

[20200529 in KLMS Q&A Board]

Title: [HW4] OfficeHour Material - expected FAQ

Q/A 1.

Q: what is 'tagged' sentences?

A:

tagged == marked

(just some identifying information that shows that the marked sentences are among the 80 sentences)

Q/A 2.

Q: System performance for the 80%?

A:

Case by case learning이 아닌 이상 100% 성능이 나오지 않을 가능성이 큼니다. 그러나, 100%가 나오는 경우도 있습니다. 성능에 있어 두 경우 모두에 대해 감점 요인이 되지는 않습니다.

It is unlikely that 100% performance will be achieved unless it is through "case by case" learning. However, there may be some cases where such perfect performance can be obtained. There is no penalty for performance in either cases.

Q/A 3.

Q: generality necessary?

A:

We will not give minus points if it is not the most general solution, but general solution would be deemed to have better quality. If you want to shoot for the best, then,

Please try to avoid

X and Y pair occurring more than once in your annotation sentences

For example, if one annotated sentence includes a triple “X Action1 Y”, then please try to avoid triples “Y Action1 X” or “X Action2 Y” or “Y Action2 X” in other sentences.

Such duplications can be a cause for the minus points.

Q/A 4.

Q: only the five verbs?

A:

Yes they may be the easiest cases.

To explain in detail about the selection of 5 verbs:

1. activate

AND

2. inhibit

AND

3. Bind

AND

4. (accelerate OR augment OR induce OR stimulate OR require OR up-regulate)

AND

5. (abolish OR block OR down-regulate OR prevent)

For each of the five verbs, find 20 sentences that contain it. So, in total, 100 sentences.

If two or more verbs are included in the same sentence (like inhibit and activate that occur in two examples in the HW4 description), they can be randomly assumed to belong to one of the two verb categories but not both.

=>한 문장 안에 동사 2개 이상이 들어가면 두 카테고리 중에 하나에 임의 배정 가능 단 중복 사용 안됩니다.

Q/A 5.

Q: should we consider the set of 80% as training data? and 20% as test data?

A:

it can be seen as follows

80 = training (annotated, seen by your system)

20 = evaluation (annotated, unseen by your system)

(Unknown to you: test = TA has it)

Q/A 6.

Q: nested structure?

A:

In a <X ACTION Y> triple, the word ACTION should be a verb. The verb should not be in the form of to-infinitive, gerunds, noun (such as inhibition), and past participle phrase.

The named entities X and Y should be noun phrases (without a prepositional phrase) but not a tensed phrase (such as "that S" phrase), gerunds, and to-infinitive.

For example,

"X prevents Y to activate Z"

<X, prevents, Y>: Okay

<X, prevents, Y to activate Z>: No (this is because it can complicate the problem very much, so we advise you to avoid this types of output)

"X prevents the activation of Y by Z"

<X, prevents, the activation>: Okay

<X, prevents, the activation of Y by Z>: No, because the preposition phrase can be very long and it can be very confusing.

"All vasodilators activated K-Cl cotransport in LK SRBCs and HYZ in VSMCs, and this activation was inhibited by calyculin and genistein, two inhibitors of K-Cl cotransport."

<calyculin, inhibit, K-Cl cotransport>: No, the noun form of the verb (inhibitors) will complicate the problem too much.

"X action Y and Z"

please only put <X, action, Y and Z>

“X action Y or Z”

please only put <X, action, Y or Z> rather than <X, action, Y> and <X, action, Z>.

(We will not severely penalize either of the outputs <X, action, Y> and <X, action, Z>, but if you need a goal to shoot for, please shoot for <X, action, Y or Z> as the answer)

“X and Y action Z and W”

please put only <X and Y, action, Z and W> not any other coordination.

“X action1 Y but action2 Z”

two triples <X, action1, Y> and <X, action2, Z> are okay.

“X action Y but Z”

only <X, action, Y> is okay.

Q/A 7.

Q: number limit per year/publisher?

A:

To explain in detail the sentence “starting with the year 2020, limiting up to 30 sentences per year, up to 10 sentences per journal, and up to two sentences per organization, as identified the affiliation of the corresponding author.”,

**** limitation on the number of sentences per year ****

2020 - up to 30 sentences

2019 - up to 30 sentences

2018 - up to 30 sentences

...

**** limitation on the number of sentences per paper (abstract) ****

up to 2 sentences in each paper (abstract part)

**** limitation on the number of sentences per organization (corresponding author) ****

up to 2 sentences (corresponding author: the author listed as 'corresponding author' among the authors, which usually appears in the last position.

**** limitation on the number of sentences per journal (publisher) ****

up to 10 sentences. For example, there may be several papers in the journal 'Nature'. If you select 5 different papers from 'Nature' and draw 2 sentences each, now you have to pick out a sentence from another journal, such as 'Science'. Each sister magazine is considered a different journal.

****년도 당 문장 개수 제한****

2020년 - 최대 30개 문장 까지

2019년 - 최대 30개 문장 까지

2018년 - 최대 30개 문장 까지

...

****1개 논문당 문장 개수 제한****

2개

****1개 기관(교신저자 기관) 별 문장 개수 제한****

2개

교신 저자=저자 중 'corresponding author'라고 표기된 저자 — 저자명 중 가장 마지막에 주로 등장

****1개 퍼블리셔당 문장 개수 제한****

10개 문장

예를 들어 '네이처'에는 여러 논문이 있을 수 있는데 '네이처'에서 5개 다른 논문을 뽑아서 각 2문장씩 뽑았다면 추가 문장은 다른 퍼블리셔 예를 들어 '사이언스'에서 문장을 찾아야 함. 자매지의 경우에는 다른 publisher로 간주함.

If you are not sure with your judgment whether or not your output is correct, please ask us in advance.