We participated in the NTCIR-17 FairWeb-1 Task. We utilize several different methods in all 5 submitted runs including reranking, learning-to-rank, and search result diversification algorithms to deal with the group fairness problem in web search. The official results indicate that our approaches achieve the best results on all relevance and fairness metrics.

Our Methods

Run 1: Sparse Retrieval
We choose two classic sparse retrieval algorithms, BM25 and QLD. We respectively conduct the retrieval using the two algorithms for Q-queries, D-queries, and QD-queries.

Run 2: LightGBM
We employ MonoBERT and MonoT5 models to rerank all retrieved documents of the three types of queries in Run 1. For each query, we use the 12 sparse retrieval scores from Run 1 and the 6 neural reranker scores as the features to conduct learning-to-rank through a lightweight learning-to-rank model, LightGBM.

Run 3: Query Augmentation
We incorporate fairness information into the semantics of queries by simply adding the entity attribute information to the query text. For example, for movie topics that need to consider regional fairness we simply add a suffix "...from Africa/America/Antarctica..." to the query. We generate a ranking list via the MonoT5 reranker for each value of an attribute. We combine ranking lists of different values of the same attribute by random sampling and then utilize RRF to merge the results of different attributes.

Run 4 & Run 5: PM2 & xQuAD
We attempt two different ways of estimating attribute scores of each candidate document. One is to extract possible entities and obtain attribute information about them through web crawlers. Scores are calculated from the ratio of the attribute values. The other is simply approximating the document’s attribute distribution through the relative proportions of related term appeared in the document. Then we try two search result diversification algorithms, PM2 and xQuAD, to balance both relevance and fairness factors.

Our methods outperform others on all relevance metrics and fairness metrics.
Run 2, 3, and 4 are significantly better than Run 1 and Run 5 thanks to the powerful fine-grained reranker even in the zero-shot scenario.

If a method performs well in terms of relevance, it also has strong performance in terms of fairness. Relevance and fairness can be jointly optimized within a certain degree. Search results with higher relevance contain more relevant entities. These large amounts of randomly distributed related entities can facilitate further optimization towards fairness.

Conclusions
We participate in the NTCIR-17 FairWeb-1 task and submit 5 runs using various methods. We achieve first place in all metrics. Our results indicate that relevance and fairness are not in opposition to some degree and it is possible to achieve their joint optimization.