

MMR Clustering Rewrite vs Query Rewrite Analysis

Method	Strategy	R@1	R@3	R@5	R@10	nDGC@1	nDGC@3	nDGC@5	nDGC@10
BM25 (PyTerrier)	Query Rewrite	0.090	0.189	0.250	0.349	0.211	0.200	0.223	0.264
	MMR Cluster (k5)	0.072	0.165	0.219	0.298	0.165	0.171	0.192	0.225
	MMR Cluster (k10)	0.077	0.175	0.232	0.320	0.179	0.182	0.205	0.242
	MMR Cluster (k15)	0.087	0.180	0.239	0.311	0.200	0.189	0.213	0.243
	MMR Cluster (k20)	0.082	0.180	0.234	0.317	0.188	0.187	0.208	0.244
BGE-base 1.5	Query Rewrite	0.153	0.305	0.381	0.498	0.358	0.326	0.354	0.404
	MMR Cluster (k5)	0.130	0.270	0.342	0.448	0.300	0.284	0.313	0.358
	MMR Cluster (k10)	0.140	0.282	0.357	0.464	0.320	0.297	0.326	0.371
	MMR Cluster (k15)	0.140	0.277	0.355	0.460	0.326	0.297	0.326	0.372
	MMR Cluster (k20)	0.142	0.284	0.354	0.459	0.326	0.303	0.329	0.373
Elser	Query Rewrite	0.187	0.372	0.476	0.608	0.429	0.399	0.438	0.495
	MMR Cluster (k5)	0.171	0.347	0.448	0.559	0.384	0.368	0.407	0.455
	MMR Cluster (k10)	0.178	0.358	0.464	0.577	0.407	0.381	0.422	0.470
	MMR Cluster (k15)	0.166	0.352	0.455	0.577	0.388	0.373	0.412	0.464
	MMR Cluster (k20)	0.166	0.353	0.451	0.566	0.383	0.373	0.409	0.460

Analysis

- **Consistent Underperformance:** The MMR Cluster strategy consistently underperforms the standard Query Rewrite baseline across all retrieval methods (Sparse, Dense, and Learned Sparse) and almost all metrics.
- **Noise Introduction:** Applying MMR clustering likely filters out relevant information captured by the standard Query Rewrite.
- **Method-Specific Trends:**
 - **BM25:** Performance stagnates with increased k .
 - **BGE-base 1.5:** Slight improvement with higher k , but remains significantly below baseline (~8% drop in nDCG).
 - **Elser:** Performance peaks at $k10$ and degrades with higher k , suggesting too much diversity hurts this learned sparse retriever.

Lambda Parameter Experiments (k=10)

Method	Strategy	R@1	R@3	R@5	R@10	nDGC@1	nDGC@3	nDGC@5	nDGC@10
BM25 (PyTerrier)	Query Rewrite	0.090	0.189	0.250	0.349	0.211	0.200	0.223	0.264
	MMR Cluster ($\lambda = 0.3$, k10)	0.076	0.174	0.226	0.298	0.174	0.178	0.199	0.229
	MMR Cluster ($\lambda = 0.5$, k10)	0.079	0.175	0.230	0.313	0.182	0.182	0.204	0.239
	MMR Cluster ($\lambda = 0.85$, k10)	0.079	0.169	0.225	0.309	0.187	0.179	0.201	0.236
	MMR Cluster ($\lambda = 0.9$, k10)	0.077	0.172	0.225	0.311	0.179	0.178	0.199	0.235
BGE-base 1.5	Query Rewrite	0.153	0.305	0.381	0.498	0.358	0.326	0.354	0.404
	MMR Cluster ($\lambda = 0.3$, k10)	0.139	0.280	0.362	0.462	0.324	0.300	0.331	0.374
	MMR Cluster ($\lambda = 0.5$, k10)	0.140	0.274	0.349	0.455	0.320	0.295	0.323	0.369
	MMR Cluster ($\lambda = 0.85$, k10)	0.141	0.283	0.358	0.469	0.330	0.303	0.330	0.378
	MMR Cluster ($\lambda = 0.9$, k10)	0.143	0.287	0.361	0.468	0.333	0.307	0.334	0.379
Elser	Query Rewrite	0.187	0.372	0.476	0.608	0.429	0.399	0.438	0.495
	MMR Cluster ($\lambda = 0.3$, k10)	0.175	0.362	0.457	0.585	0.393	0.383	0.417	0.472
	MMR Cluster ($\lambda = 0.5$, k10)	0.169	0.354	0.455	0.567	0.389	0.374	0.413	0.461
	MMR Cluster ($\lambda = 0.85$, k10)	0.177	0.356	0.451	0.579	0.408	0.380	0.415	0.470
	MMR Cluster ($\lambda = 0.9$, k10)	0.174	0.349	0.455	0.566	0.400	0.375	0.415	0.464

Lambda Experiment Analysis

- **Best Lambda Values:**
 - **BM25:** $\lambda = 0.5$ performs best overall, though still below Query Rewrite baseline
 - **BGE:** $\lambda = 0.9$ performs best
 - **ELSER:** $\lambda = 0.85$ performs best
- **Lambda Impact:** Higher lambda values (0.85, 0.9) generally improve performance for dense (BGE) and learned sparse (ELSER) retrievers, suggesting that prioritizing relevance over diversity helps these methods.
- **Still Underperforming:** Even with optimal lambda values, MMR Cluster strategy still underperforms Query Rewrite baseline across all methods.