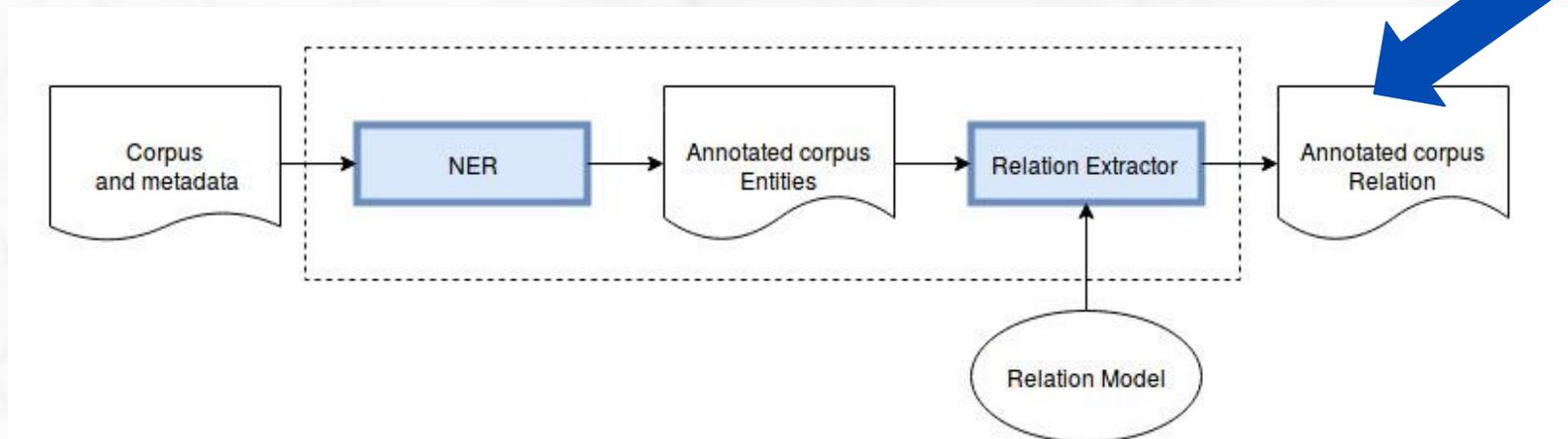


# **Text-mining methods used for information extraction in plant scientific papers**

## **6. Information extraction evaluation**

# Evaluation of the automatically annotated corpus



# How to assess the quality of predictions?

One of the simplest methods is a comparison:

- what a human would have done when annotating by hand (gold annotations)
- and predictions of information extraction systems

The principle is to take a sample of texts, and to calculate the difference between the two: the human, and the artificial intelligence system

We usually used indices such as are recall, precision and F-measure.

# Evaluation indices

Recall: number of correct entities / relations found relative to the number of correct entities / relations **existing** in total.

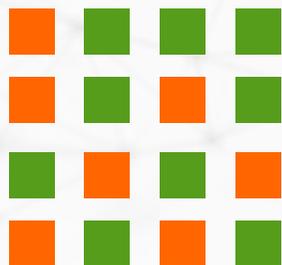
Precision: number of correct entities / relations found relative to the number of entities / relations **found** in total.

F-measure: harmonic mean: combines precision and recall

# Evaluation indices

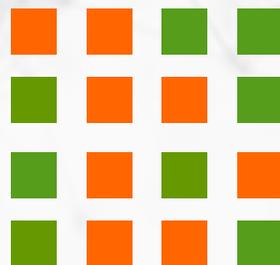
## Example on the correct identification of a Tissue-type entity :

Manual annotation :

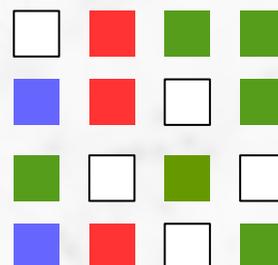


Green : Entity tagged as "Tissue"  
Orange : Entity poorly or not annotated

Automatic annotation :



Comparison: we look at what the system brought back and what it should have brought back (Note: we do not look at what it was right not to bring back)



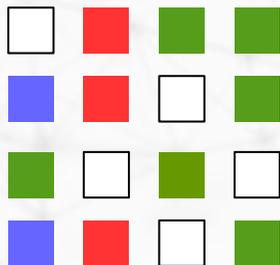
Red: Entity not returned (but should have been): false negative  
Blue: entity that was mistakenly returned: false positive  
Green: entity that has been correctly returned: true positive  
White: entity that has not been returned (and rightly so): true negative

# Evaluation indices

Recall:  $\frac{\text{True Positive}}{(\text{True Positive} + \text{False Negatives})}$  (Correctly found)  
 (Total to find)

Precision:  $\frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$  (Correctly found)  
 (Total recovered)

Comparison:



Recall :  $\frac{6}{(6+3)}$   $\frac{6}{9}$

Precision :  $\frac{6}{(6+2)}$   $\frac{6}{8}$

Red: Entity not returned (but should have been): false negative

Blue: entity that was mistakenly returned: false positive

Green: entity that has been correctly returned: true positive

White: entity that has not been returned (and rightly so): true negative