Annotation Guidelines for Named Entities and Relation in Malaysian English News Article

Research Title: **Relation Extraction from Malaysian English News Article**

Prepared by Mohan Raj Chanthran (32301065)

21/10/2022

Version 3

Introduction **2**
Malaysian English **3**
Named Entity **3**
Entity Label 4
1) PERSON (PER) 4
2) LOCATION (LOC) 6
3) ORGANIZATION (ORG) 8
4) NORP 10
5) FACILITY (FAC) 10
6) PRODUCT 11
7) EVENT 12
8) WORK OF ART 14
9) LAW 14
10) LANGUAGE 15
11) ROLE 15
12) TITLE 16
Descriptor 18
Summary Table 19
Relations **23**
Relation from DocRED 24
Relation from ACE05 24
Conclusion 24

Introduction

This project aims to extract entities and relations from Malaysian English news articles. In order to achieve this objective, we are developing a dataset called *Malaysian English News Article Dataset*. The articles are extracted from a few prominent Malaysian English news article publishers. In this document, we have compiled some guidelines to educate and aid annotators when they annotate Malaysian English News Articles with entities and relations between them.

Malaysian English

1 Malaysian English (ME) is one of the English dialects that evolved from Standard English with some mixture of localized words and grammatical constructions (Ismail et al., 2007). ME is a widely spoken and utilized language in Malaysia, where it combines elements of standard English with Malay, Chinese, and Indian languages in terms of vocabulary, pronunciation, and grammar.

2 Words in Malaysian English are adaptions from Malay, Chinese, and Tamil. We could see Morphosyntactic Adaptions and Semantical Adaptions from foreign languages to English. Some of the morphosyntactic adaptions are

Morphosyntactic Adaptation	**Examples**
Loan Words: Words adapted from Malay, Chinese, and Tamil.	*kenduri, Hari Raya, imam, Ustaz, ustazah, alhamdullilah, Insyaallah, Ramadan, bumiputera, orang asli, Datin, Datuk, Makcik, Pakcik, Tabung Haji, Dewan Negara, Datuk Bandar, Menteri Besar, Yang di-Pertua* and *pasar malam.*
Compound Blends: Words combined from two different words. It can be a combination of Malay and English, Chinese and English, or Tamil and English.	*pondok school, Orang Asli Affairs, Tabung Haji Board, kampung house* and *nasi lemak stall.* **Yellow**: English Word
Derived Words: A morphosyntactic adaption to make new words.	*datukship (adding sux -ship), Johorean (adding sux -ean)* and *non-halal (adding prex non)*
Inflected Words: A morphosyntactic adaptions to make word forms of a word(s).	*ang pows, bomohs, and cheongsams.*

3 Semantical adaptions of Malay loanwords to the Malaysian English lexicon are done by retaining original meanings with some additional meanings for the Malay words.

Figure 1 shows some examples of semantically adapted Malay loanwords in Malaysian English.

Loanword	Meaning(s) in Bahasa Malaysia	Meaning(s) in Malaysian English
dadah	**(1)** A substance used as a component of a medication. **(2)** A chemical substance, such as a narcotic or hallucinogen, that affects the central nervous system	A chemical substance, such as narcotic or hallucinogen, that affects the central nervous system
rotan	**1** Any of various climbing plantsof tropical Asia, having long, tough,slender stems. **2** The stems of any ofthese plants, used to make wickerwork, canes, and furniture. **3** A cane made from these plants. **4** Judicially-sanctioned caning inMalaysia.	**1** A rattan cane used for inflicting judicially-sanctioned corporal punishment in Malaysia. **2** Any of various climbing plants of tropical Asia, having long, tough,slender stems. **3** A cane. **4** Judicially-sanctioned caning in Malaysia
roti	Bread	Bread, usually the local version of a whiteloaf, which is slightly sweet and has a very soft texture.

Figure 1: Semantic adaptation of Malay loanwords in Malaysian English

Named Entity

An **entity** is a word that is used to represent a noun. The named entity is usually predefined based on the nature of the dataset that is being handled. The entity tag used for this research is adapted from "OntoNotes Release 5.0"[1]. OntoNotes has proposed entity name types and entity value types. For this research, we will be using 9 Entity Tags from OntoNotes entity name types, 1 Entity Tag that has been modified from OntoNotes entity name types, and 2 newly added Entity Tags. In the following chapters, we will provide guidelines for the 12 Entity Tags.

Entity Label

1) PERSON (PER)

This entity type is adapted from OntoNotes 5.0. Entity tag PERSON is used to refer:

1. Name of person (including first, middle, and last name). Example:
    a. [**Ismail Sabri Yaakob**], [**Mohammed Agus Yusoff**]
2. Names of fictional people.
    a. [**Upin**] & [**Ipin**] have always …, and [**Boboi Boy**]
3. Nicknames
    a. [*Health Minister* [**Khairy Jamaluddin**]], in a … so far, [**Khairy**] said.
4. Names with Honorary Titles (like *Datin, Dato, Tan Sri, Tengku*, and more):
    a. wife [*Datin Seri* [**Pauline Leong**]], [*Kelantan Perhilitan director* [**Mohammad Hafid Rohani**]], [*Tan Sri* [**Azman Mokhtar**]], and [*Opposition Leader* [*Datuk Seri* [**Anwar Ibrahim**]]].
5. Names with Role/Position in Organization/Special Social Status
    a. [*Prime Minister* [*Datuk Seri* [**Ismail Sabri Yaakob**]]], [*Tengku Laksamana Selangor* [*Tengku* [**Sulaiman Shah**]], [*Chief Minister* [*Datuk Seri* [**Sulaiman**

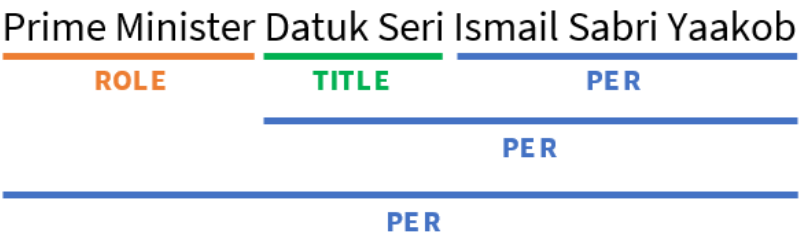
Md Ali]], [*Raja Permaisuri Agong* [*Tunku* [**Azizah Aminah Maimunah Iskandariah**]]], [*Johor Bahru Mayor* [*Datuk* [**Mohd Noorazam Osman**]]], [*ustaz* [**Hussain**], and [*Yang di-Pertuan Agong* [**Al-Sultan Abdullah Ri'ayatuddin Al-Mustafa Billah Shah**]]

b. **AVOID ROLE**:
    i. Makcik, Pakcik, Atuk, Nenek or Anak.

6. Names of Titles with Academic Degrees and Titles
    a. [*Dr.* [**John Raj**]], and [*Associate Professor* [**Deva**]]
7. Names and designations of God.
    a. [*Allah*], [*God*], [*God Almighty*], and [*Insya-Allah*] there will be …
8. Names with Family Name Affixes (like bin, binti, al-, A/L, A/P). Example:
    a. [*Yang Berhormat* [Puan [[**Kasthuriraani**] a/p[Patto]]]] and [[*Tun Hajah Rahah] binti [Tan Sri Haji Mohamed Noah*]]

Notes

1. In some scenarios, the ROLE & TITLE will be embedded together with the entity mentioned PERSON.
2. In some scenarios, annotators will find NESTED ENTITY. In this scenario, the annotator can annotate **BOTH ENTITIES**.
    a. Example: *Tun Hajah Rahah binti Tan Sri Haji Mohamed Noah*
        i. The annotator can annotate "***Tun Hajah Rahah binti Tan Sri Haji Mohamed Noah***", "***Tun Hajah Rahah***", and "***Tan Sri Haji Mohamed Noah***" separately.
3. In some scenarios, the PERSON (PER) name is included as part of the LOC, ORG, and FAC names. Example:
    a. [*Taman Tun Dr Ismail*], [*Perumahan Haji Malik*].
        i. This type of entity should **NOT** be **annotated as a PERSON.**
4. In some scenarios, the ROLE & TITLE will be embedded together with the entity mentioned PERSON. Example:
    a. [**Prime Minister** [**Datuk Seri** [**Ismail Sabri Yaakob**]]]
        i. In this scenario, the Annotator can annotate [**Prime Minister Datuk Seri Ismail Sabri Yaakob**], [**Datuk Seri Ismail Sabri Yaakob**], and [**Ismail Sabri Yaakob**] AS **PERSON.**
        ii. Also, Annotator needs to annotate:
            1. **Prime Minister** as **ROLE**
            2. **Datuk Seri** as **TITLE**
        iii. The annotation for the above example will look like the below:



5. If the name of the PERSON is not complete, unknown, undefined, or not clear then the entity is **NOT LABELLED** as a PERSON. Example:
    a. *The Dr is not feeling well …*
        i. This should not be labelled, because it is not clear which Dr are they referring to.
6. In some scenarios, punctuations or prepositions will be included within an entity name. In that case, the entity should be annotated together. Example:
    a. [Tun [V.T. Sambanthan]]
7. In some scenarios, the acronym of the word will be included beside the entity name. In this scenario, the full name of the entity should be annotated first, followed by the abbreviation. Example:
    a. John F.Kennedy (JFK) … → [John F.Kennedy] [(JFK)] …
8. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
9. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

2) LOCATION (LOC)

This entity type is adapted from OntoNotes 5.0. A LOCATION is any place that can be occupied by or has been occupied by someone in this EARTH and outside of EARTH. Entity tag LOCATION is used to refer

1. Name of a State
    a. [*Wilayah Persekutuan Kuala Lumpur*], [*Federal Territory of Kuala Lumpur*], [*Selangor*] and [*Selangor Darul Ehsan*].
2. Name of a City/Town/Suburban
    a. [*Kuala Lumpur*], [*Shah Alam*], [*Genting Highlands*], [*Petaling Jaya*] [*Taman Tun Dr Ismail*], [*Mont Kiara*], [*Ampang*] [Klang Valley], and [*Georgetown*].
3. Name of a Region
    a. [*Semanajung Malaysia*], [*Peninsular Malaysia*], [*East Coast Malaysia*], [*Central Region*], [*Southern Region*], [*Northern Region*], [*West Malaysia*], [*East Malaysia*], and [*West & East Malaysia*], [*Southeast Asia*]
4. Name of Districts
    a. [*Gombak*], [*Hulu Langat*], [*Hulu Selangor*], [*Klang*], [*Sepang*], [*Kuala Terrenganu*], and [*Cameron Highlands*]
5. Name of Villages

    a. [*Kampung Baru Lanjut*], [*Kampung Kikandis Gorowot*], *Kuala Penyu*, and [*Kampung Teluk Intan*]
6. Island with Local Authority
    a. [*Pulau Langkawi*], [*Pulau Rendang*], [*Pulau Pinang*], and [*Penang*]
7. Names of continents
    a. [*Asia*], [*Europe*], [*Africa*], [*Antarctica*], and [*America*].
8. Names of celestial bodies (Objects in Space)
    a. [*Earth*], [*Bumi*], and [*Moon*].
9. Names of hills or mountains or caves
    a. [*Gunung Ulu Semangkuk*] in [*Fraser's Hill*], [*Pahang*]
10. Names of natural bodies of water (*ocean, river, valley, waterfall, beach, lakes, glacier*).
    a. [*Pantai Bagan Lalang*], [*Black Sea*], [*Air Terjun Pisang*] and [Air Terjun Serendah]
11. Names of jungle or forest
    a. [*Hutan Lipur Gunung Nuang*], and [*Hutan Lipur Sungai Congkak*].
12. Name of natural tourist spots:
    a. [*Taman Eko Rimba Kanching*], and [*Batu Caves*]
13. Name of Residential Areas
    a. [*Perumahan Haji Malik*], [*Taman Salak Kasturi*], and [*Taman Sri Siantan*]
14. Name of Fictional Places.
    a. [*Gotham City*], [*The Emerald City*], and [*The Galaxy*]

Notes

1. Annotators can accept LOCATION (LOC) names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios the PERSON (PER) name is included as part of the LOCATION (LOC) name. Example:
    a. [*Perumahan Haji Malik*].
        i. This type of entity should be **annotated as LOCATION.**
3. In some scenarios, the simpler name of the LOCATION (LOC) name will be used. Example:
    a. Selangor Darul Ehsan can also be called SELANGOR or Selangor.
    b. Pulau Penang can also be called Penang.
        i. This simpler name can also be **annotated as LOCATION**.
4. In some scenarios, the LOCATION (LOC) name is followed after a FAC or ORGANIZATION (ORG) name. Example:
    a. *Asia Pacific University of Technology & Innovation (APU)*.
    b. *New Pantai Expressway.*
    c. *Pantai Hospital Kuala Lumpur*
        i. This type of entity should be **annotated as FAC or ORGANIZATION (ORG)**.

5. If the name of LOCATION is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as LOCATION.Example:
    a. *The city is located in 2KM away from the school*
        i. This should not be labeled, because it is not clear which city and school they are referring to.
6. In some scenarios, punctuations or prepositions will be included within an entity name. In that case, the entity should be annotated separately. Example:
    a. the shop is located in Subang Jaya, Selangor → the shop is located in [*Subang Jaya*], [Selangor]
    b. *Gunung Ulu Semangkuk*, *Pahang* → [*Gunung Ulu Semangkuk*], [*Pahang*]
7. In some scenarios, the acronym of the LOCATION name will be used. Example:
    a. Johor Bahru can also be called JB.
    b. Selangor Darul Ehsan can also be called Sel.
        i. This acronym should **NOT** be **annotated as a PERSON**.
8. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
9. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

3) ORGANIZATION (ORG)

This entity type is adapted from OntoNotes 5.0. ORGANIZATION is a group of people with a specific purpose. Entity tag ORGANIZATION is used to refer to

1. Government and Non-Government Organization (NGO)
    - [*Ministry of Finance*], [*Kementrian Dalam Negeri*] [*(KDN)*], [*KDN*], [*Parliment of Malaysia*] [*Global Peace Foundation*], [*Dewan Negara*], and [Ministry of Defense].
2. Corporate Organization
    - [*Tabung Haji*], [*UEM Groups*], [*Telekom Malaysia Berhad*][*(TM)*], and [*TM*]
3. Political Organization/Party
    - [*United Malays National Organization*] [*(UMNO)*], [*UMNO*], [*Pakatan Nasional*], [*BN*], and [*Democratic Action Party*] [*(DAP)*].
4. Non-Profit Organization
    - [*WWF*] *Malaysia,* [*Malaysian Institute of Human Resource Management (MIHRM)*], and [*MIHRM*]
5. Educational Institution
    - [*National University of Malaysia*], [*University Kebangsaan Malaysia*] [*(UKM)*], [*Multimedia University*] *Cyberjaya*, and [*Multimedia University*] *Melaka*
6. Professional Bodies
    - [*Malaysian Institute Of Human Resource Management*] [*(MIHRM)*], and [*MIHRM*]
7. Sports Organization

- [*Manchester United*], [*Arsenal*], [*Pahang FA*], and [*Selangor FC*]

8. Fans Club
    - [*Arsenal Supporters Malaysian*], [*Johor Bikers Club*], and [*JDT Supporters*].
9. Religious Organization
    - [*Malaysia Hindu Sangam*] [*(MHS)*], [*MHS*], [*Christian Federation Malaysia*], and [*Church of England*].
10. Medical, Law, Media Organization
    - [*Malaysian Red Cresent Society*], [*Academic of Malaysian in Malaysia*], [*Malaysian AIDS Council*], [*New Straits Times*] ([*NST*]), and [*Ram Singh & Co*]
11. Fictional Organization
    - [*Minister of Magic*], [*League of Peoples*], and [*Indigo Tribe*]

Notes

1. Annotators can accept ORGANIZATION (ORG) names in **ENGLISH** and **BAHASA MALAYSIA**.
2. Most of the ORGANIZATION (ORG) will have the suffix **SDN BHD / SDN / BHD / BERHAD**. In that case, **annotate together with the suffix**. Example:
    a. … convicted former prime minister Datuk Seri Najib Razak on charges of misappropriation of [*SRC International Sdn Bhd*] funds …
    b. … recognition ceremony for LBS Bina Group Berhad, here today …
3. In some scenarios, the PERSON (PER) name is included as part of the ORGANIZATION (ORG) name. Example:
    a. [*Ram Singh & Co*].
        i. This type of entity should be **annotated as ORGANIZATION.**
4. In some scenarios, the ORGANIZATION (ORG) name is followed by a LOCATION (LOC) name. Example:
    a. *University Multimedia*, *Cyberjaya*.
        i. This type of entity should be **annotated as ORGANIZATION (ORG)**.
5. If the name of ORGANIZATION is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as ORGANIZATION.Example:
    a. Ministry has been told …
        i. This should not be labeled, because it is not clear which Ministry are they referring to.
6. In some scenarios, the name of ENTITY can be **annotated as either ORGANIZATION (ORG) or FACILITY (FAC).** Example:
    a. All the ministers are attending the session in [*Dewan Negara*] …
    b. The students are going to [*Sekolah Menengah Sungai Besar*]
        i. In this scenario, annotators must choose the label based on the context. For the above example, FAC will be more suitable than ORG.
7. In some scenarios, an **ORGANIZATION (ORG)** name will start with the word “THE”. The definite article should not be included when annotating the entity:
    a. Example: *The New York Times* → *The* [*New York Times*]
8. In some scenarios, punctuations or prepositions will be included within an entity

name. In that case, the entity should be annotated separately. Example:
    a. … component parties comprising [*UMNO*], [*MCA*], [*MIC*], and [*Parti Bersatu Rakyat Sabah*] ([*PBRS*]). The convention would also be attended by a coalition party …
9. In some scenarios, the abbreviations or acronyms of the word will be included beside the entity name. In this scenario, the full name of the entity should be annotated first, followed by the abbreviation. Example:
    a. the member of party Pakatan Harapan (PH) has … → the member of the party [*Pakatan Harapan*] *(*[*PH*]*)* has …
        i. **DO NOT** include parenthesis (or bracket) symbols when annotating entity names.
10. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
11. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

4) NORP

This entity type is adapted from OntoNotes 5.0. NORP is the abbreviation for the term *Nationality, Religious or Political* group. Entity tag NORP is used to refer to

1. Words related to LOCATION (LOC)
    - [*Selangorian*], [*Johorean*], and [*Penangite*]
2. Words related to PEOPLE of certain Religion or Tribe
    - [*Non-Bumiputera*],[*Bumiputera*], [*Malays*], [*Indians*], [*Chinese*], and [*Orang Asli*]
3. Words related to PEOPLE of certain Political Group
    - [*Communist*], and [*Democratic*]

Notes
1. Annotators can accept NORP names in **ENGLISH** and **BAHASA MALAYSIA**.
2. If the name of NORP is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as NORP.
3. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
4. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

5) FACILITY (FAC)

This entity type is adapted from OntoNotes 5.0. FACILITY refers to man-made structures. Entity tag FAC is used to refer to

1. Name of Residential Place
    - [*Apartment Kinara*], [*Menara Rajawali*], [*Hotel Seri Malaysia*], and [*Sapphire*

Residence]

2. Name of Non-Residential Place
    - [*Menara Tabung Haji*], [*Menara Bank Rakyat*], [*KLCC*], and [*Putra World Trade Centre*] [*PWTC*]
3. Name of Public Place/ Physical Tourists Places
    - [*Anytime Fitness*], [*Taman Negara National Park*], [*Sunway Pyramid*], and [*Zoo Negara*]
4. Name of Transportation Infrastructure
    - [*Petaling Street*], [*Kuala Lumpur International Airport* [(*KLIA*)]], [*LRT Asia Jaya*], [*PLUS*], [*Kesas Highway*] and [*Cyberjaya Bus Station*]

Notes

1. Annotators can accept FACILITY (FAC) names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios the name of FACILITY (FAC) will have the name of PERSON (PER) or ORGANIZATION (ORG) or LOCATION (LOC). Example:
    a. [*Jalan Tun Razak*], [*Menara Kuala Lumpur*], and [*Menara Tabung Haji*]
        i. In this scenario, the annotator needs to label that as a **FACILITY (FAC)**
3. If the name of FACILITY (FAC)is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as FACILITY (FAC). Example:
    a. We went to the Zoo last weekend …
        i. This should not be labeled, because it is not clear which Zoo they are referring to.
4. In some scenarios, punctuations or prepositions will be included within an entity name. In that case, the entity should be annotated separately. Example:
    a. the shop is located in Sunway Pyramid, Petaling Jaya → the shop is located in [*Sunway Pyramid*], Petaling Jaya
5. In some scenarios, the abbreviation or acronym of words will be included beside the entity name. In this scenario, the full name of the entity should be annotated first, followed by the abbreviation or acronym. **Don't annotate both full name and abbreviation or acronym together.** Example:
    a. The [*Kuala Lumpur Convention Centre*] ([*KLCC*]) is implementing a RM2,500 minimum wage …
6. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
7. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

6) PRODUCT

This entity type is adapted from OntoNotes 5.0. Entity tag PRODUCT is used to refer to

1. Name of General Products
    - [*Acer Aspire Nitro Black Edition*], [*Samsung Galaxy S9*], and [*Ramly Chicken*

Burger]

2. Name of Mode of Transportation
    - [*Suzuki SX4*], [*Boeing 747*], and [*Perodua Myvi*]
3. Name of Services or Plan
    - [*PTPTN* loan], [*Amanah Saham Bumiputera*], [*Hotlink Postpaid + Fibre*], [*Maxis Home 4G WiFi*] [*Facebook*], and [*WhatsApp*]

Notes

1. Annotators can accept PRODUCT names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios, the name of FACILITY (FAC) will have the name of PERSON (PER) or ORGANIZATION (ORG), or LOCATION (LOC). Example:
    a. [*Acer Aspire Nitro Black Edition*], and [*Suzuki SX4*]
        i. In this scenario, the annotator needs to label that as a **PRODUCT.**
3. If the name of PRODUCT is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as PRODUCT.
4. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
5. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

7) EVENT

This entity type is adapted from OntoNotes 5.0 and ACE 2005. An EVENT is a reference to an organized or unorganized incident. Examples of organized EVENT like sports events, festival seasons, and public events. Examples of unorganized EVENT like natural disasters, and war. Entity tag EVENT is used to refer to

1. Names of natural disasters
    a. [*Indian Ocean earthquake and tsunami*]
    b. Any nouns referring to Natural Disaster
        i. [*flood*], [*flash flood*]
        ii. [*crisis*]
        iii. [*tornado*], [*storm*]
        iv. [*earthquake*]
        v. [*landfall*], [*landslide*]
        vi. [*tsunami*]
    c. What is considered a natural event?
        i. These events may include earthquakes, floods, hurricanes, tornadoes, tsunamis, landslides, wildfires, volcanic eruptions, and extreme temperatures.
2. Name of revolution, or war
    a. [*Ukraine-Russia War*], and [*World War II*]
    b. Any nouns referring to War. Example:

        i. [*terror attack*]
        ii. [*gun-shoot*], [*gun-fire*]
        iii. [*war*]
3. Name of Public Incident or Demonstration
    a. [*9/11*], [*May 13*], [*Bersih 3.0*], [*Russian invasion*], [*Pilihan raya negeri Melaka ke-15*]
    b. Any nouns referring to Public Incident.
        i. [*Pilihan Raya*]
        ii. [*demonstration*]
        iii. [*protests*]
        iv. [*ceremony*]
        v. [*riot*]
4. Name of Public Event or Performance
    a. [*Oscar 2022*], [*UEFA Champions League*], [*FA Cup*], and [*Sudirman Cup*]
5. Names of festive or holiday season
    a. [*Hari Raya Aidilfitri*], [*Deepavali*], and [*Chinese New Year*]
6. Event Nouns:
    a. Examples, related to PERSON:
        i. [*birth*], [*marriage*], [*divorce*], [*shot*], [*assassination*], [*sentenced*]
    b. Examples, related to ORGANIZATION:
        i. [*acquisition*], [*signing ceremony*], [*bankrupt*], [*bankruptcy*], [*meeting*].
7. Other allowed EVENT mentions:
    a. pandemic.

Notes

1. Annotators can accept EVENT names in **ENGLISH** and **BAHASA MALAYSIA**.
2. **Verbs should not be included as EVENT.**
3. Annotators will **annotate** the **EVENT based** on the **context** of the **ARTICLE**.
4. In some scenarios, the Event Noun and Event Name will be appearing together. Example:
    a. … like we are going to face [*PRU-15*] (next [*general election*]) tomorrow …
        i. **In this scenario, annotate both the Event name and Event noun. Both PRU-15 and the general election should be annotated.**
5. If the name of EVENT is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as EVENT. Example of incomplete EVENT name:
    a. Instead of *Hari Raya Aidilfitri*, they mention *Raya*
6. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
7. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

8) WORK OF ART

This entity type is adapted from OntoNotes 5.0. Entity tag WORK_OF_ART is used to refer to

1. Titles of Books, Songs, Films
    - [*Salina*], [*Harry Potter and The Chamber of Secrets (1998)*], [*Hantu Kak Limah*] and [*The Grayman*]
2. Titles of Awards
    - [*Ballon D'OR*], [*Piala Sudirman*]

Notes

1. Annotators can accept WORK_OF_ART names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios the name of WORK_OF_ART will have the name of PERSON (PER) or ORGANIZATION (ORG), or EVENT, LOCATION (LOC). Example:
    a. The child falls asleep while reading the novel [*Hantu Kak Limah*] …
        i. In this scenario, the annotator needs to label that as a **WORK_OF_ART**
    b. The team won [*Piala Sudirman*] last night …
        i. In this scenario, annotators must choose the label based on the context. For the above example, WORK_OF_ART will be more suitable than PERSON (PER).
3. If the name of WORK_OF_ART is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as WORK_OF_ART. Example of incomplete WORK_OF_ART name:
    a. I love to listen to *Billie Eilish* songs …
        i. Billie Eilish's song is undefined and not complete.
4. In some scenarios, the abbreviations or acronyms of words will be included beside the entity name. In this scenario, the full name and abbreviations or acronyms of the entity should be annotated together. Example:
    a. [ABCD (Any Body Can Dance)]
5. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
6. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

9) LAW

This entity type is adapted from OntoNotes 5.0. Entity tag LAW is used to refer to

1. Name of legalized document
    - [*Akta Kesalahan Pilihan Raya 1954*], [*Akta Bandaraya Kuala Lumpur 1971*], [*Akta Cukai Perjudian*], [*Akta Angkatan Tentera*], [*Article 153 – Special Position of Bumiputras and Legitimate Interests of Other Communities*], [*Perkara 48 Perlembagaan Malaysia*] and [*Standing Order 11(3)*].

Notes

1. Annotators can accept LAW names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios the name of LAW will have the name of PERSON (PER) or ORGANIZATION (ORG), or LOCATION (LOC). Example:
    a. [*Akta Bandaraya Kuala Lumpur 1971*] …
        i. In this scenario, the annotator needs to label that as a **LAW.**
3. If the name of LAW is not complete, unknown, undefined, or not clear then the entity **NOT BE LABELED** as LAW.
4. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
5. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

10) LANGUAGE

This entity type is adapted from OntoNotes 5.0. Entity tag LANGUAGE is used to refer to

1. Name of language spoken by people:
    - [*Bahasa Malaysia*], [*Mandarin*], and [*Tamil*]

Notes

1. Annotators can accept LANGUAGE names in **ENGLISH** and **BAHASA MALAYSIA**.
2. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
3. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

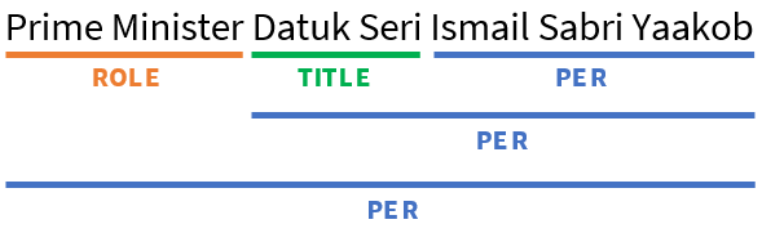
11) ROLE

This entity type is added to our relation label list. Entity tag ROLE is used to define the position or function of the PERSON in an ORGANIZATION. This ROLE is often used as part of the entity mentioned PERSON. Entity tag ROLE is used to refer to

1. Position in Government Organization
    a. [*Prime Minister*] Datuk Seri Ismail Sabri Yaakob, [*Tengku Laksamana Selangor*] Tengku Sulaiman Shah, [*Chief Minister*] Datuk Seri Sulaiman Md Ali, [*Raja Permaisuri Agong*] Tunku Azizah Aminah Maimunah Iskandariah, [*Johor Bahru Mayor*] Datuk Mohd Noorazam Osman.
2. Position in Corporate Organization
    a. [*managing director*] Datuk Azmir Merican, Toyoda [*chief executive officer*] Toyoda Misako, the great-granddaughter … and … Axiata Group Bhd's [*president*] and group [*chief executive officer*] ([*CEO*]) Datuk Izzaddin Idris is stepping down …

Notes

1. Annotators can accept ROLE names in **ENGLISH** and **BAHASA MALAYSIA**.
2. The entity label ROLE will be the **prefix** of entity **PERSON**.
3. You are allowed to annotate the ROLE that is without the name of PERSON.
4. In some scenarios, the TITLE will be between the ROLE and PERSON entity mentioned. Example:
    a. [*Prime Minister*] [*Datuk Seri*] Ismail Sabri Yaakob.
        i. **In this scenario, annotate ROLE and TITLE separately. The annotated sample will look like the below:**



5. In some scenarios, the ORGANIZATION will be followed with ROLE. Example:
    a. Toyoda [*chief executive officer*] Toyoda Misako.
        i. In this scenario, the name of **ORGANIZATION should NOT BE annotated together with ROLE.**
6. In some scenarios, a PERSON can be assigned with multiple ROLE or TITLE. Example:
    a. Axiata Group Bhd's [*president*] and group [*chief executive officer*] ([*CEO*]) Datuk Izzaddin Idris …
        i. In this scenario, the **ROLE** should be annotated separately like above.
7. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
8. If unsure about the mention, kindly refer to Google for more clarity on the entity mentioned and decide on the suitable Entity Label.

12) TITLE

This entity type is added to our relation label list. Entity tag TITLE is used to define the honorific title of the PERSON. This TITLE is often used as part of the entity mentioned PERSON. Entity tag TITLE is used to refer to

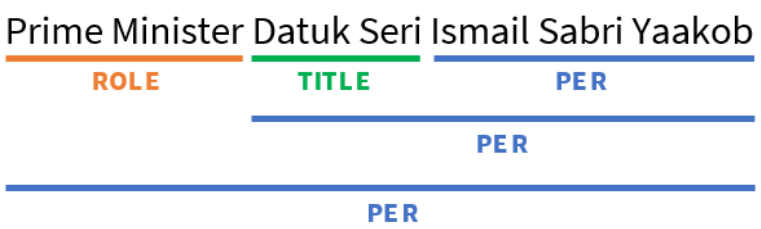
1. Common Title. Example:
    a. … to tell the world that [*Mr*] Depp is not the abuser [*Miss*] Heard said he is and to …
    b. More Examples can be viewed here: Link
2. Academic Title
    a. [*Dr*] Noor Hisham said the number of recoveries …

b. More Examples can be viewed here: Link

3. Religious Title
    a. [*ustaz*] **Hussain**
    b. More Examples can be viewed here: Link
4. Title in Military/Police
    a. [*Leftenan-General*] (Rtd) Datuk Sri Abdul Aziz Ibrahim said that it should …
5. Federal/ State Titles
    a. … managing director [*Datuk*] Yeoh Seok Hong, Yes' 5G for All was the company's commitment …
    b. … for Buddhism, Christianity, Hinduism, Sikhism, and Taoism (Limas) chairman [*Dato*]' Teng Chang Khim estimated …
    c. More Examples can be viewed here: Link
6. Malay Royalty Titles
    a. [*Yang di-Pertuan Agong*] Al-Sultan Abdullah Ri'ayatuddin Al-Mustafa Billah Shah
    b. officially launched by the [*Sultan of Selangor*] [*Sultan*] Sharafuddin Idris Shah.
    c. More Examples can be viewed here: Link

Notes

1. Annotators can accept TITLE names in **ENGLISH** and **BAHASA MALAYSIA**.
2. The entity label TITLE will be the **prefix** of entity **PERSON**.
3. You are allowed to annotate the TITLE that is without the name of PERSON.
4. In some scenarios, the TITLE will be between the ROLE and PERSON entity mentioned. Example:
    a. [*Prime Minister*] [*Datuk Seri*] Ismail Sabri Yaakob.
        i. **In this scenario, annotate ROLE and TITLE separately. The annotated sample will look like the below:**



5. In some scenarios, the entity mentioned ROLE and TITLE will be similar. Example:
    a. … and [*Pahang Mufti*] Datuk Seri Dr Abdul Rahman Osman were also present …
        i. **In this scenario, MUFTI is a religious title. The word MUFTI is annotated as TITLE, while Pahang will be LOCATION.**
6. In some scenarios, the entity mentioned for TITLE can be part of the ORGANIZATION name. Example:

    a. [*Jabatan Mufti Negeri Pulau Pinang*]
        i. In this scenario, this should be annotated as ORGANIZATION instead of TITLE.
7. In some scenarios, part of the entity name will be in lowercase. The lowercase word also should be included when annotating the entity.
8. If unsure about the mention, kindly refer to Google for more clarity on the entity mention and decide on the suitable Entity Label.

Descriptor

Descriptors are words used to describe or identify something. In this section, we will provide a few descriptors that annotators can use to identify the ENTITY name in the article.

Descriptors	**Entity**
orang	PERSON (PER)
,says	PERSON (PER)
agency	ORGANIZATION (ORG)

Descriptors	**Entity**
appointed as	PERSON (PER)
man	PERSON (PER)
association	ORGANIZATION (ORG)

persatuan/society	ORGANIZATION (ORG)
forces	ORGANIZATION (ORG)
public fund	ORGANIZATION (ORG)
university/universiti	ORGANIZATION (ORG)
committee/jawatankuasa	ORGANIZATION (ORG)
institutes/institut	ORGANIZATION (ORG)
valley/lembah	GPE/LOCATION (LOC)/ FACILITY (FAC)
station/stesen	GPE/LOCATION (LOC)/ FACILITY (FAC)
hill/mountain/gunung/ bukit	GPE/LOCATION (LOC)
residential/perumahan	GPE/LOCATION (LOC)
province/wilayah	GPE
parking	FACILITY (FAC)
bridge/jambatan/jejantas	FACILITY (FAC)
hsopital	FACILITY (FAC)
station/stesen	FACILITY (FAC)
street/jalan	FACILITY (FAC)
mosque/masjid/surau	FACILITY (FAC)
condominium/kondominium	FACILITY (FAC)
temple/rumah ibadat/tokong	FACILITY (FAC)
the	PRODUCT

bank	ORGANIZATION (ORG)
group/kumpulan	ORGANIZATION (ORG)
kolej/college	ORGANIZATION (ORG)
politeknik/polytechnic	ORGANIZATION (ORG)
company/syarikat	ORGANIZATION (ORG)
area/kawasan	GPE/LOCATION (LOC)
island/pulau	GPE/LOCATION (LOC)
city/bandar	GPE/LOCATION (LOC)
lake/tasik	LOCATION (LOC)
park/taman	GPE/LOCATION (LOC)
forest/jungle/hutan	GPE/LOCATION (LOC)
state/negeri	GPE
village/kampung	GPE
at/near	GPE/LOCATION (LOC)/ FACILITY (FAC)
airport/lapangan terbang	FACILITY (FAC)
factory/kilang	FACILITY (FAC)
highway/lebuh raya	FACILITY (FAC)
menara/tower	FACILITY (FAC)
apartment/pangsapuri	FACILITY (FAC)
residence/flat	FACILITY (FAC)
orang	NORP
celebrate/celebration	EVENT
session	EVENT

Summary Table

This table will be a summary of the different entity labels we have covered.

#	Entity Label	Description	Example
1	PERSON (PER)	Refers to the name of people. This includes the names of people (first name, middle name, and last name), nicknames, fictional people names, names of people with title or position, and names of gods.	1) [*Health Minister* [**Khairy Jamaluddin**]] 2) [**Ismail Sabri Yaakob**] 3) [**Upin**] & [**Ipin**] 4) [*Yang di-Pertuan Agong* [**Al-Sultan Abdullah Ri'ayatuddin Al-Mustafa Billah**

			Shah]] 5) [*Allah*] 6) [*Dr.* [**John Raj**]]
2	LOCATION (LOC)	Refers to places in and out of EARTH. This includes the name of a city, state, country, districts, villages, residential areas, natural bodies of water, island, fictional places, continents, jungle, tourist spots, and fictional places.	1) [*Selangor*] 2) [*Klang*] 3) [*Mont Kiara*] 4) [*Klang Valley*] 5) [*Kampung Baru Lanjut*] 6) [*Taman Salak Kasturi*],
3	ORGANIZATION (ORG)	Refers to a group of people with a specific purpose. This includes the names of government bodies, names of non-government bodies, names of political bodies, names of non-profit organizations, names of corporate organizations, names of educational institutions, names of sports organizations, names of fans clubs, names of religious organizations, names of medical organizations, and names of fictional organizations.	1) [*Ministry of Finance]* 2) [*Tabung Haji*] 3) [*Pakatan Nasional*] 4) [*University Kebangsaan Malaysia*] 5) [*MIHRM*] 6) [*Arsenal*] 7) [*Malaysia Hindu Sangam*] 8) ([*NST*]) 9) [*Indigo Tribe*]
4	NORP	Refers to the term Nationality, Religious, and Political group.	1) [*Selangorian*] 2) [*Orang Asli*] 3) [*Bumiputera*] 4) [*Democratic*]
5	FACILITY (FAC)	FACILITY refers to man-made structures. Includes, residential places, non-residential places, public places, tourist places, and transportation infrastructures.	1) [*Apartment Kinara*] 2) [*Menara Tabung Haji*] 3) [*Anytime Fitness*] 4) [*Kuala Lumpur International Airport* [(*KLIA*)]] 5) [*PLUS*]
6	PRODUCT	PRODUCT refers to the day to day things or services that we used. **ONLY** PRODUCT **with** the **brand** name will be **annotated**.	1) [*Acer Aspire Nitro Black Edition*] 2) [*Produa Myvi*]

			3) [*Hotlink Postpaid + Fibre*] 4) [*Maxis Home 4G WiFi*] 5) [*Facebook*]
7	EVENT	An EVENT is a reference to an organized or unorganized incident. Examples of organized EVENT like sports events, festival seasons, and public events. Examples of unorganized EVENT like natural disasters, and war. We are also referring to EVENT nouns.	1) [*Indian Ocean earthquake and tsunami*] 2) [*tsunami*] 3) [*Ukraine-Russia War*] 4) [*gun-fire*] 5) [*demonstration*] 6) [*Pilihan raya negeri Melaka ke-15*] 7) [*Oscar 2022*] 8) [*Hari Raya Aidilfitiri*], 9) [*born*] 10) [*dead*] 11) [*acquired*]
8	WORK OF ART	WORK OF ART refers to titles of books, songs, film-name, awards, and stock-price.	1) [*Harry Potter and The Chamber of Secrets (1998)*] 2) [*Piala Sudirman*] 3) [*NASDAQ: AAPL*]
9	LAW	Refers to legal documents and acts.	1) [*Akta Kesalahan Pilihan Raya 1954*] 2) [*Akta Bandaraya Kuala Lumpur 1971*] 3) [*Akta Cukai Perjudian*],
10	LANGUAGE	Refers to day to day language speaken by PERSON.	1) [*Bahasa Malaysia*] 2) [*Mandarin*] 3) [*Tamil*]
11	ROLE	Entity tag ROLE is used to define the position or function of the PERSON in an ORGANIZATION. This includes	1) [*Prime Minister*] Datuk Seri Ismail Sabri Yaakob 2) [*Tengku Laksamana*

		positions in government, non-government, and corporate organizations.	*Selangor*] Tengku Sulaiman Shah 3) Axiata Group Bhd's [*president*] and group [*chief executive officer*] ([*CEO*]) Datuk Izzaddin Idris
12	TITLE	Entity tag TITLE is used to define the honorific title of the PERSON.	1) [*Mr*] Depp 2) [*Dr*] Noor Hisham 3) [*Leftenan-General*] (Rtd) Datuk Sri Abdul Aziz Ibrahim 4) [*Datuk*] Yeoh Seok Hong 5) [*Yang di-Pertuan Agong*] Al-Sultan Abdullah Ri'ayatuddin Al-Mustafa Billah Shah

Relations

A Relation represents the connection between two Entities. Every Relation Instance will have at least Two Entities. The entities can be within the same sentence or across sentences. A Relation Instance will look like this: (Entity One, Relation, Entity Two). The First Step of annotating Relation, is to annotate **CORRECT** Entity. In the Phase 1, YOU have **CORRECTLY** Annotated the Entities. The Entities you have Annotated in Phase 1, we will pre-annotated in Doccanno. You will need to start from there.

In Phase 2, you **MUST** ANNOTATE Relations from TWO Relation Label groups. These two Relation Groups are from DocRED and ACE05. Having Relations from BOTH Group will help to Improve this Research Project. Phase 1 has covered Relation Labels from DocRED, but Phase 2 will cover both DocRED and ACE05.

To make easier for annotators, I have prepared TWO helping hands:

1) Relation Recommendation Tool. This tool will give suggestions on the Relation that can be possible between two Entity Pair. Please refer to the "How to Use" guide for more information.
2) Suggestion of Entity Type. Every Relation Label can be only used with a certain type of Entity Type. For example: "country", can only be used with Entity Type "PERSON", and "LOCATION". This type can be either from Entity 1 or Entity 2 or Both. In this Relation Label Guide, there will be some suggestions of Entity Type being used for every Relation Labels.

Relation from DocRED

Labels are used to define the connection between two entities. The relation labels are derived from WikiData (An Open Source Knowledge Base). Annotators can refer to the Relation Label Summary Document for further clarification on the Relation Labels allowed for this project.

Relation Label Summary - DocRED

Relation from ACE05

ACE05, has 18 Relation Labels. Annotators can refer to the Relation Label Summary Document for further clarification on the Relation Labels allowed for this project.

Relation Label Summary - ACE05

Conclusion

Annotators must annotate the relations based on labels from both DocRED and ACE05. Every relation label has its description and meaning. However, some Relation Labels from DocRED and ACE05 have similar meaning. The Annotators can refer the table below to find the relation labels that share similar meaning:

No	ACE05	DocRED	Description
1	Physical.Located	location, country, located in the administrative territorial entity, located in the statistical territorial entity	Located captures the physical location of an entity.
2	Physical.Near	country	Indicates that an entity is explicitly near another entity.
3	Part-Whole.Subsidiary	country of citizenship	Captures the ownership, administrative, and other hierarchical relationships between organizations and between organizations and GPEs.
4	Per-Social.Business	partnership with, has part or parts	Captures the connection between two entities in any professional relationships.
5	Per-Social.Family	father, mother	Captures the connection between one entity and another entity in family relations.
6	Org-Aff.Employment	work location, chief executive officer	Captures relationship between Person and their employers.
7	Org-Aff.Ownership	chairperson, owned by, board member	Captures relationship between a Person and an Organization owned by that PERSON
8	Org-Aff.Sports-Affiliation	member of sports team, coach of sports team	Captures relation between Player, Coach, Manager with their affiliated Sport ORG
9	Org-Aff.Shareholder	has part or parts, parent organization	Captures the relation between an agent and an Organization
10	Org-Aff.Membership	position held, subject has	Membership captures relation

		role, member of political party, has part or parts, member of	between an entity and organization which the entity is a member of
11	Agent-Artifact.UOIM	connects with	When an entity own an artifact, uses an artifact or caused an artifact to come into being.
12	Gen-Aff.CRRE	country of citizenship, country of origin, place of birth, ethnic group	When there is a relation between PER and 1) LOC in which they have citizenship 2) LOC they live 3) religious ORG or PER 4) LOC or PER entity that indicates their ethnicity
13	Gen-Aff.Loc-Origin	headquarters location	Captures the relation between an organization and the LOC where it is located.

TIPS: You can annotate the Relation Label that matches the similar meaning in DocRED and ACE05. This means, if you identify the Relation between entities as 'country', you can choose to annotate that entity pair either with Physical.Located or Physical.Near. Please make sure the relation label suits the context of the article.