

APPENDIX C“README file”

README - Evaluation Toolkit SKG-LLM-RAG

=====

This supplementary package provides evaluation code and datasets for the paper:

"Leveraging Survey Metadata for LLM Reasoning via Knowledge Graphs"

The code implements four evaluation tasks with extended functionality:

1. Semantic Retrieval
2. Link Prediction
3. Topic Discovery

Files included

- eval_tasks_plus_api.py

Python module implementing the Plus evaluators with sentence-embedding support (auto-fallback to TF-IDF), per-query rankings, and bootstrap CIs.

- Example dataset files (referenced in examples below):

- * b17017_schema_kg_prov.jsonl (knowledge graph with provenance)
- * b17017_skg_nodes.csv (SKG node table)
- * b17017_queries.csv (semantic retrieval query set)
- * b17017_bm25_corpus.jsonl (BM25 flat corpus)
- * b17017_text_encoder_corpus.jsonl (text-encoder flat corpus)
- * b17017_llm_rag_flat_corpus.jsonl (LLM-RAG flat corpus)
- * qa_outputs_demo.jsonl (demo QA outputs with citations)