Supplementary Materials

Anonymous Author(s)
Submission Id: 903

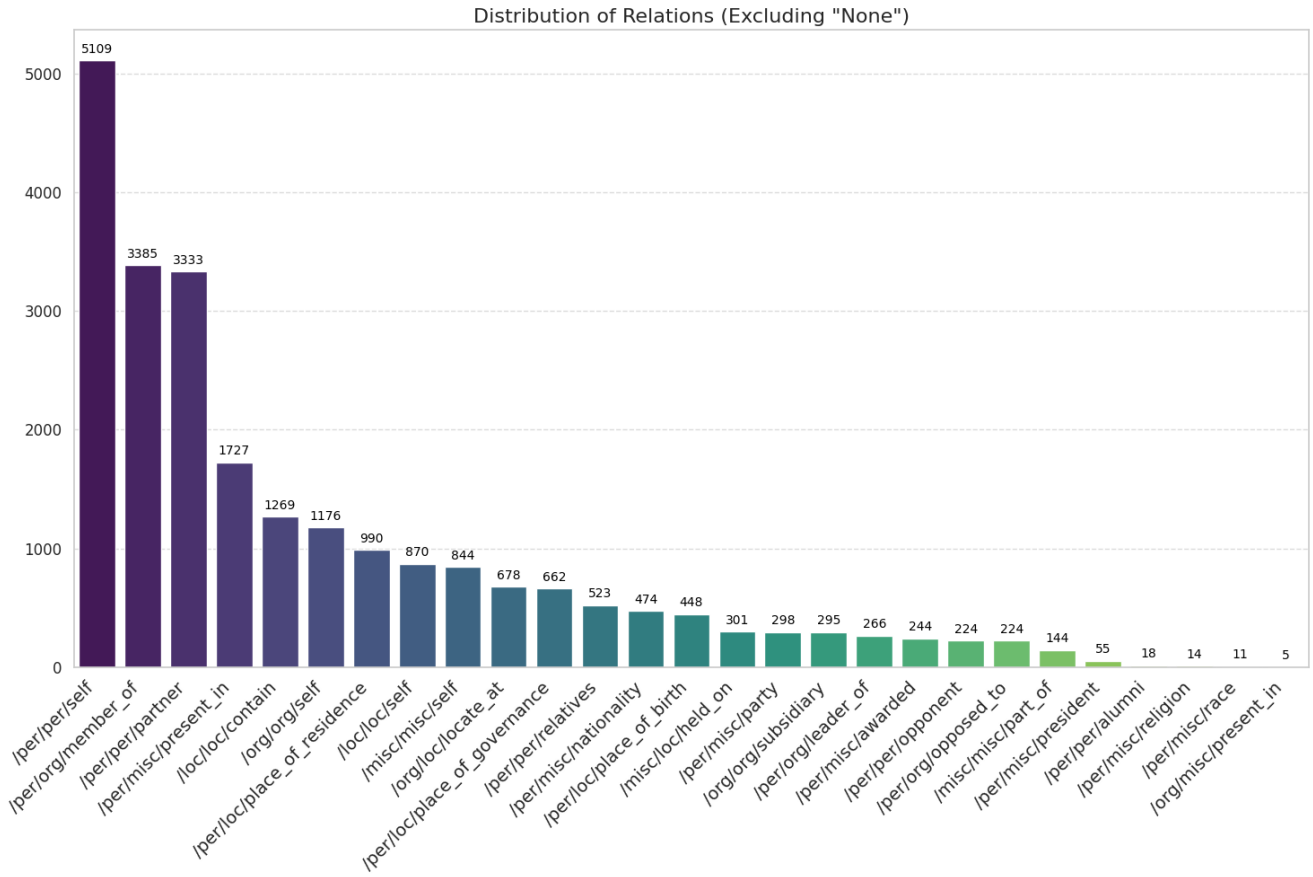

Figure 1: The distribution of relations.

A Overview of the UMRE Dataset

The UMRE dataset comprises 27 frequently-used relations, excluding **"None"**, as illustrated in Fig. 1. These relations encompass a wide range of categories, including personal social relationships and geographical location relationships, etc. We provide detailed explanations of the relations included, as described below.

A.1 Relation Interpretation

- **none**: Denotes no clear relations between entities.
- **/per/loc/place_of_governance:** Denotes the location where a person holds authority.
- **/per/misc/party:** Indicates the political party to which a person is affiliated.
- **/per/org/member_of:** Indicates that a person is a member of an organization.
- **/per/per/self:** Represents that the two entities are the same person.
- **/per/misc/nationality:** Shows the nationality of a person.
- **/loc/loc/self:** Represents that the two entities are the same location.
- **/per/misc/present_in:** Indicates that a person is present at or involved in a specific event or occasion.
- **/per/loc/place_of_residence:** Shows the place where a person resides.
- **/org/org/self:** Represents that the two entities are the same organization.
- **/misc/misc/self:** Represents that the two entities are the same miscellaneous.
- **/per/per/opponent:** Indicates opposition or competition between people.
- **/per/loc/place_of_birth:** Shows the birthplace of a person.
- **/per/per/partner:** Indicates cooperation or partnership between people.

- **/per/org/opposed_to:** Shows opposition between a person and an organization.
- **/loc/loc/contain:** Indicates that one location contains another location.
- **/org/loc/locate_at:** Shows the location where an organization is situated.
- **/per/misc/president:** Indicates that a person is the president of a location.
- **/misc/loc/held_on:** Shows that an event is held at a specific location.
- **/per/org/leader_of:** Indicates that a person is the leader of an organization.
- **/org/org/subsidiary:** Represents the subsidiary relationship between organizations.
- **/per/per/relatives:** Shows the familial relationship between people.
- **/per/misc/awarded:** Indicates that a person has been awarded something.
- **/misc/misc/part_of:** Shows that something is part of a whole.
- **/per/misc/race:** Indicates the race of a person.
- **/per/per/alumni:** Shows that the two men are alumni of the same school.
- **/per/misc/religion:** Indicates the religious belief of a person.
- **/org/misc/present_in:** Shows that an organization is present at or involved in a specific event or occasion.

B Prompt Template

In this appendix, we present three essential prompt templates used to guide the extraction and processing of multimodal data for the extended MORE and MNRE datasets. The Prompt Template for Knowledge Enhancement (Named Entity Recognition) focuses on identifying and enriching named entities in text by providing contextual information. The Prompt Template for Visual Grounding aligns text entities with visual objects in images by describing their spatial positions and generating precise bounding box coordinates. Lastly, the Prompt Template for Image Caption generates detailed descriptions of images, capturing key elements and their relationships. Together, these templates ensure accurate and comprehensive extraction of information, enhancing the diversity and quality of the dataset.

Prompt Template for Knowledge Enhancement (Named Entity Recognition)

Here are some content that people post on Twitter, and these content are composed of original text and image descriptions of the original text. Please note that the text and image descriptions here may or may not be relevant, so make your own judgment. Please follow the data annotation style and method reflected in the example I provided, comprehensively analyze the image description and the original text, determine which named entities and their corresponding types are included in the original text, and explain the reason for your judgment. Notice: just in 'Text', not include 'Image descriptions', don't change the writing style and format of entity names, and words after the @ sign are not counted.

Text: #Sports Connor Cook could not satisfy NFL teams' questions about leadership.
Image descriptions: Michigan state football player in uniform.
Question: Comprehensively analyze the Text and the Image description, which named entities and their corresponding types are included in the Text? Explain the reason for your judgment.
Answer:

- **Named entities:**
  (1) Connor Cook (person/player)
  (2) NFL (league/organization)
- **Reasoning:** The sentence mentions Connor Cook, a former Michigan State football player who was drafted by the NFL. The NFL is the highest-level professional football league in the world. The image of a Michigan state football player may be related to Connor Cook, but the image description is not directly relevant to the named entities in the text.

Text: #NBA Enhanced: OKC Thunder to win is NOW 10/3.
Image descriptions: Kevin Durant and Russell Westbrook.
Question: Comprehensively analyze the Text and the Image description, which named entities and their corresponding types are included in the Text? Explain the reason for your judgment.
Answer:

- **Named entities:**
  (1) OKC Thunder (team/franchise)
  (2) NBA (league/organization)
- **Reasoning:** The sentence mentions OKC Thunder, a professional basketball team based in Oklahoma City. The NBA is mentioned as the organization that the OKC Thunder are playing in. It is possible that the text and image are related in the sense that the odds for OKC Thunder to win may have been influenced by their past performances with the team.

Text: The Premier Leagues battle for #ChampionsLeague, Europa qualifying.
Image descriptions: A soccer player in action on the field.
Question: Comprehensively analyze the Text and the Image description, which named entities and their corresponding types are included in the Text? Explain the reason for your judgment.
Answer:

Prompt Template for Visual Grounding

Task1: Entity Expansion Expressions

Image Background: The image features a large stadium with a crowd of people gathered inside the arena. The stadium is filled with people of different ages who are all engaged in watching a soccer (or football) match. The atmosphere of the event is lively and exciting, capturing the spirit of a live sporting event.
Text: 'Premier League stadiums: Every top flight ground ranked by age'
Question: In the context of the provided information, tell me briefly what is the 'Premier League (ORG)' in the text?
Answer: A football (soccer) league in England.

Image Background: This picture features the same woman wearing a blue sweater, with the difference being that in the first picture, she has a blond hairstyle, whereas in the second picture, her hair is red. The image showcases a comparison of her appearance over time, emphasizing the changing nature of physical features, hairstyle, and fashion trends. The message is a reminder that people are constantly evolving and changing, and it is important to embrace these changes and adapt as they come.
Text: '19 things Taylor Swift does that no one else could ever get away with'
Question: In the context of the provided information, tell me briefly what is the 'Taylor Swift (PER)' in the text?
Answer: A woman singer-songwriter.

Image Background: The image depicts an indoor greenhouse filled with plants of various species. It appears as a tropical rainforest setting, with trees and shrubs of different sizes and colors. The plants are arranged in a way that creates an enchanting environment, with the greenhouse being surrounded by glass walls and windows. In this lush environment, there are numerous potted plants scattered throughout. Some of them are located close to each other, while others are placed at various distances from one another. Additionally, there are benches placed around the area, providing comfortable seating for visitors to appreciate the beauty and serenity of the greenhouse.
Text: 'The geometry of plants. Garfield Park Conservatory'
Question: In the context of the provided information, tell me briefly what is the 'Garfield Park Conservatory (LOC)' in the text?
Answer:

Task2: Visual Entailment

Please answer the following question based on the provided text and image:
Text Content: {Text}
Image: {Image_path}
Entity: {Entity Name}: {Entity Expansion Expressions}

Question: Is "{Entity Name}" included in the picture? If so, please describe him/her/it in the picture in detail. Otherwise, please print 'not in the image'.
Answer:

Prompt Template for Image Caption

Text: {Text}
Image: {Image_path}

Based on the given image and text, identify the image type and provide a detailed description:

(1) **If the image contains a person:**
- Identify the person in the image, describe their physical features (e.g., gender, age, clothing), and recognize who they are. Tell me their names.
- **Output:** The person's name and a brief description of their features.

(2) **If the image contains a location:**
- Describe the key features of the location (e.g., landmarks, scenery), and identify where it is.
- **Output:** The name of the location and key features.

(3) **If the image contains a flag or logo:**
- Identify the flag or logo, and determine which country or sports team it represents. Tell me their names.
- **Output:** The country name (for flags) or sports team name (for logos), along with a brief description of the flag/logo.

(4) **If the image contains text (perform OCR):**
- Extract meaningful text from the image itself (ignore noisy or distorted text).
- **Output:** The extracted text from the image.

Notes:
- Do not treat the social media text (Text) as part of the image OCR content.
- Only output the relevant information based on the image type.
- If the image type cannot be identified, output "Unable to identify the image type."
- If the image contains text, ensure that only relevant, readable content is extracted.

Answer: