



Evaluation of Named Entity Recognition

The extension of Precision, Recall, and the F measure to sequences



The Named Entity Recognition Task

Task: Predict entities in a text

Foreign **ORG**

Ministry **ORG**

spokesman **O**

Shen **PER**

Guofang **PER**

told **O**

Reuters **ORG**

:



Standard
evaluation
is per entity,
not per token



Precision/Recall/F1 for IE/NER

- Recall and precision are straightforward for tasks like IR and text categorization, where there is only one grain size (documents)
- The measure behaves a bit funnily for IE/NER when there are *boundary errors* (which are *common*):
 - First Bank of Chicago announced earnings ...
- This counts as both a fp and a fn
- Selecting *nothing* would have been better
- Some other metrics (e.g., MUC scorer) give partial credit (according to complex rules)



Evaluation of Named Entity Recognition

The extension of Precision, Recall, and the F measure to sequences