We Know What You're Looking For: Recommendation for Large-Scale Open Source Software

我们懂你所寻：大规模开源软件推荐

崔星 吴敬征 凌祥 罗天悦

19th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM 2025)

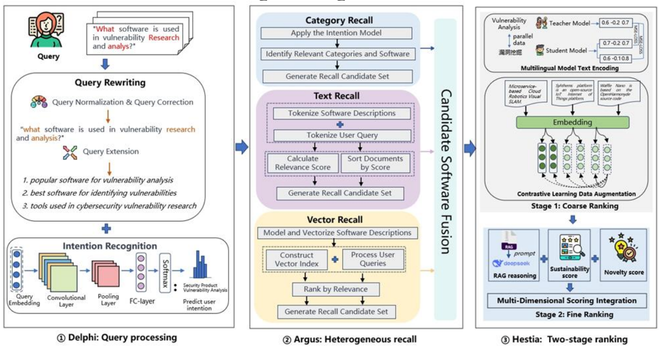
联系人：崔星，13051316652，cuixing@iscas.ac.cn

Background

In recent years, open-source software (OSS) has become central to software development. However, existing recommendation methods struggle with attribute modeling, multilingual support, and cold-start or data scarcity issues. To address these, we propose AthenaRec, an OSS recommendation system with three components: Delphi for intent analysis, Argus for heterogeneous retrieval, and Hestia for two-stage ranking—using contrastive learning and LLM enhancement. Evaluated on 7,500 real-world queries, AthenaRec improves performance by 10.9% over state-of-the-art methods, achieving 98.27% Hits@20, 95.60% MAP@20, 95.05% NDCG@20, and 92.92% MRR. Its key innovations include holistic attribute modeling, multilingual support, LLM-augmented ranking, and contrastive learning for cold-start resilience, making it a scalable and effective solution for large-scale OSS recommendation.

Methodology

We design AthenaRec with three modules: (1) Delphi for refining user queries, (2) Argus for recalling and fusing candidate software, and (3) Hestia for hierarchical two-stage ranking based on user needs.



Evaluation and results

Table 1: Performance comparison of AthenaRec and other methods

Model	Hits@10	Hits@20	MAP@10	MAP@20	NDCG@10	NDCG@20	MRR
LRMF	28.42±0.23	33.62±0.32	26.97±0.19	32.27±0.31	25.20±0.34	30.04±0.29	12.62±0.41
Req2Lib	35.62±0.27	41.91±0.47	36.44±0.23	42.65±0.43	37.75±0.18	43.86±0.38	23.75±0.46
RepoLike	41.73±0.25	45.80±0.28	41.27±0.19	45.30±0.37	38.67±0.18	39.28±0.38	34.62±0.24
DeepSeek-R1	84.77±0.48	86.95±0.29	83.86±0.23	85.01±0.34	83.10±0.27	84.91±0.42	78.42±0.38
ChatGPT-4o	78.10±0.27	79.71±0.37	78.98±0.17	80.15±0.32	78.85±0.33	80.73±0.50	68.13±0.29
Qwen2.5-7B	76.14±0.20	78.61±0.22	75.88±0.16	78.19±0.20	73.56±0.23	76.92±0.29	67.02±0.27
LLaMA-3.1-8B	72.96±0.23	75.42±0.19	73.85±0.21	76.02±0.18	70.27±0.25	73.41±0.26	63.88±0.31
Mistral-7B-v0.3	61.18±0.17	63.31±0.23	62.68±0.16	64.79±0.19	58.61±0.31	60.79±0.33	50.95±0.22
AthenaRec (Ours)	**95.29±0.13**	**98.27±0.09**	**93.17±0.08**	**95.60±0.10**	**93.39±0.07**	**95.05±0.11**	**92.92±0.13**

Table 2: Ablation analysis assessing the impact of multilingual modeling on the efficacy of coarse ranking

ID	Manual	DIKE	Accuracy	Precision	Recall	F1-Score
1	258	243	194 (75.2%)	0.798	0.752	0.774
2	263	257	196 (74.5%)	0.763	0.745	0.754
3	264	276	213 (80.7%)	0.772	0.807	0.789
AVG	262	259	201 (76.7%)	0.778	0.768	0.772

Table 3: Ablation study of fine ranking components

Model	Hits@10	Hits@20	MAP@10	MAP@20	NDCG@10	NDCG@20	MRR
w/o RAG Ranking	93.02±0.35	96.51±0.38	90.15±0.24	93.45±0.32	89.21±0.28	92.76±0.35	87.73±0.26
w/o Sustainability	94.24±0.24	97.11±0.18	90.94±0.29	94.59±0.23	90.30±0.25	93.95±0.28	88.75±0.32
w/o Novelty	94.90±0.38	98.01±0.35	91.46±0.27	94.71±0.20	90.47±0.17	94.16±0.38	89.27±0.34
w/o Integrated	91.79±0.23	94.51±0.16	89.36±0.14	91.66±0.12	87.92±0.15	90.79±0.16	87.06±0.20
AthenaRec (full)	**95.29±0.13**	**98.27±0.09**	**93.17±0.08**	**95.60±0.10**	**93.39±0.07**	**95.05±0.11**	**92.92±0.13**

Table 4: Cold start recommendation performance under ablation settings

Configuration	Hits@10	Recall@10	Visibility	NDCG@10
Coarse Ranking Only	62.50±0.35	55.10±0.28	49.30±0.12	59.60±0.29
w/o RAG Score	67.80±0.30	60.20±0.17	54.50±0.33	66.10±0.16
w/o Novelty Score	71.40±0.29	64.60±0.25	58.20±0.41	70.50±0.34
w/o Sustainability Score	69.50±0.11	62.70±0.24	56.80±0.20	68.30±0.27
AthenaRec(full)	**76.80±0.27**	**71.10±0.12**	**65.30±0.38**	**75.40±0.30**

Findings:

- ✓ **AthenaRec efficiency**: AthenaRec outperforms benchmarks with 98.27% Hits@20 and 95.60% MAP@20 by integrating Delphi, Argus, and Hestia modules.
- ✓ **Multi-Strategy benefits**: Combining various recall methods and fusion ranking improves accuracy and reduces bias in recommendations.
- ✓ **LLM impact**: Using LLMs with RAG enhances semantic accuracy and reasoning, significantly boosting OSS recommendation quality.

Contributions

- ➢ We introduce AthenaRec, a multi-language OSS recommendation system to enhance development efficiency and simplify software selection
- ➢ AthenaRec integrates heterogeneous retrieval with two-stage LLM ranking, using cross-lingual learning and RAG, plus sustainability and novelty scoring.
- ➢ AthenaRec achieves 98.27% Hits@20, 95.60% MAP@20, 95.05% NDCG@20, and 92.92% MRR on 7,500 queries, outperforming existing methods by 10.9%. It also includes a validated VSCode plugin.

Conclusion

This paper proposes AthenaRec, a large-scale OSS recommendation framework that addresses limitations in existing methods such as missing project attributes, multilingual extraction difficulties, and cold-start problems. It enhances accuracy through three modules: Delphi for query understanding, Argus for candidate retrieval, and Hestia for multi-stage ranking. Future work will further integrate LLMs to improve ranking, query processing, and feature modeling.

This paper is supported by Open Source Map