

Thai Nested Named Entity Recognition Guideline

Key Principles

If you are uncertain if a phrase is a named entity, please examine whether it is a proper noun that has a clear, specific, and identifiable reference.

For example, the government of Yingluck (รัฐบาลยิ่งลักษณ์) is considered a named entity because it doesn't just refer to the government of Ms. Yingluck (รัฐบาลของนางสาวยิ่งลักษณ์). Rather, it refers to a group of people consisting of Minister A + Minister B + ... etc.

On the other hand, Thai soldiers (ทหารไทย) are considered a named entity only in certain contexts. For instance, “a Thai soldier” in “a Thai soldier being arrested (ทหารไทยโดนจับ)” is not considered as a named entity because it only refers to a Thai soldier while “Thai soldiers” in “Thai soldiers forbid not to leave home (ทหารไทยสั่งห้ามออกจากบ้าน)” is a named entity because it specifically refers to a military institution.

If you come across a word or phrase that doesn't match any of the labels or you are not sure whether to label or not, please **leave it**.

If you're unsure of which labels should be masked, please ask your project manager.

The same phrase can be of several categories. So, it must be considered mainly from the context, for example, Chulalongkorn University (จุฬาฯ) will be org:edu when it appears to refer to the institution, not a place, like in “graduating from Chulalongkorn University(จบการศึกษาจากจุฬาฯ).” But the same word will be loc:others when it appears that it is used to refer to a place such as in “come to open the event at Chulalongkorn University (เดินทางไปเปิดงานที่จุฬาฯ).”

If the same named entity is found repeatedly in a sentence or other passage, mark it as well, according to the context in which it was found.

It is possible to find a named entity with the wrong syllable cut, causing some irrelevant messages jutting out. For example, the Party Matchimathipataya (พรรคพรรค

มัชฌิมาธิปไตย) cannot be marked as only the Matchimathipataya Party (พรรค
มัชฌิมาธิปไตย). In this case, there is no need to label. On the contrary, if the syllable is cut
wrong, but still able to cover the whole named entity, you can mark it as usual. For example,
Mr. Kamol (นาย กมล) was mistakenly cut to Mr. K amol (นาย ก มล).

This is what you will consider when you find a specific name that is misspelled, such
as the case of BOT(อปท.). If spelled incorrectly without dots, for example, written in the
BOT (อปท), mark it as usual. But, if there are other misspellings, please choose not to mark.

If there is more than one label that covers the exact same text, choose a more specific
one. For example, Tomyantee (ทมิยันตี) can be both person and pseudonym. In this case,
please mark it as pseudonym.

Frequently important cases

****NOTE:** In the example in this guideline, the largest named entity is at the bottom,
but in the actual platform (Datasaur), the largest named entity is at the top.**

Cases where mislabels frequently happen:

The case of overlapping tags

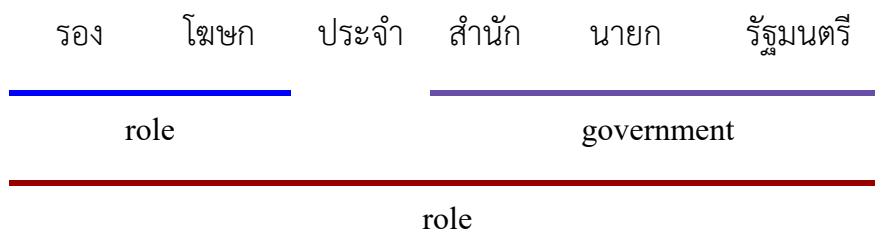


Deputy Spokesperson for the Prime Minister's Office
(รองโฆษกประจำสำนักนายกรัฐมนตรี)

Overlapping tags means that the sub-tags have overlaps with the parent tag, but their
parts jut out.

As an example, it can be seen that the Deputy Spokesman for the Prime Minister's
Office (รองโฆษกประจำสำนักนายกรัฐมนตรี) and the Spokesman for the Prime Minister's Office
(โฆษกประจำสำนักนายกรัฐมนตรี) are nested, but some parts are overlapping.

Solution: mark each subtag in descending order so that it doesn't overlap the larger tag.



The Deputy Spokesman (รองโฆษก) is marked as role.

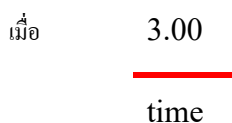
The Prime Minister's Office (สำนักนายกรัฐมนตรี) is marked as government.

Deputy Spokesman for the Prime Minister's Office
(รองโฆษกประจำสำนักนายกรัฐมนตรี) is marked as role.

Do not overlap the tags. For example, do not tag the Deputy Spokesperson of the Prime Minister's Office (รองโฆษกสำนักนายกรัฐมนตรี) and the Prime Minister's Office (สำนักนายกรัฐมนตรี). In such cases, always choose the longest tag.

In case of [number + unit]

2.1. When numbers or numbered words appear alone and they represent the named entity (time, day, month, year, etc.), label it as such.



In “at 3.00”, 3.00 is marked as time.

2.2 If there are other words indicating their own category, label the number + unit as that category and label the number appearing alone as cardinal, for example:



12 is marked as cardinal.

o'clock is marked as unit.

12 o'clock (12 นาฬิกา) is marked as time.

If there are both numbers and units/classifiers, label as in the second case **only**.

In case of [noun indicating categories + proper noun]

If there is a noun indicating categories + a proper noun, superimpose it. If the name inside is a name that already knows the category to be labeled, mark it as such.

For example, Khon Kaen University (มหาวิทยาลัยขอนแก่น), in a sentence, Khon Kaen University has ordered the suspension of teaching and learning (มหาวิทยาลัยขอนแก่นมีคำสั่งหยุดการเรียนการสอน), will be marked as follows:

มหาวิทยาลัย	ขอนแก่น
<hr/>	
province	
<hr/>	
org:edu	

Mark Khon Kaen (ขอนแก่น) as province.

Mark Khon Kaen University (มหาวิทยาลัยขอนแก่น) as org:edu.

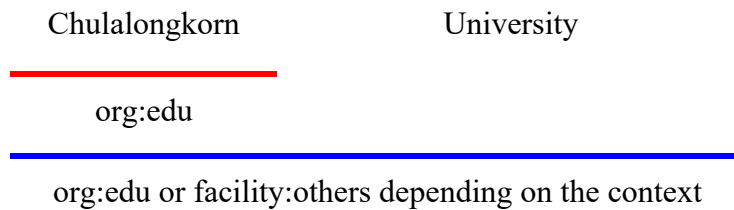
In the case that the name inside is not in another category, mark it in the same category. For example, Nang Ram Beach (หาดนางรำ) shall be marked as follows:

หาด	นางรำ
<hr/>	
loc:others	
<hr/>	
loc:others	

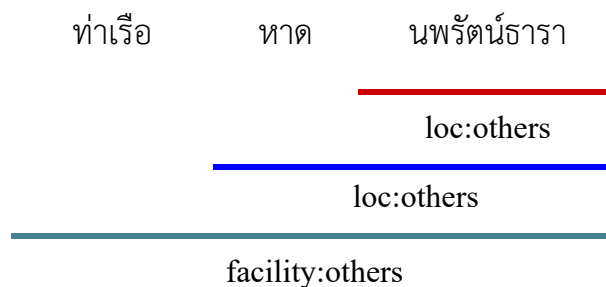
Nang Ram (นางรำ) shall be marked as loc:others.

Nang Ram Beach (หาดนางรำ) shall also be marked as loc:others.

If the name inside appears in several categories, e.g. Chulalongkorn (จุฬาลงกรณ์) in Chulalongkorn University (จุฬาลงกรณ์มหาวิทยาลัย). It may be firstname (the name of King Rama 5) or it may be org:edu. The label is based on the meaning that occurs more frequently, which is org:edu in this case.



This type of structure can be nested in several layers. Label the words in every layer. For instance, Nopparat Thara Beach Pier (ท่าเรือหาดนพรัตน์ธารา) is marked as follows:



Nopparat Thara (นพรัตน์ธารา) is marked as loc:others.

Nopparat Thara Beach (หาดนพรัตน์ธารา) is marked as loc:others.

Nopparat Thara Beach Pier (ท่าเรือหาดนพรัตน์ธารา) is marked as facility:others.

In case of address

Please label as wide as possible (which still refers to an address) by including the word “house number (บ้านเลขที่)” or “number (เลขที่)”, etc., if any, but do not include directions such as

“Headed from Huai Khwang Intersection. Go straight and then you will see a junction. After that, turn left and go straight for a bit. The shop will be on the right-hand side in front of the gold shop (มาจากแยกห้วยขวาง ตรงเข้ามาจะเจอ 3 แยก เลี้ยวซ้ายตรงเข้ามานิดนึงร้านจะอยู่ด้านขวามือหน้าร้านทอง).”

In this case, it is not considered an address.

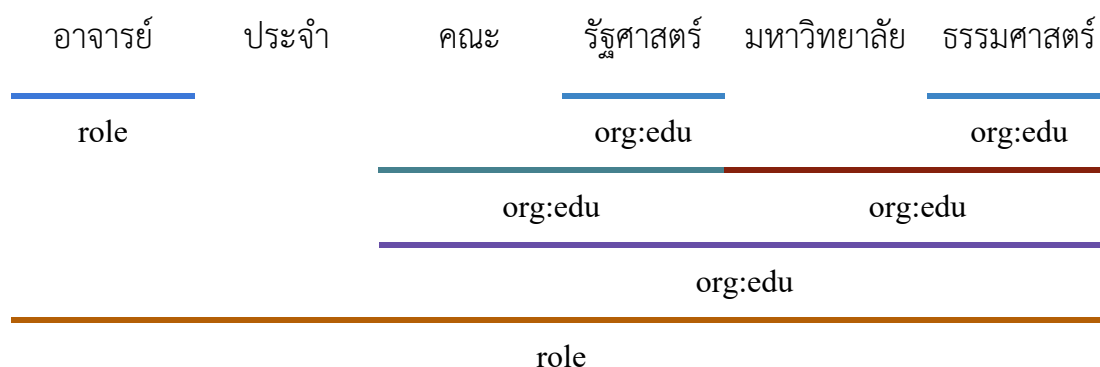
In the case of prefix + position name (ตำแหน่ง+คำขยาย)

In the case of having prefixes such as vice(รอง)_____, assistant(ผู้ช่วย)_____, or former(อดีต)_____, include these prefixes and label the nested positions that do not contain these prefixes inside. (See example number 1)

In the case of position + affiliation (ตำแหน่ง+สังกัด)

*** (Changed from the original guideline to prevent overlapping tags) ***

If you find the case of positions + many affiliations or the affiliations being nested in many layers, mark sub-position as a role and label affiliations separately in different layers. Then, mark the position as a role in the larger layer. However, do not overlap them, as in the example:



A lecturer (อาจารย์) is marked as role.

Political Science (รัฐศาสตร์) is marked as org:edu.

Thammasat (ธรรมศาสตร์) is marked as org:edu.

Faculty of Political Science (คณะรัฐศาสตร์) is marked as org:edu.

Thammasat University (มหาวิทยาลัยธรรมศาสตร์) is marked as org:edu.

Faculty of Political Science, Thammasat University (คณะรัฐศาสตร์

มหาวิทยาลัยธรรมศาสตร์) is marked as org:edu.

Professor of Faculty of Political Science, Thammasat University

(อาจารย์ประจำคณะรัฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์) is marked as role.

In the case of nested organizations in the organization or organization + affiliation

*** (Changed from the original guideline to prevent overlapping tags) ***

For example, in the Faculty of Arts, Chulalongkorn University, all can be labelled as follows:

Faculty of Arts, Chulalongkorn University

Faculty of Arts

Chulalongkorn University

They are all org:edu.

In case of tags in the NORP category

All tags in the NORP category are tags used to define noun extensions and the nouns are proper nouns. If there is a name that falls under the category of nationality, norp:political, and religion, and the name also does not denote nouns (people of (ชาว or คน)....., in the group of (กลุ่ม)), leave it blank.

In case the tag doesn't match the label

If you're not certain if a tag matches an existing label, remove tags, especially the tags that are other (see the guideline definitions).

In case of [name of a place + name of a location]

as in the following texts

Home-style restaurant serving fresh seafood from Wonnapha Beach Pier, Bangsaen.
(ร้านอาหารสไตล์บ้าน ๆ มีอาหารทะเลสด ๆ จากท่าเรือหาดวอนนภา บางแสน)

If there is a place name Bangsaen (บางแสน), next to Wonnapha Beach Pier (ท่าเรือหาดวอนนภา), please consider whether the name after is a part of the first name. In this case, Bangsaen is not part of Wonnapha Beach Pier, but acts as an extension only. Therefore, Wonnapha and Bangsaen Beach Pier are marked separately. (not Wonnapha Beach Pier, Bangsaen as in 1 named entity).

But in a case like the following, Chiang Mai in Chiang Mai University is part of the university's name. So, Chiang Mai University is marked as a named entity and Chiang Mai as a place's name is nested inside. (See example in the next case)

If you are not sure if the location's name is part of the name, consider it isn't.

In case of [noun indicating categories + proper noun]

If there is a noun indicating categories + a proper noun, superimpose it. If the name inside is a name that already knows the category to be labeled, mark it as such.

For example, Khon Kaen University (มหาวิทยาลัยขอนแก่น), in a sentence, Khon Kaen University has ordered the suspension of teaching and learning (มหาวิทยาลัยขอนแก่นมีคำสั่งหยุดการเรียนการสอน), will be marked as follows:

มหาวิทยาลัย	ขอนแก่น
<hr/>	
province	
<hr/>	
org:edu	

Mark Khon Kaen (ขอนแก่น) is marked as province.

Mark Khon Kaen University (มหาวิทยาลัยขอนแก่น) is marked as org:edu.

In the case that the name inside is not in another category, mark it in the same category. For example, Nang Ram Beach (หาดนางรำ) shall be marked as follows:

หาด	นางรำ
<hr/>	
loc:others	
<hr/>	
loc:others	

Nang Ram (นางรำ) shall be marked as loc:others.

Nang Ram Beach (หาดนางรำ) shall be marked as loc:others.

If the name inside appears in several categories, e.g. Chulalongkorn (จุฬาลงกรณ์) in Chulalongkorn University (จุฬาลงกรณ์มหาวิทยาลัย). It may be firstname (the name of King Rama 5) or it may be org:edu. The label is based on the meaning that occurs more frequently, which is org:edu in this case.

Chulalongkorn

University

org:edu

org:edu or facility:others depending on the context

This type of structure can be nested in several layers. Label the words in every layer.
For instance, Nopparat Thara Beach Pier (ท่าเรือหาดนพรัตน์ธารา) is marked as follows:

ท่าเรือ

หาด

นพรัตน์ธารา

loc:others

loc:others

facility:others

Nopparat Thara (นพรัตน์ธารา) is marked as loc:others.

Nopparat Thara Beach (หาดนพรัตน์ธารา) is marked as loc:others.

Nopparat Thara Beach Pier (ท่าเรือหาดนพรัตน์ธารา) is marked as facility:others.

In case of [number + unit]

2.1. When numbers or numbered words appear alone and they represent the named entity (time, day, month, year, etc.), label it as such.

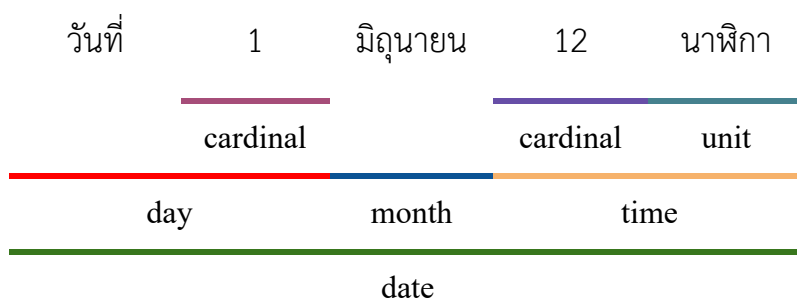
เมื่อ

3.00

time

In “at 3.00”, 3.00 is marked as time.

If there are other words indicating their own category, label the number + unit as that category and label the number appearing alone as cardinal, for example:



1 is marked as cardinal

The 1st (วันที่ 1) is marked as day.

June (มิถุนายน) is marked as month.

The 1st of June (วันที่ 1 มิถุนายน) is marked as date.

12 is marked as cardinal.

o'clock is marked as unit.

12 o'clock (12 นาฬิกา) is marked as time.

Categories of label

Person

Note: If the name is in a foreign language and cannot be separated or you are not sure about the name's part, please mark all of the name as person and do not separately label

person

The name of a person (human) that may or may not exist (fiction).

If the name of other living beings, use animate (MISC category). For example, the names in this category are Pot Phaholyothin (พจน์ พหลโยธิน), Thanathorn

Juangroongruangkit (ธนาธร จึงรุ่งเรืองกิจ), Thaksin Shinawatra (ทักษิณ ชินวัตร), King

Vajiralongkorn (พระบาทสมเด็จพระวชิรเกล้าเจ้าอยู่หัว), etc. This label is used as the biggest tag. The tag includes prefixes, suffixes, ranks, etc., but does not include positions and affiliations.

title

Name prefixes such as

- Mr (นาย)
- Mrs (นาง)
- Miss (ด.ญ. หรือ นางสาว)
- Master (ด.ช.)

This also includes the name prefixes that indicate the titles associated with the royal family, academic ranks, priests' ranks, and ranks in army, police as well as in other government agencies such as

- Khun [in Khun Pum Jensen (คุณพุ่ม เจนเซน)]
- Khun Ying (คุณหญิง)
- Mom Chao (M.C)
- Mom Rajawongse (M.R.)
- Phrakrupalad (พระครูปลัด)
- Chao Khun (เจ้าคุณ)
- Professor (ศาสตราจารย์)
- Dr. (ดร.)

If there are multiple prefixes, mark them separately, **not nest**.

ศาสตราจารย์กิตติคุณ	ดร.	ปราณี	กุลลวณิชย์
title	title	firstname	last
person			

Pranee is marked as firstname.

Kullavanijaya is marked as last.

Prof. and Dr. are marked as title.

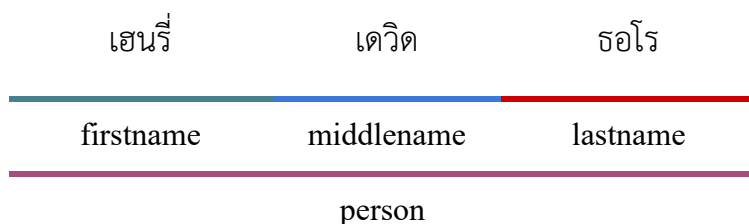
Prof. Dr. Pranee Kullavanijaya is marked as person.

firstname

First names such as Somchai (สมชาย), Sommai (สมหมาย), Somying (สมหญิง), Somsong (สมพร).

middlename

Middle names



Henry is marked as firstname.

David is marked as middlename.

Thoreau is marked as lastname.

Henry David Thoreau is marked as person.

last

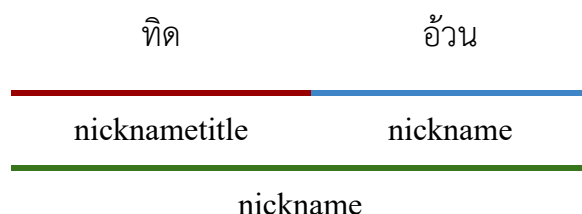
Last names such as Chan-o-cha (จันทโรชา), Shinawatra (ชินวัตร)

nickname

Nicknames such as Tee Lek (ตี้เล็ก), as well as informal nicknames given by others, such as Big Pom (ปิ๊กป้อม).

nicknametitle

Prefixs of a nickname such as Jae (เจ๊), Tee (ตี้), Thit (ทิด).



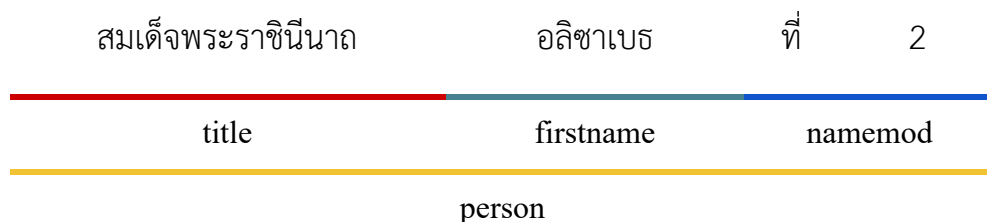
Thit (ทิด) is marked as nicknametitle.

Uan (อ้วน) is marked as nickname.

Thit Uan (ทิด อ้วน) is marked as nickname.

namemod

- Junior, Senior
- This is used for II (ที่ 2) in Queen Elizabeth II as well as V (ที่ 5) in Rama V. If it's the number in a common noun, it doesn't count as namemod.



Queen is marked as title.

Elizabeth is marked as firstname.

II (ที่ 2) is marked as namemod.

Queen Elizabeth II is marked as person.

pseudoname

It refers to religious names such as monks' nicknames, which are as Panyananta Bhikkhu (ปัญญานันทภิกขุ), pen names, handles, and usernames, including Tomyanti (ทมยันตี), J.K. Rowling.

role

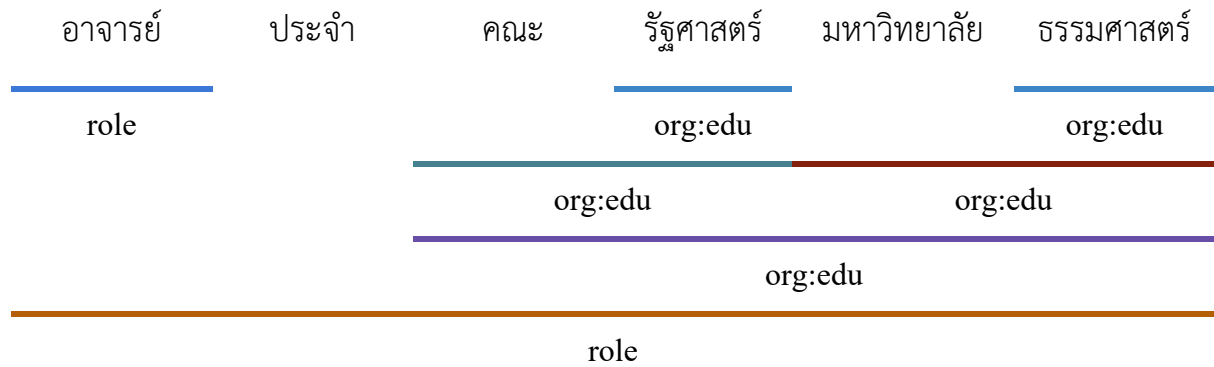
Positions such as Prime Minister, Minister, Executive Committee

For a position that can be a title, such as Commander-in-Chief (ผู้บัญชาการทหารบก), it usually has an abbreviation, such as Commander-in-Chief (ผบ.ทบ.). In this case, look at it's position in the sentence. If it's before the name, it's title; if it's after, it's role.

~~In case of position + affiliation and with many affiliations, mark it only with the affiliation that is the most immediate. For example, a lecturer of the Faculty of Political Science, in a lecturer of the Faculty of Political Science, Thammasat University, is marked as role.~~

*** (Changed from the original guideline to prevent overlapping tags) ***

If you find the case of positions + many affiliations or the affiliations being nested in many layers, mark sub-position as a role and label affiliations separately in different layers. Then, mark the position as a role in the larger layer. However, do not overlap them, as in the example:



A lecturer (อาจารย์) is marked as role.

Political Science (รัฐศาสตร์) is marked as org:edu.

Thammasat (ธรรมศาสตร์) is marked as org:edu.

Faculty of Political Science (คณะรัฐศาสตร์) is marked as org:edu.

Thammasat University (มหาวิทยาลัยธรรมศาสตร์) is marked as org:edu.

Faculty of Political Science, Thammasat University (คณะรัฐศาสตร์

มหาวิทยาลัยธรรมศาสตร์) is marked as org:edu.

Professor of Faculty of Political Science, Thammasat University
(อาจารย์ประจำคณะรัฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์) is marked as role.

Location

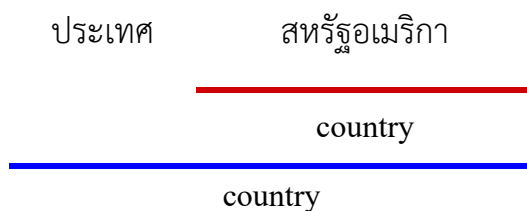
continent

Continents including Asia, Africa, Europe, etc.

- For America, make sure whether it is a country or a continent.
- North-South continents, which are North America and South America, shall not be nested.

country

The names of countries. If there is the word “the country of (ประเทศ)”, also mark it in another layer.



The United States of America (สหรัฐอเมริกา) is marked as country.

When it has the prefix to show status to be a country (ประเทศ), it is also marked as country in another layer.

state

States, in the federal constitutional republic (USA), such as Ohio, Michigan, California, etc.

city

Major cities (that are not provinces, districts, sub-districts, or other categories, mostly in foreign countries), such as New York, Berlin.

province

Provinces, including Bangkok

district

Districts (อำเภอ or เขต) or those other countries

subdistrict

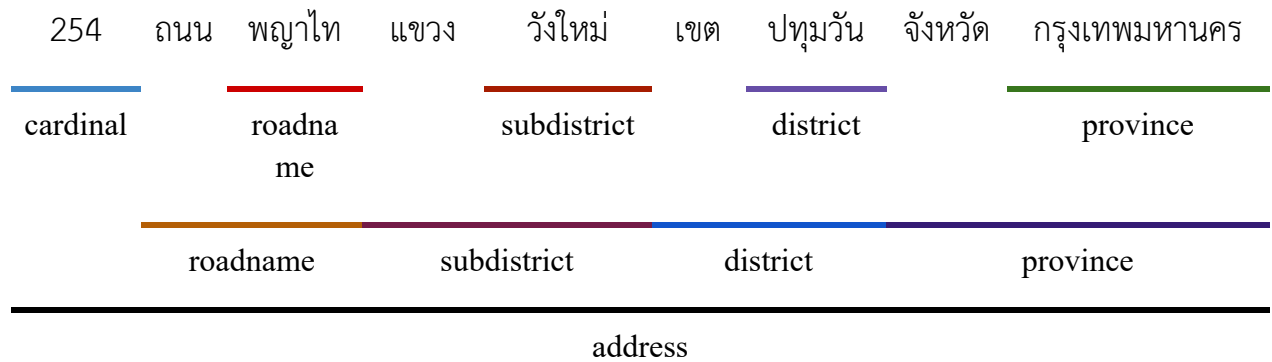
Sub-district (ตำบล or แขวง) or those in other countries

roadname

The names of roads, such as Sukhumvit (สุขุมวิท)

address

Address. Drag the whole address. Then, label the sub-tags as appropriate.



254 is marked as cardinal 1.

Phaya Thai (พญาไทย) is marked as roadname.

Phaya Thai Road (ถนนพญาไทย) is marked as roadname.

Wang Mai (วังใหม่) is marked as subdistrict.

Wang Mai Sub-district (แขวงวังใหม่) is marked as subdistrict.

Pathumwan (ปทุมวัน) is marked as district.

Pathumwan District (เขตปทุมวัน) is marked as district.

Bangkok (กรุงเทพมหานคร and จังหวัดกรุงเทพมหานคร) is marked as province.

All of them are marked as address.

soi

Names of soi or Soi + the number of soi

latitude

Latitude

longitude

Longitude

postcode

Postal codes, such as 10240

ocean

Names of oceans

island

Names of islands

mountain

Names of mountains

river

Names of rivers

space

Names of things in space, such as stars, planets, and satellites

restaurant

Names of restaurants or beverage shops or places that mainly sell food and beverages.

loc:others

Names of other places, areas, and proper nouns that are not in the above category. Also, they are not used for any specific purpose, such as beaches, creeks. If there is any specific purpose, it is considered as facility:others.

Time

date

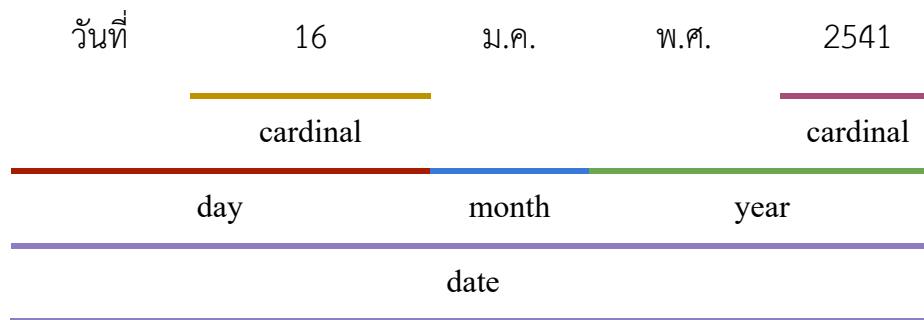
A day, month, and year, for example, April 15, 2018 (วันที่ 15 เมษายน 2561), or only day + month. (month only or month + year is not considered)

If there is a day, month, and year together, do not specify day+month as date again. Mark it in the biggest tag only once. If it contains the words date (วันที่), month (เดือน), year (ปี), etc., label them together as one layer.

Example of a date label:

The date of 16 of January, B.E. 2541 (วันที่ 16 ม.ค. พ.ศ. 2541)

If there is a word date (วันที่) or B.E. (พ.ศ.), A.D. (ค.ศ.), R.E. (ร.ศ.) and the year of (ปี), it shall be labelled as follows:



16 is marked as cardinal.

2541 is marked as cardinal.

The date of 16 is marked as day.

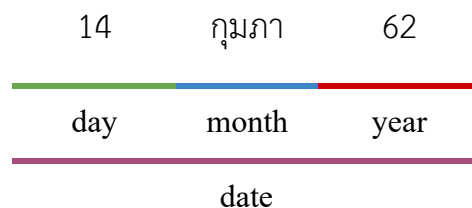
January is marked as month.

B.E. 2541 is marked as year.

The whole message is marked as date.

If not, label it like this example.

14 Feb 62 (14 กุมภาพันธ์ 62)



14 is marked as cardinal.

Feb is marked as month.

62 is marked as year.

The whole message is marked as date.

year

Years such as 2009, B.E. 2541

month

Months

day

Dates (วันที่) or the days in a week (Monday, Tuesday, etc.)

time

The time of day, which is a specific time such as 5 o'clock, 15:00, 8:30 a.m., noon, midnight, 2 p.m. But if it is written as the morning (เช้า), afternoon (บ่าย), evening (เย็น or ค่ำ), it is not considered as time because it covers the wide range of time.

Mark only the time part. For example, at 10:00 AM on Oct 1 (เมื่อเวลา 10:00 น. ของวันที่ 1 ต.ค.), label the time 10.00 AM (เวลา 10:00 น.) and 10.00 AM and 10.00 as cardinal and label Oct 1 separately.

duration

Intervals that do not match other labels, such as 4 minutes, 5 hours, three days, weeks, quarters, or time a - time b/time a to time b, or the age.

periodic

Repetitive intervals, such as every 3 days, every month, monthly, annually (You can consider from context whether it refers to a routine or that it happens repeatedly.)

season

Seasons which refer to summer, spring, excluding the use of the word season (ฤดู) to modify other names. such as the flu season (ฤดูไข้หวัดระบาด), the sales season (ฤดูเซลสินค้า)

rel

Time relative to the present, such as three days ago, tomorrow, yesterday, now (ตอนนี้), at present (ปัจจุบัน), in the past (อดีต), etc.

last (ที่ผ่านมา), coming (ที่กำลังจะมาถึง), and those days (วันนั้น ๆ) are not rel

Organisation

org:corp

Names of companies, including the Board of Directors

org:edu

Names of educational organizations, Student Councils, Alumni Associations, Parents' Associations, Board of Directors in Educational Organizations or any other organization affiliated with educational institutions. This also includes the names of faculties, offices, research units, and academic associations.

org:political

Names of political and economic organizations that are independent of the government but smaller than the government, such as the name of a political party or an NGO

that works in politics. If you come across an organization that aims for freedom and rights or the change of state or institutional policy, also mark it as org:political.

org:religious

Names of religious organizations or religious sects

government

Names of government and organization affiliated with the state (herein including both legislative, executive and judicial organizations), such as the government of China, ministries, courts, the House of Representatives, SAO (สภา.), and agencies or a committee affiliated with these organizations. It also includes state enterprises and public organizations.

army

Names of the army, armed forces (including police forces) or named sub-units

sports-team

Names of sports teams.

media

Names of mass media organizations and news agencies such as BBC, Prachatai, The Standard, Thairath

hotel

Names of hotels

museum

Names of museums

hospital

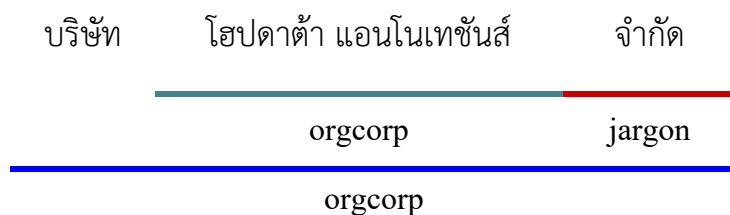
Names of hospitals or health care centers

band

Names of bands

jargon

The extensions of organization's names such as Limited, Public Company Limited



Hope Data Annotations is marked as orgcorp.

TLD is marked as jargon.

Hope Data Annotations Co., TLD. is marked as orgcorp.



Charoen Pokphand Foods is marked as orgcorp.

Public Company Limited is marked as jargon.

Charoen Pokphand Foods Public Company Limited is marked as orgcorp.

stock-exchange

Names of stock market or the name that identifies the stock market or names of stock

index

Names of stock indexes

fund

Names of funds or foundations

org:other

Names of any organization that does not match another label or of organizations at a higher rank than the government, e.g., Labor unions, World Health Organization, UNESCO

NORP

Noun extensions

nationality

Names of nationalities or ethnicities, such as Karen, Shan, which are used as extensions of nouns or noun phrases.

- If it is the same as the name of a country, such as Portugal (โปรตุเกส), label it according to its function in the sentence. If it's an extension of a noun, it's a nationality such as Portuguese (คนโปรตุเกส). If it's a generic noun or preceded by the word the country of (ประเทศ), it is used as country.
- If it is a word that is directly transliterated from a foreign adjective, such as American (อเมริกัน), label only this tag.

religion

Noun extensions that are names of religions, sects, or doctrines.

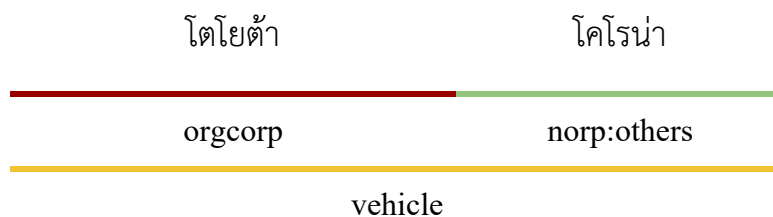
norp:political

Political noun extensions such as liberal, democratic, feminist, slim (สลิม), Seua Daeng (เสื้อแดง), Seua Som (เสื้อส้ม) (which does not refer to any specific group of people). Use this tag when the name is used to describe a party, group, movement, etc., without having to tag the word party (ฝ่าย), group (กลุ่ม), movement (ขบวนการ), etc.

norp:others

Other noun extensions that are named entities (does not fall into other categories). They are also used when there is a modifier in a proper noun. But when they are extracted, those words do not imply any proper noun, for example:

Toyota Corona (as a car) [(รถยนต์) โตโยต้า โคโรน่า]



Toyota is marked as orgcorp.

Corona is marked as norp:others.
Toyota Corona is marked as vehicle.

The word Corona is ambiguous because it can be a car model, virus, and beer. In this case, it is considered as norp:others.

Also, mark the extensions of any person or group of people that are not in another category as such, for example, labelling Isan in Isan people as norp:others.

Facility

airport

Names of airports

port

Names of ports in all sizes, both in deepwater and shallow water areas

bridge

Names of bridges

building

Names of buildings

stadium

Names of stadiums

station

Names of stations and terminals, such as Mo Chit (หมอชิต), Phrom Phong (that refers to the BTS station)

facility:other

Names of buildings, houses, places, constructions or areas that are clearly restricted to be built or set up for any purpose, such as Chinese cemetery (ฮวงซุ้ย), police stations (สน. or สภ.), shopping malls, religious sites, dams, etc.

Event

sports-event

Names of sports events

concert

Names of concerts

natural-disaster

Names of natural disaster incidents, including occasional incidents such as the 1995 Bangkok floods (เหตุน้ำท่วมกรุงเทพฯ พ.ศ. 2538)

war

Names of wars

event:other

Names of festivals, events, either important or non-important, the important days with specific names that do not fall in the above categories. It also includes both historical events such as the events of October 14 (เหตุการณ์ 14 ตุลา) and important days or moments that are annually named such as Visakha Bucha Day, Christmas, New Year's Day, the anniversary of the death of Sueb Naksathien (วันครบรอบการตายของสืบ นาคะเสถียร).

WOA

book

Names of books, articles, print media that is equivalent to a book

film

Names of films

song

Names of songs

tv-show

Names of TV shows or radio programs

woa

Names of other artworks or creative writings

MISC

animate

Names of non-human beings (that exist or made up) such as Khan Kluay (ก้านกล้วย) and Garfield (การ์ฟิลด์)

game

Names of games, such as Mario, Pokemon Red

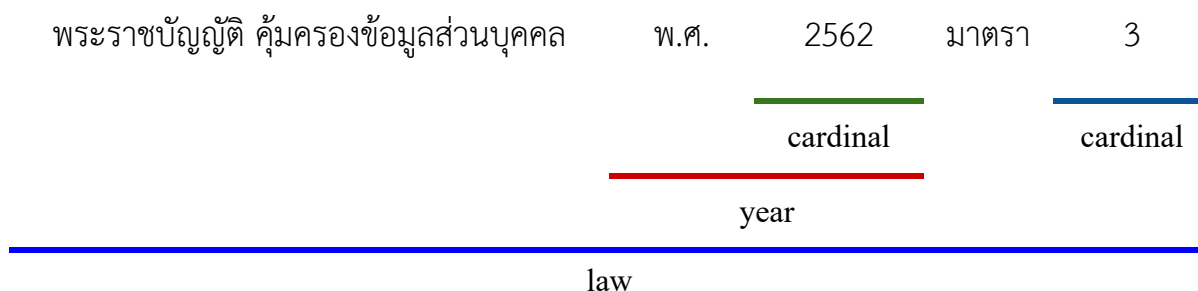
language

Names of languages

law

Names of the legal articles, such as the Personal Data Protection Act B.E. 2562, Section 3, paragraph 2. Mark the name of the law and the year together and label the year nested.

Sections and paragraphs are not part of the law's name.



2562 is marked as cardinal.

B.E. 2562 is marked as year.

3 is marked as cardinal.

The Personal Data Protection Act B.E. 2562 is marked as law.

award

Names of awards

electronics

Names of electronic devices, including softwares, programs, applications (excluding the names of websites or pages)

weapon

Names of the weapons with specific names, such as M-sixteen, (excluding guns, swords, halberds, grenades).

Example: In .38 revolver (ปืนพกสั้น .38), mark .38 and .38 revolver are marked as weapon

vehicle

Names of vehicles, makes or models, e.g., Honda and Titanic

disease

Names of diseases, viruses or the bacterias that cause disease

god

Names of gods, deities, God, etc., such as Poseidon, Yahweh, Allah, Vishnu, and Kuan Yin (พระแม่กวนอิม)

sciname

Scientific names of plants, animals, and chemicals, such as Homo erectus and sodium chloride

food:ingredient

Ingredients with specific names, such as Champagne, Kurobuta, and Gruyere Cheese

product:food

Names of menus, products or food brands, such as Lay, Meiji

product:drug

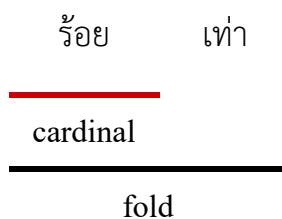
Trade names or generic names of medicinal products such as Paracetamol, Zara, Tylenol

animal-species

Names of species, non-common species or names of any particular animal such as catfish, white shark, black chicken (if it is the animal's name given by people, such as Khan Kluay (ก้านกล้วย), it is considered as animate)

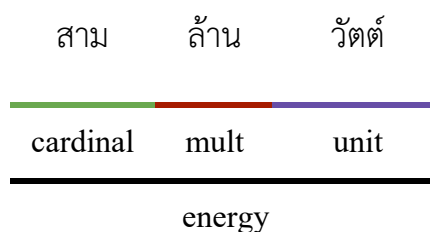
NUM

Example:



Hundred (ร้อย) is marked as cardinal.

Hundred fold is marked as fold.

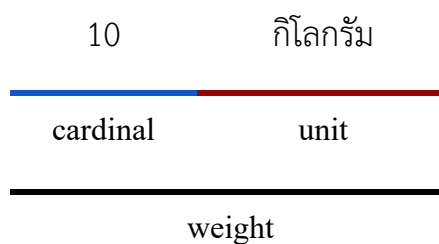


Three is marked as cardinal.

Million is marked as mult.

Watt is marked unit.

Three Million Watt is marked as energy.



10 is marked as cardinal.

Kilogram is marked as unit.

10 Kilograms is marked as weight.

cardinal

- Numerals, such as one, two, three, including 1 2 3 4 (in this case including numbers with a decimal point).
- Mark only numbers that refer to the amount. (do not include id, url or websites)

- Apply only to counting numbers, which means it excludes numbers that are dates, months, years, or time when they stand without time's indicators. (please see the time section for an example).
- For hundred, thousand, ten thousand, hundred thousand, if it is used to refer to 100, 1000, 10000, 100000, 1000000, it is cardinal.
 - If it's a multiplicative unit, for example, one million, thirty (สามสิบ), it's mult.
 - For twenty, eleven, twelve, thirteen, all are cardinal.
- If you wonder whether the word should be tagged as cardinal or a more specific category, label it in the more specific category.
- The phone number must be labeled as cardinal as well.

mult

Multiplicative numbers such as tens, hundreds, thousands, millions, hundred thousands, dozens.

fold

Multiplication, such as threefold, fivefold

money

An amount of money, such as ten dollars, five hundred yuan

energy

An amount of energy, such as twenty watts, one megawatt

speed

Speed, such as eight miles per second or nine kilometers per hour

distance

Distance, length, width, or any distance between two points, such as ten miles, five inches, one foot, two yards, ten mile (หนึ่งโหล่น), three kilometers, seven light years

weight

Weight, such as 10 pounds, 20 kilograms, 3 tons

quantity

Amount, quantity, or volume, e.g. two liters, one bucket, one cup, two teaspoons, two people, five items (label cardinal and unit nested inside)

percent

Percent or percentage, such as 5 percent

temperature

Temperature, such as ten degrees, one hundred degrees Fahrenheit, thirty degrees Celsius, Kelvin

unit

Classifiers or units used for counting with numbers or numbers (See example above).

Foreign language structure, e.g., 12 sports (12 ชนิดกีฬา), 15 best Thai people (15

สุดยอดคนไทย), is not labeled as unit.