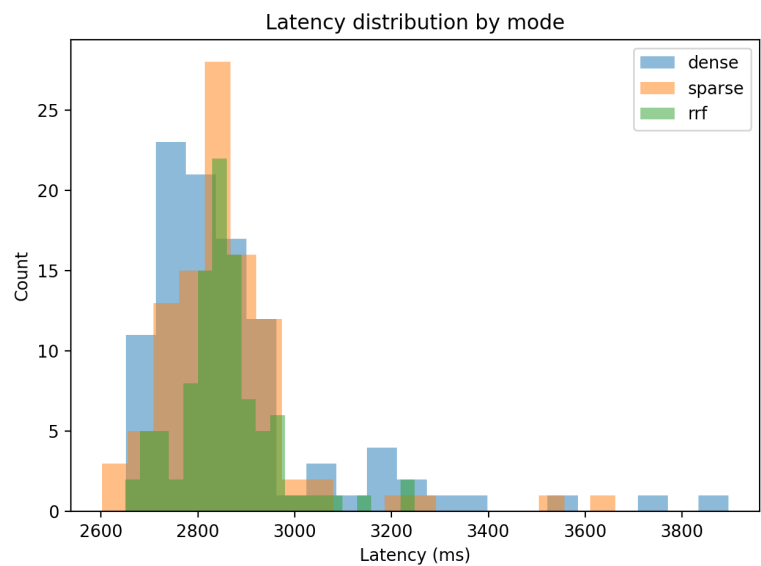
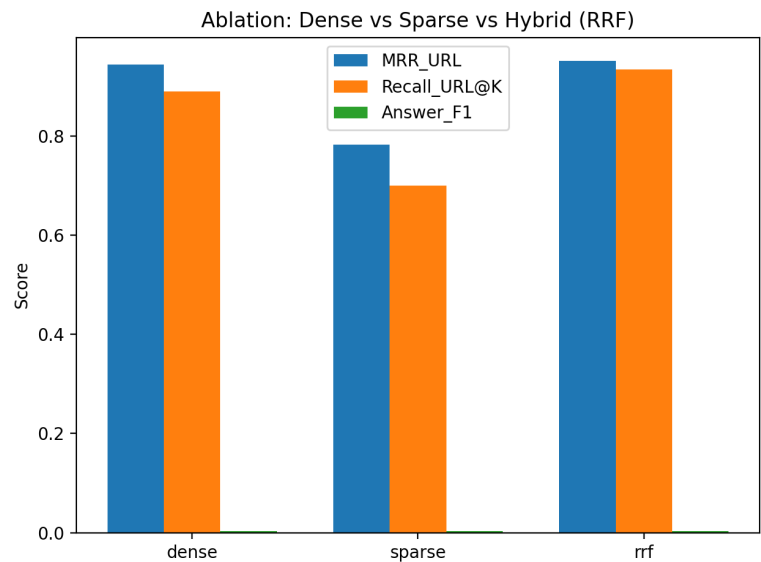


Hybrid RAG Evaluation Report

Metrics: URL-level MRR + URL Recall@K + Answer F1 + Ablation + Error analysis

Overall Summary (by retrieval mode)

- dense: MRR_URL=0.9450, Recall=Recall_URL@4=0.8900, F1=0.0036, Lat(ms)=2879.0
- sparse: MRR_URL=0.7833, Recall=Recall_URL@4=0.7000, F1=0.0036, Lat(ms)=2855.6
- rrf: MRR_URL=0.9517, Recall=Recall_URL@4=0.9350, F1=0.0036, Lat(ms)=2855.7



Error Analysis (Examples)

ID=q001 | Cat=factual | Fail=generation_exception | Rank=1 | F1=0.000

Q: What is Bantu languages?

GT: the total number of bantu languages is estimated at between 440 and 680 distinct languages, depending on the definition of

Pred: I don't know

ID=q002 | Cat=factual | Fail=generation_exception | Rank=2 | F1=0.000

Q: What is Computer architecture?

GT: it can sometimes be a high - level description that ignores details of the implementation.

Pred: I don't know

ID=q003 | Cat=factual | Fail=generation_exception | Rank=1 | F1=0.000

Q: What is Bryant Wright?

GT: he earned a bachelor of arts in english from the university of south carolina in 1974 and a master of divinity from the southern

Pred: I don't know

ID=q004 | Cat=factual | Fail=generation_exception | Rank=3 | F1=0.154

Q: What is Ion Caraion?

GT: i - iii, 1974 - 1976), canadian (1978) and american (1979).

Pred: I don't know

ID=q005 | Cat=factual | Fail=generation_exception | Rank=1 | F1=0.000

Q: What is David Hajjar?

GT: from 1978 to 1980, he completed postdoctoral training in pathology at weill cornell medicine / cornell university in new york

Pred: I don't know

ID=q006 | Cat=factual | Fail=generation_exception | Rank=1 | F1=0.000

Q: What is Ariel (moon)?

GT: = = orbit = = among uranus's five major moons, ariel is the second closest to the planet, orbiting at the distance of about 19

Pred: I don't know