CQGen-MAS: A Multi-Agent System for Competency Questions Generation from Ontology

Extended Abstract

Fei Du
BSH Home Appliances Holding (China) Co., Ltd
Nanjing, China
Department of Computer Science and Technology, Tsinghua University
Beijing, China
Rick.Du@bshg.com

Huilong An
BSH Home Appliances Holding (China) Co., Ltd
Nanjing, China
Huilong.An@bshg.com

Feng Li
BSH Home Appliances Holding (China) Co., Ltd
Nanjing, China
Feng.LiF@bshg.com

Li Chen
BSH Home Appliances Holding (China) Co., Ltd
Nanjing, China
Li.Chen@bshg.com

Weidong Liu*
Department of Computer Science and Technology, Tsinghua University
Beijing, China
liuwd@tsinghua.edu.cn

ABSTRACT

Competency questions (CQs) delineate an ontology's intended scope and use, yet many published ontologies are released without CQs, hindering interpretation and reuse. We present CQGen-MAS, a multi-agent system that automatically generates CQs from an existing ontology. We evaluate CQGen-MAS on a diverse set of ontologies and compare it with baseline approaches. Experimental results show that CQGen-MAS consistently performs favorably across diverse ontologies, indicating improved coverage and stronger semantic alignment with their conceptual scope.

KEYWORDS

Competency Questions; Multi-Agent System; Ontology; Large Language Model

ACM Reference Format:
Fei Du, Huilong An, Feng Li, Li Chen, and Weidong Liu. 2026. CQGen-MAS: A Multi-Agent System for Competency Questions Generation from Ontology: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/RRYE6579

1 INTRODUCTION

Competency Questions (CQs) are user-oriented interrogatives that define an ontology's intended scope and help engineers select relevant concepts and relations, while also validating whether the ontology can answer key information needs [8–10, 12]. However, many published ontologies are released without associated CQs [2, 7, 14], making it difficult for downstream users to assess intended use cases and, in turn, limiting reuse and interoperability. This motivates the task of deriving CQs directly from an existing ontology, turning implicit conceptual intent into explicit, testable checks.

Recent work explores using large language models (LLMs) to generate CQs [1, 3, 4, 11], often with carefully designed prompts, but two practical limitations remain. First, many methods rely on coarse-grained inputs such as corpora [3], concept lists [11], or isolated triples [1], which provide insufficient structural context. As a result, they struggle to generate complex, multi-hop CQs and instead produce mostly surface-level, single-hop retrieval questions, which conflicts with CQ design practice [8]. Second, most approaches lack executable validation and explicit coverage control, treating CQ generation as a one-off step rather than an iterative process that refines questions against the ontology.

To address these gaps, we introduce **CQGen-MAS**, a multi-agent system for automatically generating CQs from existing ontologies. CQGen-MAS targets (i) complex multi-hop CQ generation under LLM context constraints and (ii) progressive improvements in conceptual coverage via iterative refinement, yielding CQ sets that better reflect an ontology's scope.

2 PROBLEM STATEMENT

Given an ontology O, formalized as a graph of interconnected classes and properties representing domain knowledge, our goal is to automatically generate a set of competency questions CQs that makes the ontology's conceptual scope explicit and testable. We require CQs to satisfy two criteria. (1) *Semantic relevance*: each generated question must be conceptually scoped to O, i.e.,

$$\forall cq_i \in CQs,\ \exists q_i = \tau(cq_i) \text{ s.t. } \mathrm{domain}(q_i) \subseteq \mathcal{D}_O, \tag{1}$$

where $\mathcal{D}_O$ denotes the set of concepts (classes/properties) defined in O, $\tau(\cdot)$ maps a natural-language CQ to a formal query (e.g., SPARQL), and $\mathrm{domain}(q_i)$ extracts the ontology concepts referenced in q_i. (2) *Conceptual coverage*: the question set should cover

*Corresponding Author.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/RRYE6579

a sufficient portion of O's conceptual space, i.e.,

$$\mathrm{Cov}(CQs, O) = \frac{\left| \bigcup_{cq_i \in CQs} \mathrm{domain}(q_i) \right|}{|\mathcal{D}_O|} \geq \alpha, \qquad (2)$$

where $\alpha \in (0, 1]$ is a coverage threshold. Overall, the task targets CQs that are both semantically aligned with O and sufficiently comprehensive to improve interpretability and reuse.

3 APPROACH

CQGen-MAS coordinates three agents—Ontology Segmenter, CQ Generator, and SPARQL Validator—in a closed loop to produce CQs.

Ontology Segmenter. Feeding a full ontology to an LLM is often impractical: context limits can cause semantic compression, washing out structural cues and degrading CQ quality. Yet segmenting too finely (e.g., generating from isolated triples) tends to bias the output toward surface-level, single-hop questions, making complex multi-hop CQs hard to obtain. The Segmenter therefore preprocesses the ontology by selecting an appropriate segmentation granularity and producing semantically coherent subgraphs $\{G_i\}$. For heterogeneous ontologies, it can choose among graph partitioning methods (e.g., METIS [6], Louvain [5], Leiden [13]) to adapt the subgraph scale and coherence.

Competency Questions Generator. Given each subgraph G_i, the Generator produces CQs that are semantically grounded in the ontology and span both simple single-hop and complex multi-hop requirements. Since CQ style and intent are domain-dependent, we use few-shot prompting with a small set of high-quality in-domain CQ exemplars (plus a brief ontology description) to stabilize phrasing and encourage appropriate complexity.

SPARQL Validator. Generator alone does not guarantee CQ-set quality, so we introduce a Validator to operationalize the two requirements in our problem statement: semantic relevance and conceptual coverage. The Validator translates each CQ into an executable SPARQL query and executes it against the ontology. Queries that execute successfully and return non-empty answers provide evidence of consistency with the ontology's structure, while the referenced classes are used to track conceptual coverage. Uncovered concepts and validation feedback are then returned to the Generator to trigger targeted expansion, forming an iterative generate–validate–expand loop that improves both correctness and coverage.

Table 1: Recall of Competency Questions: Simple vs. Complex.

Method	Type	OneM2M	SAREF4ENV	VGO	VicinityCore
Ours	Simple	**84.2**	**80.0**	**88.9**	**77.8**
	Complex	**77.8**	**66.7**	**71.5**	**47.6**
LLM4KE [11]	Simple	16.7	31.1	55.5	24.5
	Complex	13.3	3.8	21.9	21.4
RETROFIT-CQs [1]	Simple	81.7	74.0	81.4	68.6
	Complex	70.4	38.5	38.3	**28.6**

4 EVALUATION

We compare CQGen-MAS against the following two baseline approaches: **LLM4KE** [11] and **RETROFIT-CQs** [1]. Three methods are evaluated on a set of heterogeneous ontologies that vary in domain, size, and CQ complexity, in order to assess whether the proposed method can reliably generate valid and representative competency questions. While our primary goal is coverage, we report Recall here to ensure a fair and direct comparison with baseline approaches following their evaluation protocols. Additionally, we stratify this comparison across complexity levels (simple vs. complex) to highlight performance on complex CQs.

As shown in Table 1, CQGen-MAS achieves higher recall on expert-authored benchmark CQs across four heterogeneous ontologies, with particularly strong performance on complex CQs that require multi-class and multi-relation reasoning. In addition to the complexity-level comparison reported here, we also evaluated different LLM-Method pairs and conducted ablation studies on key components of CQGen-MAS. Due to extended abstract space constraints, these results are omitted. We refer interested readers to the full paper for the complete experimental results and detailed analyses.

5 CONCLUSION

In this paper, we introduced CQGen-MAS, the first multi-agent framework for generating CQs from existing ontologies. Our system contributes a new iterative generation–validation paradigm, and demonstrates how SPARQL-based validation can be effectively operationalized within a MAS setting.

We evaluated the CQGen-MAS over four diverse ontologies. Experimental results show that our method outperforms existing baseline approaches, achieving superior recall and F1 scores in most cases. In addition, our approach exhibits a clear advantage in formulating complex questions.

Despite its strong performance, CQGen-MAS has room for improvement. We treat expert-authored CQ sets as ground truth in our evaluation and do not further assess their own conceptual coverage or semantic alignment with the ontology; in practice, we observed that some expert CQ sets may not fully match the ontology's scope. Future work will explicitly analyze this issue and introduce dedicated experiments and metrics for coverage (and alignment) of CQ sets. In addition, Ontology Segmentor could benefit from combining structural and semantic cues to avoid separating related entities. Furthermore, the development of more reliable benchmarks and alternative evaluation criteria beyond semantic similarity are essential for a faithful assessment of generation performance.

REFERENCES

[1] Reham Alharbi, Valentina Tamma, Floriana Grasso, and Terry Payne. 2024. An experiment in retrofitting competency questions for existing ontologies. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*. 1650–1658.

[2] Reham Alharbi, Valentina Tamma, Floriana Grasso, and Terry R Payne. 2024. Investigating Open Source LLMs to Retrofit Competency Questions in Ontology Engineering. In *Proceedings of the AAAI Symposium Series*, Vol. 4. 188–198.

[3] Mary-Jane Antia and C Maria Keet. 2023. Automating the generation of competency questions for ontologies with AgOCQs. In *Iberoamerican Knowledge Graphs and Semantic Web Conference*. Springer, 213–227.

[4] Fiorela Ciroku, Jacopo de Berardinis, Jongmo Kim, Albert Meroño-Peñuela, Valentina Presutti, and Elena Simperl. 2024. RevOnt: Reverse engineering of competency questions from knowledge graphs via language models. *Journal of*

Web Semantics 82 (2024), 100822.

[5] Nicolas Dugué and Anthony Perez. 2015. *Directed Louvain: maximizing modularity in directed networks.* Ph.D. Dissertation. Université d'Orléans.

[6] George Karypis and Vipin Kumar. 1998. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing* 20, 1 (1998), 359–392.

[7] C Maria Keet, Zola Mahlaza, and Mary-Jane Antia. 2019. CLaRO: a controlled language for authoring competency questions. In *Research Conference on Metadata and Semantics Research*. Springer, 3–15.

[8] Glaice Kelly Q Monfardini, Jordana S Salamon, and Monalessa P Barcellos. 2023. Use of competency questions in ontology engineering: a survey. In *International Conference on Conceptual Modeling*. Springer, 45–64.

[9] Natalya F Noy, Deborah L McGuinness, et al. 2001. Ontology development 101: A guide to creating your first ontology.

[10] María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. 2022. LOT: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence* 111 (2022), 104755.

[11] Youssra Rebboud, Lionel Tailhardat, Pasquale Lisena, and Raphael Troncy. 2024. Can LLMs Generate Competency Questions?. In *European Semantic Web Conference*. Springer, 71–80.

[12] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernandez-Lopez. 2015. The NeOn Methodology framework: A scenario-based methodology for ontology development. *Applied ontology* 10, 2 (2015), 107–145.

[13] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports* 9, 1 (2019), 1–12.

[14] Dawid Wiśniewski, Jedrzej Potoniec, Agnieszka Ławrynowicz, and C Maria Keet. 2019. Analysis of ontology competency questions and their formalizations in SPARQL-OWL. *Journal of Web Semantics* 59 (2019), 100534.