

Annotation Guidelines

This is an instruction manual for annotating named entities and sentiments in a Twitter / Weibo corpus. For clarifications, please contact Perry Lam <perry_lam@mymail.sutd.edu.sg>.

According to the language survey done previously, some of you are chosen to do annotations for Twitter, and some Weibo. Those who did not respond are given English Twitter (note: because the data is from Singapore, some may be a mix of local languages). Please go to Project > Annotation Links and click on your student ID to get your assigned data.

These guidelines apply to both Twitter and Weibo as they have similar content.

1. INTRODUCTION

In this annotation task, we will annotate named entities and sentiment polarity towards each named entity in a Twitter Corpus. We will explain the definition of named entities and sentiment polarity in the following sections as well as the way to identify them.

2. NAMED ENTITY

2.1 Definition

A named entity is a real world object such as person, location, organization, product, etc., that can be denoted with a proper name. It can be abstract or physical. Examples of named entities include Barack Obama, New York, Volkswagen Golf. Named entities can simply be viewed as entity instances (e.g. New York is an instance of a city). For the purpose of this annotation task, we need to annotate the outermost named entity phrases found in the corpus. Being the outermost noun phrase means it is not part of any larger noun phrases. For example, in the text “I went to the Bank of China”, we will tag Bank of China instead of China, since China is part of a larger phrase.

2.2 Identification

A good method to determine if a target phrase is a named entity is to check if it is a name of **person**, **organization**, **location**, or **product**. These are the four categories we select for annotation (in other systems, there may be more, like animal names or events, but we simplify the tagging for our project). Note that, in one Twitter post, there may be several named entities to annotate:

Tesla chief **Elon Musk** did not like **Singapore** when the government taxed its **Model S** car.

Here, **Tesla** is an organization, **Elon Musk** is a person, **Singapore** is a location and **Model S** is a product.

Some posts may not contain any entities, e.g. “feeling happy today”.

2.3 Special cases

- Hashtags are named entities only if they are part of the sentence (not tacked on to the end)

✓ The Final Episode Of **#TheBachelor** Draws Season Long Fans

✗ Just Watched The Final Episode Of My Favourite Series!! **#TheBachelor**

- @-mentions are named entities only if they are not usernames

✓ round 2 (@ **Club Tryst** in **Singapore** w/ @coldsteel_lex)

✗ round 2 (@ **Club Tryst** in **Singapore** w/ @coldsteel_lex)

- Annotate abbreviations and typos if they are entities when expanded / corrected

Boarding **SQ** at **Teminal 2** !!

- Titles are persons, but any name for a group with multiple people is an organization

SUTD Provost is discussing with **Root**

- Titles must be absolute, not relative

LKY is our founding **Prime Minister** ← Prime Minister is an absolute title

LKY is our Godfather ← Godfather is a relative title

- Names for songs, movies, and other works of art are products
- The following will not be considered named entities:
 - Dates ("30 Sep"), times ("4pm"), seasons ("autumn")
 - Addresses ("8 Somapah Road")
 - But subspans may be entities, e.g. "133 **Simei** Street 1"
 - Web trends like #ThrowbackThursday or #FollowFriday
- Don't annotate overlapping spans; if there are overlapping spans, choose the longer one.
- Punctuation should be annotated only if they are part of the entity

Going to **Marks & Spencer**

Linda & James not coming

Goodbye @ **T1, Changi Airport**

Have to go back **HQ 3 Div / Jurong Camp** sianz

- Quotes surrounding an entity are not included in the span.

✓ Reading " **Water at the Roots** "

✗ Reading " **Water at the Roots** "

2.4 Disambiguation

- Event names (like “World Cup”) should be classified as organizations
- Use context and uniqueness to differentiate between organizations and locations
 - I am walking to **SUTD** ← SUTD is a location (context is walking)
 - I am a student in **SUTD** ← SUTD is an organization (context is studying)
 - I am at **Subway** ← Subway is an organization (location is not unique)
 - I am at **Subway Changi City Point** ← Subway Changi City Point is a unique location
 - I am at **Subway** in **Changi City Point** ← Subway is not unique but Changi City Point is
- God is a person

3. SENTIMENT POLARITY

3.1 Definition

Sentiment Polarity is the sentiment expressed on the named entity: positive, neutral or negative.

Note that it is not about how you (the annotator) feel about the entity, it is what the poster is trying to express.

3.2 Examples

- **Stephen Curry** is very popular in **NBA** now
 - Sentiment on **Stephen Curry** is positive because the sentence mentions he is very popular. Sentiment on **NBA** is neutral, because there is no comment on it.
- **Changi Village** food normal what
 - Normal is a neutral word, hence sentiment on Change Village is neutral.
- Same **Repubs** who condemned **Bill Clinton** for unpresidential moral standards, now want **Donald Trump** for **POTUS**: WE CANNOT TRUST THEM
 - Sentiments on **Repubs** and **Donald Trump** are negative
 - Sentiment on **Bill Clinton** is neutral (the Repubs condemned him, not the poster)
 - Sentiment on **POTUS** is neutral (no comment on the president post itself)