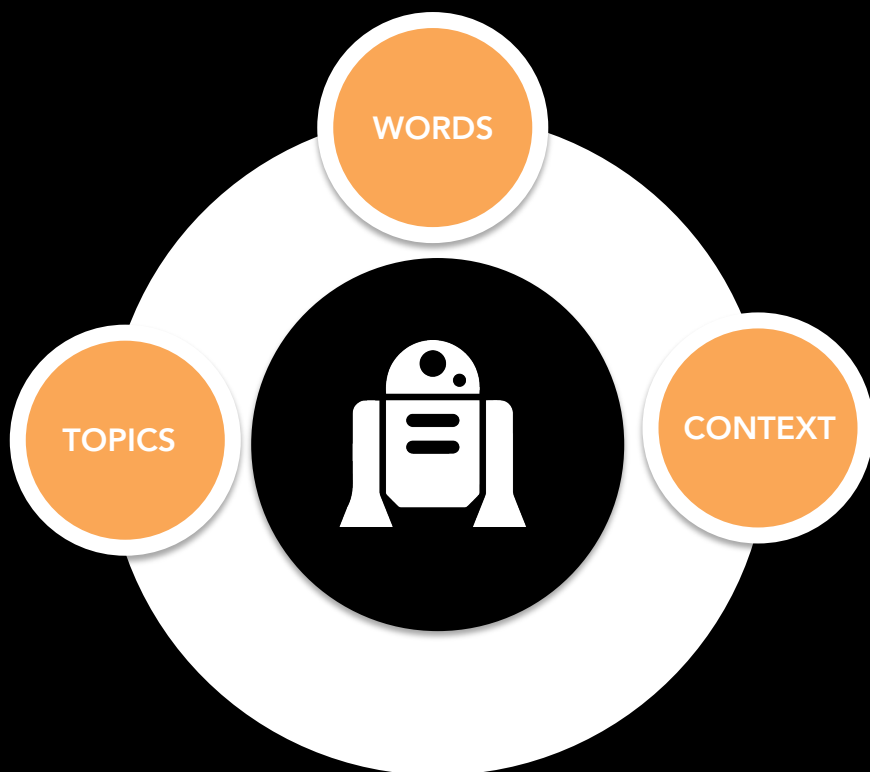


## MODEL



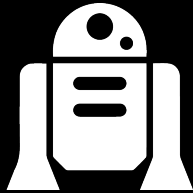
## PERCEPTION, NEUROLINGUISTICS, ETC.





WORDS

TOPICS



CONTEXT



BODY OF KNOWLEDGE

WORDS

DIMENSION

GESTALT

TOPICS

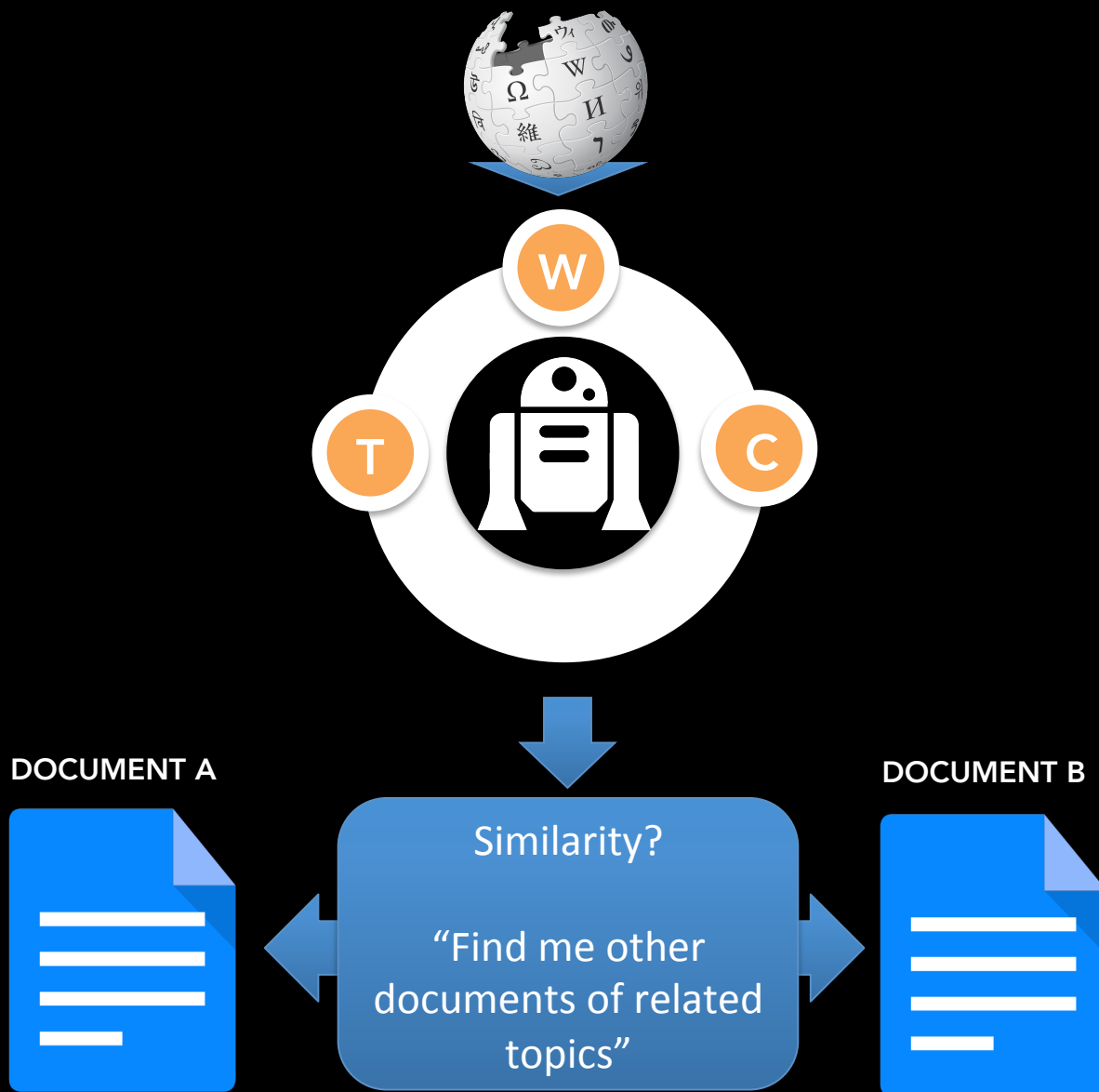


CONTEXT

FEATURES

MEANING

THRESHOLD



Model	Advantage	Disadvantage
Latent Semantic Indexing	-maps terms and documents to help differentiate against polysemy, synonymy	-lacks solid probabilistic foundation (i.e. machine learning)
Latent Dirichlet Allocation	-first fully probabilistic model for text clustering	-maps terms and documents to help differentiate against polysemy, synonymy
Word2Vec	-does not use co-occurrence models; instead uses surroundings words to derive meaning	<b>Hard to say until we explore (implementation may be difficult and error prone)</b>