

# Rocchio in Practice

- Negative (non-relevant) examples are not very important (why?)
- Often truncate the vector (i.e., consider only a small number of words that have highest weights in the centroid vector) (efficiency concern)
- Avoid “over-fitting” (keep relatively high weight on the original query weights) (why?)
- Can be used for relevance feedback and pseudo feedback ( $\beta$  should be set to a larger value for relevance feedback than for pseudo feedback)
- Usually robust and effective