Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.4299

GDRF: An Innovative Graph-Based Rank Fusion Method for Enhancing Diversity in Image Metasearch

José Solenir Lima Figuerêdo [**University of Feira de Santana** | ***jslfigueredo@ecomp.uefs.br***]
Ana Lúcia Lima Marreiros Maia [**University of Feira de Santana** | ***anamarreiros@gmail.com***]
Rodrigo Tripodi Calumby [**University of Feira de Santana** | ***rtcalumby@uefs.br***]

Postgraduate Program in Computer Science, University of Feira de Santana, Av. Transnordestina, s/n, Novo Horizonte, Feira de Santana, BA, 44036-900, Brazil.

Received: 22 March 2024 • **Published:** 13 January 2025

Abstract Metasearch technique combines a set of ranked images retrieved with different search engines to build a unified ranking in order to improve relevance. For this purpose, rank aggregation methods have been widely used, which also can improve the result provide by ambiguous or underspecified queries through process named diversification. However, current aggregation methods assume that the input rankings are built only according to the relevance of the items, disregarding the inter-relationship between images in each ranking. Consequently, these methods tend to be inadequate for diversity-oriented retrieval. The aggregated ranking may not improve results, mainly when considered a diversity optimization. To address this problem, we propose a diversity-aware rank fusion method, which was validated in the context of diverse image metasearch. Our method was compared with several order-based and score-based aggregation methods. The experimental findings indicate that the proposed method significantly improves the overall diversity of metasearch results. This result demonstrates the potential of the proposed method and paves the way for further research to explore the development of new methods implementing new aware-diversity heuristics.

Keywords: Search Result Diversification, Ranking Aggregation, Metasearch, Diversity

1 Introduction

In information retrieval tasks, given a user information need, various rankings can be defined for the same data collection, e.g., considering different search engine configurations, feature representations, and ranking criteria. Hence, rank fusion methods can successfully combine multiple rankings into a unified result. Beyond that, given alternative search systems may present different results for the same user information need, the metasearch technique combines numerous search systems to build a final aggregated ranking. Since those independent results tend to complement each other, metasearch is expected to generate final rankings with improved relevance [Aslam and Montague, 2001].

Besides rank fusion, when dealing with complex queries, a technique called diversification is widely used to attenuate some ranking challenges [McDonald *et al.*, 2022]. Specifically, diversification has been demonstrated beneficial to maximizing intent coverage for broad, ambiguous, or underspecified queries, enhancing content-based recommendation systems, handling the redundancy among the retrieved items (e.g., near-duplicate images/documents), and improving user-system information transfer in interactive retrieval sessions [Calumby *et al.*, 2017].

Diversification of image search results is a hot research problem in multimedia. Many search engines, such as google image search, are fostering techniques that allow for providing the user with a diverse representation of search results, rather than providing redundant information, e.g., the same perspective of a monument or location [Ramírez-de-la-Rosa *et al.*, 2018]. Figure 2 illustrates an example of relevance-oriented and diversity-oriented results. A relevance-oriented retrieval (Figure 1-a) usually results in a potentially redundant set of images with low coverage of the query subtopics, limiting the user experience. In contrast, by including diversity as an additional retrieval criterion, redundancy can be reduced (Figure 1-b and c) as well as subtopic coverage can be improved. For instance, diverse results are composed of relevant images with different perspectives of an object of interest or variations in shape or color [Figuerêdo and Calumby, 2022]. Therefore, diversification aims at ensuring that at least some items (e.g., documents, images, products) related to different user intentions, interpretations or query aspects are placed at the top positions of the ranking [Yigit-Sert *et al.*, 2020].

The problem addressed by ranking aggregation methods regards the combination of a set of candidate relevance-oriented rankings so that the final combination includes more relevant items than any individual candidate list [Dwork *et al.*, 2001]. Those methods consider relevance as directly related to ranking positions, i.e., the higher the relevance of an item, the higher its ranking position. In contrast, in diversified results, ranking positions do not hold a strict direct relationship to the relevance of the items, i.e., an item in a subsequent lower position is not necessarily less relevant than the previous one, but may just contribute less to the ranking diversity (up to that position) considering the other items in higher positions.

Although fusion strategies have achieved significant gains in terms of relevance improvement, there are still some open

Figure 1. Query example. In (a) we have a possible result considering only the relevance of the items. In (b) and (c) we have the result also considering the visual diversity of the set of items in the result. From Figuerêdo and Calumby [2022].

challenges regarding diversified rankings. In general, the fusion methods proposed so far consider that the results to be merged were built only on the relevance of the objects, which is not always true. Therefore, by not considering the interrelationship between items, the diversity of the aggregated results can be under-optimized.

For the image metasearch task, some diversity improvements have been reported with the use of relevance and position-based rank aggregation methods [Figuerêdo and Calumby, 2019]. Nevertheless, to the best of our knowledge, no previous work has explicitly integrated the concept of diversity and inter-image positional relationship into the rank aggregation procedure itself.

Given the aforementioned challenges, we developed a Graph-based Diversity-aware Rank Fusion method (GDRF) that explicitly considers the concept of diversity in the rank fusion process. For this, we propose a diversity-aware preference graph structure, that stores the positional preference relations between each pair of images in a ranking. The preference graphs generated for each input ranking are combined to produce a new diversity-oriented ranking score. The proposed method suggests a template for ranking diversity representation and is also completely unsupervised. A detailed description of the method is presented in section 3. This study extends our previous work [Figuerêdo *et al*., 2023], improving problem formalization, presenting a more comprehensive analysis of related work, including new experiments with other aggregation algorithms, and providing real result examples for two queries using our method and the best baseline.

The remainder of this article is organized as follows: Section 2 presents the related works and Section 3 describes the proposed method. In turn, Section 4 presents the experimental process. The results and discussions are presented in Section 5. Finally, Section 6 brings the conclusions and future work.

2 Related Works

In general, aggregation algorithms fall into two main categories: score-based and order-based. In the former, the fusion procedure takes as input the ranking scores associated with each object in the original rankings. In the latter, order-based algorithms consider only the position of the items in the ranking to perform the fusion process. Some of the most widespread score-based methods are: CombMAX, CombMIN, CombSUM, CombANZ, CombMNZ). In turn, BordaCount, Median Rank Aggregation (MRA) and Reciprocal Rank Fusion (RRF) are popular order-based methods [Vargas Muñoz *et al*., 2015]. However, although these algorithms have been used in many applications, they do not consider diversification explicitly. Based on the premise that the fusion process itself can ensure wide coverage of relevant items, some studies have been developed. Two of the main works were developed by [Liang *et al*., 2014; Ozdemiray and Altingovde, 2015; Xu and Wu, 2017; Kaur *et al*., 2018].

In Liang *et al*. [2014] diversification is performed in three stages. Initially, the fusion is executed using the Comb-

SUM and CombMNZ methods. Then an inference of latent subtopics is made. Finally, the result generated by the two previous steps is submitted to the diversification process using the method PM-2. PM-2 is an explicit diversification method that determines, iteratively, for each position in the ranked result list, the topic that best maintains the overall proportionality. It then selects the best document on this topic for this position [Dang and Croft, 2012]. On the other hand, Xu and Wu [2017], instead of merging already diversified results, chose a direct diversification approach. It also includes three stages: i) Generation of results using search algorithms based only on relevance; ii) Fusion of these results using any algorithm, such as CombMNZ; and iii) Application of an explicit diversification method such as PM-2.

To address the problem of diversification, Ozdemiray and Altingovde [2015] proposed approaches based on score and order aggregation methods. Regarding score-based methods, the authors adapted CombSUM and CombMNZ. On the other hand, for order-based, they employed simple voting and Borda voting, as well as Markov chain-based approaches [Dwork *et al.*, 2001]. In summary, using these aggregation methods, the authors proposed optimizations for the relevance score normalization and novelty estimation components of xQuAD, which is a technique for explicit result diversification [Dang and Croft, 2012]. Among the proposed approaches, the method namely as mix_CombSum, that is essentially a variant of xQuAD, stands out. However, xQuAD uses a greedy algorithm to select documents one by one, while mix_CombSum applies a linear weighted summation of the scores of all documents involved. Then, documents are re-ranked by their total scores. This new strategy is cheaper in terms of the computational cost [Ozdemiray and Altingovde, 2015].

Another study explored the ant colony optimization, in order to enhance the aggregation of rankings [Kaur *et al.*, 2018]. This metaheuristic approach is utilized to optimize Spearman's footrule and Kendall's tau distance measures, which are used to compare ranking methods. The authors evaluated five aggregation methods, including Borda count, Markov chain, scaled footrule, PageRank and mean-by-variance. In general, in comparison to other optimization approaches, such as genetic algorithms, the developed approach demonstrated superior effectiveness. Among the evaluated aggregation methods, the Borda count achieved the best results.

While previous work has focused on the analysis of diversification through fusion methods in the context of web page retrieval, the investigation of such methods in other multimedia scenarios (e.g., image or video retrieval) is still incipient. Furthermore, the applied fusion methods do not consider that the retrieved results may have come from systems that already consider ranking diversification. This work aims at filling this gap by explicitly considering diversified rankings as input to a metasearch approach.

3 Proposed method

The GDRF has three main steps. The first step corresponds to the representation of input rankings as preference graphs, followed by the attribution of edge weights. The preference graph is a position-guided structure with a directed edge between every pair of nodes in the ranking. Each edge characterizes the preference relationship between the connected items. Therefore, rank diversity is captured as multiple preference links between images. Figure 2 illustrates this representation process. In our context, each node corresponds to an image present in the considered ranking. For example, if the ranking contains an image (Img1) in a higher position than another image (Img2), the graph will contain a preference edge directed from node Img2 to node Img1.

Each edge between a pair of images has a weight (W), as illustrated in Figure 1. The weight assignment can be performed with different strategies. In our method, the attribution of weights follows Eq. 1. Considering any pair of images *(x,y)*, $\boldsymbol{\mu}$ corresponds to the average position occupied by them in the input rankings. The position is designated in descending order. For instance, if the ranking has 50 images, the image at the first position would have 50 as its position score, while the last image scores 1. With this assignment, the images that occupy the first positions are considered more relevant to the query than others. Considering a pair *(x,y)*, $\boldsymbol{\alpha}$ would be the number of nodes (images) that are preferable to y, that is, they are above y in the ranking. In turn, $\boldsymbol{\beta}$ represents the number of nodes between the higher node *(y)* and the lower node *(x)*, which indicates how many pairs of images are preferable to the pair being evaluated *(x,y)*.

$$W = 1 - \frac{1}{1 + \frac{\mu}{\alpha+\beta}} \tag{1}$$

Equation 1 aims at capturing the diversity existing in the rankings. Therefore, it favors pairs of diverse images, considering that the base rankings, in addition to being generated considering the relevance, also used diversity as a simultaneous ranking criterion. In the ranking illustrated in Figure 1, assuming Img1 as a reference, Img2 is considered the most relevant to the query while is also more diverse than the others. Otherwise, it wouldn't be the second on the list. In turn, Img3, while possibly more relevant to the query than Img2, contributes less to diversity maximization than Img2. Therefore, the score that aims to capture the degree of diversity between images is greater for the pair Img2-Img1 than for Img3-Img1.

In the second step of the GDRF, an aggregated graph ($\boldsymbol{AG}$) is constructed considering the individual graphs formed in the previous step. The $\boldsymbol{AG}$ relies on the combination of the preference relations obtained from the individual graphs. The resulting graph contains as vertices all the images that appear in at least one of the input rankings. The combination of the weights of an edge *(x,y)* is calculated according to Equation 2.

$$AG_{xy} = \sum G_{xyk} \tag{2}$$

In equation 2, the summation runs through all the individual graphs that provide preference relations for the *(x,y)* pair. $\boldsymbol{G_{xyk}}$ denotes the preference edge weight from x to y in the preference graph corresponding to input ranking $\boldsymbol{k}$. Then, step 3 begins, which corresponds to obtaining the final ranking. The induction of the final ranking is carried out from the combined preference relations stored in the $\boldsymbol{AG}$. For this, as

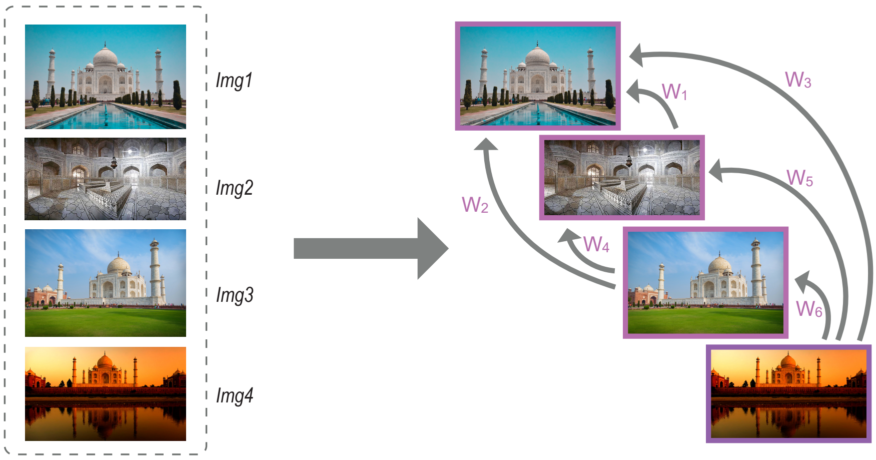


Figure 2. Converting a diversified input ranking to a preference graph.

different approaches could be followed, we report preliminary experiments with the best performance occurring by sequentially selecting the main nodes, i.e., the ones with the highest accumulated preference weights.

4 Experimental setup

For the experimental evaluation of the GDRF, the collection provided by the Information Fusion for Social Image Retrieval & Diversification Task [Ramírez-de-la-Rosa *et al.*, 2018] was used. This collection includes results from many image search systems proposed and evaluated between 2013 and 2016 in the MediaEval Retrieving Social Images tasks. There are ranked results for numerous queries. In addition, it includes relevant and diverse results with different levels of quality. The dataset is organized into development, validation, and test sets. In this work, we consider only the development set, given the unsupervised nature of the GDRF. Specifically, we pooled *devset1* (39 candidate rankings for 346 queries) and *devset2* (56 candidate rankings for 60 queries). Thus, all analyses were performed on this combined set.

Precision and Cluster-Recall measures were used for effectiveness assessment. Precision represents the quality of the ranking in terms of relevance. The Cluster-Recall measure computes the percentage of conceptual clusters that were represented in a diversified result. For effectiveness analysis, these measures were computed up to the 50^{th} position of the ranking. As baselines, we utilized both score-based and order-based methods. In the former, we used CombMAX, CombMIN, CombSUM, CombANZ and CombMNZ. In the latter, Borda Count, Median Rank Aggregation (MRA), and Reciprocal Rank Fusion (RRF) were used. For the strict comparison of the effectiveness results, the GDRF was compared to the baselines using Wilcoxon's Signed Rank Test in order to assess the statistical significance of the results.

5 Results and Discussions

Table 1 shows the effectiveness of the proposed method and baselines. The highest values are highlighted in boldface. Considering the relevance of the final rankings, the baseline RRF algorithm achieved the best Precision@N results for initial ranking positions ($P@5$, $P@10$ and $P@20$), whereas CombMNZ was superior in deeper positions ($P@30$, $P@40$ and $P@50$). However, considering diversification as the main objective in this study, the Cluster-Recall measure plays an important role. The proposed method achieved numerically superior performance over all considered baselines, except for RRF and CombSUM for $N = 5$ and $N = 10$, respectively.

Considering the algorithms from the Comb family, the CombMNZ and CombSUM methods generally yield the best results, as indicated in He and Wu [2008] and Ozdemiray and Altingovde [2015]. This was also verified in our experimental results, with CombMNZ achieving the higher effectiveness in terms of $P@N$, and CombSUM in relation to $CR@N$. Additionally, although in some scenarios the Borda count outperformed other strategies [Kaur *et al.*, 2018], in our study it achieved the worst result among all baselines.

In Table 2 we present the results of Wilcoxon's Signed Rank Test. Green cells represent statistical superiority, white cells mean equivalence, while pink cells represent inferiority against the baseline. Regarding Precision@N, the GDRF was statistically inferior to CombMNZ, RRF, CombSUM ($P@40$ and $P@50$), and MRA ($P@30$ and $P@50$). On the other hand, considering diversity (CR) the GDRF was statistically superior at multiple ranking levels. As the relevance-diversity trade-off is a central and long-lasting challenge in this task, the results reported here suggest that the GDRF is preferable to the baselines, for scenarios in which diversity maximization is a key factor, while further investigations should be performed on how to better optimize trade-off towards simultaneously better relevance results.

Nevertheless, although in the best scenario, the same method should provide the best results for both objectives, for some applications diversity is of great importance. For example, in an e-commerce system, strategically, it may be better to present diverse results with different product models, different shapes, variety of colors, among other characteristics. In such a scenario, a user would be exposed to a wider set of options, even if a few cases of non-relevant items appear in the search result. In addition, by improving diversity, there is an indirect minimization of the redundancy of

Table 1. Results for the GDRF and baselines. Top values are highlighted in boldface.

						Devset1 + Devset2						
Method	P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50
Borda Count	0.5915	0.5908	0.5995	0.6052	0.6049	0.5978	0.1735	0.2859	0.4352	0.5552	0.6401	0.7065
CombANZ	0.7149	0.7244	0.7219	0.7138	0.7040	0.6848	0.2263	0.3701	0.5527	0.6826	0.7678	0.8178
CombMAX	0.7577	0.7495	0.7515	0.7392	0.7255	0.7065	0.2363	0.3927	0.5789	0.7026	0.7863	0.8357
CombMIN	0.6264	0.6383	0.6486	0.6551	0.6524	0.6415	0.1875	0.3293	0.5059	0.6388	0.7221	0.7808
CombMNZ	0.8527	0.8346	0.8144	**0.7959**	**0.7742**	**0.7460**	0.2593	0.4139	0.6000	0.7155	0.7966	0.8459
CombSUM	0.8343	0.8286	0.8067	0.7916	0.7693	0.7414	0.2538	**0.4176**	0.5982	0.7160	0.7982	0.8486
MRA	0.8318	0.8271	0.8082	0.7927	0.7691	0.7418	0.2357	0.3890	0.5854	0.7030	0.7881	0.8413
RRF	**0.8567**	**0.8391**	**0.8154**	0.7949	0.7716	0.7455	**0.2618**	0.4115	0.5989	0.7161	0.7929	0.8416
GDRF	0.8308	0.8239	0.8056	0.7879	0.7666	0.7384	0.2580	0.4167	**0.6026**	**0.7258**	**0.7998**	**0.8550**

Table 2. Wilcoxon's Signed Rank Test. Green cells represent statistical superiority, white cells equivalence, while pink represent inferiority.

						Devset1 + Devset2						
Pair	P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50
GDRF vs Borda Count	green	green	green	green	green	green	green	green	green	green	green	green
GDRF vs CombANZ	green	green	green	green	green	green	green	green	green	green	green	green
GDRF vs CombMAX	green	green	green	green	green	green	green	green	green	green	green	green
GDRF vs CombMIN	green	green	green	green	green	green	green	green	green	green	green	green
GDRF vs CombMNZ	pink	pink	pink	pink	pink	pink	white	white	white	green	white	green
GDRF vs CombSUM	white	white	white	white	pink	pink	white	white	white	white	white	green
GDRF vs MRA	white	white	white	pink	white	pink	green	green	green	green	white	green
GDRF vs RRF	pink	pink	pink	pink	pink	pink	white	white	white	green	green	green

the results, which is important for a better user experience.

Our experimental results demonstrated that the GDRF was able to capture intrinsic quality information, thereby enhancing the discovery of implicit query subtopics. Figure 3 presents real retrieval results comparing the ranking from the best baseline (RRF) and the diverse ranking generated with the GDRF. In Figure 3a and 3b we present examples of significant improvements in terms of relevance and diversity, respectively.

As evident in Figure 3a, the proposed method retrieved more relevant images than the RRF, with the advantage of representing images from two different semantic groups (A and B). While both methods retrieved several irrelevant images, the proposed method achieved greater diversification and relevance, attaining a gain of 200% for $P@10$.

Referring to Figure 3b, although all images retrieved by RRF are relevant, many belong to the same semantic group. For instance, out of 10 retrieved images, 4 belong to group D, 3 to group C, 1 to group A, B and E, covering images from 5 different groups. In contrast, the proposed method retrieved images from 9 distinct groups, representing a gain of 80%. This outcome demonstrates that the GDRF was capable of capturing intrinsic relationships between images, allowing greater diversification, even though the input rankings were already the result of a previous diversification process.

6 Conclusion

This work introduces a novel graph-based diversity-aware rank fusion method validated in the context of metasearch. The method was compared using both score-based and order-based approaches. In terms of the relevance of the metasearch result, the proposed method achieved competitive results, but not enough to outperform the best baseline. On the other hand, the experimental findings indicate that the proposed method allowed superior results in terms of diversity at different ranking levels compared to the baselines. While alternatives should be investigated to more effectively balance the relevance-diversity trade-off, this proposal provides a significant contribution to the field by explicitly considering the diversity concept integrated into a rank aggregation strategy.

Future work should investigate, e.g., specific weighting procedures for the input rankings, given that the metasearch is performed over systems with different quality. Additional, novel strategies for assigning weights to preference relationships and other ranking-to-graph and graph-to-ranking transformations could also be proposed.

Acknowledgements

We thank to the University of Feira de Santana for providing the necessary infrastructure to conduct the experiments.

Funding

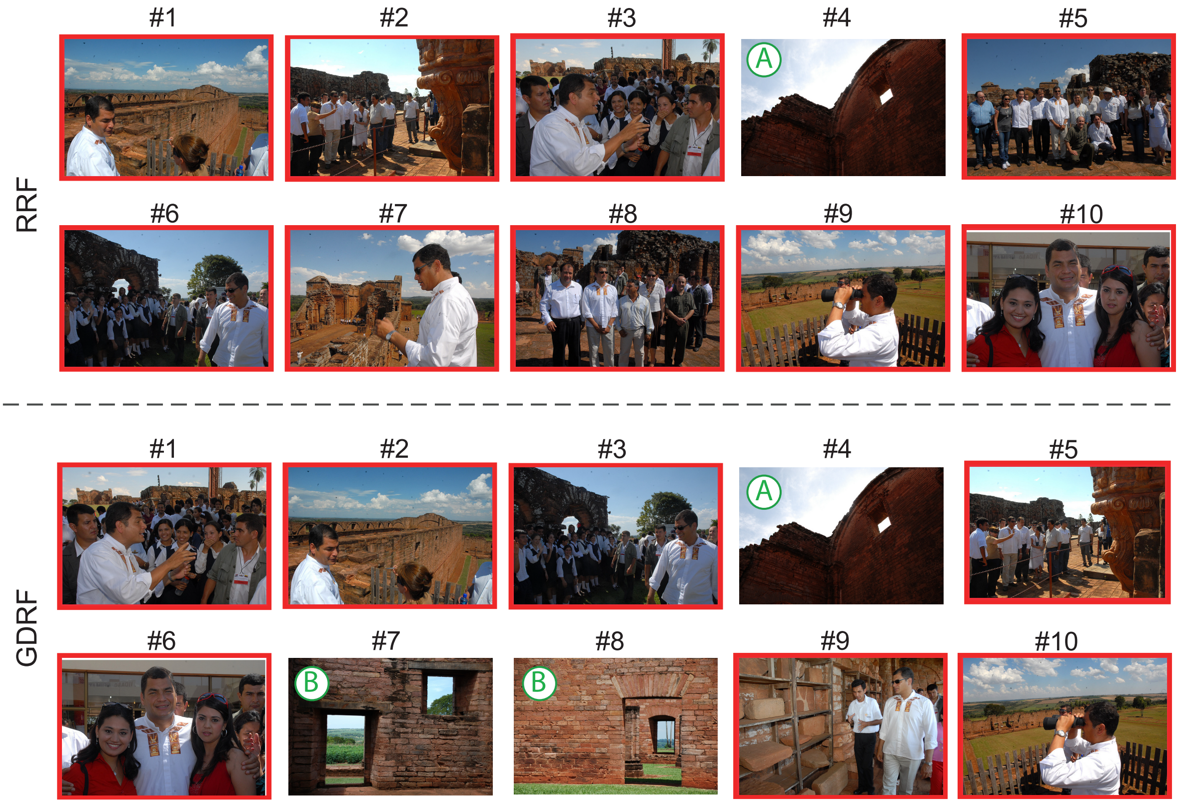
This work was partially supported by UEFS AUXPPG 2023 and CAPES PROAP 2023 grants.

Authors' Contributions

José Soenir Lima Figuerêdo contributed to the conceptualization, methodology, performed the experiments, results analysis, writing (original draft), data curation, investigation and writing (review & editing). **Ana Lúcia Lima Marreiros Maia** contributed to writing (review & editing) and supervision. Finally, **Rodrigo Tripodi Calumby** contributed to the conceptualization, writing (review & editing), results analysis, investigation and supervision.

Competing interests

The authors declare that they have no known competing financial



(a) Top-10 results for the query "La Santisima Trinidad de Parana Paraguay". At the top we have the result of the best baseline ($P@10 = 0.1000$ and $CR@10 = 0.1111$). At the bottom the result of our method ($P@10 = 0.3000$ and $CR@10 = 0.2222$).

(b) Top-10 results for the query "Mission San Carlos Borromeo de Carmelo". At the top we have the result of the baseline ($P@10 = 1.000$ and $CR@10 = 0.2500$). At the bottom the result of our method ($P@10 = 1.0000$ and $CR@10 = 0.4500$).

Figure 3. Real result examples for two queries in the execution with the RRF and the proposed method. In (a), the query with the greatest gain in relation to the RRF in terms of P@10. In (b), the query with the greatest gain in relation to the RRF in terms of CR@10. The image clusters from ground-truth (visual subtopics) are represented by letters. Non relevant images are highlighted in red.

interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and materials

The datasets used in this study are available in Ramírez-de-la-Rosa *et al*. [2018].

References

Aslam, J. A. and Montague, M. (2001). Models for metasearch. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 276–284, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/383952.384007.

Calumby, R. T., Gonçalves, M. A., and da Silva Torres, R. (2017). Diversity-based interactive learning meets multimodality. *Neurocomputing*, 259:159–175. Multimodal Media Data Understanding and Analytics. DOI: https://doi.org/10.1016/j.neucom.2016.08.129.

Dang, V. and Croft, W. B. (2012). Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, page 65–74, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/2348283.2348296.

Dwork, C., Kumar, R., Naor, M., and Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web*, WWW '01, page 613–622, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/371920.372165.

Figuerêdo, J. and Calumby, R. (2019). Unsupervised rank fusion for diverse image metasearch. In *Anais do XXXIV Simpósio Brasileiro de Banco de Dados*, pages 265–270, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbbd.2019.8834.

Figuerêdo, J. and Calumby, R. (2022). Unsupervised query-adaptive implicit subtopic discovery for diverse image retrieval based on intrinsic cluster quality. *Multimedia Tools and Applications*, 81:42991–43011. DOI: 10.1007/s11042-022-13050-4.

Figuerêdo, J., Maia, A., and Calumby, R. (2023). A novel graph-based diversity-aware rank fusion method applied to image metasearch. In *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 324–329, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbbd.2023.233400.

He, D. and Wu, D. (2008). Toward a robust data fusion for document retrieval. In *2008 International Conference on Natural Language Processing and Knowledge Engineering*, pages 1–8. DOI: 10.1109/NLPKE.2008.4906754.

Kaur, P., Singh, M., Josan, G., and et al. (2018). Rank aggregation using ant colony approach for metasearch. *Soft Computing*, 22:4477–4492. DOI: 10.1007/s00500-017-2723-3.

Liang, S., Ren, Z., and de Rijke, M. (2014). Fusion helps diversification. SIGIR '14, page 303–312, New York, NY, USA. ACM. DOI: 10.1145/2600428.2609561.

McDonald, G., Macdonald, C., and Ounis, I. (2022). Search results diversification for effective fair ranking in academic search. *Information Retrieval Journal*, 25(1):1–26. DOI: 10.1007/s10791-021-09399-z.

Ozdemiray, A. M. and Altingovde, I. S. (2015). Explicit search result diversification using score and rank aggregation methods. *Journal of the Association for Information Science and Technology*, 66(6):1212–1228. DOI: https://doi.org/10.1002/asi.23259.

Ramírez-de-la-Rosa, G. *et al*. (2018). Overview of the multimedia information processing for personality & social networks analysis contest. In *ICPR'18, Beijing, China, August 20-24*, pages 127–139. DOI: 10.1007/978-3-030-05792-3_12.

Vargas Muñoz, J. A., da Silva Torres, R., and Gonçalves, M. A. (2015). A soft computing approach for learning to aggregate rankings. page 83–92, New York, NY, USA. ACM. DOI: 10.1145/2806416.2806478.

Xu, C. and Wu, S. (2017). The early fusion strategy for search result diversification. ACM TUR-C '17, New York, NY, USA. ACM. DOI: 10.1145/3063955.3064803.

Yigit-Sert, S., Altingovde, I. S., Macdonald, C., Ounis, I., and Özgür Ulusoy (2020). Supervised approaches for explicit search result diversification. *Information Processing & Management*, 57(6):102356. DOI: https://doi.org/10.1016/j.ipm.2020.102356.