

# INTERACSPARQL: An Interactive System for SPARQL Query Refinement Using Natural Language Explanations

Xiangru Jian, Zhengyuan Dong, M. Tamer Özsu

1. Cheriton School of Computer Science, University of Waterloo



Code



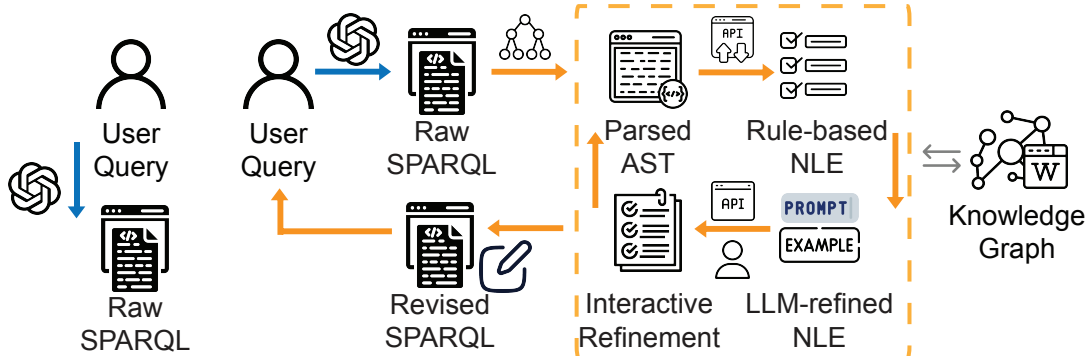
Paper

## Motivation

SPARQL is difficult to write, verify, and debug for non-expert users.

Single-turn  
NL2SPARQL

Interac  
SPARQL



## Takeaway

### Transparent refinement.

INTERACSPARQL converts SPARQL into structured natural-language explanations, making query logic easier to inspect and revise.

### Effective correction.

Iterative self-refinement substantially outperforms one-shot generation, showing that explanation-guided correction is practical and robust.

### Structured guidance matters.

The two-stage NLE design provides both accuracy and readability, and consistently outperforms simpler explanation variants.

### Grounded interaction matters.

Tool-assisted entity and property search is critical for resolving incorrect IRIs and achieving accurate query refinement.

## Pipeline

INTERACSPARQL parses a query into an AST, generates structured rule-based explanations, refines them with an LLM, and uses interactive feedback plus tool calls to iteratively correct the query.

### Input



Raw SPARQL

```
SELECT ?tvShow WHERE {
  ?tvShow wdt:P31 wd:Q5398426;
  wdt:P161 wd:Q7278;
  wdt:P1113 4;
  wdt:P580 ?startDate.
  FILTER(YEAR(?startDate) = 1983)
}
```

### Output



Revised SPARQL

```
SELECT ?tvShow WHERE {
  ?tvShow wdt:P31 wd:Q5398426;
  wdt:P161 wd:Q23760;
  wdt:P2437 ?seasons;
  wdt:P580 ?startDate.
  FILTER(?seasons = 4)
  FILTER(YEAR(?startDate) = 1983)
}
```

### InteracSPARQL System

#### ① Parsing AST

Pattern 1. Basic Graph Patterns (BGPs)

| Subject | Predicate                  | Object                   |
|---------|----------------------------|--------------------------|
| ?tvShow | {prefix}/prop/direct/P31   | {prefix}/entity/Q5398426 |
| ?tvShow | {prefix}/prop/direct/P161  | {prefix}/entity/Q7278    |
| ?tvShow | {prefix}/prop/direct/P1113 | 4 (Literal)              |

Pattern 2. Filter

| Operator | Argument 1       | Argument 2     |
|----------|------------------|----------------|
| =        | Year(?startDate) | 1983 (Literal) |

prefix=http://www.wikidata.org

Correct or Not?

#### ④ Interactive Refinement

The query **does not return any results**, which suggests there might be an issue with the criteria or the properties used.

Verify that Rowan Atkinson's entity ID is correct (wd:Q7278) and that the properties for 'number of episodes' (wdt:P1113) and 'start time' (wdt:P580) are appropriate for the dataset.

The entity for Rowan Atkinson is incorrect; it should be wd:Q23760 instead of wd:Q7278.

#### ② Rule-based NLE

Patterns: The following conditions should be satisfied:

1 The Basic Graph Pattern includes following statements:

...

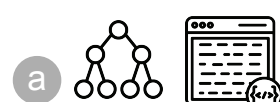
1.2 The entity 'tvShow' has the property '{prefix}/prop/direct/P161 (cast member)', and its value is '{prefix}/entity/Q7278 (political party)'.

1.3 The entity 'tvShow' has the property '{prefix}/prop/direct/P1113 (number of episodes)', and its value is '4'.

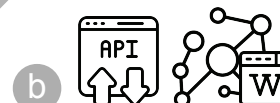
#### ③ LLM-refined NLE

| Module      | SPARQL                                                                                                                                                             |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| bgp         | ?tvShow wdt:P31 wd:Q5398426                                                                                                                                        |
| Explanation | Identify entities (?tvShow) that are instances of [television series] <wd:Q5398426>. Here, <wd:P31> is the property for [instance of].                             |
| bgp         | ?tvShow wdt:P161 wd:Q7278                                                                                                                                          |
| Explanation | Ensure that the television series (?tvShow) has a [cast member] associated with a [political party] <wd:Q7278>. Here, <wd:P161> is the property for [cast member]. |
| bgp         | ?tvShow wdt:P1113 4                                                                                                                                                |
| Explanation | Ensure that the television series (?tvShow) has a [number of episodes] equal to 4. Here, <wd:P1113> is the property for [number of episodes].                      |

### Tools



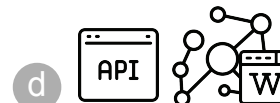
AST Parser



NL label Retrieval



Query Execution Endpoint



Entity & Property Search

## Experiment

### Main Results

Self-refinement substantially improves over one-shot generation, while accurate NLEs reveal a much higher performance ceiling.

| Dataset    | Metric    | Raw   | Self-refine | Upper bound |
|------------|-----------|-------|-------------|-------------|
| QALD-10    | Precision | 0.135 | 0.389       | 0.93        |
|            | F1        | 0.136 | 0.393       | 0.93        |
| QALD-9-WD  | Precision | 0.28  | 0.581       | 0.833       |
|            | F1        | 0.264 | 0.561       | 0.837       |
| QALD-9-DBP | Precision | 0.473 | 0.549       | 0.931       |
|            | F1        | 0.467 | 0.532       | 0.931       |

### Ablation: NLE Design

The full NLE design consistently performs best, confirming the value of structured AST-based guidance in both direct generation and self-refinement.

| Task Setting      | Variant | F1    | Precision | Recall |
|-------------------|---------|-------|-----------|--------|
| Direct Generation | OD      | 0.93  | 0.93      | 0.93   |
|                   | BQB     | 0.767 | 0.766     | 0.768  |
|                   | NFS     | 0.818 | 0.819     | 0.817  |
|                   | BQFS    | 0.376 | 0.376     | 0.376  |
| Self-Refinement   | OD      | 0.393 | 0.389     | 0.408  |
|                   | BQB     | 0.332 | 0.329     | 0.344  |
|                   | NFS     | 0.378 | 0.374     | 0.391  |
|                   | BQFS    | 0.294 | 0.291     | 0.308  |

### Ablation: Components

Both NLE and external tools are important, but removing external tools causes the largest performance drop.

| Component Setting  | F1   | Precision | Recall |
|--------------------|------|-----------|--------|
| Original           | 0.39 | 0.389     | 0.408  |
| w/o NLE            | 0.35 | 0.347     | 0.37   |
| w/o External Tools | 0.11 | 0.112     | 0.123  |
| w/o Both           | 0.09 | 0.086     | 0.098  |

### Ablation: Iteration Budget

Most gains arrive within five refinement iterations, while increasing the budget to ten yields only modest improvement.

| Dataset | Max Iterations | F1    | Precision | Recall |
|---------|----------------|-------|-----------|--------|
| QALD-10 | 1              | 0.307 | 0.303     | 0.326  |
|         | 5              | 0.393 | 0.389     | 0.408  |
|         | 10             | 0.417 | 0.41      | 0.442  |

### Human Evaluation

Participants preferred original design for completeness, usefulness, and overall clarity.

| Aspect                 | Summary                                                         |
|------------------------|-----------------------------------------------------------------|
| Overall preference     | Participants consistently preferred the original design (OD).   |
| Best dimensions for OD | Completeness, usefulness, and overall clarity.                  |
| BQFS                   | More visually appealing, but missed important semantic details. |
| NFS                    | Semantically complete, but less readable and more mechanical.   |
| BQB                    | Lowest ratings due to missing structure and weak utility.       |