**Natural Language Understanding Annotation Schema/Protocol**

-- Dr Xingkun Liu, Dr Verena Rieser, HWU

updated on 01/08/2017

Your task is to annotate the user's utterance with a label reflecting the user’s goal, i.e. what s/he wants to achieve by saying a specific sentence. This label will be used for training a Machine Learning algorithm. The dataset we have contains sentences in different home scenarios, where users interact with their home robot.

More specifically, the annotated data will be used to train the Natural Language Understanding component for Intent classification and Named Entity Recognition. This annotation task requires to modify the pre-defined intents when appropriate and annotate the key information in the utterances/sentences as the named entity. **The existing scenario names, intent names and entity names are listed at the end of this document.**

The data was collected via online user typed-in but please keep in mind that in the real situation the input to the robot will come from the ASR (Automatic Speech Recogniser) when the user speaks to the robot.

The data is in CSV format and columns are separated by semi-colon “;”. The file has following columns (**Columns in Red color below means Do Not change their contents**)

* userid: the user id
* answerid: the answer id
* scenario: scenario name, e.g. alarm, music, etc.
* intent: pre-defined user intent
* status: status of the annotated utterance, whether it is Irrelevant, Modified etc.
* answer\_annotation: your annotated utterance
* notes: your notes if you need to explain something.
* answer\_normalised: pre-normalised utterance
* answer: user's original utterances
* question: originally designed question for users

For example, for a scenario talking about home alarm setting, the content for scenario, intent and answer\_annotation:

alarm | set | set an alarm for nine am

This assumes the user is talking about scenario "alarm" and s/her intent is to set alarm "set" to “nine am” (which is called the entity). The task to annotate/label the entity and modify the intent as necessary. This example should be labelled as:

alarm | set | set an alarm for [time : nine am]

The detailed task requirements as followed below.

**1. About CSV columns to work on**

You are required to work on column: answer\_annotation, status and notes, maybe on scenario and intent for some cases. Please do not change other columns (see above in red color), they are there for your references when you are doing the annotations.

**2 Annotating utterances**

To think about the user sentence (in answer\_annotation) to abstract the key information and label them. Along with the predefined Scenario and Intent, this information will enable the robot to take actions for fulfilling user's goal.

**The format required**: along with the value (e.g. “nine am”), put the label inside square bracket "[" and "]". The value and the label are separated with a colon ":". Spaces around the colon doesn't matter but will be better to put one space there, e.g. [time : nine am], but **there should a space between a word and "[" as well as between "]" and a word.**

If you need to come up with your own label name, it should be in one word which can be a combined word. The label names should not be too specific, they should be shared across utterances or scenarios. e.g. the labels "query\_detail" and "relation" in following examples.

contacts | query | look up for the [query\_detail : residential address] of my [relation : team leader]

contacts | query | please tell me the [query\_detail : living address] of my [relation : uncle].

**3. Correction and/or extension**

**About the intent: scenario and intent column**

There may be some cases where the user answers may not correspond to the defined intents. In these cases you can change the scenario and/or intent column to reflect it.

For example: the question is to ask the user to give sentence examples for querying calendar, so the predefined intent is calendar\_query, i.e. the scenario column is “calendar” and the intent column is “query”, the user may give “is there any email from emilia” which is clearly to query email. So you can change “calendar” to “email”.

Another example, if predefined “alarm” “query” but the utterance is “set my alarm for six am tomorrow” which clearly is to set alarm, so change “query” to “set”.

**Please Do Not delete any line! Please Do Not change the values of “userid” and “answerid”!**

***If you have added new utterance lines Please leave the userid and answerid empty or fill them with “null”.***

You may want to add a note in the “notes” column if you have changed the scenario and/or intent column.

**About the sentence: answer\_annotation column**

The sentences were typed in by users and automatically normalised by scripts, so there may be some obvious typos or incorrect normalisation. please correct them if you notice any. You can put notes in the “notes” column if you want to explain something.

The normalised sentence may have place-holders **<unk>**, please replace them with concrete “values” you can think of, e.g. “john”, “seven thirty pm” or a music name etc. when appropriate. Please refer to the original user answer for user's real intentions.

Sometimes you may need to split a sentence into multiple ones. E.g.

User original answer: "what is the sum/total/square root/etc of \_\_\_\_ (and\_\_\_)"

Normalised answer: "what is the sum total square root etc of <unk> <unk>";

This should be changed (in answer\_annotation column) to

“what is the sum of twenty three and thirty one?”

“what is the total of twenty three and thirty one?”

“what is the square root of twenty four?”

and you can add even more similar sentences if you want. You can insert new rows below the current row of the CSV file for the extra sentences you added, and just fill the columns of scenario, intent, status and/or notes, other columns can be empty. The status for newly added sentence is “ADD” (meaning newly added)

You are encouraged to extend the sentence in answer\_annotation column, e.g.

news | query | iphone ==> news | query | please tell me the iphone news

because the question was to request the user to give examples on how they would ask for the latest news on a topic, but this lazy user just typed in “iphone”!

Some incorrect normalised sentences: e.g. the user answer: “Delete an first event of 22/03/2017” was normalised as “delete an first event of twenty two divided by”, here the date is clearly normalised wrong. **Please also correct** them as “twenty second march two thousand seventeen”

**4. About the status column**

For the status column, you can put:

IRR: means “irrelevant” because the sentence is some kind of garbage sentence like “ad adfa df adsf” or if it is Non-English.

ADD: for a sentence you've added in.

MOD: for a sentence you've modified. This is Optional. You could leave it as empty.

**5. Sentence without any entity**

Some sentences may not contain any entity, so there is no need to annotate them, e.g.

transport | train | book me a train

there is no entity to be labelled in this sentence.

**6. For math scenario**

We would suggest to label the whole math expression as one entity: for examples

“what is four plus five?” ==> “what is [math\_expression : four plus five? ]”, this will enable us to handle more complicated maths like (3 + 4 +5 + 6 …) x (….) / (…) etc.

**7. Please use lower-case**

for all label names. (the post processor will convert them to lower-case anyway).

**8. CSV file saving format**

When save the csv file, please select to Quote Text Filed, so the text field will be quoted like " blahblah.....". and use semi-colon ";" as delimiter though using comma is also ok but main thing is to quote the text.

**Current scenarios covered (21 scenarios)**:

[alarm, audio, audiobook, calendar, cooking, datetime, email, game, general, iot, lists, music, news, podcasts, qa, radio, recommendation, social, takeaway, transport, weather]

**The Intent names (63 intents):**

The intent name used for the intent classification is combined from the scenario column and intent column in the csv file. The existing intent names followed by an example annotated sentence are listed below (in some cases, some short utterances are separated by a bar “|” ):

Intent => Examples

-------- --------------

alarm\_set => wake me up at [time : five am] [refer : this] [date : week]

alarm\_remove => remove my alarm for [date : tomorrow] [timeofday : morning]

alarm\_query => did i set an alarm to [alarm\_type : wake up] in the [timeofday : morning]

audio\_volume\_down => decrease the volume [change\_amount : to ten percent]

audio\_volume\_mute => please be quiet for [time\_range : another hour]

audio\_volume\_up => increase the volume of the [device\_type : left speaker] [change\_amount : by ten]

audio\_volume\_other => show volume settings

=> change volume [change\_amount : to thirty five percent]

audiobook\_play => please [player\_setting : continue playing] [audiobook\_name : odd thomas] by [audiobook\_author : dean koontz]

calendar\_delete => cancel [event\_name : business meeting] on [date : wednesday]

calendar\_query => do i have [event\_name : meeting] with [person\_name : steve] [time : this week]

calendar\_set => schedule a [event\_name : chat] with [person\_name : adam] on [date : thursday] [time : afternoon]

cooking\_recipe => whats the recipe for [food\_type : pasta sauce]

datetime\_convert => it is [time\_start : five forty five am] in [location\_start : liverpool] what [query\_detail : time] is now [location\_stop : new york]

datetime\_query => what's the [query\_detail : time] in [location : australia]

=> what's [query\_detail : date] [date : today]

email\_add\_contact => add [person\_name : carl's] [person\_info : email] to my contact

email\_query => are there [refer : new] emails for me

email\_query\_contact => what's the [person\_info : email address] of [person\_name : silvia]

email\_send\_email => send email to [person\_name : martha] saying [email\_content : last evening was wild]

game\_play => i want to play [game\_name : rock paper scissors] let's play

general\_affirm => yes please | that's correct | you got it

general\_command\_stop => would you stop please | stop it

general\_confusion => please repeat | what did you meant | i want to hear that again

general\_conversation => we had great [conversation\_topic : dinner] [date : today]

general\_dontcare => i do not care | don't really care | I don't mind

general\_greet => how are you | how are you doing | good morning | hey how are you

general\_joke => whats a [joke\_type : silly] joke you know

general\_negate => that's incorrect | you got me wrong | no that's not what i asked | that was wrong

general\_praise => [thank\_you : thanks] that was awesome | very helpful [thank\_you : thanks]

=> [thank\_you : thank] you | that was great | well done | [thank\_you : thank] you that was great

iot\_cleaning => robot vacuum the [house\_place : living room] now | vacuum the [house\_place : kitchen] | robot vacuum the [house\_place : hallway] now

iot\_coffee => can i have an [coffee\_type : espresso] please

iot\_hue\_changecolor => please set the lights to [color\_type : green]

=> set lights [change\_amount : to twenty percent]

=> make the [house\_place : house] lights [color\_type : amber] at [time : six pm]

iot\_hue\_lightdim => dim the lights in the [house\_place : hall]

=> can you turn down the lights [change\_amount : by twenty percent]

=> dim the [house\_place : bedroom] lights [change\_amount : to half]

iot\_hue\_lightoff => turn off the lights in the [house\_place : bathroom]

iot\_hue\_lightup => increase brightness [change\_amount : by twenty percent]

=> brighten the lights on the [house\_place : front porch]

iot\_wemo\_off => turn off my [device\_type : smart plug] [time : in thirty minutes]

iot\_wemo\_on => make the [device\_type : wemo plug socket] turned on

lists\_create\_add => add [list\_entry : apples] to [list\_name : shopping] list

lists\_delete => remove [list\_entry : eggs] for the [list\_name : shopping] list

lists\_query => do i have [list\_entry : banana] in my [list\_name : shopping] list

music\_dislikeness => i don't like this song

=> that song in the [music\_descriptor : background] is annoying

music\_likeness => my favourite music band is [artist : queen]

=> add this song to my [preference : favorite] list

=> i like [music\_genre : rock] music

music\_play => play [song\_name : why do you sing] by [artist : kirk franklin]

=> play [music\_descriptor : old fashioned christmas] music

=> play my [preference : favorite] song

music\_query => [querymusic\_song : what music] is this

=> tell me the [querymusic\_artist : artist] of this song

music\_settings => [player\_setting : repeat] this song [int\_number : three] times

news\_query => tell me latest [media\_type : bbc] news

=> whats happening in [news\_topic : football] [date : today]

=> who is [time : currently] [query\_detail : leading] in the [news\_topic : parliamentary election]

podcasts\_play => [player\_setting : play] [refer : next] [podcast\_name : doctor who] episode

qa\_definition => what is the definition of [definition\_word : forensic]

qa\_maths => what is [math\_expression : two hundred divided by ten]

qa\_query => what is the [query\_detail : financial capital] of [location : canada]

qa\_stock => show me the stock price for [business\_name : google]

radio\_play => play a [radio\_name : pop station] on the [media\_type : radio]

recommendation\_events => what can i do in [location : berlin]

recommendation\_locations => tell me [food\_type : chinese] [business\_type : restaurants] [location : nearby]

recommendation\_movies => tell me movies playing in [location : oceanside] [date : tonight]

social\_post => post a status [post\_content : happy birthday mom] on [media\_type : facebook]

social\_query => may i know what is trending in [media\_type : facebook]

takeaway\_order => place order of [int\_number : one] [food\_type : pizza]

takeaway\_query => i want to know if the [business\_name : olive garden] provides [order\_type : takeaway]

transport\_query => what [query\_detail : time] is the [refer : earliest] [transport\_type : train] to [location\_stop : glasgow]

transport\_taxi => call me a [transport\_type : taxi] to the [location\_stop : airport] at [time : five am]

transport\_ticket => book a [transport\_type : train] tickets for me

transport\_traffic => what is level of traffic at [location : washington]

weather\_query => is it [weather\_descriptor : raining] [location : outside]

=> what is the weather in [location : paris]

=> is it going to [weather\_descriptor : rain] in the [timeofday : evening]

**Entity Names:**

[action\_type, addressee, alarm\_type, animal\_type, app\_name, artist, audiobook\_author, audiobook\_name, business\_name, business\_type, celebrity, change\_amount, city\_name, coffee\_type, color\_type, conversation\_topic, cooking\_type, country\_name, date, date\_start, date\_stop, device\_type, drink\_type, email\_account, email\_address, email\_content, email\_folder, email\_subject, event\_name, food\_type, game\_name, game\_type, general\_frequency, house\_place, ingredient, int\_number, joke\_type, list\_entry, list\_name, location, location\_start, location\_stop, media\_type, month, movie\_name, movie\_type, music\_album, music\_descriptor, music\_genre, news\_topic, object\_name, object\_type, order\_type, person\_info, person\_name, player\_setting, playlist\_name, podcast\_descriptor, podcast\_episode, podcast\_name, post\_content, preference, program\_name, query\_detail, querymusic\_album, querymusic\_artist, querymusic\_band, querymusic\_favoritegenre, querymusic\_favoritesong, querymusic\_from, querymusic\_genre, querymusic\_lyrics, querymusic\_recorded, querymusic\_release, querymusic\_song, querymusic\_writer, querymusic\_yesno, radio\_name, refer, relation, reminder\_date, reminder\_time, song\_name, sport\_type, thank\_you, time, time\_range, time\_start, time\_stop, time\_zone, time\_zone\_start, time\_zone\_stop, timeofday, traffic\_detail, train\_detail, transport\_agency, transport\_descriptor, transport\_name, transport\_type, weather\_descriptor, year]

================= END OF FILE =============================