



Michael Ekstrand
Boise State University

Graham McDonald
University of Glasgow

Amifa Raj
Boise State University

Isaac Johnson
Wikimedia Foundation

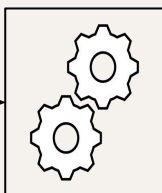
TREC 2022 FAIR RANKING TRACK OVERVIEW



Query



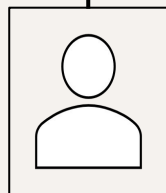
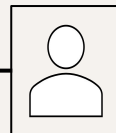
Ranking



Documents



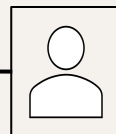
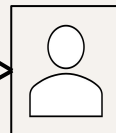
Protected Attributes



User

Exposure

Attention



Do relevant articles associated with protected attributes get fair exposure?

Fair Ranking Track 2022 vs Previous Years

- 2019-2020
 - Domain: Scholarly articles
 - Tasks:
 - Task 1: Ad-hoc retrieval
 - Task 2: Re-ranking
- 2021
 - Domain: Wikimedia
 - Tasks: Ad-hoc retrieval
 - Task 1: Single ranking
 - Task 2: Stochastic ranking
- 2022
 - Same tasks & domain
 - **New:** many more dimensions! And intersectionality evaluation

Equal Expected Exposure

“given a fixed information need, no item should have an expected exposure more or less than any other item of the same relevance.”

WikiProject



- Main page
- Contents
- Current events
- Random article
- About Wikipedia
- Contact us
- Donate
- Contribute
- Help
- Learn to edit
- Community portal
- Recent changes
- Upload file
- Tools
- Special pages
- Printable version
- Languages

Special page

Search results

Only searching in pages whose title starts with "Wikipedia:WikiProject" (Search all pages)

Q Agriculture

Advanced search:

Search in:

There is a page named **"Agriculture"** on Wikipedia

Wikipedia:WikiProject Agriculture

WikiProject **Agriculture**. Wikipedians have formed this collaboration resource and group dedicated to improving Wikipedia's coverage of **agriculture** and the **14 KB (0 words) - 16:11, 23 September 2021**

Wikipedia:WikiProject Agriculture/Participants

Cattle, other stock, **agricultural** biodiversity, traditional pasture systems, sustainable **agriculture**, semi-natural **agricultural** habitats etc.[] IvoShandor **25 KB (2,480 words) - 06:30, 28 October 2021**

Wikipedia:WikiProject Agriculture/Assessment

the **Agriculture** WikiProject! This department focuses on assessing the quality of Wikipedia's articles about **Agriculture** or the people of **Agriculture**. While **10 KB (1,855 words) - 06:30, 28 October 2021**

Wikipedia:WikiProject Agriculture

From Wikipedia, the free encyclopedia

"WP:AG" redirects here. For the guide for administrators, see Wikipedia:Administrators' guide.

“ The toils of agriculture will here be rewarded with a greater variety of valuable productions... ”

—Manasseh Cutler
The First Map and Description of Ohio



This is a **WikiProject**, an area for focused collaboration among Wikipedians. New participants are welcome; please feel free to participate!

[Guide to WikiProjects](#) · [Directory of WikiProjects](#)

[Shortcuts](#)
[WP:AG](#)
[WP:FARM](#)

? About WikiProject Agriculture [edit]

Welcome to **WikiProject Agriculture**. Wikipedians have formed this collaboration resource and group dedicated to improving Wikipedia's coverage of **agriculture** and the organization of information and articles on this topic. This page and its subpages contain their suggestions and various resources; it is hoped that this project will help to focus the efforts of other Wikipedians interested in the topic. If you would like to help, please [join](#) the project, inquire on the [talk page](#) and see the [to-do list](#) below.

Goals [edit]

This **Project** aims primarily to provide a consistent article structure for agricultural related topics while striving to develop and improve said agriculture articles. The goal is to make Wikipedia a comprehensive source of factually accurate, neutral articles that include relevant, credible facts.

? Scope [edit]

See also: List of covered agriculture subjects

This WikiProject strives to develop and improve articles at Wikipedia related to crop production, **livestock** management, **aquaculture**, **dairst farming** and **forest management**. The project also covers related areas, including both governmental and **NGO** regulatory agencies, **agribusiness**, support agencies such as **4H**, agricultural products including **fertilizers** and **herbicides**, pest management,

WikiProject Agriculture



Project information

Portal: [Agriculture and Agronomy Portal](#)

Shortcuts: [WP:FARM](#)
[WP:AG](#)
[WP:AGRI](#)

Banner: {{WPFarm}}

Page templates: {{Apples}}
 {{Cereals}}
 {{Cherries}}

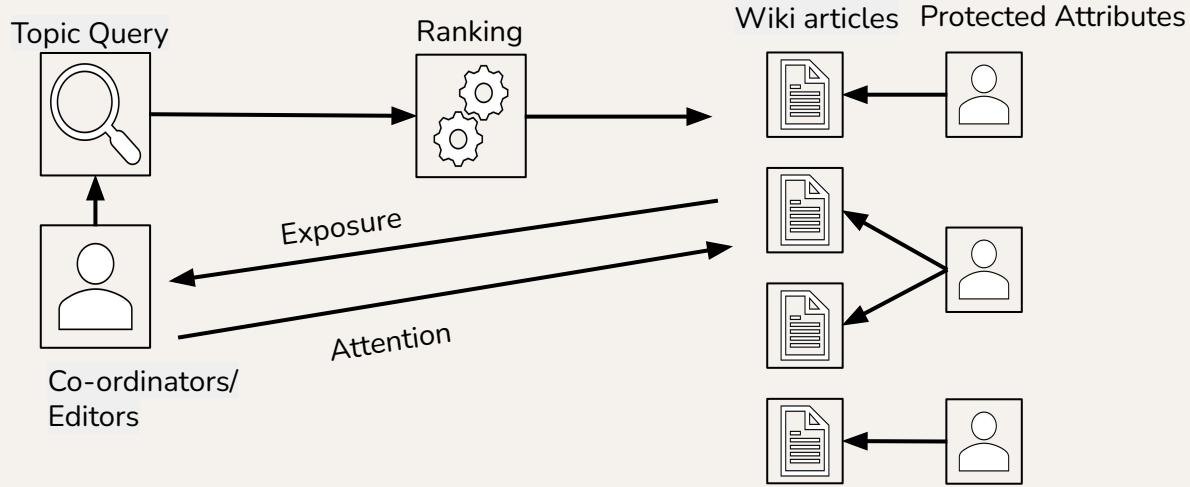
Userbox: {{User WP Agriculture}}

Core article: [Agriculture](#)
Open Tasks: [Agriculture tasks](#)

[view](#)

[edit this panel](#)

WikiProject



Documents: A subset of English language Wikipedia articles

Query: Topic form WikiProject

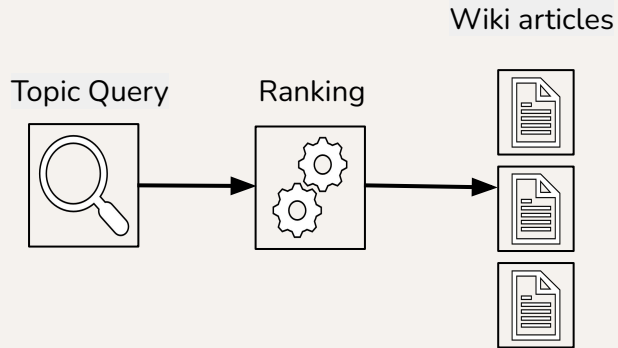
Fairness Objective: Ensure groups of articles associated with protected attributes get fair exposure

Task: ad-hoc retrieval

Task 1

Use Case: Help Wikiproject coordinators in finding articles for editors.

Single ranking per query



Evaluated based on



Relevance



Geographic



Demographic

Fairness

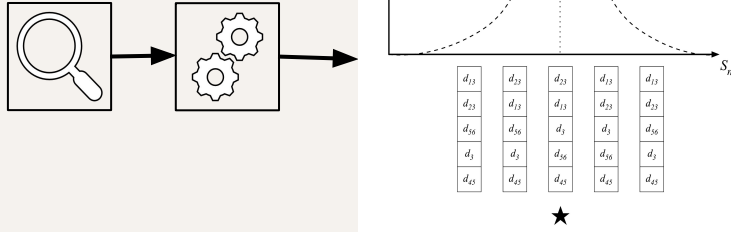
Task 2

Use Case: Help Wikiproject editors finding articles associated with a project (saved search)

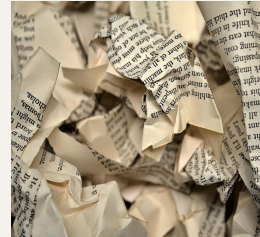
Distribution over ranking

Evaluated based on

Topic Query Ranking



Relevance



Quality



Geographic



Demographic

Fairness

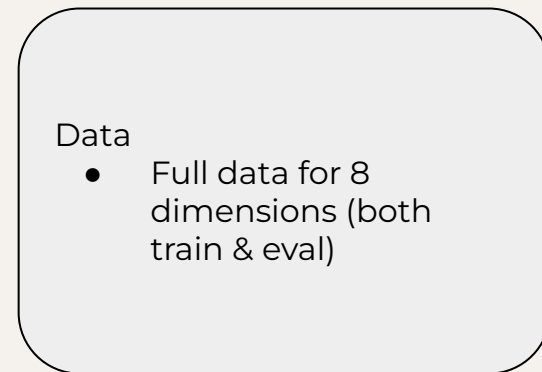
We Provided



WikiProject topic



English Wiki articles



Ranking Objectives

- Relevant documents come before irrelevant documents
- Fairness goal: Group exposure is fairly distributed according to the average of the distribution of relevant documents and the distribution of global population
- Relevant documents are sorted in nonincreasing order of work needed (Task 2)
 - Articles that need more editing will be on top
- For each query participants have to submit:
 - Task 1: single ranking consisting of 500 articles
 - Task 2: 100 rankings each consisting of 20 articles

Annotations

NIST assessors annotated the articles with binary relevance score.
Annotations were incomplete because:

- Task 2 generates massive data (~700,000 article-topic pairs)
- Incomplete articles and not having enough information

We obtained assessment through tiered-pooling

- The first 20 items of all rankings for Task 1 (all queries)
- The first 5 items of the first 25 rankings from every submission to Task 2 (about 75% of the queries).

Evaluation Strategy

- Fairness with respect to **8 dimensions** simultaneously

$$G^* = G_1 \times \cdots \times G_8$$

- “unknown” as a separate group where needed
- 11M total intersectional dimensions
- Log discounting for attention weight
- Measure exposure of group
- Compare with target exposure

Target Exposure

Ranked list



Exposure vector

Relevant set from
Wikipedia

Global population

average

- World population for geography
- Equality for gender
- No averaging for other attributes
- Multidimensional: outer product of single dimensions

- Wikipedia has well-documented bias
- Imbalance in group distribution in topic relevance
- No data on ideal distribution for any particular topic

Metrics for Single Ranking

- **Relevance Metric:** NDCG
- **Fair Ranking Metric:** Attention Weighted Rank Fairness (AWRF)
 - Jensen -shannon divergence between target and given exposure

$$AWRF(L) = 1 - d_{JS}(d_L, d_q^*)$$

d_L cumulated exposure a list gives to each group

d_q^* target exposure

- Combine fairness metric and relevance metric

$$M_1(L) = AWRF(L) \times NDCG(L)$$

Fairness Metric for Stochastic Rankings

Expected Exposure Metric

$$\pi(\rho, q) = P(\rho, q)$$

π ranking policy

q user query

ρ document ranking

$$EE(\Pi, q) = \|\epsilon - \epsilon^*\|_2^2$$

ϵ expected exposure

ϵ^* optimal expected exposure

Fairness Metric for Stochastic Rankings

Expected Exposure Metric for Group Fairness

$\gamma_g = \sum_d y_d a_{dg} \epsilon_d$ aggregate exposure for relevant docs in group

a_{dg} alignment of document with group

γ $|G| \times 1$ group exposure vector

$$EE_G(\pi, q) = \|\gamma - \gamma^*\|_2^2$$

**Expected-Exposure
Disparity (EE-D)**

$$EE_G(\pi, q) = \|\gamma\|_2^2 - 2\gamma^T \gamma^* + \|\gamma^*\|_2^2$$

**Expected-Exposure
Relevance (EE-R)**

This is **Expected Exposure Loss (EE-L)**

Submissions

5 Teams Participated:

- We received
 - Submissions from 5 teams for Task 1 (27 runs total)
 - Submissions from 2 teams for Task 2 (11 runs total)
- Approaches

Task 1

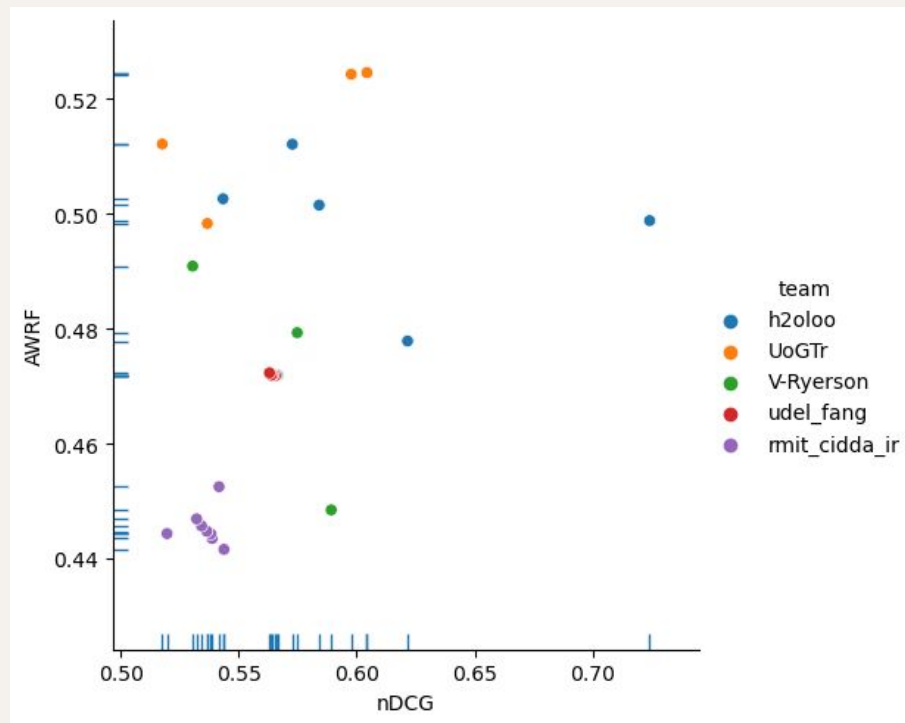
- Explicit Diversification
- Implicit diversification
- Data Fusion
- Constraint-based approaches
- Relevance-only baselines

Task 2

- Multi-Armed Bandits
- Colbert-based
- Tailored Diversification
- Relevance-only

Result: Task 1

run_name	nDCG	AWRF	Score	95% CI
tmt5	0.724215	0.498812	0.362626	(0.326, 0.397)
UoGRelvOnlyT1	0.604420	0.524572	0.325393	(0.284, 0.372)
UoGTrT1ColPRF	0.604420	0.524572	0.325393	(0.283, 0.369)
UoGTrExpE2	0.597741	0.524305	0.322974	(0.280, 0.368)
0mt5	0.621612	0.477848	0.298972	(0.267, 0.332)
0mt5_p	0.584138	0.501518	0.294887	(0.262, 0.326)
tmt5_p	0.572786	0.512117	0.294553	(0.260, 0.327)
FRT_constraint	0.574858	0.479295	0.278210	(0.245, 0.312)
bm25_p	0.543421	0.502616	0.277292	(0.241, 0.312)
UoGTrQE	0.536764	0.498313	0.273444	(0.240, 0.309)
UoGTrExpE1	0.517629	0.512153	0.271641	(0.238, 0.308)
UDInfo_F_bm25	0.566569	0.471891	0.270826	(0.236, 0.302)
ans_bm25	0.566126	0.471885	0.270617	(0.237, 0.303)
UDInfo_F_mlp2	0.565497	0.471797	0.270307	(0.235, 0.302)
FRT_attention	0.589254	0.448431	0.270190	(0.231, 0.311)
UDInfo_F_lgbm2	0.564489	0.471885	0.269847	(0.235, 0.302)
UDInfo_F_mlp4	0.563832	0.471889	0.269505	(0.234, 0.301)
UDInfo_F_lgbm4	0.563080	0.472303	0.269334	(0.235, 0.302)
FRT_diversity	0.530537	0.490888	0.264083	(0.229, 0.299)
rmit_cidda_ir_5	0.541712	0.452475	0.248452	(0.215, 0.282)
rmit_cidda_ir_1	0.543797	0.441563	0.243317	(0.210, 0.277)
rmit_cidda_ir_4	0.538810	0.443463	0.243134	(0.209, 0.278)
rmit_cidda_ir_7	0.538240	0.444274	0.242620	(0.209, 0.276)
rmit_cidda_ir_3	0.536504	0.444729	0.241975	(0.208, 0.275)
rmit_cidda_ir_6	0.534297	0.445653	0.241763	(0.208, 0.276)
rmit_cidda_ir_8	0.532212	0.446878	0.241549	(0.208, 0.276)
rmit_cidda_ir_2	0.519678	0.444310	0.234528	(0.201, 0.269)



Higher score is better

Task 1 - Individual Attributes

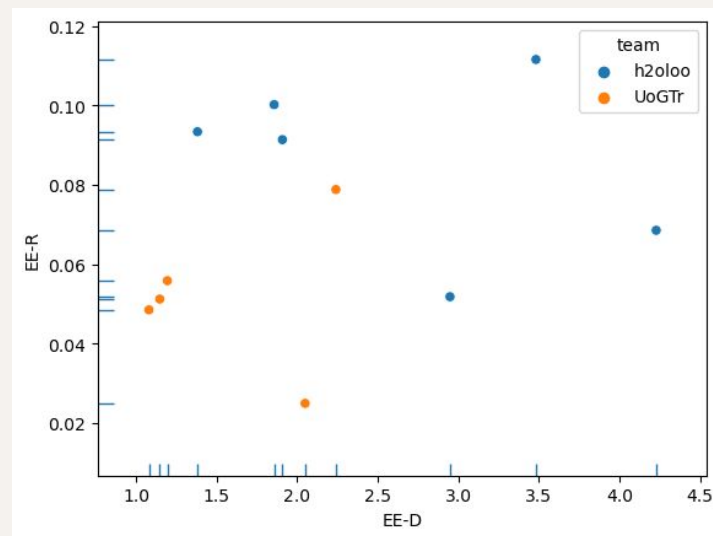
	Overall	age	alpha	gender	langs	occ	pop	src-geo	sub-geo
run_name									
tmt5	0.362626	0.686027	0.719043	0.679535	0.678615	0.682496	0.631264	0.645316	0.644979
UoGRelvOnlyT1	0.325393	0.584286	0.591560	0.526585	0.589594	0.526671	0.580183	0.554831	0.538912
UoGTrT1CoIPRF	0.325393	0.584286	0.591560	0.526585	0.589594	0.526671	0.580183	0.554831	0.538912
UoGTrExpE2	0.322974	0.579672	0.586859	0.545275	0.584931	0.544321	0.576468	0.548206	0.534440
0mt5	0.298972	0.583271	0.616369	0.583422	0.581721	0.584468	0.533271	0.552885	0.549564
0mt5_p	0.294887	0.555188	0.580112	0.551882	0.552981	0.551437	0.511165	0.531507	0.535161
tmt5_p	0.294553	0.548149	0.569553	0.543740	0.546267	0.542118	0.507668	0.527100	0.530302
FRT_constraint	0.278210	0.551129	0.571171	0.533015	0.549395	0.535801	0.522011	0.514404	0.505925
bm25_p	0.277292	0.517810	0.539497	0.515627	0.516485	0.514714	0.476814	0.496864	0.500517
UoGTrQE	0.273444	0.521577	0.529086	0.481319	0.522546	0.480685	0.497549	0.482825	0.490564

More on this tomorrow...

Result: Task 2

	EE-R	EE-D	EE-L	EE-L 95% CI
run_name				
UoGTrMabWeSA	0.0485	1.0791	1.1231	(0.9790, 1.3096)
UoGTrMabSaWR	0.0512	1.1462	1.1847	(1.0258, 1.3739)
UoGTrMabSAED	0.0558	1.1935	1.2228	(1.0507, 1.4185)
tmt5_p_e	0.0934	1.3803	1.3345	(1.1409, 1.5793)
0mt5_p_e	0.1002	1.8563	1.7968	(1.5600, 2.0779)
bm25_p_e	0.0914	1.9077	1.8659	(1.5343, 2.2897)
UoGTrMabSaNR	0.0249	2.0486	2.1397	(1.8286, 2.6452)
UogTRelvOnlyT2	0.0788	2.2398	2.2231	(1.8645, 2.7786)
0mt5_e	0.0518	2.9483	2.9856	(2.5022, 3.7271)
tmt5_e	0.1116	3.4819	3.3997	(3.0171, 3.8588)
ans_bm25_e	0.0685	4.2286	4.2324	(2.7795, 8.3580)

Lower EE-L is better



Lower EE-D is better; higher EE-R is better

Task 2 - Individual Attributes

	Overall	age	alpha	gender	langs	occ	pop	src-geo	sub-geo
run_name									
UoGTrMabWeSA	1.123	7.215	6.798	15.163	8.756	15.127	22.197	4.798	5.925
UoGTrMabSaWR	1.185	7.194	7.259	15.291	8.978	15.226	22.674	4.979	6.303
UoGTrMabSAED	1.223	7.136	7.111	15.636	9.127	15.551	21.423	4.884	6.314
tmt5_p_e	1.334	21.024	6.554	18.282	23.936	16.420	31.123	11.142	14.070
0mt5_p_e	1.797	23.243	6.266	18.579	24.876	16.516	37.273	11.962	16.320
bm25_p_e	1.866	25.273	7.607	20.687	26.782	18.361	39.964	13.278	16.995
UoGTrMabSaNR	2.140	8.637	9.343	15.981	8.640	16.078	42.181	5.894	7.790
UogTRelvOnlyT2	2.223	11.152	11.037	19.964	11.686	19.738	22.193	6.136	9.654
0mt5_e	2.986	27.765	11.149	22.881	27.967	20.988	43.410	14.288	19.946
tmt5_e	3.400	28.021	10.255	35.293	35.676	32.521	38.150	19.642	23.982
ans_bm25_e	4.232	29.440	14.830	27.922	32.793	25.886	50.670	16.335	25.838

Limitations

- Fairness Criteria
 - Geography: incomplete location information
 - Gender: possibility of misgendering
 - Multidimensional fairness counts all attributes as independent
- Task Definition
 - Doesn't consider missing or deleted articles
 - There are more important protected attributes

Lessons Learned

- Increasing dimensions significantly increases complexity of computations, especially for determining target distributions
 - Final space was 11M dimensions!
 - Many points have no instances
 - Memory-intensive computations
- Performing well on multidimensions does not always equal good performance on individual dimensions.

Fair Ranking Track Plenary Session is on November 15th (9am - 12pm)

Thanks

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**



grant IIS 17-51278