

**Word embedding
model generated
from publications**



**(1)
NLP
filtering**

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

**(2)
Initial labeling**

Distance
UBS



**(3)
Sampling
strategies**

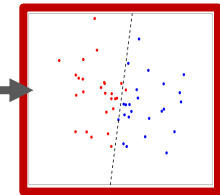
Random
UBS
Distance
UBS

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

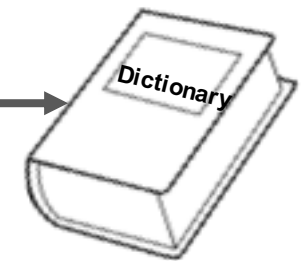


Labeled
entity
referents

**(5)
Candidate
discrimination**



**Digital
dictionary**



Active learning

