

Selección de datasets para evaluación de métodos QPP

Evaluación comparativa de métodos de Query Performance Prediction (QPP) para búsquedas Ad-hoc utilizando métricas de correlación

La siguiente tabla describe los datasets seleccionados junto con su relevancia para el proyecto y su respectiva justificación basada en información encontrada en línea. Estos fueron seleccionados desde `ir_datasets` cumpliendo los siguientes criterios: disponibilidad, diversidad, relevancia para el proyecto de métodos QPP sin IA y en búsquedas ad-hoc.

Dataset	Descripción	Relevancia para el Proyecto	Justificación
BEIR-TREC-COVID	Es un dataset parte de BEIR, centrado en la recuperación de información sobre COVID-19, basado en consultas médicas y científicas.	Permite evaluar la capacidad de los métodos QPP para manejar consultas en un dominio médico, ofreciendo un contexto especializado y desafiante.	Thakur, N., et al. (2021). "BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models." <i>arXiv preprint arXiv:2104.08663</i> .
Cranfield	Conjunto de datos clásico que incluye resúmenes de artículos científicos con consultas predeterminadas, es pequeño, ideal para pruebas controladas.	Su simplicidad y tamaño reducido permiten realizar pruebas rápidas y controladas, facilitando la comparación directa de métodos QPP.	Cleverdon, C. W. (1967). "The Cranfield tests on index language devices." <i>Aslib Proceedings</i> , 19(6), 173-194.
MS MARCO (Passage)	Contiene millones de pasajes derivados de búsquedas reales en motores, con relevancia asignada por evaluadores humanos.	Ofrece un entorno diverso con consultas reales, crucial para probar la capacidad de los métodos QPP en un contexto de búsqueda cotidiana y variada.	Bajaj, P., et al. (2016). "MS MARCO: A Human Generated Machine Reading Comprehension Dataset." <i>InCoCo@NIPS</i> .

Antique/Test	Dataset enfocado en la recuperación de respuestas basadas en opiniones.	Es útil para evaluar cómo los métodos QPP manejan consultas subjetivas, lo que añade una capa de complejidad en la predicción del rendimiento de la consulta.	Hasibi, F., et al. (2018). "ANTIQUE: A Non-factoid Question Answering Benchmark." <i>arXiv preprint arXiv:1811.03502</i> .
CAR (Complex Answer Retrieval) versión Fold 0	Subconjunto manejable del TREC CAR, enfocado en consultas complejas derivadas de la estructura de la Wikipedia.	Permite evaluar los métodos QPP en la recuperación de información compleja, manteniendo la diversidad en tipos de consultas sin abarcar el dataset completo.	Dietz, L., et al. (2017). "TREC Complex Answer Retrieval Overview." <i>TREC</i> .