



**NIST Trustworthy and Responsible AI
NIST AI 100-8**

RUFEERS (Recognizing Ultra Fine-grained Entities, Events, and Relations) Annotation Guidelines

Shudong Huang
Song Chen
Justin Mott

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.100-8>

NIST Trustworthy and Responsible AI
NIST AI 100-8

RUFEERS (Recognizing Ultra Fine-grained Entities, Events, and Relations)
Annotation Guidelines

Shudong Huang
Information Access Division
Information Technology Laboratory

Song Chen
Linguistic Data Consortium

Justin Mott
Independent Researcher

This publication is available free of charge from:
<https://doi.org/10.6028/NIST.AI.100-8>

May 2026



U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Craig Burkhardt, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

NIST AI 100-8
May 2026

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

NIST Technical Series Policies

[Copyright, Use, and Licensing Statements](#)

[NIST Technical Series Publication Identifier Syntax](#)

Publication History

Approved by the NIST Editorial Review Board on 2026-04-14

How to Cite this NIST Technical Series Publication

Huang, Shudong, et al. (2026) RUFEEERS (Recognizing Ultra Fine-grained Entities, Events, and Relations) Annotation Guidelines. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Trustworthy and Responsible AI, NIST AI 100-8. <https://doi.org/10.6028/NIST.AI.100-8>

Author ORCID iDs

Shudong Huang: 0009-0000-3357-6119

Contact Information

shudong.huang@nist.gov

Abstract

A key area in natural language processing and by extension, in artificial intelligence, is a system's ability to extract entities (things in the world), events (who does what and when), and relations (who is related to whom in what way) and to transform unstructured data into structured data enabling automated analysis, knowledge graph construction, and deeper understanding for tasks like trend analysis, threat detection, personalized recommendations, and building medical databases by identifying who, what, where, and when in text, and the connections between them. To measure a system's performance, it is often the case that evaluation reference data is created by human annotators. This document provides overall annotation guidelines for annotators as part of the RUFERS evaluation.

Keywords

data annotation; entity; entity type; event; event type; ontology; relation ; relation type

Table of Contents

1. Introduction	1
1.1. Events, Relations, Entities and Their Mentions	1
1.1.1. Events Overview	1
1.1.2. Relations Overview.....	2
1.1.3. Entities Overview.....	3
2. Annotation Overview	4
3. Taggability	7
3.1. Event Taggability	7
3.2. Relation Taggability.....	8
3.3. Entity Taggability.....	9
4. Event Annotation	11
4.1. Event Handles	12
4.2. Event Types	12
4.3. Event Mention Spans	13
4.3.1. Verbs as Event Mention Spans	13
4.3.2. Adjectives as Event Mention Spans.....	14
4.3.3. Nouns and Pronouns as Event Mention Spans	14
4.4. Event Mention Attributes	15
4.5. Event Arguments and Argument Roles.....	17
4.5.1. Linking entity mentions to events as event arguments	18
4.5.2. Linking event mentions to events as event arguments	20
4.5.3. Thing: event argument placeholder	20
4.5.4. Attributes for event arguments.....	21
4.6. Temporal Information: Start/End	21
4.7. Event Coreference.....	23
5. Relation Annotation	25
5.1. Relation Handle.....	26
5.2. Relation Types.....	26
5.3. Relation Mention Span	27
5.4. Relation Mention Attributes	28
5.5. Relation Arguments	29
5.6. Temporal Information: Start/End	30
5.7. Relation Coreference	31
6. Entity Annotation	32

6.1. Entity Handle.....	33
6.2. Entity Mention Types.....	34
6.3. Entity Mention Span	34
6.3.1. NAM Mention Span.....	35
6.3.2. NOM Mention Span.....	36
6.3.3. PRO Mention Span	37
6.3.4. Conjoined Mentions.....	38
6.4. Entity Types.....	38
6.4.1. Scope of Context	39
6.4.2. Tag for Usage/Meaning.....	40
6.4.3. Multiple Entity Types.....	40
6.5. Entity coreference.....	41
6.6. Argument Filler Annotation	42
7. Towards Successful Annotation and Future Improvements.....	44
7.1. Annotation Guidelines and Ontology.....	44
7.2. Talent Acquisition	44
7.3. Annotation Tool	44
7.4. Quality Control.....	45
7.5. Looking forward	45
7.5.1. Cross-document linking and cross-lingual annotation.....	45
7.5.2. Event relations.....	46
7.6. Conclusion.....	46
References.....	47
Appendix A. Annotation Procedure	48
Appendix B. Sample Source Document	49
Appendix C. Glossary	51

List of Tables

Table 1. Event Mention Frame of Event “Attack”	4
Table 2. Event “Attack” Frame.....	4
Table 3. Mention 1 of “Zelensky Being President of Ukraine” Relation.....	4
Table 4. Mention 2 of “Zelensky Being President of Ukraine” Relation.....	5
Table 5. Mention 5 of Entity “Ukraine”	5
Table 6. Entity “Ukraine” and Its Mentions	5
Table 7. Event Mention Frame.....	11

Table 8. Sample Event Linking Frame	19
Table 10. Event as Argument	20
Table 11. Attaching Temporal Information to Event Coref Chain.....	23
Table 12. Coreferential Event Mentions with Different Attributes	24
Table 13. Relation Mention Frame.....	25
Table 14. Relation Mention Example	26
Table 15. Relation Mention with Handle Example	26
Table 16. Relation Argument Example	29
Table 16. Symmetric Relation Argument Example	30
Table 17. Mandatory Relation Argument	30
Table 18. Coreferential Relation Mentions.....	31
Table 19. Entity Mention Frame	32
Table 20. Entity Handle Examples	33
Table 21. Sample Document	49

1. Introduction

RUFEERS (**R**ecognizing **U**ltra **F**ine-grained **E**ntities, **E**vents, and **R**elation**S**) is an expansion of RUFES (**R**ecognizing **U**ltra **F**ine-grained **E**ntitie**S**) (2022-2023), which itself is a continuation of the Message Understanding Conference (MUC) [10], the Automatic Context Extraction evaluations [9], and the NIST Text Analysis Conference (TAC)'s Knowledge Base Population (KBP) and Entity Discovery and Linking (EDL) tasks [1]. The original focus of RUFES was on fine-grained (broader and deeper) entity types. RUFEERS significantly expands the scope to focus on events and relations while retaining the same set of entity types as RUFES (2023).

The act of annotation is making explicit the understanding that humans glean about categories of interest in a particular medium or media and encoding this information in a format that machines can read. The human-annotated data can then be used to develop and train systems as well as evaluate how well systems perform. In RUFEERS, the goal of the annotation is to represent events that occurred in the world in a structured way – to identify and understand not just the types of events, but also their participants, when and where they occur and whether they are real or not. In addition, RUFEERS annotation captures background information about the entities by annotating relations between them. With the primary focus on events and relations, entity annotations in RUFEERS are limited to entities that are arguments (participants and localities) of a tagged event or relation.

Human annotators use guidelines to understand and perform annotation tasks, and annotation guidelines are typically updated continuously to address new issues encountered during annotation. To help stakeholders and researchers better understand the RUFEERS annotation process, NIST has revised the internal annotation guidelines to the NIST publication standards and makes them available to the public in general.

Note if examples are from the sample document in Appendix B, they are prefixed with S followed by the sentence number as shown in the appendix.

1.1. Events, Relations, Entities and Their Mentions

The RUFEERS annotation is composed of three interwoven annotation tasks: event annotation, relation annotation, and entity annotation. These targets are defined as follows:

- An Event is something that happens at a certain time and place, often described as a change of state.
- A Relation is a relationship that holds between two entities.
- An Entity is an object or concept in the world.

1.1.1. Events Overview

An event is some activity that happens in the world at a specific time at a specific place – for instance, Russia invaded Ukraine on Feb 24th, 2022. An event mention is a particular reference to an event that can be anchored to a word or phrase in the document being annotated. For

the purposes of RUFEEERS, this consists of a mention of the event in written prose language. An event can be mentioned more than once in a given document. In RUFEEERS, mentions of all taggable events¹ must be annotated, and those mentions referring to the same event are grouped together. For example, the following bolded words are coreferential mentions of the same military event:

*S1. Russia unleashes military **assault** on Ukraine.*

*S4. Russia on Thursday launched an **attack** on cities and military installations across Ukraine...*

*S5. The **assault** was startling in its scope, as missiles and rockets rained down across Ukraine's vast territory and Russian military vehicles rolled across the border from multiple directions.*

*S26. The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified **attack** by Russian military forces.*

Understanding the difference between an event, an activity that happens in the world, and a mention of that event in text is crucial for the RUFEEERS annotation. For each taggable event,

- annotate all the mentions that refer to the event by building a frame for each mention, including mention span, event type, attributes, arguments, and its start and end times;
- build a coreference (henceforth coref) chain by linking all mentions referring to the same event; and
- update event type and start and end times as needed to ensure that coreferential event mentions have the same event type and the same start and end times.

1.1.2. Relations Overview

A relation is a relationship that holds between two entities, for instance, the person entity “Volodymyr Zelenskyy” has a relationship with the country entity “Ukraine”, such that “Volodymyr Zelenskyy” is the leader of the country “Ukraine.” For RUFEEERS, a relation mention is a specific occurrence of a taggable relation in a text document. Similar to events, a relation can be mentioned more than once within the same document, and all mentions of a relation that occur in a document must be annotated.

*Ukrainian President **Volodymyr Zelensky** has condemned the shelling, calling it a “war crime.”*

*The **Ukrainian leader** called for Russia to be condemned for the attack.*

Understanding the difference between a relation and a mention of that relation is important for the RUFEEERS annotation. For each taggable relation,

¹ See section 3 Taggability below.

- annotate all the mentions that refer to the relation by building a frame for each mention, including mention span, relation type, attributes, arguments, and its start and end times;
- build a relation coref chain by linking all mentions referring to the same relation; and
- update relation types and start and end times as needed to ensure that coreferential relation mentions have the same relation type and start and end times.

1.1.3. Entities Overview

An entity is a unique physical or abstract object or set of physical or abstract objects in the world – for instance, a specific person (Adam Smith), place (United States), organization (NIST), treaty (Treaty on the Non-Proliferation of Nuclear Weapons, or NPT), etc. An entity mention is a linguistic expression that refers to, or describes, an entity. Such linguistic expressions can be a proper name (Named Mention, or NAM), a common noun phrase (Nominal Mention, or NOM), or a pronoun (Pronominal Mention, or PRO). In RUFEEERS annotation, we explicitly mark the linguistic category used to refer to the entity. For example, the following are all mentions of the person entity “Adam Smith” that one may see in a document.

NAM: Adam Smith, Smith, AS

NOM: the young senator from NY, the senator

PRO: he, his

In RUFES, an entity mention is taggable if the entity it refers to is of a type in the entity ontology. As its focus is on events and relations, RUFEEERS only annotates mentions of entities that are event or relation arguments AND all their coreferential mentions in the document. In other words, an entity mention is taggable if:

- it is an argument of an event/relation mention; or
- it is coreferential with an entity mention that has been tagged as an argument of an event/relation mention.

For each taggable entity,

- annotate all the mention spans that refer to the entity by building a complete frame for each mention, including mention span, mention level and entity type; and
- build an entity coref chain by linking all mentions referring to the same entity.

2. Annotation Overview

Consider again the four event mentions given earlier that all refer to the same military event.

*S1. Russia unleashes military **assault** on Ukraine.*

*S4. Russia on Thursday launched an **attack** on cities and military installations across Ukraine...*

*S5. The **assault** was startling in its scope, as missiles and rockets rained down across Ukraine’s vast territory and Russian military vehicles rolled across the border from multiple directions.*

*S26. The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified **attack** by Russian military forces.*

The event mention frame for the mention in S4 is shown below.

Table 1. Event Mention Frame of Event “Attack”

Span	attack
Type	military_activity.military_strike ²
Attribute	ACTUAL (see section 4.4 Event Mention Attributes below)
Argument roles	Agent: Russian Patient: cities Patient: installations Place: Ukraine
Temporal information	Start on: 2022-02-24 End: unknown

After the frames for all 4 mentions, “assault” in S1, “attack” in S4, “assault” in S5 and “attack” in S26 are built, an event coref chain is formed to indicate that they are mentions of the same military event.

Table 2. Event “Attack” Frame

Handle	Russia invaded Ukraine on Feb 24, 2022
Event mentions	Mention 1 “assault” in S1 Mention 2 “attack” in S4 Mention 3 “assault” in S5 Mention 4 “attack” in S26

Relation annotation follows the same procedure: first a mention frame is built for each relation mention and then a relation coref chain is built over the mentions referring to the same relation. The following sentence has two relation examples, and they happen to refer to the same relation, “Zelensky is the president of Ukraine.”

*S4. ... as **Ukraine’s president** called for **his nation** to fight in the streets and NATO leaders said Europe’s security had been fundamentally altered.*

Table 3. Mention 1 of “Zelensky Being President of Ukraine” Relation

Handle	Zelensky is the president of Ukraine
---------------	--------------------------------------

² A period is used between a parent type and a child type (subtype) to indicate the hierarchical structure of event/relation/entity ontology.

Type	organizational_relation.leader_of.government_head_of
Mention span	's
Attribute	ACTUAL
Arguments	arg1: Zelensky arg2: Ukraine
Temporal information	Start: unknown End: unknown

Table 4. Mention 2 of “Zelensky Being President of Ukraine” Relation

Handle	Zelensky is the president of Ukraine
Type	organizational_relation.leader_of.government_head_of
Mention span	's
Attribute	ACTUAL
Arguments	arg1: his arg2: nation
Temporal information	Start: unknown End: unknown

In RUFFEERS, entity annotation revolves around taggable events and relations, namely, only entities that are event or relation arguments are annotated. However, at the mention level, all mentions in the document referring to the same taggable entity must be annotated and an entity coref chain must be established. The following bold-faced mentions are all identified to refer to the entity “Ukraine”:

*S1. Russia unleashes military assault on **Ukraine**.*

*S4. Russia on Thursday launched an attack on cities and military installations across **Ukraine**...*

*S5. The assault was startling in its scope, as missiles and rockets rained down across the **country**'s vast territory and Russian military vehicles rolled across the border from multiple directions.*

*S9. Lines formed at a border crossing between **Ukraine** and Poland.*

*S16. **Ukraine**'s Western backers scrambled to react.*

An entity mention frame looks like the table below.

Table 5. Mention 5 of Entity “Ukraine”

Span	Ukraine
Type	GPE.Country
Level	NAM

With mentions of “Ukraine” in S1, “Ukraine” in S4, “country” in S5, “Ukraine” in S9 and “Ukraine ” in S16, a coref chain for the entity “Ukraine” can be built as follows to indicate that they are mentions of the same entity.

Table 6. Entity “Ukraine” and Its Mentions

Handle	Ukraine
Entity mentions	Mention 1 “Ukraine” in S1 Mention 2 “Ukraine” in S4

Mention 3 "country" in S5
Mention 4 "Ukraine" in S9
Mention 5 "Ukraine" in S16

3. Taggability

Taggability refers to what is taggable; it is task-specific, confined by the task ontology and other requirements. In RUFES, for example, any entity mention is taggable if the entity it refers to is an instance of one of the entity types defined in the RUFES ontology. However, in RUFEEERS, if an entity is not an argument in any event or relation in the document, none of its mentions is taggable.

The RUFEEERS ontology is a structural representation of event, relation and entity types curated from various resources such as Wikidata with new entries not found in known ontologies. Each phase of annotation and evaluation has its own version of ontology with updates and additions in later versions. The set of event, relation or relation types for each version is frozen prior to annotation. Having a closed ontology is important for human annotation. Otherwise, all prior annotations must be re-visited if any change is made to the ontology during annotation. By extension, systems participating in the evaluation use the same ontology.

The entire ontology is accessible in the annotation GUI (INCEpTION) and annotators can navigate the nesting structure to find each type and its associated definition. It is not distributed with this publication due its length, but the Excel version can be released upon request.

3.1. Event Taggability

For an event mention to be taggable,

- it must be overtly mentioned in the document, i.e., anchored to a text span in the document (see section 4.3 for more details of identifying event span);
- it must be able to be categorized by a type of any level in the event ontology an argument; and
- the event categorization must meet the specification of such type.

As an example, consider the following paragraph from the sample article.

S22-S24. "Peace on our continent has been shattered," he told reporters in Brussels. "This is a deliberate, coldblooded and long-planned invasion. Russia is using force to try to rewrite history.

The verb "told" refers to an event of conversation type, defined as "interactive communication between two or more people", and the noun "invasion" to an event of military_strike type, defined as "act of taking control of a place using military force." Hence, both event mentions are tagged.

This paragraph also has several examples where an event mention such as "shattered", "planned" or "rewrite" refers to a real-world event, but the type is not covered by the current RUFEEERS ontology. Therefore, they are not taggable.

For mentions to be taggable, they not only need to be categorizable by the type of any level in the ontology but must also meet the **specification** of the type, for example required arguments.

In the example below, “said” might be understood as an `information_exchange` event, defined as “transfer of information between entities.” Since it is a government official who “said” in a statement, it does not meet the specification of `information_exchange`, as there is no recipient of the information, hence it is not taggable as an `information_exchange` event.

S26. “The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified attack by Russian military forces,” a government official said in a statement late Wednesday.

Note words such as “said” and “declared” in general are not tagged when the context does not show that there is transfer of information between entities. Instead, the entity who is doing the “saying” is captured as a communicator argument if what is being said contains a taggable event.

S17. European leaders vowed crushing sanctions against Russia...

In this sentence, the mention “sanction” refers to a boycott event, and the entity referred to by “leaders” is the communicator argument of the boycott event. So, the “vowing” event is not tagged.

Event mentions are exhaustively tagged, meaning every taggable event mention in the document must be tagged.

3.2. Relation Taggability

For a relation mention to be taggable,

- it must be overtly mentioned in the document;
- it must be able to be categorized by a type of any level in the relation ontology; and
- the relation categorization must meet the specification of such type.

Unlike an event, which is always triggered by a text span in the document, a relation can be triggered by a syntactic structure (such as a possession, premodification) in addition to a text span (such as a noun, a verb or a preposition). Regardless, the relation must be attested in the document either lexically, semantically or morphosyntactically but cannot be inferred.

As an example, consider this sentence from our sample article.

S26. The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified attack by Russian military forces.

The relation between “people” and “Ukraine” is triggered by the preposition “of” and can be categorized as “citizenship.” The relation between “forces” and “Russian” is triggered by “Russian” modifying “military forces” and can be categorized as “`organizational_relation.subsidiary`” with “forces” as a subsidiary of “Russia.”

RUFEEERS does not annotate relations between two arguments if both are arguments of a taggable event and their relationship is implied or expressed via the event specification. In the following example,

John bought a book yesterday at the airport.

“John” owns the “book” after the purchase (both entity mentions would be annotated as arguments of a purchase event during event annotation), but the ownership is inferred and not directly mentioned, so the ownership relation between “John” and “book” is not taggable. “John” and “book” also have relations with “airport”, as both “John” and the “book” are located in the “airport” at the time when the “buying” event happened. Since the location relations are inferable from the event and these entity mentions are already linked to the buying event as event arguments, the location relation between “John” and “airport” or between “book” and “airport” is not taggable.

For relations to be tagged and classified correctly, they must also meet the **specification** of a given type in the relation ontology. In the S26 example, “people” and “Ukraine” seem to be related and may be categorized as “geopolitical_relation” or “organizational_relation.” However, the definition of “geopolitical_relation” is defined as “strictly for relations between two (or more) GPE entities”, and “people” is not a GPE entity, so it should not be tagged as “geopolitical_relation.” On the other hand, “organizational_relation” is defined as “a person or a group of persons or an organization (arg1) has relation with another person or an organization or GPE (arg2).” Therefore, the relation between “people” and “Ukraine” can be categorized as “organizational_relation.” In fact, “organizational_relation” has a finer subtype, “citizen_of”, defined as “a person entity (arg1) is the citizen of a GPE (arg2)”, which can be applied more precisely to the relation between “people” and “Ukraine.

As with event mentions, relation mentions are also exhaustively tagged, meaning every taggable relation mention in the document must be tagged.

3.3. Entity Taggability

For RUFEEERS, all mentions of an entity must be annotated if at least one of its mentions appears in a taggable event or relation mention. If an entity is not an argument of a taggable event or relation, none of its mentions is taggable. In the following example:

S13. Ukraine is defending itself and will not give up its freedom, no matter what Moscow thinks.

“Ukraine” does not fill any argument slot of a taggable event or relation mention, but it is coreferential to “Ukraine” in S1, which is a place argument of the invasion event, so “Ukraine” in the example above is a taggable entity mention.

On the other hand, “prayers” in the following example is not an argument of any taggable event or relation in the entire document, so it is not a taggable entity mention.

S26. The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified attack by Russian military forces.

Like events and relations, entities must be categorized by a type in the entity ontology. However, many entity arguments in taggable events or relations cannot be categorized in the original RUFES entity ontology. To alleviate this issue, the RUFEEERS entity ontology adds a

place-holder type, “argument_filler”, to be used for any entity that is yet to be covered by the current entity ontology but fills an event argument slot. For example, a job title such as “CEO” can fill a “position” argument for an organizing_activity event type. While job titles are not one of the entity types in the current ontology, “CEO” can be categorized by TTL (title) type under the place-holder type “argument_filler.” Similarly, a monetary value such as “5 million dollars” can fill a “money” argument of an economic_activity event and can be categorized by the MON type under “argument_filler.” See section 6.4 for more details of entity types.

4. Event Annotation

All mentions referring to a taggable event are annotated, with each mention anchored to its own text span, typically a word, in the document. Each event mention is represented by a frame with the following elements: type, span, attribute, arguments, start time and end time.

A coreference chain for each event is built by linking all the event mentions that refer to the same event. Each event will have its own coreference chain, and the same event mention cannot be linked to more than one event coreference chain. For each event, the annotator must provide an event handle, a brief description of the event in well-formed English, which must be easily distinguishable from handles of other events, especially similar events. In summary, each event coref chain consists of:

- **Handle:** A straightforward sentence or phrase in well-formed English that provides the crucial information about the event.
- **Event Mentions:** All event mentions referring to the event.

A typical event mention is represented in a form-like frame populated with details as shown in the following table.

Table 7. Event Mention Frame

Frame Field	Description
Event Mention ID	A unique ID assigned to the event mention. The ID is assigned automatically by the annotation GUI.
Type	The most specific type (furthest level in the ontology) that categorizes the event. Event mentions that refer to the same event are of the same type.
Mention Span	The string of text span that indicates to a human reader that the event mention occurs. Each event mention has a unique span. Two event mentions cannot have the same span in the same document.
Attribute	Features of the event mention that indicate its specificity or whether it happened. See section 4.4 Event Mention Attributes below.
Arguments	The entities that participate in the event and the roles the entities play in the event. Each event type has a list of argument roles, and each role can be filled by one or more entity mentions.
Temporal information	The start and end time of the event indicating when the event happens.

For event mentions that are linked to the same event coref chain, the event mention frame may be simplified as the following to avoid redundancies during annotation.

- **Type:** Only assign the type to one of the event mentions in the coref chain, preferably the first one in the document.
- **Mention Span:** Required for each event mention.
- **Attribute:** Required **only if** the attribute of the mention is not ACTUAL.
- **Arguments:** Only select the arguments within the event mention scope that are new to the event (see section 4.5 for more details).

- **Temporal information:** Only provide start and end time for one of the event mentions, preferably the first one. The temporal information needs to be as specific as possible based on the context.

In what follows, each field in the event and event mention frames are discussed in more detail.

4.1. Event Handles

Each event coref chain is described with an **event handle**, a straightforward sentence or phrase in well-formed English that provides the crucial information about the event. A good event handle neatly summarizes the pertinent information about an event in a way that makes it distinct from other events in the document.

- **Good (✓):** “Russian planes dropped bombs on eastern Ukraine.”
- **Bad (X):** “Russia attacked.” (insufficient information)
- **Worse (X X):** “RU 🛩️ 💣 UA” (insufficient information AND not in natural language)

A handle should make clear what specific type of event it is, clearly refer to the core participants in the event and include any other information (e.g. temporal information) needed to disambiguate it from other events.

4.2. Event Types

An **event type** is a broad category of events, such as accident, biological process, movement, etc. Event types are multi-tiered such that an event type at a high level may have finer-grained subtypes, subsubtypes, and so on. The event ontology is presented as an expandable “tree” in the annotation GUI, allowing annotators to navigate the hierarchical structure of the nested levels of event types to select the most specific one that categorizes the event.

Annotators rely on context and the annotation ontology to determine the event type of a taggable event mention. They first decide the top-level type, then decide whether a finer type can capture the event more precisely. Annotators will always select the lowest level of type that best categorizes the event whenever possible. For example, in:

*S4. ...Ukraine’s president called for his nation to **fight** in the streets.*

“Fight” can either be a type of **violence** or **military_activity**. Based on the definition of these two event types (violence is defined as “use of physical force or power for coercion or as punishment” and **military_activity** as “human activity by or related to military forces”), violence appears to be a more suitable type, as “his nation to fight in the streets” refers to people in Ukraine, not necessarily Ukrainian military forces, fighting in the street against Russian invaders. The annotators will then check if a more specific subtype under **violence** can describe the fighting:

- **killing:** act of killing a living being
- **torture:** intentional infliction of physical or mental suffering upon a person or an animal

- **attack**: action to injure another organism

Since none of the subtypes of violence captures the “fight” in the context, the type **violence** is selected for “fight” in this sentence.

Sometimes an event mention may be categorized by more than one type. As the current RUFFERS specification only allows one type to be annotated for any event, the type selected must be the one that is more relevant to the context. As an example, consider the following sentence.

John killed Mary with a knife.

The “killed” event mention can either be tagged as “illegal_activity.crime.violent_crime.murder”, defined as “unlawful killing of a human with malice aforethought”, or tagged as “violence.killing.homicide”, defined as “killing of a human being by another human being.” If the document containing this sentence is talking about the killing in the context of illegal activity, the “illegal_activity.crime.violent_crime.murder” type should be selected, even though the “violence.killing.homicide” type is a close contender.

4.3. Event Mention Spans

An event **mention span** is the smallest string of text that reasonably indicates to a human reader that the event occurs, which can be verbs (most common), nouns, pronouns, or even adjectives.

NOTE: in the examples that follow, bold face will be used to indicate the word or phrase to be chosen as the span of a taggable event mention.

4.3.1. Verbs as Event Mention Spans

In many cases, the main verb in the sentence or clause describes the event. For example:

*The explosion **killed** 7 and **injured** 20.*

*Milosovic was **indicted** yesterday for committing war crimes.*

*Fifteen Palestinians were **injured** this morning in the town of Rafah when a rocket hit a local market.*

Auxiliary verbs supporting main verbs to express tenses, moods, aspects, or voices are not included in the mention span. In the following example, “have died” indicates the death event, and since “have” is an aspectual verb, the main verb “died” is selected as the extent of the “death” event.

*More than 200 people have **died** in the landslide.*

Verbs such as “want”, “promise”, “try”, “claim” etc. can be used to express modality. Usually, it is the verb following them that triggers an event while those “volition” verbs may indicate an event attribute. Similar to auxiliary modal verbs, they are not included in the mention span. For

example, “attempted” in the following sentence indicates the event is hypothetical or in the future, so the assassination event has an “irrealis” (not factual) attribute.

*With Confederate armies near collapse across the South, John Wilkes Booth attempted to **assassinate** Abraham Lincoln.*

The extent for a phrasal verb (e.g. break down, lay off) must include both the verb and the following particle unless they are separated by an object, in which case only the head verb is included in the extent.

*Jane was **laid off** by XYZ Corp after her argument with her boss.*

*XYZ Corp **laid** Jane off after her heated argument with her boss.*

This is different from prepositional verbs (e.g. “look into”, “arrive at”) where the preposition is never included in the extent.

4.3.2. Adjectives as Event Mention Spans

Predicative adjectives describe the state of an entity, which may be a result of an event. For example, when a person is dead, there was a “dying” event leading to the state of being dead. If the event is not expressed explicitly in the same sentence or clause, the stative adjective is treated as the event trigger.

A frequent pattern in English is “verb + NP (entity mention) + adjective”:

The explosion left at least 30 dead and dozens injured.

For all such cases, the adjective describing the resulting state from a taggable event is tagged as the event anchor. This is called Stand-Alone Adjective Rule:

Stand-Alone Adjective Rule: When a verb and an adjective are used together to express the occurrence of an event, the adjective will be chosen as the event mention span whenever it can stand alone in expressing the resulting state brought about by the event.

So, in the above example, there are two different events with “dead” and “injured” as mention spans respectively.

4.3.3. Nouns and Pronouns as Event Mention Spans

Nouns can be used to describe events:

*S4. Russia on Thursday launched an **attack** on cities and military installations across Ukraine*

*S20. The alliance planned an emergency **summit** of its leaders.*

Sometimes in a verbal phrase consisting of a verb and a noun, the verb and the noun may appear to “compete” to be the mention span as in S4 above. While the verb “launch” does

trigger an event, it is the noun “attack” that refers to the event semantically. In such cases the Stand-Alone Noun Rule below is followed to identify the mention span:

Stand-Alone Noun Rule: In cases where both a verb and noun compete to be the event mention span, the noun will be selected whenever that noun can be used by itself to refer to the event.

Below is another example showing how the Stand-Alone Noun Rule is applied.

*As job cuts affect several employees, city council held a **meeting** in Piggott to discuss severance packages.*

An event mention can also be anchored to a pronoun, either via an antecedent referring to a taggable event or via directly referring to an event in the world.

*John was accused of **murdering** his wife and he denied **it**.*

*Putin said in the speech, by turns cold and angry. “We simply weren’t given any other option to defend Russia and our people other than **that which** we will use today.”*

In the first example above, the pronoun “it” refers back to the gerund “murdering”, which by itself refers to a taggable event. In the second example, the pronoun “that” refers to Russian military operation that Russia was launching in Ukraine, so it is the span of a `military_activity.military_strike` event mention. Note the pronoun “which” must also be tagged as referring to the same event, similar to “it” in the first example.

4.4. Event Mention Attributes

An **event mention attribute** is an additional piece of information about an event at the **mention** level. Most events discussed so far are actual events that took place at some point in the real world. However, events may also be referenced in other ways, such as events that **did not** occur, events that **might have** occurred, or events that **will** occur in the future. Events may also be specific or generic. Such distinctions are made via event mention attributes as follows.

- **Actual:** an event that did or will happen.

Hannity texted Meadows and Rep. Jim Jordan (R-Ohio).

- **Negated:** an event that did not or will not happen.

The government says the incident was a failed attempt to assassinate Maduro planned by Colombia, the United States and the Venezuelan opposition involving drones carrying explosives.

- **Hedged:** an event that may or may not happen or have happened.

*Police have identified the man found dead inside a house and are now considering opening a **murder** investigation.*

- **Irrealis:** a hypothetical or future event.

*The national security team, which is working with NATO allies in Europe to try to avert a Russian **invasion** of Ukraine.*

- **Generic:** a non-specific event, meaning the event is not a singular or finite set of occurrences.

*The delay is sure to be a bitter disappointment to the many parents who have been waiting anxiously for inoculations for children younger than 5 — the only people in the United States still not eligible to receive Coronavirus **vaccinations**.*

*Salat Hassen called on countries that give **aid**.*

*The group specialized in **transporting** illegal weapons.*

*Israel says, it has lifted the internal restrictions that barred Palestinians from **moving** among West Bank towns and villages.*

*One core role of inter-provincial enterprises is to implement **mergers**.*

The default attribute for event mentions is Actual; there is no need to explicitly mark this during annotation.

It is important to note that event mention attributes are not always mutually exclusive. For instance, a negated event can also be hedged and/or irrealis. For example:

Putin said that Russia did not intend to occupy Ukraine.

The event here is `military_activity.military_occupation`, which is negated as conveyed by “not.” Additionally, the event is also irrealis, as indicated by “intend to.”

Likewise, in the following example, the event denoted by “occupy” is both negated and irrealis since it describes an event that will not happen in the future.

*He added that Moscow would not **occupy** Ukraine and that it had to end eight years of conflict in eastern Ukraine, where Kyiv government forces have been fighting pro-Moscow separatist militants.*

The attributes Actual and Negated are mutually exclusive, and only one can appear in the frame for an event mention. In cases where the document asserts both the occurrence and non-occurrence of an event, two event mentions are created: one with an Actual attribute and one with a Negated attribute.

*Russian officials denied reports that the Moskva was sunk by a Ukrainian missile **strike**, attributing **it** to “stormy seas” in the Black Sea.*

One may be tempted to tag such instances as Hedged, rather than creating two separate mentions. This is incorrect; the Hedged attribute is used when the document asserts that the event may or may not have occurred. In the example above, however, there are two conflicting assertions, and each of these assertions must be captured with a different mention attribute.

4.5. Event Arguments and Argument Roles

Event arguments are usually entities, but sometimes another event can also fill an argument in an event. Below is the list of event argument roles in RUFEEERS:

- **Agent:** “doer” or “instigator” of the action or event.
- **Patient:** animate “undergoer” of the action or event.
- **Artifact:** inanimate “undergoer” of the action or event.
- **Participant:** participant of the event when it is not clear whether the participant is a doer or undergoer.
- **Beneficiary:** entity that benefits from the action or event.
- **Communicator:** entity that communicates the action or event.
- **Topic:** “for what” or “about what” slot of the event or action. This argument role can be filled by either entity or event.
- **Place:** location where the action or event occurs.
- **Origin:** location or entity from which something moves. This is reserved for movement and financial transaction event types to indicate the origin of the movement or transaction.
- **Destination:** location or entity to which something moves. This is reserved for movement and financial transaction event types to indicate the destination of the movement or transaction.
- **Crime:** reserved for legal_process event types to indicate the crime/offense involved in a legal process. This argument role can be filled by either entity or event.
- **Illness:** illness involved in the event or action. This is reserved for medical_procedure and death/illness events.
- **Instrument:** medium by which the action or event is carried out.
- **Money:** financial asset such as money involved in an event or action.
- **Position:** job or official position involved in an event or action.
- **Sentence:** reserved for legal_process event types to indicate the sentence in a legal process. This argument role is usually filled by argument_filler, a placeholder for any entity that does not yet have a predefined type in the entity ontology.
- **Thing:** argument placeholder. For a given event, if any entity is involved in the event, but the existing argument roles do not capture the role the entity plays in the event, this role can be selected. For example, in a manufacturing event, an agent manufactures an artifact with an instrument for a beneficiary using some kind of materials or components; here the materials or components can fill the “thing” argument role.

To check whether the entities correctly fill the argument roles in an event, one may plug in the entities and the event types in a template to see if the formulaic sentence based on the template is a good paraphrase of the event in question. For example, by plugging in the arguments for the “attack” event mention as shown in **Table 1. Event Mention Frame of Event “Attack”** into the template below,

- **military_activity.military_strike**
 - Argument roles: agent, patient, instrument, place
 - Template: <agent> strikes <patient> or <artifact> using <instrument> at <place>

the reference to the event can be paraphrased as follows.

Russia struck cities and installations in Ukraine

Compared to the original sentence below, the paraphrase accurately captures the event:

S4. Russia on Thursday launched an attack on cities and military installations across Ukraine.

Below are examples to two more event types and their canonical argument roles and templates:

- **biological_process.death**
 - Arguments roles: patient, agent, instrument, illness, place
 - Template: <patient> dies of <agent> with <instrument> or from <illness> at <place>
- **information_exchange.dissemination.broadcast**
 - Arguments roles: agent, patient, instrument, topic, place
 - Template: <agent> broadcasts to <patient> about <topic> using <instrument> at <place>

If there are multiple entities filling in the same argument role, all of them must be tagged. For example, the instrument role is played by three arguments, “missiles”, “rockets” and “vehicles” in the “assault” event referenced by the sentence below.

*S5. The **assault** was startling in its scope, as missiles and rockets rained down across Ukraine’s vast territory and Russian military vehicles rolled across the border from multiple directions.*

As most event argument roles are filled by entities, refer to Section 6 “Entity Annotation” with regards to what information needs to be annotated for each entity and entity mention.

4.5.1. Linking entity mentions to events as event arguments

Although event arguments are being linked to specific event mentions during annotation, they are related to the event at the event level. What this means in practice is that there is no need to link every argument found in an event mention if it has already been added in a previously

annotated mention of the same event. In other words, annotators will only need to link arguments that are new to the event coref chain during annotation.

It is possible (though rare) that an argument to an event may not appear in any event mention, e.g., a location is only specified in the title of the news article and the title itself does not contain an event mention. In such cases, annotators should link the argument to the closest event mention detected.

Just as an event may have multiple arguments assuming the same role, an entity may also serve as arguments in different events.

To demonstrate how the above annotation principles are applied, consider the following 4 sentences from the sample document where the “Russia attacking Ukraine” event is mentioned.

*S1. Russia unleashes military **assault** on Ukraine that US officials call ‘premeditated war’.*

*S4. Russia on Thursday launched an **attack** on cities and military installations across Ukraine...*

*S5. The **assault** was startling in its scope, as missiles and rockets rained down across the country’s vast territory and Russian military vehicles rolled across the border from multiple directions.*

*S26. The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified **attack** by Russian military forces.*

The four mentions, including the span, event type, attribute and argument roles, are summarized in the following table:

Table 8. Sample Event Linking Frame

	Event Mention 1	Event Mention 2	Event Mention 3	Event Mention 4
Span	assault	attack	assault	attack
Type	military_activity.military_strike			
Attribute	ACTUAL	ACTUAL	ACTUAL	ACTUAL
Argument roles	Agent: Russia Patient: Ukraine	Patient: cities Patient: installations Place: Ukraine	Instrument: missiles Instrument: rockets Instrument: vehicles Place: territory	Agent: Russian military forces

These four event mentions are coreferential and linked to the same event coref chain. Since “Russia” already fills the agent role in Event Mention 1, there is no need to add “Russia” in S4 for the same agent role in Event mention 2. S26 has a new argument, “Russian military forces”, filling another agent role. So, it must be included in the list of arguments in Event Mention 4.

On the other hand, “Ukraine” in Event Mention 1 is assigned the role of patient. But in Event Mention 2, its role is place.

It should be noted that it is possible for an entity mention to be linked to more than one event.

4.5.2. Linking event mentions to events as event arguments

Some event argument roles, such as topic and crime in RUFEEERS, may be filled by events. Similar to entity mentions, an event mention must be annotated first before it can be selected to fill in the topic or crime argument slot. For example, in

*S17. European leaders vowed crushing **sanctions** against Russia and planned to meet later Thursday to **discuss them**.*

“discuss” is an event mention of the conversation subtype under the interaction type. The topic to be discussed by the European leaders is “sanctions”, which in itself is an event of the legal_process.punishment.sanction type. The sanction event has two mentions in this sentence, “sanctions” and “them”, which must be annotated first in order to be included in the list of arguments for the “discuss” event as follows:

Table 9. Event as Argument

Span	discuss
Type	information_exchange.communication.human_communication.interpersonal_communication.conversation
Attribute	Irrealis
Argument roles	Agent: European leaders Topic: them (sanctions) Place: Brussels

4.5.3. Thing: event argument placeholder

Sometimes there may not be an applicable argument role defined in the current argument role inventory for a potential event argument. Consider the entity mention “local residents” in the example below.

*Local residents watched anxiously when Russia started their military **operation** for the “demilitarization and denazification” of eastern Ukraine.*

A military_activity event such as the one anchored on the mention “operation” can have a number of argument roles such as agent, patient, instrument and place. However, there is not an argument role for people witnessing the event. To include the local residents in the event, the annotator may assign the role of “thing” to the entity mention “local residents.” It should be noted though that if “watching” were a taggable event, the “watching” entity would be an argument of that event. A future iteration of RUFEEERS may include such event types (and by extension entity types such as “witness”) as they would be of interest to legal/crime events.

4.5.4. Attributes for event arguments

Similar to event mentions being assigned attributes, attributes are also assigned to event arguments, reflecting the nature of the argument's participation in an event. There are three attributes applicable to event arguments:

- **Actual:** the argument **did/does/will** participate in the event.
- **Negated:** the argument **did/does/will not** participate in the event.
- **Hedged:** the argument **may or may not** have participated in the event.

Arguments with the attribute Actual are asserted to have participated in the event. For example:

Meadows received texts from Fox News hosts Laura Ingraham and Brian Kilmeade, as well as Hannity, according to the newly released communications.

Here, all recipient arguments are asserted as participants in the contact.communication event and so are assigned the attribute Actual.

Hedged is used to indicate that the document is uncertain whether the argument participated in the event.

Eyewitness reports differ on whether the tragic explosion at the Al-Ahli Arab Hospital was caused by Hamas or the Israeli Defense Forces.

In this example, the explosion may have been caused by Hamas or the IDF. To reflect this uncertainty, both arguments are tagged as arguments and assigned the attribute Hedged.

A negated argument is one that is asserted to have **not** been a participant in an event. For example,

Israel Defense Forces (IDF) spokesperson Jonathan Conricus told CNN on Wednesday that the "first packet of information" was "evidence that clearly supports the fact that the explosion could not have been the result of an Israeli attack."

Here, IDF asserts first that the Attack event happened, and that it was **not** perpetrated by Israel.

The default attribute for all arguments is Actual. Therefore, only Negated and Hedged attributes are explicitly assigned.

4.6. Temporal Information: Start/End

Specific events start and end at specific points in time. To capture the temporal information for events, the most specific calendar dates that can be determined from the context are annotated. For each event, one of the following four options is selected to indicate when the event starts.

- **StartOn:** the event starts on this date.
- **StartBefore:** the start date is at least as early as this date.

- **StartAfter:** the start date is at least as late as this date.
- **StartUnknown:** the start date is unknown in this document.

Likewise, one of the following four options is selected to indicate when the event ends.

- **EndOn:** the event ends on this date.
- **EndBefore:** the start date ends at least as early as this date.
- **EndAfter:** the end date is at least as late as this date.
- **EndUnknown:** the end date is unknown in this document.

In RUFEEERS dates are recorded in the format of YYYY-MM-DD, for example, 2022-02-22 for February 22, 2022.

A time expression in the document may be expressed in ways other than a calendar date, for example, Thursday, yesterday, last week, etc. Annotators may need to combine the time expression with the context as well as the publication date of the document to determine a specific date. Looking up a calendar may be necessary but checking sources such as other news reports or Wikipedia is not allowed.

Consider for example the following sentence from a news article published on March 1, 2022.

*When Russia launched its **invasion** Thursday, it fired more than 100 missiles into Ukraine, primarily at airfields and other military targets.*

The previous Thursday before March 1, 2022 was February 24. Thus, the date on which this military_strike event occurred should be recorded as 2022-02-24.

Sometimes an event may not have a calendar date specified by the news article on when it starts or ends. In such cases, the letter X is used to denote the “missing” value in the date. For example,

*Putin added Moscow would not occupy Ukraine and that it had to end **eight years** of conflict in eastern Ukraine, where Kyiv government forces have been fighting pro-Moscow separatist militants.*

According to this news article, Putin gave the speech on February 22, 2022, which means the conflict must have started sometime in 2014. But the month and date cannot be determined. So, the format 2014-XX-XX is used to denote the StartOn date of the conflict. Similarly, 2023-10-0X for “the first week of October 2023”, and 2022-0X-XX for “in the weeks before February 24, 2022.”

Temporal information is at the event level. If two events look similar but with different time stamps, the annotator must determine if they are two distinct events or the same event due to difference of granularity between time stamps. If multiple time stamps are seen at the mention level, the annotator must consolidate them to ensure the most detailed and precise time stamps are selected and preferably are attached to the first event mention. In other words, in an event coref chain, the annotator only needs to annotate one piece of temporal information but must modify it if new information warrants a more precise time stamp.

Consider again the entity argument linking example in section 4.5.1. The StartOn timestamp appears in the second event mention, but as shown in the expanded table below, it is placed in the frame of the first event mention.

Table 10. Attaching Temporal Information to Event Coref Chain

	Event mention 1	Event mention 2	Event mention 3	Event mention 4
Span	assault	attack	assault	attack
Type	military_activity.military_strike			
Attribute	ACTUAL	ACTUAL	ACTUAL	ACTUAL
Argument roles	Agent: Russia Patient: Ukraine	Patient: cities Patient: installations Place: Ukraine	Instrument: missiles Instrument: rockets Instrument: vehicles Place: territory	Agent: Russian military forces
Temporal	StartOn: 2022-02-24 EndOn: unknown			

4.7. Event Coreference

Event coreference is the act of grouping together individual event mentions that refer to the same event. To be linked to the same event, event mentions must meet the following criteria:

- intuitively the same event
- same attested scope, including both temporal and location scopes
- same type, down to the finest-grain level (event mentions of the same event cannot be different even if they are under the same hierarchy)

But individual mentions of the same event do not necessarily have to share

- same extent text
- same arguments
- same attributes (except for genericity – a generic event and a specific event cannot co-refer)

As shown in Table 10 above, coreferential mentions may have different texts and arguments (including entity arguments at different levels of granularity). Coreferential event mentions can also have different attributes. Consider the following political_election event mention:

White’s doctor told him if he might lose the election if he did not quit smoking.

Since the election had yet to happen according to this sentence (context), the event mention is assigned the attribute Irrealis. But a later event mention in the following sentence has the attribute Actual because the election had already ended:

White has vociferously complained about NBC44’s coverage of the election after losing his bid.

Despite referring to the event in different ways (i.e. having different attributes), these two event mentions still refer to the same election, so are coreferential.

Table 11. Coreferential Event Mentions with Different Attributes

	Event mention 1	Event mention 4
Handle	White lost the election	
Type	politics.voting.political_election	
Attribute	IRREALIS	ACTUAL

Event mentions of the same event can have arguments that are at different levels of granularity, but the event mentions themselves must be at the same level of granularity. An aggregate event is not coreferential with its component events or subevents.

*The group has been blamed for some of the country's worst terror **attacks**, including the **firebombing** of a passenger ferry in Manila Bay that killed more than 100 people last year.*

In the above example, “the **firebombing** of a passenger ferry in Manila Bay” is a component event of the aggregate event “some of the country's worst terror **attacks**.” Since they are at different levels of granularity with different scopes, they must be tagged as different events.

5. Relation Annotation

A relation refers to the way in which two or more entities (concrete or abstract) are connected. Both events and relations describe how entities in the world interact, but events are usually dynamic whereas relations are generally static.

In RUFEEERS, relations are limited to two entities (or an entity and entity filler), known as argument 1 and argument 2, or simply as arg1 and arg2. That means relations in RUFEEERS are all binary. Most relations defined in RUFEEERS are antisymmetric (in one direction) while some are symmetric (in either direction).

Consider for example the following sentence:

Volodymyr Zelensky is the president of Ukraine.

In this sentence, there exists a relationship between the person named “Volodymyr Zelensky” and the country named “Ukraine” such that “Volodymyr Zelensky” serves as the head of that country. In RUFEEERS, this relation type is defined as `organizationAffiliation.Leadership.HeadOfState`.

A relation mention is a particular reference to a relation between a pair of non-coreferential entity mentions. A relation can be referred to by different relation mentions:

Ukraine’s president called for his nation to fight in the streets.

Ukrainian President Volodymyr Zelensky said in a national address...

In the first sentence, the `HeadOfState` relationship is signaled by the `NOM` reference of Volodymyr Zelensky and the possessiveness of Ukraine. In the second sentence, the relationship lies in the title.

RUFEEERS tags all mentions that refer to a taggable relation, with each mention anchored to a text span (typically a word) in the document. Like event mentions, a typical relation mention is represented in a form-like frame that needs to be populated with the required information, including type, span, attribute, arguments, start time, and end time.

Table 12. Relation Mention Frame

Frame Field	Description
Relation Mention ID	A unique ID assigned to the relation mention. The ID is assigned automatically by the annotation GUI.
Type	The most specific type (furthest level in the relation ontology) that categorizes the relation. Relation mentions that refer to the same relation must be of the same type.
Mention Span	The string of text span that indicates to a human reader that the relation holds. Unlike event mentions, two or more relation mentions can have the same span in the same document.
Attribute	Features of the relation mention that indicate how the relation holds.
Arguments	The entities that participate in the relation and the roles the entities play in the relation. Each relation type has two argument roles. The definition of a relation also specifies which entity fills which argument role.
Temporal information	The start and end time of the relation indicating when the relation holds.

Consider the following sentence from the sample document:

S5. The assault was startling in its scope, as missiles and rockets rained down across the country's vast territory and Russian military vehicles rolled across the border from multiple directions.

In this sentence, there is a spatial relation between the territory where Russia launched its attack and the country Ukraine, namely, the territory is inside Ukraine. This relation mention can be represented as follows:

Table 13. Relation Mention Example

Type	spatial_relation.inside_of
Mention Span	's
Attribute	ACTUAL
Arguments	arg1: territory arg2: country
Temporal information	StartOn: unknown EndON: unknown

Relation mentions referring to the same relation are grouped into a relation coref chain. During the annotation, a relation handle describing the relation must be included with one of the coreferential relation mentions, typically the first one.

5.1. Relation Handle

A **relation handle** is a straightforward sentence or phrase in well-formed English that provides the crucial information about the relation. A good relation handle must clearly indicate the relation type as specifically as possible as well as both arguments. In a relation coref chain with multiple relation mentions, only one of them – typically the first mention – needs an explicit handle attached. The example in the preceding section can be expanded to include a handle as follows:

Table 14. Relation Mention with Handle Example

Handle	The territory is located inside Ukraine
Type	spatial_relation.inside_of
Mention Span	's
Attribute	ACTUAL
Arguments	arg1: territory arg2: country
Temporal information	StartOn: unknown EndON: unknown

5.2. Relation Types

A **relation type** is a broad category of relation, such as ownership, social_relation, spatial_relation, etc. Like events, relations are also broken down into smaller, more fine-grained

categories. Whenever possible, the most specific label available in the relation ontology is assigned. The relation ontology is presented as an expandable “tree” in the annotation GUI, allowing annotators to navigate the hierarchical structure of the nested levels of relation types to select the most specific relation type that categorizes the relation.

Annotators rely on the context and their knowledge of the annotation ontology to determine if a taggable relation type is present in a sentence. They first decide the top-level type and then decide whether a more specific/lower-level type can apply to the relation more precisely. The lowest level of relation type that best categorizes the relation is selected whenever possible. Consider the two relation mentions in the following sentence:

*S4. **Ukraine’s president** called for **his nation** to fight in the streets.*

The mention “Ukraine’s president” of the relation “Volodymyr Zelensky is Ukraine’s president” has arguments “president” and “Ukraine.” The type is determined by the very word “president.” According to the relation ontology, the finest type, `state_head_of`, under the type `organizational_relation.leader_of` best categorizes this relation. The mention “his nation”, also of the relation “Volodymyr Zelensky is Ukraine’s president”, has arguments “his” and “nation”, and is assigned the same relation type.

5.3. Relation Mention Span

A relation **mention span** is the string of text that indicates the presence of a relation in the document. For example:

*Such tactics are notoriously horrifying, trapping civilians under fire as an invading force **encircles** a city and prevents food, ammunition and medical supplies from entering.*

Here, there is a mention of the type `spatial_relation` between an army referenced by “an invading force” and a place referenced by “a city.” This relation mention is indicated by the verb “encircles”, which is selected as the mention span.

Unlike event mentions, relation mentions do not always have explicit words that solely anchor a relation. Rather, the relation is often indicated by morphology (such as adjectives, possessive markers), syntax (such as modification, word order), semantics (meaning of participating arguments such as “president”, “parent”, “friend”, etc.), and world knowledge.

*The **Ukrainian Interior Ministry** said cities ranging from Ivano-Frankivsk in the west to Odessa and Mariupol in the south were under assault.*

In the above sentence, there is an `organizational_relation.subsidiary` relation between Interior Ministry and Ukraine. The clue is morphosyntactic – the ministry is modified by the adjectival form of Ukraine.

The following **ordered** principles are used to determine mention spans of relation mentions:

1. If there is a word or phrase other than the arguments signaling the relation, use that word or phrase.

*an invading force **encircles** a city*

*Ivano-Frankivsk **in** the west*

*Odessa and Mariupol **in** the south*

*Mayor **of** New York City*

2. If the two arguments are in possessive constructions, and
 - a. if there is an apostrophe, select the apostrophe (and the following letter s if there is one)

Ukraine's president

the Johnsons' house

- b. if one of the arguments is a possessive pronoun, select the pronoun

his nation

my colleague

3. If there is no non-argument linguistic expression (except for 2.b above), and
 - a. if the relation is antisymmetric, select argument 1 in the relation template

*Russian **military***

***Kyiv**, Ukraine*

- b. if the relation is symmetric, select the first argument appearing in the text

***John** and Mary from Team USA*

So, for “Ukrainian Interior Ministry”, the mention span is anchored on “Interior Ministry” according to 3a.

An event mention span can only anchor one event mention. However, relation mention spans can anchor more than one non-coreferential relation mentions.

*Odessa and Mariupol **in** the south*

In the above phase, there are two inside_of relations: “Odessa is inside the south (Ukraine)” and “Mariupol in the south (Ukraine).” The mention span “in” anchors both relation mentions.

5.4. Relation Mention Attributes

A **relation mention attribute** is an additional piece of information about a relation. In addition to being positively asserted, relations may also be referenced in other ways, such as they do or might not hold. The following three attributes are applied to relation mentions in RUFEEERS:

- **Actual:** a relation that did or will hold.

*It taught me the power of the young **producers** at **Fox**.*

Handle: Producers employed by Fox

Type: organizational_relation.employed_by

Attribute: Actual

- **Negated:** a relation that did not or will not hold.

*He added that **Moscow** would not occupy **Ukraine**.*

Handle: Russia occupies Ukraine

Type: geopolitical_relation.controlled_by

Attribute: Negated

- **Hedged:** a relation that did not or will not hold.

*We can't tell whether the surface-to-air **missile** is **Ukrainian** or **Russian**.*

Handle: Missile owned by Ukraine

Type: ownership.owned_by.state_owned_by

Attribute: Hedged

Handle: Missile owned by Russia

Type: ownership.owned_by.state_owned_by

Attribute: Hedged

The Actual attribute is the default category and does not need to be explicitly marked. So, only Hedged and Negated attributes are tagged if applicable.

Note that attributes Generic and Irrealis as seen in events are not applicable for relations in RUFEEERS. In addition, actual, negated and hedged are all mutually exclusive properties. Therefore, all mentions of a relation will have the same attribute.

5.5. Relation Arguments

Relation arguments are participants in a relation. In RUFEEERS, every relation has exactly two arguments. There are no special labels for argument roles other than arg1 (argument 1) and arg2 (argument 2). However, for antisymmetric relations, it is important to distinguish between the two because arg1 being antisymmetrically related to arg2 excludes the possibility of arg2 being related to arg1 in the same way. The definition of the relation type explains which entity/filler can fill which argument for antisymmetric relations.

***Ukrainian** President **Volodymyr Zelensky** has condemned the shelling, calling it a “war crime.”*

In this example, there is a leadership relation between Ukraine and Zelensky, specifically the organizational_relation.leader_of.state_head_of relation, defined as “A person entity (arg1) is the head of a nation (arg2).” So, the person entity is arg1 and the GPE entity is arg2.

Table 15. Relation Argument Example

Type	organizational_relation.leader_of.state_head_of
Mention Span	President

Arguments	arg1: Volodymyr Zelensky arg2: Ukrainian
------------------	---

For symmetric relations, it is recommended to assign arg1 to the first occurring argument in the first occurring mention of the relation and continue to assign arg1 to the same argument in later mention of the same relation.

The following is an example of symmetric relation:

...the country's second-largest city, which lies close to the border with Russia.

Table 16. Symmetric Relation Argument Example

Type	spatial_relation.close_to
Mention Span	close to
Arguments	arg1: city arg2: Russia

RUFEERS requires that for every relation mention tagged, both arguments must be included in the relation mention frame. That does not mean both arguments are always in the same sentence where the relation mention is detected.

Jane married John 10 years ago. She ended the marriage last month.

The word “marriage” in the second sentence indicates that Jane was John’s spouse as of last month. But there is no entity mention referring to John in this sentence. In such cases, the annotator should link the entity/filler mention closest to the relation mention span. So “John” in the first sentence is linked to the relation mention and fills the arg2 role.

Table 17. Mandatory Relation Argument

Type	person_association.interpersonal_relation.relative_of.spouse_of
Mention Span	marriage
Attribute	Negated
Arguments	arg1: Jane arg2: John

An entity/filler mention can also serve as an argument in more than one relation, even in the same sentence. Consider again the two “inside_of” relation example, repeated below.

*Odessa and Mariupol **in** the south*

The location “south (Ukraine)” appears as arg2 in two relation mentions, “Odessa is in southern Ukraine” and “Mariupol is in southern Ukraine.”

5.6. Temporal Information: Start/End

Annotating temporal information for relations follows the exact same practice as for annotating event temporal information. Both start and end timestamps must be tagged, each with a choice of on, before, after, or unknown.

Likewise, temporal information is at the relation level. So, only one relation mention needs to be annotated with timestamps. If there is more than start or end timestamp, the annotator

must make sure they are for the same underlying relation and select the correct one (e.g. earlier “start on” or later “end on”). If there are two very different timestamps for the seemingly same relation, it is more likely there are two relations of the same relation type involving the same arguments but over two different periods of time.

For details on how to annotate temporal information, refer to Section 4.6.

5.7. Relation Coreference

Relation coreference is the grouping together of individual relation mentions that refer to the same relation. Two or more relations are coreferential if and only if they

- are of the same relation type (down to the finest level),
- have the same arguments with the same argument roles, and
- have/assume the same temporal information

The annotator will link all the mentions of a relation as a relation coref chain. In the following two sentences from the sample document, there are three relation mentions of the relation “Zelensky is Ukraine’s president”: “Ukraine’s president”, “his nation” and “Ukrainian President Volodymyr Zelensky.”

S4. Ukraine’s president called for his nation to fight in the streets.

S14. Ukrainian President Volodymyr Zelensky said in a national address

The coref chain can be represented in the following table.

Table 18. Coreferential Relation Mentions

	Mention 1	Mention 2	Mention 3
Handle	Zelensky is Ukraine’s president		
Type	organizational_relation.leader_of.state_head_of		
Mention Span	’s	his	President
Arg1	president	his	Volodymyr Zelensky
Arg2	Ukraine	nation	Ukrainian

6. Entity Annotation

An entity is a unique object or concept or set of objects or concepts in the world – for instance, a person, place, organization, etc. For example:

- **Person:** Adam Smith
- **Organization:** World Health Organization
- **Location:** North America
- **Geopolitical Entity:** United States
- **Facility:** Philadelphia International Airport
- **Weapon:** nuclear bomb
- **Vehicle:** U.S.S. Cole
- **Consumer goods:** facia mask
- **Astro object:** the moon
- **Treaty:** Treaty on the Non-Proliferation of Nuclear Weapons

An entity mention is a single occurrence of a name, nominal phrase, or pronominal phrase that refers to, or describes, an entity. For example, the person entity George Washington can be mentioned as President George Washington, President Washington, the first US president, or he/him/I in different contexts, all of which are various mentions of the same entity.

In RUFES (**R**ecognized **U**ltra-**F**ined **E**ntities), all entities that can be categorized by the RUFES entity ontology must be annotated. RUFES limits taggable entities to those that must fill event/relation argument roles in the document.

All mentions of taggable entities must be annotated even if a particular mention does not appear in a taggable event or relation mention as an argument. A coref chain for each entity is built by linking all entity mentions that refer to the same entity. Each entity has its own coref chain, and no entity mention can be linked to more than one entity coref chain.

A complete entity mention can be represented in a form like frame as follows.

Table 19. Entity Mention Frame

Frame Field	Description
Entity Mention ID	A unique ID assigned to the entity mention. The ID is assigned automatically by the annotation GUI.
Entity Handle	An unambiguous noun phrase such as a proper name or common noun phrase depicting the entity – only required of one (preferably the first one) mention in a coref chain.
Mention Type	One of three noun phrase types: NAM, NOM and PRO
Mention Span	The string of text span that the entity mention is anchored to. Each entity mention has a unique span and two entity mentions cannot have the same span in the same document.

Entity type	The finest type from the entity ontology that categorizes the entity – only required of one (preferably the first one) mention in a coref chain.
--------------------	--

An entity coref chain will then consist of the handle, the entity type and all the entity mentions that refer to the same entity.

6.1. Entity Handle

Each entity coref chain is described with an **entity handle**, which can be a proper name (e.g. a person’s legal name) or a variant of it (e.g. last name for a person, acronym for an organization, etc.) for that entity, or a phrase that accurately describes the entity if a proper name is not available for the entity (e.g. the man with a red hat, the building burnt down, etc.). The following shows some examples of acceptable entity handles:

Table 20. Entity Handle Examples

Entity	Handle(s)
William Jefferson Clinton	Bill Clinton, President Clinton, Former President Clinton
Hillary Diane Rodham Clinton	Hillary Clinton, Secretary Clinton, Mrs. Clinton
World Health Organization	WHO
National Institute of Standards and Technology	NIST
The apartment complex burnt down from the fire last Saturday	the building burnt down
Cities in Ukraine	Ukrainian cities
Cities in Russia	Russian cities

Annotators must avoid ambiguity when the seemingly same entity mention refers to an entity of different types. For example, “Harvard University” may refer to the university as an educational institution (ORG type) or as a college campus (FAC type) depending on the context. If both references are seen in the document, there are really two different entities, requiring two different handles.

*John graduated from **Harvard University**.*

Handle: Harvard_ORG

*John drove by **Harvard University**.*

Handle: Harvard_FAC

*John drove by **Harvard University**. He graduated from **that university** 20 years ago.*

- **Harvard University** (as FAC): Harvard_FAC
- **that university** (as ORG): Harvard_ORG

Similarly, if a company operates multiple facilities in multiple locations, it is best to include the location for each individual facility and add a descriptor when it refers to the organization itself, for example, McDonald’s at Tysons Corner, McDonald’s in Silver Spring, McDonald’s_ORG.

A pronoun cannot be used as an entity handle. The annotator will resolve a pronoun's reference and come up with a handle similar to the examples above. In the case of a pronoun making a generic reference, e.g. "You shall not steal", a handle like "generic people" is a better option.

When a reporter puts themselves in the news article by using the first-person pronoun "I" or "we" but nowhere in the article is the reporter's name mentioned, a handle like "author of article" can be used.

6.2. Entity Mention Types

Entity mention types – not to be confused with entity types (see section 6.4) – are semi-grammatical classifications of noun phrases. RUFEEERS distinguishes between three types of mentions – named mention, nominal mention, and pronominal mention.

- A named mention (**NAM**) is a proper name, or any of its variants such as acronym, nickname, abbreviation, or alias, etc. that refers to an entity. For example, James Comey, the FBI, the American Civil Liberties Union, the ACLU, Smith, and Congress are all NAMs. Proper adjectives, namely adjectives derived from proper nouns, are also taggable (see section 6.3.1 for more details).
- A nominal mention (**NOM**) is a noun phrase headed by a common noun, for example, the former FBI director referring to James Comey.
- A pronominal mention (**PRO**) is a noun phrase consisting of a pronoun, for example, the third person singular pronoun "he" referring to someone previously referenced by his full name.

The following is a sentence showing all three types of mentions referring to the same entity, the Ukrainian president Volodymyr Zelensky.

S4. Ukraine's president Volodymyr Zelensky called for his nation to fight in the streets.

- NAM: Volodymyr Zelensky
- NOM: president
- PRO: his

6.3. Entity Mention Span

An entity mention is a string of text referring to an entity. While one might simply say an entity mention corresponds to the entire noun phrase referring to the entity, treating the entire phrase as the mention span regardless of whether the mention is NAM, NOM, or PRO has negative impact on other taggable entity mentions, as nested mentions (a mention within another mention) are not allowed. How a mention span is marked in RUFEEERS depends on the mention's type.

6.3.1. NAM Mention Span

The taggable span of a NAM mention is the entire proper name excluding any determiner (except where the determiner is an integral part of the name), or any modifier that appears before or after the proper name (premodifier or postmodifier) but is not an integral part of the name.

the FBI -> the [FBI], not [the FBI]

Volodymyr Zelensky -> [Volodymyr Zelensky], not [Volodymyr] [Zelensky]

a defiant Volodymyr Zelensky -> a defiant [Volodymyr Zelensky]

Certain proper nouns in English have adjectival forms known as proper adjectives. These adjectives are treated as NAM and are tagged the same way as their nominal counterparts.

the American Ambassador to Japan -> the [American] Ambassador to Japan

Russian-made -> [Russian]-made

NAM mentions are atomic, meaning no additional mention should be tagged within the mention span even if some words may seem to be mentions referring to other entities.

the University of Maryland -> the [University of Maryland], not the [University] of [Maryland]

the United States Department of State -> the [United States Department of State], not the [United States] [Department of State], nor the [[United States] Department of State]

But modifiers that are not part of the name should not be included in the mention span as shown below.

Georgetown University in Northwest D.C. -> [Georgetown University] in [Northwest D.C.]

University of Oregon in Eugene, Oregon -> [University of Oregon] in [Eugene] , [Oregon]

Sometimes it can be difficult to decide if a modifier is an integral part of the name, for example “the Chinese Foreign Ministry” vs “China’s Foreign Ministry.” Intuitively, “Chinese” feels much closer to “Foreign Ministry” than “China’s” as shown by where an option adjective modifier can be placed grammatically.

A defiant Chinese Foreign Ministry

China’s defiant Foreign Ministry

With this modifier test, the pair of examples are best annotated as follows.

the Chinese Foreign Ministry -> the [Chinese Foreign Ministry]

China’s Foreign Ministry -> [China]’s [Foreign Ministry]

Similarly,

the ACLU’s Washington Legislative Office -> the [ACLU] ’s [Washington Legislative Office], not the [ACLU’s Washington Legislative Office]

the ACLU's little known Washington Legislative Office -> the [ACLU] 's little known [Washington Legislative Office]

Titles, honorifics, professional credentials, etc. that precede or follow personal names not integral parts of personal names and thus should not be included in the mention span.

President [George Bush]

[Jane Doe], M.D.

On the other hand, generational suffixes such as Jr., Sr, II, etc. are an integral part of personal names and must be included in the mention span.

Henry Ford III -> [Henry Ford III], not [Henry Ford] III

Note that although determiners are in general excluded from the mention span, there are exceptions, for example, "The" in "The Rock" (a stage name for the wrestler/actor Dwayne Johnson) or in "The Hague" (a European city). In such special cases, the determiner must be included in the mention span.

6.3.2. NOM Mention Span

For a nominal mention, i.e. a noun phrase headed by a common noun, only the head is selected as the mention span.

S4. Russia on Thursday launched an attack on [cities] and military [installations] across Ukraine, forcing thousands of [civilians] to flee.

It was mentioned earlier that nested mentions are not allowed. Since only the head of a NOM mention is selected as the mention span, taggable mentions within the full boundaries of a NOM mention can and must be annotated. In the following two examples, both FBI and Bush are taggable NAM entity mentions.

the new FBI director -> the new [FBI]_{NAM} [director]_{NOM}

the Bush administration -> the [Bush]_{NAM} [administration]_{NOM}

A compound noun consists of two or more words. There are three types of compound nouns. A closed compound noun such as "supermarket" or "firefighter" is always treated as one word as there is no white space separating "super" and "market" or "fire" and "fighter." For open compounds, it is not always easy to tell them apart from a noun with a modifier. For simplicity, only the head of an open compound is selected as the mention span.

vice [president]

assistant [professor]

attending [nurse]

chief [justice]

[attorney] general

[editor] in chief

Note the head of a compound in English is not always at the end of the compound as show in the last two examples above.

However, hyphenated compounds are treated as “atomic” words, same as closed compounds.

[editor-in-chief]

[father-in-law]

A common noun phrase is not always headed by a noun. For example, both adjectives and numbers can be the “head” of a noun phrase, where the “true” head is omitted. In such cases, the adjective or number is selected as the mention span:

The [rich] are getting richer.

There were five [victims]; [two] were children.

[One] of the five [victims] was a child.

Of the 7,160 patients whose chronic illness status was known through health records, [184] died, and [173] of them had an underlying condition, the CDC said.

Among all the cases analyzed, [10.9%] of patients had diabetes mellitus, [9.2 percent] had chronic lung disease, and [9 percent] had cardiovascular disease.

It should be noted that a proper noun can be used in a common noun phrase, for example, a family name referring to a member of the family, a vehicle manufacturer or model name referring an individual product, or the name of an ethnic group referring to a member from the group.

He met a [Kennedy]_{NOM}

She drove a [Toyota]_{NOM}

John talked to a [Guatemalan]_{NOM} attempting to immigrate to the U.S.

Note how the last sentence differs from the following one:

John met a [Guatemalan]_{NAM} [immigrant]_{NOM} yesterday.

6.3.3. PRO Mention Span

A pronominal mention is a noun phrase consisting of a pronoun. The span to be selected for a taggable pronominal mention is simply the pronoun itself. The difficulty lies in anaphoric resolution. This topic of coreference will be covered in more details in section 6.5.

The following is an (almost) exhaustive list of potentially taggable pronouns:

- Personal pronouns
 - Subjective: *I, we, you, he, she, they, it*
 - Objective: *me, us, you, him, her, them, it*
 - Possessive

- Possessive determiners: *my, our, your, his, her, their, its*
- Possessive pronouns: *mine, ours, yours, his, hers, theirs, its*
- Demonstrative pronouns: *this, that, these, those*
- Reflexive pronouns: *myself, ourselves, yourself, yourselves, himself, herself, themselves, themselves, itself, oneself*
- Reciprocal pronouns: *each other, one another (both words are selected as the head)*
- Relative pronouns: *who, whom, whose, which, that, whoever, whomever*
- Dual pronouns: *both, either, neither*
- Positive indefinite pronouns: *all, some, one, many, everything, everybody*
- Negative indefinite pronouns: *no one, nobody, nothing, nowhere*
- Proadverbs and relative adverbs: *there, where*

6.3.4. Conjoined Mentions

When two or more mentions are conjoined, they must be tagged separately even though they function as a single grammatical unit such as the subject or object of a sentence:

Boeing and Lockheed Martin -> [Boeing] and [Lockheed Martin]

Mention selection must be continuous. Two discontinuous words cannot be selected to form a single mention span. In the example below, “George” and “Bush” cannot form a span.

George and Laura Bush -> [George] and [Laura Bush]

Conjoined mentions refer to separate entities. If they are arguments of an event, they can be annotated as separate participants in the same event. Furthermore, labeling them as separate mentions may reveal a taggable relation if it exists as in the case of the Bush example above.

6.4. Entity Types

The RUFEEERS entity types are the same as RUFES entity types. They are organized in three levels, and the annotator must find the applicable type of the lowest level. There is a total of 356 entity types: 19 at the very top, 115 at the second, fine-grained level, and 356 at the third, ultra-fine-grained level. The detailed types, subtypes, and sub-subtypes including their definitions can be found in the entity ontology document, and annotators have access to them in the annotation tool as well.

There are two closely related restrictions on what entity type label(s) to assign to an entity: context and usage.

6.4.1. Scope of Context

Just as a news reporter expects the reader to have some level of world knowledge (for example, an American reader would know what “FBI” refers to), selecting a proper entity type also assumes some level of world knowledge on the part of the annotator. Most readers indeed know that the FBI is a US government agency, hence will assign ORG.Government.Agency as its entity type.

However, in RUFEEERS (and RUFES), world knowledge is not the first and foremost basis for entity type labeling for two main reasons: (1) an entity in the real world can be categorized into many different types without context and (2) an entity may not be known to the annotator. Instead, the annotator will assign an entity type or entity types at the mention level based on what it means and how it is used in the local context, defined as the sentence in which it appears. Consider the following paragraph as an example.

[Zelenskyy] grew up as a native Russian speaker in Kryvyi Rih, a major city of Dnipropetrovsk Oblast in central Ukraine. Prior to [his] acting career, [he] obtained a degree in law from the Kyiv National Economic University.

The mention of “Zelenskyy” in the first sentence would only be assigned “PER.” In the second sentence, there are two PRO mentions of him, “his” and “he.” The first PRO mention indicates he was once an actor. So, it is assigned the type of PER.Artist. The second PRO mention is assigned PER.Student because he studied at Kyiv National Economic University. All three types are assigned independent of each other. None of them is assigned the type PER.Politician.HeadOfNation.

Below is the opening sentence from a news report about Amy Coney Barrett’s Supreme Court Justice nomination hearings.

Senate Republicans moved swiftly Thursday toward confirming Supreme Court nominee [Amy Coney Barrett] before the Nov. 3 elections.

What entity type should be assigned to the mention “Amy Coney Barrett”? If the annotator follows the news, they will know she has been a circuit judge since 2017. So, they might be tempted to tag it as PER.CivilServant.Judge. But the sentence itself does not disclose this information. In fact, being a current judge or even a lawyer is not even one of the qualifications of a Supreme Court Justice nominee. Therefore, this mention’s entity type should only be labeled as PER (or as PER.PoliticalNominee if this entity were included in the RUFEEERS entity ontology).

A later sentence however indicates Amy Coney Barrett is a judge.

“Judge [Barrett] as Justice [Barrett] may well cast the deciding vote to overturn the Affordable Care Act, with potentially disastrous consequences for a majority of Americans,” Sen. Christopher A. Coons (D-Del.) said Thursday.

So, in this sentence, PER.CivilServant.Judge should be assigned.

In general, information external to the news article should be avoided to determine entity types. There are exceptions, though. For example, if an unfamiliar location or landscape in a

foreign country appears in an article but the annotator cannot determine if the type should be GPE or LOC based on the context, they may research external sources to make a better choice. But this approach should not be used liberally.

6.4.2. Tag for Usage/Meaning

Choice of entity type is not only confined by the context, but also by the intended meaning. This has been known as the “tag for usage” or “tag for meaning” rule in various entity detection annotation tasks.

A very common figure of speech in journalist writings is metonymy, where another name closely associated with the intended entity is used instead of the entity’s official name, for example, “Pentagon” referring to the US military, “the White House” referring to the US government executive branch, etc. In such cases, the entity type assigned to the mention is determined by the intended entity being referred to.

The [White House]ORG.Government.Administration just announced sanctions against Russia for its invasion of Ukraine.

The ceremony will be held in the [White House]FAC.Building.GovernmentBuilding.

In this case, the two entity mentions above are not coreferential and cannot be linked to the same entity coref chain. Refer to section 6.5 for details on coreference.

Other common examples of metonymy include the capital city’s name referring to the country or state government, a city name referring to the city’s sports team(s), etc.

Facilities are operated by organizations, and often the same name may be used for both kinds of entities. For example, McDonald's may refer to the corporation (as in “McDonald's pays a large sum of bonus to its executives”), which is of the type ORG.CommercialOrganization, or one of its restaurants around the corner (as in “Jonny bought a kid’s meal at McDonald’s”), which is of the type FAC.Building.StoreShop (the RUFEEERS entity ontology does not have restaurant as an entity type). Other examples are hospitals, educational institutes, etc.

Sometimes a generic reference to people is intended to refer to their organization, their country, etc. For example, “the Democrats” is used to refer to the Democratic Party (as in “The Democrats won the House”), or “the Americans” to the United States (as in “The Americans declared war on the Japanese on December 8, 1941”), or “(government) officials” to the government, etc.

6.4.3. Multiple Entity Types

Given the right context, more than one entity type may be applicable to an entity mention.

CEO [Mike Johnson] arrived at INOVA in Fairfax, Virginia, Friday evening after he started to experience a low-grade fever, chills, nasal congestion and cough.

In the sentence above, the title “CEO” indicates Mike Johnson was a company executive. So, the mentions “Mike Johnson” and “he” can be assigned PER.Executive.BusinessExecutive. He was

experiencing some sickness and checked into a hospital. So, the type PER.SituationalRole.Patient is also applicable. In RUFES, both types must be labeled for the two entity mentions. As the focus in RUFEEERS is on events and relations, only one entity type – the most salient type for the context – is required for a given entity mention for simplicity. In this example, “Mike Johnson” being a patient is more salient as there are two events involving him as a patient. Therefore, the two entity mentions are labeled as “PER.SituationalRole.Patient” in this context.

6.5. Entity coreference

Coreference is a relationship between mentions – if two or more mentions in an article refer to the same entity, they corefer, for example, “James Comey” and “the new FBI director”, “the ACLU” and “the American Civil Liberties Union”, etc.

A major difference between entity mention coreference and event/relation mention coreference is that different mentions of the same entity can be assigned different entity types depending on the context where the mention occurs and its usage whereas mentions of the same event or relation must be of the exact same event or relation type. For example, “Bill Clinton” may be tagged as PER.Politician.Governor in one paragraph and as PER.Politician.HeadOfNation in another in the same document. Both mentions refer to the same person and are thus coreferential.

But there is one condition: coreferential entity mentions must be of the same top-level entity type. Otherwise, they are not coreferential but refer to different entities. For example, “Philadelphia” as GPE (as in “Philadelphia raised its property tax rate”) is not coreferential with “Philadelphia” as ORG (as in “Philadelphia won the football game”). Likewise, “Harvard” as ORG cannot be coreferential with “Harvard” as FAC.

While mentions of different subtypes may be coreferential, the opposite is not necessarily true. For example, “Kyiv” tagged as GPE.Country referring to the country “Ukraine” (as in “Kyiv will not negotiate with Moscow on its sovereignty”) is different from “Kyiv” tagged as GPE.City referring to the capital city of Ukraine (as in “the Mayor of Kyiv”). They refer to two different entities in such cases even though they are of the same top-level GPE type.

Annotators must pay close attention to pronoun resolution. Pronouns are classified as exophoric (referring to something in the context of the speaker), which is not common in written language, or endophoric (referring to something mentioned in the surrounding text), which can be either anaphoric (backward reference) or cataphoric (forward reference).

Although RUFEEERS does not require every entity mention to be assigned a handle so long as one of the coreferential mentions (preferably the first one in the document) is tagged with a handle, it is highly recommended that annotators add a handle to every PRO mention. This will make linking them to the correct coref chain much easier.

6.6. Argument Filler Annotation

In addition to the entity types described above, arguments of events and relations can also be **argument fillers**, abbreviated to **fillers**. Fillers differ from entities in that they cannot be categorized by the entity types in the current entity ontology but still can fill argument roles for events and relations. Argument fillers are annotated differently from entities in three ways:

- Additional mentions of fillers that have been tagged as an argument of an event/relation are **not** exhaustively annotated. That is, only fillers that are themselves arguments are tagged.
- Mention levels of argument fillers by default are set as NOM.
- Argument fillers do not undergo coreference annotation.

Argument_fillers as place-holder types are included in the list of entities in the RUFEEERS entity ontology. They were not annotated in RUFES but are added to RUFEEERS to support event and relation annotation.

The following are the argument filler types taggable in RUFEEERS.

- **INF**: Information Object. The text extent of INF is the string that describes the content or topic of some communication. It might be a noun, noun phrase, or part of a verb phrase.
She asked him [to memorize the speech by Monday].
He lied about [not knowing the victim].
leaving the lawyers to negotiate [the terms of the deal]
- **MON**: Money. The extent of a MON mention includes quantifiers, amounts, and the currency unit, all of which may be optional.
[Nearly \$400 Million] was given to the charity in 2010.
[20 Euros] bought him a nice meal in Granada!
Imagine if you wanted to pay for the cows with [bitcoins].
We paid [400] for it.
- **RES**: Result of an election/vote. The text extent of RES is the string that describes the result of an election/vote.
Multiple violations of [the results] of the referendum were observed.
- **SEN**: Sentence. The text extent of SEN is the string that describes the sentence of a legal_proceeding event.
Ms. Ellis agreed to be sentenced to [five years of probation], pay [\$5,000 in restitution] and perform [100 hours of community service].
- **TTL**: Title. The text extent of TTL is the extent of the title or job name, which will generally be either a noun or a compound noun, or may be a noun phrase.
[Vice President] Dick Cheney

[spokesperson] Mary Gillette

7. Towards Successful Annotation and Future Improvements

At the time of the first public release of the annotation guidelines, RUFEEERS is in the second year of annotation. Prior to RUFEEERS, NIST had completed three years of RUFES annotation. This section summarizes experiences and lessons learned and suggests some future directions.

7.1. Annotation Guidelines and Ontology

RUFEEERS system evaluation is ground truth based, which requires “gold standard” evaluation data to be created by human annotators. Annotators need detailed annotation guidelines that are easy to understand and to follow. They cannot begin without well-established guidelines, and guidelines also need to be continuously updated to address newly discovered issues during annotation. But human languages are complicated and there are always exceptions to patterns and rules. Annotators may have different interpretations of lesser-known linguistic phenomena or deviations from norms. To achieve a high-level agreement among annotators, annotators are encouraged to use a chat group to discuss difficulties they encounter during annotation.

RUFEEERS uses a closed ontology set at each phase. The entity ontology includes a filler type “Thing” to accommodate the scenario when there is no existing entity type for an entity argument in a taggable event or relation. But there is no such “general purpose” event or relation type as it will lead to “over-annotation.” If the annotator thinks an event or relation not covered by the existing ontology is important for the document, the current practice is that they make a note so that a new type will be incorporated into the next phase’s ontology.

It may be worth exploring a different approach: exhaustive annotation of all events and relations with an open ontology on a dataset with much shorter text documents.

7.2. Talent Acquisition

Although the source data for RUFEEERS are news stories and native speakers with basic educational background have no trouble reading and understanding them, the annotation task involves annotators’ understanding of rather sophisticated linguistic concepts to be able to annotate the data competently. A lesson learned from the first year’s RUFES annotation – when existing accessors were tasked to do the annotation was that it is difficult if not impossible for annotators without linguistic background to be trained in a short period of time to acquire the needed knowledge.

Starting from the second year of RUFES, NIST has been recruiting college students majoring in linguistics as well as professional linguists with prior annotation experience for the tasks of RUFES and RUFEEERS. Having a team of long-term qualified annotators can not only speed up the process but also reduce costs.

7.3. Annotation Tool

Every data annotation task needs an annotation tool. Organizations may develop proprietary tools for specific tasks, but they are expensive to build and support and may not be easily

modified for different tasks. For example, RUFES used a tool provided by a university research lab, but it was designed for entity detection and linking only. It was not possible to adapt it for annotating events and relations without significant technical investment.

Fortunately, researchers in natural language processing (NLP) and information retrieval (IR) have been developing open-source tools that can be configured and scaled for different linguistic annotations. For RUFEEERS, NIST selected [INCEption](#), a semantic annotation platform developed by the Ubiquitous Knowledge Processing (UKP) Lab at the Department of Computer Science, Technische Universität Darmstadt, Germany. It meets RUFEEERS' needs but with one caveat: its heavily visual annotation interface requires large screen real estate. This is particularly challenging for linking coreferential entity, event or relation mentions that are far away from each other in the document, even though source documents longer than 600 words are already truncated.

7.4. Quality Control

There are different approaches to annotation quality control. Dual annotation followed by adjudication is considered the best option, but it is also a very expensive option as it requires three different annotators for each document. RUFEEERS uses a simpler approach, a first pass annotation followed by a second quality control (QC) pass whereby a different annotator (usually with experience) inspects the first pass annotation, adds misses, and corrects errors.

To make the QC process more efficient, the original INCEption output is converted into separate entity, event, relation and argument files in tab-separated value (tsv) format that are easier for humans to read and interpret. The QC annotator can sort the files in different ways, look for obvious mistakes, and correct them in the annotation interface.

The conversion itself checks for errors from not conforming to annotation guidelines (e.g. time stamps in incorrect format). These errors are saved in a separate file, and the QC annotator corrects them in the annotation interface. The conversion is run multiple times until no further such errors exist.

In the future, it is worthwhile looking into adding some error prevention features to the annotation tool, depending on its capabilities. For example, currently time stamps are manually entered. The annotator will be less likely to make formatting errors on specific time stamps if they can click on a calendar and select the date.

7.5. Looking forward

The focus of RUFEEERS (and RUFES) is on fine-grained types. But there are other areas of entities, relations, and events (ERE) worth exploring in the future.

7.5.1. Cross-document linking and cross-lingual annotation

Currently, coreference chain annotation in RUFEEERS is at the document level, primarily because of the single source of input data (i.e., Washington Post news articles). Cross-document

entity/event/relation linking requires a knowledge base (KB) where specific entities, events and relations are stored so that the annotator can add entries to the KB and link new mentions to KB entries if they already exist. This will call for diverse source data, possibly in other media such as audios and videos as well, and careful data selection so that the same entities, events and relations are found in different sources.

Another area worth exploring is to expand RUFEEERS to other languages. While ontologies are language-independent, languages differ on how event, relation and entity mentions are realized not just phonologically and orthographically but also morphologically and syntactically, which means language-specific guidelines on how to annotate them will be needed.

7.5.2. Event relations

One feature that can be added into the next round of RUFEEERS is event relations. In the current implementation of RUFEEERS, events are annotated as if they were isolated. In the real world, many events are connected to each other in one way or another. A macro event may consist of many micro events, for example, Russia's invasion involving ground troops moving into towns and cities, missile attacks on military and civilian targets, etc. An event may result from another event, for example, a person's death from bombing. Or two events may occur one after another without causal effect or they occur at the same time. Extracting relations between events is a worthwhile effort.

7.6. Conclusion

This publication is based on internal RUFEEERS annotation guidelines, which are less formal and contain directive language (e.g. frequent use of first and second person pronouns). It is intended for stakeholders and potential performers in RUFEEERS evaluation who are interested in learning more about how evaluation data is annotated. As the annotation is ongoing, some details in the internal version(s) may have changed and may not be reflected in this release.

Phase-specific task specifications and ontologies are distributed directly to system performers. No ontology is included in this publication, but the latest version is available upon request.

References

- [1] Aguilar, Jacqueline, et al. (2014) *A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards*. (Proceedings of the second workshop on EVENTS: Definition, detection, coreference, and representation).
- [2] Song, Zhiyi, et al. (2015) *From light to rich ERE: Annotation of entities, relations, and events* (Proceedings of the 3rd workshop on EVENTS: Definition, detection, coreference, and representation).
- [3] National Institute of Standards and Technology (2022) *RUFES 2022 Annotation Guidelines (V1.0)*. Available at <https://tac.nist.gov/2022/KBP/RUFES/guidelines/TAC-KBP2022-RUFES-AnnotationGuidelinesV1.0.pdf>
- [4] Walker, Christopher, et al. (2006) *ACE 2005 Multilingual Training Corpus LDC2006T06* (Linguistic Data Consortium, Philadelphia, PA). <https://doi.org/10.35111/mwxc-vh88>
- [5] Chen, Song, et al. (2023) *DEFT English Light and Rich ERE Annotation LDC2023T04* (Linguistic Data Consortium, Philadelphia, PA). <https://doi.org/10.35111/ctq4-mn59>
- [6] Chen, Song, et al. (2025) *KAIROS Schema Learning Complex Event Annotation LDC2025T07* (Linguistic Data Consortium, Philadelphia, PA). <https://doi.org/10.35111/g7sc-nt96>
- [7] The INCEpTION Team *INCEpTION User Guide* (version 40.0). Available at <https://inception-project.github.io/releases/40.0/docs/user-guide.html>
- [8] Birnbaum, Michael, et al. (2022, February 24) *Russia unleashes military assault on Ukraine* (The Washington Post). <https://www.washingtonpost.com/national-security/2022/02/23/russia-attack-ukraine/>
- [9] Doddington, George, et al. (2004) *The Automatic Content Extraction (ACE) Program – Tasks, Data, and Evaluation*. (Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal).
- [10] Grishman, Ralph (1994) *Whither Written Language Evaluation?* (Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994).

Appendix A. Annotation Procedure

Below is a recommended procedure for RUFEEERS annotation using INCEpTION:

1. Quickly read through a document to get a sense of the topic and likely taggable events/relations. Reject document if there are no taggable events/relations.
2. Quickly read through a document to get a sense of the topic and likely taggable events/relations. Reject document if there are no taggable events/relations.
3. Begin with the first sentence of text, identify the taggable events/relations in the sentence.
4. Create the event/relation mentions by selecting the span of taggable event/relations.
5. Link coreferential event mentions as event coref chain and populate the event frame for the first mention of the coref chain: event handle → type → Label Attribute → Label Start and End times. For singleton (event mentions that do not have any coreferential mentions), the mention frame needs to be fully populated.
6. Link coreferential relation mentions as relation coref chain and populate the relation frame for the first mention of the coref chain: event handle → type → Label Attribute → Label Start and End times. For singleton (relation mentions that do not have any coreferential mentions), the mention frame needs to be fully populated.
7. Annotate arguments/entities: create entity mentions that are event arguments by selecting mention spans. For each entity mention, also select mention level.
8. Link entity mentions to event or relation mention as arguments: select argument role → link entity to the argument role.
9. If an argument has negated or hedged attribute, create the mention at the argument_attribute level and link the argument_attribute mention as arguments.
10. Annotate entity mentions that are not arguments of an event or relation mention, but coreferential with any existing entity mentions that are arguments of an event or relation mention.
11. Link coreferential entity mentions as entity coref chain and populate the entity frame for the first mention of the coref chain: entity handle → Select entity type.
12. For a mention in the same entity coref chain, decide the entity type based on context if it is different from the entity type of the first mention.
13. Review all annotations to confirm completeness and correctness and move on to next document

Appendix B. Sample Source Document

The following is the sample document presented in a table format where most examples in the annotation guidelines come from.

Table 21. Sample Document

Sentence ID	Sentence Text
1	Russia unleashes military assault on Ukraine
2	By Michael Birnbaum, Mary Ilyushina, Paul Sonne and Isabelle Khurshudyan
3	Updated February 24, 2022 at 8:51 a.m. EST Published February 23, 2022 at 7:43 p.m. EST
4	Russia on Thursday launched an attack on cities and military installations across Ukraine, forcing thousands of civilians to flee, as Ukraine’s president called for his nation to fight in the streets and NATO leaders said Europe’s security had been fundamentally altered.
5	The assault was startling in its scope, as missiles and rockets rained down across the country’s vast territory and Russian military vehicles rolled across the border from multiple directions.
6	It came shortly after Russian President Vladimir Putin declared on state television that Russia was beginning a military operation for the “demilitarization and denazification” of eastern Ukraine.
7	He said that Russia did not intend to occupy the country.
8	Flashes could be seen and explosions could be heard in Kyiv, as highways became clogged with people trying to get out of the city of 3 million.
9	Lines formed at a border crossing between Ukraine and Poland.
10	A senior Ukrainian official said the capital’s main airport was under assault.
11	Explosions could also be heard in Kharkiv, the country’s second-largest city, which lies close to the border with Russia.
12	The Ukrainian Interior Ministry said cities ranging from Ivano-Frankivsk in the west to Odessa and Mariupol in the south were under assault.
13	“Ukraine is defending itself and will not give up its freedom, no matter what Moscow thinks.
14	For Ukrainians, independence and the right to live on their land according to their will is the highest value,” Ukrainian President Volodymyr Zelensky said in a national address.
15	He said that his country was being attacked “from the north, east and south.”
16	Ukraine’s Western backers scrambled to react.
17	European leaders vowed crushing sanctions against Russia and planned to meet later Thursday to discuss them.
18	NATO ambassadors gathered for emergency consultations in Brussels to bolster the alliance’s own defenses against Russia and also to discuss how to further support Kyiv.
19	“It will be a new Europe after the invasion we saw today,” NATO Secretary General Jens Stoltenberg said after a first round of meetings.
20	The alliance planned to hold an emergency summit of its leaders.
21	Calling Russia’s moves “a brutal act of war,” Stoltenberg said there will be more alliance forces in the east “in the coming days and weeks.”
22	“Peace on our continent has been shattered,” he told reporters in Brussels.

23	“This is a deliberate, coldblooded and long-planned invasion.
24	Russia is using force to try to rewrite history.”
25	The explosions marked the beginning of a conflict that Ukraine and its allies had worked for months to avoid.
26	“The prayers of the entire world are with the people of Ukraine tonight as they suffer an unprovoked and unjustified attack by Russian military forces,” Biden said in a statement late Wednesday.
27	“President Putin has chosen a premeditated war that will bring a catastrophic loss of life and human suffering.
28	Russia alone is responsible for the death and destruction this attack will bring.”

Appendix C. Glossary

Annotation

Act of explicitly marking elements of interest within a document. Also known as tagging.

Argument

An entity, event or relation mention that fills certain argument roles of an event or relation mention to indicate its participation in the event or relation.

Argument Filler

Something that fills an argument role in an event or relation that is not categorized with the entity types in the current entity ontology, such as a job title, an amount of money, etc. Often abbreviated to filler.

Coreference

A relationship between two or more mentions such that they refer to the same entity, relation or event.

Document

For this task, a document is the entire text of a news article (possibly truncated).

Entity

A physical object or a conceptual reality that exists in some universe (typically the real world in which we live).

Extent/Span

The span of text that an entity, relation or event mention that anchors to in a given text document.

Extent/Span

The span of text that an entity, relation or event mention that anchors to in a given text document.

Event

Something that happens in the world at a specific time at a specific place.

Handle

A brief description of the event, entity or relation.

Head

The word in a phrase that determines the syntactic category of a phrase. In a noun phrase, the head is the noun that refers to the same entity that the whole phrase refers to.

Mention

A word or phrase in text or speech referring to an entity, relation or event.

Ontology

A model that organizes structured and unstructured information through entities, relations and events, their properties, and the way they relate to one another.

Relation

A relationship that holds between two entities.

Scope

The section of the document from which information about event and relation mentions can be drawn; used to delimit which of multiple mentions an argument should be added to.

Type

A type is a general category of event, entity or relation. In RUFEEERS, types are structured as “nodes in a tree” such that a given type may have subtypes that are more specific categories of the event/entity/relation type, each subtype may have subsubtypes that are even more specific categories of their parent subtype, and so on and so forth. In RUFEEERS annotation, these are referred to as Levels 1, 2, 3, etc.