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
“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.
We only get famous and high h-index people…
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

Related people

Susan
C. J. van Rijsbergen
Ricardo Baeza-Yates
Gerard Salton
Calvin Mooers
Let’s consider group fairness

Attribute set: **HINDEX**

Target distribution: give more exposure to junior researchers!

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>x &lt; 10</td>
<td>10 ≤ x &lt; 30</td>
<td>30 ≤ x &lt; 50</td>
<td>50 ≤ x</td>
</tr>
</tbody>
</table>

Attribute set: **GENDER**

Target distribution: equal opportunities for different genders!

- Group 1: he
- Group 2: she
- Group 3: pronoun not found

Approximates true gender: only reflects whether “he” or “she” is found in the researcher bio.
Web search that considers group fairness

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](#), [Sakai21CIKM](#)

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

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

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

- **JSD** = 0.3651
- **NMD** = 0.2000
- **RNOD** = 0.5477

Closer to target

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>h-index</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

JSD = 0.3651

RNOD = 0.6000

NMD = 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”

System turns' achieved distribution

Target distribution

1st system turn

2nd system turn

1st user turn

2nd user turn

Please list IR evaluation experts

Can you name a few more?

User model
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

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

• M-topic

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

• 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."
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)

n topics
Subtask workflow

Annotators

Organisers

Participants

Release the topic set

Submit runs (SERP for each topic)

Form a depth-k pool for each topic

n topics

n topics

SERP SERP SERP
Subtask workflow

Annotators
- Annotate up to 3 relevant entities from each document

Organisers
- Release the topic set
- Form a depth-k pool for each topic

Participants
- Submit runs (SERP for each topic)
Subtask workflow

Annotators
- Annotate up to 3 relevant entities from each document

Organisers
- Release the topic set
- Form a depth-k pool for each topic
- Derive relevance + group membership of each page, compute evaluation measures for each run

Participants
- Submit runs (SERP for each topic)
- Web page
  - Relevant entity
  - Page relevance
- Group membership
Relevant entity schema (key attributes are **underlined**)

R

(GoogleScholarURL, h-index, ResearcherName, BioURL, he/she/other)

M

(MovieName, IMDbURL, #ratings, countries)

Y

(YouTubeContentURL, #subscribers)

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

d: page
E(d): set of relevant entities extracted from d (|E(d)| ≤ 6)
r(e) ∈ \{1, 2\}: relevance level of e ∈ E(d)

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

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

Page relevance level = max entity relevance level within page
Deriving page group membership

\[ C = \{ C_1, \ldots, C_{|C|} \} : \text{attribute set} \]

\[ F(e, C_i) : \text{flag that maps } e \text{ to exactly one group} \]

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

\[ F(e, C_1) = 1, \ F(e, C_2) = F(e, C_3) = F(e, C_4) = 0 \]

An 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:

\[
P(d, C_i) = \begin{cases} 
\frac{1}{|C|}, & (E(d) = \emptyset); \\
\frac{\sum_{e \in E(d)} F(e, C_i)}{\sum_i \sum_{e \in E(d)} F(e, C_i)}, & \text{(otherwise)}. 
\end{cases}
\]
Deriving page group membership

\[ C = \{C_1, \ldots, C_{|C|}\} \]: attribute set
\[ \text{ORIGIN}(e) (\subseteq C) \]: set of geo regions
for movie \( e \in E(d) \) (\( m = |\text{ORIGIN}(e)| (\geq 1) \))

Soft group membership wrt \text{ORIGIN} for movie entities

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

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

For \( e \), if \text{Countries of Origin} = \{UK, Russia\}
UK → Europe
Russia → Asia, Europe
⇒ \text{ORIGIN}(e) = \{Asia, Europe\}

Uniform for nonrelevant page
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

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

Probability that users will reach k and finally get satisfied

\[ \text{Decay}_{L@k} = \begin{cases} p_{sat}^{L@1} & \text{if } k = 1 \\ p_{sat}^{L@k} \prod_{j=1}^{k-1} (1 - p_{sat}^{L@j}) & \text{otherwise.} \end{cases} \]

\[ \text{GFR}(L) = \sum_{k=1}^{|L|} \text{Decay}_{L@k} \left( w_0 \text{Utility}_{L@k} + \sum_{m=1}^{M} w_m \text{DistrSim}_m^{L@k} \right) \]

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

Page relevance level

0, ¼, ¾ in our task

Similarity between achieved distribution@k and target
**Decay:** probability that the users will abandon the SERP at rank \( k \)

**ERR (Expected Reciprocal Rank) user model**

- \( P_{L@1} = 0 \)
- \( P_{L@2} = 1 \times \left( \frac{1}{4} \right) = \frac{1}{4} \)
- \( P_{L@3} = 1 \times \left( \frac{3}{4} \right) \times 0 = 0 \)
- \( P_{L@4} = \left( \frac{3}{4} \right) \times 1 \times \left( \frac{3}{4} \right) = \frac{8}{16} \)

\[
P_{\text{sat}} = \frac{2g(d_{L@k}) - 1}{2G}
\]

- \( \frac{2^1 - 1}{2^2} = \frac{1}{4} \)
- \( \frac{2^2 - 1}{2^3} = \frac{3}{4} \)
Utility: how useful was the top k of the SERP?

\[ Utility_{ERR}^{L@k} = \frac{1}{k} \]

\[ Utility_{L@k}^{IRBU} = \Phi^k \]

(\(\Phi = 0.99\))

Nonrelevant
Partially rel
Nonrelevant
Relevant

\[
\begin{align*}
1/2 &= 0.50 \\
1/4 &= 0.25 \\
0.99^2 &= 0.98 \\
0.99^4 &= 0.96
\end{align*}
\]
**DistrSim**: Similarity between achieved distribution@k and target

For the m-th attribute set:

For attribute sets containing **nominal groups**:

\[ \text{Divergence} = \text{JSD} \text{ (Jensen-Shannon Divergence)} \]

For attribute sets containing **ordinal groups**:

\[ \text{Divergence} = \text{NMD} \text{ (Normalised Match Distance)} \text{ or } \text{RNOD} \text{ (Root Normalised Order-aware Divergence)} \]

**Similarity**: larger=better

\[ \text{DistrSim}^m_{L@k}(D^m_{L@k} \| D^m_*) = 1 - \text{Divergence}^m_{L@k}(D^m_{L@k} \| D^m_*) \]

0.1 0.7

Achieved distribution at rank k of SERP

0.7 0.1 0.1

Target distribution

For the m-th attribute set:

**Similarity**: larger=better

\[ \text{DistrSim}^m_{L@k}(D^m_{L@k} \| D^m_*) = 1 - \text{Divergence}^m_{L@k}(D^m_{L@k} \| D^m_*) \]
JSD etc. are not suitable for ordinal groups

KLD(D1||D1') = 
0.1\text{log}(0.1/0.4) + 0.7\text{log}(0.7/0.4) = 0.3651
KLD(D*||D1') = 
0.7\text{log}(0.7/0.4) + 0.1\text{log}(0.1/0.4) = 0.3651
JSD = \frac{0.3651+0.3651}{2} = 0.3651

KLD(D2||D2') = 
0.1\text{log}(0.1/0.4) + 0.7\text{log}(0.7/0.4) = 0.3651
KLD(D*||D2') = 
0.7\text{log}(0.7/0.4) + 0.1\text{log}(0.1/0.4) = 0.3651
JSD = \frac{0.3651+0.3651}{2} = 0.3651

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

Log with base 2
NMD (Normalised Match Distance)

NMD = ( |0.1-0.7| + |0.8-0.8| + |0.9-0.9| + |1.0-1.0| )/3
= 0.6/3
= 0.2000

Closer to the target!

Cumulative: (0.1, 0.8, 0.9, 1.0)
Achieved: D1

Cumulative: (0.7, 0.8, 0.9, 1.0)
Target: D*

Cumulative: (0.1, 0.2, 0.3, 1.0)
Achieved: D2

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

aka Earth Mover’s Distance
For computing RNOD

\[ \text{DW} \quad \text{(Distance-Weighted sum of squares)} \]

\[ DW_i = \sum_{j=1}^{\left| C \right|} \left|i - j\right| \left(P_j - P_j^*\right)^2 \]

For computing RNOD
RNOD (Root Normalised Order-aware Divergence)

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

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

\[
\begin{align*}
OD &= \frac{(0.36 + 0.36 + 1.08 + 1.80)}{4} = 0.90 \\
RNOD &= \text{SQRT}(0.90/3) = 0.5477 \\
OD &= \frac{(1.08 + 1.08 + 1.08 + 1.08)}{4} = 1.08 \\
RNOD &= \text{SQRT}(1.08/3) = 0.6000
\end{align*}
\]
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:

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:

U: Please name more.

S:
Tetsuya Sakai, https://scholar.google.com/citations?user=vi56gz4AAAAJ

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
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:


I hope this helps! Let me know if there’s anything else I can help you with.
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=A0073fekFJk
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
Alfie, https://www.youtube.com/watch?v=ZVMIk3xYaYo

U: Can you name a few more?

S:
Walk On By, https://www.youtube.com/watch?v=A0073fekFJk

Duplicate entities will not be rewarded

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
Subtask workflow

**Annotators**
- Annotate all entities (with URLs) in each conversation

**Organisers**
- Release the topic set
- Collect all submitted conversations for each topic
- Derive relevance + group membership of each conversation, compute evaluation measures for each run

**Participants**
- Submit runs (a textual conversation for each topic)
Annotation interface (R topic)

Current conversation

Conversation selection panel

Conversation viewer

Karen Sparck Jones,

Position = p if this URL is the p-th word

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

Annotator can highlight a region and enter: Researchername, GoogleScholarURL, h-index, BiographyURL, gender
Position (white-space-based)

| 1U: | 2Please | 3return | 4a | 5list | 6of | 7covers | 8of | 9Bacharach/David | 10songs. | 11Each | 12cover | 13must | 14be | 15accompanied | 16by | 17a | 18YouTube | 19URL. | 20Format: | 21<youtube title>, | 22<youtube URL> | 23S: | 24Here | 25are | 26the | 27covers | 28of | 29songs | 30written | 31by | 32Burt | 33Bacharach | 34and | 35Hal | 36David | 37that | 38are | 39available | 40on | 41YouTube: | 42Walk | 43On | 44By, | 45https://www.youtube.com/watch?v=A073fekFJk |

| 1U: | 2Please | 3return | 4a | 5list | 6of | 7covers | 8of | 9Bacharach/David | 10songs. | 11Each | 12cover | 13must | 14be | 15accompanied | 16by | 17a | 18YouTube | 19URL. | 20Format: | 21<youtube title>, | 22<youtube URL> | 23S: | 24Here | 25are | 26the | 27covers | 28of | 29songs | 30written | 31by | 32Burt | 33Bacharach | 34and | 35Hal | 36David | 37that | 38are | 39available | 40on | 41YouTube: | 42Walk | 43On | 44By, | 45https://www.youtube.com/watch?v=A073fekFJk |

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

First relevant entity

Position of 1st relevant entity
GFRC: GFR for Conversations [Sakai23EVIA]

\[ C = (U_1, S_1, \ldots, U_T, S_T) : T \text{-round user-system conversation} \]

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

Relevance score

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

Normalisation factor

Gain value of nugget

Position-based weight of nugget (position defined wrt C)
GFRC: R instantiation

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

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)

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

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

Entities early in the conversation are rewarded more

When every word is a relevant nugget (unattainable)
GFRC: GFR for Conversations

Group Fairness score

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

- 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)
GFRC: GF instantiation

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

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

\[ GF^m(C) = \frac{1}{|S|} \sum_{S_i \in S} DistrSim^m(D^m(S_i) \parallel D^*_m) \]

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

Simply average DistrSim’s over relevant system turns
GFRC: examples from Sakai23EVIA
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.

**The Terminator (1984):** 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.

**Groundhog Day (1993):**

**Interstellar (2014):**

**The Time Machine (1960):**

I hope this helps! Let me know if there’s anything else I can help you with.
Here are some sci-fi movies that feature time travel, listed in IMDb:

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

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

  - RATINGS: (0,0,1,0) #661K
  - ORIGIN: (0,1,0,0,0,0,0,0) #United States

- **Interstellar (2014)**: https://www.imdb.com/title/tt0816692/
  - 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/
  - RATINGS: (0,0,0,1) #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.
Here are some sci-fi movies that feature time travel, listed in IMDb:

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

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

  - RATINGS: (0,0,1,0) #661K
  - ORIGIN: (0,1,0,0,0) #United States

- **Interstellar (2014):** [https://www.imdb.com/title/tt0816692/](https://www.imdb.com/title/tt0816692/)
  - RATINGS: (0,0,0,1) #2M
  - ORIGIN: (0,2/3,0,0,0,1/3,0,0) #United States, United Kingdom, Canada

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

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

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

#RATINGS: (0,0,1,0) #661K
#ORIGIN: (0,1,0,0,0,0,0,0) #United States

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

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

<table>
<thead>
<tr>
<th>Movie title</th>
<th>IMDb URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>#RATINGS: (0,0,1,0) #634K</td>
<td></td>
</tr>
<tr>
<td>#ORIGIN: (0,1,0,0,0,0,0,0)</td>
<td>United States</td>
</tr>
<tr>
<td>#RATINGS: (0,0,1,0) #508K</td>
<td></td>
</tr>
<tr>
<td>#ORIGIN: (0,1,0,0,0,0,0,0)</td>
<td>United States,Canada</td>
</tr>
<tr>
<td>#RATINGS: (0,0,1,0) #592K</td>
<td></td>
</tr>
<tr>
<td>#ORIGIN: (0,1/2,0,1/2,0,0,0,0)</td>
<td>United States,China</td>
</tr>
<tr>
<td>#RATINGS: (0,0,1,0) #711K</td>
<td></td>
</tr>
<tr>
<td>#ORIGIN: (0,1,0,0,0,0,0,0)</td>
<td>United States,Canada</td>
</tr>
<tr>
<td>#RATINGS: (0,0,1,0) #294K</td>
<td></td>
</tr>
<tr>
<td>#ORIGIN: (0,1/2,0,0,0,0,1/2)</td>
<td>Australia,United States</td>
</tr>
</tbody>
</table>

I hope this helps! Let me know if there’s anything else I can help you with.
### R scores

\[
p_{\text{w}(n)} = \max(0, 1 - \frac{W_{\text{C}(n)} - 1}{L})
\]

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

<table>
<thead>
<tr>
<th>Google</th>
<th>pw(n)</th>
<th>g(n)</th>
<th>pw(n)*g(n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back to the Future</td>
<td>0.5960</td>
<td>1</td>
<td>0.5960</td>
</tr>
<tr>
<td>Interstellar</td>
<td>0.5528</td>
<td>0.5</td>
<td>0.2764</td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td>0.0014</td>
</tr>
</tbody>
</table>

- **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)

<table>
<thead>
<tr>
<th></th>
<th>Membership vectors of 5 movies:</th>
<th>Achieved distribution</th>
<th>DistrSim (RNOD)</th>
<th>GF RATINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>0 0 0 1 0 0 0 0 1</td>
<td>0.6 0.6</td>
<td>0.6773</td>
<td>0.5785</td>
</tr>
<tr>
<td>$S_2$</td>
<td>0 0 0 1 0 0 0 0 1</td>
<td>0.4 0.4</td>
<td>0.4796</td>
<td></td>
</tr>
</tbody>
</table>

Gold distribution: uniform

<table>
<thead>
<tr>
<th></th>
<th>Membership vectors of 2 movies:</th>
<th>Achieved distribution</th>
<th>DistrSim (RNOD)</th>
<th>GF RATINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>0 0 0 1 0 0 0 0 1</td>
<td>- - - - -</td>
<td>- - - - -</td>
<td>- - - - -</td>
</tr>
<tr>
<td>$S_2$</td>
<td>0 0 0 0 1 0 0 0 1</td>
<td>0.4 0.4</td>
<td>0.4049</td>
<td>0.4049</td>
</tr>
</tbody>
</table>

Google returns famous movies only
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)