Re^2-DocRED: Revisiting Revisited DocRED for Joint Entity and Relation Extraction

Chen Kim Heng[1], Shao Wen Tong[1], Julian Wei Sheng Wong[1]
[1]KLASS Engineering and Solutions
Correspondence: chenkim.heng@klasses.com.sg

Abstract

Document-level Joint Entity and Relation Extraction (JERE) benchmarks such as DocRED, Re-DocRED, and DocGNRE suffer from pervasive False Negatives (FN), undermining training and evaluation. In this paper, we introduce SiftingLogic – a training-free annotation pipeline that leverages LLMs with user-specifiable reasoning, enriched inverse/co-occurring relation schemas, and novel entity-level constraints to systematically address FN gaps. Applying SiftingLogic and our enriched schema of inverse and co-occurring relations, we add 29,580 verified triplets to Re-DocRED (train/dev, +27%) and over 9,700 verified triplets to DocGNRE test (+49.89%), yielding the enhanced Re^2-DocRED dataset. Beyond English datasets, we also apply our SiftingLogic to REDFM Mandarin test set, resulting in a significant increase in triplets from 663 to 1,391 (+109.8%) demonstrating our pipeline's generalisability across languages and datasets. Our experiments show that recall scores of models trained on existing public datasets drop notably on our revised splits, whereas our enriched training set mitigates this, underscoring persistent FN gaps and motivating our proposed SiftingLogic and Re^2-DocRED. To facilitate further research and reproducibility of our work, the Re^2-DocRED dataset is released at https://github.com/klassessg/re2-docred.

1 Introduction

DocRED (Yao et al., 2019) pioneered document-level JERE with a Wikidata-based benchmark, surpassing earlier datasets (Zhang et al., 2017; Walker et al., 2006; Mitchell et al., 2005; Roth and Yih, 2004; Gurulingappa et al., 2012; Gardent et al., 2017; Riedel et al., 2010). However, its reliance on distant supervision (Mintz et al., 2009) and limited human validation led to a significant amount of False Positives (FP) (Xiao et al., 2020) and False Negatives (FN) (Tan et al., 2022b).

Re-DocRED (Tan et al., 2022b) mitigates FN issues by using KD-DocRE (Tan et al., 2022a), DocuNET (Zhang et al., 2021), and ATLOP (Zhou et al., 2020) to generate candidate relations, which are refined by human annotators while retaining all original DocRED annotations. While these methods help mitigate FN issues to a certain degree, the problem remains only partially addressed (Li et al., 2023).

Li et al. (2023) note that Re-DocRED's annotations inherit DocRED's relation distribution, limiting long-tail coverage. They use GPT-3.5 (OpenAI, 2023) to generate triplets, construct hypothesis triplets using the entities in GPT output and verify them using T5-XXL (Raffel et al., 2019), producing a distantly supervised train set and an improved human-verified test set, which they release as DocGNRE. While DocGNRE further mitigates FN issues, we show in our paper that DocGNRE – along with Re-DocRED – remains incomplete, with substantial FN persisting.

In this paper, to address the prevailing FN issues in existing JERE datasets, we introduce SiftingLogic – a training-free LLM-based approach that utilizes novel entity-level constraints along with enriched inverse/co-occurring relation schemas to systematically tackle FN issues. We apply the SiftingLogic pipeline to re-annotate both the DocGNRE and REDFM (Huguet Cabot et al., 2023) Mandarin test sets, introducing our enhanced dataset, Re^2-DocRED. We then present comparative statistics between Re^2-DocRED and Re-DocRED, along with results for the revised REDFM test set.

To summarize, our main contributions are:

- **SiftingLogic** – a novel, training-free LLM-based annotation pipeline introducing explainable and user-specifiable reasoning for Joint Entity and Relation Extraction.
- A new concept of **entity-level constraints** derived from relation definitions to rigorously

Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics
Volume 1: Long Papers, pages 4585–4621

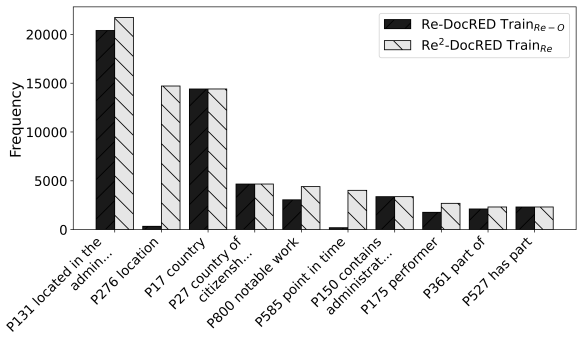

Figure 1: Top-10 Relation Frequency in Training Split. **Train**$_{Re}$ refers to train split of our proposed Re^2-DocRED; augmented from Re-DocRED train split.

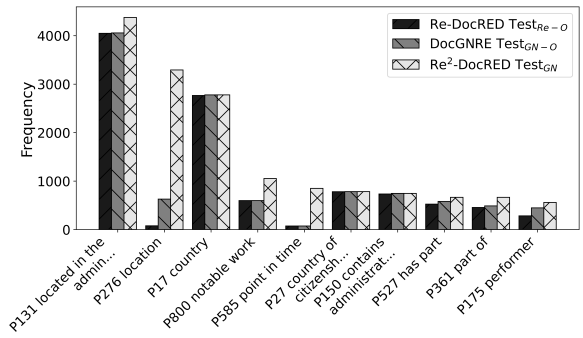

Figure 2: Top-10 Relation Frequency in Test Split. **Test**$_{GN}$ refers to test split of our proposed Re^2-DocRED; augmented from DocGNRE test set (**Test**$_{GN-O}$).

validate candidate entity pairs.

- **Re^2-DocRED** – a substantially enhanced dataset addressing false negatives in Re-DocRED and DocGNRE, including a revised Mandarin test split for REDFM.

In addition, as part of our proposed pipeline, we are introducing:

- Relation-specific compatible entity type pairs for Re-DocRED
- Relation verbalizations for Re-DocRED and REDFM
- Enhanced inverse and co-occurring relations rules originally proposed in Re-DocRED

2 Rule-Based Expansion of Re-DocRED and DocGNRE

Despite Re-DocRED's (Tan et al., 2022b) attempt to mitigate FN issues, many triplets are observed to be missing. We show that through leveraging our enhanced inverse and co-occurring relation rules (Appendices D, E), we add over 24,000 triplets to the training set (+28.87%). Figure 1 compares the top relation frequencies in the revised training set (**Train**$_{Re}$) of our proposed Re^2-DocRED with the training set of Re-DocRED(**Train**$_{Re-O}$).

DocGNRE (Li et al., 2023) test set (**Test**$_{GN-O}$) added over 2,000 triplets to Re-DocRED's test set (**Test**$_{Re-O}$) . However, these additional triplets are also insufficient to fully address FN issues. Using the same rules as above, we add over 5,000 triplets (+27.69%), forming **Test**$_{GN}$. Figure 2 compares top relation frequencies in the enhanced test set (**Test**$_{GN}$) of our proposed Re^2-DocRED with **Test**$_{Re-O}$ and **Test**$_{GN-O}$ respectively.

Through rule-based expansion alone, we add over 27% more triplets across Re-DocRED train, dev, test, and DocGNRE test splits, forming part of our Re^2-DocRED. Table 1 shows the number of triplet additions to these existing datasets, highlighting the persistent FN issues in them.

3 Recent Works

For Relation Extraction (RE), some notable datasets include TACRED (Zhang et al., 2017), ACE (Walker et al., 2006; Mitchell et al., 2005), CoNLL (Roth and Yih, 2004), ADE (Gurulingappa et al., 2012), WebNLG (Gardent et al., 2017), NYT (Riedel et al., 2010), DocRED (Yao et al., 2019), Re-DocRED (Tan et al., 2022b), and DocGNRE (Li et al., 2023). TACRED, ACE, CoNLL, and ADE provide sentence-level annotations from newswire, web, and medical texts, while WebNLG and NYT are knowledge base-driven (DBpedia Auer et al., 2007, Freebase Bollacker et al., 2008) but are also sentence-level. In contrast, DocRED, Re-DocRED, and DocGNRE are Wikidata-based (Vrandečić and Krötzsch, 2014) with document-level annotations. In this paper, we focus on mitigating FN gaps in Re-DocRED and DocGNRE.

With the advent of LLMs such as ChatGPT (OpenAI, 2022) and the Llama series (Touvron et al., 2023a; Touvron et al., 2023b; Grattafiori et al., 2024), several studies explore zero- or few-shot JERE annotation. Approaches include single-prompt, two-step, or hybrid strategies. Sivarajkumar et al. (2023) evaluate prompting on GPT-3.5 (OpenAI, 2023), Google BARD (Google, 2023), and Llama2 (Touvron et al., 2023b), finding heuristic and chain-of-thought prompts most effective. Jahan et al. (2024) and Zhu et al. (2024) incorporate relation lists in prompts to LLMs including GPT-3.5, GPT-4 (OpenAI et al., 2024), PaLM-2 (Anil et al., 2023), Claude-2 (Anthropic, 2023),

Split	Current number of triplets	Number of additional triplets	% increase
Re-DocRED Train	85392	24810	28.87%
Re-DocRED Dev	17284	4770	27.59%
Re-DocRED Test	17448	4844	27.76%
DocGNRE Test	19526	5407	27.69%

Table 1: Additional triplets introduced to Re-DocRED (Tan et al., 2022b) and DocGNRE (Li et al., 2023) via rule-based expansion.

and Llama-2. Extensions by Brokman (2024) and Piano et al. (2024) add in-context examples. Piano et al. (2024) use Llama3-70B for generation and smaller models (Llama3-8B, Phi3-mini, Phi3-small (Abdin et al., 2024), Anita-8B (Polignano et al., 2024)) for majority voting to improve accuracy.

Two-step prompt methods first perform Named Entity Recognition (NER) and then Relation Extraction (RE) to generate triplets. Sainz et al. (2022) reformulate RE as a Textual Entailment (TE) task with label verbalization, using an NER model to identify entities before a TE model assesses relations. Cai et al. (2024) extract entities via `spaCy` software, classify them with synonyms, and then apply a RE model. Recently, Wu et al. (2025) use GPT-3.5-turbo and GPT-4 to first extract entities and then generate triplets from them.

Hybrid approaches include Bi et al. (2024) and Park et al. (2024), who leverage GPT-3.5 code generation to extract triplets. Wei et al. (2024) first generate relations from text and then identify entity pairs for each relation. Li et al. (2024) adopt a two-step process: generating and cleaning triplets, followed by formatting the output.

Training-free RE methods include Li et al. (2023), Gao et al. (2024), and Möller and Usbeck (2024). Li et al. (2023) use GPT-3.5 to generate triplets from entities and context, then T5-XXL (Raffel et al., 2019) for textual entailment filtering within a predefined relation set. Gao et al. (2024) prompt local LLMs for relations and employ data programming to reduce false positives. Möller and Usbeck (2024) enhance relation classification by integrating entity types and relation definitions.

Existing training-free annotation methods often suffer from hallucinations, schema violations, and limited explainability, particularly in LLM-based approaches (Li et al., 2023). Our method addresses these issues by using open-source LLMs to provide explicit reasoning at each step, incorporating fine-grained entity-level constraints derived from relation definitions. Unlike prior works, our pipeline enables traceability, debuggability, and user-specified constraints. It validates candidate triplets step-by-step for each relation rather than generating them directly, ensuring adherence to relation schemas and reducing hallucinations.

4 Methodology

4.1 Problem Definition

A triplet is defined as (subject, r_i, object) or $(e_{head}, r_i, e_{tail})$ where e_{head} and e_{tail} represent subject and object entities respectively, with relation r_i. In our task formulation, we consider a document D consisting of M sentences where $M \geq 1$. Given this document D and a set of entity-entity relations R, our objective is to predict a set of triplets $(e_{head}, r_i, e_{tail})$ such that $e_{head}, e_{tail} \in E$ where E is set of entities that can be found in document D and $r_i \in R$.

Figure 3 shows our proposed framework, SiftingLogic. We first retrieve relevant relations R^* and entities E from the context in Stage 0. This is followed by generating initial triplets T_1 in Stage 1 based on relation head–tail types (Appendix F). In Stage 2, we validate T_1 with a small LLM using entity constraints and relation verbalizations, producing triplets T_2. In Stage 3, we employ a large LLM which performs stricter validation to produce the final triplets T_3, marking the end of our SiftingLogic pipeline. In addition, we include a post-processing step which applies human verification to triplets T_3 and rule-based expansion with inverse and co-occurring relations (Appendices D, E), yielding the final human-verified triplets T_4 to further reduce false negatives.

The following sections describe the different stages of our pipeline in more detail.

4.2 Stage 0: Relevant Relation Retrieval and Named Entity Recognition via small LLM

Using the entity-entity relations set R and context text as inputs, our pipeline first retrieves a context-relevant relation set R^* using a small LLM, prompted with the context and Wikidata relation definitions (Vrandečić and Krötzsch, 2014). The resulting relation set R^* consists of reasoning for each relation. Entities are then extracted in two

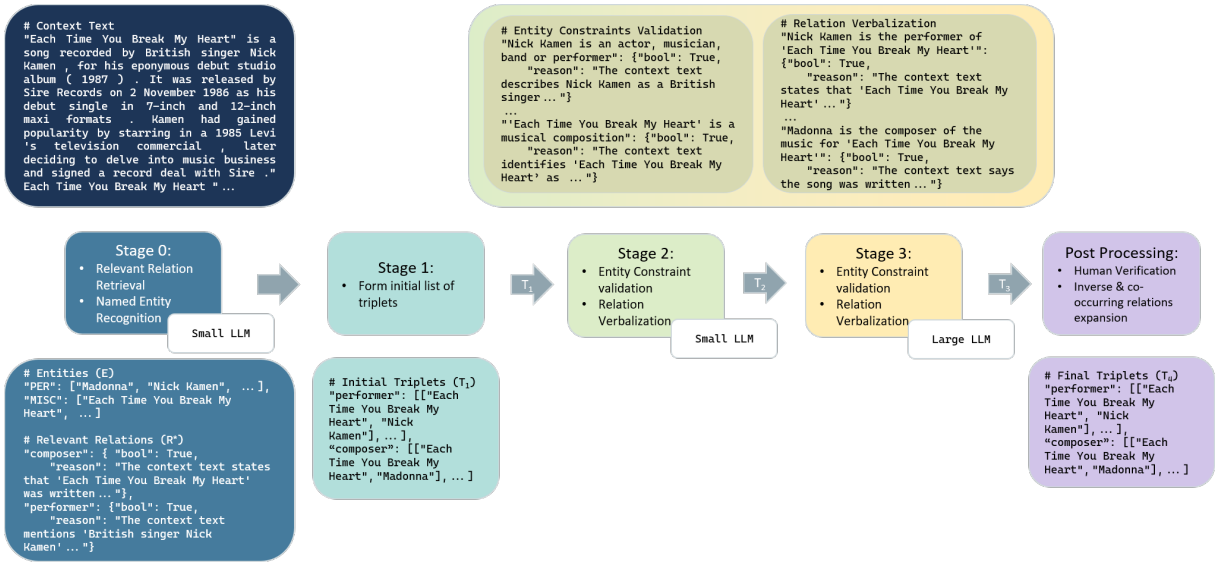

Figure 3: Flow of the SiftingLogic pipeline with post-processing. Stages 0–3 form SiftingLogic: Stage 0 uses a small LLM for relation retrieval and NER; Stage 1 generates initial triplets; Stage 2 validates them with a small LLM using entity constraints and relation verbalizations; Stage 3 applies a large LLM for refined validation. Post-processing then applies human verification and rule-based expansion via inverse and co-occurring relations to produce the final verified triplet set.

steps: the LLM is prompted with DocRED (Yao et al., 2019) entity type definitions followed by the context, and then re-prompted with relation reasoning to improve relevance. Results from both steps are deduplicated and merged, generating the entities set E. Full prompt details are in Appendix A.

4.3 Stage 1: Construction of Initial Candidate Triplets via Relation - Entity Type Pair Compatibility

This stage takes the relevant relations set R^* and entities set E extracted from the previous stage (Section 4.2) as inputs. As also shown in Figure 3, entities are categorized into different types such as PER for person, ORG for organizations, and, so on (refer to Appendix A.2 for more details). Here, we generate the initial candidate triplet set T_1 by pairing entities with relations according to our curated mapping of each relation to their compatible head–tail entity types (Appendix F).

4.4 Stage 2: Entity Constraint and Relation Verbalization Validation via small LLM

In this stage, initial triplets extracted from previous stage (Section 4.3), T_1, are first validated using entity-level constraints derived from relation definitions (Appendix G). For example, for `composer` (definition: person(s) who wrote the music [for lyricist, use "lyrics by" (P676) (Wikidata-Contributors, 2025)]), the entity constraints are:

- `[subject] is a musical composition: True`
- `[object] is a person: True`
- `[object] is a lyricist: False`

These constraints, sourced from Wikidata (Vrandečić and Krötzsch, 2014) definitions, provide finer granularity than prior work (Sainz et al., 2022) which uses entity types. Our validation process uses a two-step multi-turn LLM prompt (Li et al., 2024), discarding any triplet that violates any constraint to form triplets T_{21}.

Next, triplets in T_{21} are validated via relation verbalizations (Appendix H), following Sainz et al. (2022) and Li et al. (2023), using the same two-step prompt procedure. Using relation `composer`, an example verbalization would be:

- `composer: [object] is the composer of the music for [subject]`

Triplets labeled `False` are removed, yielding the revised triplet set T_2.

In this stage, we employ a small LLM that allows our pipeline to remain both precise and computationally efficient. Full prompt details are provided in Appendix B for reference.

4.5 Stage 3: Entity Constraint and Relation Verbalization Validation via large LLM

Given triplets T_2 obtained from the previous stage (Section 4.4), in this stage, a large LLM re-validates

	% triplets
5/5	79.41%
4/5	90.10%
3/5	93.92%
2/5	96.78%
1/5	100%

Table 2: Percentage of Sifting Logic (SL) triplets with ≥n annotators agreeing, restricted to triplets marked correct by at least one annotator

Split	Number of triplets	% increase
Train$_{Re}$	110202	28.87%
Dev$_{Re}$	22054	27.59%
Test$_{Re}$	22292	27.76%
Test$_{GN}$	24933	27.69%
Test$_{GN+SL}$	29268	49.89%
Test$_{RED(zh)}$	1391	109.80%

Table 3: Number of examples in Re^2-DocRED splits

entity-level constraints and relation verbalizations using the same prompts, which improves precision by reducing hallucination-induced false positives. This creates the final triplet set T_3. Across all stages of SiftingLogic, intermediate outputs are available for review or refinement, which ensures that our pipeline remains model-agnostic, modular and adaptable to future LLMs.

4.6 Post Processing

Following prior works (Yao et al., 2019; Tan et al., 2022b; Li et al., 2023), human annotators verify the triplets T_3 generated from our pipeline to form the final human-verified triplets T_4. The details are as follows:

1. Five university-educated, English-proficient annotators perform the triplet verification task (annotation scheme in Appendix C).
2. Annotators validate each triplet by reviewing the context with highlighted entities, the triplet itself, and the relation definition, marking it correct or incorrect.
3. Each triplet is annotated by five annotators, yielding a Krippendorff's Alpha-Reliability (α) coefficient of 0.88, indicating a high level of inter-annotator reliability.
4. Only triplets rated correct by all five annotators (unanimous agreement) are retained as true positives; the rest are treated as false positives. Table 2 shows that even with the strictest criterion, ∼80% of triplets marked correct by at least one annotator are retained, indicating minimal exclusion of potentially valid triplets.

Using our SiftingLogic with human verification and rule-based expansion (Appendices D, E), we reduce false negatives, and enhance Re-DocRED and DocGNRE splits to form our Re^2-DocRED splits, compositions of which are described below:

- {**Train**$_{Re}$, **Dev**$_{Re}$, **Test**$_{Re}$}: Train, Dev and Test splits of our Re^2-DocRED dataset, obtained by revising the original splits of Re-DocRED dataset, {**Train**$_{Re-O}$, **Dev**$_{Re-O}$, **Test**$_{Re-O}$}, with rule-based expansion.
- {**Test**$_{GN}$, **Test**$_{GN+SL}$}: Test splits of our Re^2-DocRED dataset obtained by revising the original split of DocGNRE dataset, **Test**$_{GN-O}$, with rule-based expansion (**Test**$_{GN}$) and addition of SiftingLogic triplets followed by rule-based expansion (**Test**$_{GN+SL}$) respectively.

Note that, **Test**$_{GN+SL}$ is created by applying our SiftingLogic to the DocGNRE test set (Li et al., 2023), followed by human verification. Verified triplets are then deduplicated from DocGNRE test set using gpt-4o-mini (OpenAI, 2024) using the following strict criteria in the prompt:

- Triplet is not an exact duplicate of an existing triplet
- Triplet does not express the same meaning as any existing triplets — including paraphrases, synonyms, abbreviations, or formatting differences
- Triplet contributes a new fact or relation not already captured in existing triplets

The resultant deduplicated triplets are then merged with DocGNRE test set and expanded via our refined inverse and co-occurring relation rules (Appendices D, E).

Similarly, for REDFM (Huguet Cabot et al., 2023), we revise the original Mandarin test set, **Test**$_{RED(zh)-O}$ by applying similar procedure with three independent, university-educated Mandarin-proficient annotators, without rule-based expansion, forming **Test**$_{RED(zh)}$[1][2] . Improvement statistics for our Re^2-DocRED splits and our revised REDFM Mandarin test split are reported in Table 3, with detailed relation distributions in Appendix K.

[1]We did not apply rule-based expansion to REDFM as the authors explicitly note that inverse relations are collapsed (Huguet Cabot et al., 2023 Section 3.1) by design.

[2]We use existing entities from **Test**$_{RED(zh)-O}$ as input to SiftingLogic to create **Test**$_{RED(zh)}$.

	$\textbf{Train}_{Re-O}$	$\textbf{Train}_{Re}$
	P / R	P / R
$\textbf{Test}_{Re-O}$	65.58 / 48.29	47.54 / **49.95**
$\textbf{Test}_{GN-O}$	66.33 / 43.51	49.36 / **46.02**
$\textbf{Test}_{Re}$	66.58 / 38.24	65.92 / **51.75**
$\textbf{Test}_{GN}$	67.36 / 34.49	66.72 / **46.67**
$\textbf{Test}_{GN+SL}$	68.55 / 29.82	67.08 / **39.93**

Table 4: AutoRE (Xue et al., 2024) Precision (P) and Recall (R) results across test and training sets. $\textbf{Test}_{Re-O}$ and $\textbf{Test}_{GN-O}$ are original test sets of Re-DocRED and DocGNRE, while $\textbf{Train}_{Re-O}$ is the original training set of Re-DocRED. Top scores are in **bold**.

	$\textbf{Train}_{Re-O}$	$\textbf{Train}_{Re}$
	P / R	P / R
$\textbf{Test}_{Re-O}$	57.36 / 42.85	44.87 / **43.57**
$\textbf{Test}_{GN-O}$	57.57 / 38.43	45.95 / **39.87**
$\textbf{Test}_{Re}$	57.61 / 33.68	58.28 / **44.30**
$\textbf{Test}_{GN}$	57.85 / 30.24	58.50 / **39.76**
$\textbf{Test}_{GN+SL}$	58.60 / 26.00	59.23 / **34.05**

Table 5: TaG (Zhang et al., 2023) Precision (P) and Recall (R) results across test and training sets. Top scores are in **bold**.

	$\textbf{Train}_{Re-O}$	$\textbf{Train}_{Re}$
	P / R	P / R
$\textbf{Test}_{Re-O}$	57.63 / **45.33**	48.02 / 43.42
$\textbf{Test}_{GN-O}$	54.12 / 40.48	47.12 / **40.51**
$\textbf{Test}_{Re}$	52.12 / 37.63	57.32 / **47.57**
$\textbf{Test}_{GN}$	48.06 / 33.59	53.12 / **42.68**
$\textbf{Test}_{GN+SL}$	43.12 / 29.32	47.57 / **37.18**

Table 6: REBEL (Huguet Cabot and Navigli, 2021) Precision (P) and Recall (R) results across test and training sets. Top scores are in **bold**.

	P / R
$\textbf{Test}_{RED(zh)-O}$	41.14 / 44.80
$\textbf{Test}_{RED(zh)}$	51.80 / 22.96

Table 7: mREBEL_{32}^{T} (Huguet Cabot et al., 2023) Precision (P) and Recall (R) results. $\textbf{Test}_{RED(zh)-O}$ is the original REDFM Mandarin test set.

5 Experimental Results

This section first discusses our chosen JERE models and their evaluation on original test splits of Re-DocRED (Tan et al., 2022b) and DocGNRE (Li et al., 2023), $\textbf{Test}_{Re-O}$ and $\textbf{Test}_{GN-O}$, respectively. We further evaluate on their corresponding revised test splits, $\textbf{Test}_{Re}$, $\textbf{Test}_{GN}$ and $\textbf{Test}_{GN+SL}$, present in our Re^2-DocRED. We then move to evaluating the original REDFM (Huguet Cabot et al., 2023) Mandarin test set, $\textbf{Test}_{RED(zh)-O}$, and our corresponding revised split, $\textbf{Test}_{RED(zh)}$.

We then detail the LLMs used within SiftingLogic and report inference precision and recall on our revised splits $\textbf{Test}_{GN+SL}$ and $\textbf{Test}_{RED(zh)}$. Finally, we present an ablation study analyzing each SiftingLogic stage's effect on FP reduction and the individual contributions of SiftingLogic and rule-based expansion in mitigating FN.

5.1 JERE Evaluation on Original and Our Revised Test Splits

For dataset evaluation, we select AutoRE (Xue et al., 2024), a state-of-the-art document-level JERE model, TaG (Zhang et al., 2023) and REBEL (Huguet Cabot and Navigli, 2021). All are trained with default settings on both original Re-DocRED (Tan et al., 2022b) training set, $\textbf{Train}_{Re-O}$, and our revised training set $\textbf{Train}_{Re}$, then evaluated on Re-DocRED, DocGNRE, and Re^2-DocRED test splits. To our knowledge, this represents the first JERE evaluation on DocGNRE. Results are summarized in Tables 4, 5 and 6.

As shown in Table 4, recall on Re-DocRED train set ($\textbf{Train}_{Re-O}$) steadily decreases as test sets become more complete (i.e., with fewer FNs), reflecting persistent FN issue in the training data. With our expanded Re^2-DocRED training set, $\textbf{Train}_{Re}$, which mitigates FN through rule-based expansion, average recall across $\textbf{Test}_{Re}$, $\textbf{Test}_{GN}$, and $\textbf{Test}_{GN+SL}$ improves by 11.93 points. Notably, models trained on $\textbf{Train}_{Re}$ achieve consistently higher recall across both existing and revised test sets, underscoring that reducing FN in training data yields broader recall gains. The lower precision of the model trained on $\textbf{Train}_{Re}$ on the original test sets ($\textbf{Test}_{Re-O}$, $\textbf{Test}_{GN-O}$) is attributable to FN, as noted by the recovery of precision on the revised benchmarks ($\textbf{Test}_{Re}$, $\textbf{Test}_{GN}$). Similar trends are observed for TaG and REBEL, as shown in Tables 5 and 6. Additional analysis on long-tail relations is provided in Appendix J.

$\textbf{Test}_{GN+SL}$ represents the most challenging evaluation set in Table 4, exhibiting the lowest recall (29.82%) when models are trained on $\textbf{Train}_{Re-O}$. This pronounced drop highlights the effectiveness of our pipeline in producing more complete triplet annotations. Through SiftingLogic and rule-based expansion, $\textbf{Test}_{GN+SL}$ offers a stricter benchmark that reveals model limitations and encourages training approaches that address FN issues.

For the REDFM (Huguet Cabot et al., 2023) Mandarin test sets evaluations, Table 7 shows that on our revised test set $\textbf{Test}_{RED(zh)}$, mREBEL_{32}^{T}

	Test$_{GN-O}$+SL (P)	**Test**$_{GN-O}$+SL (R)	**Test**$_{GN+SL}$ (R)	**Test**$_{GN-O}$ (R)
GPT-3.5 with NLI with relation description (Li et al., 2023)	-	-	-	15.32
SiftingLogic (SL)	32.57	35.52	-	19.01
SiftingLogic (SL) + rule	-	-	49.39	23.40

Table 8: Precision (P) and Recall (R) of SiftingLogic (SL) across multiple evaluation settings. Note, **Test**$_{GN-O}$+SL represents the original DocGNRE test set, **Test**$_{GN-O}$, supplemented with human-verified SiftingLogic (SL) outputs. Results are reported without and with rule-based expansion (rule). A triplet is considered a true positive (TP) if it achieves unanimous annotator agreement or is present in an existing dataset.

	Test$_{RED(zh)}$ (P)	**Test**$_{RED(zh)}$ (R)	**Test**$_{RED(zh)-O}$ (R)
SiftingLogic (SL)	43.40	77.35	52.49

Table 9: Precision (P) and Recall (R) of SiftingLogic (SL) evaluated on original and our revised Mandarin test set in REDFM, **Test**$_{RED(zh)-O}$ and **Test**$_{RED(zh)}$, respectively. A triplet is considered a true positive (TP) if it achieves unanimous annotator agreement or is present in an existing dataset.

	Stage 1	**Stage 2.1**	**Stage 2.2**	**Stage 3.1**	**Stage 3.2**	**Post-Processing**
located in the administrative territorial entity	100	16.18	5.65	4.52	3.24	1.79
location	100	48.01	14.77	7.11	4.42	1.90
country	100	29.13	10.44	6.54	4.99	3.12
has part	100	53.75	12.17	9.57	4.87	1.89
part of	100	53.92	13.17	10.3	5.27	1.90
point in time	100	88.51	18.01	16.43	6.51	2.95
notable work	100	42.97	9.95	7.65	3.21	1.83
participant	100	27.45	10.93	3.93	2.45	1.14
participant of	100	17.89	7.22	2.69	1.67	1.00
all	100	31.4	7.43	5.09	2.34	1.24

Table 10: Triplet retention (in %) at each stage of SiftingLogic on selected major inverse and co-occurring relations.

achieves higher precision (41.14 → 51.80) but lower recall (44.80 → 22.96). This reflects the revised set's more complete triplet coverage, mitigating false negatives while revealing model overestimation under the original benchmark.

5.2 SiftingLogic with Rule-based Expansion Inference Performance

For SiftingLogic, we evaluate several small LLMs – including Llama3.1-8B (MetaAI, 2024a), Qwen2.5-7B (QwenTeam, 2024), AllenAI-Tulu3-8B (Allen Institute for AI, 2024), InternLM3-8B (ShanghaiAILaboratory, 2025), and LGAI-Exaone3.5-8B (LGAIResearch et al., 2024) – and found Llama3.1-8B best suited for structured JSON output. Hence, we pick Llama3.3-70B (MetaAI, 2024b) as the large LLM in our pipeline for prompt consistency. Both models are Q6_K-quantized and run via `llama-cpp-python` on an NVIDIA A100 GPU.

Referring to Table 8, on **Test**$_{GN-O}$+SL (DocGNRE original test set **Test**$_{GN-O}$ supplemented with human-verified SiftingLogic (SL) output), SiftingLogic attains a precision of 32.57% and recall of 35.52%. These results indicate that even without post-processing, our pipeline is capable of generating reliable set of plausible triplets under the strict criterion of unanimous annotator agreement. With post-processing through human verification and rule-based expansion, recall reaches 49.39% on **Test**$_{GN+SL}$ demonstrating the overall pipeline's ability to uncover correct triplets. As shown in Table 9, SiftingLogic further generalizes to REDFM, achieving 43.40% precision and 77.35% recall on **Test**$_{RED(zh)}$. This cross-linguistic performance highlights the robustness and adaptability of our method across diverse datasets and languages.

Table 8 compares SiftingLogic's verified triplet overlap with the original DocGNRE test set (**Test**$_{GN-O}$) against the GPT-3.5 based solution from Li et al. (2023) as baseline[3] and shows that SiftingLogic achieves an ∼3-point higher overlap (19.01% versus 15.32%) in comparison. Across Tables 3 and 8, we further show that SiftingLogic

[3]Baseline method (indicated as "GPT-3.5 with NLI with relation description" in Table 8) generates triplets using GPT-3.5, construct hypothesis triplets using the entities in GPT output and validates them via Natural Language Inference (NLI) with relation descriptions.

	Test$_{GN-O}$	**Test**$_{GN-O}$ + rule	**Test**$_{GN-O}$+SL	**Test**$_{GN+SL}$
located in the administrative territorial entity	1	1.0789	1.0155	1.1324
location	1	5.2353	1.1876	5.4674
has part	1	1.1466	1.7828	2.8724
part of	1	1.3627	1.8381	3.3770
point in time	1	11.4730	1.9054	11.8784
performer	1	1.2500	1.0223	1.3304
participant	1	1.1939	1.2319	1.6540
participant of	1	1.0064	1.1731	1.4263
member of	1	1.2175	1.0056	1.2373
all	1	1.2769	1.1146	1.4989

Table 11: Triplet count improvement ratios on original DocGNRE test **Test**$_{GN-O}$ for baseline, rule-based expansion (rule), SiftingLogic (SL), and their combination. Values are normalized to the baseline (i.e., **Test**$_{GN-O}$ = 1).

matches 3,711 triplets (19.01%)[4] from **Test**$_{GN-O}$ while adding 9,742 new triplets, a 49.89% increase over the original set. This shows that SiftingLogic triplets are complimentary to existing dataset. For the REDFM Mandarin original test set (**Test**$_{RED(zh)-O}$), Table 9 shows overlap is measured at 52.49% (348 triplets), with 728 additional triplets, corresponding to a 109.80% increase. This highlights the serious FN issue plaguing the current test set.

5.3 Ablation

5.3.1 SiftingLogic stages

To address the inherent sparsity of RE, we evaluate SiftingLogic's candidate filtering across stages to isolate plausible pairs. Table 10 reports retention for major inverse and co-occurring relations alongside overall averages.

Entity-constraint validation using a small LLM in Stage 2 (Section 4.4), addresses sparsity by significantly narrowing the set of candidate triplets—for example, *located in the administrative territorial entity* drops to 16.18% and *participant of* to 17.89%. On average, 70% of candidates are pruned in this step, allowing subsequent stages to focus on a smaller, more plausible set efficiently.

By Stage 3 (large LLM validation; Section 4.5), fewer than 8% of original candidates remain. Early small LLM stages focus attention on the most plausible triplets, reducing computational burden and enabling efficient, high-quality large LLM processing. Ultimately, only ∼2% of triplets reach post-processing (Section 4.6), with overall retention at 1.24%. Relations like *participant* and *participant of* fall below 1.2%, showing that prior filtering efficiently reduces human review workload.

The ablation study shows that each stage of the pipeline is vital for handling RE's inherent sparsity. Entity-constraint validation filters unlikely triplets early, while relation-verbalization and large-LLM verification further refine results, reducing human review. The staged filtering design—small LLM followed by large LLM—focuses resources on the most promising candidates, demonstrating that every component meaningfully enhances both efficiency and effectiveness.

5.3.2 SiftingLogic vs Rule-based Expansion

Table 11 shows that SiftingLogic and rule-based expansion increase extraction by 11.46% and 27.69%, respectively. Combined, they yield a 49.89% gain on the DocGNRE test set, indicating synergistic coverage. Further ablation disentangling their roles as silver labels is provided in Appendix I.

6 Conclusions

In this work, we demonstrate that existing document-level RE datasets—DocRED (Yao et al., 2019), Re-DocRED (Tan et al., 2022b), and DocGNRE (Li et al., 2023)—remain incomplete. To address this, we enrich Re-DocRED and DocGNRE via improved relation rules and **SiftingLogic**, a novel training-free, LLM-based pipeline utilizing explainable reasoning and Wikidata-derived **entity-level constraints**. Our **Re^2-DocRED** dataset adds 9,742 verified triplets (+49.89%) to the DocGNRE test set and 728 (+109.80%) to the REDFM (Huguet Cabot et al., 2023) Mandarin set highlighting SiftingLogic's strong cross-lingual generalizability and enabling models to outperform prior baselines. We release the datasets to facilitate reliable JERE evaluation.

[4]Using entities from Test$_{GN-O}$, based on output from Stage 2, SiftingLogic achieved an overlap of 8172 triplets (41.85%) with Test$_{GN-O}$.

Limitations

Our framework, SiftingLogic, relies on LLMs to generate structured outputs at each stage. In practice, there were instances where the LLMs generated ill-formatted outputs. We used the same small LLM to correct the format of the output given the ill-formatted output as input. For the entity recognition stage, we note that the entities generated via the small LLM at times differ in granularity or formatting from the entities in DocGNRE (Li et al., 2023). In order to merge the human-verified generated triplets with DocGNRE, we used gpt-4o-mini (OpenAI, 2024) to identify duplicate triplets and entities. On the novel entity level constraints derived from DocRED relation definitions introduced in this work, the constraints are manually crafted based on the relation definitions. A method to semi or fully automate the generation of these constraints would smooth the process of applying this pipeline in other datasets. Finally, though the pipeline achieved a respectable precision and recall score, human verification of triplets are still necessary and time consuming. Exploration of automated checking via GPT and its error tradeoffs would minimise the time taken for human verification.

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, and 110 others. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *Preprint*, arXiv:2404.14219.

Allen Institute for AI. 2024. Tülu 3 opens language model post-training up to more tasks and more people. https://allenai.org/blog/tulu-3. Accessed: 2025-05-18.

Rohan Anil, Andrew M. Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, Eric Chu, Jonathan H. Clark, Laurent El Shafey, Yanping Huang, Kathy Meier-Hellstern, Gaurav Mishra, Erica Moreira, Mark Omernick, Kevin Robinson, and 109 others. 2023. Palm 2 technical report. *Preprint*, arXiv:2305.10403.

Anthropic. 2023. Claude 2. https://www.anthropic.com/news/claude-2. Accessed: 2025-03-20.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The Semantic Web*, pages 722–735, Berlin, Heidelberg. Springer Berlin Heidelberg.

Zhen Bi, Jing Chen, Yinuo Jiang, Feiyu Xiong, Wei Guo, Huajun Chen, and Ningyu Zhang. 2024. Codekgc: Code language model for generative knowledge graph construction. *Preprint*, arXiv:2304.09048.

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250.

Aviv Brokman. 2024. *On Generative Models and Joint Architectures for Document-level Relation Extraction*. University of Kentucky.

Erica Cai, Olga Simek, Benjamin A. Miller, Danielle Sullivan-Pao, Evan Young, and Christopher L. Smith. 2024. From low resource information extraction to identifying influential nodes in knowledge graphs. *Preprint*, arXiv:2401.04915.

Chufan Gao, Xulin Fan, Jimeng Sun, and Xuan Wang. 2024. PromptRE: Weakly-supervised document-level relation extraction via prompting-based data programming. In *Proceedings of the 1st Workshop on Towards Knowledgeable Language Models (KnowLLM 2024)*, pages 132–145, Bangkok, Thailand. Association for Computational Linguistics.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating training corpora for NLG micro-planners. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 179–188, Vancouver, Canada. Association for Computational Linguistics.

Google. 2023. Bard ai. https://bard.google.com. Accessed: 2025-03-20.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 542 others. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

Harsha Gurulingappa, Abdul Mateen Rajput, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. 2012. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of biomedical informatics*, 45(5):885–892.

Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Pere-Lluís Huguet Cabot, Simone Tedeschi, Axel-Cyrille Ngonga Ngomo, and Roberto Navigli. 2023. REDfm: a filtered and multilingual relation extraction dataset. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4326–4343, Toronto, Canada. Association for Computational Linguistics.

Israt Jahan, Md Tahmid Rahman Laskar, Chun Peng, and Jimmy Xiangji Huang. 2024. A comprehensive evaluation of large language models on benchmark biomedical text processing tasks. *Computers in Biology and Medicine*, 171:108189.

LGAIResearch, Soyoung An, Kyunghoon Bae, Eunbi Choi, Kibong Choi, Stanley Jungkyu Choi, Seokhee Hong, Junwon Hwang, Hyojin Jeon, Gerrard Jeongwon Jo, Hyunjik Jo, Jiyeon Jung, Yountae Jung, Hyosang Kim, Joonkee Kim, Seonghwan Kim, Soyeon Kim, Sunkyoung Kim, Yireun Kim, and 14 others. 2024. Exaone 3.5: Series of large language models for real-world use cases. *Preprint*, arXiv:2412.04862.

Junpeng Li, Zixia Jia, and Zilong Zheng. 2023. Semi-automatic data enhancement for document-level relation extraction with distant supervision from large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5495–5505, Singapore. Association for Computational Linguistics.

Yinghao Li, Rampi Ramprasad, and Chao Zhang. 2024. A simple but effective approach to improve structured language model output for information extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5133–5148, Miami, Florida, USA. Association for Computational Linguistics.

MetaAI. 2024a. Introducing llama 3.1: Our most capable models to date. https://ai.meta.com/blog/meta-llama-3-1/. Accessed: 2025-05-18.

MetaAI. 2024b. Llama 3.3 model card and prompt formats. https://www.llama.com/docs/model-cards-and-prompt-formats/llama3_3/. Accessed: 2025-05-18.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Alexis Mitchell and 1 others. 2005. Ace 2004 multilingual training corpus ldc2005t09. Web Download. LDC Catalog No. LDC2005T09.

Cedric Möller and Ricardo Usbeck. 2024. Incorporating type information into zero-shot relation extraction.

OpenAI. 2022. Chatgpt. Accessed: 2025-03-20.

OpenAI. 2023. Gpt-3.5. https://openai.com. Accessed: 2025-03-20.

OpenAI. 2024. Gpt-4o mini: advancing cost-efficient intelligence. Accessed: 2025-05-19.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Chaelim Park, Hayoung Lee, and Ok-ran Jeong. 2024. Leveraging medical knowledge graphs and large language models for enhanced mental disorder information extraction. *Future Internet*, 16(8).

Leonardo Piano, Alessia Pisu, Sandro Gabriele Tiddia, Salvatore Carta, Alessandro Giuliani, and Livio Pompianu. 2024. Llimoniie: Large language instructed model for open named italian information extraction.

Marco Polignano, Pierpaolo Basile, and Giovanni Semeraro. 2024. Advanced natural-based interaction for the italian language: Llamantino-3-anita. *Preprint*, arXiv:2405.07101.

QwenTeam. 2024. Qwen2.5: A party of foundation models! https://qwenlm.github.io/blog/qwen2.5/. Accessed: 2025-05-18.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, abs/1910.10683.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III 21*, pages 148–163. Springer.

Dan Roth and Wen-tau Yih. 2004. A linear programming formulation for global inference in natural language tasks. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004*, pages 1–8, Boston, Massachusetts, USA. Association for Computational Linguistics.

Oscar Sainz, Haoling Qiu, Oier Lopez de Lacalle, Eneko Agirre, and Bonan Min. 2022. ZS4IE: A toolkit for zero-shot information extraction with simple verbalizations. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations*, pages 27–38, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

ShanghaiAILaboratory. 2025. Internlm: Open-source large language models. https://github.com/InternLM/InternLM. Accessed: 2025-05-18.

Sonish Sivarajkumar, Mark Kelley, Alyssa Samolyk-Mazzanti, Shyam Visweswaran, and Yanshan Wang. 2023. An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing. *Preprint*, arXiv:2309.08008.

Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. 2022a. Document-level relation extraction with adaptive focal loss and knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1672–1681, Dublin, Ireland. Association for Computational Linguistics.

Qingyu Tan, Lu Xu, Lidong Bing, Hwee Tou Ng, and Sharifah Mahani Aljunied. 2022b. Revisiting DocRED - addressing the false negative problem in relation extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8472–8487, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023a. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, and 49 others. 2023b. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. Ace 2005 multilingual training corpus ldc2006t06, 2006. *URL https://catalog. ldc. upenn. edu/LDC2006T06*.

Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, Yong Jiang, and Wenjuan Han. 2024. Chatie: Zero-shot information extraction via chatting with chatgpt. *Preprint*, arXiv:2302.10205.

Wikidata-Contributors. 2025. Property:p86. Accessed: 2025-04-09.

Ling-I Wu, Yuxin Su, and Guoqiang Li. 2025. Zero-shot construction of chinese medical knowledge graph with gpt-3.5-turbo and gpt-4. *ACM Transactions on Management Information Systems*, 16(2):1–17.

Chaojun Xiao, Yuan Yao, Ruobing Xie, Xu Han, Zhiyuan Liu, Maosong Sun, Fen Lin, and Leyu Lin. 2020. Denoising relation extraction from document-level distant supervision. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3683–3688, Online. Association for Computational Linguistics.

Lilong Xue, Dan Zhang, Yuxiao Dong, and Jie Tang. 2024. AutoRE: Document-level relation extraction with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 211–220, Bangkok, Thailand. Association for Computational Linguistics.

Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 764–777, Florence, Italy. Association for Computational Linguistics.

Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021. Document-level relation extraction as semantic segmentation. *CoRR*, abs/2106.03618.

Ruoyu Zhang, Yanzeng Li, and Lei Zou. 2023. A novel table-to-graph generation approach for document-level joint entity and relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10853–10865.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2020. Document-level relation extraction with adaptive thresholding and localized context pooling. *CoRR*, abs/2010.11304.

Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2024. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *Preprint*, arXiv:2305.13168.

A Stage 0 Prompts

A.1 Stage 0: Relevant Relation Retrieval - LLM User Prompt

```
|im_user|
# Context Text
{context_text}

# Hypothesis:
Given the following relation definition:

- '{relation}': {relation_definition}

The relation '{relation}' is mentioned in the
context.

# Task:
Based on the context text, is the hypothesis
<[True]> or <[False]>? Provide a short reasoning.
```

Figure 4: Stage 0.1: Relevant Relation Retrieval Prompt

A.2 Stage 0: Named Entity Recognition - LLM User Prompt 0

```
# Entity Type Definitions
PER:
- People, including fictional
ORG:
- Companies, universities, institutions,
political or religious groups, etc.
LOC:
- Geographically defined locations, in-
cluding mountains, waters, etc.
- Politically defined locations, includ-
ing countries, cities, states, streets,
etc.
- Facilities, including buildings, muse-
ums, stadiums, hospitals, factories, air-
ports, etc.
TIME:
- Absolute or relative dates or periods.
NUM:
- Percents, money, quantities
MISC:
- Products, including vehicles, weapons,
etc.
- Events, including elections, battles,
sporting events, etc.
- Laws, cases, languages, etc.

# Task:
Based on the context text and entity
type definitions, extract all the enti-
ties from the context text according to
the entity types in a Python dict.
An example format is

{
    "PER": [...],
    "ORG": [...],
    "LOC": [...],
    "TIME": [...],
    "NUM": [...],
    "MISC": [...]
}

Return only the python dictionary.
Each value in the dictionary must be a
python list.
Each list can only contain python string.
Every string in the list needs to be ter-
minated correctly.
If the list is empty, return [].
The dictionary ‘{‘ needs to be closed cor-
rectly with ‘}‘.
Dictionary MUST NOT include python com-
ments.

# Context Text
The Battle of Stalingrad, fought between
July 17, 1942, and February 2, 1943, was a
pivotal confrontation during World War II
between the Soviet Union and Nazi Germany
over control of Stalingrad, now known as
Volgograd.  The battle resulted in one
of the deadliest encounters in military
history, with an estimated 2 million ca-
sualties. The Wehrmacht initially gained
ground, but the Soviet Red Army, under the
command of Vasily Chuikov, launched Oper-
ation Uranus, encircling and ultimately
defeating the German 6th Army, led by
Friedrich Paulus.  The battle was marked
by intense urban warfare, with fighting
taking place in key locations such as the
Grain Elevator and Pavlov's House. The So-
viet victory decisively shifted the tide
of the Eastern Front, weakening Germany's
strategic position and bolstering Allied
morale.
```

Figure 5: Stage 0: Named Entity Recognition - LLM User Prompt 0

A.3 Stage 0: Named Entity Recognition - LLM Assistant Response 0

```
{
    "PER": ["Vasily Chuikov", "Friedrich
Paulus"],
```

```
    "ORG": ["Soviet Union", "Nazi Germany", "Wehrmacht", "Red Army"],
    "LOC": ["Stalingrad", "Volgograd", "Grain Elevator", "Pavlov's House"],
    "TIME": ["July 17, 1942", "February 2, 1943"],
    "NUM": ["2 million"],
    "MISC": ["Battle of Stalingrad", "World War II", "Operation Uranus", "6th Army", "Eastern Front"]
}
```

Figure 6: Stage 0: Named Entity Recognition - LLM Assistant Response 0

B Stage 2 Prompts

B.1 Stage 2: Entity Constraint Validation - LLM User Prompt 0

```
# Context Text
{context_text}

# Statements
{list_entity_level_constraints}

# Task
Based on the context text, for each statement, reason whether the statement is True or False. Follow the following format:

{
    "statement content": {
        "bool": True or False,
        "reason": ...},
    ...
}

# Example
## Example Context Text
The Battle of Stalingrad, fought between July 17, 1942, and February 2, 1943, was a pivotal confrontation during World War II between the Soviet Union and Nazi Germany over control of Stalingrad, now known as Volgograd. The battle resulted in one of the deadliest encounters in military history, with an estimated 2 million casualties. The Wehrmacht initially gained ground, but the Soviet Red Army, under the command of Vasily Chuikov, launched Operation Uranus, encircling and ultimately defeating the German 6th Army, led by Friedrich Paulus. The battle was marked by intense urban warfare, with fighting taking place in key locations such as the Grain Elevator and Pavlov's House. The Soviet victory decisively shifted the tide of the Eastern Front, weakening Germany's strategic position and bolstering Allied morale.
## Example statements
{
    "1. Red Army defeated 6th Army.": {
        "bool": True,
        "reason": "The text states that the Soviet Red Army, under the command of Vasily Chuikov, launched Operation Uranus, which led to the encirclement and ultimate defeat of the German 6th Army. This confirms that the Red Army was responsible for defeating the 6th Army."},
    "2. Wehrmacht launched Operation Uranus against the Soviet Union.": {
        "bool": False,
        "reason": "The text specifies that the Soviet Red Army launched Operation Uranus, not the Wehrmacht. The operation was a counteroffensive by the Soviets that led to the encirclement of German forces, not an attack initiated by the Wehrmacht against the Soviet Union."},
    "3. The Soviet Union's Red Army launched the operation known as 'Operation Uranus'.": {
        "bool": True,
        "reason": "According to the text, the Soviet Red Army initiated Operation Uranus, which encircled and ultimately defeated the German 6th Army."}
}
```

Figure 7: Stage 2: Entity Constraint Validation - LLM User Prompt 0

B.2 Stage 2: Entity Constraint Validation - LLM User Prompt 1

```
Given the assistant output, return python dictionary of statements and their True/False without the reason.
Follow the following format:

{
    "statement content": True/False,
    ...
```

```
}


Return only the python dictionary without
any explanations


# Example
## Example input
{
    "1. Red Army defeated 6th Army.": {
        "bool": True,
        "reason": "The text states that
the Soviet Red Army, under the command
of Vasily Chuikov, launched Operation
Uranus, which led to the encirclement and
ultimate defeat of the German 6th Army.
This confirms that the Red Army was re-
sponsible for defeating the 6th Army."},
    "2.  Wehrmacht launched Operation
Uranus against the Soviet Union.": {
        "bool": False,
        "reason":  "The text specifies
that the Soviet Red Army launched Oper-
ation Uranus, not the Wehrmacht. The op-
eration was a counteroffensive by the So-
viets that led to the encirclement of Ger-
man forces, not an attack initiated by the
Wehrmacht against the Soviet Union."},
    "3.  The Soviet Union's Red Army
launched the operation known as 'Opera-
tion Uranus'.": {
        "bool": True,
        "reason": "According to the text,
the Soviet Red Army initiated Operation
Uranus, which encircled and ultimately
defeated the German 6th Army."}
}


## Example output
{
    "1.  The Red Army defeated the 6th
Army.": True,
    "2. The Wehrmacht launched Operation
Uranus against the Soviet Union.": False,
    "3.  The Soviet Union's Red Army
launched the operation known as 'Opera-
tion Uranus'.": True
}
```

Figure 8: Stage 2: Entity Constraint Validation - LLM User Prompt 1

B.3 Stage 2: Verbalized Relation Validation - LLM User Prompt 0

```
# Context Text
{context_text}


# Relations and Definitions
{list_relation_definitions}


# Statements
{list_relation_verbalizations}


# Task
Based on the context text and relations'
definitions, for each statement, reason
whether the statement is True or False.
If no information is provided in the con-
text text or not enough data, the state-
ment is False.
Follow the following format:


{
    "statement content": {
        "bool": True or False,
        "reason": ...},
    ...
}


# Example ## Example Context Text
The Battle of Stalingrad, fought between
July 17, 1942, and February 2, 1943, was a
pivotal confrontation during World War II
between the Soviet Union and Nazi Germany
over control of Stalingrad, now known as
Volgograd.  The battle resulted in one
of the deadliest encounters in military
history, with an estimated 2 million ca-
sualties. The Wehrmacht initially gained
ground, but the Soviet Red Army, under the
command of Vasily Chuikov, launched Oper-
ation Uranus, encircling and ultimately
defeating the German 6th Army, led by
Friedrich Paulus.  The battle was marked
by intense urban warfare, with fighting
taking place in key locations such as the
Grain Elevator and Pavlov's House. The So-
viet victory decisively shifted the tide
of the Eastern Front, weakening Germany's
strategic position and bolstering Allied
morale.
## Example statements
{
```

```
    "1. Red Army defeated 6th Army.": {
        "bool": True,
        "reason": "The text states that
the Soviet Red Army, under the command
of Vasily Chuikov, launched Operation
Uranus, which led to the encirclement and
ultimate defeat of the German 6th Army.
This confirms that the Red Army was re-
sponsible for defeating the 6th Army."},
    "2.  Wehrmacht launched Operation
Uranus against the Soviet Union.": {
        "bool": False,
        "reason":  "The text specifies
that the Soviet Red Army launched Op-
eration Uranus, not the Wehrmacht. The
operation was a counteroffensive by the
Soviets that led to the encirclement of
German forces, not an attack initiated by
the Wehrmacht against the Soviet Union."}
}
```

Figure 9: Stage 2: Verbalized Relation Validation - LLM User Prompt 0

B.4 Stage 2: Verbalized Relation Validation - LLM User Prompt 1

```
Given the assistant output, return python
dictionary of statements and their True
or False without the reason.
Follow the following format:

{
    "statement content": True or False,
    ...
}

Return only the python dictionary without
any explanations

# Example
## Example input
{
    "1. Red Army defeated 6th Army.": {
        "bool": True,
        "reason": "The text states that
the Soviet Red Army, under the command
of Vasily Chuikov, launched Operation
Uranus, which led to the encirclement and
ultimate defeat of the German 6th Army.
This confirms that the Red Army was re-
sponsible for defeating the 6th Army."},
    "2.  Wehrmacht launched Operation
Uranus against the Soviet Union.": {
        "bool": False,
        "reason":  "The text specifies
that the Soviet Red Army launched Op-
eration Uranus, not the Wehrmacht. The
operation was a counteroffensive by the
Soviets that led to the encirclement of
German forces, not an attack initiated by
the Wehrmacht against the Soviet Union."}
}
## Example output
{
    "1. Red Army defeated the 6th Army.":
True,
    "2.  Wehrmacht launched Operation
Uranus against the Soviet Union.": False
}
```

Figure 10: Step 2: Verbalized Relation Validation - LLM User Prompt 1

C Post Processing - Human Annotator Verification

The annotators verifying validity of triplets generated by SiftingLogic was selected amongst our colleagues. They were screened to natively speak and educated in the required language and have minimum qualification of university-level education. They were given a tutorial on the task and were explicitly advised to be as truthful as possible in their answer. To further mitigate any bias from the annotators, we only select triplets with unanimous agreements among the annotators to be included in the improved datasets.

In the annotation software (Figure 11), the annotators were presented with the context text with subject and object entities highlighted, the triplet to be evaluated and the definition of the relation in the triplet. They must indicate whether the triplet is correct or wrong based on the context text provided.

The definitions of the relations are presented below:

- **head of government**: head of the executive power of this town, city, municipality, state, country, or other governmental body
- **country**: sovereign state that this item is in (not to be used for human beings)
- **place of birth**: most specific known (e.g. city instead of country, or hospital instead of city)

Figure 11: Screenshot of software used by annotators to verify triplets output from SiftingLogic. The same software is used for both Re^2-DocRED and revised REDFM mandarin test set.

birth location of a person, animal or fictional character

- **place of death**: most specific known (e.g. city instead of country, or hospital instead of city) death location of a person, animal or fictional character
- **father**: male parent of the subject. For stepfather, use "stepparent" (P3448)
- **mother**: female parent of the subject. For stepmother, use "stepparent" (P3448)
- **spouse**: the subject has the object as their spouse (husband, wife, partner, etc.). Use "unmarried partner" (P451) for non-married companions
- **country of citizenship**: the object is a country that recognizes the subject as its citizen
- **continent**: continent of which the subject is a part
- **instance of**: that class of which this subject is a particular example and member; different from P279 (subclass of); for example: K2 is an instance of mountain; volcano is a subclass of mountain (and an instance of volcanic landform)
- **head of state**: official with the highest formal authority in a country/state
- **capital**: seat of government of a country, province, state or other type of administrative territorial entity
- **official language**: language designated as official by this item
- **position held**: subject currently or formerly holds the object position or public office
- **child**: subject has object as child. Do not use for stepchildren—use "relative" (P1038), qualified with "type of kinship" (P1039)
- **author**: main creator(s) of a written work (use on works, not humans); use P2093 (author name string) when Wikidata item is unknown or does not exist
- **member of sports team**: sports teams or clubs that the subject represents or represented
- **director**: director(s) of film, TV-series, stageplay, video game or similar
- **screenwriter**: person(s) who wrote the script for subject item
- **educated at**: educational institution attended by subject
- **composer**: person(s) who wrote the music [for lyricist, use "lyrics by" (P676)]
- **member of political party**: the political party of which a person is or has been a member or otherwise affiliated
- **employer**: person or organization for which the subject works or worked
- **founded by**: founder or co-founder of this organization, religion or place
- **league**: league in which team or player plays or has played in
- **publisher**: organization or person responsible for publishing books, periodicals, printed music, podcasts, games or software
- **owned by**: owner of the subject
- **located in the administrative territorial entity**: the item is located on the territory of the following administrative entity. Use

P276 for specifying locations that are non-administrative places and for items about events. Use P1382 if the item falls only partially into the administrative entity

- **genre**: creative work's genre or an artist's field of work (P101). Use main subject (P921) to relate creative works to their topic
- **operator**: person, profession, or organization that operates the equipment, facility, or service
- **religion**: religion of a person, organization or religious building, or associated with this subject
- **contains administrative territorial entity**: (list of) direct subdivisions of an administrative territorial entity
- **follows**: immediately prior item in a series of which the subject is a part, preferably use as qualifier of P179 [if the subject has replaced the preceding item, e.g. political offices, use "replaces" (P1365)]
- **followed by**: immediately following item in a series of which the subject is a part, preferably use as qualifier of P179 [if the subject has been replaced, e.g. political offices, use "replaced by" (P1366)]
- **headquarters location**: city or town, where an organization's headquarters is or has been situated. Use P276 qualifier for specific building
- **cast member**: actor in the subject production [use "character role" (P453) and/or "name of the character role" (P4633) as qualifiers] [use "voice actor" (P725) for voice-only role]
- **producer**: person(s) who produced the film, musical work, theatrical production, etc. (for film, this does not include executive producers, associate producers, etc.) [for production company, use P272, video games - use P178]
- **award received**: award or recognition received by a person, organization or creative work
- **creator**: maker of this creative work or other object (where no more specific property exists)
- **parent taxon**: closest parent taxon of the taxon in question
- **ethnic group**: subject's ethnicity (consensus is that a VERY high standard of proof is needed for this field to be used. In general this means 1) the subject claims it themselves, or 2) it is widely agreed on by scholars, or 3) is fictional and portrayed as such)
- **performer**: actor, musician, band or other performer associated with this role or musical work
- **manufacturer**: (main or final) manufacturer or producer of this product
- **developer**: organization or person that developed the item
- **series**: series which contains the subject
- **sister city**: twin towns, sister cities, twinned municipalities and other localities that have a partnership or cooperative agreement, either legally or informally acknowledged by their governments
- **legislative body**: legislative body governing this entity; political institution with elected representatives, such as a parliament/legislature or council
- **basin country**: country that have drainage to/from or border the body of water
- **located in or next to body of water**: body of water on or next to which a place is located
- **military branch**: branch to which this military unit, award, office, or person belongs, e.g. Royal Navy
- **record label**: brand and trademark associated with the marketing of subject music recordings and music videos
- **production company**: company that produced this film, audio or performing arts work
- **location**: location of the object, structure or event. In the case of an administrative entity as containing item use P131. For statistical entities use P8138. In the case of a geographic entity use P706. Use P7153 for locations associated with the object

- **subclass of**: this item is a subclass (subset) of that item; all instances of this item are instances of that item; different from P31 (instance of), e.g.: K2 is an instance of mountain; volcano is a subclass of mountain (and an instance of volcanic landform)
- **subsidiary**: subsidiary of a company or organization; generally a fully owned separate corporation. Compare with "business division" (P199). Opposite of parent organization (P749)
- **part of**: object of which the subject is a part (if this subject is already part of object A which is a part of object B, then please only make the subject part of object A), inverse property of "has part" (P527, see also "has parts of the class" (P2670))
- **original language of work**: language in which a film or a performance work was originally created. Deprecated for written works and songs; use P407 ("language of work or name") instead
- **platform**: platform for which a work was developed or released, or the specific platform version of a software product
- **mouth of the watercourse**: the body of water to which the watercourse drains
- **original network**: network(s) or service(s) that originally broadcast a radio or television program
- **member of**: organization, club or musical group to which the subject belongs. Do not use for membership in ethnic or social groups, nor for holding a political position, such as a member of parliament (use P39 for that)
- **chairperson**: presiding member of an organization, group or body
- **country of origin**: country of origin of this item (creative work, food, phrase, product, etc.)
- **has part**: part of this subject; inverse property of "part of" (P361). See also "has parts of the class" (P2670)
- **residence**: the place where the person is or has been, resident
- **date of birth**: date on which the subject was born
- **date of death**: date on which the subject died
- **inception**: time when an entity begins to exist; for date of official opening use P1619
- **dissolved, abolished or demolished**: point in time at which the subject (organisation, building) ceased to exist; see "date of official closure" (P3999) for closing a facility, "service retirement" (P730) for retiring equipment, "discontinued date" (P2669) for stopping a product
- **publication date**: date or point in time when a work was first published or released
- **start time**: time an entity begins to exist or a statement starts being valid
- **end time**: moment when an entity ceases to exist or a statement stops being valid
- **point in time**: date something took place, existed or a statement was true; for providing time use the "refine date" property (P4241)
- **conflict**: battles, wars or other military engagements in which the person or item participated
- **characters**: characters which appear in this item (like plays, operas, operettas, books, comics, films, TV series, video games)
- **lyrics by**: author of song lyrics
- **located on terrain feature**: located on the specified (geo)physical feature. Should not be used when the value is only political/administrative (P131) or a mountain range (P4552)
- **participant**: person, group of people or organization (object) that actively takes/took part in an event or process (subject). Preferably qualify with "object has role" (P3831). Use P1923 for participants that are teams
- **influenced by**: this person, idea, etc. is informed by that other person, idea, etc., e.g. “Heidegger was influenced by Aristotle”
- **location of formation**: location where a group or organization was formed

- **parent organization**: parent organization of an organization, opposite of subsidiaries (P355)
- **notable work**: notable scientific, artistic or literary work, or other work of significance among subject's works
- **separated from**: subject was founded or started by separating from identified object
- **narrative location**: the narrative of the work is set in this location
- **work location**: location where persons or organisations were actively participating in employment, business or other work
- **applies to jurisdiction**: the item (institution, law, public office, public register...) or statement belongs to or has power over or applies to the value (a territorial jurisdiction: a country, state, municipality, ...)
- **product or material produced**: material or product produced by an organization, industry, facility, or process
- **unemployment rate**: portion of a workforce population that is not employed
- **territory claimed by**: administrative divisions that claim control of a given area
- **participant of**: event in which a person or organization was/is a participant; inverse of P710 or P1923
- **replaces**: person, state or item replaced. Use "structure replaces" (P1398) for structures. Use "follows" (P155) if the previous item was not replaced or predecessor and successor are identical
- **replaced by**: other person or item which continues the item by replacing it in its role. Use P156 ("followed by") if the item is not replaced nor identical, but adds to the series (e.g. books in a series)
- **capital of**: country, state, department, canton or other administrative division of which the municipality is the governmental seat
- **languages spoken, written or signed**: language(s) that a person or a people speaks, writes or signs, including the native language(s)
- **present in work**: this (fictional or fictionalized) entity, place, or person appears in that work as part of the narration (use P2860 for works citing other works, P361/P1433 for works being part of other works, P1343 for entities described in non-fictional accounts)
- **sibling**: the subject and the object have at least one common parent (brother, sister, etc. including half-siblings); use "relative" (P1038) for siblings-in-law (brother-in-law, sister-in-law, etc.) and step-siblings (step-brothers, step-sisters, etc.)
- **named after**: entity or event that inspired the subject's name, or namesake (in at least one language). Qualifier "applies to name" (P5168) can be used to indicate which one
- **occupation**: occupation of a person. See also "field of work" (Property:P101), "position held" (Property:P39). Not for groups of people. There, use "field of work" (Property:P101), "industry" (Property:P452), "members have occupation" (Property:P3989).
- **shares border with**: countries or administrative subdivisions, of equal level, that this item borders, either by land or water. A single common point is enough.
- **sport**: sport that the subject participates or participated in or is associated with

D Inverse Relations

Refer to Table 12.

E Co-occuring Relations

Refer to Table 13.

F Compatible Entity Types for each DocRED Relation

Refer to Table 16.

G Entity-level Constraints based on Relation Definition

Refer to Table 17 for entity constraints for Re^2-DocRED and REDFM.

Relation	Inverse Relations
author	notable work
performer	notable work
producer	notable work
composer	notable work
director	notable work
lyrics by	notable work
participant	participant of
participant of	participant
has part	part of
part of	has part
sibling	sibling
series	has part
spouse	spouse
characters	present in work
conflict	participant
parent organization	subsidiary
subsidiary	parent organization, owned by
follows	followed by
followed by	follows
father	child
mother	child
replaced by	replaces
replaces	replaced by
head of government	applies to jurisdiction
legislative body	applies to jurisdiction
head of state	applies to jurisdiction
sister city	sister city
capital	capital of
creator	notable work
screenwriter	notable work
capital of	capital

Table 12: List of inverse relations

Relation	Co-occuring Relations
country	located in the administrative territorial entity, location
conflict	participant of
composer	creator
author	creator
cast member	performer
member of sports team	member of
member of political party	member of
dissolved, abolished or demolished	end time
parent organization	owned by
inception	start time, point in time
start time	point in time
date of birth	point in time
publication date	point in time

Table 13: List of co-occurring relations

H Relation Verbalization

Refer to Table 18 for relation verbalizations for Re^2-DocRED and REDFM.

I Supplemental Ablation

We conduct an additional ablation study using TaG (Zhang et al., 2023) to isolate the effects of Sifting-Logic (SL) and rule-based expansion (rule). Due to computational constraints, models were trained on a subset of 803 train examples and evaluated on the full test sets. From Table 14, we note:

1. TaG trained on **Subset Train**$_{Re-O}$ **+ SL + Rule** achieves the highest recall across all train and test sets, demonstrating the effectiveness of SL and rule-based approaches without human intervention.

2. Notably, **Subset Train**$_{Re-O}$ **+ SL + Rule** outperforms **Subset Train**$_{Re-O}$ **+ Rule**, confirming SL's specific contribution to the training set.

J Evaluation on Long-tail relations

We define a long-tail relation as relations with frequency < 100 in Re-DocRED test set; Test$_{Re-O}$. From Table 15, we observe approximately similar pattern of results compared to the results on the full test sets (Table 4), namely:

1. Recall on Train$_{Re-O}$ steadily decreases as test sets improve in completeness through FN reduction.

2. Our improved Train$_{Re}$ delivers recall gains on the improved test sets (Test$_{Re}$, Test$_{GN}$, Test$_{GN+SL}$), emphasizing that reducing FN in training data yields recall gains in more complete test sets.

K Re^2-DocRED Relation Frequency Distribution

For Test$_{GN+SL}$ relation distribution, refer to Table 19.

For Test$_{RED(zh)}$ relation distribution, refer to Table 20.

	Subset Train$_{Re-O}$	**Subset Train**$_{Re-O}$ **+ rule**	**Subset Train**$_{Re-O}$ **+ SL + rule**
	R	R	R
Test$_{Re-O}$	38.31	38.55	**40.18**
Test$_{GN-O}$	34.32	35.29	**37.09**
Test$_{Re}$	30.12	40.11	**41.74**
Test$_{GN}$	27.03	36.00	**37.98**
Test$_{GN+SL}$	23.60	30.87	**34.11**

Table 14: TaG (Zhang et al., 2023) Recall (R) results across subsets of Train$_{Re-O}$ as baseline, with rule-based expansion (rule) and with SiftingLogic and rule-based expansion (SL + rule). Top scores are in **bold**.

	Train$_{Re-O}$	**Train**$_{Re}$
	P / R	P / R
Long-tail Test$_{Re-O}$	51.95 / **36.51**	17.68 / 31.88
Long-tail Test$_{GN-O}$	53.98 / **26.49**	20.42 / 25.68
Long-tail Test$_{Re}$	52.36 / 16.82	62.84 / **46.75**
Long-tail Test$_{GN}$	54.58 / 14.27	64.21 / **39.02**
Long-tail Test$_{GN+SL}$	55.89 / 12.94	63.19 / **34.17**

Table 15: AutoRE (Xue et al., 2024) Precision (P) and Recall (R) results across long-tail test sets. Top scores are in **bold**.

Table 16: Relation and Compatible Entity Types

Relation	Entity Type Pairs
head of government	('LOC', 'PER'), ('ORG', 'PER'), ('MISC', 'PER')
country	('ORG', 'LOC'), ('MISC', 'LOC'), ('LOC', 'LOC')
place of birth	('MISC', 'LOC'), ('PER', 'LOC')
place of death	('MISC', 'LOC'), ('PER', 'LOC')
father	('PER', 'PER')
mother	('PER', 'PER')
spouse	('PER', 'PER')
country of citizenship	('PER', 'LOC')
continent	('PER', 'LOC'), ('ORG', 'LOC'), ('MISC', 'LOC'), ('LOC', 'LOC')
instance of	('PER', 'MISC'), ('LOC', 'MISC'), ('ORG', 'MISC'), ('NUM', 'MISC'), ('TIME', 'MISC'), ('MISC', 'MISC')
head of state	('LOC', 'PER')
capital	('LOC', 'LOC')
official language	('PER', 'MISC'), ('ORG', 'MISC'), ('MISC', 'MISC'), ('LOC', 'MISC')
position held	('PER', 'MISC')
child	('PER', 'PER')
author	('MISC', 'PER')
member of sports team	('PER', 'ORG'), ('PER', 'MISC')
director	('MISC', 'PER')
screenwriter	('MISC', 'PER')
educated at	('PER', 'ORG')
composer	('MISC', 'PER')
member of political party	('PER', 'ORG'), ('PER', 'MISC')
employer	('PER', 'PER'), ('PER', 'ORG')
founded by	('ORG', 'PER'), ('ORG', 'ORG'), ('ORG', 'LOC'), ('ORG', 'MISC'), ('LOC', 'PER'), ('LOC', 'ORG'), ('LOC', 'LOC'), ('LOC', 'MISC'), ('MISC', 'PER'), ('MISC', 'ORG'), ('MISC', 'LOC'), ('MISC', 'MISC')
league	('PER', 'MISC'), ('ORG', 'MISC'), ('MISC', 'MISC')
publisher	('MISC', 'PER'), ('MISC', 'ORG'), ('MISC', 'MISC')
owned by	('LOC', 'PER'), ('MISC', 'PER'), ('ORG', 'PER'), ('ORG', 'ORG'), ('LOC', 'ORG'), ('MISC', 'ORG'), ('ORG', 'LOC'), ('LOC', 'LOC'), ('MISC', 'LOC'), ('MISC', 'MISC'), ('ORG', 'MISC'), ('LOC', 'MISC')
located in the administrative territorial entity	('PER', 'LOC'), ('ORG', 'LOC'), ('MISC', 'LOC'), ('LOC', 'LOC')
genre	('PER', 'MISC'), ('MISC', 'MISC')
operator	('LOC', 'ORG'), ('MISC', 'ORG'), ('LOC', 'MISC'), ('MISC', 'MISC'), ('LOC', 'PER'), ('MISC', 'PER')
religion	('ORG', 'ORG'), ('MISC', 'ORG'), ('LOC', 'ORG'), ('PER', 'ORG'), ('PER', 'MISC'), ('ORG', 'MISC'), ('MISC', 'MISC'), ('LOC', 'MISC')
contains administrative territorial entity	('LOC', 'LOC')
follows	('ORG', 'ORG'), ('TIME', 'TIME'), ('LOC', 'LOC'), ('MISC', 'MISC'), ('NUM', 'NUM'), ('PER', 'PER')

Relation	Entity Type Pairs
followed by	('ORG', 'ORG'), ('TIME', 'TIME'), ('LOC', 'LOC'), ('MISC', 'MISC'), ('NUM', 'NUM'), ('PER', 'PER')
headquarters location	('ORG', 'LOC')
cast member	('MISC', 'PER')
producer	('MISC', 'PER')
award received	('PER', 'MISC'), ('MISC', 'MISC'), ('ORG', 'MISC')
creator	('MISC', 'PER'), ('MISC', 'ORG')
parent taxon	('MISC', 'MISC')
ethnic group	('ORG', 'ORG'), ('PER', 'ORG'), ('PER', 'MISC'), ('ORG', 'MISC')
performer	('PER', 'PER'), ('MISC', 'PER'), ('MISC', 'ORG')
manufacturer	('MISC', 'ORG'), ('MISC', 'PER')
developer	('MISC', 'PER'), ('MISC', 'ORG')
series	('LOC', 'MISC'), ('ORG', 'MISC'), ('TIME', 'MISC'), ('NUM', 'MISC'), ('PER', 'MISC'), ('MISC', 'MISC')
sister city	('LOC', 'LOC')
legislative body	('LOC', 'ORG'), ('LOC', 'MISC')
basin country	('LOC', 'LOC')
located in or next to body of water	('LOC', 'LOC')
military branch	('MISC', 'MISC'), ('ORG', 'MISC'), ('PER', 'MISC'), ('ORG', 'ORG'), ('MISC', 'ORG'), ('PER', 'ORG')
record label	('MISC', 'ORG'), ('MISC', 'MISC')
production company	('MISC', 'ORG')
location	('LOC', 'LOC'), ('MISC', 'LOC')
subclass of	('MISC', 'MISC')
subsidiary	('ORG', 'ORG')
part of	('PER', 'ORG'), ('PER', 'LOC'), ('PER', 'MISC'), ('ORG', 'ORG'), ('ORG', 'LOC'), ('ORG', 'MISC'), ('LOC', 'LOC'), ('LOC', 'MISC'), ('NUM', 'MISC'), ('TIME', 'TIME'), ('TIME', 'MISC'), ('MISC', 'PER'), ('MISC', 'ORG'), ('MISC', 'LOC'), ('MISC', 'TIME'), ('MISC', 'MISC')
original language of work	('MISC', 'MISC')
platform	('MISC', 'MISC')
mouth of the watercourse	('MISC', 'LOC'), ('LOC', 'LOC')
original network	('MISC', 'ORG'), ('MISC', 'MISC')
member of	('ORG', 'ORG'), ('PER', 'ORG'), ('LOC', 'ORG'), ('MISC', 'ORG'), ('ORG', 'MISC'), ('PER', 'MISC'), ('LOC', 'MISC'), ('MISC', 'MISC')
chairperson	('MISC', 'PER'), ('ORG', 'PER')
country of origin	('MISC', 'LOC')
has part	('ORG', 'PER'), ('LOC', 'PER'), ('MISC', 'PER'), ('ORG', 'ORG'), ('LOC', 'ORG'), ('MISC', 'ORG'), ('LOC', 'LOC'), ('MISC', 'LOC'), ('MISC', 'NUM'), ('TIME', 'TIME'), ('MISC', 'TIME'), ('PER', 'MISC'), ('ORG', 'MISC'), ('LOC', 'MISC'), ('TIME', 'MISC'), ('MISC', 'MISC')
residence	('PER', 'LOC')
date of birth	('PER', 'TIME')
date of death	('PER', 'TIME')

Relation	Entity Type Pairs
inception	('ORG', 'TIME'), ('LOC', 'TIME'), ('MISC', 'TIME')
dissolved, abolished or demolished	('MISC', 'TIME'), ('ORG', 'TIME'), ('LOC', 'TIME')
publication date	('MISC', 'TIME')
start time	('ORG', 'TIME'), ('LOC', 'TIME'), ('MISC', 'TIME')
end time	('ORG', 'TIME'), ('LOC', 'TIME'), ('MISC', 'TIME')
point in time	('ORG', 'TIME'), ('LOC', 'TIME'), ('MISC', 'TIME'), ('PER', 'TIME')
conflict	('PER', 'MISC'), ('ORG', 'MISC'), ('LOC', 'MISC'), ('MISC', 'MISC')
characters	('MISC', 'PER'), ('MISC', 'MISC')
lyrics by	('MISC', 'PER'), ('MISC', 'ORG')
located on terrain feature	('PER', 'LOC'), ('ORG', 'LOC'), ('MISC', 'LOC'), ('LOC', 'LOC')
participant	('MISC', 'PER'), ('MISC', 'ORG'), ('MISC', 'MISC')
influenced by	('MISC', 'PER'), ('PER', 'MISC'), ('PER', 'PER'), ('MISC', 'MISC')
location of formation	('ORG', 'LOC'), ('MISC', 'LOC')
parent organization	('ORG', 'ORG')
notable work	('ORG', 'MISC'), ('PER', 'MISC'), ('MISC', 'MISC')
separated from	('ORG', 'ORG'), ('LOC', 'LOC'), ('MISC', 'MISC')
narrative location	('MISC', 'LOC')
work location	('PER', 'LOC'), ('ORG', 'LOC'), ('MISC', 'LOC')
applies to jurisdiction	('PER', 'LOC'), ('ORG', 'LOC'), ('MISC', 'LOC')
product or material produced	('LOC', 'MISC'), ('MISC', 'MISC'), ('ORG', 'MISC')
unemployment rate	('LOC', 'NUM')
territory claimed by	('LOC', 'ORG'), ('LOC', 'MISC')
participant of	('PER', 'MISC'), ('ORG', 'MISC'), ('MISC', 'MISC')
replaces	('PER', 'PER'), ('MISC', 'PER'), ('PER', 'MISC'), ('ORG', 'ORG'), ('ORG', 'MISC'), ('MISC', 'ORG'), ('MISC', 'MISC')
replaced by	('PER', 'PER'), ('MISC', 'PER'), ('PER', 'MISC'), ('ORG', 'ORG'), ('ORG', 'MISC'), ('MISC', 'ORG'), ('MISC', 'MISC')
capital of	('LOC', 'LOC')
languages spoken, written or signed	('PER', 'MISC'), ('ORG', 'MISC'), ('MISC', 'MISC')
present in work	('PER', 'MISC'), ('ORG', 'MISC'), ('LOC', 'MISC'), ('TIME', 'MISC'), ('MISC', 'MISC')
sibling	('PER', 'PER')

Table 17: Relation and Entity Constraints

Relation	Entity Constraints
head of government	1. [subject] is a town, city, municipality, state, country, or other governmental body: True 2. [object] is a head of government: True
country	1. [object] is a sovereign state: True 2. [subject] is a human being: False
place of birth	1. [subject] is a person, animal or fictional character: True 2. [object] is a location: True
place of death	1. [subject] is a person, animal or fictional character: True 2. [object] is a location: True
father	1. [object] is a male: True 2. [object] is a parent: True
mother	1. [object] is a female: True 2. [object] is a parent: True
spouse	1. [subject] is a husband, wife or partner: True 2. [object] is a husband, wife or partner: True
country of citizenship	1. [object] is a country: True 2. [subject] is citizen of a country: True
continent	1. [object] is a continent: True
instance of	1. [subject] is a particular example and member of [object]: True
head of state	1. [object] is an official: True 2. [subject] is a country or state: True
capital	1. [subject] is a country, province, state or other type of administrative territorial entity: True 2. [object] is a capital: True
official language	1. [object] is a language: True
position held	1. [object] is a position or public office: True
child	1. [subject] is a parent: True
author	1. [subject] is a written work: True 2. [object] is creator of a written work: True

Relation	Entity Constraints
member of sports team	1. [object] is a sports team or club: True
director	1. [subject] is a film, TV-series, stageplay, video game or similar: True 2. [object] is a director: True
screenwriter	1. [subject] is a film, television show or other visual media: True 2. [object] is a screenwriter: True
educated at	1. [object] is an educational institution: True
composer	1. [subject] is a musical composition: True 2. [object] is a person: True 3. [object] is a lyricist: False
member of political party	1. [subject] is a person: True 2. [object] is a political party: True
employer	1. [subject] is an employee of a person or organization: True 2. [object] is a person or organization: True
founded by	1. [subject] is an organization, religion or place: True 2. [object] is a founder or co-founder: True
league	1. [object] is a league: True 2. [subject] is a team or player in a team: True
publisher	1. [subject] is a book, periodical, printed music, podcast, games or software: True 2. [object] is an organization or person: True
owned by	1. [object] is an owner: True
located in the administrative territorial entity	1. [object] is an administrative territorial entity: True 2. [subject] is located in a specific administrative territorial entity: True 3. [subject] is an item about events: False 4. [object] is a non-administrative place: False
genre	1. [subject] is a creative work or artist: True 2. [object] is a genre or field of work: True
operator	1. [subject] is an equipment, facility, or service: True 2. [object] is a person, profession, or organization: True

Relation	Entity Constraints
religion	1. [subject] is a person, organization or religious building: True 2. [object] is a religion: True
contains administrative territorial entity	1. [subject] is an administrative territorial entity: True 2. [object] is an administrative territorial entity: True
follows	1. [subject] and [object] is part of the same series: True 2. [subject] and [object] are of the same type: True 3. [subject] replaces [object]: False
followed by	1. [subject] and [object] is part of the same series: True 2. [subject] and [object] are of the same type: True 3. [subject] is replaced by [object]: False
headquarters location	1. [subject] is an organization: True 2. [object] is a city or town: True
cast member	1. [object] is an actor: True 2. [subject] is a production: True
producer	1. [subject] is a film, musical work, or theatrical production: True 2. [object] is a producer: True 3. [object] is a production company: False 4. [subject] is a video game: False
award received	1. [subject] is a person, organization or creative work: True 2. [object] is an award or recognition: True
creator	1. [subject] is a creative work or other object: True 2. [object] is a maker of a creative work or object: True
parent taxon	1. [subject] is a taxon: True 2. [object] is a taxon: True
ethnic group	1. [subject] has an ethnicity: True 2. [object] is an ethnic group: True
performer	1. [subject] is a role or musical work: True 2. [object] is an actor, musician, band or performer: True

Relation	Entity Constraints
manufacturer	1. [subject] is a product: True 2. [object] is a manufacturer or producer: True
developer	1. [subject] is an item: True 2. [object] is an organization or person: True
series	1. [subject] is part of a series: True 2. [object] is a series: True
sister city	1. [subject] is a town, city, municipality or locality: True 2. [object] is a town, city, municipality or locality: True
legislative body	1. [object] is a legislative body: True 2. [subject] is an entity governed by a legislative body: True
basin country	1. [subject] is a body of water: True 2. [object] is a country: True
located in or next to body of water	1. [object] is a body of water: True 2. [subject] is a location: True
military branch	1. [subject] is a military unit, award, office, or person: True 2. [object] is a military branch: True
record label	1. [subject] is a music recording or music video: True 2. [object] is a brand or trademark: True
production company	1. [subject] is a film, audio or performing arts work: True 2. [object] is a production company: True
location	1. [subject] is an object, structure or event: True 2. [object] is a location: True
subclass of	1. [object] is a broader category than [subject]: True
subsidiary	1. [object] is a fully owned corporation: True 2. [subject] is a company or organization: True
part of	1. [object] is a whole that contains parts: True 2. [subject] is a part of the whole: True
original language of work	1. [subject] is a film or a performance work: True 2. [object] is a language: True

Relation	Entity Constraints
platform	1. [subject] a software product or work: True 2. [object] is a platform or platform version: True
mouth of the watercourse	1. [subject] is a watercourse: True 2. [object] is a body of water: True
original network	1. [subject] is a radio or television program: True 2. [object] is radio or television broadcaster: True
member of	1. [object] is an organization, club or musical group: True 2. [subject] is a member: True 3. [object] is an ethnic group: False 4. [object] is a social group: False 5. [object] is a political position: False
chairperson	1. [subject] is an organization, group or body: True 2. [object] is a chairperson: True
country of origin	1. [subject] is a creative work, food, phrase or product: True 2. [object] is a country: True
has part	1. [subject] is a whole that contains parts: True 2. [object] is a part of the whole: True
residence	1. [subject] is a person: True 2. [object] is a place: True
date of birth	1. [object] is a date: True 2. [subject] is a person: True
date of death	1. [object] is a date: True
inception	1. [object] is a point in time: True
dissolved, abolished or demolished	1. [object] is a point in time: True 2. [subject] is an organisation or building: True
publication date	1. [object] is a date or point in time: True 2. [subject] is a published or released work: True
start time	1. [object] is a point in time: True
end time	1. [object] is a point in time: True
point in time	1. [object] is a point in time: True

Relation	Entity Constraints
conflict	1. [object] is a battle, war or military engagement: True 2. [subject] is a person or item: True
characters	1. [subject] is a play, opera, operetta, book, comic, film, TV series or video game: True 2. [object] is a character: True
lyrics by	1. [subject] is a song: True 2. [object] is a lyricist: True
located on terrain feature	1. [object] is a (geo)physical feature: True 2. [object] is an administrative territorial entity: False
participant	1. [subject] is an event or process: True 2. [object] is a person, group of people or organization: True
influenced by	1. [subject] is a person or idea,: True 2. [object] is a person or idea: True
location of formation	1. [subject] is group or organization: True 2. [object] is a location: True
parent organization	1. [subject] is an organization: True 2. [object] is an organization: True
notable work	1. [object] is a notable scientific, artistic or literary work, or other work of significance: True 2. [subject] has a notable work: True
separated from	1. [subject] was part of [object]: True
narrative location	1. [object] is a location: True 2. [subject] is a narrative work: True
work location	1. [subject] is a person or organization: True 2. [object] is a location: True
applies to jurisdiction	1. [subject] is an institution, law, public office or public register: True 2. [object] is a territorial jurisdiction such as a country, state or municipality: True
product or material produced	1. [subject] is an organization, industry, facility, or process: True 2. [object] is a material or product: True

Relation	Entity Constraints
unemployment rate	1. [subject] has a workforce population: True 2. [object] is an unemployment rate: True
territory claimed by	1. [object] is an administrative division: True 2. [subject] is an area: True
participant of	1. [object] is an event: True 2. [subject] is a person or organization: True
replaces	1. [object] is a person, state or item: True 2. [subject] is a person, state or item: True 3. [subject] adds to the series [object] is part of: False
replaced by	1. [object] is a person, state or item: True 2. [subject] is a person, state or item: True 3. [object] adds to the series [subject] is part of: False
capital of	1. [object] is a country, state, department, canton or other administrative division: True 2. [subject] is a municipality: True
languages spoken, written or signed	1. [object] is a language: True 2. [subject] is a person or people: True
present in work	1. [subject] is a (fictional or fictionalized) entity, place, or person: True 2. [object] is a narrative work: True
sibling	1. [subject] is a sibling: True 2. [object] is a sibling: True
named after	1. [object] existed, was known, or was named prior to [subject]: True
occupation	1. [subject] is a person: True 2. [object] is a type of job or profession: True
shares border with	1. [subject] exists as a defined country or administrative subdivision: True 2. [object] exists as a defined country or administrative subdivision: True 3. [subject] is at the same administrative level as [object] (e.g., both countries or administrative subdivisions): True
occupation	1. [object] is recognized as a sport: True

Table 18: Relations and their verbalizations

Relation	Verbalization
head of government	[object] is the head of government of [subject]
country	[subject] is located in [object]
place of birth	[subject] was born in [object]
place of death	[subject] died in [object]
father	[object] is the father of [subject]
mother	[object] is the mother of [subject]
spouse	[subject] is married to [object]
country of citizenship	[subject] is a citizen of [object]
continent	[subject] is located in [object]
instance of	[subject] is an instance of [object]
head of state	[object] is the head of state of [subject]
capital	[object] is the capital of [subject]
official language	[object] is the official language of [subject]
position held	[subject] currently holds the position of [object]
child	[object] is a child of [subject]
author	[object] is the author of [subject]
member of sports team	[subject] is a member of the [object] sports team
director	[object] is the director of [subject]
screenwriter	[object] is the screenwriter of [subject]
educated at	[subject] was educated at [object]
composer	[object] is the composer of the music for [subject]
member of political party	[subject] is a member of [object] political party
employer	[object] is the employer of [subject]
founded by	[object] founded [subject]
league	[subject] plays in [object] league
publisher	[object] is the publisher of [subject]
owned by	[subject] is owned by [object]
located in the administrative territorial entity	[subject] is located in the administrative territorial entity of [object]
genre	[subject] belongs to the genre of [object]
operator	[object] is the operator of the [subject]
religion	[subject] practices [object]'s religion
contains administrative territorial entity	[subject] contains [object] as an administrative territorial entity
follows	[subject] follows [object] in a series
followed by	[subject] is followed by [object] in the sequence
headquarters location	[subject]'s headquarters is located in [object]
cast member	[object] is a cast member in [subject]
producer	[object] produced the [subject]
award received	[subject] has received [object] as an award
creator	[object] is the creator of [subject]
parent taxon	[object] is the closest parent taxon of [subject]
ethnic group	[subject] is a member of the [object] ethnic group
performer	[object] is the performer of [subject]
manufacturer	[object] manufactures [subject]
developer	[object] developed [subject]
series	[object] is a series that contains [subject]
sister city	[subject] is a sister city of [object]
legislative body	[object] is the legislative body of [subject]

Relation	Verbalization
basin country	[object] is a country that drains into [subject]
located in or next to body of water	[subject] is located in or next to [object]
military branch	[subject] belongs to the [object] military branch
record label	[object] is the record label of [subject]
production company	[object] is the production company that created [subject]
location	[subject] is located in [object]
subclass of	[subject] is a subclass of [object]
subsidiary	[object] is a subsidiary of [subject]
part of	[subject] is part of [object]
original language of work	[object] is the original language of the work [subject]
platform	[subject] was developed for [object]
mouth of the watercourse	[object] is the mouth of the watercourse of [subject]
original network	[subject] was originally broadcast on [object]
member of	[subject] is a member of [object]
chairperson	[object] is the chairperson of [subject]
country of origin	[subject] originated in [object]
has part	[object] is a part of [subject]
residence	[subject] resides in [object]
date of birth	[subject] was born on [object]
date of death	[subject] died on [object]
inception	[subject] began to exist at [object]
dissolved, abolished or demolished	[subject] was dissolved on [object]
publication date	[subject] was first published on [object]
start time	[subject]'s start time is [object]
end time	[subject]'s end time is at [object]
point in time	[subject] occurred at a point in time in [object]
conflict	[subject] participated in [object]
characters	[object] is a character in [subject]
lyrics by	[object] wrote the lyrics for [subject]
located on terrain feature	[subject] is located on the terrain feature [object]
participant	[object] is a participant in [subject]
influenced by	[subject] is influenced by [object]
location of formation	[object] is the location where [subject] was formed
parent organization	[object] is the parent organization of [subject]
notable work	[object] is a notable work by [subject]
separated from	[subject] was founded by separating from [object]
narrative location	[object] is the location where the story of [subject] takes place
work location	[subject] works at [object]
applies to jurisdiction	[subject] governs [object]
product or material produced	[object] is the product produced by [subject]
unemployment rate	The unemployment rate of [subject] is [object] percent
territory claimed by	[object] claims the territory of [subject]
participant of	[subject] is a participant of [object]
replaces	[subject] replaces [object]
replaced by	[object] is the replacement for [subject]

Relation	Verbalization
capital of	[subject] is the capital of [object]
languages spoken, written or signed	[subject] speaks, writes, or signs [object]
present in work	[subject] is present in [object]
sibling	[subject] and [object] share at least one common parent
named after	[subject] was named after [object]
occupation	[subject] works as a [object]
shares border with	[subject] and [object] share a common border
sport	[subject] participates in [object]

Table 19: Test$_{GN+SL}$ Relation Distribution

Relation	Frequency
located in the administrative territorial entity	4592
location	3439
country	2929
has part	1666
part of	1648
notable work	1028
contains administrative territorial entity	933
point in time	879
country of citizenship	809
performer	596
participant of	445
member of	438
participant	435
present in work	397
creator	384
publication date	364
country of origin	293
inception	289
applies to jurisdiction	275
work location	267
start time	252
owned by	252
date of birth	220
residence	212
record label	211
continent	207
located in or next to body of water	207
cast member	206
developer	187
date of death	183
conflict	168
place of birth	166
series	162
author	146
characters	143
sibling	137
instance of	135
employer	128
child	121
award received	121
parent organization	120
follows	120
located on terrain feature	118
producer	113
educated at	107
member of sports team	107
followed by	105
publisher	103
platform	97

Relation	Frequency
mouth of the watercourse	92
place of death	91
spouse	90
father	89
headquarters location	88
founded by	86
subsidiary	84
military branch	83
capital of	83
director	81
head of government	80
location of formation	80
languages spoken, written or signed	80
head of state	79
league	78
capital	73
composer	71
legislative body	68
religion	67
operator	66
influenced by	64
genre	61
end time	60
subclass of	58
screenwriter	56
dissolved, abolished or demolished	55
member of political party	53
production company	53
manufacturer	48
official language	45
chairperson	43
narrative location	41
original network	38
mother	37
original language of work	36
territory claimed by	34
ethnic group	33
product or material produced	32
basin country	32
lyrics by	31
replaced by	29
replaces	25
position held	21
parent taxon	9
separated from	4
unemployment rate	1

Table 20: Test$_{RED(zh)}$ Relation Distribution

Relation	**Frequency**
part of	332
country	245
location	215
instance of	118
located in or next to body of water	79
inception	62
country of citizenship	50
participant	47
genre	46
manufacturer	40
notable work	38
child	37
cast member	32
owned by	32
follows	23
headquarters location	22
named after	20
shares border with	19
capital	18
author	16
occupation	14
member of	14
league	14
characters	13
sport	13
spouse	13
sibling	11
replaces	11
place of birth	10
founded by	10
director	10
mouth of the watercourse	5