:::::::::::: Wikitheoria data set ::::::::::::

The Wikitheoria data set consists of ~1000 English sentence pairs, built starting from existing book glossaries.
Each sentence pair is annotated for relatedness in meaning between the two elements.

The Wikitheoria data set is used in Wikitheoria Project - Task: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness

File Structure: tab-separated text file

Fields:

- term ID

- term

- sentence A

- sentence B

- semantic relatedness score (on a 1-5 continuous scale)
    - 1 totally unrelated
    - 5 totally related

Some Sentence Pair Scoring Samples:

Sentence A: The dog is snapping at some droplets of water
Sentence B: A girl in a band is playing an instrument
Relatedness score: 1

Sentence A: A dog is jumping into the air in the country
Sentence B: A dog is jumping in the shade of some trees
Relatedness Score: 3.7

Sentence A: A surfer is surfing a big wave
Sentence B: A surfer is surfing a huge wave
Relatedness Score: 4.9

Sentence A: A man in a jersey which is black is standing in a gym
Sentence B: A man in a black jersey is standing in a gym
Relatedness Score: 4.9

Sentence A: Two dogs are playing in a forest
Sentence B: A brown dog is running after a black dog on a rocky shore
Relatedness Score: 2.3

Sentence A: A person is eating in a restaurant
Sentence B: A person is making a bed
Relatedness Score: 1.2

Sentence A: A man is buttering a piece of bread
Sentence B: A woman is buttering a piece of bread
Relatedness Score: 3.5

Sentence A: A person is dancing
Sentence B: A woman is exercising
Relatedness Score: 3

Sentence A: A woman is slicing a fish
Sentence B: There is no woman slicing a fish
Relatedness Score: 2.7