

SoccEval Annotation Guideline

I. Introduction

The underlying goal for our project is to evaluate the ability and performance of a soccer player from both objective and subjective descriptions of them and their actions. More specifically, we hope to be able to use these evaluations to summarize the salient attributes of a player, to distinguish between good, mediocre, and poor players, and to use these judgments to compare the media's portrayal of a player to that player's objective statistics.

II. A Brief Summary of the Annotations

The annotation is to find phrases describing the performance of a player in a soccer team. You will be given either match reports or season reviews, and you will try to identify phrases that describe either factual events or opinions linked to a specific player and their performance. From there you will assign time, description and sentiment values to those phrases. You will also link the target player with the performance description in MAE.

III. Structure of the Scheme

Our scheme has five tags -- four substantive tags (Player, Coref, Fact and Opinion), and one link tag (TargetLink).

1. Player : first mention of a unique player

span: the text span that mentions a player by name

playerID: a unique ID needs to be assigned to each unique player

name: optional tag

2. Coref: secondary mentions of a previously mentioned player

span: the text span that refers to the player. This can include pronouns and other noun or noun phrases.

playerID: needs to be the same as the player being referred to

3. Fact

span: the text span that describes an event

type: goal, assist, pass, shot, movement, position on field, substitute out, substitute in, injury, tackle, save, other

time: distant past, last season, current season, last match, present, future

factID: a unique ID needs to be assigned to each different event

4. Opinion

span: the text span that expresses an opinion with attitude

type: figurative language, soccer skill, general attribute, impact on team, growth/decline, accomplishment, other

polarity: positive, negative

time: distant past, last season, current season, last match, present, future

hypothetical: counter-factive, prediction

reported: false, true

5. TargetLink

fact/opinion: tag ID (generated by MAE) of a Fact or Opinion

target: tag ID (generated by MAE) of the relevant Player or Coref

IV. Explanation of Tags

1. Player:

For players, you just need to assign an ID for each unique player. Our task does not include annotation on coaches, sponsors or anyone who is currently not a player. Preprocessing would be made to highlight the names of most players beforehand, so that it would be easier to tag players. There will still be some player names that were not caught during preprocessing; If you are not sure if a name belongs to a player or not, confirm it by searching on google or wikipedia. Make sure he is a soccer player during a specific season.

We recommend that the *playerID* number start from 1. Assign a new ID whenever you notice a new player name.

There is also an optional attribute, which is *name*. The purpose of this attribute is primarily to help you, the annotator, keep track of which player is which. This might be useful in cases where the document mentions a large number of players and the document does not always use a player's full name.

2. Coref:

Any time a player is referred to by something other than his name, it should be tagged as *Coref*. Each *Coref* tag has a *playerID* attribute that should be given the same value as the one given to the player it refers to. Pronoun anaphors have been automatically tagged and must either be given the correct *playerID* or deleted if they refer to an entity other than a player.

While there are many types of non-pronoun anaphors, the most common are those that refer to a player's position (such as the "forward or the goalkeeper") and those that refer to a player's nationality (such as "the Spaniard").

3. Fact

The *Fact* tag is specifically for phrases describing the the actions of a player or events involving the players. These phrases describe objective facts and do not carry a sentiment with them.

For type attribute, we have created categories for the following events:

- *goal*: phrase describes a player scoring a goal
- *assist*: phrase describes that a player helps another player score a goal
- *pass*: phrase describes players pass the ball
- *shot*: phrase describes an attempt to score a goal, including a head-shot and others
- *movement*: phrase describes players moving from one place to another on the field
- *positioning*: phrase describes a player's position on the field
- *substitute out*: phrase describes the team making a substitution; this tag is for the outgoing player
- *substitute in*: phrase describes the team making a substitution; this tag is for the incoming player
- *injury*: phrase describes an injury event
- *tackle/block/defend*: phrase describes a defensive action
- *save*: phrase describes a goalkeeper preventing a goal from being scored
- *foul*: phrase describes a player committing a foul
- *other*: phrase describes any other player-related events in the game

Beside from figuring out the type of the event, you will also need to identify the *time* the event actually happened. It can be: *last match*, *current season* (previous matches during the current season), *last season*, *distant past* (before the previous season), or less likely, *present* or *future*.

In addition, you also need to assign an *factID* for each phrase. Phrases that describe the same event will get the same ID. The conventions for this are the same as for the *playerID*.

4. Opinion

The *Opinion* tag is specifically for sentences where an attitude is expressed about the performance of a player.

For the *type* attribute, so far we have categorized them into the following groups:

- *Soccer skill*: phrase expresses an evaluation or sentiment about a specific soccer-related skill. These will most often be noun phrases.
 - “snazzy dribbles”
 - “ingenious passes”
 - “decisive finishes”
- *Accomplishment*: phrase makes note of some kind of accomplishment. These will usually be full phrases
 - Can be either a concrete statistic (“scored most goals in the Premier League this season”)

- or more vague evaluations (“this is his best season yet”)
- *General attribute*: phrase expresses a more general evaluation of a player than “soccer skill”. These will most often be adjectives, nouns, or noun phrases
 - “trickery”
 - “dynamism and deadliness”
 - “agility”
- *Impact on team*: phrase expresses a sentiment about how a player either helped or hindered the team. These will often be full phrases.
 - “led, masterfully, the title-winning defence for every minute of every match”
 - “tidying up messes by the goofs around him”
 - “added to Van Gaal’s [the coach] problems”
 - “Tottenham fans were almost able to forget about [the player]”
- *Growth/decline*: phrase expresses an opinion about a player’s improvement or decline in performance over a period of time. Usually a full phrase.
 - “endured an iffy spell”
 - “in top form”
- *Figurative language*: phrase represents an overall, pure opinion about the player as a whole, without describing any of the above categories. These phrases often use flowery or metaphorical language. Can be noun phrases or longer phrases.
 - “always a joy to watch”
 - “a Winston Wolfe figure at Manchester United”
 - “delivered performances that ridiculed the suggestion he was past it”
 - “this nifty mischief-maker”

In some cases where a phrase can be tagged with multiple tags, follow the following priority: *soccer skill*, *accomplishment* > *general attribute* > *impact on team* > *growth/decline* > *figurative language*

Sentiment polarity needs to be assigned based on the positive or negative attitude toward the player.

The *time* attribute is the same as that in *Fact*; you need to identify if it belongs to the performance of *last match*, *current season* (previous matches in current season), *last season*, or more *distant past*, or less likely, *present* or *future*.

For the *hypothetical* attribute, there are two cases where it should be used. One is for cases when the phrase makes a prediction about something related to the player. In such cases, tag it as a *prediction*. The other is for phrases speculating on what might have happened had something occurred differently. In such cases, tag them as *counterfactive*.

Examples:

prediction: "Keep Kompany fit until the end of the season and more likely than not they will be lifting the Premier League trophy aloft come May"
counterfactive: "United would certainly not be in the Champions League without him"

The *reported* attribute is tagged as *false* by default. But if the phrase in the span is reported by someone within the story (ie. a secondary source that isn't the writer of the article), make it *true*.

5. Other issues regarding *Fact* and *Opinion* tags

How much text to tag

- Tag as much text as is connected to a specific event.
 - If the *Fact/Opinion* refers to multiple instances of an event (such as "two goals"), be sure to tag the number as well.
- However, if within a phrase, the phrase describes two separate but related events/opinions, split the phrase into the two sections.
- For *general attributes* and *soccer skills*, it is often enough to just tag the noun phrase:
 - i.e. "the [agility]_{OPINION: GENERAL ATTRIBUTE} of their goalkeeper."

Possible forms a Fact or Opinion might take:

- Verb phrase
- Noun phrase (i.e. "penalty", "dynamism and deadliness")
- Adjective (i.e. "absent")
- Possessive (i.e. "His unbelievable passes")

6. TargetLink

TargetLink links a fact or opinion to a player or coref. Therefore it takes two arguments, one for fact ID or opinion ID generated by MAE, the other for player ID or coref ID generated by MAE. In general, Fact/Opinion tags will be linked to Player/Coref tags within the same sentence; in most cases the Player/Coref is subject of the sentence.