Precision, Recall, F1:
Interpretation of true positives, false positives and false negatives for the geoparsing coursework
[Updated with additional clarifications on 28th December]

For NER tasks:
True Positives
We count a true positive when the classifier identifies a toponym correctly, I.e. when the detected toponym matches the ground truth toponym.
TP will then be the total number of toponyms that were correctly identified.

True Positives
We count a false positive when the classifier determines that a word in the caption is a toponym when it does not match the ground truth toponym for that caption.

Note that if in a caption the classifier does not detect any toponyms there will be no true positives or false positives for that caption.

False Negatives
The false negatives for the NER task is the number of toponyms that the classifier failed to identify correctly, taking account of all 12 captions.

For an individual caption if the classifier does not find a true positive then for that caption, count a false negative.

Not that if the classifier identified two entities as a toponym and one of them matched the ground truth, there would be one true positive and one false positive, but there would NOT be a false negative for that caption.

In this coursework there is a total of 12 true positives. Your number of false negatives will be equal to 12 minus the number of true positives. Thus, if you have 12 true positives then you will have no false negatives.

In summary for the NER tasks:

> If the classifier identifies an entity as a toponym and that entity matches the ground truth then count a true positive.
>
> If the classifier identifies an entity as a toponym and that entity does not match the ground truth then count a false positive.
>
> If for a caption the classifier does not find any entity that matches the ground truth, then count a false negative.
> Also note that in this coursework the number of false negatives equals 12 minus the number of true positives.

As a consequence of the above three rules for the NER tasks:
For an individual caption it is possible to have both a true positive and a false positive.
For an individual caption if there is a true positive then there is not a false negative.
For an individual caption if there is a false positive and there is no true positive then there is a false negative.

For the geocoding tasks:

For each caption, only attempt to disambiguate a toponym that matches the ground truth toponym.

True Positives
A true positive is counted when the classifier correctly disambiguates a toponym, i.e. the predicted toponym coordinates are within the tolerance distance of the guide coordinates for that toponym.

False Positives
A false positive is counted when the classifier predicts the wrong instance of the toponym to be disambiguated. That instance will not be within the tolerance distance of the guide coordinates for that toponym.

Note that if no prediction of a disambiguated toponym is made for a particular caption then there will be no true positive or false positive counted for that caption.
That could occur when there was no correctly identified toponym for the caption and hence nothing to disambiguate.

False negatives
A false negative is counted when the classifier fails to identify a correct instance of the toponym in a caption.
In this coursework this could correspond to a situation where the classifier generated a false positive (selected the wrong instance) or where no prediction was made by the classifier.

Note that in this coursework, when computing ***recall*** for the geocoding tasks the denominator of (TP + FN) should be equal to the total number of toponyms to be disambiguated, i.e. 12.
Your number of false negatives will be equal to 12 minus the number of true positives. Thus, if you have 12 true positives then you will have no false negatives.

In summary for the geocoding tasks:

For a toponym that matches the ground truth toponym, if the predicted toponym coordinates are within the tolerance distance of the guide coordinates for that toponym, then count a true positive.

For a toponym that matches the ground truth toponym, if the predicted toponym coordinates are NOT within the tolerance distance of the guide coordinates for that toponym, then count a false positive.

If the classifier finds a false positive then also count a false negative. Note that this rule applies here on the assumption that the classifier only selects a *single instance of the toponym as the disambiguated toponym* (as instructed in Questions 3 and 4).
Thus in Question 3 if the toponym selected by the geocoder is not within the tolerance distance count both a false positive and a false negative.
In Question 4 your selected toponym instance should be the one that is closest to the guide coordinates, but if that candidate toponym is not within the tolerance distance then count both a false positive and a false negative.
In both Questions 3 and 4, if the classifier fails to predict a disambiguated toponym for a particular caption, because the NER task did not find an entity that matched the ground truth toponym, then count a false negative.

As a consequence of the above three rules *for the geocoding tasks*:
For an individual caption it is NOT possible to have both a true positive and a false positive.
For an individual caption if there is a true positive then there is not a false negative.
For an individual caption if there is a false positive then there is also a false negative.

Please note that the above interpretations of precision and recall for the tasks of toponym recognition and toponym resolution are similar to the definitions used in the paper "D. Weissenbacher, A. Magge, K. O'Connor, M. Scotch, G. Gonzalez-Hernandez (2019). SemEval-2019 Task 12: Toponym Resolution in Scientific Papers" Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019), pages 907–916

Example
Note that in the following the ground truth and predicted toponym could refer either to the case of detecting toponyms with NER or to geocoding toponyms where in the latter case the predicted toponym is the predicted disambiguated instance of a toponym.

Ground Truth Toponym	Predicted Toponym(s)	TP	FP	FN
A	A	1		
B	X		1	1
C	C	1		
D	D	1		
E	E	1		
F				1
G				1
H	H, I	1	1	

Note that for the geocoding tasks there can only ever be one predicted (disambiguated) toponym.
TP = 5
FP = 2
FN = 3

Precision is the proportion of all predictions that are correct.
Here 7 predictions were made of which 5 were correct.

Precision = TP / (TP + FP) = 5 / (5 + 2) = 5/7 = 0.71

Recall is the proportion of instances of the ground truth class in the whole dataset that were predicted correctly.
In the table above there are 8 ground truth instances of the target toponym, of which 5 have been predicted correctly.

Recall = TP / (TP + FN) = 5 / (5 + 3) = 5/8 = 0.625

F1 is the harmonic mean of precision and recall – a mix of the two.

F1 = 2 * Precision * Recall / (Precision + Recall) = 2 * 0.71 * 0.625 / (0.71 + 0.625) = 0.66

Note that in the case of the NER tasks, it is possible that your classifier might not make any false predictions, in which case the total FP value then would be 0 and the precision would be 1.