (Expected Reciprocal Rank) user model

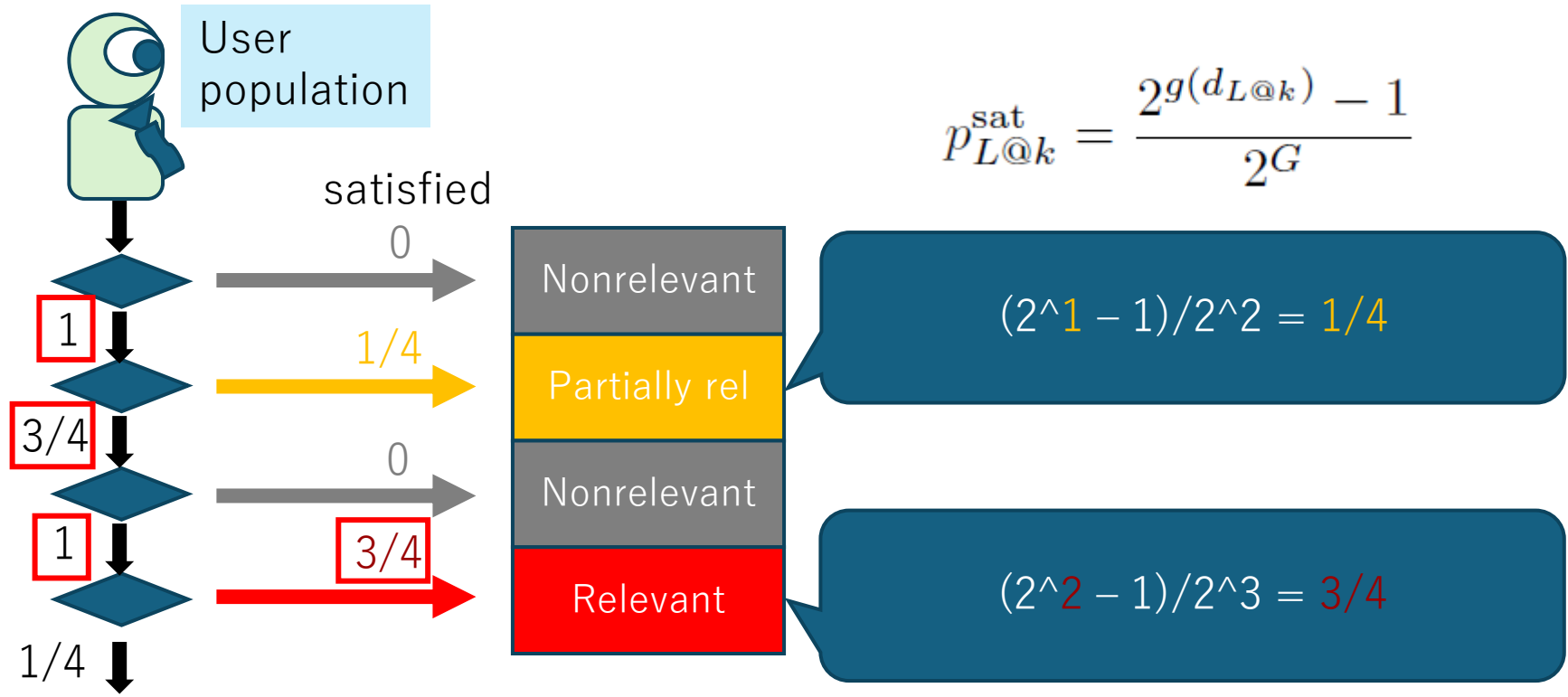
$$p_{L@k}^{\text{sat}} = \frac{2^{g(d_{L@k})} - 1}{2^G}$$

Decay $L@1 = 0$

Decay $L@2 = 1 * (1/4) = 1/4$

Decay $L@3 = 1 * (3/4) * 0 = 0$

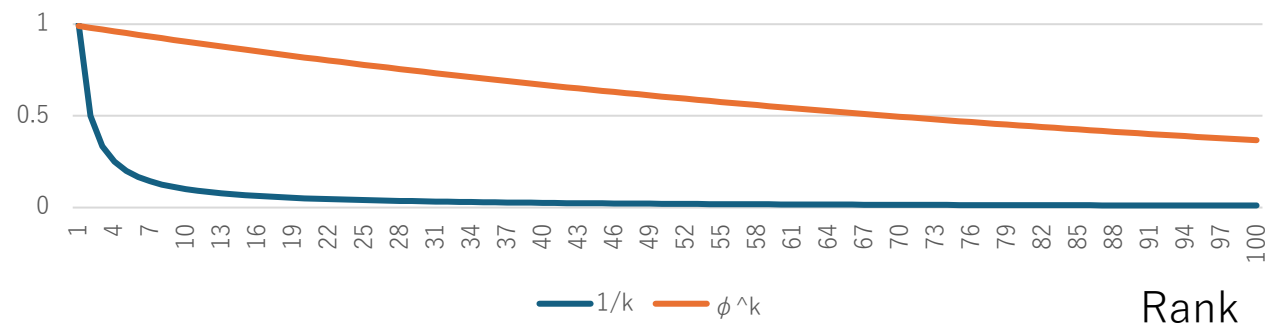
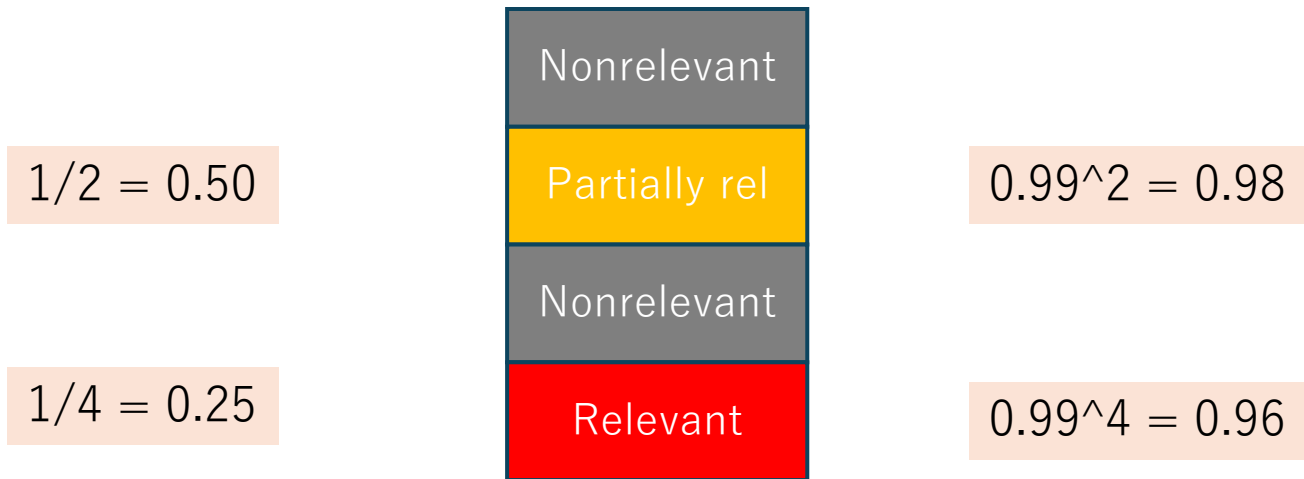
Decay $L@4 = 1 * (3/4) * 1 * (3/4) = 8/16$



Utility: how useful was the top k of the SERP?

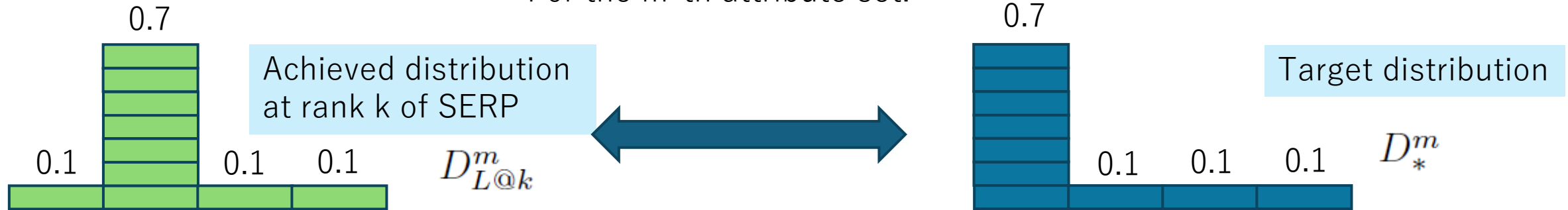
$$Utility_{L@k}^{ERR} = 1/k$$

$$Utility_{L@k}^{iRBU} = \phi^k \quad (\phi = 0.99)$$



DistrSim: Similarity between achieved distribution@k and target

For the m-th attribute set:

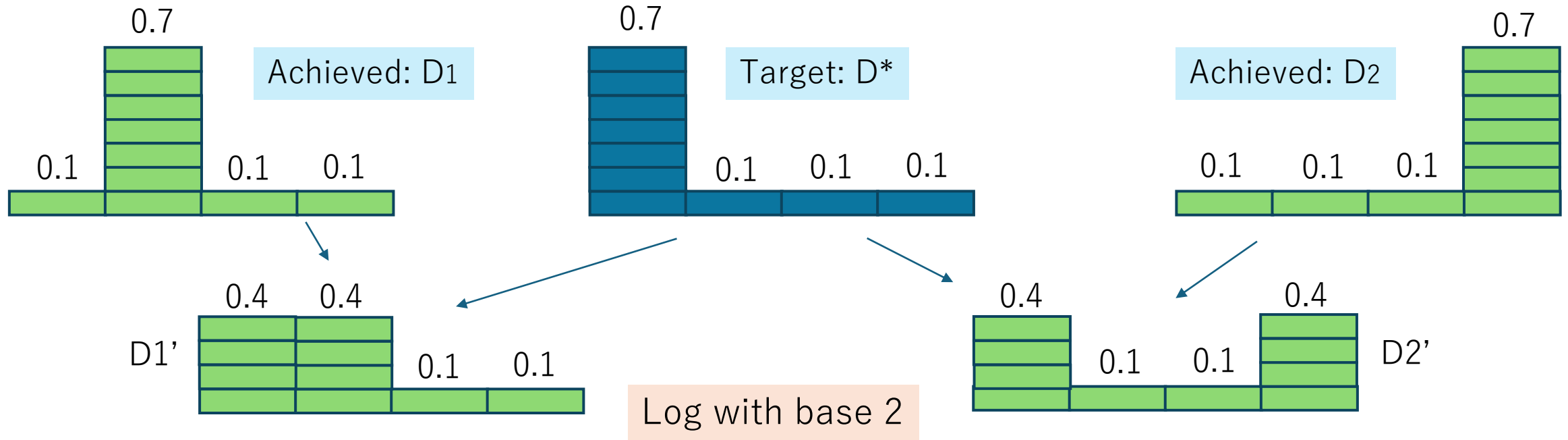


Similarity: larger=better

$$DistrSim_{L@k}^m(D_{L@k}^m \parallel D_*^m) = 1 - Divergence_{L@k}^m(D_{L@k}^m \parallel D_*^m)$$

- For attribute sets containing **nominal groups**:
Divergence= **JSD** (Jensen-Shannon Divergence)
- For attribute sets containing **ordinal groups**:
Divergence= **NMD** (Normalised Match Distance) or
RNOD (Root Normalised Order-aware Divergence)

JSD etc. are not suitable for ordinal groups



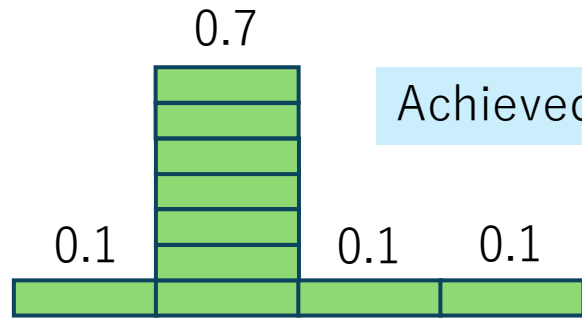
$$\begin{aligned} \text{KLD}(D_1 || D_1') &= \\ 0.1 \log(0.1/0.4) + 0.7 \log(0.7/0.4) &= 0.3651 \\ \text{KLD}(D^* || D_1') &= \\ 0.7 \log(0.7/0.4) + 0.1 \log(0.1/0.4) &= 0.3651 \\ \text{JSD} &= (0.3651 + 0.3651) / 2 = \mathbf{0.3651} \end{aligned}$$

$$\begin{aligned} \text{KLD}(D_2 || D_2') &= \\ 0.1 \log(0.1/0.4) + 0.7 \log(0.7/0.4) &= 0.3651 \\ \text{KLD}(D^* || D_2') &= \\ 0.7 \log(0.7/0.4) + 0.1 \log(0.1/0.4) &= 0.3651 \\ \text{JSD} &= (0.3651 + 0.3651) / 2 = \mathbf{0.3651} \end{aligned}$$

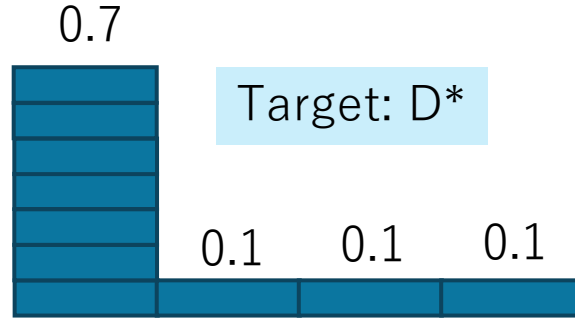
D1 (not too bad) and D2 (terrible) considered equivalent

aka Earth Mover's Distance

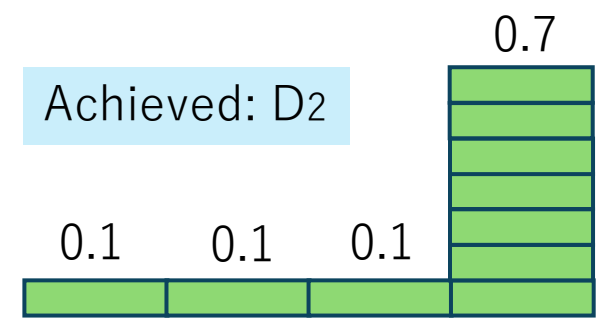
NMD (Normalised Match Distance)



Achieved: D1



Target: D*



Achieved: D2

Cumulative:

(0.1, 0.8, 0.9, 1.0)

Cumulative:

(0.7, 0.8, 0.9, 1.0)

Cumulative:

(0.1, 0.2, 0.3, 1.0)

$$\begin{aligned} \text{NMD} &= \\ & (|0.1-0.7| + |0.8-0.8| + \\ & |0.9-0.9| + |1.0-1.0|) / 3 \\ &= 0.6 / 3 \\ &= 0.2000 \end{aligned}$$

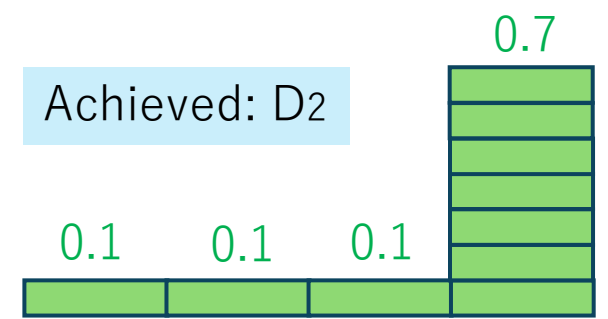
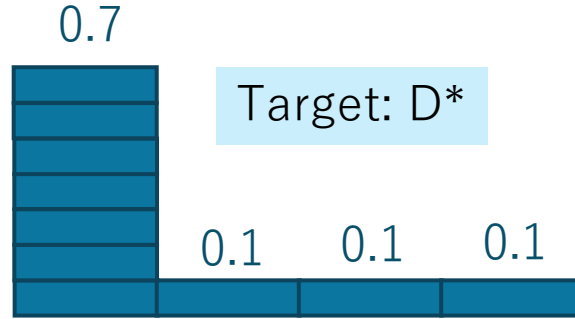
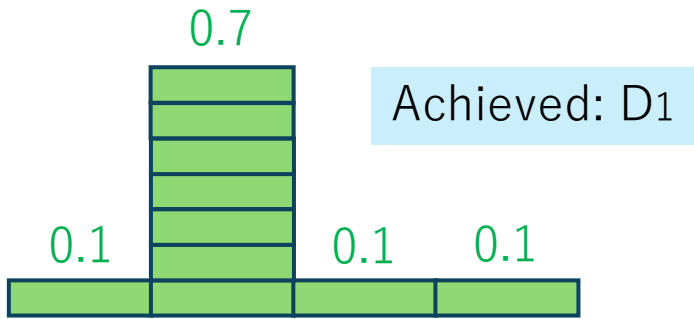
Closer to the target!

$$\begin{aligned} \text{NMD} &= \\ & (|0.1-0.7| + |0.2-0.8| + \\ & |0.3-0.9| + |1.0-1.0|) / 3 \\ &= (0.6+0.6+0.6) / 3 \\ &= 0.6000 \end{aligned}$$

For computing RNOD

DW (Distance-Weighted sum of squares)

$$DW_i = \sum_{j=1}^{|C|} |i - j| (P_j - P_j^*)^2$$



$$DW_1 = \overset{j=2}{1 \cdot (0.7 - 0.1)^2} + \overset{j=3}{2 \cdot (0.1 - 0.1)^2} + \overset{j=4}{3 \cdot (0.1 - 0.1)^2}$$

$$= 0.36$$

$$DW_2 = \overset{j=1}{1 \cdot (0.1 - 0.7)^2} + \overset{j=3}{1 \cdot (0.1 - 0.1)^2} + \overset{j=4}{2 \cdot (0.1 - 0.1)^2}$$

$$= 0.36$$

$$DW_3 = \overset{j=1}{2 \cdot (0.1 - 0.7)^2} + \overset{j=2}{1 \cdot (0.7 - 0.1)^2} + \overset{j=4}{1 \cdot (0.1 - 0.1)^2}$$

$$= 1.08$$

$$DW_4 = \overset{j=1}{3 \cdot (0.1 - 0.7)^2} + \overset{j=2}{2 \cdot (0.7 - 0.1)^2} + \overset{j=3}{1 \cdot (0.1 - 0.1)^2}$$

$$= 1.80$$

$$DW_1 = \overset{j=2}{1 \cdot (0.1 - 0.1)^2} + \overset{j=3}{2 \cdot (0.1 - 0.1)^2} + \overset{j=4}{3 \cdot (0.7 - 0.1)^2}$$

$$= 1.08$$

$$DW_2 = \overset{j=1}{1 \cdot (0.1 - 0.7)^2} + \overset{j=3}{1 \cdot (0.1 - 0.1)^2} + \overset{j=4}{2 \cdot (0.7 - 0.1)^2}$$

$$= 1.08$$

$$DW_3 = \overset{j=1}{2 \cdot (0.1 - 0.7)^2} + \overset{j=2}{1 \cdot (0.1 - 0.1)^2} + \overset{j=4}{1 \cdot (0.7 - 0.1)^2}$$

$$= 1.08$$

$$DW_4 = \overset{j=1}{3 \cdot (0.1 - 0.7)^2} + \overset{j=2}{2 \cdot (0.1 - 0.1)^2} + \overset{j=3}{1 \cdot (0.1 - 0.1)^2}$$

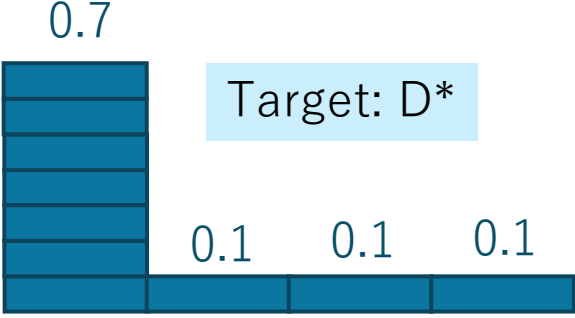
$$= 1.08$$

RNOD (Root Normalised Order-aware Divergence)

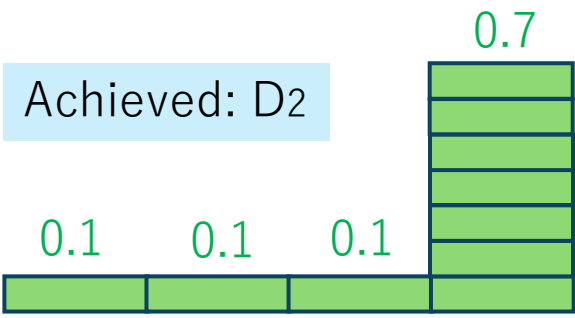


Achieved: D1

DW (0.36, 0.36, 1.08, 1.08)



Target: D*



Achieved: D2

DW (1.08, 1.08, 1.08, 1.08)

OD =
 $(0.36 + 0.36 + 1.08 + 1.08) / 4$
 = 0.90

$$OD(D \parallel D^*) = \frac{\sum_{i \text{ s.t. } C_i \in C^*} DW_i}{|C^*|}$$

OD =
 $(1.08 + 1.08 + 1.08 + 1.08) / 4$
 = 1.08

RNOD =
 $\text{SQRT}(0.90 / 3)$
 = 0.5477

$$RNOD(D \parallel D^*) = \sqrt{\frac{OD(D \parallel D^*)}{|C| - 1}}$$

RNOD =
 $\text{SQRT}(1.08 / 3)$
 = 0.6000

Closer to the target!

Evaluating intersectional group fairness

- R topics
relevance

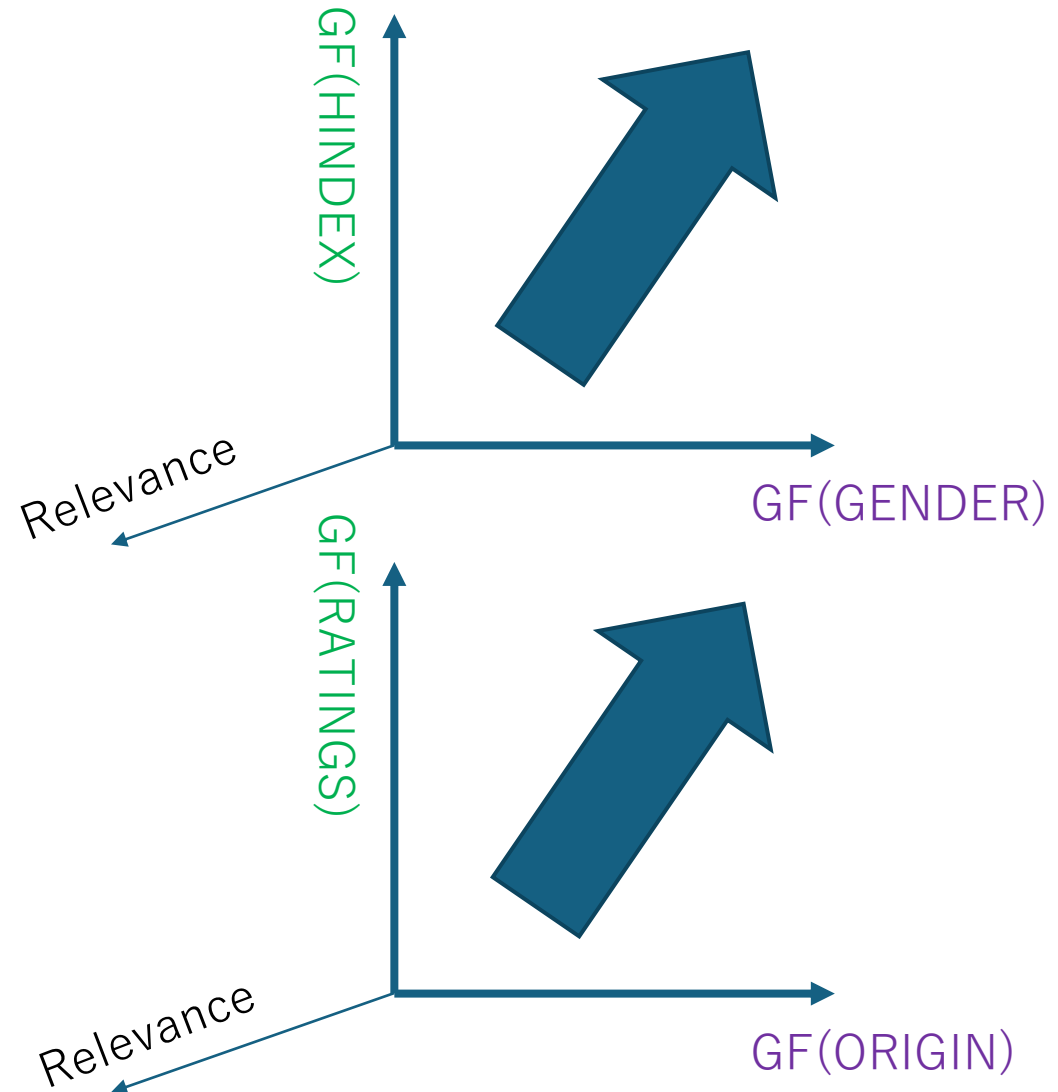
HINDEX (ordinal)

GENDER (nominal)

- M topics
relevance

RATINGS (ordinal)

ORIGIN (nominal)



How is GFR different from the single-ranking measure used at TREC 2022?

https://fair-trec.github.io/docs/Fair_Ranking_2022_Participant_Instructions.pdf

Main diffs:

- **Decay**: TREC uses nDCG decay (relevance-unaware); we use ERR decay (relevance-aware).
- **Achieved distribution**: TREC computes a single distribution for a SERP; we obtain a distribution at every rank with a relevant doc.
- **Divergence**: TREC uses JSD; we use JSD for **nominal groups** but NMD and RNOD for **ordinal groups**.
- **Combining relevance and group fairness**: TREC multiplies the two; we average relevance (ERR or iRBU score) and one or two GF scores.

OUTLINE

1. Motivation
2. Web Search Subtask
3. Conversational Search Subtask
4. Summary

Output: user and system turns (R-topic: researchername + Google Scholar URL)

Example R-topic: “Researchers who have coauthored at least one paper with Stephen E. Robertson, the inventor of BM25.”

Example output:

Must include both user turns (prompts) and system turns

U: Please return a list of researchers who have coauthored at least one paper with Stephen E. Robertson. Each researcher name must be accompanied by a Google Scholar URL.

Format: <researcher name>, <Google Scholar URL>

S:

Karen Sparck Jones, <https://scholar.google.com/citations?user=HzDsaGYAAAAJ>

Gareth Jones, https://scholar.google.com/citations?user=YJuN_H8AAAAJ

U: Please name more.

Muti-turn (optional)

S:

Tetsuya Sakai, <https://scholar.google.com/citations?user=v>

Output may be single-turn or multi-turn

Each user turn must include “U:” as a prefix.
Each system turn must include “S:” as a prefix.
Each output must begin with a “U:”
Up to 1,250 words (≐ 5 minutes reading time)
#words will be counted based on white spaces

Output: user and system turns (M-topic: movienename + IMDbURL)

Example M-topic: “sci-fi movies that feature time travel, listed in IMDb.”

Example output:

U: Please return a list of sci-fi movies that feature time travel. Each movie must be accompanied by an IMDb URL.

Format: <movietitle>,<IMDb URL>

S:

Here are some sci-fi movies that feature time travel, listed in IMDb:

Back to the Future (1985), <https://www.imdb.com/title/tt0088763/>

The Terminator (1984), <https://www.imdb.com/title/tt0088247/>

Groundhog Day (1993), <https://www.imdb.com/title/tt0107048/>

Interstellar (2014), <https://www.imdb.com/title/tt0816692/>

The Time Machine (1960), <https://www.imdb.com/title/tt0054387/>

I hope this helps! Let me know if there's anything else I can help you with.

Each user turn must include “U:” as a prefix.
Each system turn must include “S:” as a prefix.
Each output must begin with a “U:”
Up to 1,250 words (≅ 5 minutes reading time)
#words will be counted based on white spaces

Output: user and system turns (Y-topic: videotitle + youtubeURL)

Example Y-topic: “Covers of songs written by Burt Bacharach and Hal David, available on YouTube. Only Bacharach/David collaborations (just these two writers) are relevant”

Example output:

U: Please return a list of covers of Bacharach/David songs. Each cover must be accompanied by a YouTube URL.

Format: <youtube title>, <youtube URL>

S: Here are the covers of songs written by Burt Bacharach and Hal David that are available on YouTube:

Walk On By, <https://www.youtube.com/watch?v=AO073fekFJk>

I Say a Little Prayer, <https://www.youtube.com/watch?v=KtBbyglq37E>

Raindrops Keep Fallin' on My Head, <https://www.youtube.com/watch?v=5DmYLRxR0Y8>

What the World Needs Now Is Love, <https://www.youtube.com/watch?v=YUaxVQPohIU>

Alfie, <https://www.youtube.com/watch?v=ZVMlk3xYaYo>

U: Can you name a few more?

S:

Walk On By, <https://www.youtube.com/watch?v=AO073fekFJk>

Each user turn must include “U:” as a prefix.
Each system turn must include “S:” as a prefix.
Each output must begin with a “U:”
Up to 1,250 words (≐ 5 minutes reading time)
#words will be counted based on white spaces

Duplicate entities will not be rewarded

Subtask workflow

Annotators

Organisers

Participants

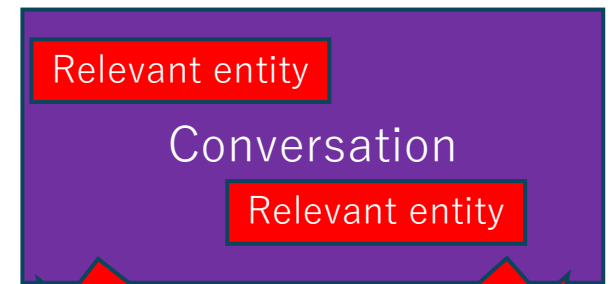
Release the topic set

Submit runs (a textual conversation for each topic)

Collect all submitted conversations for each topic

Annotate all entities (with URLs) in each conversation

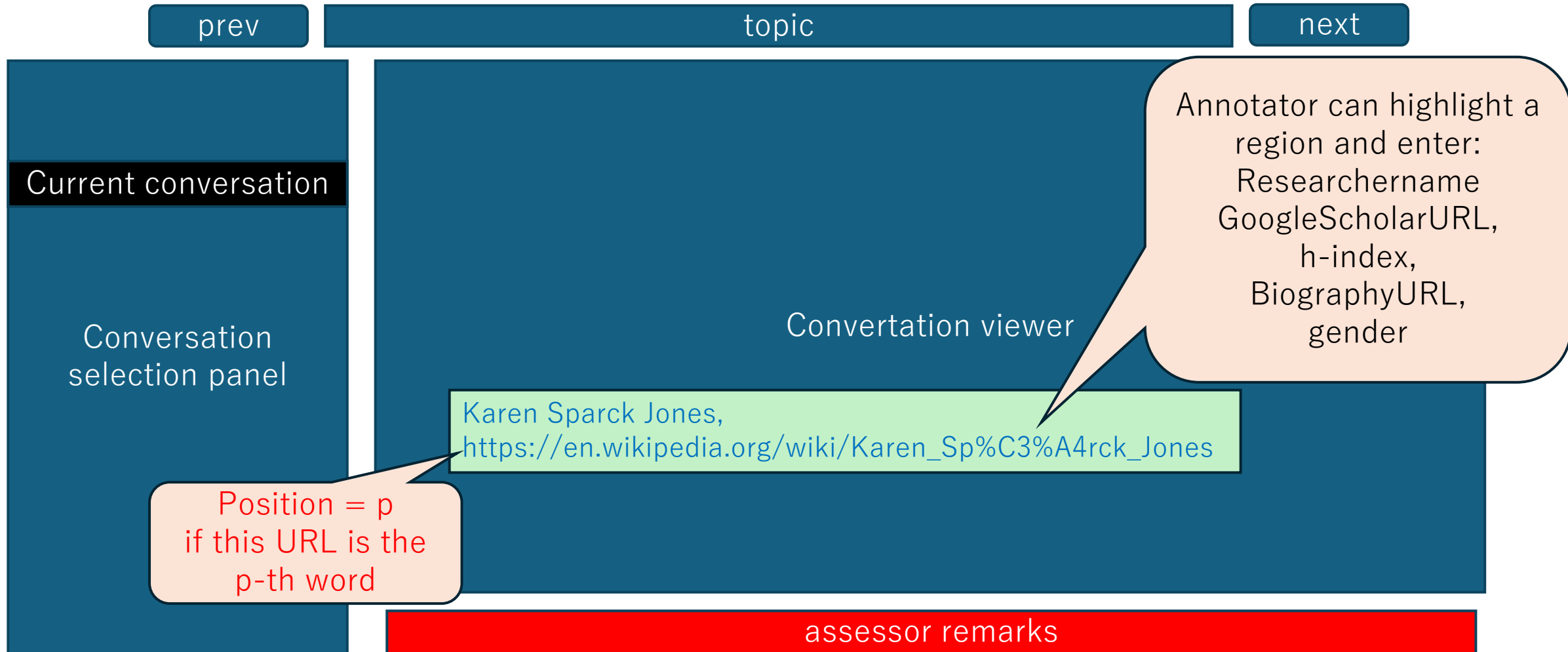
Derive relevance + group membership of each conversation, compute evaluation measures for each run



Relevance

Group membership

Annotation interface (R topic)



Backend records: <topicID, docID, position, ResearcherName, GScholarURL, h-index, BioURL, he/she/other>

Position (white-space-based)

1 U:	24 Here
2 Please	25 are
3 return	26 the
4 a	27 covers
5 list	28 of
6 of	29 songs
7 covers	30 written
8 of	31 by
9 Bacharach/David	32 Burt
10 songs.	33 Bacharach
11 Each	34 and
12 cover	35 Hal
13 must	36 David
14 be	37 that
15 accompanied	38 are
16 by	39 available
17 a	40 on
18 YouTube	41 YouTube:
19 URL.	42 Walk
20 Format:	43 On
21 <youtube title>,	44 By,
22 <youtube URL>	45 https://www.youtube.com/watch?v=AO073fekFJk
23 S:	

User prompt is included in the word count (to reflect user effort):
Try to make it short

Position of 1st relevant entity

First relevant entity

GFRC: GFR for Conversations [\[Sakai23EVIA\]](#)

$C = (U_1, S_1, \dots, U_T, S_T)$: T-round user-system conversation

n_{ij} : j-th nugget (=relevant piece of text) in i-th system turn

=relevant entity

Relevance
score

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

Gain value of nugget

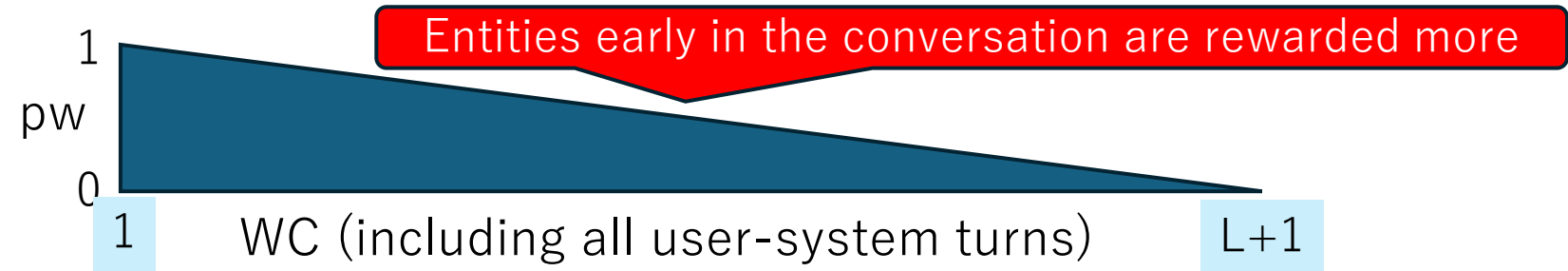
Normalisation factor

Position-based weight of **nugget**
(position defined wrt C)

GFRC: R instantiation

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

Average reading speed (English): 250 words/minute
 Given 5 minutes, up to $L=1,250$ words can be read
 (Nuggets after 1,250 words are worthless)



When every word is a relevant nugget (unattainable)

$$pw(n) = \max\left(0, 1 - \frac{WC(n) - 1}{L}\right)$$

$$\mathcal{N} = \sum_{l=1}^L \left(1 - \frac{l-1}{L}\right) = \frac{L+1}{2}$$

$$R(C) = \frac{2}{L+1} \sum_{i=1}^T \sum_{n_{ij} \in S_i} \max\left(0, 1 - \frac{WC(n_{ij})}{L}\right)g(n_{ij})$$

GFRC: GFR for Conversations

Group Fairness score

$$GF(C) = \frac{1}{N'} \sum_{i=1}^T PW(S_i) \sum_{m=1}^M w_m \text{DistrSim}^m(D^m(S_i) \parallel D_*^m)$$

Normalisation factor

Position-based weight of **turn**
(position defined wrt C)

Weight for the
m-th attribute
set

Distribution
similarity between
achieved and
target (for the m-th
attribute set)

When there are M attribute sets to consider

GFRC: GF instantiation

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



$$GF(C) = \frac{1}{m} \sum_{m=1}^M GF^m(C)$$

Simply average DistrSim's over relevant system turns

$$GF^m(C) = \frac{1}{|\mathcal{S}|} \sum_{S_i \in \mathcal{S}} \text{DistrSim}^m(D^m(S_i) \parallel D_*^m)$$

As in Web Search, use JSD for nominal and NMD/RNOD for ordinal groups

GFRC: examples from
Sakai23EVIA

1st system turn

Bing

Here are some sci-fi movies that feature time travel, listed in IMDb:

Back to the Future (1985): <https://www.imdb.com/title/tt0088763/>

#L2

#RATINGS: (0,0,0,1) #1.3M

#ORIGIN: (0,1,0,0,0,0,0) #United States

The Terminator (1984): <https://www.imdb.com/title/tt0088247/>

#L2

#RATINGS: (0,0,1,0) #898K

#ORIGIN: (0,1/2,0,0,0,1/2,0,0) #United Kingdom,United States

Groundhog Day (1993): <https://www.imdb.com/title/tt0107048/>

#L2

#RATINGS: (0,0,1,0) #661K

#ORIGIN: (0,1,0,0,0,0,0) #United States

Interstellar (2014): <https://www.imdb.com/title/tt0816692/>

#L1

#RATINGS: (0,0,0,1) #2M

#ORIGIN: (0,2/3,0,0,0,1/3,0,0) #United States,United Kingdom,Canada

The Time Machine (1960): <https://www.imdb.com/title/tt0054387/>

#L2

#RATINGS: (0,0,1,0) #43K

#ORIGIN: (0,1,0,0,0,0,0) #United States

I hope this helps! Let me know if there's anything else I can help you with.

wc=35

wc=39

wc=43

wc=46

wc=51

Google

Sure, here are some sci-fi movies that feature time travel, listed in IMDb, with their IMDB links and posters:

Back to the Future (1985): A nerdy teenager accidentally travels back in time to 1955 and meets his future parents, jeopardizing his own existence.

Back to the Future (1985) movie posterOpens in a new window

www.imdb.com

Back to the Future (1985) movie poster

Pulp Fiction (1994): A series of interconnected stories in Los Angeles's underworld. One of the stories involves a mob hit that goes wrong and the hitman's attempt to fix it by traveling back in time.

Pulp Fiction (1994) movie posterOpens in a new window

www.originalfilmart.com

Pulp Fiction (1994) movie poster

A 452-word response that does not follow the specified format

1st system turn

Nuggets in blue
Annotations in red

Bing

Here are some sci-fi movies that feature time travel, listed in IMDb:

Highly relevant

wc=35

wc=39

#L2

#RATINGS: (0,0,1,0) #898K

#ORIGIN: (0,1/2,0,0,0,1/2,0,0) #United Kingdom,United States

wc=51

I hope this helps! Let me know if there's anything else I can help you with.

Google

Sure, here are some sci-fi movies that feature time travel, listed in IMDb, with their IMDB links and posters:

Back to the Future (1985): A nerdy teenager accidentally travels back in time to 1955 and meets his future parents, jeopardizing his own existence.

Back to the Future (1985) movie poster [Opens in a new window](#)

www.imdb.com

Back to the Future (1985) movie poster

Pulp Fiction (1994): A series of interconnected stories in Los Angeles's underworld. One of the

A 452-word response that does not follow the specified format

1st system turn

Nuggets in blue

Bing

Here are some sci-fi movies that feature time travel, listed in IMDb:

Back to the Future (1985): <https://www.imdb.com/title/tt0088763/>

#L2
#RATINGS: (0,0,0,1) #1.3M
#ORIGIN: (0,1,0,0,0,0,0) #United States

wc=35

The Terminator (1984): <https://www.imdb.com/title/tt0088247/>

#L2
#RATINGS: (0,0,1,0) #898K
#ORIGIN: (0,1/2,0,0,0,0,1/2,0,0) #United Kingdom,United States

wc=36

The Terminator (1984): <https://www.imdb.com/title/tt0088247/>

#L2
#RATINGS: (0,0,1,0) #898K
#ORIGIN: (0,1/2,0,0,0,0,1/2,0,0) #United Kingdom,United States

Group 1: $x < 100$
Group 2: $100 \leq x < 10K$
Group 3: $10K \leq x < 1M$
Group 4: $1M \leq x$
898K ratings on IMDb \Rightarrow Group 3

Groundhog Day (1999): <https://www.imdb.com/title/tt0101413/>

#L2
#RATINGS: (0,0,0,1) #2M
#ORIGIN: (0,2/3,0,0,0,1/3,0,0) #United States,United Kingdom,Canada

wc=40

The Time Machine (1960): <https://www.imdb.com/title/tt0054387/>

#L2
#RATINGS: (0,0,1,0) #43K
#ORIGIN: (0,1,0,0,0,0,0,0) #United States

wc=51

A 452-word response that does not follow the specified format

I hope this helps! Let me know if there's anything else I can help you with.

1st system turn

Bing

Here are some sci-fi movies that feature time travel, listed in IMDb:

Back to the Future (1985): <https://www.imdb.com/title/tt0088763/>

#L2

#RATINGS: (0,0,0,1) #1.3M

#ORIGIN: (0,1,0,0,0,0,0) #United States

wc=35

The Terminator (1984): <https://www.imdb.com/title/tt0088247/>

#L2

#RATINGS: (0,0,1,0) #898K

#ORIGIN: (0,1/2,0,0,0,1/2,0,0) #United Kingdom,United States

wc=39

The Terminator (1984): <https://www.imdb.com/title/tt0088247/>

#L2

#RATINGS: (0,0,1,0) #898K

#ORIGIN: (0,1/2,0,0,0,1/2,0,0) #United Kingdom,United States

Groundhog Day (1999): <https://www.imdb.com/title/tt0169858/>

#L2

#RATINGS: (0,0,1,0) #898K

#ORIGIN: (0,1,0,0,0,0,0) #United States

wc=40

Interstellar (2014): <https://www.imdb.com/title/tt0816692/>

#L1

#RATINGS: (0,0,0,1) #2M

#ORIGIN: (0,2/3,0,0,0,1/3,0,0) #United States,United Kingdom,Canada

wc=51

The Time Machine (1960): <https://www.imdb.com/title/tt0054387/>

#L2

#RATINGS: (0,0,1,0) #43K

#ORIGIN: (0,1,0,0,0,0,0,0) #United States

I hope this helps! Let me know if there's anything else I can help you with.

Group 1: Africa
Group 2: America
Group 3: Antarctica
Group 4: Asia
Group 5: Caribbean
Group 6: Europe
Group 7: Middle East
Group 8: Oceania
Countries of origin on IMDB:
US + UK ⇒ Groups 2 and 6

A 452-word response that does not follow the specified format

2nd system turn

U2: Can you name a few more?

Bing

Sure! Here are some more sci-fi movies that feature time travel, listed in IMDb:

12 Monkeys (1995): <https://www.imdb.com/title/tt0114746/>

#L2

#RATINGS: (0,0,1,0) #634K

#ORIGIN: (0,1,0,0,0,0,0) #United States

The Butterfly Effect (2004): <https://www.imdb.com/title/tt0289879/>

#L2

#RATINGS: (0,0,1,0) #508K

#ORIGIN: (0,1,0,0,0,0,0) #United States,Canada

Looper (2012): <https://www.imdb.com/title/tt1276104/>

#L2

#RATINGS: (0,0,1,0) #592K

#ORIGIN: (0,1/2,0,1/2,0,0,0,0) #United States,China

Edge of Tomorrow (2014): <https://www.imdb.com/title/tt1631867/>

#L2

#RATINGS: (0,0,1,0) #711K

#ORIGIN: (0,1,0,0,0,0,0) #United States,Canada

Predestination (2014): <https://www.imdb.com/title/tt2397535/>

#L2

#RATINGS: (0,0,1,0) #294K

#ORIGIN: (0,1/2,0,0,0,0,0,1/2)#Australia,United States

I hope this helps! Let me know if there's anything else I can help you with.

wc=91

wc=96

wc=99

wc=104

wc=107

U2: Adhere to the output format I specified:
"Format: movie title, iMDB URL"

Google

Sure, here are the sci-fi movies that feature time travel, listed in IMDb, in the specified format:

Movie title iMDB URL

Back to the Future (1985) <https://www.imdb.com/title/tt0088763/>

#L2

#RATINGS: (0,0,0,1) #1.3M

#ORIGIN: (0,1,0,0,0,0,0) #United States

Back to the Future (1985) movie posterOpens in a new window

www.imdb.com

Back to the Future (1985) movie poster

Pulp Fiction (1994) <https://www.imdb.com/title/tt0110912/>

#L0

Pulp Fiction (1994) movie posterOpens in a new window

www.originalfilmart.com

Pulp Fiction (1994) movie poster

Looper (2012) <https://www.imdb.com/title/tt1853739/>

#L0 # wrong URL "You're Next"

Looper (2012) movie posterOpens in a new window

www.imdb.com

Looper (2012) movie poster

Interstellar (2014) <https://www.imdb.com/title/tt0816692/>

#L1

#RATINGS: (0,0,0,1) #2M

#ORIGIN: (0,2/3,0,0,0,1/3,0,0) #United States,United Kingdom,Canada

:

wc=506

Wrong URL

wc=560

R scores

$$pw(n) = \max(0, 1 - \frac{WC(n) - 1}{L})$$

Bing

	pw(n)	g(n)	pw(n)*g(n)
Back to the Future	0.9728	1	0.9728
The Terminator	0.9696	1	0.9696
Groundhog Day	0.9664	1	0.9664
Interstellar	0.9640	0.5	0.4820
The Time Machine	0.9600	1	0.9600
12 Monkeys	0.9280	1	0.9280
The Butterfly Effect	0.9240	1	0.9240
Looper	0.9216	1	0.9216
Edge of Tomorrow	0.9176	1	0.9176
Predestination	0.9152	1	0.9152
		<i>R</i>	0.0143

Google

	pw(n)	g(n)	pw(n)*g(n)
Back to the Future	0.5960	1	0.5960
Interstellar	0.5528	0.5	0.2764
		<i>R</i>	0.0014

No relevant entities in 1st turn

Not really time travel

Only about 9.8% of Bing's R score

GF scores - RATINGS (Bing vs Google)

Membership vectors of
5 movies:

0	0	0	1
0	0	1	0
0	0	1	0
0	0	0	1
0	0	1	0

Gold distribution: uniform

Achieved distribution

DistrSim
(RNOD)

S_1	0	0	0.6	0.4	0.6773
S_2	0	0	1	0	0.4796
$GF^{RATINGS}$					0.5785

Membership vectors of 2 movies:

0	0	0	1
0	0	0	1

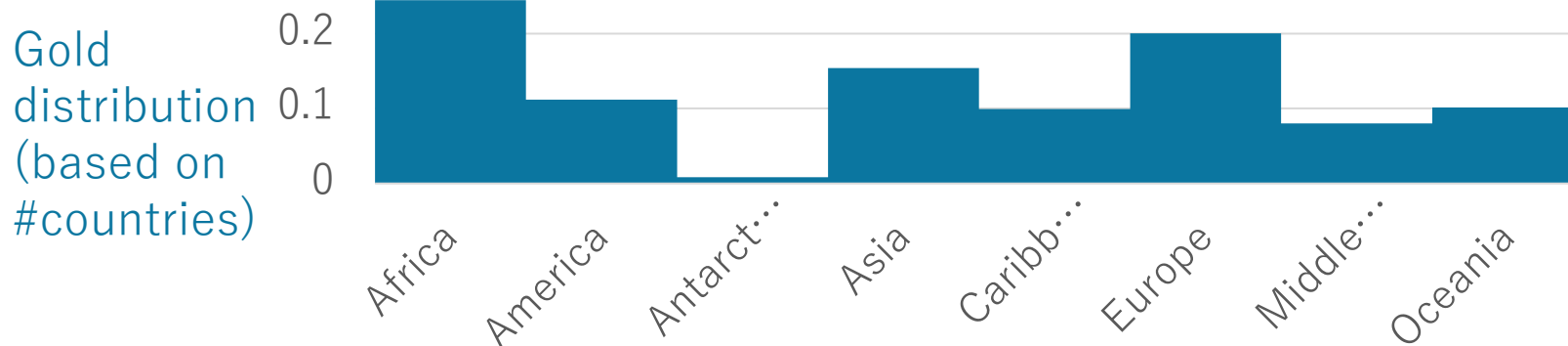
Achieved distribution

DistrSim
(RNOD)

S_1	-	-	-	-	-
S_2	0	0	0	1	0.4049
$GF^{RATINGS}$					0.4049

Google returns famous movies only

GF scores - ORIGIN (Bing vs Google)



	Africa	America	Antarctica	Asia	Caribbean	Europe	Middle East	Oceania		Africa	America	Antarctica	Asia	Caribbean	Europe	Middle East	Oceania		
S1	0	0.8333	0	0	0	0.1667	0	0	0.4303	-	-	-	-	-	-	-	-	-	-
S2	0	0.8	0	0.1	0	0	0	0.1	0.4682	0	0.8333	0	0	0	0.1667	0	0	0	0.4303
	GF^{ORIGIN}								0.4493	GF^{ORIGIN}								0.4303	

Google returns only American and European movies so the GF score is low. Bing does better (also returns Asian and Oceanian movies).

OUTLINE

1. Motivation
2. Web Search Subtask
3. Conversational Search Subtask
4. Summary

Timeline (tentative)

Nov 1, 2024: test topics released; task registrations due

Dec 15, 2024: run submissions due

Dec 2024-Jan 2025: Entity annotation + evaluation

Feb 1, 2025: Evaluation results + draft overview released

June 10-13, 2025: NTCIR-18@NII, Tokyo

We know you care about both fairness and IR.
That's why we know you will participate in this task!

Links

- [NTCIR-17 FairWeb-1 page \(with official results, relevant papers, etc.\)](#)
- [NTCIR-18 FairWeb-2 page \(with relevant papers, etc.\)](#)
- [Lots of videos and slides on divergences for quantification tasks, i.e., for comparing two probability mass functions \(from ACL 2021\)](#)