Shared Task Information Science data split

In this document, we explain and motivate the decision making process for creating the train-test split that we use for testing the results of our experiments. This is because it allows for transparency and the possibility to look back on earlier decisions, for instance when needed in writing the report.

Selection of templates

In total, there are seven templates in the original training data.[1]

1. I do not like X, I prefer Y   [no hypernym relations possible]
I do not like animals, I prefer pigs. [INVALID]
I do not like pigs, I prefer animals. [INVALID]

2. I like X, except Y   [X is superset of Y]
I like animals, except pigs. [VALID]
I like pigs, except animals. [INVALID]

3. I like X more than Y   [no hypernym relations possible]
I like animals more than pigs. [INVALID]
I like pigs more than animals. [INVALID]
I like jewelry more than jazz. [VALID]

4. I like X, and more specifically Y   [X is superset of Y]
I like animals, and more specifically pigs. [VALID]
I like pigs, and more specifically animals. [INVALID]

5. I like X, an interesting type of Y   [Y is superset of X]
I like animals, an interesting type of pig. [INVALID]
I like pigs, an interesting type of animal. [VALID]

6. I like X, and Y too   [no hypernym rel]
I like animals, and pigs too. [INVALID]
I like pigs, and animals too. [INVALID]

7. I like X, but not Y   [Y cannot be superset of X]
I like animals, but not pigs. [VALID]
I like pigs, but not animals. [INVALID]

---

[1] Verbs and nouns can vary in these templates, we use 'like', 'X' and 'Y' in these examples.

It is our aim to create a test set that measures performance on a variety of templates and nouns that are not seen in the training set.

We find that there are three main template relations possible:
1. [X is superset of Y]
2. [no hypernym rel]
3. [Y cannot be superset of X]

We want these to be present in the test set. Thus, we select the following templates for it:
1. **I like X, and more specifically Y.**
2. **I like X, and Y too.**
3. **I like X, but not Y.**

Additionally, we want there to be nouns in the training set that do not occur in the training set. Our way to realize this is to filter all sentences that contain the word 'use' and 'met'. The first is only used for **materials** and the second is only used for **people**. However, we noticed that these are also used with other verbs. For instance, we see that these nouns are also used with different verbs, for instance:

   2888, "I do not like plastic , I prefer glass . "

So, we use 'use' and 'met' only to extract nouns and then filter out these nouns. As a result, we also make sure that the test set contains some examples that were previously seen in the training set, as we expect this is also the case for the official test set. In summary, the data split looks as follows.

TRAIN	TEST
Templates: I do not like X, I prefer Y. I like X, except Y. I like X more than Y. I like X, an interesting type of Y. **Nouns:** All except materials and persons	**Templates:** I like X, and more specifically Y. I like X, and Y too. I like X, but not Y. All templates with nouns and materials. **Nouns:** Materials Persons