

**Word embedding
model generated
from publications**



Cut numbers,
cut adjectives,
trim :.,;- etc.

**(1)
NLP
filtering**

**(2)
Initial bootstrapping
and labeling**

Distance
UBS



Random
UBS
Distance
UBS

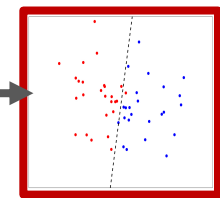
**(3)
Sampling
strategies**

**(4)
Expert and untrained
crowd labeling**

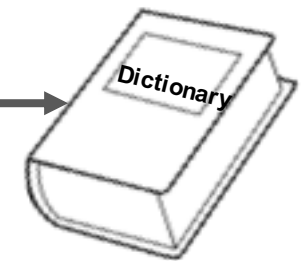


Labeled
entity
referents

**(5)
Candidate
classification**



**Digital
dictionary**



Active learning

