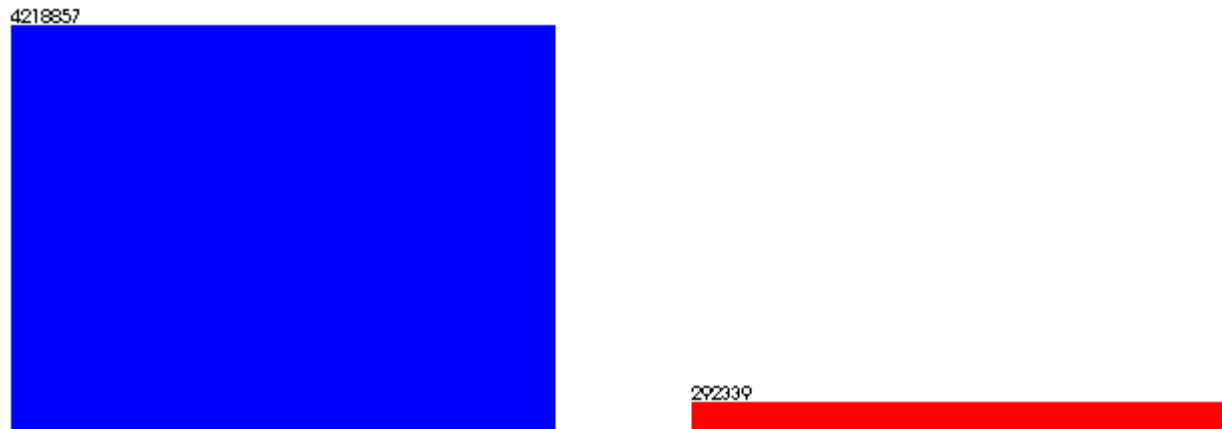# Morphological disambiguation by using machine learning methods

Josef Bušta

12. 11. 2012

Input examples (1)

- Czes (465 102 710)
- Context: word1 word2 word3 class word4 word5 word6;
  class $\in$ {k3c4, k7c7}
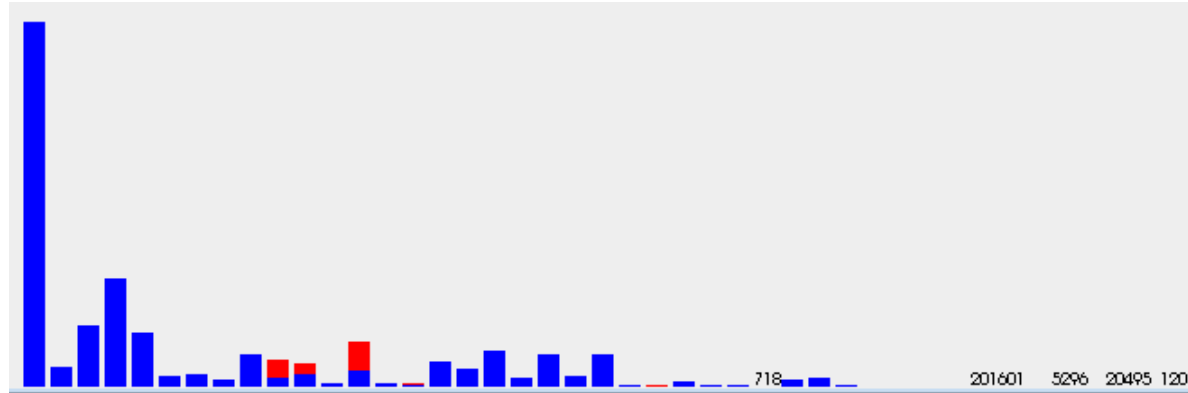
4218857

292339

Baseline: 93,5197
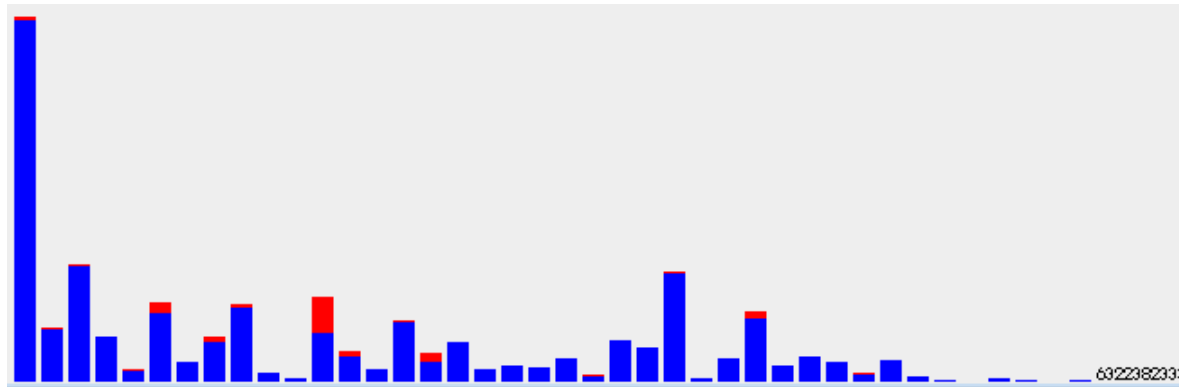Total: 4 511 196
k3c4: 4 218 857
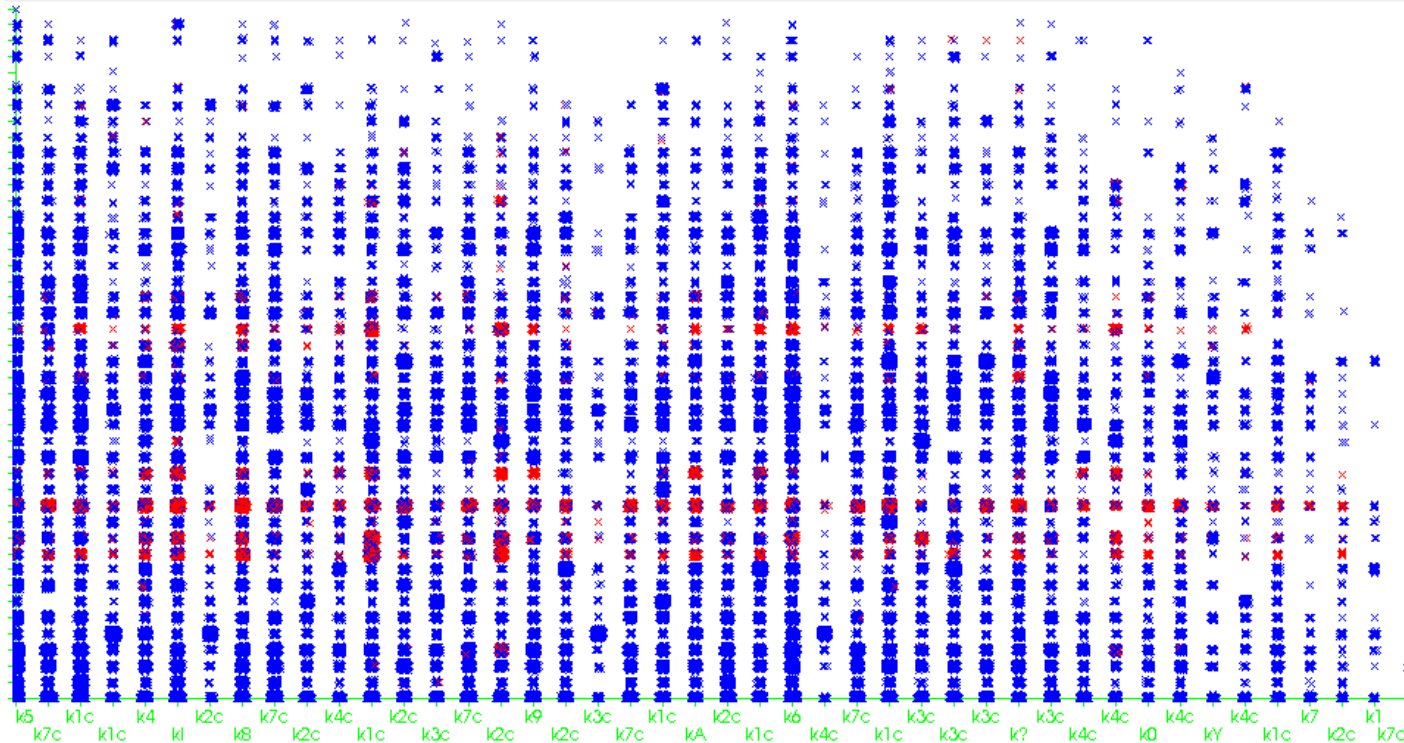k7c7: 292 339

Input examples (2)

Significant values
of attribute (word4):
k1c7, k2c7,k3c7

Significant values
of attribute (word5):
k1c7

Input examples (3)

X: word5, Y: word4

word4 $\in$ {k4c7, k4, k1c7, k3c7, k2c7} $\rightarrow$ k7c7

Results (1)

	ZeroR	J48	Id3	RF	PART	JRIp	NB	SMO	IB3	BFTree
4500	93.5	98.2	96.95	98.2	**98.3**		96.9			
1000	93.8	98.1			98.2					
100	93.8	97.5	95.1	97.6	97.7		96.8			
10	93.8	97	93.6	96.7	96.7	96.4	96.8	**97.8**	96.4	97.2

word1, **word2**, word3, **word4**, word5, **word6**, class
k1c1, **k7c2**, k1c2, **k2c7**, k1c7, **kI**, k7c7
k1c1, **kI**, k5, **k1c2**, k2c2, **k1c2**, k3c4

Results (2)

	J48	NB
4500*	97	96.2
4500**	97	96.1

* pos4, **case4**, pos5, **case5**, pos6, **case6**, class
k2, **c7**, k1, **c7**, kl, **?**, k7c7
k1, **c2**, k2, **c2**, k1, **c2**, k3c4

** **word4**, word5, **word6**, class
k2c7, k1c7, **kl**, k7c7
k1c2, k2c2, **k1c2**, k3c4

Further work

- Inductive Logic Programming
- Data stream mining
- Testing different parameter settings for the algorithms