

Distribution of three topic categories in Reuters-21578 data set – "ModLewis Split"

	Test Set	6188	Train Set	13625	Total	19813
earn	1088	17.6%	2817	20.7%	3905	19.7%
acq	719	11.6%	1615	11.9%	2334	11.8%
money-fx	180	2.9%	519	3.8%	699	3.5%

Combinations

acq & money-fx	1
earn & money-fx	0
acq & earn	16

Note

The "ModLewis" split assigns documents from April 7, 1987 and before to the training set, and documents from April 8, 1987 and after to the test set.