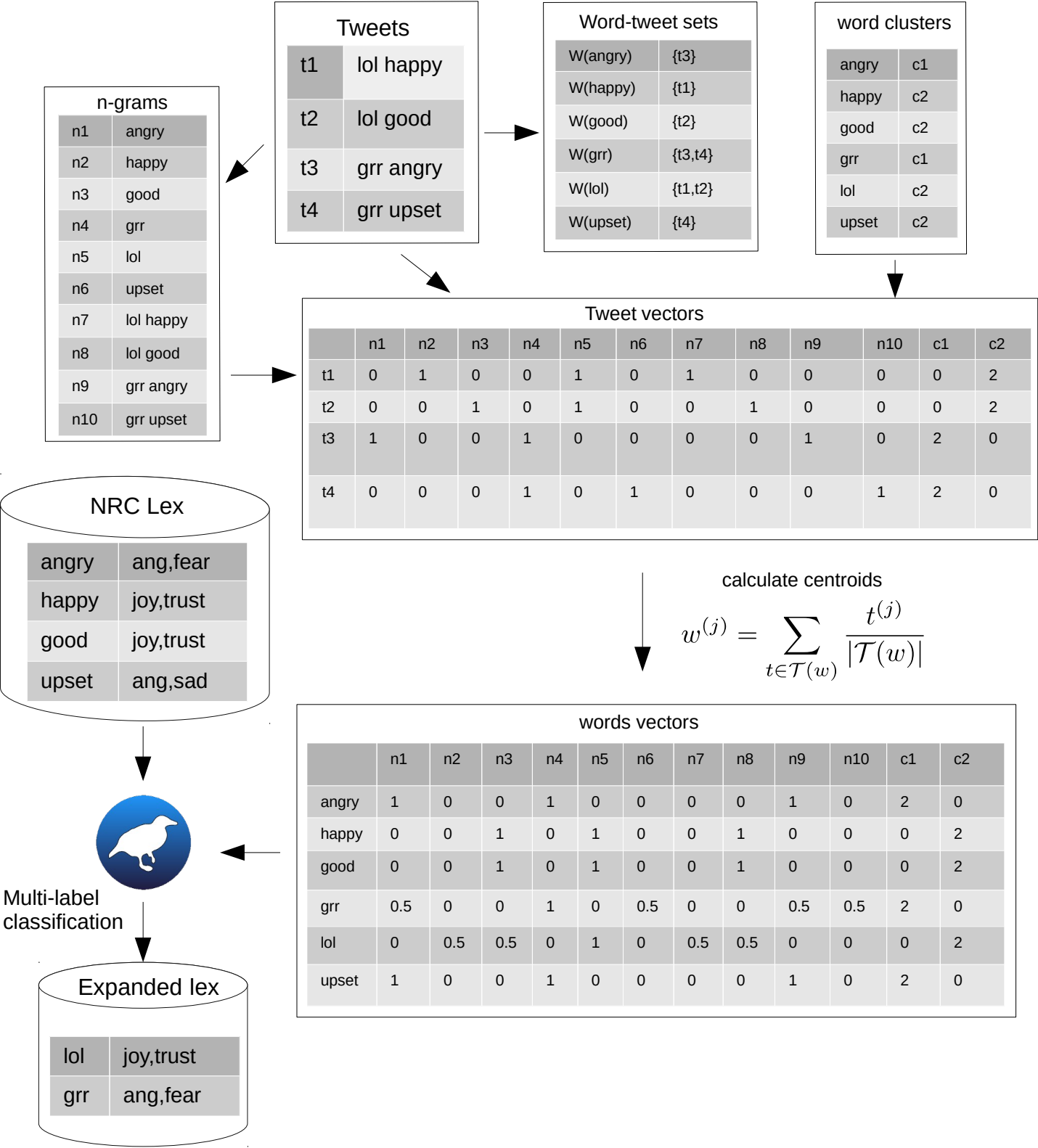
Tweets
t1 lol happy
t2 lol good
t3 grr angry
t4 grr upset
Word-tweet sets
W(angry) {t3}
W(happy) {t1}
W(good) {t2}
W(grr) {t3,t4}
W(lol) {t1,t2}
W(upset) {t4}
word clusters
angry c1
happy c2
good c2
grr c1
lol c2
upset c2
n-grams
n1 angry
n2 happy
n3 good
n4 grr
n5 lol
n6 upset
n7 lol happy
n8 lol good
n9 grr angry
n10 grr upset
Tweet vectors
n1 n2 n3 n4 n5 n6 n7 n8 n9 n10 c1 c2
t1 0 1 0 0 1 0 1 0 0 0 0 2
t2 0 0 1 0 1 0 0 1 0 0 0 2
t3 1 0 0 1 0 0 0 0 1 0 2 0
t4 0 0 0 1 0 1 0 0 0 1 2 0
NRC Lex
angry ang,fear
happy joy,trust
good joy,trust
upset ang,sad
calculate centroids
$w^{(j)} = \sum_{t \in \mathcal{T}(w)} \frac{t^{(j)}}{|\mathcal{T}(w)|}$
words vectors
n1 n2 n3 n4 n5 n6 n7 n8 n9 n10 c1 c2
angry 1 0 0 1 0 0 0 0 1 0 2 0
happy 0 0 1 0 1 0 0 1 0 0 0 2
good 0 0 1 0 1 0 0 1 0 0 0 2
grr 0.5 0 0 1 0 0.5 0 0 0.5 0.5 2 0
lol 0 0.5 0.5 0 1 0 0.5 0.5 0 0 0 2
upset 1 0 0 1 0 0 0 0 1 0 2 0
Multi-label classification
Expanded lex
lol joy,trust
grr ang,fear