Material for the Semantic Collocations task

Task   Decide whether the given word pair forms a semantic collocation

Objects   Word pairs (identified by two lemmas L1 and L2)

Target class   YES / NO

Basic statistics (preprocessed data)

attribute	type	definition
L1	string	lemma of the first word
L2	string	lemma of the second word
A	integer	frequency of the bigram (L1, L2)
B	integer	frequency of all bigrams of the type (L1, ¬L2)
C	integer	frequency of all bigrams of the type (¬L1, L2)
D	integer	frequency of all bigrams of the type (¬L1, ¬L2)

Feature vector

feature	type	definition
A1	continuous	Pointwise mutual information
A2		First Kulczynski coefficient
A3		Unigram subtuples measure
A4		Ngram word coocurrence
A5		Reverse confusion probability
A6		Reverse cross entropy
A7		Right context phrasal entropy
A8		Log frequency biased mutual dependency
A9		Cosine context similarity in boolean vector space
A10		Dice context similarity in tf.idf vector space
A11	categorical	POS1:POS2

Legend

L1	lemma of the first word
L2	lemma of the second word
¬L1	a lemma different from L1
¬L2	a lemma different from L2
POS1	part of speech of L1
POS2	part of speech of L2