Human-Friendly Explanation for Ontology-based Concept Similarity: Design and Development

Watanee Jearanaiwongkul[1], Teeradaj Racharak[1]

[1]*Advanced Institute of So-Go-Chi (Convergence Knowledge) Informatics, Tohoku University, Miyagi, Japan*

Abstract

While recent neuro-symbolic approaches have enabled interpretable computation of concept similarity in ontologies, translating these formal explanations into human-friendly forms remains a challenge. In this work, we investigate how large language models, particularly ChatGPT and Gemini, can be prompted to generate natural language explanations that justify similarity results in a way that is understandable to end users. Building on a neuro-symbolic framework for measuring concept similarity in Description Logic (DL) ontologies, we explore two types of human-friendly explanations i.e., node-based and path-based explanation. Furthermore, we evaluate LLMs's ability to generate each component of these path-based explanations using our small curated dataset. We evaluate the effectiveness of prompting approaches along different dimensions such as clarity, informativeness, and perceived usefulness through both qualitative analysis and user studies. Our results show the potential and limitations of using LLMs as a tool to bridge the gap between formal similarity reasoning and human interpretability, paving the way for more transparent ontology-driven systems.

Keywords

Explainable AI, Explanation generation, Ontology Concept Similarity, LLMs, Interpretable AI

1. Introduction

The increasing deployment of AI systems in high-risk decision-making domains (e.g. healthcare, finance, and laws) has gained high attention to their explainability to ensure transparency and accountability [1, 2]. Considering Knowledge Graphs (KGs), AI models can effectively predict similarity scores between concepts but explaining why two concepts are considered similar remains a challenging task and requires a human-friendly and interpretable for users to understand and trust the system [3].

Logic-based explanation methods have proposed promising approaches, offering formal and structured reasoning steps that faithfully reflect the underlying computational logic [4]. These methods can directly trace the paths and nodes within a KG that contribute to a similarity score. In our work, we specifically utilize the logic-based explanations produced by a prior work [5], which introduces a concept similarity measure in description logic $\mathscr{ELH}$ with pre-trained word embeddings. While this approach provides sound and rigorous explanations, their explanations that support the reasoning are symbolic, and still difficult for non-expert users to interpret.

To address this gap, we focus on creating human-friendly explanations that retain the structure of the original logic-based reasoning, but are simpler and easier to read. We define two explanation types that correspond to the key components of a KG: 1) Node-based Explanation and 2) Path-based Explanation. In this work, we focus on readability and clarity and omit the full details of similarity computation from the explanation trees due to their complexity and length.

To generate these explanations, we leverage Large Language Models (LLMs), specifically ChatGPT 4o, by evaluating its performance in generating "path-based" explanations and analyze its strengths and limitations using metrics such as precision, recall, and F1 score. Our study is an initial exploration toward transforming symbolic reasoning outputs to natural language explanations.

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan

watanee@tohoku.ac.jp (W. Jearanaiwongkul); racharak@tohoku.ac.jp (T. Racharak)
0000-0003-2101-1625 (W. Jearanaiwongkul); 0000-0002-8823-2361 (T. Racharak)

2. Dataset Preparation

Our goal was to construct a dataset consisting of original logic-based explanation and their corresponding human-friendly explanations. To achieve this, we designed two explanation formats according to homomorphism-based semantic similarity introduced in [5]: (1) **node-based explanation** (which describes how a similarity score was calculated by using the comparison of similarity between two nodes) and (2) **path-based explanation** (which describes by using the comparison of two paths instead).

To evaluate these formats, we conducted a survey with 20 participants (10 per each explanation type) and used their feedback to improve and select one design for our experiment. The survey provided participants with: (1) similarity score between two concepts, (2) background knowledge represented as knowledge graphs, (3) summary explanation, (4) detailed explanation, and (5) table explanation.

Participants were asked to review the similarity score and explanation of why they are considered similar, then respond to eight questions. The participants were recruited via Prolific[1], compensated fairly, and represented a wide range of demographics (ages 19–65, educational backgrounds from high school to PhD, and professions including data analysis, engineering, IT, and research). All participants were fluent in English and resided in countries such as the US, South Africa, Australia, Mexico, Canada, and France. Table 1 presented the main results from six questions. The responses were rated on a 0–7 Likert Scale (0 = lowest satisfaction, 7 = highest). While the node-based explanations were easier to “read”, path-based explanations were more easier to “understand” and made the explanation more sufficient as indicated by lower score on the questions about the need of additional explanation. Therefore, we selected to experiment with path-based explanation approach and prepared a dataset w.r.t. this format. A total of 20 well-constructed explanations were created and used for this study, examples of which are shown in Table 2.

Question	Average		Median	
	Node-based	Path-based	Node-based	Path-based
Q1. Does the explanation provide understandable reasons?	4.5	5.8	5	6
Q2. Is the explanation easy to read?	5.4	4.4	5	4.5
Q3. Is the explanation sufficient to answer why the 2 words are similar?	4.9	5.4	5	5.5
Q4. Are you interested in more explanation of why they are similar?	5.4	4.3	5.5	4.5
Q5. Is the explanation enough to answer how the 2 words are similar?	5.2	5.5	5	5.5
Q6. Are you interested in more explanation on how the similarity score was calculated?	5.6	4.5	5.5	4.5

Table 1
Comparison of survey results between node-based explanation and path-based explanation.

3. Experiments and Results

3.1. Experimental setting

This work conducted an experiment using OpenAI API GPT-4o model, configured with a temperature 0 and top-p of 0.05 to ensure deterministic and focused outputs. The objective was to evaluate the model’s performance in generating human-friendly explanations from the given logic-based explanations. As aforementioned, our study involves three types of explanations: summary, detailed, and table-based. Due to space limitation, a shortened version of the one-shot prompt is provided in Prompt 1, while the complete prompt is publicly available on this GitHub repository[2].

[1]https://www.prolific.com/
[2]https://github.com/realearn-people/sim-elh-explainer-to-text

<table>
<tr><th>Item</th><th>Detail</th></tr>
<tr><td>Description graphs representing the meaning of two concepts</td><td>Graph 1: SeaFood
– (Menu, contain, Fish)
– (Menu, contain, Shrimp)
– (Menu, serveWith, LemonSauce)

Graph 2: FriedFood
– (Menu, contain, Shrimp)</td></tr>
<tr><td>Original explanation (from [5])</td><td>[G1*SeaFood][G2*FriedFood] → (0.52, G1*Menu, G1*contain.Fish, G1*contain.Shrimp, G1*serveWith.LemonSauce, (contain, serveWith))
[G1*Fish][G2*Shrimp] → (0.7, G1*Fish, { }, (Fish,Shrimp))
[G1*Shrimp][G2*Shrimp] → (1.0, G1*Shrimp, { }, { })
[G1*LemonSauce][G2*Shrimp] → (0.02, G1*Shrimp, { }, (LemonSauce, Shrimp))</td></tr>
<tr><td>Summary explanation</td><td>They are 52% similar because
Their Top Parent node comparison: 1 of 1 matched 100%
Their Path comparison: 1 of 3 matched 100%</td></tr>
<tr><td>Detailed explanation</td><td>They are 52% similar because
They share top parent node:
100% match: Menu
They share exactly the same “path” from Graph SeaFood to Graph FriedFood:
100% match: (Menu, contain, Shrimp) and (Menu, contain, Shrimp)
They share the similar “path” from Graph SeaFood to Graph FriedFood (partly computed using the used embeddings):
70% match: (Menu, contain, Fish) and (Menu, contain, Shrimp)
2% match: (Menu, serveWith, LemonSauce) and (Menu, contain, Shrimp)</td></tr>
<tr><td>Table explanation</td><td><table>
<tr><th>#</th><th>Node or Path from Graph 1: SeaFood</th><th>Node or Path from Graph 2: FriedFood</th><th>Similarity Result</th></tr>
<tr><td>1</td><td>Menu</td><td>Menu</td><td>Same</td></tr>
<tr><td>2</td><td>(Menu, {contain}, Fish)</td><td>(Menu, {contain}, Shrimp)</td><td>70% match</td></tr>
<tr><td>3</td><td>(Menu, {contain}, Shrimp)</td><td>(Menu, {contain}, Shrimp)</td><td>Same</td></tr>
<tr><td>4</td><td>(Menu, {serveWith}, LemonSauce)</td><td>(Menu, {contain}, Shrimp)</td><td>2% match</td></tr>
</table></td></tr>
</table>

Table 2
Various types of explanations and graph representations for two concepts (namely, SeaFood and FriedFood). Note that asterisk indicates super-sub structures of graph's elements.

Prompt 1

Input structure: You are given 2 things: 1. Graph: A pair of knowledge graphs. 2. Explanation: A list of structured tuples that shown similarity scores and the reason how the score was calculated from: ...
Example of input: Graph 1: (...), **Graph 2:** (...)
Explanation:

- [G1*ActivePlace][G2*Mangrove] → (0.62, {G1*Place}, {∃{G1*canWalk, G1*canMoveWithLeg}.Trekking, ∃G1*canTravelWithSail.Kayaking}, {(canMoveWithLeg, canTravelWithSail)})
- [G1*Trekking][G2*Trekking] → (1.0, {G1*Trekking}, { } , { })
- [G1*Kayaking][G2*Trekking] → (0.97, {G1*Kayaking}, { } , {(Kayaking,Trekking)})

Output format of human-friendly explanation: ActivePlace and Mangrove are 62% similar meaning.

- Summary Explanation: ...
- Detail Explanation: ...
- Table Explanation: ...

Your task: generate a friendly, human-readable explanation based on the input below, using the exact format provided. **[Input:]**

3.2. Results and Discussion

Table 3 summarizes the results of the human-friendly explanation generation based on LLM approach. The evaluation results compare metrics across three explanation types: 1) Summary Explanation, 2) Detailed Explanation, and 3) Table Explanation. For the Summary Explanation, the exact match rate is high for the top parent node (0.95) but lower for path comparison (0.5), indicating that generating explanations for top parent node is the easiest task to tackle by LLMs, while generating for the whole path is more challenging. Similarly, the table also shows high exact match rate for generation explanations for top parent node in Detailed Explanation (Item 2.1 with 0.95 on exact match) and Table Explanation (Item 3.1 with 1 on exact match). In the Detailed Explanation section, precision and recall are high for listing the “same” path comparison (0.824), indicating balanced performance with minimal false

Explanation type	Item in each type of explanation	Metrics							
		Exact match		Precision		Recall		F1 score	
		CGPT	Gemini	CGPT	Gemini	CGPT	Gemini	CGPT	Gemini
1. Summary Explanation	1.1) # of exact match top parent node	0.95	1	-	-	-	-	-	-
	1.2) # of exact match path comparison	0.5	0.4	-	-	-	-	-	-
2. Detailed Explanation	2.1) List of same top parent nodes	0.95	1	-	-	-	-	-	-
	2.2) List of same path comparisons	-	-	0.824	0.895	0.824	0.94	0.824	0.919
	2.3) List of similar path comparisons	-	-	1	1	0.5	0.7	0.67	0.824
3. Table Explanation	3.1) List of same top parent nodes	1	1	-	-	-	-	-	-
	3.2) List of same path comparisons	-	-	0.833	0.9	0.882	1	0.857	0.947
	3.3) List of similar path comparisons	-	-	1	1	0.6	0.65	0.75	0.788

Table 3
Comparison of path-based explanation generation between ChatGPT-4o (CGPT) and Gemini-1.5-flash.

positives and false negatives. Moreover, the "similar" path comparison achieves perfect precision (1) but a lower recall (0.5), resulting in a moderate F1 score (0.67), which indicates that many true matches are missed. For Table Explanation, the performance is slightly better, with precision (0.833) and recall (0.882) for the "same" path comparison leading to an F1 score of 0.857, and the "similar" path comparison shows perfect precision (1) and a recall of 0.6, indicating the same trend of missing true matches as observed in Detailed Explanation. Overall, these results suggest that Table Explanation provides the most balanced performance across metrics, particularly for listing the results of the "same" path comparisons.

3.3. Related work

Logic-based explainability has recently emerged as a rigorous and interpretable alternative to post-hoc explainers. Marques-Silv [6] provides a comprehensive survey on logic-based explanations in trustable AI/ML which highlighting the needs of rigorous definitions, rigorous computation of explanations, but also expressivity of explanations. The paper also discusses ongoing challenges and outlines future directions, including the integration of symbolic and sub-symbolic methods and the need for scalable, user-centric explanations. In [7], they focused on explaining the predictions of machine learning models using interpretable concepts and logic rules. Specifically, explanation is provided in simple first-order logic format for its expressiveness. Considering explaining concept similarity in ontologies, Racharak [5] introduced an $\mathscr{ELH}$-based similarity scoring framework that combines description logic with (pre-trained) embedding-based representations and offers structured symbolic explanations as output. These methods offer precise explanations but the output form can be further optimized for clarity to non-expert users. This work fulfills this gap by investigating how to generate human-friendly explanation from the neural-symbolic structures for ontology-based concept similarity.

4. Conclusion

This study presents an initial step to explore human-friendly explanation generation for similarity score of concepts. By defining node-based and path-based explanations and focusing on the latter in a ChatGPT-based experiment, we demonstrated both the potential and challenges of using LLMs for interpretable AI explanations. Our findings show that ChatGPT can effectively generate explanations for single node comparison, particularly top-parent nodes in all three types of explanations. However, a limitation remains in capturing true matches for similar path comparisons in both Detailed Explanations and Table Explanations. Upon examining the errors, we observed that LLMs tend to miss correct matches when the size of description graphs increase, i.e., the number of nodes and the graphs' depth increase. Another error happened when each edge was labeled by a set in the explanation graphs. Our future steps are to enlarge the dataset, collect more feedback to improve the explanation format, as well as experiment with other LLMs and deep learning-based models' construction for text generation. We also plan to fine-tune LLMs with our dataset for our explanation-graph-to-text translation in future.

Declaration of Generative AI and AI-assisted Technologies

During the preparation of this work, the authors used ChatGPT for grammatical check. After using this tool, the authors reviewed and edited the content as needed. Thus, they take full responsibility for the content of the publication.

References

[1] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608 (2017).

[2] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (xai), IEEE access 6 (2018) 52138–52160.

[3] C. Molnar, Interpretable machine learning, Lulu. com, 2020.

[4] J. Marques-Silva, Logic-based explainability in machine learning, in: Reasoning Web. Causality, Explanations and Declarative Knowledge: 18th International Summer School 2022, Berlin, Germany, September 27–30, 2022, Tutorial Lectures, Springer, 2023, pp. 24–104.

[5] T. Racharak, On approximation of concept similarity measure in description logic ELH with pre-trained word embedding, IEEE Access 9 (2021) 61429–61443. URL: https://doi.org/10.1109/ACCESS.2021.3073730. doi:10.1109/ACCESS.2021.3073730.

[6] J. Marques-Silva, Logic-based explainability: past, present and future, in: International Symposium on Leveraging Applications of Formal Methods, Springer, 2024, pp. 181–204.

[7] G. Ciravegna, P. Barbiero, F. Giannini, M. Gori, P. Lió, M. Maggini, S. Melacci, Logic explained networks, Artificial Intelligence 314 (2023) 103822.