

Supplementary Material for
CLARE: Context-Aware, Interactive Knowledge Graph
Construction from Transcripts

Ryan Henry, Jiaqi Gong

September 25, 2025

Contents

S1 Model configurations	3
S2 Prompt templates	3
S2.1 Prompt S1. Entity Extraction	3
S2.2 Prompt S2. Entity Disambiguation	4
S2.3 Prompt S3. Relation extraction	4
S3 Source code, versions, and environment	6
S3.1 Source Code Location	6
S3.2 Runtime environment	6

S1 Model configurations

Table S1: Models and configurations used in CLARE, including task assignments, providers, access methods, and decoding parameters.

Task	Model name	Provider	Date Accessed	Access method	Temp.	Top-p
Entity Extraction	GPT-4o-mini	OpenAI	2025-05-22	API	0.0	0.5
Entity Disambiguation	GPT-4o-mini	OpenAI	2025-05-22	API	0.0	0.5
Relation Extraction	GPT-4o-mini	OpenAI	2025-05-22	API	0.0	0.5
Token Embedding	all-mpnet-base-v2	Sentence-Transformers	2025-05-22	Python Package	N/A	N/A

S2 Prompt templates

Each template is given verbatim. Replace placeholders like `{num_words}` or `{max_length}` as appropriate.

S2.1 Prompt S1. Entity Extraction

Task. For each text cluster, representative samples are provided to the LLM, which extracts up to `{num_topics}` entities, each no longer than `{max_length}` words. To reduce redundancy, the list of previously extracted entities is included in subsequent prompts.

Listing 1: Entity extraction prompt

```

1 You are an advanced AI assistant specialized in summarizing information.
2 Determine the core theme from the following fragments (delimited by triple backticks).
3 Your answer must be a comma-separated list of at most {num_topics} keywords, each at most
  {max_length} words.
4 Write your answer in {prompt_language} without full sentences.
5 These keywords have already been chosen so do not repeat them: <{",".join(avoid_content)
  }>
6 Information: '''{','.join(chosen_texts)}'''
7 Your response:

```

S2.2 Prompt S2. Entity Disambiguation

Task. A comma-separated list of keywords is supplied to the LLM. The model refines this list by removing duplicates and semantically similar terms, constraining each keyword to at most {max_length} words, and discarding extraneous symbols. Terms may be split or reorganized as needed, and the output is returned as a clean set of keywords without additional labels.

Listing 2: Entity extraction prompt

```
1 You are an advanced AI assistant specialized in analyzing various pieces of information
  and providing precise summaries.
2 In the following task, you will be provided with a list of keywords, each separated by a
  comma.
3 You may split or reorganize the terms as needed.
4 Rules:
5     1. Remove duplicates or semantically similar terms.
6     2. Each keyword must be at most {max_length} words.
7     3. Do not include extra symbols.
8     4. Do not add any prefix or label (such as "Processed Keywords:" or "Raw Keywords:")
       to your response.
9 Examples:
10     Raw Keywords: Mineral processing EPC, Mineral processing, EPC Mineral processing.,
      circulation, suction, hollow shaft., alkali-acid purification
11     Processed Keywords: Mineral processing EPC, circulation, suction, hollow shaft,
      alkali-acid purification
12 Raw Keywords: {"", ".join(core_list)}
13 Processed Keywords:
```

S2.3 Prompt S3. Relation extraction

Task. This prompt instructs the LLM to extract relational triples from candidate entity pairs. For each pair, the model determines subject–object directionality, proposes up to three precise predicates, and justifies each choice with a short rationale. The output is returned as a structured JSON object containing pair indices, relation directions, predicates, and rationales, with no additional commentary.

Listing 3: Relation extraction prompt

```
1 You are an expert relation extractor. Follow these rules exactly:
2     1. For each entity pair determine which is the actor (subject) and which is the
      receiver (object). If roles are unclear, set direction to "none".
3     2. Propose one or more precise predicates that describe distinct relationships
      between subject and object. Avoid vague verbs like led to, part of, or caused by.
4     3. If there are multiple meaningful relations, list up to three. Each must have its
      own direction and rationale.
```

4. For each relation include a brief rationale (15 words) explaining your choice of subject, object, and predicate.

5. Output exactly one JSON object of the form:

```
{
  "results": [
    {
      "pair_index": 1,
      "direction": "forward",
      "relationship": "fought against",
      "rationale": "Nelson Mandela appears as grammatical subject and initiates
the action"
    },
    {
      "pair_index": 1,
      "direction": "reverse",
      "relationship": "surrendered to",
      "rationale": "United States is the object receiving the action of surrender"
    },
    {
      "pair_index": 2,
      "direction": "reverse",
      "relationship": "commander of",
      "rationale": "Ulysses Grant is identified as leading Union's Army in the
sentence"
    },
    {
      "pair_index": 2,
      "direction": "none",
      "relationship": "",
      "rationale": "No clear subject/object roles"
    }
  ]
}
```

6. Do not include any other keys or commentary.

Context:

```
{chunk_text}
```

Pairs to process:

```
{pairs_text}
```

S3 Source code, versions, and environment

S3.1 Source Code Location

- The MINE dataset is available in the KG-Gen repository: <https://github.com/stair-lab/kg-gen>
- The commit hash of the KG-Gen repository that was used for this experiment is: 4316e3621cd54d573b380d221fde625ff78da570
- The source code and GUI executable for CLARE are available at: <https://github.com/ryanwaynehenry/CLARE>
- The commit hash used for CLARE used in this experiment is: 7b1179711f58f6914e12ab45677142868c7fa2cb

S3.2 Runtime environment

- Windows 11
- Python 3.10.16, litellm 1.65.1, sentence-transformers 4.1.0
- CLARE configured with 30 entities per cluster and a dynamic threshold of 50%