

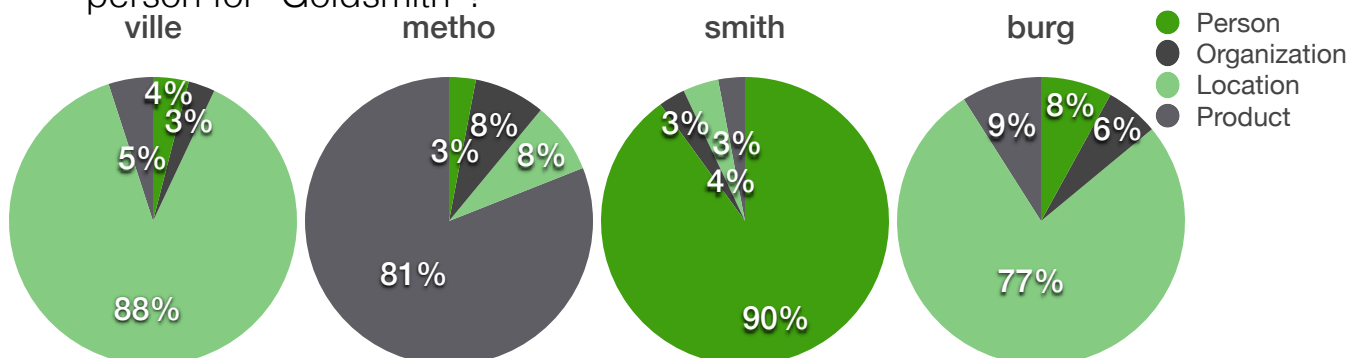
5.1 RAT

1. Compute precision, recall, and F_1 score for the following guess of an NER system.

	the	University	of	Vermont	College	of	Medicine
GOLD	B-ORG	I-ORG	I-ORG	I-ORG	I-ORG	I-ORG	I-ORG
GUESS	O	B-ORG	O	B-PERSON	I-PERSON	O	B-GPE

2. The IOB format categorizes tagged tokens as I, O and B. Why are three tags necessary? What problem would be caused if we used I and O tags exclusively? Give an example.

3. Which of the following is a good feature to use to predict the class person for “Goldsmith”?



4. Say we are using IOB encoding with nine named entity types (reproduced below). If we are building a first order sequence model with class sequence features of previous class and current class over all possible classes in our encoding scheme, how many of these class sequence features will the model have? Don't forget to include the 'O'.

NE Type	Examples
ORGANIZATION	Georgia-Pacific Corp., WHO
PERSON	Eddy Bonte, President Obama
LOCATION	Murray River, Mount Everest
DATE	June, 2008-06-29
TIME	two fifty a m, 1:30 p.m.
MONEY	175 million Canadian Dollars, GBP 10.40
PERCENT	twenty pct, 18.75 %
FACILITY	Washington Monument, Stonehenge
GPE	South East Asia, Midlothian