??RAG???????????

> ???? Step6 ???????????? outputs/ ? docs/TABLE_*.md?

??

?????FinDER??????????????????????????????????+??RAG ???????LLM API??full dev Recall@10: 0.3246?0.3772??????????????????MRR????????????????????????????????

???????????????????????????????????

---

1 ??

?????FinDER?????????????????????/??+?????????? ?????????????????

?????????????RAG?Retrieval-Augmented Generation??????????????????????????????LLM API??

???????
??????????
- **??????????**??? gap ????????????????????????????????????
???????????????/????????????????????????????????
- **????????????**??? full dev / complex dev / numeric dev ??????????????????????????

---

2 ????

???????????????QA??

???????RAG????RAG?????????????????/??/?????????????QA??????

??/??????????????????????????????????????gap???????????????????

???RAG?Agentic RAG?????????????????????????????????????

---

3 ??

??????????????**?????**?**??????**?**???????**?**??????**?

?1 ?????????

Query ? Query Understanding ? Step-1 Retrieval ? Gap Detector
? (gap & gate) Refined Query Retrieval ? Merge & Rank
? Calculator (optional) ? Template Answer

3.1 ????
??gap?????????????

3.2 ??????
????????????top-k??????gap???????????????
Gap Detector??????????????????
- **Gate**?? gap_conf < min_gap_conf ?????????? query ???

Merge Strategy?maxscore ? step1_first ??????????
- **Stop Criteria**??? max_steps ??????????

????????
???? T?max_steps?
- ???? top-k?top_k_each_step?

???? top-k_f?top_k_final?
- ???? $ au$?min_gap_conf?

Step6 ????????T=2?top_k_each_step=10?merge=maxscore?novelty_threshold=0.0?stop_no_new_steps=1?min_gap_conf=0.3?

3.3 ???????
??????????????????????
YoY/??/??/????????????????????????????????Step6 ?????min_conf=0.2?allow_task_types=[]???????
?????

3.4 ????
?????? status=ok ????????????????????????????????? fallback ???

---

4 ????

4.1 ??????
?? FinDER ????? train/dev/test ??????????? qid?query?answer ? evidences?

4.2 ????
complex_dev????????????????+?????
- **numeric_dev**?????????/???/??/??/??????

4.3 ???????
?????Recall@k?MRR@k?QA???EM/F1??????Numeric-EM?RelErr?Coverage?

?????

$$\mathrm{Recall@k}=
rac{1}{N}\sum_{i=1}^N \mathbb{I}(ext{gold}_i \in ext{Top-}k) ag{1}$$

$$\mathrm{MRR@k}=
rac{1}{N}\sum_{i=1}^N
rac{1}{\mathrm{rank}_i} ag{2}$$

$$\mathrm{RelErr}=
rac{|\hat{y}-y|}{\max(|y|,\epsilon)} ag{3}$$

---

5 ???????

5.1 ????
?1 ????full dev ? complex dev?

label	run_id	full_r10	full_mrr10	complex_r10	complex_mrr10
pre_ft_baseline	20260130_234540_ae7cdf_m01	0.3246	0.2030	0.3457	0.2330
post_ft_baseline	20260130_234540_ae7cdf_m02	0.3772	0.2601	0.3909	0.2960
post_ft_multistep_best	20260130_234540_ae7cdf_m03	0.3772	0.2601	0.3909	0.2961
post_ft_baseline_calc_best	20260130_234540_ae7cdf_m04	0.3772	0.2601	0.3909	0.2960
post_ft_multistep_calc_best	20260130_234540_ae7cdf_m05	0.3772	0.2601	0.3909	0.2961
post_ft_multistep_T1_calc_best	20260130_234540_ae7cdf_m06	0.3772	0.2601	0.3909	0.2960

?????complex dev??

baseline(post-ft) vs best multistep?Recall@10 0.3909465 ? 0.3909465?MRR@10 0.2960138 ? 0.2960873

-
- 5.2 ??????
- **?2 ??????numeric dev?**
-

label	run_id	num_em	num_rel	num_cov
pre_ft_baseline	20260130_234540_ae7cdf_m01	0.3791	2874.5248	0.6202
post_ft_baseline	20260130_234540_ae7cdf_m02	0.3838	683.3536	0.6266
post_ft_multistep_best	20260130_234540_ae7cdf_m03	-	-	-
post_ft_baseline_calc_best	20260130_234540_ae7cdf_m04	0.3838	683.3536	0.6266
post_ft_multistep_calc_best	20260130_234540_ae7cdf_m05	0.3838	683.3536	0.6266
post_ft_multistep_T1_calc_best	20260130_234540_ae7cdf_m06	0.3838	683.3536	0.6266

-
-
- ?????numeric dev??
- baseline(post-ft) vs best calc gate?Numeric-EM 0.3838 ? 0.3838?RelErr(mean) 683.3536 ? 683.3536?Coverage 0.6266 ? 0.6266

5.3 ????

?3 ?????

label	run_id	full_r10	full_mrr10	complex_r10	complex_mrr10
post_ft_multistep_T1_calc_best	20260130_234540_ae7cdf_m06	0.3772	0.2601	0.3909	0.2960

5.4 ?????????3??

??1?qid=8c8c8c34?

Query?Hasbro (HAS) 2023 one-time charges impact on operating profitability vs historical trends and cap allocation implications.

- Gold Answer?????In 2023, Hasbro’s operating result turned from a profit in prior years (407.7 million in 2022 and 763.3 million in 2021) to an operating loss of 1,538.8 million. A key driver behin...

Step0 Top3?008beea7_e0_c0, 8c8c8c34_e0_c2, f8aec91a_e0_c1

- Step1 Top3?008beea7_e0_c0, f8aec91a_e0_c1, 8c8c8c34_e0_c2

gap/stop?MISSING_ENTITY / MAX_STEPS?final_topk_size=10

-
- ????????????/????????????????????????? gap ??? query ?????????????
-
- **??2?qid=52e25ec7?**
-
- Query?Impact on net investing cash flows from EUC sale cash inflow offsets vs acquisition outflows, AVGO.

Gold Answer?????The $3,485 million inflow from the sale of the EUC business helped to partially offset the significantly higher cash expenditures related to acquisitions. Specifically, Broadcom’s ...

- Step0 Top3?506e7d1e_e0_c0, 52e25ec7_e0_c0, e4661352_e0_c3

Step1 Top3?506e7d1e_e0_c0, 52e25ec7_e0_c0, 1c47856d_e0_c1

- gap/stop?MISSING_ENTITY / MAX_STEPS?final_topk_size=10

????????????/????????????????????????? gap ??? query ?????????????

??3?qid=ed746c33?

Query?Cash flow & cap alloc implications of IRM's ASC 842 storage rev rec vs other lines.

- Gold Answer?????For its Global Data Center Business, Iron Mountain recognizes storage revenues under ASC 842 (leases) rather than ASC 606 (contracts with customers), which is applied to its other ...

Step0 Top3?ed746c33_e0_c0, 2a8785e8_e0_c15, a68b8600_e0_c5

- gap/stop?NO_GAP / NO_GAP?final_topk_size=10

????????????/????????????????????????? gap ??? query ?????????????

---

6 ??

?????????
?????????

- Query ?????????

?????????????

- ??????/???????????????

?????????????? query ??????????????????????????? agent ????

---

7 ??

???????????RAG??
gap ?????????

????

[1] ???????