

EMW Information Extraction Annotation Manual

Introduction.....	1
Terminology.....	2
General Notes for the annotators.....	2
Linguistic Information Annotation.....	5
Event Characteristics.....	5
Type of the Event.....	5
Event Mention.....	8
Title of the Article.....	10
Date and Time of an Event.....	10
Place of an Event.....	11
The Facility where Event Takes Place.....	12
Centrality (Rural, Urban) of the Event Place/Facility.....	13
Participant Characteristics.....	13
Types of Participants.....	13
Name of the Participants.....	15
Participant Count.....	15
Ideology of the Participants.....	15
Ethnic identity of participants.....	16
Religious Identity of Participants.....	16
Caste of the Participants.....	16
The Nature of Participants' Employment Status.....	16
Socioeconomic Status of the Participants.....	17
Organizer Characteristics.....	17
Distinguishing between organizers and participants.....	18
The event organizer type.....	19
The Name of the Event Organizer.....	19
The Ideology of the Event Organizer.....	20
The Ethnicity of the Event Organizer.....	20
The Religion of the Event Organizer.....	20
Caste of the Event Organizer.....	21
Socioeconomic Status of the Organizer.....	21
Target Individual/Organization Characteristics.....	21
Target Type.....	22
Target Name.....	22
References.....	22

Introduction

The WP2 of the EMW Project aims at creating a database of protest events (See the terminology section below for definition of Event) through extraction of information from news articles. This manual specifies the key event information that should be annotated and automatically extracted from text of these articles. The event information will be first annotated manually on a sample of news articles based on this manual. Then the annotated text will be used to develop and evaluate automatic classification and information extraction methods.

This manual describes the annotation process of news articles for EMW. We benefit from and built upon available annotation standards, such as ACE, CAMEO, and TimeML to identify a set of information types that can be targeted and determine annotation details of these set of information types.

The ultimate aim of this manual is to describe how to mark explicit linguistic information that signal the relevant information types related to the project. The whole range of relevant expressions will be captured and be available with their context to be automatically learned for further refinement for social sciences studies.

Each relevant information type and its related tag(s) is described briefly, and sample expressions are provided for the languages and countries in the scope of EMW project. For instance, en_IN presents the examples in English language for India. Description of each information type ends, if available, with example annotations and notes for the annotators.

The tag names are presented in lower case and multi-word tags are separated with underscore. These tags are presented with title case and spaces instead of underscores in the annotation environment.

Extensive details about the information that we will annotate is listed in the following documents that will be provided with this manual:

Codebook_tr.xlsx: A detailed list of event types, participants, organizers, ideology, ethnicity, location, facility, and targets. This list is specific for Turkish. You should use a country specific list for these information types.

Event labeling manual: This manual provides the event definition we facilitate at document level. You should be aware of this definition in order to be able recognize the event related expressions in the documents.

List of the tags: *document_title, event_time, urban_location_identifier, rural_location_identifier, event_place, facility_type, event_type, ...* . The annotation tool will contain this list.

The annotation process of the linguistic information is performed in the “Annotation editor” of the annotation environment (FLAT) and consists of four parts; event, participant, target, and organizer characteristics. These are named as focus and found under “Annotation Focus” in FLAT. The document opens by default in the “Annotation editor” and “Event” focus.

Terminology

- Event: Events that fall in the scope of contentious politics, these are mainly protests and conflicts among groups. Refer to *event labelling manual* for a detailed description of what we consider as event.
- Event-type: Word or words that express an in-scope event in the text.
- Event-mention: Words which refer to and/or stand for the events that are already tagged with event type
- Event sentence: In the news article, “event sentences” are all the sentences which contain the event_type and event_mention information (Please see “Type of Event” section for further information about event_type and event mention). They are generally located in the first paragraph of the news article and usually contain vital information such as event time, place and main participants etc.
- FLAT-sentence: The annotation environment may represent text fragments as partial sentences, single sentences, or several sentences in the same line, which is given a number on the environment. These lines are referred as *FLAT-sentence*.

General Notes for the annotators

1. Read the manual carefully and report any issue you may observe or any suggestion you may have to the advisors.
2. First read the whole of the news article you are going to annotate before starting the annotation.
3. We have a particular event definition (See the “*Event Labeling Manual*” for further details on this definition) and only the news articles that conform to this definition should be annotated. If you think the news article does not contain an event as defined by the “Event Labelling Manual”, do not do any annotation on that news article. Simply mark the “Event” field in the Metadata Editor of the annotation environment, FoLiA Linguistic Annotation tool (FLAT), as “No” and notify the advisors.
4. In case a tag is not used for an article, we will assume that this information was not provided in the article. Therefore, we rely on you to not miss any relevant information. Guessing may cause issues that will cost time and quality. In case you are not sure, you may prefer to give a lower confidence score to your annotation and add a comment which explains your doubts rather than leaving the information untagged. You will find the confidence level bar once you check the “confidence” box in the annotation window which pops up when you select a phrase to annotate it. Please do not skip explaining the reasons of the low confidence by adding a comment in order to avoid future confusion.
5. If the tagged expression occurs multiple times in the article, keep in mind that our priority is the information directly related with the event(s) in the article. We aim at annotating the most relevant entities to the event. This information is usually provided in the document title and/or in the event sentence(s) (see see the terminology section for

the definition of *event sentence*”). Therefore, please follow the steps below to avoid unnecessary tagging or missing any relevant information:

- 5.1. Identify the title of the news article and annotate all event-related information in the title.
- 5.2. Identify the event sentence(s) and annotate all event-related information in those sentences. In cases where the tagged word is used more than once in the event sentence(s) then all those words should be tagged. E.g. “the problem rose in the village after a tdp supporter won the third ward of the village in the panchayat election held on august 14.” Both occurrences of “village” in the sentence should be tagged as *rural_location_identifier* (See the relevant section below for further details on how to use *rural_location_identifier* tags).
- 5.3. There will be no annotation outside the event sentences, that is sentences which do not contain an event_type or event_mention tag will not be annotated.

Observe the following sentences as an example of this principle:

“Samajwadi Party workers also held a **demonstration** at **jantar mantar** **on thursday morning** demanding resignation of up chief minister rajnath singh and union home minister l k advani for their negligence which allowed the assassination.

“SP general secretary usha yadav said the party leaders also visited phoolan devi's residence at ashoka road to express condolences to the bereaved family.”

In this example, only the first sentence is event related. The entities which are related to the protest (in the first sentence) such as SP, are not tagged in the second sentence because they occur in an irrelevant context.

6. The punctuation marks, including quotation marks and parentheses, at the beginning and end of the expressions that are tagged should not be annotated. This principle does not apply to punctuation marks in the middle of the expressions tagged.
7. Sometimes certain expressions or parts of them give pieces of information which correspond to more than one tag. Overlapping annotation refers to tagging the same expression (or a part of the expression) with more than one tag. In the rules enumerated below, you will find the cases where overlapping annotation is required as well as cases where it is not allowed. Note that the rules are exhaustive, that is, overlapping annotation will be allowed for only the cases mentioned below.
 - 7.1. Tags on all kinds of event information in the document will overlap with the document_title tag.
 - 7.2. Participant type expressions which also relay information about other participant characteristics such as ideology, religion, ethnicity, caste and SES will be tagged with all applicable tags.

e.g. “Teachers are determined to continue their strike unless their demands are met.”
[Teachers]=participant_type, participant_SES

e.g. “The city witnessed a massive rally of thousands of angry Maoists today”
[Maoists]=participant_type, participant_ideology

- 7.3. Organizer type or name expressions which contain information which are annotated with other tags in the organizer focus, i.e. organizer, ideology, religion, SES, caste and ethnicity.
e.g. “Over 1.000 people were present in the Communist Party of India (Marxist) rally today” [Communist Party of India (Marxist)]=organizer_name, [Communist]=organizer_ideology, [Marxist]=Organizer_ideology
- 7.4. Named entity expressions will only be annotated with “name” tags. There will be no “type” tags overlapping with facility, organizer and participant name expressions.
e.g. [Safdarjung Hospital]=facility_name (i.e. “Hospital” will not be tagged with facility_type)
e.g. [Karnataka State Government Employees Association]=organizer_name (i.e. no other tag will be used in this expression.)
- 7.5. Expressions which give different information which belong to more than one separate event, which is mentioned in the same sentence, will have all tags necessary. For instance, if the participants of event 1 are targets of event 2, overlapping annotation will be allowed for participant and target tags with their respective event number comments.
8. Named entities or expressions referring to a specific group of people may be expressed with different expressions after their first occurrence. The different occurrences should be annotated with the tag(s) used in their first occurrence if the alternative expression also denotes the same class of entity (event type, participant type, sector etc.). E.g. If a tagged expression such as “teachers’ strike”, where the word “teachers” is tagged with *participant_type* and *participant_SES*, is followed by a phrase such as “agitators want a higher salary” where “agitators” refers to teachers, the word “agitators” is to be tagged as *participant_type* only as the word “agitator” denotes a type of participant but denotes neither the sector nor the SES.
9. Groups of people, regardless of whether they are organizers, participants, or targets may be expressed with pejorative terms, e.g., “miscreants”. This kind of terms should be annotated with the tag corresponding to their role in the event, that is, organizers, participants or targets.
10. Prepositions before *event_places* names will not be tagged. However, prepositions before *facility types* and *names* are tagged (see below the sections Place of an Event and The Facility Where the Event Takes Place).
11. We annotate pieces of information even when the documents use qualifiers such as supposed(ly) or alleged(ly) before them. E.g. “the police have reportedly rounded up 11 members of the bku for allegedly attacking one kishan singh”
12. While annotating a multiple word entity, make sure the tag span does not exceed the end of a *FLAT-sentence* as represented in FLAT. You can recognize these sentences as numbered on the left side of the document like rows in a table. If the phrase overflows to the next sentence, annotate the part of the phrase in the second sentence with a new tag.
13. General tag span rule: As a rule of thumb, we must aim to include in the tag span as few words as possible. The more words are in a tag span, the less the likelihood of inter-

annotator agreement. For many tags, coding single words is ideal. Adjective, adverbial, noun or prepositional qualifiers are generally left out of the tag spans (exceptions are in tag descriptions below). Yet there are many instances where a single word will not indicate an entity in a meaningful way and inclusion of more words becomes necessary. The most general exception to the one word tag span principle is compounds. Compounds are single units of meaning that are formed by combining more than one word such as “someone”, “likewise”, “camcorder” etc. Some compounds, such as “floppy disk” and “ice cream”, however, are written separately and look and act like noun or adjective (or other) phrases, . These might be difficult to discern from phrases which entail descriptions of qualified words and are thus not single units of meaning as a whole. Consider “floppy disk” which designates a concept and is not a description (i.e. it does not mean “a disk that is floppy”). Prominent examples of compounds that we encounter often in protest annotation are phrasal verbs such as “*flare up*” and “*walk out*” or expressions such as “*hunger strike*”, “*take to the streets*”, “*toddy tapping*” and the like. Since compounds are single units of meaning, i.e. designate single concepts, when they are tagged, the tag spans must cover all the words that make the compound (i.e. as if they were written in a conjoined way). On the contrary, modifier words in phrases must be left out of tag span as they do not change the nature of the concept that is tagged but merely describe it (e.g. school *teachers*, landmine *blast* etc.). To distinguish compounds from ordinary phrases you can keep in mind that the multiple words which form them usually lose their original meanings and the compound means something different from each of its word-parts. You can also remember that they usually have their own definitions in dictionaries unlike noun or adjective phrases. Below, under tag definitions, you will find tag span rules spelled out for those individual tags that require special treatment. Indefinite articles (a, an) are not to be included in the tag when they occur before or after the expressions to be tagged. Definite articles (“the”) will **only** be tagged if it is part of the official name of an entity, e.g. “The Supreme Court”. Lowercase “the” will not be tagged.

Linguistic Information Annotation

Linguistic annotation is the process of marking part of a text with tags relevant to the task we want to achieve. Our objective in this task is to identify and extract the event-related information (such as type of an event; participants, organizers, targets of the events and, if available, their religious, ethnic characteristics and ideological stances).

The **yellow** mark and bold face are used to indicate the part that should be annotated.

Event Characteristics

These information types are about characteristics of an event.

Type of the Event

The event has a central role in our annotation efforts. You should first detect the event in a news article before you do any annotation. Events that can be included in the broad term

contentious politics are in our scope (for the sake of simplicity, we use protest events to refer to contentious politics). Contentious politics refers to any form of grassroots political action, and actions of political and non-governmental organizations that are aimed at mobilizing the public in the name of political demands and grievances. Such actions are outside the institutional and legal forms of political participation such as voting, litigation etc. For details as to our event definition and various event types, refer to the *event labelling manual*.

A number of examples of protest events can be given as follows:

Demonstrations, rallies, marches, press declarations that take place outdoors in public, gatherings, meetings (when referring to public rally-like meetings), sit-ins, unofficial commemorations with mass participation (e.g. funerals or burial site visits), acts of civil disobedience and nonviolent resistance, collective petitions, collecting signatures, strikes and any other work related protest events (slow-downs pickets etc.), hunger strikes, death fasts, lynchings, lynch attempts, kidnappings, occupations, boycotts, cyber-attacks, hacktivism, hijacking, self-immolation, armed and unarmed violent attacks (by non-state actors), armed militancy actions against state security forces, unarmed clashes with security forces, clashes among groups of civilians caused by political issues, and the like.

Event types that are not among these examples should be evaluated based on their relation to definitions provided in the event labeling manual. Any doubt should be communicated to the experts.

The protest events that are detected in the news articles must be current or past events in order to be annotated. Events that are announced, threatened, reported to take place in the future in any other way will not be annotated. The only exception to this is when a threat of violence is issued by an actor, or a violent action attempt is thwarted, e.g. a death threat or a planted bomb that is diffused before going off. Self-harm threats and attempts are to be considered as violence attempt/threat, e.g. self-immolation attempt or threat. Such threats or attempts are thought to have the effect of events carried to completion and are thus regarded as events themselves. General or conceptual mentions to events, instead of specific happenings, will not be annotated either.

The type of the event should be annotated as follows:

event_type:

en_EN: **strike**, clash

en_IN: **Akhand Path**, **dharna**, **bandh**, **Chalo assembly**

tr_TR: göstericiler polisle **çatıştı**,

Some other examples from the ACE event annotation manual are:

1. Thousands of people **rioted** in Port-au-Prince, Haiti over the weekend.
2. The union began its **strike** on Monday.
3. Protesters **rallied** on the White House lawn.
4. The **rioting** crowd broke windows and overturned cars.

5. A crowd of 1 million **demonstrated** Saturday in the capital, San'a, protesting against Israel, the United States and Arab leaders regarded as too soft on Israel.
6. In Ramallah, around 500 people **took to the town's streets** chanting slogans denouncing the summit and calling on Palestinian Authority leader Yasser Arafat not to take part in it.
7. For weeks Italian Jewish groups, World War II veterans and leftist political parties have staged **protests** against a meeting between the pope and Haider, arguing that a papal encounter would lend the Austrian politician legitimacy.
8. More than 40,000 workers were back at their jobs Thursday following a 1-day **walkout** that closed social welfare offices and crippled public medical services. During the work **stoppage** Wednesday, local residents were unable to register marriages or get documents for real estate transactions.
9. Shah's supporters also had gone on the **rampage** outside the principal's office on Tuesday.

Notes:

1. Tag Span Rules:

- i. Adjective, adverbial, prepositional, noun or any other qualifier word will not be included in the span. *e.g.* landmine **blast**, token **strike**, forcibly **stopped**, communal **flare-up**, indefinite **hunger strike**, fierce **encounters**, twin militant **attacks**
 - ii. In verb event types, only the verb will be in the span. *e.g.* **boycotted** a meeting, **rioting** with weapons, forcibly **stopped** the auto, **fight** it with bombs, landmine **exploded**, **hit** by crude bombs, **refused to work** extra hours.
 - iii. Auxiliary verbs will not be included, except when the verbs are in negative form *e.g.* have **blocked off**, will **protest**, are **demonstrating**, **did not start work**,
 - iv. In noun event types, only the noun will be in the span *e.g.* laid an **ambush**, forced a **bandh**.
 - v. If the verb or noun itself is not meaningful or is dubious as a protest action, include the minimum required words. Since tag spans cannot be broken, this might necessitate inclusion of non event type words. *e.g.* **Violent reaction**, **violent culmination**, **took to the town's streets**, **did not start work**, **refused to work**, **shouted slogans**, **written slogans**.
2. If the article contains multiple events, mark all. If there are multiple events in a news story, the events that have different times, places, facilities and participants from other events should be annotated separately. In such cases, after the first event, the second and following events should have a comment that starts with "Event 2", "Event 3", etc. The first event will be considered as main event and will not have an event number comment. The numbering of different events will be based on the order in which they are presented in the article. All related tags to following events should contain the same comment. For example, place of the second event should contain the comment of "Event 2." However, if the events in a news story share the same time, place, facility and participants (or a subset of the same participants), they

should be counted as one event and share the same event number in the comment section. Note that if a tag doesn't have an event number comment, it will be assumed that it belongs to the first event. Also remember the rule which says that there will be no annotation outside event sentences here. If a sentence contains information about an event which **doesn't** have an event type or event mention in that sentence, that sentence will not be tagged for such event even if it is the event sentence of another event.

3. Demonstrations, gatherings, electoral rallies, etc. organized by political parties, organizations, or representatives of political parties should be included. However, events happening under the auspices of the parliament such as politicians leaving parliamentary sessions, disrupting them or having heated arguments with each other are accepted as usual proceedings within parliaments and thereby not considered as events. **BUT** if politicians engage in a protest action which is unusual and outside the formal parliamentary procedures and proceedings, (e.g. such as hunger strikes, sit-ins) these will be considered as protest events.
4. Spontaneous gatherings or demonstrations with no organizer or leader should be annotated as events.
5. Be aware that Hindu rituals such as idol immersions and akhand paths are not protest events in themselves. However they can become means or scenes of protest events. In an expression like "**Clashes** broke out after idol immersion", only "clashes" should be annotated as event type but in a sentence like "An **akhand path** was organized to protest government's decision", this ritual is deployed as a protest itself.
6. Public interest litigations and writ petitions are to be tagged as event types in case they have political motivations, and are participated by multiple members of the public.

Event Mention

Any word that refer to the event type which occurs after the first occurrence of the event_type will be tagged with the event_mention tag. Pronouns and token words which refer to the event type are included in the tag's scope. Token event words are generic expressions which denote the event (i.e. words like incident, event, protest and agitation).

Notes

1. For every separate event, there will only be one event_type tag which will be the first occurrence of the event word, provided it is not a token event word (see below). All the following event words and pronouns that stand for the event type will be tagged event_mention.
2. Event words in the document title will be tagged with event_type tag. The first occurrence of the event word outside the document title will also be tagged with event_type tag.
3. Token event words which stand for events, will be tagged with event_mention tag even if they are the first occurring references to events (see the examples below). In principle, you can distinguish a token event word from a true event type word by the

fact that the former, on its own, means a broad class of protest actions which can correspond to multiple categories of events defined in the Event Labeling Annotation Manual, namely, demonstrations, industrial actions, group clashes and political violence and militancy. The words like incident, event, protest and agitation can stand for events from all (or more than one) of these categories. Such words will always be tagged event_mention. The only exception to this is the case when token words are the only event words that are used in the article to refer to the events. In such a case, the general rule will hold and the first occurrence of the token word will be tagged event_type and the following references will be tagged event_mention.

4. Event separation logic will be the same as any other tag. That is, any event mention tag that is used for each separate event will be numbered in a comment.
5. The sentences which contain the event_mention will become event sentences, thus any relevant event information in these sentences will be tagged even if the event_type tag occurs in other sentences.

Examples

CPI(M) stages protest **rally** in Bhavnagar. the bhavnagar unit of communist party of india cpi(m) on friday staged a **demonstration** opposite the local post office here...

–The word “rally”, the first occurrence of the event word, is tagged event_type, demonstration which refers to this event, is to be tagged with event_mention instead of event_type.

Commenting on the **strike** which was flagged off on Monday, the union secretary stated “**it** will continue as long as our demands are not met.”

–The word “it” stands for the event_type “strike” and thus is tagged with event_mention.

In a separate **incident** in the neighbouring area of Awantipore, an airforce vehicle was damaged and two security force personnel were injured when a landmine **exploded**.

–Here, the main event word is “exploded” and incident

The students organized a **protest** by **marching** against the payment seat decision.

–Here the token event word “protest” is tagged event_mention as any contentious politics event can be referred to as protest while the word “march” is tagged event_type as it directly indicates what the protest is, and thus is a true event type.

A variety of NGO’s expressed their support for the **agitation** of teachers in the capital.

–Considering this sentence is the only one which mentions the event in the article, the token word agitation is tagged event_type.

Title of the Article

The news article will have a title by default most probably at the top of the page. You should mark it with *document_title* tag. The following cases are taken into account:

1. If the document title is on top of the news article, mark it. This is the case most of the time.
2. Some articles may contain more than one news report independent of each other. Some of these reports may not include any protest events, in which case their titles will not be tagged. Mark only titles of the articles that include at least one event. You should insert a note in the comments section of the tag for the second and following titles in this case. The comments should start with "Event 2" for the second title and go on Event 3... if there are more titled reports.
3. In multiple report cases, the titles that follow the first one might not be as easily discernible as the first title, which is usually on the top. The flat environment does not contain font size or colour variety that distinguishes the titles from the text body. Try to single out the titles from punctuation, inconsistent upper or lower case use etc.
4. We use this tag only to identify the news article titles as they were headlined by the newspapers. In case there is no title at the top of the article, do not search for it in between the lines. Simply notify the advisors about the missing title.
5. The title generally contains key information about the event. You should first annotate the words or phrases that contain event information with relevant tags, and then annotate the entire title with the *document_title* tag.

Date and Time of an Event

The publication date of the news article will be provided as it is extracted from the HTML file of the news article. Therefore, it will be at the top or bottom of the document most of the time. Please mark it with the *event_time_published* tag. And make sure that you mark the given expression in full, including the date, time and time zone - if provided. Don't forget to mark the punctuation marks (commas) within the time expression.

Any other event-related time expression in the article should be marked with the *event_time* tag. When annotating time expressions, make sure that you mark their complete versions, including for instance the related prepositions, such as 'on' in the case of 'on Friday'.

event_time_published:

en_EN:

en_IN: "Sep 8, 2001, 23:34 IST",

tr_TR: "05 Şubat 2018 Pazartesi"

event_time:

en_EN: on Friday, yesterday, in the morning, on January 23, over the past week, in the early hours of Saturday, on saturday evening, late on saturday night

en_IN:

tr_TR: Cuma, akşam saatlerinde

Notes:

1. Expressions like “75th anniversary”, “75th birthday” should not be annotated.
2. Duration expressions are not in the scope of our study. You should not annotate the duration in expressions like “after 102 hours of hunger strike”. Namely, “after 102 hours” should not be annotated as *event_time*. In an expression like “the junior doctors strike entered the fourteenth day on Sunday,” only “on Sunday” should be annotated as *event_time*.

Place of an Event

Event places are officially defined administrative divisions pertaining to each country. That is to say a **place name** will be tagged if it corresponds to one of those official divisions such as state, province, city, district, neighborhood, village, taluk, and/or panchayat, etc.

The tag *event_place* will be used for labelling the place names in the event sentence while the tag *event_place_published* will be used for labelling the publication place in the meta information, provided next to the title and publishing time of the news article.

Streets, avenues, boulevards, market places, names of buildings are not administrative units. Therefore, they should be tagged not as *event_places* but as *facility_types* or *facility_names*.

Please note that in phrases such as Vadodara city or Mattihalli village, only the names (“Vadodara” and “Mattihalli”) will be tagged as *event_place*. The nouns city, village etc. will be tagged as location identifiers (See the section “Centrality”).

event_place:

en_EN: London, Soho neighbourhood, Poonch district

en_IN: Gujarat, Ahmedabad, Kutch, Chak ram singh wala village, mattihalli village in harapanahalli taluk of davanagere district, Sector 25 in the city.

tr_TR: Istanbul, Ankara, Kadıköy ilçesi, Acıbadem Mahallesi, Gecek Köyü, Dağlıca.

event_place_published:

en_IN: “Students clash on Monday” TNN | Sep 18 , 2001 , 11.50 PM IST vadodara :

Notes:

1. Tag span rules:
 - i. Only the Place name will be tagged. Generic settlement categories (village, city etc.) and prepositions will not be included e.g. Chapadu mandal, near Charar-e-Sharief, in Bihar’s impoverished Aurangabad district, Gecek köyü
 - ii. Direction words like east west will be included only when they are parts of the official names e.g. Imphal west district, Salem West district, West Bengal
 - iii. The conjunction “and” will only be included if it is part of the official name of the place, if it is used to connect different places, it will not be a part of the span

and places it connects will have separate tags (this rule is valid for all the tags) e.g. Jammu and Kashmir, Karnataka and Gujarat states

2. If the article mentions events from multiple places, mark all events and the event related places and note the relevant event number in the comments section.
3. Events in countries other than the country we focus on at that moment should not be annotated and reported to the advisors. Moreover, please follow the instructions about relevant country field in the implicit information labelling manual to mark it appropriately. Country names will never be tagged.
4. Prepositions that are preceding a place name should **not** be covered in *event_place* annotation.

The Facility where Event Takes Place

Facilities are human made structures or built environments that serve as the location the events take place. These facilities can be streets, factories, airports, universities, vicinities of prominent buildings or sites (e.g. in front of a party office or an organization) etc. When tagging facilities, the prepositions that are linked to them will be included in the tag. E.g. "the strike had crippled all services in the hospital", "the strikers gathered in front of the hospital". This also holds for facility names (see below).

facility_type:

en_EN: street, factory, airport, university, in front of the party office,

en_IN: District commissioner's office, in front of a Gurudwara,

tr_TR: Meydanda, havaalani, ana yolu.

Sample sentences:

1. Owners of contract buses, autorickshaws and taxis stayed off the roads.
2. shah 's supporters also had gone on the rampage outside the principal 's office on Tuesday .
3. The strike had crippled all services in the hospital and most patients had been discharged.
4. two individuals died in the bomb explosions organized by radicals in places of worship
5. Services in government hospitals were paralysed due to the strike.

In case these facilities are mentioned with their named entity, the tag for this information type should be *facility_name*:

facility_name:

en_EN: In Hyde Park,

en_IN: Safdarjung Hospital,

tr_TR: Taksim Meydani, GATA, Gezi Parkı

Notes

1. Tag span rules (for both facility name and type):

I. Prepositions will be included e.g. on the Osmania University campus, at Cusrow Wadia Institute of Technology, into the university campus.

Centrality (Rural, Urban) of the Event Place/Facility

The linguistic information that relays information about the centrality of the place in which the event takes place will be labelled with these tags. Event places can be urban or rural areas. The expressions which indicate this information about the event place will be tagged with Urban or Rural Location Identifier tags. The expressions such as town, city, village, panchayat which directly indicate this will be tagged. In addition, expression which identify built environments (human made surroundings as settings for human activity) such as plaza, avenue, slum, (village) well, that are invariably associated with urban or rural settings will be tagged.

urban_location_identifier:

en_EN: City, town, urban, avenue, plaza

en_IN: slum, municipality,

tr_TR: şehir merkezi, Çardaközü Köyü

rural_location_identifier:

en_EN: village, rural, irrigation

en_IN: Panchayat,

tr_TR: kırsal, köy, mezra,

Notes:

1. The place name is not an urban or rural identifier. Therefore, place names such as “New Delhi”, “Delhi” etc. should not be tagged with these tags.
2. No other type of expression than the ones provided above will be tagged. Types of buildings or institutions, types of economic activity etc. will **not** be tagged with LI tags.
3. If expressions such as avenue or plaza are facilities of protest events and are tagged with facility tags, they will **not** be tagged with LI tags. In other words, there will not be overlapping LI tags.

Participant Characteristics

Types of Participants

The participants are the main actors of a protest event. They can be individuals or groups that engage in a protest. Groups may be referred to as masses (common people, public), armed or violent political actors (attackers, militants, militias, guerillas), unarmed non-violent political actors (activists), blue-collar workers (factory workers, low skilled workers) white-collar workers/professionals (lawyers, doctors, nurses, corporate employees,

teachers, engineers), the unemployed, students, women, LGBTI Individuals, peasants (including landless peasants and sharecroppers), villagers, war veterans and their relatives, politicians (members of the parliament, political leaders, party members), academicians, journalists, public figures, artists, intellectuals, artisans and shopkeepers, the disabled (physically challenged), prisoners and their relatives, refugees/immigrants, consumers, hackers or hactivists, ultras, sports spectators, supporters, relatives of political violence victims, prisoners, kidnapped, disappeared or assassinated individuals, and the like.

The tag for this information type is participant_type.

participant_type:

en_EN: teaching and non-teaching staff went on strike on Tuesday

en_IN: hundreds of karamcharis staged a dharna in Delhi

en_IN: skirmishes occurred between the bharatiya kisan union (bku)-ekta activists and the police at various places in the district, on Monday.

en_IN: senior congress leaders including kamal nath, oscar fernandes, narendra nath, dpcc chief subhash chopra and party legislators participated in the demonstration.

en_IN: more than 800 agitators participated in the protest

tr_TR: gazeteciler insan hakları anıtı önünde basın açıklaması yaptı

Notes:

1. Tag span rules:

i. Fewest possible words which indicate the actual participants meaningfully. No qualifiers will be included in the tag span e.g. rival party supporters, resident doctors, Muslim women, jobless diamond workers, secondary school teachers, inter-city bus drivers, angry karamcharis.

ii. As an exception to the previous rule, some qualifiers which modify certain occupational groups change the nature of those occupations so significantly that they become a different category of employees with different socioeconomic status and/or organizational characteristics. In these instances the qualifiers should be included in the tag. It is not straightforward to determine these occupations as they require certain unique country characteristics to be taken into account. When a likely case is encountered, don't hesitate to ask the project experts about the coverage of the tag spans. Prominent examples are as follows: auto-rickshaw drivers, college teachers, safai karamcharis. Note that the words, drivers, teachers and karamcharis are the only words to be included in the tag spans in other instances, see the previous note.

iii. Words like mob, group, crowd, will be tagged only if they are the only words standing for participants. When they are parts of phrases with more definite participant words, they will not be included e.g. a group of about 100 people, group of locals, group of militants, group of students, the angry mob, the crowds gathered in front of the supreme court.

Name of the Participants

The name of the participant(s) may be mentioned. This can be a first name and/or last name of an individual who is engaging in the protest. Please use the tag *participant_name* for annotating these names.

participant_name:

en_EN:

en_IN: Senior congress leaders **rosaiah**, **venkataswamy**, **sudhakar reddy** and others participated in the meeting.

En_IN: Senior congress leaders including **kamal nath**, **oscar fernandes**, **narendra nath**, dpcc chief **subhash chopra** and party legislators participated in the demonstration.

tr_TR: Ünlü sinemacı **Ahmet Pınar** eylem sırasında kalabalığa seslendi.

Participant Count

Any expression that indicates the total quantity of individuals who participated in the event should be annotated with the tag *participant_count*.

participant_count:

en_EN: a large number of people, many people

en_IN: hundreds of BJP workers, four agitators

en_IN: as many as 50 employees gathered at the dc's office

en_IN: more than 800 agitators participated in the protest

en_IN: about 700 members of the bku are sitting on a dharna,

tr_TR: binlerce gösterici, iki çevre aktivisti, 1000'den fazla işçi

Notes:

1. In Expressions, such as “more than a hundred demonstrators”, “nearly fifty assailants”, “as many as five thousand workers”, and “about twenty militants”, that give an estimation of the number of participants, only the number expressions will be tagged; that is the estimation qualifiers will not be included in the tag span.
2. Sometimes quantity expressions refer to a *part* of all participants (e.g. the police arrested 11 of the protesters). These expressions should not be tagged with *participant_count* as they do not give information on the total quantity participants

Ideology of the Participants

In case an explicit mention of participant ideology is mentioned, this is tagged as *participant_ideology*. The participant ideology can be communist, socialist, anarchist, social democrat, left-wing, right-wing, feminists, liberal, nationalist, religious fundamentalist, conservative, far right, green or environmentalist, animal rights, pro-LGBTI, and any expression similar which identifies the ideology or world view of participants. Please tag any word/phrase that indicates the ideology of the participants.

participant_ideology:

en_EN: Leftwing militants blocked the road in Caracas.
en_IN: Maoists staged a rally in the city centre.
tr_TR: Çevreci aktivistler, kaza ile ilgili basın açıklaması düzenledi.

Ethnic identity of participants

Expressions which identify the ethnic identity of participants are marked with *participant_ethnicity*.

participant_ethnicity:

en_EN: Catalan protesters took to the streets of Barcelona.
en_IN: Two of the activists were identified as Tamil militants.
tr_TR: Kürt gençler Diyarbakır merkezde oturma eylemine destek verdi.

Religious Identity of Participants

Any expression that identifies the religious affiliation of participants are tagged with *participant_religion*.

participant_religion:

en_EN: Some Catholics are protesting what they regard as a blasphemous stage production that disrespects God and the Catholic faith.
en_IN: Nationalistically aroused Sikhs are leading demonstrations to ban the sale of cigarettes since the use of tobacco is forbidden to Sikhs.
tr_TR: Ülkenin güneyinde Müslümanlar, Yahudiler ve Hristiyanlar birlikte eylem yaptı

Example: Muslim workers protested against the ban of Friday praying in factories in Mumbai.

Caste of the Participants

Expressions which designate the caste of participants will be annotated with *participant_caste* tag. This tag is specific for annotation of news from India.

participant_caste:

en_IN: The safai karamcharis went on strike on monday, after five of them were handed suspension orders
en_IN: They also explained the aggression of the nayaks on harijans on every trivial issue

The Nature of Participants' Employment Status

The tag *participant_employment* is used for the expressions that denote the employment status of the protest participants. Any words or phrases that indicate whether the participant is employed in the formal or informal sector, works for a subcontractor, domestic and unpaid work, or is unemployed should be annotated with this tag. The formality of participant employment may not be clear most of the time. If there is an expression which

indicates that workers do not have social security or any form of job related insurance or that they work in the black market (shadow economy), these will be tagged with `employment_status`.

participant_employment:

en_EN: Hundreds of thousands of public sector workers have gone on strike to protect their `pensions`.

en_IN: The protesting `informal` sector workers in India have been demanding basic *workers'* rights for several years.

tr_TR: İşten çıkarılan `taşeron` işçiler eylem yaptı

Notes:

- Subcontracted employees (also called subcontractors) are those who do not work for a company but perform duties through a liaison contract and/or business to fulfill part or all of the obligations of a given company. The expressions that are signaling subcontractor status of the protest participants should be annotated with *"participant_employment" tag*.

Example: `Subcontracted` workers joined the strike led by the SEWA union in Bengal.

Socioeconomic Status of the Participants

The expressions which give information as to the social standing or class of participants will be tagged with *participant_SES*. Any word or phrase that indicate the class position, socio-economic status, or income or education level of participants should be annotated.

participant_SES:

en_EN:

en_IN:

tr_TR:

Notes:

1. The expressions which designate, if indirectly, the socio-economic status of participants, they will be tagged. Examples can be "starvation", "malnutrition", "slums", gated community etc.
2. If the participant type is a category implying participant SES (like it is the case for teachers, doctors, workers) then mark them with both participant type and participant SES tags. The words "politicians", "activists", "students", "housewives" do not imply any SES information and hence will not be tagged *participant_SES*.

Organizer Characteristics

Organizer is another actor category which engages in contentious politics. Organizers organize and/or lead the event and participants. Some events do not have organizers because they are spontaneous events, in which participants take action without being led by

anyone. In other cases, an organization, institution, a group of people or single individuals might organize or lead the protest. Any organization which takes part in the protest and individuals who lead or organize the event will be tagged with organizer tags. Individuals can be organizer only in cases in which it is explicitly stated that they *lead* or *organize* the event.

Distinguishing between organizers and participants

The difference between organizer and participant of an event is very important. First, the same participant group can join two events organized by two different organizations and the result may be totally different. For example, workers can join a left-wing rally organized by unions or a right-wing lynching against migrants, organized by a racist organization. Second, some events are organized by some organizations or it can happen spontaneously without any organizing effort. Theoretically, events have one of the following organizer-based characteristics:

- There is not any explicit organizer, spontaneous events
- There are organizers and participants, coordinated events. Unions leading a strike, for example.
- There are organizers, but there are not any participants: This is not very common. We see some parties/organizations coming together to organize an event, in which no one participated.
- There are no organizers and participants mentioned. This could be an explosion, for example, where no one was identified as responsible.

Some examples that clarifies this distinction are as follows:

- 1- In case there is a boycott at a university campus, the organizer is the student union and the participants are the students.
- 2- The potential organizer of a strike is a labor union and the participants are the employees.
- 3- Some events are reported as led by a particular person, in such a case the person is the organizer, e.g. "the protesters led by **shelia dikshit**", where "sheila dikshit" is the *organizer_name*. If the person is not clearly specified as leading the event, it should be annotated with the appropriate participant tags. For instance, an assassin is still a participant of an assassination event.
- 4- A phrase like "association members" should be tagged as "association": *organizer_type* and "members": *participant_type*.
- 5- Some cases mention names of an organization and its leader, such as "bjp leader madan lal khurana , architect of many anti-government protests , lent his voice , and a broad smile .". These cases, in case they are in the scope of an event, should be annotated as "bjp": *organizer_name*, "madan lal khurana": *participant_name*.
- 6- Some organizations or people express their support to an event. These entities should not be tagged with any tag in case there is not any indication of organization or participation effort. So, sole support is not enough to be annotated.

The event organizer type

The organizer(s) of an event can be Political party, Non-governmental organization, associations, charities, Labor union, Non-party political organizations (student organization, womens' organizations etc...), Anti-state armed groups such as terrorist organizations, resistance movements, guerilla groups., Pro-state armed groups such as paramilitary organizations, Social communities (tribes, religious groups), Local public authority (municipalities, governorates etc...) or Multiple organizations (more than one party, NGO and other type of organization coordinated the event). Please tag any type of the organizer with *organizer_type*.

organizer_type:

en_EN: Left **parties** organized a silent march,

en_IN: A prominent banned **religious outfit**, Deendar Anjuman, exploded a mine on Sunday

tr_TR: Feminist **dernekler** cumartesi günü bir basın açıklaması gerçekleştirdi.

Examples:

- 1- "when the protesters, led by senior youth congress leader ish kohli, held a demonstration just before the beginning of the council meeting, firdous came down and received a memorandum demanding not to handover the property to the private organisation". youth congress: *organizer_name*, leader: *organizer_type*, protesters: *participant_type*, ish kohli: *organizer_name*.
- 2- "the three-day recitation of the guru granth sahib was organised by the delhi transporters ' unions , hoping for an extension in the cng conversion deadline which expires this month." delhi transporters unions: *organizer_name*.

Notes

1. Tag span rules:
 - i. Like any other tag span, the shortest possible type expression will be tagged, i.e. words like association, party, union, employees association. If the word association stands for a workers union, which is generally the case in India, the word before the association will be included, i.e. employees association.

The Name of the Event Organizer

In case the name of the organizer of the event is mentioned, please tag it/them with *organizer_name* tag.

organizer_name:

en_EN: "**Frankie Boyle** has started a hunger strike in solidarity with Guantánamo Bay prisoner Shaker Aamer", "**The Labour Party** organized a rally on Sunday",

en_IN: "A prominent banned religious outfit, **Deendar Anjuman**, exploded a mine on Sunday"

en_IN: **The University Teachers' Association** joined the campaign and signed the anti-austerity statement.

tr_TR: **DİSK**'in çağrısı üzerine bir araya gelen sendikalar yasa tasarısını protesto etti.

Notes

1- Tag span rules:

- i. Full organizer name will be included. If the abbreviation of the name is present within brackets, it will be tagged separately without the brackets **e.g.** Datta Gule's auto union, All India Bank Officers' Association, (PUCL), Indian National Trade Union Congress (INTUC), Joint Front of Trade Unions of the GIC and the LIC

The Ideology of the Event Organizer

The organizer of a protest may have an explicit ideology, which can be at least one of Communist/Socialist, Anarchist, Social democrat, Mixed left (including feminists, environmentalists etc...), Other left-wing ideologies, Conservative, Liberal, Nationalist, Religious/Conservative, Far right, Other right-wing ideologies, Mixed right, Feminist, Green movement/Environmentalism, Animal rights, pro-LGBTI, Mixed right/left together, feminists, socialists, or anarchists and environmentalists etc. Mixed right includes right-wing groups that organize together: nationalists and conservatives. Please examine the country specific ideologies list and annotate the occurrence of an ideology with *organizer_ideology*.

organizer_ideology:

en_EN: Civil rights organisations and Left parties on Friday staged a protest rally

en_IN:

tr_TR:

Example:

The Ethnicity of the Event Organizer

If an organizer is described using its ethnic identity, mark the identity with *organizer_ethnicity*.

organizer_ethnicity:

en_EN:

en_IN:

tr_TR:

Example:

The Religion of the Event Organizer

Organizer's religious identity is annotated with *participant_religion*.

organizer_religion:

en_EN:

en_IN:

tr_TR:

Example:

Caste of the Event Organizer

In case caste of the organizer is provided, it should be annotated with *organizer_caste* tag. This tag is specific for annotation of news from India.

organizer_caste:

en_IN:

Example:

Socioeconomic Status of the Organizer

If the news article contains information about whether the organizers of the event are poor, rich, or middle-class people annotate this with the tag *organizer_SES*. Please annotate all mentions.

organizer_SES:

en_EN:

en_IN:

tr_TR:

Notes:

- If there are SES indicative phrases related to organizers, mark them with this tag. These can be “starvation”, “malnutrition”.

Example:

Target Individual/Organization Characteristics

The organization, institution, group(s) or individual(s) that is the antagonist of the protest event is annotated with *target* tags. Targets are the entities that are opposed by those who organize or participate in the protest. Certain protest actions can aim at the targets directly for instance when they are attacked by the protesters. In other instances, the entities towards which the actions are directed are not the targets but nevertheless become the objects of the protests. In the latter case, care should be taken not to tag the objects of the protest as targets. Consider the examples where “small businesses closing their *shops* to protest some **government** regulation,” “workers blocking *roads* to protest their **employer**.” “a majority caste members attack the *settlements* where members of a **minority caste** live. The entities *italicised* are the objects of protest actions but the true targets are those expressions written in **bold**. In brief, target implies the entity that is opposed by the protesters in the abstract. Most of the time, the objects of the protest events are entities annotated with *facility* tags. Facilities, as defined above, are human made structures or spaces which the protests take place in or even target at times and should not be confused with targets. In the examples already mentioned, *shops*, *roads* and *settlements* are facilities even though the actions are directed at them; whereas the **government**, the **employer** and the **minority caste** are targets, i.e. the antagonists of the protest.

Target Type

The type of entities that are the targets of the protest events, such as organizations, individuals, groups countries etc. should be identified and annotated with *target_type*. The examples of the organizations are Government (governing party and its politicians, ministries), state, the army, police, employer & business, mayor or governor (including local governments, municipalities and the governorate), Other political parties and politicians (except for the governing party), judiciary, civilians and non-political social groups (students, women, workers etc...), minority groups (ethnic, religious, racial, caste etc.), non-party political organizations (including non-state armed groups, activist organizations), political activists, politically affiliated individuals, intellectuals and alike, public institutions, bureaucracy and bureaucrats (regulatory agencies, public schools and hospitals), the parliament, foreign countries, NGOs, transnational organizations, etc..., unions, armed paramilitary organizations, criminal organizations (drug trafficking, mafia etc...). In case the event did not occur against a specific individual or institution or it is not clear what the target is, there will not be any annotation.

The target type can be:

target_type:

en_EN: Activists accused **the minister** of violating the international law.

en_IN: More than 2000 persons staged a protest near Vijay Theatre ground against the **district police, contract bus association chief** harish sabharwal was booed by the protestors

en_IN: trouble is brewing at the ahmedabad district co-operative **bank** (adcb) over the bank's investments for the revival the madhavpura mercantile co-operative bank ltd (mmcb).

en_IN: **Priests** among 7 shot dead in Kashmir

tr_TR: Göstericiler TBMM önünde **hükümeti** protesto ettiler.

Target Name

The name of the target should be annotated with the *target_name* tag.

target_name:

en_EN: NATO, World Bank, Prime Minister Theresa May

en_IN: contract bus association chief **harish sabharwal** was booed by the protestors

en_IN: after 102 hours of hunger strike, there has been no effort from the **adcb** chairman to resolve the issue.

tr_TR: Savaş karşıtı kuruluşlar **NATO**'yu protesto etti.

References

[1] ACE English Events Guidelines version 5.4.3

[2] Yıldırım, Kerem. (2017): Emerging Welfare: Content Analysis Project, Instruction Manual. Koç University.